



A computer vision approach to continuously monitor fatigue during resistance training

Justin Amadeus Albert ^{*}, Bert Arnrich

Digital Health - Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Rudolf-Breitscheid-Strasse 187, Potsdam, 14482, Brandenburg, Germany

ARTICLE INFO

Keywords:

Computer vision
Machine learning
Deep learning
Fatigue prediction

ABSTRACT

Monitoring fatigue during resistance training is essential to avoid injuries caused by overtraining. Fatigue can be comprehensively quantified by the external and internal load, where the external load is the work done by the athlete, and the internal load is the psychological and physiological response to the external load. This paper proposes a computer vision method to continuously monitor fatigue during resistance training by predicting external and internal parameters, namely the generated power and the rating of perceived exertion. We utilize the human pose estimation from two Microsoft Azure Kinect cameras to capture the movement of athletes while performing stationary exercises. Our method processes the obtained kinematic data, computes skeleton features to train traditional machine learning algorithms, and constructs feature maps to train convolutional neural network-based models to predict the load parameters. For evaluation, we recorded a dataset of 16 subjects who performed squat exercises on a Flywheel and rated their perceived exertion after each set. A measuring unit integrated into the Flywheel provided power readings for each repetition. The results show that our method achieves good results in predicting both parameters. Gradient Boosting Regression Trees best predicted perceived exertion with a mean absolute percentage error of 8.08% and a Spearman's $\rho = 0.74$. Multi-layer Perceptron performed best in predicting power with a mean absolute error of 23.13 Watts and $\rho = 0.79$. Our findings show that our approach delivers promising external and internal load quantifications for fatigue, with great potential to provide external feedback to coaches or athletes.

1. Introduction

Resistance training is beneficial for athletes to generate muscular strength [1] and is one of the most critical factors in preventing frailty in the elderly population [2]. However, fatigue can occur during exercise activities and might lead to a deviation of the movement strategies, increasing the injury risk [3,4]. Therefore, continuous fatigue monitoring is beneficial for preventing injuries caused by overtraining. Training workload comprises external and internal loads. The external load refers to the performed physical work, e.g., the generated power or the distance traveled during the training session. Internal load is the psychological and physiological stress imposed on the athlete by the external load, such as heart rate, lactate concentrations, or RPE [5]. Internal response varies between individuals; e.g., running the same distance at the same speed might feel different for individual athletes or have different heart rate responses. RPE are numerical scales where athletes rate their perceived level of exertion. A typical example is the Borg scale, which ranges from 6 (not exhausted) to 20 (extremely exhausted) [6]. RPE has proven to be an effective method to assess the internal load, as it can be easily collected and used for different exercise

modalities [7]. However, Coyne et al. [8] mentioned that RPE might suffer from a lack of compliance, dishonest reporting because of peer pressure, and influencing training plans. Also, RPE is usually assessed retrospectively after each set or training session (the so-called session RPE), which might suffer from a recall bias.

Scott et al. [7] recommend combining multiple training quantifications in a multivariate training monitoring as it offers broader insights into the fatigue state of athletes. To this end, external load parameters can help quantify the fatigue status holistically; however, they are sometimes hard to obtain as they might require specialized equipment. For example, measuring the total or high-speed running distance requires a Global Positioning System (GPS) [9]. To predict power values for overground walking requires ground reaction force plates or linear transducers [10]. Therefore, a monitoring tool that combines internal and external training variables is beneficial to deliver a holistic view of training quantification.

Research aimed to overcome the mentioned limitations of RPE by automatically predicting RPE using various sensor modalities in different scenarios. Multiple studies [11–14] investigated the individual

* Corresponding author.

E-mail address: justin.albert@hpi.de (J.A. Albert).

RPE prediction during soccer training using mainly features from GPS and Heart Rate Variability (HRV) data. Other studies investigated RPE prediction during running [15–19], mostly relying on GPS and Inertial Measurement Unit (IMU) data. For resistance training, our domain, some studies [20,21] investigated RPE prediction for stationary resistance exercises, primarily using IMU data. However, all the existing research for RPE prediction relies on data from wearable sensors while only predicting RPE as a single training quantification, sometimes even using an external load, such as GPS, as features. Therefore, this paper aims to overcome those research limitations by proposing an approach to predict internal and external load quantification using machine learning to allow for a contactless, easy-to-use fatigue monitoring approach, removing the burden of wearable sensors while offering holistic training quantification.

In more detail, this approach aims to monitor fatigue during stationary exercises by predicting RPE and power outputs. Our markerless computer vision method uses three-dimensional (3D) skeleton data from two Microsoft Azure Kinect cameras. We present a method to estimate robust and temporal consistent 3D joint positions and orientations by fusing skeleton sequences from both Kinect cameras. From there, this paper presents and compares two developed methods using traditional ML and Deep Learning (DL). The ML method incorporates domain knowledge by segmenting the skeleton sequences into individual squat repetitions and handcrafting skeleton and statistical features on the data before training GBRT, Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Regression (SVR) models to predict RPE. The DL approach potentially learns valuable hidden information in the raw data without the explicit need for feature engineering. To this end, a Convolutional Neural Network (CNN) combined with a Gated Recurrent Unit (GRU) model (CNN-GRU) is trained on transformed skeleton data as image-like structures to predict RPE. Ultimately, both approaches are extended to predict generated power as an objective external training quantification. The models are evaluated on a dataset including 16 subjects performing the squat exercise in a controlled laboratory environment and rated their fatigue level on the 6–20 Borg scale. Blood lactate analysis confirmed the induced fatigue. With the presented approach, we aim to develop a holistic training monitoring system that alleviates the need to wear body sensors and provides real-time feedback.

This paper makes the following contributions throughout the development of our markerless continuous fatigue monitoring system. (1) A computer vision approach to predict RPE and power that works completely markerless without the need for any wearables attached to the athlete. (2) A camera setup consisting of two low-cost Kinect cameras to derive robust and temporal consistent 3D skeleton data to train the machine learning models. (3) A comparison of prediction accuracy of two developed approaches, traditional ML trained on handcrafted domain-knowledge features and automatically extracted features for DL. (4) An evaluation of the approaches on a novel dataset including data from 16 subjects that performed squat exercises in a controlled environment, where fatigue was confirmed with blood lactate samples and power was obtained from a specific measurement unit.

The rest of the paper is structured as follows: Section Two presents relevant studies to our approach from multiple areas, including motion capture, Human Activity Recognition (HAR), RPE, and power prediction. Section Three presents our computer vision approach to predict RPE and generated power, which is evaluated in Section Four. Section Five summarizes our results and discusses study limitations, practical applications, and future work.

2. Related work

This section discusses relevant studies to our marker-less motion RPE and power prediction approach, including low-cost optical motion capture, related methods from HAR, and RPE and power prediction studies.

2.1. Optical marker-less motion tracking

Multi-camera marker-based motion capture systems are considered the gold standard for motion analysis. However, these systems are expensive, usually bound to laboratory settings, and require anatomical knowledge for the marker attachment. In 2010, Microsoft released the Kinect, a so-called RGB-D camera (red, green, blue, and depth) sensor, implementing a data-driven algorithm to track specific landmarks of users [22]. Since then, the research community has investigated this device for tasks such as human gait analysis or functional movement [23–25]. More recently, CNNs have significantly advanced computer vision for various tracking tasks such as face detection [26] and person identification [27], and human pose estimation [28]. In 2019, Microsoft released the latest Kinect generation, Azure Kinect, that uses CNN to improve human pose estimation performance [29].

2.2. Human activity recognition based on skeleton data

A related field to this study is HAR, which identifies specific human actions. Skeleton data has proven to be a suitable modality for this task as the skeleton data encodes sufficient information on performed actions [30]. The study by Franco et al. [31] investigated skeleton data for HAR. The authors demonstrated the importance of specific skeleton features, such as joint orientations and angles, independent of the viewpoint and normalizing the user height. With this specific feature set, the authors achieved new state-of-the-art results with a 95.0% precision and 95.0% recall on an established 12-class activity dataset. Deep learning also advanced the field of HAR, where deep models have proven effective for learning longer and more complex human activities. Yadav et al. [32] presented an approach to classify human actions using a CNN and Long Short-Term Memory (LSTM) model that learns temporal and spatial features in the skeleton data. They validated their method on a self-recorded 7-class activity dataset with Kinect v2, achieving a 98.89% accuracy.

2.3. Prediction of ratings of perceived exertion

RPE prediction studies investigated various sensor modalities in different settings, such as running, soccer, or resistance training. Pernek et al. [21] investigated RPE prediction during resistance training, recording subjects with five IMU sensors. They calculated statistical and autocorrelation features on the IMU data in sliding windows. Support Vector Machine (SVM) achieved an accuracy of 6% of prediction error on a normalized RPE, i.e., in the interval [0, 1]. Jiang et al. [20] used data from five IMU sensors and a force plate captured from subjects performing three exercises until they reached a maximum exertion level. Authors handcrafted features and trained RF and CNN models. They found the highest correlations between fatigue and displacement for the center of pressure features. Kim et al. [14] predicted RPE for soccer athletes using GPS. They handcrafted features from GPS data, including velocity, acceleration, and jerk features, to train a CNN with a GRU part. They obtain a MAE = 0.78 for the RPE 1–10 class prediction. Zhou et al. [33] predicted RPE during lifting tasks from video using tracked 2D image coordinates and extracted facial expressions. A proposed 1D Conv-LSTM network predicted RPE values with a Root Mean Squared Error (RMSE) of RMSE = 2.26.

2.4. Prediction of generated power from movement data

Power measurements deliver valuable about athlete performance. However, obtaining those measurements requires specialized equipment. Machine learning-based methods can overcome this issue by learning a mapping from other modalities to the power response variables. Jiang et al. [34] proposed a machine-learning method to predict the ankle joint power during walking using IMU sensors to deploy power calculation to outdoor scenarios without access to specialized

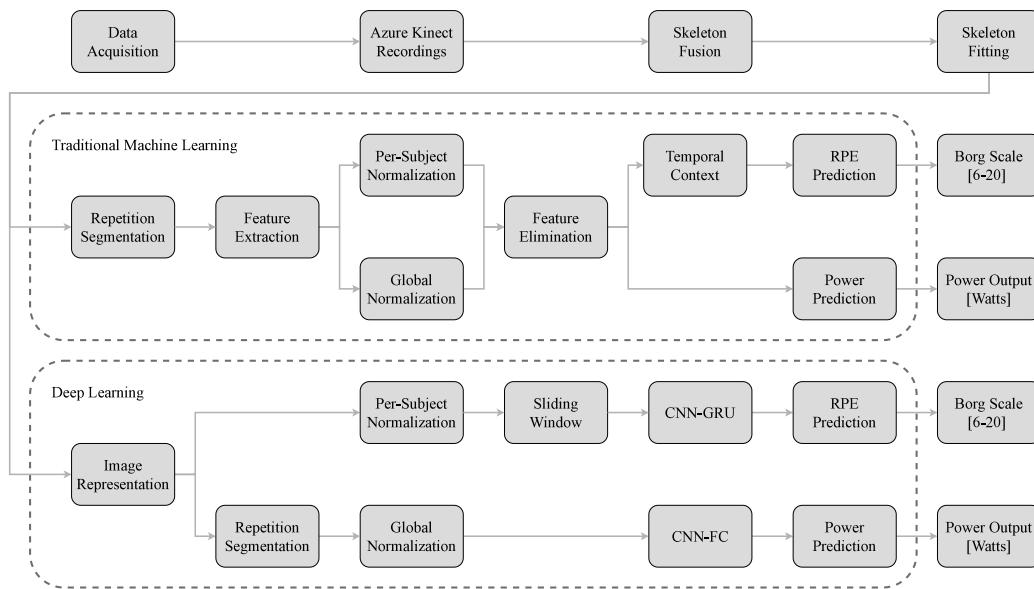


Fig. 1. Overview of the RPE and power prediction methodology based on skeleton data from two Kinect cameras. The first step is skeleton data fusion and processing. Subsequently, traditional ML and DL models are trained for both prediction tasks.

equipment. Ankle joint power is usually obtained with inverse dynamics methods that require biomechanical models. The authors train the RF model on the data from two IMU sensors to predict ankle joint power. A marker-based optical system and a force-plate-instrumented treadmill provided ground truth. Their method achieved a RMSE = 0.06 w/kg. Uddin et al. [35] investigated power prediction during roller ski skating. Thirteen subjects performed trials with roller skis on a treadmill while being recorded with seven IMU sensors attached to the upper body. After calculating statistical features on the raw IMU data, the authors trained a LSTM network to predict the mechanical power. The authors found that a subject-dependent model delivered accurate results with a relative error of 3.5%, while a general model achieved a relative error of 11.6%.

3. Materials and methods

This section presents our human pose estimation approach to predict RPE and power during resistance training. Formally, the aim is to learn a function $f : \mathbf{X} \rightarrow [6, 20]$ from given skeleton data \mathbf{X} to predict RPE values provided on the Borg scale, which ranges from 6 to 20. For power predictions, the learned function $f : \mathbf{X} \rightarrow \mathbb{R}$ shall predict power in the unit of Watts based on given skeleton data \mathbf{X} . Fig. 1 provides an overview of the here-presented approach. After initial data acquisition, skeleton sequences from two Microsoft Kinect cameras are pre-processed and fused into one skeleton representation. Robust joint orientations are estimated using a skeleton-fitting approach. From there, this paper presents machine learning and deep learning methods for both prediction tasks, namely RPE and power. For the traditional ML, skeleton sequences are segmented into individual repetitions, from which skeleton and statistical features are handcrafted to train the ML models on both prediction tasks. The data processing steps vary for the two prediction tasks, i.e., RPE prediction includes a subject-dependent normalization and a temporal context. In contrast, power prediction uses global normalization and does not require temporal context. For the DL methods, skeleton sequences are transformed into image-like representations (as explained in Section 3.5). The model for RPE prediction comprises a CNN and GRU component to model the spatial and temporal variation in the data. The power prediction model consists of a CNN followed by fully-connected (FC) layers, as power does not require a temporal context.

3.1. Data acquisition

Sixteen healthy male subjects (average age: 23.3 ± 2.9 years, body mass: 82.6 ± 4.8 kg, and height: 183.8 ± 5.3 cm) participated in our study. The protocol consisted of performing the squat exercise involving the leg muscles, the largest muscle group of the human body. Subjects performed squats on a Flywheel device (Exxentric Training, Sweden) that consists of a platform and a spinning wheel connected to a rope. The athlete is attached to the rope with a hip harness. Pulling on the rope during the upwards (concentric) phase creates inertia stored in the wheel. The athlete has to neutralize this energy during the downward movement (the eccentric movement). Unlike back squats using a weighted barbell that works against gravity, the load in Flywheel training is determined by the mass of the disc and the velocity with which the athlete pulls on the hip harness. Fig. 2 shows an athlete performing squats on the Flywheel.

The data acquisition started with five minutes of relaxation and a two-minute free warmup. Subsequently, the maximum squat velocity for each subject was tested by measuring the times for a few squat repetitions performed with maximum effort. Then, 90% of the average repetition time was targeted for the entire training session, guided by a visual metronome on a Tablet presented to the athletes during the workout. The protocol consisted of twelve sets with twelve repetitions in each set. Subjects had one minute rest after each set and three minutes rest after every third set. After each set, subjects reported their current RPE on the Borg scale. Induced fatigue was confirmed by analyzing blood lactate samples before and after the exercises.

During the experiments, data was collected from two Microsoft Azure Kinect cameras, a one-channel Electrocardiography (ECG) sensor, and six IMUs sensors with a full-body setup. The Flywheel device has a measurement unit integrated, the so-called kMeter device, which measures the wheel rotation to calculate aggregated parameters, such as repetition power, duration, or vertical squat distance. The power measurements were evaluated against a force plate as a gold-standard device [36] and served as ground truth for the power predictions based on the skeleton data. The study was approved by the Ethics Review Board (ERB) of the University of Potsdam (Application no. 21/2021). Before the experiments, subjects gave their informed consent. The recorded dataset is publicly available, and more information about the dataset can be found in our PERSIST dataset publication as provided by Albert et al. [37].

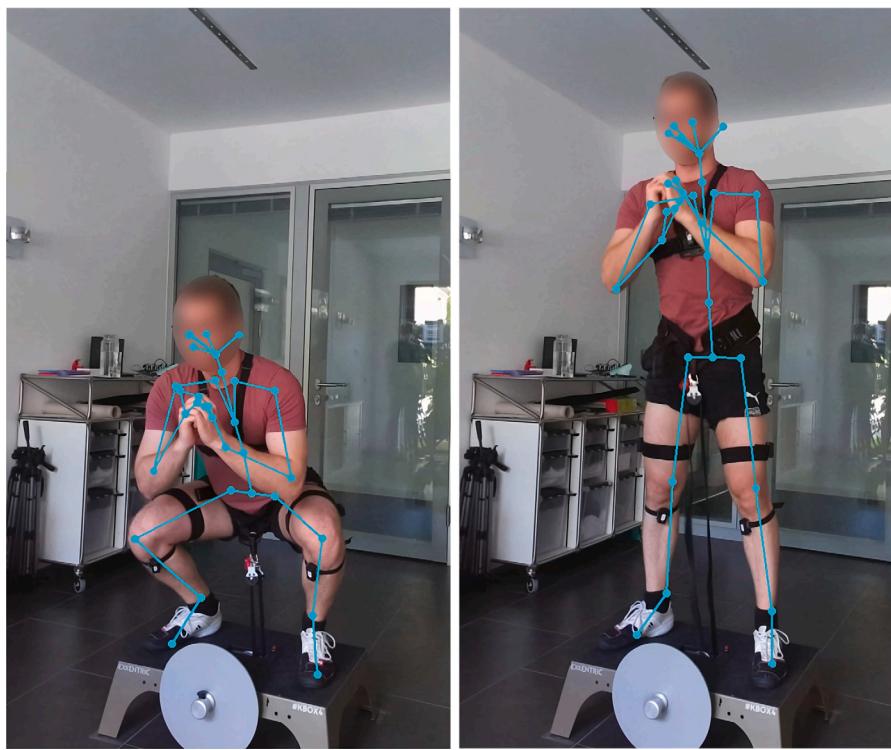


Fig. 2. Subject performing squats on the Flywheel while being tracked with the Azure Kinect skeleton tracking. Left: repetition starts by standing up and creating inertia in the wheel. The wheel stands still, see the black pin in the middle. Right: During the downward phase, the athlete neutralizes the energy.

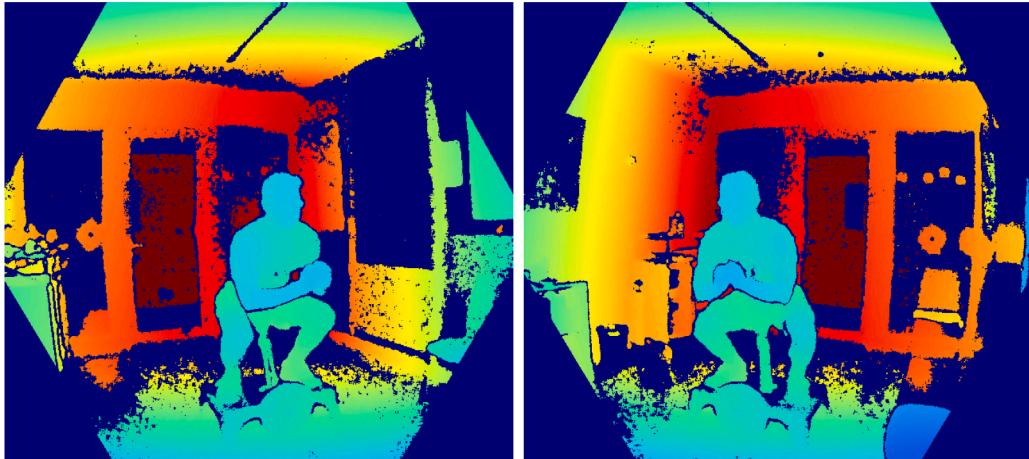


Fig. 3. Images from the two Microsoft Azure Kinect cameras. Cameras were positioned on the left and right sides of the subjects and captured data simultaneously.

Studies have shown that the human pose estimation accuracy of the Microsoft Kinect camera suffers from issues such as self-occlusion or complex movement tasks [38,39]. To this end, the camera system used in this work consists of a two-camera setup to tackle the issue of potential self-occlusion and to make the skeleton tracking more robust by compensating for the inaccuracies of the first camera with the second camera. The here-used camera system consists of two Microsoft Azure Kinect cameras placed in front of the subjects, moved to the left and right of the subjects. Fig. 3 shows the camera view of the subjects. The cameras were mounted on tripods and adjusted in height so that the subjects were properly visible in the field of view. The cameras were placed as close to the subjects while maintaining a proper field of view to increase the depth accuracy.

The Kinect cameras were temporally synchronized via a hardware mechanism using a 3.5 mm audio cable provided by the Azure Kinect devices [40]. The right camera was set as the primary device, while

the left was the secondary device. A fixed temporal offset was set between both cameras and a fixed exposure time for each camera to ensure stable data recordings, as variable exposure does alter the equidistant sampling. Both Kinect cameras recorded data at a sampling frequency of 30 Hz. Both cameras recorded raw .mkv (Matroska) files that contained the different sensor modalities from both Kinect cameras, including the color, depth, and Infra-red (IR) images. The 3D skeleton time-series data, including the 3D positions and orientations, was extracted in a post-processing step using the Microsoft Azure Kinect Body Tracking SDK in version 1.1.2, the most recent version when working on this project.

3.2. Skeleton fusion from two-camera setup

The first step was to fuse the synchronized 3D skeleton positions from the two Kinect cameras and manually estimate 3D orientations

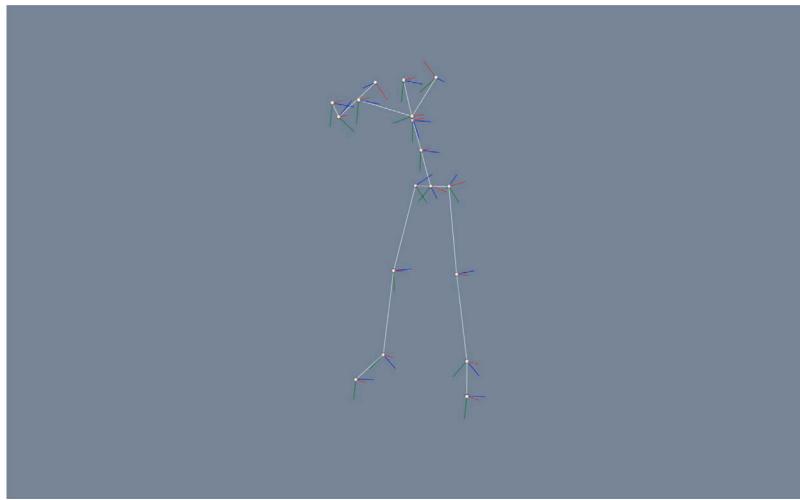


Fig. 4. Final skeleton as estimated by the skeleton fitter. The local joint orientations are depicted by the coordinate systems.

to obtain robust skeleton sequences. Generally, a skeleton sequence is defined as $\mathbf{P} = \{p_1^1, \dots, p_n^j\}$ with N frames and J joints, where each $p_n^j \in \mathbb{R}^3$ represents joint coordinate in three-dimensional space. The Azure Kinect delivers $J = 32$ joints, from which meaningless or redundant markers like nose, ears, and eyes were removed, ending up with $J = 18$ joints for each skeleton sequence. The first step was to spatially align the joint positions of both sequences by minimizing the mean squared error between corresponding joints, using the SVD-based Kabsch algorithm, as presented in our previous publication by Albert et al. [37]. Subsequently, the roughly overlapping skeleton sequences were fused by a fusion algorithm that punishes inconsistencies, i.e., larger jumps in each signal [37]. Qualitative experiments have shown that this fusion algorithm delivered better results than advanced methods such as the Extended Kalman Filter [38].

The fusion result is a somewhat temporally stable skeleton sequence. The next step was to estimate the local 3D joint orientations by fitting a standard skeleton with defined bone lengths into the measured 3D points, inspired by the approach of Mehta et al. [28]. The joint orientations are defined as $\mathbf{O} = \{o_1^1, \dots, o_n^j\}$, where each $o_n^j \in \mathbb{R}^3$ represents a 3D Euler rotation matrix. Our developed inverse kinematic solver estimates the joint orientations as Euler angles in radians within the fused skeleton sequence. To this end, each set of angles per frame o_n^j is formulated as flattened vector $\theta_i \in \mathbb{R}^{3J}$. The method starts by defining a standard skeleton with constant bone lengths $\mathbf{B} \in \mathbb{R}^l$ with $l = J - 1$, calculated as the mean distance between two adjacent joints that form a limb. This skeleton has initial local joint orientations of θ_0 that are defined to show a T-Pose. The optimization is further constrained by only allowing anatomically reasonable angle configurations within certain intervals. Successively, the optimizer fits the standard skeleton into the measured 3D skeleton points by minimizing the least squares between the current skeleton frame p_i^j and the standard skeleton parameterized with bone lengths and angles θ_i , as shown in Eq. (1). The function $D(\theta_i, \mathbf{B})$ constructs joint positions from given angles θ_i and bone lengths \mathbf{B} by using the kinematic chain of the skeleton.

$$E(\theta_i) = \left\| \left(D(\theta_i, \mathbf{B}) - p_i^j \right) \right\|_2 \quad (1)$$

Our skeleton fitting algorithm uses the Trust Region Reflective algorithm from the Scikit library for the Python programming language. For further speedup, the skeleton fitter uses the angle configuration for step t from the previous frame $t - 1$ as the initial guess. Finally, the resulting skeleton sequences were filtered using a 4th-order Butterworth filter with a cut-off frequency of 12 Hz. Fig. 4 shows a final skeleton frame, including the estimated joint orientations.

3.3. Segmentation of squat exercises

The next step is segmenting the skeleton sequences with fused 3D joint positions and estimated 3D joint orientations into individual repetitions. This process will encode specific domain knowledge into the features and simplify mapping power values for each repetition. The segmentation method works on a 1D signal of the positional data of a joint with a distinct pattern chosen as the vertical axis of the Pelvis joint [41]. A peak-finding algorithm identified potential repetition candidates in the signal due to the sinusoidal shape of the joint trajectory. Invalid repetition candidates were discarded based on manually defined rules, such as minimum vertical distance, maximum duration, and minimum standard deviation [42]. After discarding all invalid repetition candidates, our approach further segments the repetitions into the concentric phase, i.e., the upwards phase, and the eccentric phase, i.e., the downwards phase of a squat. Fig. 5 shows a time series example with detected phases. This information will further encode specific knowledge into the features.

3.4. Machine learning-based RPE and power prediction method

In the presented approach, the traditional machine learning models are generally trained on the segmented skeleton sequences from the previous step. Absolute 3D position coordinates are not suited for training due to the arbitrary positioning, the different sizes of subjects, and the different viewpoints [43]. Therefore, absolute joint positions are first converted into relative coordinates as proposed by Ahad et al. [44]. To this end, the central hip-joint p_i^{Pelvis} is subtracted from every other joint in each frame p_i^j . Subsequently, the result is divided by the distance from the neck to the spine-chest joints to make the skeletons independent from the subject size, as shown in Eq. (2). The result is a time series of $J - 1$ relative joints after neglecting the Pelvis joint only consisting of zeros.

$$p_i^j = \frac{p_i^j - p_i^{\text{Pelvis}}}{\|p_i^{\text{Neck}} - p_i^{\text{Spine}}\|} \quad (2)$$

The estimated joint orientations are already independent of the subject height and view-invariant; thus, they require no further transformations.

The next step is feature extraction, calculated for the relative skeleton positions and orientations for the concentric and eccentric phases. The statistical features include median, length, standard deviation, skewness, kurtosis, root mean square, maximum, absolute maximum, and minimum. The result from this step is a rather large data set

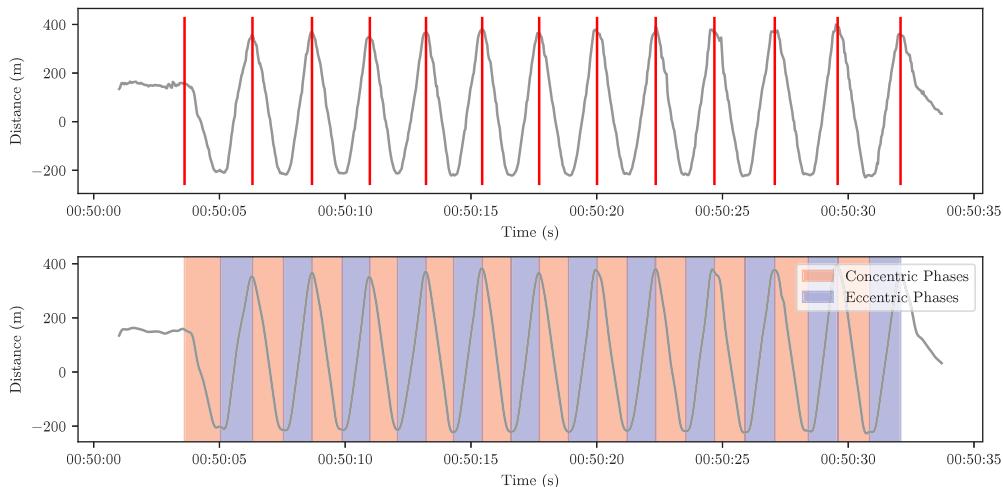


Fig. 5. Segmented squat repetitions based on a peak finding algorithm and manual rules. Top row: Raw data with repetition candidates. Bottom row: Filtered and segmented skeleton data. Concentric phases are colored in red and eccentric phases in blue.

$\mathbf{X} \in \mathbb{R}^{M \times N}$ with M samples, i.e., individual repetitions and $N = 2(\text{con+ecc}) \cdot 9(\text{stat. features}) \cdot 3(\text{axes}) \cdot (17 + 18)\text{joints} = 1890$ features. To this end, reducing the feature set size is necessary to eliminate unimportant features before training the models [16]. The first step is to remove columns with constant values, followed by removing highly correlated features, i.e., all features with a correlation higher than 0.95. Furthermore, extreme outliers in the feature set were clipped using a Z-score of 3σ to maintain the sample size in the dataset but to prevent significant outliers from slowing down the convergence of models and to improve the generalizability of unseen data. Afterward, this approach uses the Recursive Feature Elimination (RFE) to obtain the most important features for the desired number of features, choosing 50 and 100 features for all experiments.

Four traditional ML models were trained, including SVR, GBRT, RF, and MLP [45], for both prediction tasks, i.e., predicting RPE and power measurements. In each experiment, an extensive grid search was performed for each model to find the optimal set of hyperparameters. Table 1 presents the tested hyperparameters of each of the four traditional ML models. As the samples extracted in the feature set are independent, a temporal context is necessary for the task of RPE prediction. This context is essential to model the temporal relationship between the extracted features and uncover potential temporal changes. To this end, the temporal context for the RPE prediction task was implemented using rolling statistics of all features, i.e., the mean and standard deviation of each feature for the last given number of samples are added to the dataset as a new feature. Our experiments included different temporal histories, i.e., using the rolling statistics of the last 6, 9, and 12 repetitions. Those rolling statistics features were normalized independently before training the ML models.

No temporal context is needed for the power prediction task as the model individually predicts single power values for each repetition. To this end, the power prediction models are trained on the sample-independent dataset \mathbf{X} .

3.5. Deep learning-based RPE and power prediction method

This paper proposes two model architectures for the deep learning-based methods, one for each prediction task that operates on transformed raw skeleton data. The first step is transforming the recorded skeleton sequences of 3D joint positions and orientations into a well-suited representation for models processing structured data. Our approach encodes the joint signals into an image-like representation with three channels inspired by the approach presented by Chen et al. [46]. An image encoding consists of a channel with relative joint positions, a channel of joint displacement, and a channel with joint orientations.

Table 1

Hyperparameter search space of the four traditional ML models tested in the grid search approach. Those parameters were tested for both prediction tasks, i.e., RPE and power predictions.

Model	Parameter	Values
GBRT	Nr. Estimators	{50, 150, 500}
	Learning rate	{0.05, 0.1, 0.2}
	Loss	{Squared error, Absolute error}
	Nr. Iter. w/ No change	{None, 5, 50, 100}
MLP	Hidden layers	{(100,), (100, 50), (100, 150)}
	Activation	{Logistic, ReLU, TanH}
	Solver	{SGD, Adam}
	Learning rate Init.	{0.01, 0.001, 0.0001}
	Epochs	{2000, 5000}
	Early stopping	{True, False}
RF	Nr. Estimators	{50, 100, 150, 200}
	Criterion	{Squared error, Absolute error}
SVR	Kernel	{Linear, Radial basis function}
	Gamma	{0.001, 0.0001}
	C	{0.001, 0.01, 0.1}

The relative coordinates are calculated by subtracting the Pelvis joint from each joint, as presented in Eq. (3), where p_i^{Pelvis} represents the Pelvis joint.

$$\mathbf{R} = p_i^j - p_i^{\text{Pelvis}} \quad (3)$$

The joint displacement represents the joint velocity, calculated as shown in Eq. (4).

$$\mathbf{D} = p_{i-1}^j - p_i^j \quad (4)$$

The third channel is the unmodified joint orientation calculated by the skeleton fitter. The final image-like representation is denoted as feature map I consisting of the stacked three channels $\mathbf{I} = (\mathbf{R}, \mathbf{D}, \mathbf{O})$. Fig. 6 presents samples from the image-like representation that encodes the joint and temporal information used for the prediction tasks, where the channels are converted to the interval [0, 255] for viewing purposes.

Fig. 7 presents the developed DL-based model architectures for RPE and power prediction based on the calculated image encodings. Both models build upon the same CNN-based backbone architecture that shall extract meaningful spatial features from the image encodings. This backbone consists of 2D convolutional layers with a 3×3 kernel, followed by batch normalization and 2×2 max pooling. The number of filters in each convolutional layer doubles, starting with 64 or 128 filters. The RPE prediction model consists of a GRU block after the CNN backbone to learn temporal properties in the data. In contrast,

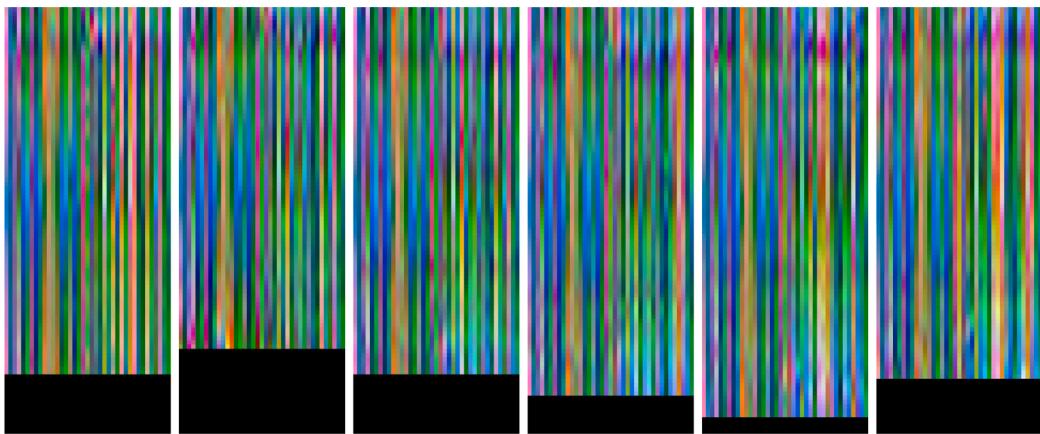


Fig. 6. Image encodings of six repetitions from the same subject. Repetitions correspond to the last repetition of the 2, 4, 6, 8, 10, and 12th set. Pixel values were normalized to [0, 255] for visualization, and black pixels show the zero-padding.

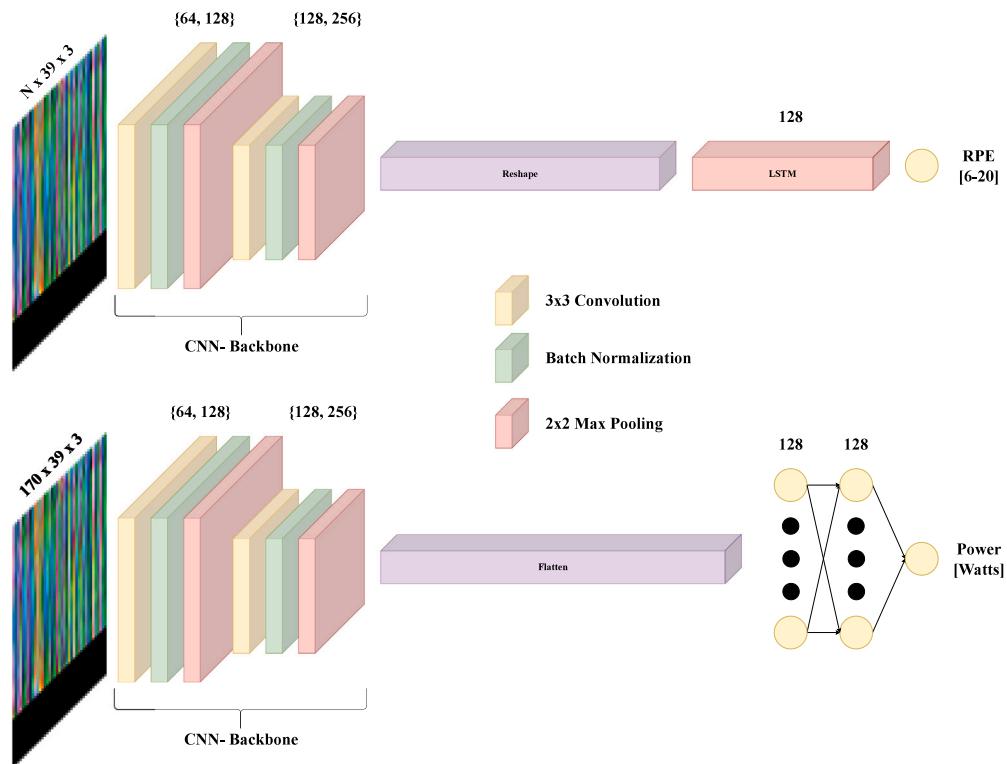


Fig. 7. The two deep learning architectures developed for this approach. The first architecture is for the RPE prediction task, and the second is for the power prediction task.

the power prediction model consists of fully connected layers after the CNN backbone. The model hyperparameters were optimized using a grid search, as shown in Table 2. The CNN backbone parameters were tested for all combinations for each GRU parameter and FC parameters, for RPE and power predictions, respectively.

As the RPE prediction task requires temporal context to model evolving fatigue over time, the CNN-GRU model, the output of the last convolutional layer is fed to a GRU recurrent layer to learn the temporal structure of the data. Due to the relatively small dataset size, the RPE prediction model utilizes a GRU layer instead of an LSTM layer, as it has a more complex structure and is usually harder to train. The image-like encodings are fed to the CNN-GRU model using a sliding window approach with different window lengths of 1, 2, 3, and 4 s, equal to 30, 60, 90, and 120 frames. Due to the small dataset size, a large window overlap of 95% was used to generate as much training data as possible.

For the CNN-FC model, the last convolutional layer is flattened and forwarded to two fully connected layers, each with 128 neurons and a dropout layer with a dropout probability of $p = 0.5$. The CNN-FC model was trained on segmented individual repetition to output the power value of the according repetition. As the CNN model expects input with fixed sizes, the rows of the image-like representations were filled using zero-padding to the maximum repetition length in the dataset, which in our case is 170 frames.

3.6. Power estimation using physics

As mentioned in Section 2, data-driven approaches are beneficial for predicting kinetics when not all necessary physical variables are available. The aim is to compare the performance of the ML models that learn a mapping to a physical model that aims to describe the experiment using all the available physical variables. The model shown

Table 2

Hyperparameter search space of the deep learning models tested in the grid search approach. Those parameters were tested for prediction tasks, i.e., RPE and power predictions.

Model	Parameter	Values
CNN backbone	Dropout	{0.5}
	Conv layers	{2,3}
	Nr. Filters 1st layer	{64, 128}
GRU (RPE prediction)	Window size	{30, 60, 90, 120}
	Overlap	{0.9}
	Nr. of neurons	{128}
FC (Power prediction)	Nr. Neurons	{64, 128}
	Nr. Layers	{2, 3}
	Dropout	{0.5}
General	Learning rate	{0.0001}
	Optimizer	{Adam}
	Epochs	{500 (early stopping)}
	Batch size	{16, 32}

Table 3

The calculation of correction factors Δr_i used for each test subject. Those values are calculated from the remaining 15 subjects.

Subject	Mean Δr_i	Std Δr_i
0	67.71	10.75
1	68.51	10.86
2	68.61	10.91
3	67.80	10.56
4	67.70	10.55
5	68.79	10.79
6	68.31	10.87
7	68.64	10.57
8	69.24	10.21
9	67.65	10.66
10	69.07	10.57
11	68.50	10.90
12	68.59	10.61
13	68.38	10.75
14	68.08	11.00
15	67.67	10.35

In Eq. (5) takes into account the inertia of the Flywheel disc T , in this case, $T = 0.025 \text{ kg m}^2$ and the radius of the disc set to $r = 0.135 \text{ m}$. Furthermore, it includes the mean velocity \bar{v} in m/s of the pelvis joint, the main actor on the flywheel belt, and the individual repetition duration denoted as d in s.

$$P = \frac{T\bar{v}^2}{2r^2d} \Delta r^2 \quad (5)$$

The presented model was verified by estimating the known radius of the disc r after rearranging Eq. (5) and incorporating the ground-truth power values. This experiment found an offset mismatch between the estimated and ground truth power values, subsequently corrected with a correction factor Δr integrated into Eq. (5). This correction factor was determined for each test subject using the leave-one-subject-out evaluation method, i.e., the correction factor Δr_i for each test subject i with $i \in \{1, \dots, 16\}$ was calculated as the average of the remaining 15 subjects. This evaluation method should prevent partial correction of test subjects included in the correction factor. Table 3 presents the used correction factors Δr_i for each test subject as used in this experiment.

4. Evaluation

This section evaluates our RPE and power prediction methods. First, it defines the used evaluation metrics specifically for regression. All presented experiments utilize the leave-one-subject-out (LOSO) cross-validation (CV), i.e., one subject is left out, and the model is trained on the remaining 15 subjects. Subsequently, the trained model is tested on the left out subject. This process is repeated for each subject, i.e., performing a 16-fold split. Finally, this section reports the mean

and standard deviation of the final test subjects. For each experiment, the grid search is performed for traditional ML and the DL model implementations as presented in Tables 1 and 2, respectively. The training process of the presented DL models uses early stopping, i.e., stopping the optimization when the validation loss increases five times in a row.

4.1. Evaluation metrics

The prediction accuracy of the machine learning and deep learning models is assessed using various error metrics between the predictions \hat{y} and the ground truth y . Those error metrics used here include Mean Absolute Error (MAE) and RMSE. Furthermore, Mean Absolute Percentage Error (MAPE) is assessed, as shown in Eq. (6). In this equation, y_i and \hat{y}_i refer to individual samples, with $i \in \{1, \dots, n\}$, where n is the total number of samples. This metric is often used to compare regression models across datasets.

$$\text{MAPE}(y, \hat{y}) = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

Furthermore, the results are evaluated using the R^2 metric, also known as the coefficient of determination. This metric indicates the proportion of the variance in the dependent variable that the independent variables in the model can explain. It is important to note that even called *R squared*, the metric's values can be negative and range from $[-\infty, 1]$, where a perfect model would achieve a score of $R^2 = 1$. If a model only predicts the constant mean value of the ground truth (if the ground truth is not constant), the model would get a score of $R^2 = 0$. As the model's performance can be arbitrarily worse, the values can also become negative. This metric is defined in Eq. (7), where \bar{y} refers to the mean of the ground truth.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Finally, the predictions are evaluated using Spearman's rank correlation coefficient ρ , a statistical measure that quantifies the strength and direction of the monotonic relationship between two variables. It is based on the ranks of the observations rather than their actual values, making it robust to outliers and non-linear associations. The following four thresholds define the quality of agreement: $\rho \geq 0.9$ excellent, $0.7 \leq \rho < 0.9$ good, $0.4 \leq \rho < 0.7$ moderate, and $\rho < 0.4$ poor correlation [47].

4.2. Prediction of ratings of perceived exertion

This first experiment is the RPE prediction with traditional ML and DL models. Even though the RPE prediction task is a regression problem, RPE labels are interpreted in this approach as categorical class labels to allow for the use of random oversampling for balancing the distribution of samples. Each set in the protocol consists of twelve repetitions but only has a single RPE value assigned. Therefore, the mean value is calculated over all predictions in each set to obtain a final RPE value. Subsequently, the previously presented evaluation metrics are used to evaluate the model on the aggregated RPE values.

4.2.1. Evaluation setup

As the fatigue development during the training session is individual for each subject, the feature sets for the RPE prediction experiments are normalized for each subject individually. To this end, for training the ML models, the feature set X is normalized using standard scaling to obtain \bar{X} by subtracting the subject mean μ_S dividing subject standard deviation σ_S , as shown in Eq. (8).

$$\bar{X}_S = \frac{X_S - \mu_S}{\sigma_S} \quad (8)$$

Analogously, the image encodings I are standard scaled for each subject before training the CNN-GRU model. Therefore, the individual channels C are normalized separately, as shown in Eq. (9).

$$\bar{C}_S = \frac{C_S - \mu_{C_S}}{\sigma_{C_S}} \quad C \in \{R, D, O\} \quad (9)$$

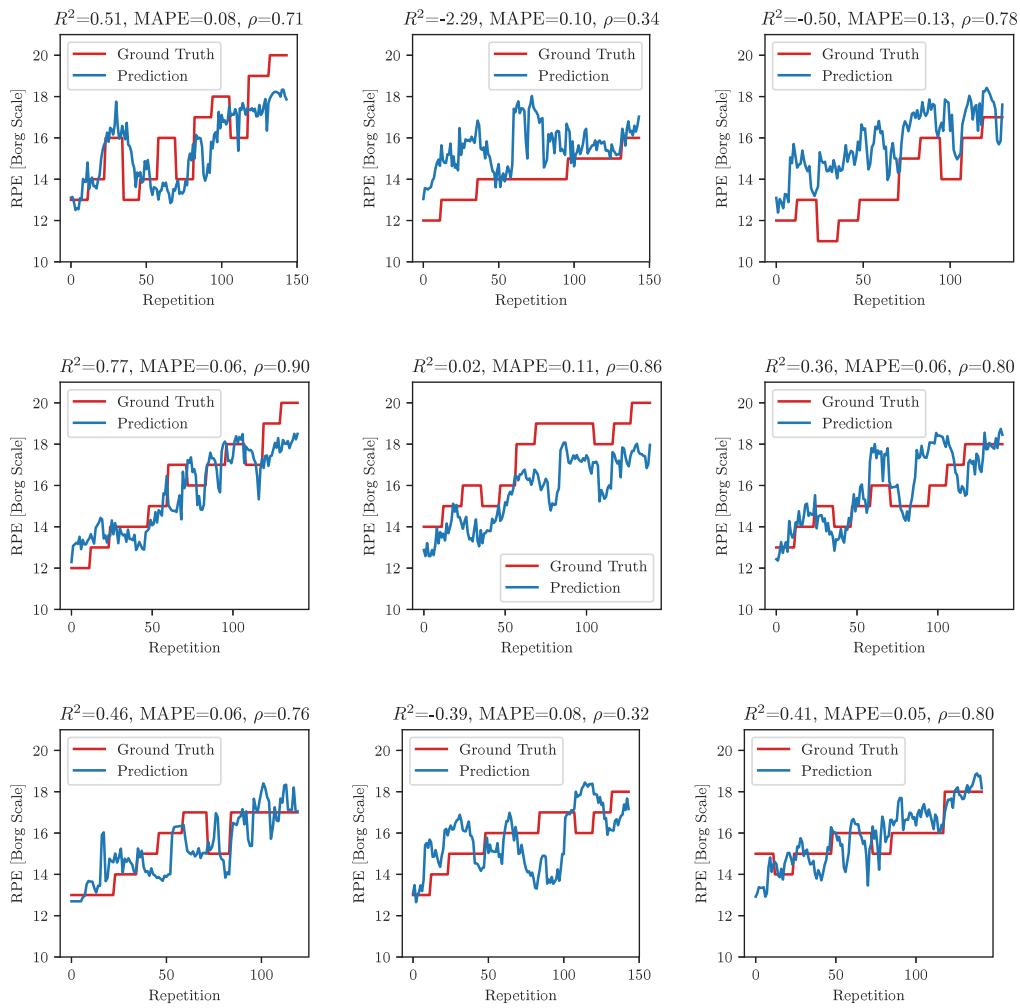


Fig. 8. Sample-based predictions from the best-performing GBRT model for nine randomly selected subjects.

Table 4

Prediction results of the four traditional ML models trained on segmented exercises and the CNN-GRU model trained on the time-series data. The GBRT model achieves the best training results.

	GBRT	MLP	RF	SVR	CNN-GRU
RMSE	1.49 ± 0.51	1.70 ± 0.46	1.58 ± 0.50	1.62 ± 0.52	2.01 ± 0.79
MAE	1.26 ± 0.46	1.43 ± 0.43	1.33 ± 0.48	1.38 ± 0.48	1.67 ± 0.71
MAPE	8.08 ± 3.16	9.17 ± 2.82	8.59 ± 3.32	8.91 ± 3.06	10.91 ± 4.96
R^2	0.17 ± 0.67	-0.08 ± 0.87	0.04 ± 0.84	0.05 ± 0.73	-0.47 ± 1.31
Spearman's ρ	0.74 ± 0.23	0.62 ± 0.27	0.73 ± 0.22	0.68 ± 0.30	0.38 ± 0.44

4.2.2. Evaluation results

The training results show that the GBRT model delivers the best performance. Table 4 presents the results of all trained models, showing the results as mean and standard deviation over all subjects. The table reveals that GBRT achieved the best results with $R^2 = 0.17 \pm 0.67$ and $RMSE = 1.49 \pm 0.51$ over all subjects. The CNN-GRU model, on the other hand, achieved the worst training results by reaching an $R^2 = -0.47 \pm 1.31$ and achieved the lowest Spearman correlation of $\rho = 0.38 \pm 0.44$. Therefore, this section evaluates the results of the best-performing GBRT model in more detail. The GBRT model achieved the best results when trained on 25 features and a temporal context of nine past repetitions.

Fig. 8 analyzes the sample-based predictions from the best-performing GBRT model in more detail for nine randomly selected subjects. The x-axis shows the individual repetitions (144 repetitions

in the optimal case) for which the model predicts the current RPE values. The ground truth line represents a step function with a single RPE value for an entire set. It becomes apparent that for all subjects, most predictions follow the trend of the reported RPE values. The model successfully estimates appropriate baselines for most subjects. However, an offset exists for some subjects, as the model has no further information about the athletes performing the exercise. Even though the model obtains low or negative $R^2 = 0.02$ values for some subjects, the model still achieves good correlations.

For completion, Fig. 9 presents Spearman's ρ for each test subject inferred with the best-performing GBRT model for all test subjects. It shows that the model performance varies across the test subjects. Seven subjects had an excellent correlation, four had a good correlation, three had a moderate correlation, and two only showed a weak correlation.

Fig. 10 further investigates the results of the GBRT model using a scatter plot (left) and residual analysis (right) of the RPE predictions for each set. The scatter and residual plots indicate data points from the same subject in the same color. The scatter plot shows ground truth on the horizontal, predictions on the vertical axis, and the theoretical ideal regression line $f(x) = x$. The regression line of the data points has a positive slope of 0.47 and a bias of 8.33, where the points form a slight cluster for intermediate RPE levels. The R^2 of the regression is $R^2 = 0.47$. The regression plot shows ground truth values on the horizontal and the difference between ground truth and predictions on the vertical axis. The analysis reveals a trend in the data, where the residuals prompt a certain distribution. It shows that for smaller RPEs

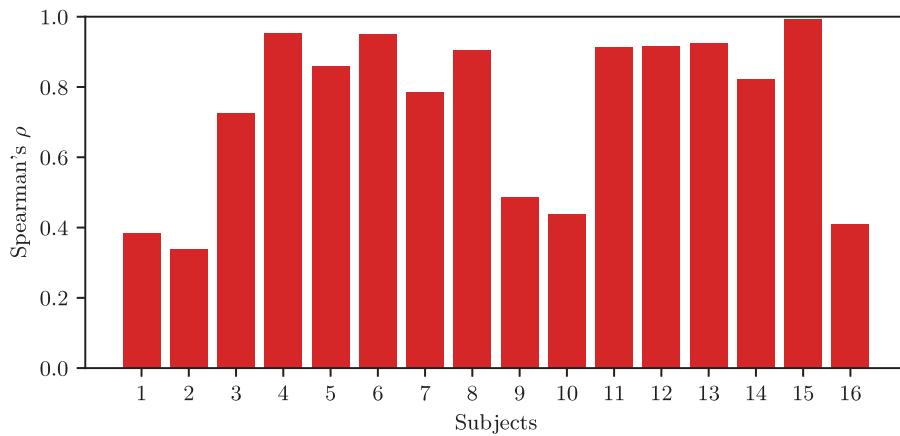


Fig. 9. Correlations between the best-performing GBRT model predictions and the ground truth RPE values.

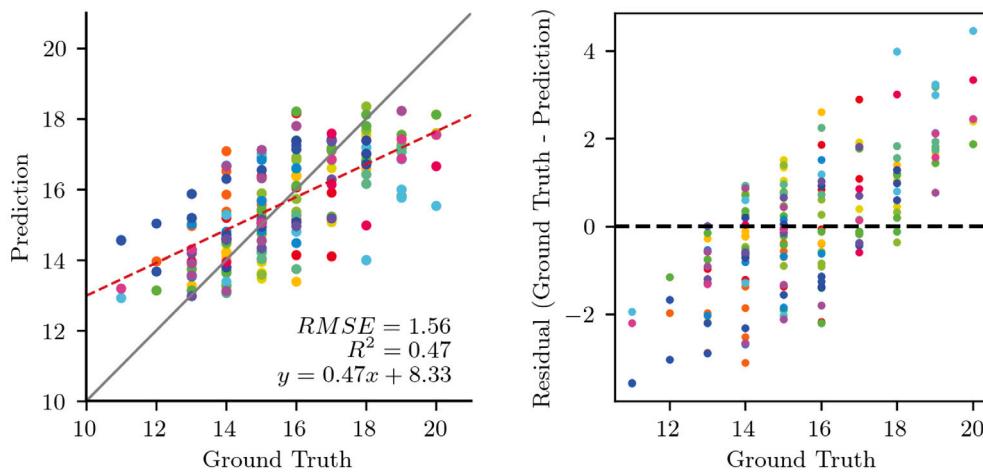


Fig. 10. Scatter-plot and residual analysis of predictions from the best-performing GBRT model. The data point colors indicate individual subjects.

(<16), the model tends to overestimate the predictions, while for higher RPEs (>16), the model tends to underestimate the RPE.

4.3. Prediction of generated power during resistance training

This section presents the evaluation of the power prediction experiments. The ground truth data was delivered by the kMeter measurement device integrated into the Flywheel training device that measures the Flywheel rotation and computes the average power for each repetition. As the power prediction does not require a temporal context, the feature set for ML does not contain the rolling statistics, and the CNN-FC model is trained.

4.3.1. Evaluation setup

As the power prediction experiment is not dependent on the spatio-temporal development of the time series but in measurements, we normalize the feature set \mathbf{X} for traditional ML using standard scaling over the entire feature set across all subjects. Similarly, the image encodings of the CNN-FC model \mathbf{I} are standard scaled for each channel independently.

4.3.2. Evaluation results

Table 5 presents the power prediction results of the four ML models, the CNN-FC model, and the physical model. The table shows that GBRT and MLP achieved the best prediction results. Judging by the highest coefficient of determination, the MLP was selected for further analysis

as the best-performing model. It obtains a $R^2 = 0.40 \pm 0.31$ and obtained the lowest MAE = 23.13 ± 8.38 . The physical model generally showed the worst performance, with an MAE of MAE = 39.41, yet showing a good correlation ($\rho = 0.75$).

Fig. 11 compares the prediction performance from the MLP, the CNN-FC, and the physical model for three randomly selected test subjects. It becomes apparent that the physical model sometimes overestimates the power values, whereas the data-driven models managed to be closer to the ground truth. Even though the physical model is off heavily, the obtained correlations are still good.

Fig. 12 investigates and compares the prediction results of the MLP, CNN-FC, and the physical model across all test subjects using scatter plots and Bland-Altman analysis. Similar to the other plots, the colors represent individual subjects. The scatter plots reveal that although the physical model achieves the best regression line, the MLP model shows the best RMSE = 31.99 and the best $R^2 = 0.82$, while the regression line of the CNN has the worst slope. The physical model only achieves a $R^2 = 0.27$. The Bland-Altman plots indicate that the MLP has the best limit of agreement with $(-66.04, 58.47)$, while the CNN-FC comes second with $(-94.41, 65.48)$, and the physical model achieves the worst agreement with $(-114.28, 133.62)$. The physical model achieves a good mean close to zero -3.79 , the CNN-FC obtains 14.46, and the physical model obtains 9.67. It also becomes apparent that the physical model overestimates one particular subject, which is present in data-driven models.

Table 5

Results of the power prediction of the four traditional ML models, the CNN-FC model, and the physical model.

	GBRT	MLP	RF	SVR	CNN-FC	PHYSICS
RMSE	29.71 ± 12.32	29.73 ± 9.89	31.96 ± 14.07	30.61 ± 11.11	36.19 ± 20.89	49.12 ± 41.62
MAE	23.80 ± 10.77	23.13 ± 8.38	25.55 ± 12.58	24.44 ± 9.80	30.61 ± 20.48	39.41 ± 38.55
MAPE	24.09 ± 12.29	24.26 ± 15.00	26.41 ± 13.98	26.01 ± 16.22	27.32 ± 12.26	34.81 ± 21.74
R^2	0.32 ± 0.61	0.40 ± 0.31	0.19 ± 0.85	0.28 ± 0.56	-0.02 ± 1.07	-0.80 ± 2.36
Spearman's ρ	0.79 ± 0.16	0.77 ± 0.17	0.78 ± 0.18	0.78 ± 0.19	0.76 ± 0.21	0.75 ± 0.19

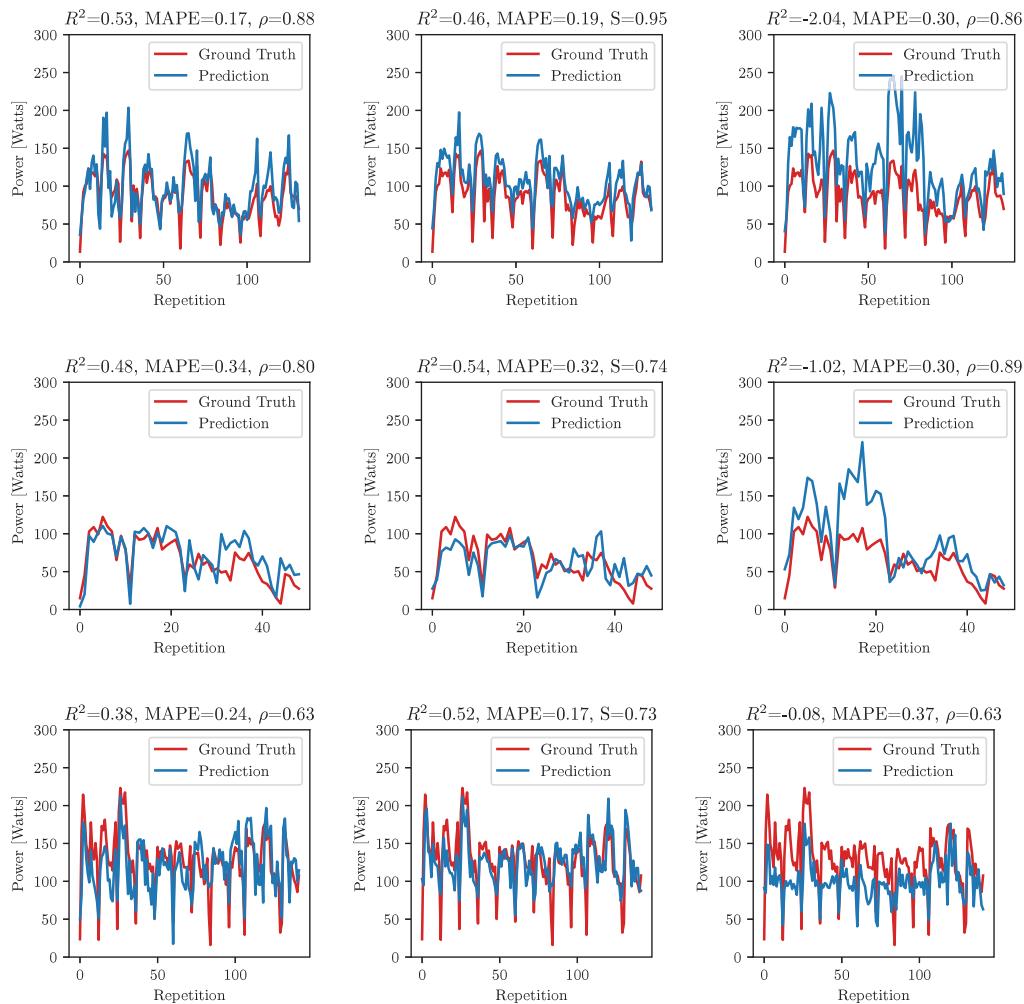


Fig. 11. Power predictions from the best-performing MLP (left column), CNN-FC (middle column), and physical model (right column) for three randomly selected subjects.

5. Discussion and future work

This paper presented a computer vision method based on human pose estimation to predict external and internal load through power and RPE during resistance training. Skeleton sequences obtained from two Microsoft Azure Kinect cameras using the Body Tracking SDK were fused to obtain robust 3D joint positions and orientations. For RPE prediction using traditional machine learning, four regression models were trained on statistical features handcrafted on segmented squat repetitions. In our deep learning approach, skeleton sequences are transformed into a suitable image-like representation to train a CNN-FC model. The RPE prediction showed promising results with good correlations of $\rho \geq 0.76$ overall test subjects. However, the RPE prediction also delivered moderate to weak outcomes for five of the 16 subjects. The results also indicated that traditional ML benefited from the handcrafted features and thus outperformed the CNN-GRU model. The second experiment concerned predicting generated power using the

same feature set. The results showed that the traditional ML models trained on the handcrafted features outperformed the CNN-FC model by achieving smaller RMSE than the other models. They also outperformed a physical model that aims to describe the power generation using a physical equation, which generally achieved the worst results. All models obtained good correlations $\rho \geq 0.75$; however, the MSE indicated that the traditional ML models delivered the most accurate results. The developed physical model most likely misses important physical variables and factors, e.g., the Flywheel belt's friction, and therefore does not appropriately describe reality, thus the lack of model accuracy.

We want to compare the results of our approach to the most related study conducted by Jiang et al. [20]. In their study, fourteen subjects performed the squat exercise and rated their exertion level on an RPE scale ranging from zero to ten. The authors evaluated their model using Pearson's correlation for all subjects individually. Their results show that 0%, 71%, 21%, and 7% of all subjects achieved

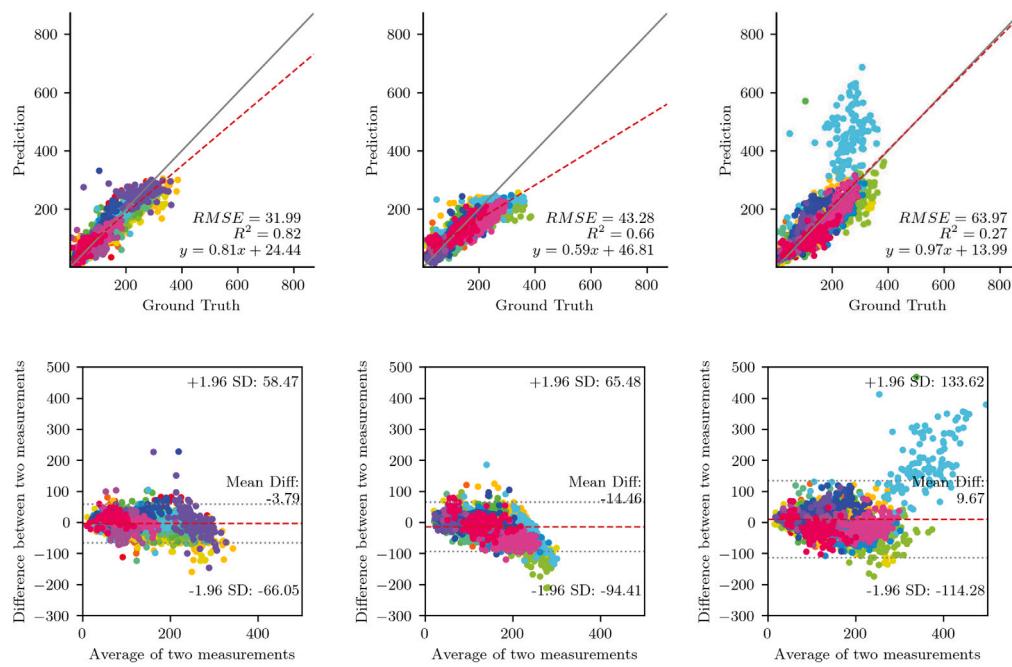


Fig. 12. Scatter plots and Bland–Altman analysis of power prediction from the best-performing MLP, the CNN-FC, and the physical model. The data point colors indicate individual subjects.

correlations categorized as excellent, strong, moderate, or weak (values do not add up to 100 due to rounding). In our approach, 44%, 25%, 25%, and 6% of subjects obtained correlations classified as excellent, strong, moderate, and weak. However, it is important to note that a direct comparison is difficult due to different metrics (Pearson's vs. Spearman), RPE scales, protocols, and sensors. Further, the results of this approach are compared to our previous study that used the same datasets but other available sensor modalities, as presented by Albert et al. [45]. This study aimed to predict RPE based on six IMUs and HRV parameters. Four machine learning models were trained on statistical features from IMU and HRV data, calculated for individual windows in a sliding window approach. The results showed that IMU data alone delivered poor results ($R^2 = 0.08 \pm 0.10$) and improved when adding HRV parameters ($R^2 = 0.52 \pm 0.06$). The Training Impulse (TRIMP) parameter significantly improved the accuracy, as this parameter is cumulative during the exercise. This computer vision approach reduced the RMSE from 4.47 and 2.23 from IMU only and IMU+HRV, respectively, to a $RMSE = 1.49$. We hypothesize that incorporating the domain knowledge into the features and adding the temporal context improved the prediction results. Finally, this approach using human pose estimation alleviates the need for wearable sensors, increasing the comfort for athletes.

This study faces certain limitations that may have impacted the findings. The Kinect camera's motion capture precision might have been influenced by environmental factors such as lighting conditions or body part occlusions, leading to potential inaccuracies in the collected skeleton data. To this end, two Microsoft Azure Kinect cameras were used to compensate for potential tracking inaccuracies of one camera by the other. Further, the sample size of the collected dataset is relatively small, which limits the generalizability of the results. Additionally, this study focuses on one specific exercise performed in a controlled laboratory environment to exclude independent experiment variables. Another limitation of this study is that RPE is a subjective quantification and requires athletes to be familiar with the RPE scale. Due to the short study time, some subjects unfamiliar with the RPE scale only had a small amount of time to learn about using the scale. Lastly, the power measurements result from the kMeter device integrated into the

Flywheel. Weakley et al. [36] validated the reliability of the kMeter sensor using a force plate gold-standard device. However, in future studies aiming to predict generated power from skeleton data, using well-established gold-standard systems such as force plates would be beneficial.

This study aims to lay the foundations for a continuous fatigue monitoring system during resistance training. Especially since the COVID-19 pandemic, training at home has become more prominent. Patients are often dismissed early from the rehabilitation center and must train at home without external supervision. Extending existing monitoring systems that provide feedback on the posture with our fatigue monitoring approach could help prevent future injuries, especially when patients are unfamiliar with using an RPE scale.

In the future, we plan on extending the conducted study. As the sample size of our dataset is relatively small and only includes a single exercise performed in a controlled environment, future work should address this issue by crafting a large, heterogeneous dataset recorded in the wild. Fatigue is also subjective, so future work should investigate personalized models trained on individuals to learn individual-specific data distributions. To this end, incorporating the subject's anthropometric information or medical history into the model is another possible research direction to obtain more accurate and individualized results.

CRediT authorship contribution statement

Justin Amadeus Albert: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Bert Arnrich:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

- [1] Stephen J. McQuilliam, David R. Clark, Robert M. Erskine, Thomas E. Brownlee, Free-weight resistance training in youth athletes: a narrative review, *Sports Med.* 50 (9) (2020) 1567–1580.
- [2] Pedro Lopez, Ronei Silveira Pinto, Regis Radaelli, Anderson Rech, Rafael Grazioli, Mikel Izquierdo, Eduardo Lusa Cadore, Benefits of resistance training in physically frail elderly: a systematic review, *Aging Clin. Exp. Res.* 30 (2018) 889–899.
- [3] Gregory C. Bogdanis, Effects of physical activity and inactivity on muscle fatigue, *Front. Physiol.* (ISSN: 1664-042X) 3 (2012) <http://dx.doi.org/10.3389/fphys.2012.00142>, URL <http://journal.frontiersin.org/article/10.3389/fphys.2012.00142/abstract>.
- [4] Michelle Karg, Gentiane Venture, Jesse Hoey, Dana Kulic, Human movement analysis as a measure for fatigue: A hidden Markov-based approach, *IEEE Trans. Neural Syst. Rehabil. Eng.* (ISSN: 1534-4320) 22 (3) (2014) 470–481, <http://dx.doi.org/10.1109/TNSRE.2013.2291327>, 1558-0210. URL <https://ieeexplore.ieee.org/document/6716986/>.
- [5] Shona L. Halson, Monitoring training load to understand fatigue in athletes, *Sports Med.* (ISSN: 0112-1642) 44 (2014) 139–147, <http://dx.doi.org/10.1007/s40279-014-0253-z>, 1179-2035. URL <http://link.springer.com/10.1007/s40279-014-0253-z>.
- [6] Gunnar Borg, Perceived exertion as an indicator of somatic stress, *Scand. J. Rehabil. Med.* (1970).
- [7] Brendan R. Scott, Grant M. Duthie, Heidi R. Thornton, Ben J. Dascombe, Training monitoring for resistance exercise: Theory and applications, *Sports Med.* (ISSN: 0112-1642) 46 (5) (2016) 687–698, <http://dx.doi.org/10.1007/s40279-015-0454-0>, 1179-2035. URL <http://link.springer.com/10.1007/s40279-015-0454-0>.
- [8] Joseph O.C. Coyne, G. Gregory Haff, Aaron J. Coutts, Robert U. Newton, Sophia Nimpithus, The current state of subjective training load monitoring—a practical perspective and call to action, *Sports Med. - Open* (ISSN: 2199-1170) 4 (1) (2018) 58, <http://dx.doi.org/10.1186/s40798-018-0172-x>, 2198-9761. URL <https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-018-0172-x>.
- [9] Yalin Liao, Aleksandar Vakanski, Min Xian, A deep learning framework for assessing physical rehabilitation exercises, *IEEE Trans. Neural Syst. Rehabil. Eng.* (ISSN: 1534-4320) 28 (2) (2020) 468–477, <http://dx.doi.org/10.1109/TNSRE.2020.2966249>, 1558-0210. URL <https://ieeexplore.ieee.org/document/8957502/>.
- [10] B.T. Crewther, L.P. Kilduff, D.J. Cunningham, C. Cook, N. Owen, G.-Z. Yang, Validating two systems for estimating force and power, *Int. J. Sports Med.* (ISSN: 0172-4622) 32 (4) (2011) 254–258, <http://dx.doi.org/10.1055/s-0030-1270487>, 1439-3964. URL <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0030-1270487>.
- [11] Youri Geurkink, Gilles Vandewiele, Maarten Lievens, Filip de Turck, Femke Ongevae, Stijn P.J. Matthys, Jan Boone, Jan G. Bourgois, Modeling the prediction of the session rating of perceived exertion in soccer: Unraveling the puzzle of predictive indicators, *Int. J. Sports Physiol. Perform.* (ISSN: 1555-0265) 14 (6) (2019) 841–846, <http://dx.doi.org/10.1123/ijssp.2018-0698>, 1555-0273. URL <https://journals.humankinetics.com/view/journals/ijssp/14/6/article-p841.xml>.
- [12] Gilles Vandewiele, Youri Geurkink, Maarten Lievens, Femke Ongevae, Filip De Turck, Jan Boone, Enabling training personalization by predicting the session rate of perceived exertion (sRPE), in: *Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2017 Workshop*, 2017, pp. 1–12.
- [13] D.L. Carey, K. Ong, M.E. Morris, J. Crow, K.M. Crossley, Predicting ratings of perceived exertion in Australian football players: methods for live estimation, *Int. J. Comput. Sci. Sport* (ISSN: 1684-4769) 15 (2) (2016) 64–77, <http://dx.doi.org/10.1515/ijcss-2016-0005>, URL <https://www.sciendo.com/article/10.1515/ijcss-2016-0005>.
- [14] Jeongbin Kim, Hyunsung Kim, Jonghyun Lee, Jaechan Lee, Jinsung Yoon, Sang-Ki Ko, A deep learning approach for fatigue prediction in sports using GPS data and rate of perceived exertion, *IEEE Access* (ISSN: 2169-3536) 10 (2022) 103056–103064, <http://dx.doi.org/10.1109/ACCESS.2022.3205112>, URL <https://ieeexplore.ieee.org/document/9881489/>.
- [15] Padraig Davidson, Peter Düking, Christoph Zinner, Billy Sperlich, Andreas Hotho, Smartwatch-derived data and machine learning algorithms estimate classes of ratings of perceived exertion in runners: A pilot study, *Sensors* (ISSN: 1424-8220) 20 (9) (2020) 2637, <http://dx.doi.org/10.3390/s20092637>, URL <https://www.mdpi.com/1424-8220/20/9/2637>.
- [16] Bjoern Eskofier, Patrick Kugler, Daniel Melzer, Pascal Kuehner, Embedded classification of the perceived fatigue state of runners: Towards a body sensor network for assessing the fatigue state during running, in: *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, IEEE, ISBN: 978-0-7695-4698-8, 2012, pp. 113–117, <http://dx.doi.org/10.1109/BSN.2012.4>, 978-1-4673-1393-3. URL <https://ieeexplore.ieee.org/document/6200548/>.
- [17] Swapnali Karvekar, Masoud Abdollahi, Ehsan Rashedi, Smartphone-based human fatigue level detection using machine learning approaches, *Ergonomics* (ISSN: 0014-0139) 64 (5) (2021) 600–612, <http://dx.doi.org/10.1080/00140139.2020.1858185>, 1366-5847. URL <https://www.tandfonline.com/doi/full/10.1080/00140139.2020.1858185>.
- [18] Mohsen Gholami, Christopher Napier, Astrid García Patiño, Tyler J. Cuthbert, Carlo Menon, Fatigue monitoring in running using flexible textile wearable sensors, *Sensors* (ISSN: 1424-8220) 20 (19) (2020) 5573, <http://dx.doi.org/10.3390/s20195573>, URL <https://www.mdpi.com/1424-8220/20/19/5573>.
- [19] Tim Op De Beeck, Wannes Meert, Kurt Schütte, Benedicta Vanwanseele, Jesse Davis, Fatigue prediction in outdoor runners via machine learning and sensor fusion, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, ISBN: 978-1-4503-5552-0, 2018, pp. 606–615, <http://dx.doi.org/10.1145/3219819.3219864>, URL <https://dl.acm.org/doi/10.1145/3219819.3219864>.
- [20] Yanran Jiang, Vincent Hernandez, Gentiane Venture, Dana Kulić, Bernard K. Chen, A data-driven approach to predict fatigue in exercise based on motion data from wearable sensors or force plate, *Sensors* (ISSN: 1424-8220) 21 (4) (2021) URL <https://www.mdpi.com/1424-8220/21/4/1499>.
- [21] Igor Pernek, Gregorij Kurillo, Gregor Stiglic, Ruzena Bajcsy, Recognizing the intensity of strength training exercises with wearable sensors, *J. Biomed. Inform.* (ISSN: 1532-0464) 58 (2015) 145–155, <http://dx.doi.org/10.1016/j.jbi.2015.09.020>, URL <https://www.sciencedirect.com/science/article/pii/S1532046415002142>.
- [22] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake, Real-time human pose recognition in parts from single depth images, in: *CVPR 2011*, IEEE, ISBN: 978-1-4577-0394-2, 2011, pp. 1297–1304, <http://dx.doi.org/10.1109/CVPR.2011.5995316>, URL <http://ieeexplore.ieee.org/document/5995316>.
- [23] Daphne J. Geerse, Bert H. Coolen, Melvyn Roerdink, Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments, *PLOS ONE* (ISSN: 1932-6203) 10 (10) (2015) e0139913, <http://dx.doi.org/10.1371/journal.pone.0139913>, URL <https://dx.plos.org/10.1371/journal.pone.0139913>.
- [24] Moataz Eltoukhy, Adam Kelly, Chang-Young Kim, Hyung-Pil Jun, Richard Campbell, Christopher Kuenze, Validation of the Microsoft Kinect® camera system for measurement of lower extremity jump landing and squatting kinematics, *Sports Biomech.* (ISSN: 1476-3141) 15 (1) (2016) 89–102, <http://dx.doi.org/10.1080/14763141.2015.1123766>, 1752-6116. URL <http://www.tandfonline.com/doi/full/10.1080/14763141.2015.1123766>.
- [25] Ross A. Clark, Yong-Hao Pua, Karine Fortin, Callan Ritchie, Kate E. Webster, Linda Denhey, Adam L. Bryant, Validity of the Microsoft Kinect for assessment of postural control, *Gait Posture* (ISSN: 09666362) 36 (3) (2012) 372–377, <http://dx.doi.org/10.1016/j.gaitpost.2012.03.033>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0966636212200128>.
- [26] Michał Wieczorek, Jakub Siłka, Marcin Woźniak, Sahil Garg, Mohammad Mehedi Hassan, Lightweight convolutional neural network model for human face detection in risk situations, *IEEE Trans. Ind. Inform.* 18 (7) (2022) 4820–4829, <http://dx.doi.org/10.1109/TII.2021.3129629>.
- [27] Marcin Woźniak, Jakub Siłka, Michał Wieczorek, Deep learning based crowd counting model for drone assisted systems, in: *Proceedings of the 4th ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and beyond, DroneCom '21*, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450387057, 2021, pp. 31–36, <http://dx.doi.org/10.1145/3477090.3481054>.
- [28] Dushyanti Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, Christian Theobalt, VNect: Real-time 3D human pose estimation with a single RGB camera, *ACM Trans. Graph.* 36 (2017) <http://dx.doi.org/10.1145/3072959.3073596>, URL <http://vgv.mpi-inf.mpg.de/projects/VNect/>.
- [29] Justin Amadeus Albert, Victor Owolabi, Arnd Gebel, Clemens Markus Brahm, Urs Granacher, Bert Arnrich, Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study, *Sensors* (ISSN: 1424-8220) 20 (18) (2020) <http://dx.doi.org/10.3390/s20185104>, URL <https://www.mdpi.com/1424-8220/20/18/5104>.
- [30] Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Ville Kyrik, Andrea Monterù, Luca Romeo, Federica Verdini, A Hidden Semi-Markov Model based approach for rehabilitation exercise assessment, *J. Biomed. Inform.* (ISSN: 1532-0464) 78 (2018) 1–11, <http://dx.doi.org/10.1016/j.jbi.2017.12.012>, URL <https://www.sciencedirect.com/science/article/pii/S1532046417302824>.
- [31] Annalisa Franco, Antonio Magnani, Dario Maio, Joint orientations from skeleton data for human activity recognition, in: *Image Analysis and Processing-ICIAP 2017: 19th International Conference*, Catania, Italy, September 11-15, 2017, *Proceedings, Part I 19*, Springer, 2017, pp. 152–162.
- [32] Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey, Shaik Ali Akbar, Skeleton-based human activity recognition using ConvLSTM and guided feature learning, *Soft Comput.* (ISSN: 1432-7643) 26 (2) (2022) 877–890, <http://dx.doi.org/10.1007/s00500-021-06238-7>, 1433-7479. URL <https://link.springer.com/10.1007/s00500-021-06238-7>.

- [33] Guoyang Zhou, Vaneet Aggarwal, Ming Yin, Denny Yu, A computer vision approach for estimating lifting load contributors to injury risk, *IEEE Trans. Hum.-Mach. Syst.* (ISSN: 2168-2291) 52 (2) (2022) 207–219, <http://dx.doi.org/10.1109/THMS.2022.3148339>, 2168-2305. URL <https://ieeexplore.ieee.org/document/9718506/>.
- [34] Xianta Jiang, Mohsen Gholami, Mahta Khoshnam, Janice J. Eng, Carlo Menon, Estimation of ankle joint power during walking using two inertial sensors, *Sensors* 19 (12) (2019) 2796.
- [35] Md Zia Uddin, Trine M. Seeberg, Jan Kocabach, Anders E. Liverud, Victor Gonzalez, Øyvind Sandbakk, Frédéric Meyer, Estimation of mechanical power output employing deep learning on inertial measurement data in roller ski skating, *Sensors* (ISSN: 1424-8220) 21 (19) (2021) <http://dx.doi.org/10.3390/s21196500>, URL <https://www.mdpi.com/1424-8220/21/19/6500>.
- [36] Jonathon Weakley, Bruno Fernández-Valdés, Liam Thomas, Carlos Ramirez-Lopez, Ben Jones, Criterion validity of force and power outputs for a commonly used flywheel resistance training device and bluetooth app, *J. Strength Cond. Res.* 33 (5) (2019) 1180–1184.
- [37] Justin Amadeus Albert, Arne Herdick, Clemens Markus Brahms, Urs Granacher, Bert Arnrich, PERSIST: A multimodal dataset for the prediction of perceived exertion during resistance training, *Data* (ISSN: 2306-5729) 8 (1) (2023) <http://dx.doi.org/10.3390/data8010009>, URL <https://www.mdpi.com/2306-5729/8/1/9>.
- [38] Sungphil Moon, Youngbin Park, Dong Wook Ko, Il Hong Suh, Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering, *Int. J. Adv. Robot. Syst.* 13 (2) (2016) 65.
- [39] Karolis Ryselis, Tautvydas Petkus, Tomas Blažauskas, Rytis Maskeliūnas, Robertas Damaševičius, Multiple Kinect based system to monitor and analyze key performance indicators of physical training, *Hum.-Centric Comput. Inf. Sci.* 10 (1) (2020) 1–22.
- [40] Qing-Jun Xing, Yuan-Yuan Shen, Run Cao, Shou-Xin Zong, Shu-Xiang Zhao, Yan-Fei Shen, Functional movement screen dataset collected with two azure kinect depth sensors, *Sci. Data* 9 (1) (2022) 1–17.
- [41] J. Sarsfield, D. Brown, N. Sherkat, C. Langensiepen, J. Lewis, M. Taheri, L. Selwood, P. Standen, P. Logan, Segmentation of exercise repetitions enabling real-time patient analysis and feedback using a single exemplar, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (5) (2019) 1004–1019, <http://dx.doi.org/10.1109/TNSRE.2019.2907483>.
- [42] A. Bevilacqua, B. Huang, R. Argent, B. Caulfield, T. Kechadi, Automatic classification of knee rehabilitation exercises using a single inertial sensor: A case study, in: 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2018, pp. 21–24, <http://dx.doi.org/10.1109/BSN.2018.8329649>.
- [43] Claudio Coppola, Diego R. Faria, Urbano Nunes, Nicola Bellotto, Social activity recognition based on probabilistic merging of skeleton features with proximity priors from RGB-D data, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, pp. 5055–5061, <http://dx.doi.org/10.1109/IROS.2016.7759742>.
- [44] Md Atiqur Rahman Ahad, Masud Ahmed, Anindya Das Antar, Yasushi Makihara, Yasushi Yagi, Action recognition using kinematics posture feature on 3D skeleton joint locations, *Pattern Recognit. Lett.* (ISSN: 01678655) 145 (2021) 216–224, <http://dx.doi.org/10.1016/j.patrec.2021.02.013>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0167865521000751>.
- [45] Justin Amadeus Albert, Arne Herdick, Clemens Markus Brahms, Urs Granacher, Bert Arnrich, Using machine learning to predict perceived exertion during resistance training with wearable heart rate and movement sensors, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, pp. 801–808, <http://dx.doi.org/10.1109/BIBM52615.2021.9669577>.
- [46] Junjie Chen, Wei Yang, Chenqi Liu, Leiyu Yao, A data augmentation method for skeleton-based action recognition with relative features, *Appl. Sci.* (ISSN: 2076-3417) 11 (23) (2021) <http://dx.doi.org/10.3390/app112311481>, URL <https://www.mdpi.com/2076-3417/11/23/11481>.
- [47] Leslie G. Portney, Foundations of Clinical Research: Applications to Evidence-Based Practice, FA Davis, 2020.