

Original Paper

Automated Assessment of Balance Rehabilitation Exercises With a Data-Driven Scoring Model: Algorithm Development and Validation Study

Vassilios Tsakanikas^{1*}, BSc, MSc; Dimitris Gatsios^{1*}, BSc, PhD; Athanasios Pardalis^{1*}, BSc, MSc; Kostas M Tsiouris^{1*}, BSc, PhD; Eleni Georga^{1*}, BSc, PhD; Doris-Eva Bamiou^{2,3*}, BSc, MSc, PhD; Marousa Pavlou^{4*}, BSc, MSc, PhD; Christos Nikitas^{5*}, BSc, PhD; Dimitrios Kikidis^{5*}, PhD; Isabelle Walz^{6*}, BSc, MSc; Christoph Maurer^{6*}, PhD; Dimitrios Fotiadis^{1,7*}, PhD

¹Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

²Ear Institute, University College London, London, United Kingdom

³Biomedical Research Centre Hearing and Deafness, University College London Hospitals, London, United Kingdom

⁴Centre for Human and Applied Physiological Sciences, King's College London, London, United Kingdom

⁵First Department of Otolaryngology-Head and Neck Surgery, Hippokrateio General Hospital, National Kapodistrian University of Athens, Athens, Greece

⁶Department of Neurology and Neuroscience, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁷Biomedical Research Institute, Ioannina, Greece

* all authors contributed equally

Corresponding Author:

Vassilios Tsakanikas, BSc, MSc

Unit of Medical Technology and Intelligent Information Systems

Department of Materials Science and Engineering

University of Ioannina

University of Ioannina Campus

Ioannina, 45110

Greece

Phone: 30 6972745067

Email: vasilistsakanikas@gmail.com

Abstract

Background: Balance rehabilitation programs represent the most common treatments for balance disorders. Nonetheless, lack of resources and lack of highly expert physiotherapists are barriers for patients to undergo individualized rehabilitation sessions. Therefore, balance rehabilitation programs are often transferred to the home environment, with a considerable risk of the patient misperforming the exercises or failing to follow the program at all. Holobalance is a persuasive coaching system with the capacity to offer full-scale rehabilitation services at home. Holobalance involves several modules, from rehabilitation program management to augmented reality coach presentation.

Objective: The aim of this study was to design, implement, test, and evaluate a scoring model for the accurate assessment of balance rehabilitation exercises, based on data-driven techniques.

Methods: The data-driven scoring module is based on an extensive data set (approximately 1300 rehabilitation exercise sessions) collected during the Holobalance pilot study. It can be used as a training and testing data set for training machine learning (ML) models, which can infer the scoring components of all physical rehabilitation exercises. In that direction, for creating the data set, 2 independent experts monitored (in the clinic) 19 patients performing 1313 balance rehabilitation exercises and scored their performance based on a predefined scoring rubric. On the collected data, preprocessing, data cleansing, and normalization techniques were applied before deploying feature selection techniques. Finally, a wide set of ML algorithms, like random forests and neural networks, were used to identify the most suitable model for each scoring component.

Results: The results of the trained model improved the performance of the scoring module in terms of more accurate assessment of a performed exercise, when compared with a rule-based scoring model deployed at an early phase of the system (k-statistic

value of 15.9% for sitting exercises, 20.8% for standing exercises, and 26.8% for walking exercises). Finally, the resulting performance of the model resembled the threshold of the interobserver variability, enabling trustworthy usage of the scoring module in the closed-loop chain of the Holobalance coaching system.

Conclusions: The proposed set of ML models can effectively score the balance rehabilitation exercises of the Holobalance system. The models had similar accuracy in terms of Cohen kappa analysis, with interobserver variability, enabling the scoring module to infer the score of an exercise based on the collected signals from sensing devices. More specifically, for sitting exercises, the scoring model had high classification accuracy, ranging from 0.86 to 0.90. Similarly, for standing exercises, the classification accuracy ranged from 0.85 to 0.92, while for walking exercises, it ranged from 0.81 to 0.90.

Trial Registration: ClinicalTrials.gov NCT04053829; <https://clinicaltrials.gov/ct2/show/NCT04053829>

(*JMIR Rehabil Assist Technol* 2022;9(3):e37229) doi: [10.2196/37229](https://doi.org/10.2196/37229)

KEYWORDS

balance rehabilitation exercises; scoring model; exercise evaluation; persuasive coaching system

Introduction

Balance rehabilitation is essential evidence-based treatment for patients with balance disorders, especially when they are at risk of falls [1]. However, it is not feasible or economically affordable to provide patients with in-hospital sessions involving a dedicated clinician for all rehabilitation sessions required [2]. Physiotherapy health services are provided in hospitals or outpatient clinics, with assessment sessions conducted in-person by clinicians, followed by unsupervised rehabilitation sessions in the patients' homes (eg, Otago Exercise Program [3]). Research groups and published reports have shown that more than 90% of all treatments are home based [4]. According to these procedures, patients are asked to report their daily activities related to the instructed exercises and actions at home. Actual progress evaluation is performed during visits to the physician [5]. Low patient motivation and adherence to the appropriate rehabilitation exercise programs have been reported, and these consequently prolong treatment times and impose higher health care costs [6]. While various factors have been identified that contribute to low compliance, lack of continuous feedback is an important factor, and accurate monitoring of patient exercises by medical professionals in a home environment is considered essential [7,8].

A typical home-based rehabilitation exercise program (with no digital tools integrated) is based on a handbook of instructions and directions about the frequency, intensity, and correct performance of physiotherapy exercises [8]. Yet, such programs do not always ensure the full recovery of patients, as compliance rates are low [9]. In turn, activity recognition and evaluation have received increasing attention in the fields of machine learning (ML) and computer vision. Especially during the COVID-19 outbreak, the need for enhancing typical home-based rehabilitation programs with sensing devices and virtual reality interaction has substantially increased [10].

Activity recognition approaches use sensing devices to collect appropriate signals and infer the performed activity. Sensing devices vary in complexity and cost, and include video sensors, inertial measurement units, and pressure sensors. Motion analysis based on video signals explores various representations, like skeleton extraction and space-time volume. While many visual techniques have been used in recent decades, large

differences in anatomy, human occlusion, and changes in perspectives often limit the capacity of the proposed models to correctly assess the performance of an exercise. Sensing technology (apart from video) has made significant progress during the last decade, especially with low-power devices, wireless communication, high computational capacity, and data processing [11]. Wearable sensors can be integrated in clothes, strips, mobile devices, and smartwatches [12]. It is important to mention that the assessment of balance rehabilitation exercises requires accurate identification of specific movements and kinematics during the execution of the exercise (eg, head movement speed and direction, and chest flexion).

In contrast to the pure recognition of an activity, in rehabilitation programs especially, the evaluation of exercise execution is of paramount importance. This is especially significant for recovery, as it demonstrates whether the patient can perform the prescribed process [13]. During the last few years, several approaches for exercise evaluation have been proposed. In a previous study [14], a smart sensor-based rehabilitation exercise recognition and evaluation system using a deep learning framework was proposed. The main limitation was data synchronization from several sensors related to activity recognition. In similar approaches, the collected data include noise and vary when different people perform the same activity [15]. Furthermore, a state probability transition is proposed to show the transition likelihoods among states to capture the hidden states of sensory data. To test rehabilitation activities, a special matrix has been introduced, and the learned classifier has been used to identify the best features of every class at various levels. The scoring functions are given for the (0-1) range of the output values tested. To train the proposed deep neural networks in rehabilitation, the resulting movement quality scores have been used [16].

A previous study [17] proposed the hidden semi-Markov model for the assessment of rehabilitation exercises. The method extracts clinically related motion features from an RGB-D camera's skeleton and proposes an abstract representation of the subject. The effectiveness of the proposed solution has been assessed by analyzing the correlation between both a clinical evaluation and dynamic time-warping algorithms. Additionally, a previous study [18] proposed the multi-path convolutional neural network (CNN) for the recognition of rehabilitation exercises. The results of the classification accuracy in the

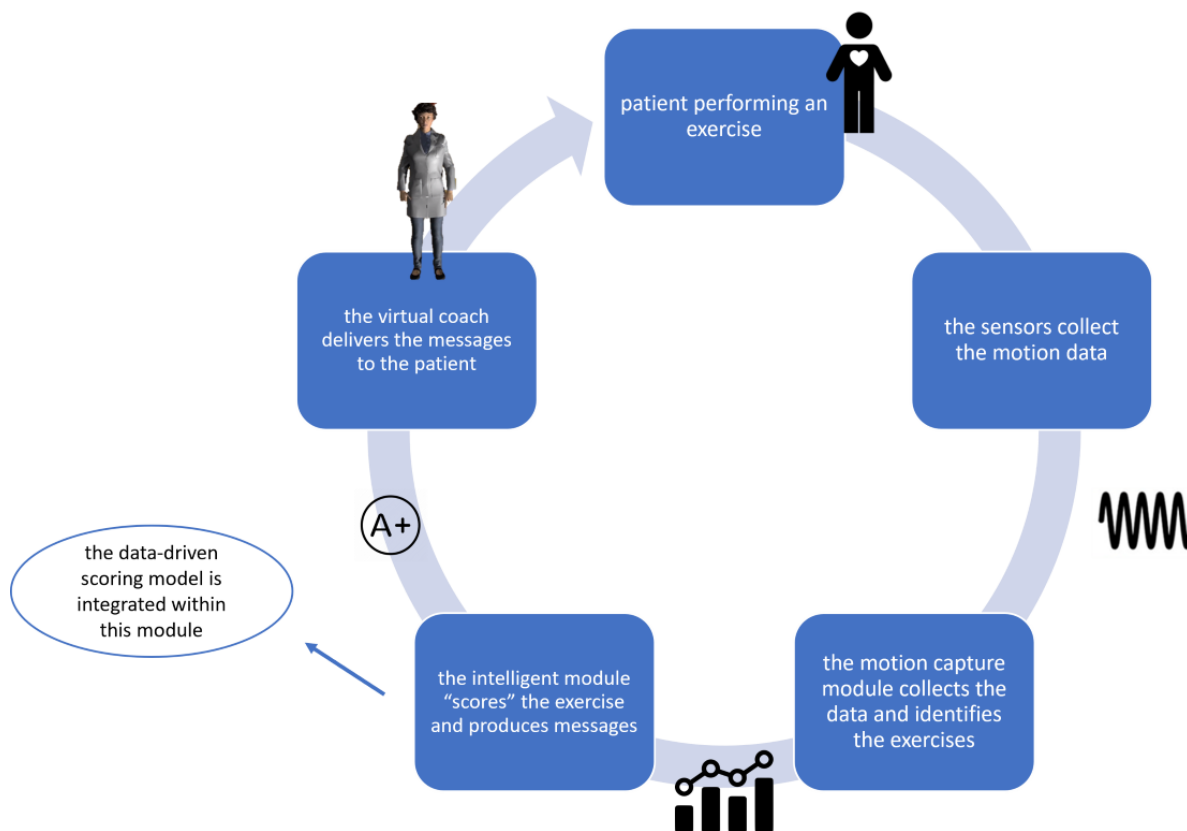
relative experiments showed that a multi-path CNN is highly efficient for sensor data acquisition. In another study [19], a deep learning-based framework for rehabilitation exercise assessment was introduced. The main modules of the system were the calculation of metrics for the quantity of motion output, the scoring of performance assessment functions for numerical motion quality ratings, and deep neural network models for quality regression of input motion through supervised learning. A previous survey [20] suggested sensor-based activity recognition by deep learning. More specifically, the survey [20] presented the recent progress in sensor-based recognition in a deep learning model, where the authors summarized the current literature (deep models and sensory techniques). Finally, a previous paper [21] assessed physical activity recognition and monitoring using Internet of Things and presented a systematic review of existing studies.

The recent development of deep learning allows high-level automated feature extraction to achieve promising performance in numerous areas [22]. Deep learning approaches for sensor-based activity recognition have been widely adopted. Further, deep learning can greatly reduce the strain on features

and can acquire much higher and meaningful features by training a neural end-to-end network. Furthermore, the deep network structure facilitates uncontrolled and incremental learning. However, compared with supervised learning approaches, deep learning models require a substantially large amount of data, which are, in general, not available in the physiotherapy domain. Thus, bearing in mind the individualities of the physiotherapy exercises, feature engineering is mandatory for each specific exercise.

In our previous work [23], we have proposed a framework for managing a balance physiotherapy program at home. This framework (Figure 1), which has been designed and developed within the Holobalance project, comprises a holographic virtual coach, presented to the patient through an augmented reality system, a motion sensing platform, and a smart engine, which assesses in real time the exercise performance. Details on the overall architecture of the system can be found elsewhere [24,25]. The technology supporting the virtual coach augmented reality module is described in several studies (eg, [26]), where information regarding augmented reality systems in rehabilitation systems can be found.

Figure 1. Virtual coaching closed-loop interaction. The proposed model is integrated into the “intelligent” module of the virtual coaching system.



The aim of this study was to design, implement, test, and evaluate a scoring model for the accurate assessment of balance rehabilitation exercises, based on data-driven techniques. More specifically, this work presents an improved model for the offline scoring function, which is not based on the knowledge-based model that was used previously [23], but is based on a data-driven model with the capacity to predict with higher accuracy the score of a performed exercise. As it is of paramount importance for a closed-loop persuasive system to

correctly evaluate the performance of an exercise, the proposed scoring model is expected to provide more robust and reliable feedback to the overall system’s reasoning engine.

Methods

Ethics Approval

This study has received institutional ethics approvals in Germany/Freiburg (reference: 265/2019) and Greece/Athens (reference: 9769/24-6-2019).

Study Design

A pilot study with 20 participants was conducted with the aim to collect the appropriate data set to develop the scoring model. After 1 dropout, 19 patients followed an 8-week balance rehabilitation program, according to the protocol described previously [27] at 2 pilot sites. Participants were elderly individuals who had experienced at least one fall during the last year. They were all informed about the context of the study and volunteered to participate, after providing their written consent regarding the willingness to use the Holobalance system in the

clinic and to have their data recorded and used for research purposes.

While the Holobalance system is designed for home use, it was installed in a clinic setup to test safety and to collect the necessary data. After recruitment of the patients, functional and cognitive assessments were performed based on the Mini-Balance Evaluation Systems Test (MINIBEST), Functional Gait Assessment (FGA), Falls Efficacy Scale International (FES-I), Montreal Cognitive Assessment (MoCA), World Health Organization Disability Assessment Schedule (WHODAS), and Activities-Specific Balance Confidence Scale (ABC), as per the clinical study protocol [27]. It is important to mention that while both the FES-I and ABC attempt to infer similar information about the patient, their outputs are not fully correlated [28]. Demographic data as well as the distribution of the tests are presented in Table 1. According to FGA results, the population of this study had mild cognitive impairment [1].

Table 1. Study participant details.

Variable	Pilot site		Total value
	Athens	Freiburg	
Participants, n	14	5	19
Age (years), median (IQR)	64.5 (15.5)	72.0 (4.0)	68.0 (11.0)
Height (cm), median (IQR)	157.5 (11.8)	170.0 (2.0)	160.0 (16.5)
Weight (kg), median (IQR)	67.0 (21.5)	69.0 (8.0)	69.0 (21.0)
Male gender, %	7.14	40.00	15.79
Mini-Balance Evaluation Systems Test score (range ^a 0-28), median (IQR)	21.5 (6.0)	21.0 (1.0)	21.0 (5.5)
Functional Gait Assessment score (range ^a 0-30), median (IQR)	21.0 (5.0)	22.0 (3.0)	21.0 (5.5)
Falls Efficacy Scale International score (range ^a 16-64), median (IQR)	27.5 (9.25)	19.0 (8.0)	27.0 (8.5)
Montreal Cognitive Assessment score (range ^a 0-30), median (IQR)	25.5 (3.75)	27.0 (4.0)	26.0 (4.0)
World Health Organization Disability Assessment Schedule score (range ^a 100-0), median (IQR)	23.0 (24.5)	17.0 (21.0)	17.0 (22.0)
Activities-Specific Balance Confidence Scale score (range ^a 0-100), median (IQR)	76.9 (20.3)	87.5 (15.0)	82.5 (19.9)

^aFor the score range a-b, “a” represents no disability and “b” represents the highest disability.

Data Set

The participants, following the balance rehabilitation program prescribed by their physicians, performed a set of exercises during 16 sessions (2 sessions per week). During each session, a set of exercises was performed according to the program. The number of exercises per session varied from 3 to 8. Participants were instructed to execute the exercises at a self-paced rate (frequency and velocity of the movements) that would make them feel comfortable, avoiding any symptoms. As the sessions progressed, the aim of the program was to increase these metrics.

The performed exercises (with the relative progression levels for each exercise), which are described in a previous paper [27], were grouped into 9 classes, according to the kinematic characteristics of each exercise. The rehabilitation protocol

included 3 types of exercises (sitting exercises, standing exercises, and walking exercises). More specifically, there were 3 sitting exercises with 3 progression levels (in terms of intensity and complexity), 4 standing exercises with 4 progression levels, and 3 walking exercises with 3 progression levels (Table 2). The exercises were designed under the rationale of progressiveness of difficulty, including both simple and complex tasks, aiming for head-eye-hand coordination through multisensory rehabilitation exercises. As reported previously [29], the system is acceptable by end users and is feasible for use in hospital and home environments.

The data set was collected from April 2020 to June 2021. In total, 1313 exercises were recorded. Table 3 summarizes the collected annotated exercises.

Table 2. Description of the available rehabilitation exercises offered within the Holobalance intervention protocol (adapted from Liston et al [27], which is published under Creative Commons Attribution 4.0 International License [30]).

Exercise type	Exercise description
Sitting 1: Yaw	Perform head rotations of 30 degrees in the yaw plane (ie, left-right) while sitting, aiming at enhancing gaze stability.
Sitting 2: Pitch	Perform head rotations of 30 degrees in the pitch plane (ie, up-down) while sitting, aiming at enhancing gaze stability and improving common vestibular symptoms such as dizziness, swimminess, and light-headedness.
Sitting 3: Bend over	Bend as if to pick up an object off the floor from the sitting position and return to the upright position, aiming at improving functional activities of daily living (ADL) tasks and mitigating vestibular symptoms if provoked through practice.
Standing 1: Maintain balance	Maintain balance while standing up and remain in the proper position, aiming at improving postural alignment and standing ability with a smaller base of support.
Standing 2: Maintain balance on foam	Maintain balance as in standing exercise 1 while standing on a cushion and remain in the proper position, aiming at promoting sensory reweighting.
Standing 3: Bend over and reach up	Bend over bringing the chin to the chest, return the head to the normal upright position on coming up, and reach up while slightly tilting the head back, aiming at improving functional ADL tasks and dizziness.
Standing 4: Turn	On site, turn to face the opposite direction (ie, 180° turn), aiming at improving functional ADL tasks and dizziness.
Walking 1: Walk to horizon	Walk across the room (back and forth) in a straight path while looking at the horizon, aiming at promoting a normal gait pattern. Minimum space of 2 meters.
Walking 2: Walk & yaw	Walk across the room (back and forth) in a straight path while turning the head left and right, aiming at improving gaze stability while walking and functional ADL walking tasks. Minimum space of 2 meters. Yaw movement as in sitting exercise 1.
Walking 3: Walk & pitch/V-shape	Walk across the room (back and forth) in a straight path while turning the head up and down, and with V-shaped movement, aiming at improving gaze stability while walking and functional ADL walking tasks. Minimum space of 2 meters. Yaw and pitch movements as in sitting exercises 1 and 2.

Table 3. Exercises according to the type and progression level (N=1313).

Exercise type	Value, n	Exercise progression
Sitting exercise	514	
Sitting exercises 1 and 2	347	All progression levels
Sitting exercise 3	167	All progression levels
Standing exercise	530	
Standing exercises 1 and 2	312	All progression levels
Standing exercise 3	97	Progression levels 0 and 1 included 46; progression level 2 included 19; progression level 3 included 32
Standing exercise 4	121	All progression levels
Walking exercise	269	
Walking exercise 1	87	All progression levels
Walking exercises 2 and 3	182	All progression levels

During the execution of the exercises, a physiotherapist monitored the patient and scored patient performance using a scoring rubric that included 4 components (frequency, amplitude, velocity, and symmetry) for the sitting and standing exercises and an additional component (gait quality) for the walking exercises. For exercises with complex kinematic characteristics, additional components were considered in the scoring. For example, if an exercise included movement of the head and walking, rubric components for head movement and for gait quality were included in the scoring process.

More specifically, for sitting exercises, frequency referred to the number of head rotations (eg, in the yaw plane for sitting exercise 1) per second, while amplitude referred to the degree

of head turn from the upfront position to the extreme points of the movement. Additionally, velocity referred to the number of seconds a patient needed to perform a movement. This metric differs from frequency, as patients usually paused for some seconds between exercise movements, especially for complex ones like sitting exercise 3.

For each component, a score from 0 to 3 was given, with a score of 0 representing the noncompletion of the exercise. On top of the rubric components, a total score for each exercise was calculated a posteriori as the average of all components (N) of an exercise.

$$Total\ score = \left\| \frac{1}{N} \sum_{i=1}^n \|rubric_component_i\| \right\| \quad (1)$$

The proposed scoring model infers the score for all the involved components of an exercise, as well as the total score, which is mainly required to provide input to adjacent modules of the persuasive coaching system.

All patients undertook training sessions to get familiarized with the system. In addition, the session physiotherapists provided specific instructions for the correct execution of the exercises to the patients, in terms of timing and kinesiology. As described previously [23], these instructions were used to create the knowledge-based scoring model of the system.

A subset of the data set described in Table 3 was annotated by 2 physiotherapists, who monitored the patients during the execution of the exercises. More specifically, 38 sessions from 4 patients, which included 90 sitting exercises, 78 standing exercises, and 59 walking exercises, were scored by 2 independent evaluators to assess the interobserver variability of the annotation process. This resulted in 665 annotated scores for the different components of the scoring rubric.

Metrics and Analytics

As presented previously [23], based on a set of sensing devices (Figure 2), the system collected temporal signals and processed them by extracting specific kinematic metrics, which were translated to exercise analytics. These analytics, along with the knowledge-based scoring model presented previously [23], were used as features in the ML models used to constitute the scoring model. Table 4 summarizes the extracted features, which were used as inputs for the ML models. The build prototype of the home-based system, including all the sensing devices, the head-mounted display, and the processing unit, costs approximately €4800 (US \$4850) (Figure 2).

The knowledge-based exercise score model (kb_score), mentioned in Table 4, refers to a rule-based model that attempts to assess the performance of an exercise based on the values of the captured motion analytics. More specifically, a group of experts established the acceptable range for each of the motion analytics (eg, 30 degrees for the head movement in sitting exercise 1). Based on these ranges, the knowledge-based model calculates the proportion of time a patient performs within these ranges, as well as how close the patient comes to the optimal range, and outputs the final kb_score. For assessing balance, sway, and stability, posture and trunk_sway metrics (Table 4) have been used.

Figure 2. The Holobalance system. (A) Sensor positioning in the Holobalance system. (B) Devices of the Holobalance system. IMU: inertial measurement unit.

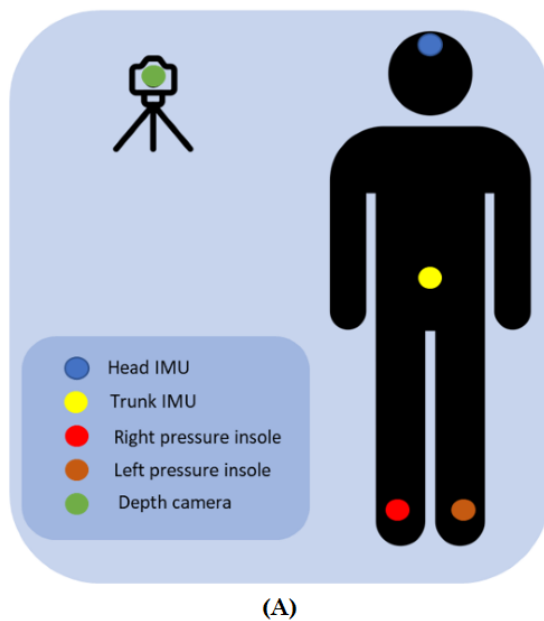


Table 4. Input features for training the machine learning models.

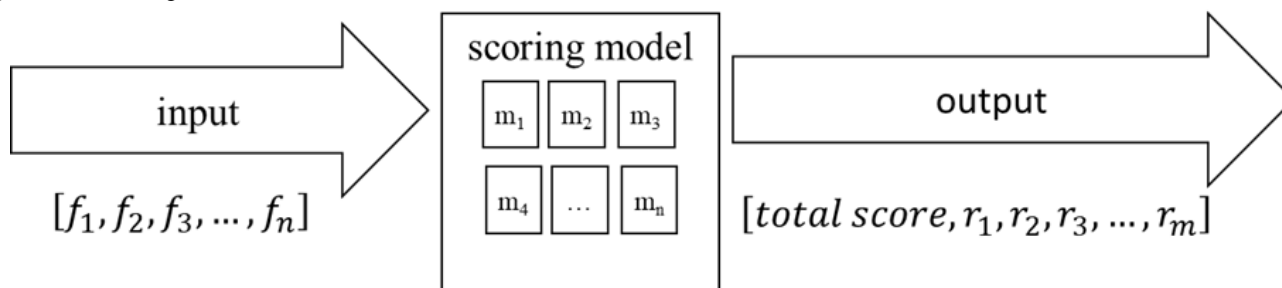
Feature	Description
kb_score	Knowledge-based exercise score as proposed previously [27]
head_movement_speed	Number of head rotations per second (mean and standard deviation) in the yaw and pitch planes
head_movement_range	Range of head rotations (mean and standard deviation) in the yaw and pitch planes
posture	Angle of the torso (sitting and standing)
trunk_sway	Mean and standard deviation of trunk sway
gait_parameters	Center of pressure on both feet (mean distance covered by the center of pressure and standard deviation per gait cycle); double support time (mean value and standard deviation per gait cycle); single support time (mean value and standard deviation per gait cycle); step duration (mean value and standard deviation per gait cycle); stride duration (mean value and standard deviation per gait cycle); cadence (mean value and standard deviation per gait cycle)

Scoring Model

The proposed data-driven exercise scoring model uses as inputs the analytics described in Table 4 and outputs a scoring vector for each exercise, as presented in Figure 3. More specifically, f_i refers to the features that describe the motion and movement of a patient during the performance of an exercise, while r_i refers

to each one of the evaluation components (frequency, amplitude, velocity, and symmetry), as expressed in each different exercise. Finally, *total score* refers to an overall assessment of the exercise. As the importance of the input features varies for the different exercise categories (Table 3), a separate model for each one of these groups of exercises and progressions has been developed and incorporated in the final scoring model.

Figure 3. The scoring model.



Aiming to identify the most relevant ML model for each rubric component (and for the total score), a set of ML models was assessed for each one of the components. The considered models were k-nearest neighbors (kNN) [31], support vector machines (SVMs) [32] (with both lineal and radial basis function), Gaussian process [33], random forests [34], neural networks [22], naïve Bayes [35], and AdaBoost [36]. These specific models were selected as they have been used in a wide set of similar data-driven problems [37].

For standing exercise 3, it was required to consider different models for different progressions owing to different kinematic characteristics in its progressions. This resulted in relatively small data sets for these cases. For this, the SMOTE (synthetic minority oversampling technique) algorithm [38] was used to

oversample the collected instances in order to obtain the necessary data to train the ML models.

The approach followed during the training of the ML models is summarized in Figure 4. More specifically, the first step was to identify data inconsistencies, like missing values, and remove them from the data set. Afterwards, min-max feature normalization was applied, aiming to improve the training process of the ML models. The next step involved an iterative process of training different ML models and evaluating them. For each model, an intermediate step for fine-tuning each parameter was applied, mainly using the grid search approach. Finally, the winning classifier for each model was selected, based on F1-score and receiver operating characteristic analysis results.

Figure 4. Machine learning (ML) model training approach. kNN: k-nearest neighbors; ROC: receiver operating characteristic; SVM: support vector machine.



Deployment Details: Integration

The winning classifiers were implemented under Python 3.8, using the scikit-learn 0.24 library. As soon as the system identifies the performed exercise, the appropriate classifier is invoked and the score of the exercise is inferred. This is now part of the Holobalance system, which is currently under evaluation.

Results

Overview

Within this section, the results of the training and evaluation of the ML models for each component of the scoring rubric are presented. All models were evaluated by applying a 10-fold cross-validation process and assessing the macro-average accuracy of the models. The training and testing data sets for each fold were created under an 80/20 ratio.

Interobserver Variability

As already mentioned earlier, almost 17.3% of the recorded exercises were scored by 2 observers to assess the interobserver

variability of the annotation process. The results of this procedure are presented in Table 5. The selected evaluation metric is Cohen kappa coefficient [39], which is calculated as follows:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (2)$$

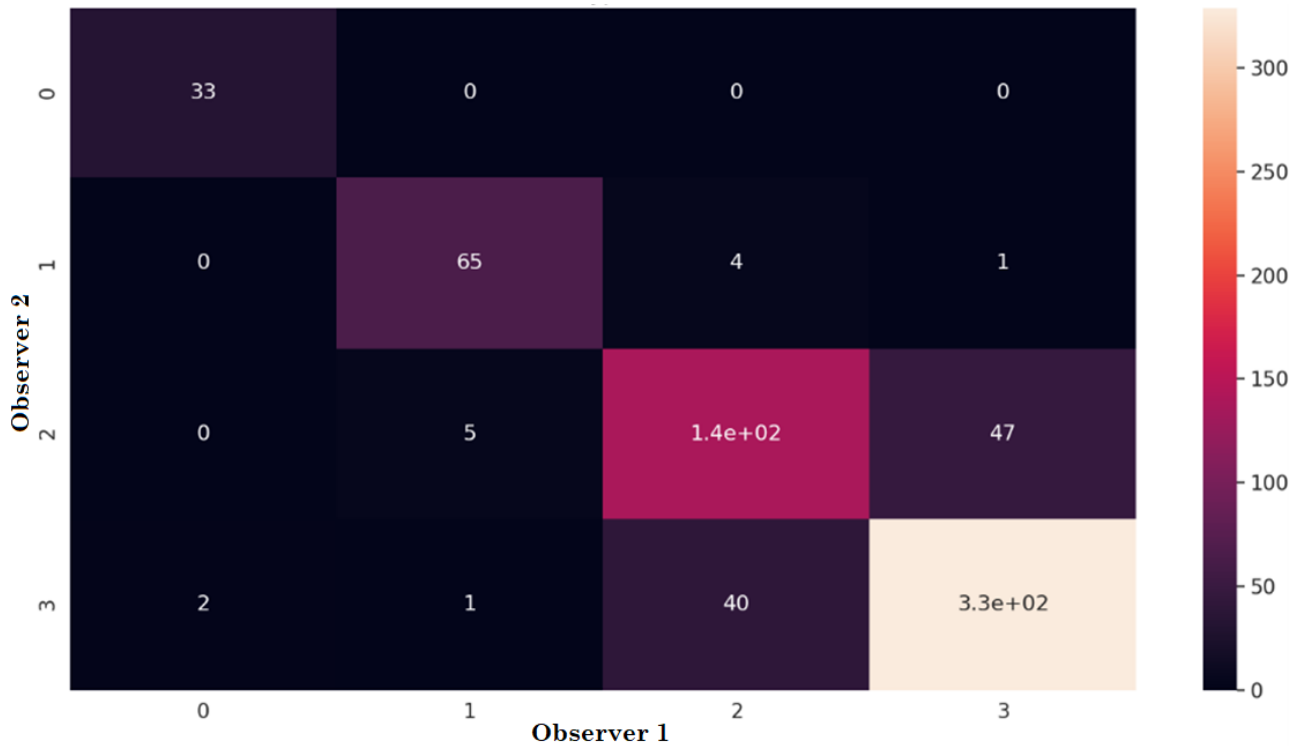
where $\Pr(a)$ is the relative observed agreement among raters and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probability of each observer randomly seeing each category. If the raters are in complete agreement, then $k=1$. If there is no agreement between the raters other than what would be expected by chance (as given by $\Pr(e)$), then $k=0$.

From a previous study [40], it can be concluded that the agreement of the observers was “good,” allowing the use of the collected data set to train reliable ML models. Figure 5 presents the confusion matrix of the annotation process (please see Multimedia Appendix 1 for more details).

Table 5. Results of interobserver variability per exercise type.

Exercise type	k statistic
All exercises	0.75
Sitting exercises	0.68
Standing exercises	0.79
Walking exercises	0.75

Figure 5. Confusion matrix. All types of exercises (N=665) in the annotation process of 2 observers.



Classification Results of Each Model

As mentioned earlier, an ML model for each component of the scoring rubric was trained and evaluated. The results are presented in Table 6, where the macro-average accuracy has been provided, along with the winning classifier for each model. The results below present a set of 40 trained classifiers, which finally constitute the system’s scoring model. More detailed results for the classification models are presented in Multimedia Appendix 1.

For the sitting and standing exercises, it can be observed that the Gaussian process is the most relevant classifier, most probably because the number of features was lower compared with that for the walking exercises. Additionally, the low number of input features was correlated with higher accuracy results, which was expected. Thus, the accuracy for sitting exercises 1 and 2 was almost 90%, while that for walking exercises 2 and 3 dropped to slightly higher than 80% (Table 6). Finally, for the total score, the random forest classifier outperformed the rest of the models for 2 exercise subgroups, while kNN and linear SVM outperformed for 1 subgroup.

Table 6. Macro accuracy results of the winning classifiers for each of the considered models.

Exercise type	Macro accuracy/winning classifier						
	Total score	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
Sitting 1 and sitting 2	0.90/Gaussian process	0.88/Gaussian process	0.90/kNN ^a	0.89/Gaussian process	N/A ^b	N/A	N/A
Sitting 3	0.87/Gaussian process	0.86/Neural network	0.91/Gaussian process	N/A	N/A	N/A	N/A
Standing 1 and standing 2	0.85/Gaussian process	0.83/Gaussian process	0.86/Gaussian process	N/A	N/A	N/A	N/A
Standing 3 (progressions 0-1)	0.91/kNN	0.91/Gaussian process	0.92/Gaussian process	0.89/kNN	0.90/Random forest	N/A	N/A
Standing 3 (progression 2)	0.87/SVM ^c (linear)	0.89/Gaussian process	0.90/Naive Bayes	0.88/Random forest	0.91/kNN	N/A	N/A
Standing 3 (progression 3)	0.91/Random forest	0.90/AdaBoost	0.88/Neural network	0.86/kNN	0.89/kNN	N/A	N/A
Standing 4	0.92/Gaussian process	0.86/Gaussian process	0.88/Gaussian process	0.80/kNN	N/A	N/A	N/A
Walking 1	0.90/Random forest	0.81/Gaussian process	0.85/Random forest	0.92/Random forest	N/A	N/A	N/A
Walking 2 and walking 3	0.81/kNN	0.74/kNN	0.75/SVM (linear)	0.78/SVM (RBF ^d)	0.71/kNN	0.75/SVM (RBF)	0.75/kNN

^akNN: k-nearest neighbors.

^bN/A: not applicable.

^cSVM: support vector machine.

^dRBF: radial basis function.

Overall Results: k-Statistic Analysis

Table 7 presents the overall results of the classification models for each individual exercise and the progression levels. In the same table, comparisons of interobserver variability, and the variability among observer 1 and the trained ML models are provided, which were performed on the testing data sets of each model. In addition, the previously used knowledge-based model [23] was compared with the annotations of the first observer.

Based on the results, the proposed framework's performance was similar to interobserver variability, thus constituting a reliable model for automated scoring of balance physiotherapy exercises. More specifically, the variability for the sitting exercises was almost identical, while there was a drop of 0.02

for the standing exercises. Finally, for the walking exercises, the decrease in the k-statistic was 0.04, which was justified due to the increased complexity of the relative exercises and the increased input features for the classification problems in these specific exercises.

When compared with the knowledge-based scoring model, the improvement in the agreement was substantial (15.9% for sitting exercises, 20.8% for standing exercises, and 26.8% for walking exercises for the k-statistic). This improvement enables the system to effectively deduce the performance of the patient, and thus, the system can not only correctly inform the clinician about the patient's status, but also enable them to design/choose correctly future rehabilitation programs.

Table 7. Overall classification accuracy and k-statistic analysis.

Exercise type	Total score (model)	k statistic (interobserver variability)	k statistic (observer 1 – ML ^a model)	k statistic (observer 1 – knowledge-based model)
Sitting		0.68	0.69	0.58
Sitting exercises 1 and 2	0.90 (Gaussian process)			
Sitting exercise 3	0.86 (Gaussian process)			
Standing		0.79	0.77	0.61
Standing exercises 1 and 2	0.853 (Gaussian process)			
Standing exercise 3 (progression level 0-1)	0.912 (kNN ^b)			
Standing exercise 3 (progression level 2)	0.8736 (SVM ^c linear)			
Standing exercise 3 (progression level 3)	0.905 (random forest)			
Standing exercise 4	0.918 (Gaussian process)			
Walking		0.75	0.71	0.52
Walking exercise 1	0.899 (random forest)			
Walking exercises 2 and 3	0.813 (kNN)			

^aML: machine learning.

^bkNN: k-nearest neighbors.

^cSVM: support vector machine.

Discussion

Principal Findings

The proposed set of ML models can effectively score the balance rehabilitation exercises of the Holobalance system. The models had similar accuracy in terms of Cohen kappa analysis, with interobserver variability, enabling the scoring module to infer the score of an exercise based on the collected signals from sensing devices. More specifically, for the sitting exercises, the scoring model had high classification accuracy, ranging from 0.86 to 0.90. Similarly, for the standing exercises, the classification accuracy ranged from 0.85 to 0.92, while for the walking exercises, it ranged from 0.81 to 0.90. From the obtained results, we observed that the lowest classification accuracies were related to the most complex exercises, in terms of required movements. While this result was anticipated, it is interesting that the same exercises also presented the highest interobserver variability, revealing that objectively scoring a complicated exercise is not a trivial task, even for expert physiotherapists. This is clearly reflected by the k-statistic analysis for almost all different exercise types. It is also important to mention that most of the misclassifications involved classes 2 and 3, meaning that poor performance (classes 0 and 1) and adequate performance (classes 2 and 3) can be assessed more accurately, by both the experts and the scoring model.

Comparison With Prior Work

The first version of the scoring module was built upon medical knowledge extracted by a group of experts [27]. The main drawback of this model was that it could not capture all possible states of a patient during the execution of a balance rehabilitation exercise. Thus, it failed in various situations to correctly grade the patient. The proposed data-driven model significantly

improves the accuracy for the performed exercises, increasing the k-statistic by 0.11 for sitting exercises, 0.16 for standing exercises, and 0.19 for walking exercises. It was noticeable that a more complex exercise was associated with higher improvement.

Strengths

The novelty of this work can be summarized in 2 main remarks. First, an annotated data set of sensor signals during the performance of about 1300 exercise sessions from 19 patients, along with the scoring of the exercises from an expert, was created. To the best of our knowledge, no such data set has been reported in the literature. Second, a scoring module, which includes several ML-supervised learning models, was developed and tested. The results clearly indicate that the proposed model appears to have similar predicting capacity considering the interobserver variability of experts who annotated the ground-truth data set.

Within the context of the Holobalance system, the capacity of the scoring module obviously enables correct exercise assessment in a rehabilitation program, as a physician can monitor the performance and progress of a patient and adopt the program accordingly. This assessment has a 2-fold advantage. First, the physiotherapist managing the patient is properly informed about the performance of the patient; thus, the next rehabilitation phases are designed based on objective information, which avoids the bias of self-reported results. Second, the virtual coach interaction with the patient is based on accurate scores, which facilitates realistic interaction with the system. More specifically, the exercise progression module is based on the scores produced by the scoring module to correctly assess whether a patient should progress to the next level of an exercise. As discussed earlier, each exercise is administered at different levels in terms of difficulty, speed,

and repetitions. Hence, the high accuracy of the scoring module enables the proper function of the exercise progression module. Additionally, the scoring module can be used for “red flagging” patients with very low performance and adherence early, thus allowing the physiotherapist to alter the rehabilitation approach. These aspects have a direct impact on the safe and effective execution of rehabilitation programs in home environments.

It is also important to stress that compared with other scoring models (eg, [41] and [42]), the output of the proposed model assesses not the recognition of the performed exercise but the quality of the performance of the exercise, a crucial aspect in the assessment of a rehabilitation program. By providing a high-accuracy exercise assessment model, as the one presented, virtual coaching systems can be equipped with the capacity to interact with patients using personalized context, thus enriching user experience.

Besides the value of a reliable scoring module within a persuasive coaching system like Holobalance, this module can be used independently as a separate module in clinical practice. One of the most important uses is objective baseline assessment of a patient, as it can support clinicians in objectively evaluating the baseline of a patient when performing an exercise during the first clinic visit. Additionally, the analysis for building the scoring module, especially the feature statistics analysis, can contribute to the design of new balance rehabilitation exercises targeting mainly the metrics that appear to have an important contribution to the score of an exercise, while eliminating aspects and kinematics related to metrics of low importance to the model. Furthermore, the scoring module can support patients who require long-term monitoring, especially those with

degenerative neurological conditions, such as ataxia or dementia, which require long-term rehabilitation and monitoring for maintenance purposes. Moreover, a reliable scoring and assessment module can facilitate the education of novice physiotherapists and physicians, enabling them to better understand the needs of different clinical populations. Finally, within the research context, the sensor-based information from this model could be used as a biomarker to monitor populations of interest over the long term (such as older adults or patients with cognitive impairments) for the early prediction of the risk of falls and early prediction of cognitive decline.

Limitations

Regarding the limitations of the proposed model, a major drawback is that the model requires knowledge of the type of exercise to assess the score for the exercise. In other words, the proposed scoring model does not have the capacity to recognize the exercise, limiting its usage to only rehabilitation programs with predefined exercise sets. Additionally, the size of the collected data set did not allow us to test deep learning models, which might show higher classification accuracies.

Future Directions

Regarding the future directions related to the scoring model, we anticipate to incorporate motion recognition algorithms, enabling the module to infer which exercise is performed. This will allow the module to support free-program exercise sessions. Finally, deploying the module to more sites will allow us to extend the exercise data set, which will provide wider validation to the proposed solution and help in the use of deep learning models, if the volume of data is adequate.

Acknowledgments

This work has received funding from the European Union’s Horizon 2020 research and innovation program (grant agreement no. 769574).

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional study results.

[\[DOCX File, 873 KB-Multimedia Appendix 1\]](#)

References

1. Bevilacqua A, Brennan L, Argent R, Caulfield B, Kechadi T. Rehabilitation Exercise Segmentation for Autonomous Biofeedback Systems with ConvFSM. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 574-579. [doi: [10.1109/embc.2019.8856428](https://doi.org/10.1109/embc.2019.8856428)]
2. Escalona F, Martínez-Martin E, Cruz E, Cazorla M, Gomez-Donoso F. EVA: EVALuating at-home rehabilitation exercises using augmented reality and low-cost sensors. *Virtual Reality* 2019 Dec 17;24(4):567-581. [doi: [10.1007/s10055-019-00419-4](https://doi.org/10.1007/s10055-019-00419-4)]
3. Shubert TE, Smith M, Goto L, Jiang L, Ory M. Otago exercise program in the United States: Comparison of 2 implementation models. *Phys Ther* 2017 Feb 01;97(2):187-197. [doi: [10.2522/ptj.20160236](https://doi.org/10.2522/ptj.20160236)] [Medline: [28204770](https://pubmed.ncbi.nlm.nih.gov/28204770/)]
4. Kertész C. Physiotherapy Exercises Recognition Based on RGB-D Human Skeleton Models. 2013 Presented at: 2013 European Modelling Symposium; November 20-22, 2013; Manchester, UK p. 21-29. [doi: [10.1109/ems.2013.4](https://doi.org/10.1109/ems.2013.4)]

5. Bonnechère B, Sholukha V, Omelina L, Van Sint Jan S, Jansen B. 3D analysis of upper limbs motion during rehabilitation exercises using the Kinect sensor: Development, laboratory validation and clinical application. *Sensors (Basel)* 2018 Jul 10;18(7):2216 [FREE Full text] [doi: [10.3390/s18072216](https://doi.org/10.3390/s18072216)] [Medline: [29996533](https://pubmed.ncbi.nlm.nih.gov/29996533/)]
6. Gu F, Khoshelham K, Valaee S, Shang J, Zhang R. Locomotion activity recognition using stacked denoising autoencoders. *IEEE Internet Things J* 2018 Jun;5(3):2085-2093. [doi: [10.1109/jiot.2018.2823084](https://doi.org/10.1109/jiot.2018.2823084)]
7. Taewong Um T, Babakeshizadeh V, Kulić D. Exercise motion classification from large-scale wearable sensor data using convolutional neural networks. 2017 Presented at: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September 24-28, 2017; Vancouver, BC, Canada p. 2385-2390. [doi: [10.1109/iros.2017.8206051](https://doi.org/10.1109/iros.2017.8206051)]
8. Khan MH, Helsper J, Boukhers Z, Grzegorzec M. Automatic recognition of movement patterns in the vojta-therapy using RGB-D data. 2016 Presented at: 2016 IEEE International Conference on Image Processing (ICIP); September 25-28, 2016; Phoenix, AZ, USA p. 1235-1239. [doi: [10.1109/icip.2016.7532555](https://doi.org/10.1109/icip.2016.7532555)]
9. Pomeroy V, Aglioti SM, Mark VW, McFarland D, Stinear C, Wolf SL, et al. Neurological principles and rehabilitation of action disorders: rehabilitation interventions. *Neurorehabil Neural Repair* 2011 Jun 25;25(5 Suppl):33S-43S [FREE Full text] [doi: [10.1177/1545968311410942](https://doi.org/10.1177/1545968311410942)] [Medline: [21613536](https://pubmed.ncbi.nlm.nih.gov/21613536/)]
10. Seron P, Oliveros M, Gutierrez-Arias R, Fuentes-Aspe R, Torres-Castro RC, Merino-Osorio C, et al. Effectiveness of telerehabilitation in physical therapy: A rapid overview. *Phys Ther* 2021 Jun 01;101(6):pzab053 [FREE Full text] [doi: [10.1093/ptj/pzab053](https://doi.org/10.1093/ptj/pzab053)] [Medline: [33561280](https://pubmed.ncbi.nlm.nih.gov/33561280/)]
11. Vakanski A, Ferguson JM, Lee S. Metrics for performance evaluation of patient exercises during physical therapy. *Int J Phys Med Rehabil* 2017 Jun;5(3):403 [FREE Full text] [doi: [10.4172/2329-9096.1000403](https://doi.org/10.4172/2329-9096.1000403)] [Medline: [28752104](https://pubmed.ncbi.nlm.nih.gov/28752104/)]
12. Nweke HF, Teh YW, Al-garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications* 2018 Sep;105:233-261 [FREE Full text] [doi: [10.1016/j.eswa.2018.03.056](https://doi.org/10.1016/j.eswa.2018.03.056)]
13. O'Reilly M, Caulfield B, Ward T, Johnston W, Doherty C. Wearable inertial sensor systems for lower limb exercise detection and evaluation: A systematic review. *Sports Med* 2018 May 24;48(5):1221-1246. [doi: [10.1007/s40279-018-0878-4](https://doi.org/10.1007/s40279-018-0878-4)] [Medline: [29476427](https://pubmed.ncbi.nlm.nih.gov/29476427/)]
14. Zhang W, Su C, He C. Rehabilitation exercise recognition and evaluation based on smart sensors with deep learning framework. *IEEE Access* 2020;8:77561-77571. [doi: [10.1109/access.2020.2989128](https://doi.org/10.1109/access.2020.2989128)]
15. Fuentes D, Gonzalez-Abril L, Angulo C, Ortega J. Online motion recognition using an accelerometer in a mobile device. *Expert Systems with Applications* 2012 Feb;39(3):2461-2465. [doi: [10.1016/j.eswa.2011.08.098](https://doi.org/10.1016/j.eswa.2011.08.098)]
16. 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 May;70:211-217. [doi: [10.1016/j.gaitpost.2019.03.008](https://doi.org/10.1016/j.gaitpost.2019.03.008)] [Medline: [30903993](https://pubmed.ncbi.nlm.nih.gov/30903993/)]
17. Capecchi M, Ceravolo MG, Ferracuti F, Iarlori S, Kyrki V, Monteriù A, et al. A Hidden Semi-Markov Model based approach for rehabilitation exercise assessment. *J Biomed Inform* 2018 Feb;78:1-11 [FREE Full text] [doi: [10.1016/j.jbi.2017.12.012](https://doi.org/10.1016/j.jbi.2017.12.012)] [Medline: [29277330](https://pubmed.ncbi.nlm.nih.gov/29277330/)]
18. Zhu Z, Lu Y, You C, Chiang C. Deep learning for sensor-based rehabilitation exercise recognition and evaluation. *Sensors (Basel)* 2019 Feb 20;19(4):887 [FREE Full text] [doi: [10.3390/s19040887](https://doi.org/10.3390/s19040887)] [Medline: [30791648](https://pubmed.ncbi.nlm.nih.gov/30791648/)]
19. Liao Y, Vakanski A, Xian M. A deep learning framework for assessing physical rehabilitation exercises. *IEEE Trans Neural Syst Rehabil Eng* 2020 Feb;28(2):468-477. [doi: [10.1109/tnsre.2020.2966249](https://doi.org/10.1109/tnsre.2020.2966249)]
20. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* 2019 Mar;119:3-11. [doi: [10.1016/j.patrec.2018.02.010](https://doi.org/10.1016/j.patrec.2018.02.010)]
21. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y. Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review. *J Biomed Inform* 2018 Nov;87:138-153 [FREE Full text] [doi: [10.1016/j.jbi.2018.09.002](https://doi.org/10.1016/j.jbi.2018.09.002)] [Medline: [30267895](https://pubmed.ncbi.nlm.nih.gov/30267895/)]
22. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018 Nov;4(11):e00938 [FREE Full text] [doi: [10.1016/j.heliyon.2018.e00938](https://doi.org/10.1016/j.heliyon.2018.e00938)] [Medline: [30519653](https://pubmed.ncbi.nlm.nih.gov/30519653/)]
23. Tsakanikas VD, Gatsios D, Dimopoulos D, Pardalis A, Pavlou M, Liston MB, et al. Evaluating the performance of balance physiotherapy exercises using a sensory platform: The basis for a persuasive balance rehabilitation virtual coaching system. *Front Digit Health* 2020 Nov 27;2:545885 [FREE Full text] [doi: [10.3389/fdgh.2020.545885](https://doi.org/10.3389/fdgh.2020.545885)] [Medline: [34713032](https://pubmed.ncbi.nlm.nih.gov/34713032/)]
24. Kouris I, Sarafidis M, Androutsou T, Koutsouris D. HOLOBALANCE: An Augmented Reality virtual trainer solution for balance training and fall prevention. 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 18-21, 2018; Honolulu, HI, USA p. 4233-4236. [doi: [10.1109/embc.2018.8513357](https://doi.org/10.1109/embc.2018.8513357)]
25. Tsiouris KM, Gatsios D, Tsakanikas V, Pardalis AA, Kouris I, Androutsou T, et al. Designing interoperable telehealth platforms: bridging IoT devices with cloud infrastructures. *Enterprise Information Systems* 2020 Apr 30;14(8):1194-1218. [doi: [10.1080/17517575.2020.1759146](https://doi.org/10.1080/17517575.2020.1759146)]
26. Mostajeran F, Steinicke F, Ariza NO, Gatsios D, Fotiadis D. Augmented Reality for Older Adults: Exploring Acceptability of Virtual Coaches for Home-based Balance Training in an Aging Population. In: CHI '20: Proceedings of the 2020 CHI

- Conference on Human Factors in Computing Systems. 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI, USA p. 1-12. [doi: [10.1145/3313831.3376565](https://doi.org/10.1145/3313831.3376565)]
27. Liston M, Genna G, Maurer C, Kikidis D, Gatsios D, Fotiadis D, et al. Investigating the feasibility and acceptability of the HOLOBalance system compared with standard care in older adults at risk for falls: study protocol for an assessor blinded pilot randomised controlled study. *BMJ Open* 2021 Feb 12;11(2):e039254 [FREE Full text] [doi: [10.1136/bmjopen-2020-039254](https://doi.org/10.1136/bmjopen-2020-039254)] [Medline: [33579762](https://pubmed.ncbi.nlm.nih.gov/33579762/)]
 28. Morgan MT, Friscia LA, Whitney SL, Furman JM, Sparto PJ. Reliability and validity of the Falls Efficacy Scale-International (FES-I) in individuals with dizziness and imbalance. *Otol Neurotol* 2013 Aug;34(6):1104-1108 [FREE Full text] [doi: [10.1097/MAO.0b013e318281df5d](https://doi.org/10.1097/MAO.0b013e318281df5d)] [Medline: [23542134](https://pubmed.ncbi.nlm.nih.gov/23542134/)]
 29. Pardalis AA, Gatsios D, Tsakanikas V, Walz I, Maurer C, Kikidis D, et al. Exploring the Acceptability and Feasibility of Providing a Balance Tele-Rehabilitation Programme to Older Adults at Risk for Falls: An Initial Assessment. 2021 Presented at: 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); November 01-05, 2021; Mexico p. 6915-6919. [doi: [10.1109/embc46164.2021.9629478](https://doi.org/10.1109/embc46164.2021.9629478)]
 30. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: <https://creativecommons.org/licenses/by/4.0/> [accessed 2022-08-29]
 31. Guo G, Wang H, Bell D, Bi Y, Greer K. KNN Model-Based Approach in Classification. In: Meersman R, Tari Z, Schmidt DC, editors. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science*, vol 2888. Berlin, Heidelberg: Springer; 2003:986-996.
 32. Pal M, Foody GM. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sensing* 2010 May;48(5):2297-2307. [doi: [10.1109/tgrs.2009.2039484](https://doi.org/10.1109/tgrs.2009.2039484)]
 33. Dudley RM. Sample Functions of the Gaussian Process. In: Giné E, Koltchinskii V, Norvaiša R, editors. *Selected Works of R.M. Dudley. Selected Works in Probability and Statistics*. New York, NY: Springer; 2010:187-224.
 34. Qi Y. Random Forest for Bioinformatics. In: Zhang C, Ma Y, editors. *Ensemble Machine Learning*. Boston, MA: Springer; 2012:307-323.
 35. Webb GI. Naïve Bayes. In: Sammut C, Webb GI, editors. *Encyclopedia of Machine Learning*. Boston, MA: Springer; 2011:713-714.
 36. Schapire RE. Explaining AdaBoost. In: Schölkopf B, Luo Z, Vovk V, editors. *Empirical Inference*. Berlin, Heidelberg: Springer; 2013:37-52.
 37. Devika R, Avilala SV, Subramaniaswamy V. Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. 2019 Presented at: 3rd International Conference on Computing Methodologies and Communication (ICCMC); March 27-29, 2019; Erode, India p. 679-684. [doi: [10.1109/iccmc.2019.8819654](https://doi.org/10.1109/iccmc.2019.8819654)]
 38. Fernandez A, Garcia S, Herrera F, Chawla NV. SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research* 2018 Apr 20;61:863-905. [doi: [10.1613/jair.1.11192](https://doi.org/10.1613/jair.1.11192)]
 39. Warrens MJ. Five ways to look at Cohen's kappa. *J Psychol Psychother* 2015;05(04):1. [doi: [10.4172/2161-0487.1000197](https://doi.org/10.4172/2161-0487.1000197)]
 40. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22(3):276-282 [FREE Full text] [Medline: [23092060](https://pubmed.ncbi.nlm.nih.gov/23092060/)]
 41. Burns DM, Leung N, Hardisty M, Whyne CM, Henry P, McLachlin S. Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch. *Physiol Meas* 2018 Jul 23;39(7):075007. [doi: [10.1088/1361-6579/aacfd9](https://doi.org/10.1088/1361-6579/aacfd9)] [Medline: [29952759](https://pubmed.ncbi.nlm.nih.gov/29952759/)]
 42. Haghighi Osgouei R, Soulsby D, Bello F. Rehabilitation exergames: Use of motion sensing and machine learning to quantify exercise performance in healthy volunteers. *JMIR Rehabil Assist Technol* 2020 Aug 18;7(2):e17289 [FREE Full text] [doi: [10.2196/17289](https://doi.org/10.2196/17289)] [Medline: [32808932](https://pubmed.ncbi.nlm.nih.gov/32808932/)]

Abbreviations

- ABC:** Activities-Specific Balance Confidence Scale
CNN: convolutional neural network
FES-I: Falls Efficacy Scale International
FGA: Functional Gait Assessment
kNN: k-nearest neighbors
ML: machine learning
SVM: support vector machine

Edited by T Leung; submitted 11.02.22; peer-reviewed by A Videira-Silva, T Szturm; comments to author 03.05.22; revised version received 23.05.22; accepted 25.06.22; published 31.08.22

Please cite as:

Tsakanikas V, Gatsios D, Pardalis A, Tsiouris KM, Georga E, Bamiou DE, Pavlou M, Nikitas C, Kikidis D, Walz I, Maurer C, Fotiadis D

Automated Assessment of Balance Rehabilitation Exercises With a Data-Driven Scoring Model: Algorithm Development and Validation Study

JMIR Rehabil Assist Technol 2022;9(3):e37229

URL: <https://rehab.jmir.org/2022/3/e37229>

doi: [10.2196/37229](https://doi.org/10.2196/37229)

PMID: [36044258](https://pubmed.ncbi.nlm.nih.gov/36044258/)

©Vassilios Tsakanikas, Dimitris Gatsios, Athanasios Pardalis, Kostas M Tsiouris, Eleni Georga, Doris-Eva Bamiou, Marousa Pavlou, Christos Nikitas, Dimitrios Kikidis, Isabelle Walz, Christoph Maurer, Dimitrios Fotiadis. Originally published in JMIR Rehabilitation and Assistive Technology (<https://rehab.jmir.org>), 31.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Rehabilitation and Assistive Technology, is properly cited. The complete bibliographic information, a link to the original publication on <https://rehab.jmir.org/>, as well as this copyright and license information must be included.