2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)

Regression-based, mistake-driven movement skill estimation in Nordic Walking using wearable inertial sensors

Adrian Derungs[1,2], Sebastian Soller[2], Andreas Weishäupl[3], Judith Bleuel[3], Gereon Berschin[3], Oliver Amft[1,2]
[1] ACTLab, Chair of eHealth and mHealth, FAU Erlangen-Nürnberg, Germany
[2] ACTLab, Chair of Sensor Technology, University of Passau, Germany
[3] Sports centre, University of Passau, Germany
{adrian.derungs, oliver.amft}@fau.de

***Abstract*—We propose a mistake-driven skill estimation approach for movement analysis in sports using wearable inertial measurement units (IMUs) and continuous regression models. As a motion-quality oriented sport, we focus on Nordic Walking. Nordic Walking involves complex body part interactions, which when wrongly performed, increase injury risk and reduce training effectiveness. We present an approach to assess three mistakes related to health risk which are typical in beginners. Based on a stride segmentation, features relevant to the mistakes were extracted and selected in a dedicated feature selection step. Subsequently, Bayesian Ridge Regression (BRR) and alternative regression models were trained for each mistake type. Models were evaluated in our pattern analysis architecture supporting parallel and continuous step-wise estimation of all mistakes in movement skill grades ranging from 1 (correct) to 3 (incorrect). We evaluated our skill estimation approach in a study including 10 Nordic Walking beginners and 11247 expert-annotated strides derived from 50 recording sessions. We investigate seven practical wearable sensor placement configuration using leave-one-participant-out (LOPO) cross-validation. Results showed that all mistakes can be estimated with an normalised RMSE of 24.15 % across all participants. Additional analysis of trends suggest that participants could improve skilfulness in training sessions during our study.**

I. Introduction

Assessing movement quality and skilfulness are vital for injury prevention, performance tracking, and training optimisation. Wearable inertial motion units (IMUs) have been frequently deployed to identify motion or determine cumulative performance statistics, such as counting motion repetitions. Estimating movement skills from wearable sensors remains an open research challenge. In many motion-focused sports, experts consider movement quality as a multi-dimensional problem and identify different mistakes by observation. In contrast to a binary decision over the complete execution into correct or incorrect, detecting mistakes could be supplemented by skill grades per mistake. Identified mistakes enable athletes to improve by focusing on specific training strategies. Mistakes, however, may involve various body parts. For example, Nordic Walking involves complex body part interactions that, when executed correctly, maximise the cardiovascular training effect. Several mistakes, often made by beginners lead to injury [1]. An observable mistake in Nordic Walking includes the dragging of poles over ground. Although, dragging of poles might not appear as health critical, characteristic Nordic Walking motion patterns could be influenced reducing the cardiovascular training effect. Wearable IMUs could provide a continuous training assistance complementing coaches during free exercise. We focus in this investigation on the estimation of skilfulness. As there are no established assessments for Nordic Walking, we begin by defining mistake types and a skill grading scheme, and subsequently evaluating the skill estimation against experts. We aim to combine expert knowledge, wearable sensor technology and machine learning algorithms, demonstrating feasibility of a mistake-driven skill estimation approach.

At the expense of detail, skill grades may be considered categorical and typical pattern classification methods could be applied. However, skill improvement is gradual thus a categorical modelling would mask minor changes around a skill grade. Instead, we chose a regression-based modelling approach to denote skill grades per mistake type. Within regression methods, the estimation of feature relevance is however more complex than in classification problems. We propose here a preprocessing to determine initial feature relevance by considering the skill grades as categorical variables.

Due to the full-body movement pattern in Nordic Walking, various sensor placements appear relevant to analyse movements. Here we investigate seven practical sensor configurations, each considering a limited set of four IMU positions at the body and walking poles.

In particular, this paper provides the following contributions:

1) We introduce a mistake-driven skill grade estimation method for Nordic Walking. Our method utilises individual *Bayesian Ridge Regression* models for each mistake, estimating skills stride-by-stride using motion data derived from wearable IMUs.
2) We evaluate our skill estimation approach in a participant study including 10 Nordic Walking beginners and compare seven practical sensor placement configurations. A recording and annotation setup is proposed and the process to obtain stride-by-stride expert annotations for each mistake type and skill grade is described.
3) We present a categorisation of three basic, health related mistake types for Nordic Walking and a skill grading scheme that provides insight into mistakes while minimising the risk of information overload.

978-1-5386-3224-6/18/$31.00

II. Related work

Most skill level and motion performance analyses originated from sports and movement rehabilitation domains. Interestingly, Nordic Walking was barley considered so far. Nordic Walking was mentioned by Pärkkä *et al.* as one of several activities classified by wearable sensor data [2], but no further motion analysis was made. Nordic Walking has become a common cardiovascular training method among recreational athletes to increase general health and well-being. The complex motion, involving diverse interactions of body parts however, needs continuous training to avoid injuries and maximise training effect. Kocur *et al.* found Nordic Walking not only being effective for healthy individuals, but also for patients [3], rendering Nordic Walking interesting in rehabilitation scenarios, i.e. in stroke rehabilitation. Subsequently we highlight skill assessment approaches used in sports and movement rehabilitation.

A. Skill assessment in sports

Various state and performance analysis methods have been proposed and used in different sports applications. For example, Kunze *et al.* investigated Tai Chi movements using gyroscopes and accelerometer [4]. Three basic forward and backward movements were selected to analyse differences between experts and amateurs. Based on the results of four analysed participants the authors found that rotational energy was a key feature to separate experts from beginners. Ahmadi *et al.* used three wearable accelerometers and gyroscopes, placed on different body locations to assess first serve skills in tennis [5]. The authors clustered first serves of four players using three distinct motion features: angular velocity of the upper arm internal rotation, wrist flexion, and shoulder rotation. For new serves, a qualitative analysis of the clusters was used to distinguish between amateurs, sub-elite, and elite players. Bächlin *et al.* developed a wearable assistant for swimmers, including acceleration sensors to continuously monitor and evaluate swimmers performance parameter (e.g. time per lane, velocity, and strokes per lane) to increase stroke technique efficiency [6]. The authors describe differences in specific performances using two parameters (body balance and body rotation) to differentiate between occasional, recreational, and elite swimmers. Although feedback to adapt swimming style to increase stroke efficiency was provided to the swimmer, the authors focused their investigation on the feedback type (visual, auditory, tactile) to be best recognised by the swimmer. Outdoor running performance was analysed by Strohrmann *et al.* [7] using a support vector machine to analyse different skill groups which were determined according their maximum running speed.

Comparing performances between different skill levels could show differences in motion patterns. While comparing performances to an expert level may reveal ways to improve, elite athletes have very specific strategies that do not suit other individuals.

Estimating overall skilfulness according to a set of defined performance attributes was investigated by Ladha *et al.*. The authors investigated climber performance using four core skills: power, control, stability and speed [8]. Core skills were derived using a Restricted Boltzman Machine to capture characteristic wrist motion patterns from two wrist-worn accelerometers. Those core skills were subsequently merged to predict an overall climbing score using a linear regression and compared to an experts' evaluated reference score. The authors showed that core skills could be combined to estimate overall skilfulness. Improvement or changes in core skills remain hidden. Michahelles *et al.* [9] showed how wearable sensors (force sensitive resistors, accelerometer, gyroscopes) combined with camera footage could be used to analyse movement of alpine skiers. Skiers' movement parameter including movement dynamics, foot force, velocity, and ski edging angles were investigated to identify strengths and weaknesses. Movement parameters and camera footage were combined using a visualisation software and provided to coaches to support their subjective assessment of skiers. Although combining different sensor modalities, the proposed system did not involve an automatic performance assessment or skill estimation. Rawashdeh *et al.* [10] proposed an IMU to count potential harmful overhead motion executions to prevent an overuse of the shoulder joint. Injury prevention strategies remain unclear.

The above investigations showed that skill levels could be differentiated in various sports. However, often cumulative quantification were used to denote the difference between levels based on motion performance, using classification or clustering approaches. Using wearable sensors and machine learning techniques to assess motion execution and potential health risks remains challenging. In particular, combining experts knowledge and skill estimation algorithms to provide insights in motion mistakes is crucial.

B. Skill assessment in movement rehabilitation

In movement rehabilitation, functional motion assessments were used to provide information about a patient's health state and potential recovery trend. Many investigations focused on estimating a clinical motor function assessment score from selected assessment items, e.g. when a patient is asked to grasp a cup from a table. The assessments typically describe a patient's ability to perform certain movements using therapy-specific descriptions as "patient needs full support", "patient needs no support". Movement speed descriptors include "slow", "normal", "fast", that were mapped into ordinal scores. As complete motion sequences are evaluated, motor assessment scores remain too coarse to directly map to improvement strategies. For example, Bento *et al.* analysed movement from selected items of the Wolf Motor Function Test with a decision tree classifier to classify movement quality and mapped the output in a value corresponding to the assessments grading scheme [11]. At each branch of the tree, decision classes were assigned based on descriptions: "Task completed", "Detected movement in involved arm", "Detected movement in non-involved arm", etc. Although the item is modelled by a decision tree, the movement description is binary. Del Din *et al.* used wearable sensors and a Random Forest algorithm including 100 trees to predict clinical scores of the Fugl-Meyer Assessment [12]. The authors determined which items (subset of assessment tasks) were best suited to predict the clinical score of the complete assessment. First approaches to quantify patients movement beyond clinical assessments were made using wearable IMUs and smartphones [13], [14].

So far, skills were not understood as mistakes given a particular, complex movement pattern, as it is the approach of our current work. Our work aims at estimating Nordic Walking skills regarding three typical mistakes in beginners with a minimal sensor setup using a regression-based approach. Our approach considers fine grained changes in motion execution, enabling Nordic Walkers to monitor motion performance and potential trends. The complex movements and the interactions of muscle groups and joints during Nordic Walking are not well investigated and described in literature and no Nordic Walking data are available for analysis.

III. Skill Estimation Approach

The Nordic Walking motion patterns have close relation to cross-country skiing and is therefore often considered as summer training by elite cross-country skiers. Following expert advise and skiing association recommendations of the Deutscher Skiverband (DSV [1]), we devised a catalogue of three health related mistakes typically observed in Nordic Walking beginners according Table I. Regularly occurring motion mistakes could harm elbow and wrist joints, reduce walking efficiency and subsequent cardio-vascular benefits promoted by Nordic Walking.

TABLE I. Description of our mistake-driven movement skills estimation approach including mistakes MS1, MS2, and MS3. Mistakes are graded with skill levels ranging from 1 (correct) to 3 (incorrect).

MS	Title	Mistake description / Skill grading
MS1	Matching gear	A correct matching gear is defined by a diagonal movement, where the right pole simultaneously hits the ground with the left foot and vice versa. MS1 either appears or not, thus graded with 1 (correct) or 3 (incorrect).
Recommendation		Focus on natural walking where arms swing naturally asymmetrical to feet.
MS2	Pole use	A correct pole usage involves establishing ground contact in the fore swing and opening of the handles in the back swing. An incorrect pole use is characterised by dragging poles over ground, missing ground contact, and continuous holding of handles. Skill grading ranges from 1 (correct) to 3 (incorrect).
Recommendation		Try to stick poles actively in the ground during fore swing and open hand after back swing.
MS3	Arm swing	A correct arm swing involves the shoulder joint to reduce stress on elbow and wrist. An incorrect arm swing involves increased elbow movement and lacks shoulder motion. Skill grading ranges from 1 (correct) to 3 (incorrect).
Recommendation		Try to stretch arms towards full extension in fore- and back-swing phases.

[1] http://www.deutscherskiverband.de

We used wearable IMUs in our skill estimation approach to identify body and limb motion and coordination between body parts. Skills were described by defining mistakes, analysed on a stride-by-stride resolution. The architecture of our mistake-driven movement skill estimation approach is illustrated in Figure 1. Calibrated IMU samples were initially preprocessed to denoise and resample data. Subsequently, Nordic Walking data were segmented into individual strides using a hill-climbing algorithm. Nordic Waking experts provided reference skill grades for each stride, based on reviewing footage recorded in the evaluation study, see details in Section V. After selecting relevant features, regression models were trained for each mistake type. Regression models aimed to estimate skill grades corresponding to the expert evaluation.

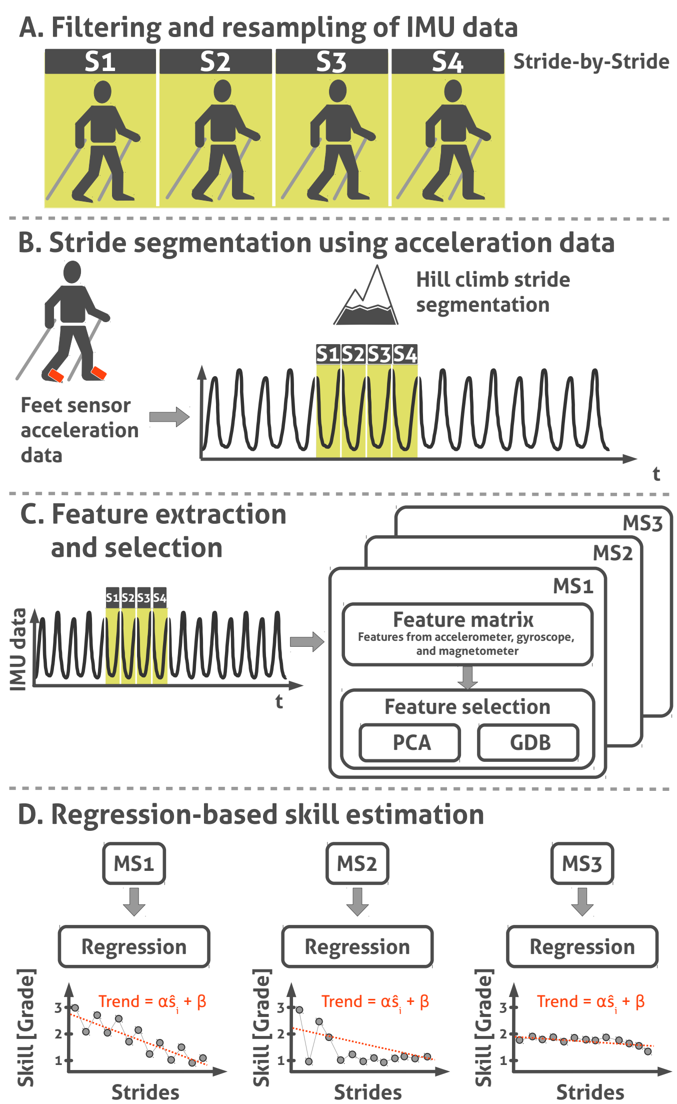

Fig. 1. Architecture of our mistake-driven movement skill estimation approach. (A.) Filtering and resampling of IMU data: IMU data are time-synchronised to align time stamps of all IMUs. (B.) Stride segmentation using IMU acceleration data: stride segmentation was performed using acceleration data from feet mounted IMUs. (C.) Feature extraction and selection: features were derived from accelerometer, gyroscope and magnetometer sensors. Principal Component Analysis (PCA) and Gradient Descent Boosting (GDB) were used for feature selection for each mistake type. (D.) Regression-based skill estimation: for each mistake type individual regression models were derived to estimate skill grades.

IV. Implementation

In this section, we detail the implementation of our mistake-driven movement skill estimation approach. In particular, we describe data preprocessing and stride segmentation, feature extraction and selection, regression-based skill estimation, and trend analysis. The framework was implemented using Python's machine learning library scikit-learn [2].

A. Data preprocessing

The initial data processing included time-synchronisation of the inertial sensor data streamed by two Xsens XBus Master and recorded with two Bluetooth enhanced notebooks to ensure correct aligned time stamps of all inertial sensor data.

B. Stride segmentation and skill annotation

Acceleration data derived from feet mounted sensors were low-pass filtered to subsequently segment individual strides. The cut-off frequency for the low-pass filter was calculated using a *Fast Fourier Transformation* to determine the dominant frequency in the acceleration signal. The cut-off frequency was derived for each participant and recording session separately. A *Hill-Climb* algorithm was subsequently used to segment individual strides [14]. Visual evaluation of the automatic stride segmentation (time stamp of the heel strike) compared with the manually segmented video data confirmed accurate match, thus no further correction for alignment were required.

C. Feature extraction and selection

We derived stride-by-stride features from IMUs' accelerometers, gyroscopes, and magnetometers x-, y, and z-axis. In addition, we derived stride-by-stride features from median-filtered raw data and the first derivative to capture potential movement pattern characteristics. Extracted features included: *min, max, mean, standard deviation, energy, percentiles, skew, min-, maxposition of peak, and zero-crossing*. For each stride the feature vector derived form one IMU included 351 features. We implemented two feature selection strategies; *Principal Component Analysis* (PCA) and *Gradient Descent Boosting* (GDB). The *PCA* [15], [16] was limited to three principal components. For the GDB algorithm [17] we incremented the number of features up to 25 with a least square loss optimisation function investigating estimation performance. A *Decision Fusion* (DF) was used to maximise skill estimation accuracy, combining estimated skill grades derived via PCA and GDB [18].

D. Continuous skill estimation

Mistakes were estimated on each individual stride using regression algorithms. We evaluated and compared different regression-based skill estimation models, maximising estimation performance while considering standard regression evaluation metrics, see Section V-E. The *Bayesian Ridge Regression* (BRR) [19] was implemented accounting for potential variation in motion execution across study participants and minimising collinearity between features. In addition, we estimated skills using *Ordinary Least Square* (OLS-linear regression), *Support Vector Regression* (SVR) and *AdaBoostR* (AdaBoost.R2 algorithm [20]). The AdaBoostR included a decision tree regressor as base estimator with a linear loss function [21]. OLS, SVR and AdaBoostR were evaluated in combination with the GDB-feature selection algorithm.

[2]http://scikit-learn.org

E. Skill trends

Potential skill development in Nordic Walking beginners were analysed using a regression over the estimated skills, i.e. each mistake type. We expected that Nordic Walking beginners would improve their skilfulness gradually over time, hence a linear regression model could describe the trend (**T**) according Eq. 1:

$$\mathrm{T} = \alpha \hat{\mathrm{s}}_{\mathrm{i}} + \beta \tag{1}$$

where parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the regression models' slope and offset; $\hat{\mathrm{s}}$ refers to the estimated skill grade derived from each stride, denoted with the index i. Improvement in skilfulness would consequently correspond to a negative regressions slope. We assessed the regression slope as either improvement or deterioration.

V. Nordic Walking Evaluation Study

A. Participants

We recruited 10 Nordic Walking beginners with no Nordic Walking experience (9 males, average 26.4 years, ranging from 18 to 34 years). All participants were introduced to the study protocol and signed a written consent form before recordings began. Being an observational study we did not advice participants in walking style before the initial recording session. However, if participants would constantly keep performing the same mistakes during the first half of a recording session, instructions for motion correction were provided by the Nordic Walking expert to prevent injuries.

B. Study design

Five recording sessions were distributed over several days for each participant. To maintain constant environmental conditions, an indoor walking strip was arranged in a large university sports hall. Participants were asked to walk 15 min per session, with a break for sensor checks after 7 min, resulting in 75 min of movement data per Nordic Walking beginner. The floor was prepared with a fitness mat to provide sufficient grip for the walking poles' tips and to avoid hall floor damages. The Nordic Walking strip was prepared according to expert recommendations with acceleration, evaluation, and deceleration sections, allowing participants to get to a constant movement pattern in the evaluation section. The strip design is illustrated in Figure 2. Two video cameras were positioned to record frontal and side view of participants. The footage was later used for offline skill annotations by the experts.

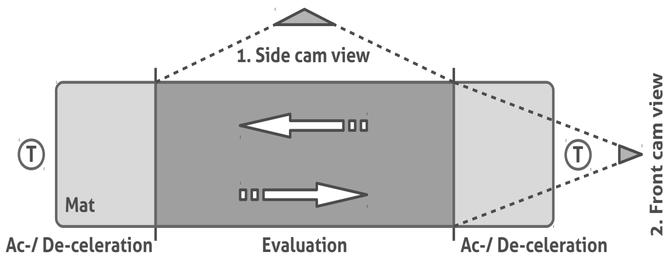

Fig. 2. Schema of the Nordic Walking strip set up for study recordings. Positions (1.) and (2.) indicate viewing angles of video cameras. Acceleration/deceleration sections were used to speed up/slow down before turning points (T). Participants were asked to maintain a convenient and constant Nordic Walking speed in the evaluation section. Each stride, recorded in the evaluation section, was offline annotated by the expert using ELAN.

C. Data recordings

Nordic Walking motion execution and performances were recorded using totally 14 IMUs (Xsens, MTx) providing inertial data (accelerometer, gyroscope, and magnetometer) in three sensor axes and orientation estimates in quaternion coordinates. The inertial sensors were symmetrically attached to a training suit using Velcro straps: wrists (RLA, LLA), upper arms (RUA, LUA), bridge of feet (RUF , LUF), shins (RLL, LLL), thighs (RUL, LUL), below Nordic Walking pole handles (RNP, LNP), as well as one sensor at the upper back (STE), and one at the lower backs' centre of mass (CEN), see Figure 3. IMUs were wired to two Xsens XBus Masters worn at the lower back. Measurements were streamed via Bluetooth to notebook computers. IMU sampling rate was 50 Hz.

Fig. 3. Example of a Nordic walker with sensor suit. Velcro strips were used for sensor attachment. A data collection box was worn on the rear. The setup was extensively tested to minimise any influences on the Nordic Walking performance due to the sensors.

To synchronise sensor data and video recordings, custom made flash light devices were attached to the ankles and connected to a force sensitive resistor sensor at the shoes' insole, thus visually indicating heel strikes of each leg in the video recordings. Each participant was asked to jump three times landing on their heels at begin and end of the recording session. The pole length was adjusted for each participant to ensure that the Nordic Walking technique was not influenced by the participants' height. While standing still, participants maintained a 90 ° angle between upper and lower arm to adjust the telescopic poles to the participants height.

D. Mistake annotations

Time-synchronised IMU and video data were annotated using the ELAN toolkit (The Language Archive [3]) [22]. ELAN was used to display both front and side-view video recordings. Each stride was manually segmented in the aligned data, as strides were observable from the flash light when heels made ground contact.

One expert (professional ski cross athlete with 15 years of experience as Universiade participant and Continental-Cup competitor) evaluated and assessed motion skills, hence provided grading reference for our skill estimation approach. To avoid inter-rater bias, a second expert randomly selected 25 recordings (50 % of all recorded footages) from all participants and annotated each stride considering all mistakes. Annotations involved identifying all mistakes on a three grade scale (1: no mistakes, 2: minor mistake, and 3: health-critical, i.e. mistake causing potential injury). The three-level grading was proposed by the sports experts involved in the project to capture skilfulness. One recording session lasting 15 min required approximately 3 h for mistake annotation by the expert and included up to 200 strides, depending on the participants' stride length. In total, 50 sessions from 10 beginners were annotated.

E. Evaluation Methodology

Our evaluation methodology included a leave-one-participant-out (LOPO) cross-validation (CV) to investigate generalisation of the proposed mistake-driven movement skills estimation approach. In addition, we defined seven sensor placement configurations hat were considered for the movement skill estimation, see Figure 4. Placement configurations were motivated considering potential mobile analysis on a smartphone and a limited amount of sensors. We derived skill grade estimation for each sensor placement configuration using the LOPO-CV. For each sensor configuration we derived stride-by-stride movement features according the stride segmentation described in Section IV-B. For the regression-based skill estimation we included strides recorded while Nordic Walking participants passed the evaluation strip, covered by the side cam (see Figure 2).

To benchmark feature extraction and selection approaches and final skill estimation, we created a baseline by sampling uniformly distributed values as input for the BRR. To analyse resulting estimation variances, we repeated the benchmarking 25 times.

We evaluated the continuous regression-based skill estimation performance by computing standardised evaluation metrics including the root-mean-square-error (RMSE), normalised RMSE (nRMSE), and the mean-absolute-error (MAE) [23]. RMSE and MAE express the estimation error in skill grades, hence the error is denoted in units of the skill grading scheme.

[3]http://tla.mpi.nl/tools/tla-tools/elan/

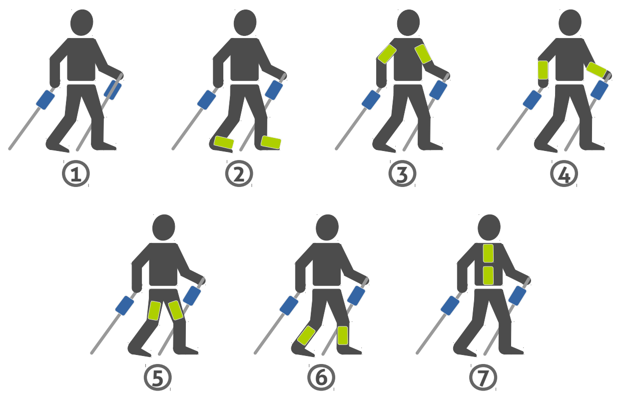

Fig. 4. Sensor placement configurations. In each of the seven configurations two or four sensors at different body positions were considered. Blue boxes indicate pole sensors that appear in every sensor placement configuration, green boxes depict, varying sensor positions. Overall considered sensor configurations: (1) poles (RNP, LNP), (2) poles + feet (RUF, LUF), (3) poles + upper arms (RUA, LUA), (4) poles + wrists (RLA, LLA), (5) poles + thighs (RUL, LUL), (6) poles + shins (RLL, LLL), (7) poles + upper back (STE) and lower back at the centre of mass (CEN).

Stride-by-stride estimated skills grades $\mathbf{\hat{s}_i}$ were compared against the experts evaluated skill grade reference $\mathbf{s_i}$ for all strides $\mathbf{N}$ and each mistake type. RMSE was derived as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(\hat{s}_i - s_i\right)^2}{N}} \tag{2}$$

To illustrate the estimation performance on a percent metric we derived the normalised RMSE (nRMSE) as:

$$\mathrm{nRMSE} = \frac{\mathrm{RMSE}}{s_{max} - s_{min}} \times 100\,\% \tag{3}$$

where s_{max} = 3 (incorrect) and s_{min} = 1 (correct).

The mean absolute error (MAE) was derived to measure the average magnitude of errors in estimated skills where all differences between estimated and reference skill grades were equally weighted. MAE was derived as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{N}|\hat{s}_i - s_i| \tag{4}$$

While intuitive, RMSE and MAE may miss errors due to bias or variance of the estimate. Thus we derived the *Pearson correlation coefficient* (r) [24] in addition to compare estimated skill grades with the experts' annotated skill grade reference using scipy [4]. Correlation was defined significant using a significance level $\alpha < .01$.

VI. Results

We provide an initial study overview on participant data and skills distribution in Section VI-A. Subsequently, we detail our evaluation on feature selection algorithms and compare regression models in Section VI-B and highlight performance of skill grade estimation when using different sensor placement configurations in Section VI-C. Skill trends are presented in Section VI-D.

[4]scipy: scipy.stats.pearsonr

A. Study overview

Table II summarises the numbers of strides per mistake and corresponding skill grades, assessed by the expert. We observed variation and unbalanced mistake grade distributions across all Nordic Walking beginner. A total of 11247 strides were recorded across all study participants. 10674 (94.9 %) were correct matching gear strides and 573 (5.1 %) were incorrect non-matching gear strides referring to MS1. A matching gear was vital to all further mistake assessment and skill estimation analysis, thus non-matching strides were excluded for the evaluation of mistakes MS2 and MS3. For MS2 3364 (31.5 %), 5543 (51.9 %) and 1767 (16.6 %) strides were graded with 1,2, and 3 respectively. For MS3 1790 (16.8 %), 5008 (46.9 %) and 3876 (36.6 %) strides were graded with 1,2, and 3 respectively.

Inter-rater reliability yielded 96 %, hence experts' annotated ground truth was considered viable for subsequent regression-based skill grade estimation. The automatic stride segmentation using the hill-climb algorithm and sensor data derived from RUF and LUF yielded an average time offset of 20 ± 5 ms compared to the manual stride segmentation using ELAN.

B. Feature selection and regression evaluation

Table III summarises performance figures including RMSE, nRMSE, MAE, ande Pearson correlation coefficients (r) for different feature selection algorithms (PCA, GDB) in a leave-one-participant-out (LOPO) cross-validation. We evaluated BRR, OLS, AdaBoostR and SVR using the GDB feature selection algorithm, which showed best performance for BRR. The evaluation was based on sensor data derived from pole and upper leg sensors (sensor placement Configuration 5), which yielded best average results for RMSE, nRMSE, MAE and Pearson correlation. For sensor placement Configuration 5, we observed that skills could be estimated with an average deviation of 0.34 skill grade compared to the experts reference when using the MAE. When using the RMSE as evaluation metric, errors were on average 0.48 skill grades.

The baseline showed a standard deviation below 0.1 % for RMSE, nRMSE, and MAE for all mistakes on average (25 repetitions). Pearson coefficients showed no correlation for the baseline for all mistakes on average: r = 0.04 ± 0.01.

TABLE III. Comparison of regression models for continuous skill grade estimation. Performance figures were derived using sensor placement Configuration 5 (RNP, LNP, RUL, LUL).

	MS1				MS2				MS3			
	RMSE	nRMSE	MAE	Pearson	RMSE	nRMSE	MAE	Pearson	RMSE	nRMSE	MAE	Pearson
	[Skill Grade]	[%]	[Skill Grade]	[-]	[Skill Grade]	[%]	[Skill Grade]	[-]	[Skill Grade]	[%]	[Skill Grade]	[-]
Bayesian Ridge Regression (BRR)												
Baseline	0.47	23.7	0.25	0.05	0.71	35.3	0.56	0.03	0.74	37.0	0.61	0.07
PCA	0.44	22.0	0.20	0.29	0.66	33.1	0.56	0.53	0.70	34.8	0.55	0.44
GDB	0.45	22.3	0.23	0.24	0.49	24.3	0.35	0.52	0.51	25.7	0.44	0.48
DF	0.45	22.6	0.23	0.24	0.50	24.9	0.37	0.51	0.52	26.0	0.44	0.50
Linear Regression (OLS)												
GDB	0.45	22.3	0.23	0.24	0.49	24.4	0.36	0.52	0.51	25.7	0.44	0.47
AdaBoostR (Decision Tree Regressor)												
GDB	0.46	23.1	0.23	0.23	0.52	26.0	0.39	0.51	0.51	25.6	0.47	0.47
Support Vector Regression (SVR)												
GDB	0.43	21.6	0.18	0.27	0.54	27.2	0.37	0.48	0.53	26.6	0.39	0.46

For MS1, the BRR baseline resulted in RMSE (0.47), nRMSE (23.7 %), and MAE (0.25). Similar to the baseline, using BRR and PCA resulted in RMSE (0.44), nRMSE (22 %), and MAE (0.20); BRR and GDB resulted in RMSE (0.45),

TABLE II. Stride distribution for each mistake according expert skill grading. The analysis of MS2 and MS3, included matching gear strides only. Mistake MS1 was graded 1 (correct matching gear stride) or 3 (incorrect non-matching gear stride), MS2 and MS3 were graded with 1, 2, or 3. NG = Not graded. ID1, . . . , ID10 = Participants.

	ID1			ID2			ID3			ID4			ID5		
	Grades			Grades			Grades			Grades			Grades		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
MS1	1059	NG	188	1151	NG	0	881	NG	36	925	NG	46	1044	NG	49
MS2	0	991	68	937	212	2	587	246	48	590	209	126	2	731	311
MS3	0	188	871	862	289	0	594	0	287	0	797	128	2	692	350

	ID6			ID7			ID8			ID9			ID10			Stride Distribution			Strides
	Grades			Grades			Grades			Grades			Grades			Grades			Total
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	-
MS1	1167	NG	2	1251	NG	26	1286	NG	185	1110	NG	0	800	NG	41	10674	NG	573	11247
MS2	237	620	310	0	509	742	263	935	88	667	438	5	81	652	67	3364	5543	1767	10674
MS3	0	705	462	0	232	1019	0	891	395	259	615	236	73	599	128	1790	5008	3876	10674

nRMSE (22.3 %), and MAE (0.23), hence no improvement could be noticed. OLS, AdaBoostR, and SVR in combination with GDB results were comparable with the BRR and the GDB for MS1. For the correlation coefficient however, we observed increased performance compared to the baseline: BRR and PCA (r = 0.28), BRR and GDB (r = 0.24), BRR and DF (r = 0.24), OLS and GDB (r = 0.24), AdaBoostR and GDB (r = 0.23), and SVR and GDB (r = 0.27). For MS2 the BRR and the GDB resulted in RMSE (0.49), nRMSE (24.3 %), and MAE (0.35) and Pearson correlation coefficient (r = 0.52). For MS3 the BRR and the GDB resulted in RMSE (0.51), nRMSE(25.7 %), and MAE (0.44) and Pearson correlation coefficient (r = 0.48).

MS2, resulted in RMSE (0.71) and Pearson correlation coefficient (r = 0.03) using BRR and the baseline, while the BRR and the GDB reduced the RMSE (0.49) and increased the Pearson correlation coefficient (r = 0.52). MS3, resulted in RMSE (0.74) and Pearson correlation coefficient (r = 0.07) using BRR and the baseline, while the BRR and the GDB reduced the RMSE (0.51) and increased the Pearson correlation coefficient (r = 0.48). See Table III for related error metrics nRMSE and MAE.

Using BRR and GDB yielded best performance for skill grade estimation across the mistake types and alternative regression models (OLS, AdaBoosT, and SVR). The decision fusion (DF) could not outperform the PCA or the GDB feature selection, hence was not further considered. GDB showed increased performance figures compared to the PCA. For sensor placement configuration and skill trend analysis we focused on BRR and GDB.

Figure 5 illustrates the skill grade estimation performance when increasing the number of features in the BRR model. Average RMSE, nRMSE, and MAE were minimal when using 7 features, while Pearson correlation was maximal. Hence, the final BRR model was restricted to 7 features.

The final regression model for all mistakes using BRR, GDB and sensor placement Configuration 5 can be found in Section IX.

C. Sensor placement configuration

Table IV summarises average performance for all seven sensor placement configurations investigated in this work. Each configuration was used to derive skill grade estimates for all mistake types.

Minor performance variations were observable across all sensor placement configurations. Configuration 5 including both pole (RNP, LNP) and upper leg (RUL, LUL) sensors resulted in minimal average errors and minimal error variance for MS1, MS2, and MS3 considering RMSE (0.48 ± 0.03), nRMSE (24.15 ± 1.74), and MAE (0.34 ± 0.11), respectively.

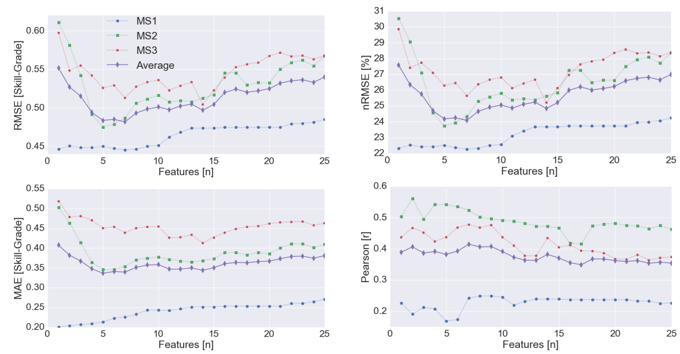

Fig. 5. Performance figures for feature selection using the *Gradient Descent Boosting* (GDB) algorithm with increasing number of features ranging from 1 to 25 for the sensor placement Configuration 5 (RNP, LNP, RUL, LUL). Performance figures (RMSE, nRMSE, MAE, and Pearson correlation) are illustrated for MS1, MS2, MS3, and the average.

TABLE IV. Averaged RMSE, nRMSE, MAE, and Pearson correlation values. Seven sensor placement configurations were evaluated for each mistake type. Each configuration included pole sensors (RNP, LNP). Configurations 2, . . . , 6 included two additional sensor at varying body positions.

Sensors	Sens. placement Configuration	RMSE [Skill Grade]	nRMSE [%]	MAE [Skill Grade]	Pearson [r]
RNP, LNP (poles only)	1	0.53 ± 0.11	26.55 ± 5.52	0.38 ± 0.18	0.33 ± 0.17
RNP, LNP, RUF, LUF	2	0.57 ± 0.11	28.53 ± 5.70	0.41 ± 0.17	0.39 ± 0.23
RNP, LNP, RUA, LUA	3	0.52 ± 0.07	26.02 ± 3.41	0.37 ± 0.14	0.33 ± 0.23
RNP, LNP, RLA, LLA	4	0.52 ± 0.11	26.13 ± 5.35	0.38 ± 0.16	0.37 ± 0.16
RNP, LNP, RUL, LUL	5	0.48 ± 0.03	24.15 ± 1.74	0.34 ± 0.11	0.41 ± 0.15
RNP, LNP, RLL, LLL	6	0.52 ± 0.07	25.91 ± 3.45	0.37 ± 0.13	0.41 ± 0.23
RNP, LNP, CEN, STE	7	0.54 ± 0.09	27.25 ± 4.43	0.40 ± 0.15	0.33 ± 0.18

Pearson correlation coefficients were similar in sensor placement Configuration 5 and 6, though sensor placement Configuration 5 showed smaller standard deviation (0.41 ± 0.15).

D. Skill trends

Trend analysis suggested that 8 participants improved their skill grades regarding MS1. For MS2, 9 participants could improve their skill grades and for MS3, 8 participants. Table V indicates the trend for each participant and mistake type. Figure 6 illustrates exemplarily estimated skill grades and trends in MS3 for all Nordic Walking study participants across all five recording sessions using sensor placement Configuration 5.

TABLE V. Skill trends for each participant and mistake type across all recording sessions where arrows indicate improvement (↑) or deterioration (↓).

	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
MS1	↑	↑	↑	↑	↑	↓	↓	↑	↑	↑
MS2	↑	↑	↑	↑	↑	↑	↑	↑	↑	↓
MS3	↑	↑	↑	↑	↓	↑	↑	↑	↑	↓

VII. Discussion

The complex motion pattern of Nordic Walking makes mistakes likely to occur. Mastering the movement pattern and improving skills, requires training. Although all Nordic Walking beginners included in our study walked mostly in

Fig. 6. Estimated skill grades for MS3 across all participants (ID1,. . . ,ID10) using *Bayesian Ridge Regression* (BRR) and the *Gradient Descent Boosting* (GDB) algorithm for feature selection. Skill grades were derived from the sensor placement Configuration 5, including sensors RNP, LNP, RUL, and LUL. Linear regression trends with negative slopes suggest improving skills during trainings.

correct matching gear (MS1), according our experts evaluation, participants struggled with pole use (MS2) and arm swing (MS3) as the stride distributions in Table II shows. Use of poles (MS2), appeared challenging for beginners as 16.6 % of the strides were graded with 3. Correct arm swing, including elbow joint extension and shoulder activity (MS3) appeared even harder, where 36.3 % of the strides were graded with 3. The study data confirms that our approach to estimate skilfulness is practically relevant.

The skill reference data revealed imbalances between mistake skill grades which are challenging for estimation systems. We consider that the skill grade distribution in our study data is representative of beginners in unconstrained, out-of-lab settings. While our choice of estimation and evaluation methods was directed to control for skill grade imbalances, an effect of imbalances on the results cannot be fully eliminated. The Bayesian regression helped us to capture information about the sample distribution while remaining less sensitive to imbalance than many classification methods.

Contrary to typical classification or clustering approaches [4], [25] the continuous variable output of regression methods is better suited to reveal gradual changes and trends across training sessions as it is often occurring when improving a certain skill. In the present investigation, most study participants improved the more advanced mistake types MS2 and MS3 across training sessions while receiving little advice. However, parameters including sportiness and motivation which could influence the skill development were not investigated. In this work, we focused on health related mistakes and evaluated the feasibility of our approach. Additional mistakes could be devised to further investigate movement patterns, e.g. considering muscle stress caused by too wide or narrow strides. Moreover, such continuous regression-based estimation could inform Nordic Walking beginners immediately, e.g. after a training session about walking performance. Thus, memorising erroneous movements could be avoided.

In our approach, mistake types were modelled as individual regression problems yielding a parallel skill grade estimation for each mistake type. To realise our approach, no additional preprocessing is needed. So far, our approach relies on a personalised low-pass filter cut-off frequency for the subsequent stride segmentation. Hence, further investigations are required to devise a stride segmentation algorithm generalising to new, unseen Nordic Walking movement patterns. The estimated skill grades for each mistake type could be presented to the user or filtered according to user preferences, training objectives, or training level (beginner, intermediate or professional). However, further analysis of feedback strategies is required to provide users targeted and focused instructions for movement optimisation as suggested by Wulf *et al.* [26], [27]. In addition, biomechanical modelling of body interactions and movement could help to understand motion interaction as suggested by Mandery *et al.* [28].

Independent of the skill estimation approach, our work showed that skill annotation is laborious to obtain for Nordic Walking. Although we optimised the post-annotation process, annotation work lasted more than 10-times longer than the recording. It may be conceivable to create larger databases of Nordic Walking recordings including additional experts for annotation. Our analysis of the inter-rater reliability confirmed that the definition of mistakes and annotation guidelines were adequate to yield a consistent reference. To ensure generalisation, we applied a LOPO-CV approach in the evaluation. The generalisation could be further investigated by including participants with different demographic background, e.g. age, body composition, geographic region, etc.. In addition, outdoor recordings with varying ground conditions could reveal changes in motion execution and subsequent skilfulness. Although, we designed the course to maintain a steady walking speed in the evaluation section of the course, varying acceleration could contribute to motion mistakes.

The result confirmed that the definition of mistake and annotation guidelines were adequate to yield a consistent reference. Hence, our approach demonstrated that combining expert knowledge, mistake-driven, and regression-based skill estimation could be used to analyse complex motion patterns. Nevertheless, experts identified challenges in skill grading. The grading of skill grades 1 or 3 could be achieved quickly due to clear distinction of ”good” and ”bad” skills. Grading movement skills with a 2 required more elaboration due to small differences in motion execution across participants. Experts reported that movements were graded as 2, although descriptions including ”a good 2” or ”a bad 2” were used to describe variances in motion execution. The proposed skill reference was devised to evaluate skills similar to traffic light categories including 1 (green, no mistake), 2 (orange, minor mistake), and 3 (red, mistake). The traffic light approach is quickly understandable for coaches and walkers too, and rendered the experts' annotation effort feasible. Grading of skills using categorical labels, e.g. ranging from 1 to 10, was considered inappropriate due to nuances in motion performance which could not be captured by human experts.

We showed that skill estimation is feasible using regression methods. Comparing different regression methods however, showed similar performance. Algorithms for feature selection, particularly the gradient boosting approach could improve estimation results as they provide an effective way to determine relevant features. Using RMSE or MAE only may be insufficient to understand the performance of regression estimates. *Chai and Draxler* [23] concluded that RMSE should be favoured as potential outliers are considered too, compared to MAE. The additional correlation coefficients revealed in our analysis that the estimation algorithms perform sensible, even in the highly imbalanced MS1, where RMSE and MAE results are masked by the most frequent skill grade.

Our results suggest some flexibility regarding sensors placement at different body positions. For example, sensor could be well integrated in Nordic Walking poles. In addition, sensor could be integrated into shoes or cloth pockets. We believe that practically viable sensor integration for Nordic Walking monitoring is feasible and could be achieved using e.g. Bluetooth enhanced IMUs in combination with a smartphone. The smartphone could be used to derive additional geo-location data via GPS, quantify and visualise skilfulness immediately after an Nordic Walking session, and provide motion recommendations to the user.

While approaches to sensor-based estimation of clinical movement assessment often focus on individual body parts or selective items of an assessment, e.g. in stroke rehabilitation [11], [12], we analysed mistakes including the complete body. Nordic Walking is characterised by complex motion patterns and body interactions, hence expert knowledge was required to define mistakes. We believe our approach could be applied in movement rehabilitation and other health domains too, using expert knowledge of clinicians. For example, motion mistakes could be devised related to impairments, limitations and coping strategies of patients. By creating mistake models and skill grades the analysis of movements and compensation could reveal changes over time, thus indicate potential recovery trends using continuous regression-based models. In other sport domains, e.g. golf or tennis, a similar approach could be devised using expert knowledge of coaches to determine mistakes which reduce effective motion execution.

VIII. Conclusion

In this work we introduced a mistake-driven skill grade estimation approach to assess skilfulness in Nordic Walking beginners. Our approach to estimate health-related mistakes using regression appears suitable as a training support.

Using a regression-based estimation, we showed that fine-grained variation in motion performance could capture potential trends in Nordic Walking beginners correlating with expected skill improvement. In our study, all participants improved skills across training sessions, considering the mistake types.

Overall, seven sensor configurations were devised considering practical placement on different body positions. Sensors placed at each pole and two body-worn sensors provided reliable skill estimation performance. In addition, sensors could be integrated in shoes or sports clothes, rendering a mobile mistake analysis viable.

Acknowledgment

We thank our Nordic Walking study participants and student assistants for supporting data recordings, and our Nordic Walking experts for support and annotation.

IX. Appendix

Regression models which resulted in best average skill grade estimation using BRR, GDB and sensor placement Configuration 5 are expressed in Eq. 5, 6, 7.

$$\begin{aligned} \mathrm{MS1} = -0.12 \cdot \mathrm{F1}_{\mathrm{acc_X}} + 0.03 \cdot \mathrm{F2}_{\mathrm{gyr_X}} - 0.03 \cdot \mathrm{F3}_{\mathrm{acc_X}} \\ -0.01 \cdot \mathrm{F4}_{\mathrm{acc_Y}} + 0.004 \cdot \mathrm{F5}_{\mathrm{gyr_Y}} - 0.03 \cdot \mathrm{F6}_{\mathrm{acc_X}} - 0.19 \cdot \mathrm{F7}_{\mathrm{gyr_X}} \end{aligned} \tag{5}$$

$$\begin{aligned} \mathrm{MS2} = -0.03 \cdot \mathrm{F8}_{\mathrm{acc_X}} - 0.13 \cdot \mathrm{F9}_{\mathrm{gyr_X}} + 0.002 \cdot \mathrm{F10}_{\mathrm{acc_Z}} \\ -0.02 \cdot \mathrm{F11}_{\mathrm{acc_X}} - 0.13 \cdot \mathrm{F12}_{\mathrm{acc_X}} - 0.03 \cdot \mathrm{F13}_{\mathrm{acc_X}} - 0.1 \cdot \mathrm{F14}_{\mathrm{acc_Z}} \end{aligned} \tag{6}$$

$$\begin{aligned} \mathrm{MS3} = 1.03 \cdot \mathrm{F15}_{\mathrm{gyr_Y}} + 0.2 \cdot \mathrm{F16}_{\mathrm{gyr_X}} - 0.02 \cdot \mathrm{F17}_{\mathrm{gyr_Y}} \\ -0.11 \cdot \mathrm{F18}_{\mathrm{gyr_X}} - 0.13 \cdot \mathrm{F19}_{\mathrm{acc_X}} + 0.12 \cdot \mathrm{F20}_{\mathrm{acc_Y}} - 0.08 \times \mathrm{F21}_{\mathrm{acc_Y}} \end{aligned} \tag{7}$$

where acc = accelerometer, gyr = gyroscope, X = x-axis, Y = y-axis, and Z = z-axis. Table VI summarises extracted features in each regression model.

TABLE VI. SELECTED FEATURES FOR THE BAYESIAN RIDGE REGRESSION USING THE GRADIENT DESCENT BOOSTING ALGORITHM. F1, . . . , F21 DENOTE FEATURES IN THE REGRESSION MODEL. SENSOR POSITIONS ARE: RIGHT POLE (RNP), LEFT POLE (LNP), RIGHT UPPER LEG (RUL), AND LEFT UPPER LEG (LUL).

			MS1 - Matching gear			
F1, LNP 5%ile	F2, RNP min	F3, RNP max	F4, RUL mean	F5, RNP minpos	F6, RUL 5%ile	F7, LNP min
			MS2 - Pole use			
F8,LNP 5%ile	F9, RUL min	F10, RNP sum	F11, LNP max	F12, LNP skew	F13, LUL mean	F14, LUL skew
			MS3 - Arm swing			
F15, RNP 5%ile	F16, RNP 5%ile	F17, LNP min	F18, LUL 5%ile	F19, LNP max	F20, LNP max	F21, LNP skew

REFERENCES

[1] K. Knobloch and P. Vogt, "Nordic Walking Verletzungen - Der Nordic-Walking-Daumen als neue Verletzungsentität," *Sportverletzung· Sportschaden*, vol. 20, no. 03, pp. 137–142, 2006.

[2] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans Inf Technol Biomed*, vol. 10, no. 1, pp. 119–128, January 2006.

[3] P. Kocur, E. Deskur-Smielecka, M. Wilk, and P. Dylewicz, "Effects of Nordic Walking training on exercise capacity and fitness in men participating in early, short-term inpatient cardiac rehabilitation after an acute coronary syndrome – a controlled trial." *Clin Rehabil*, vol. 23, no. 11, pp. 995–1004, Nov 2009. [Online]. Available: http://dx.doi.org/10.1177/0269215509337464

[4] K. Kunze, M. Barry, E. A. Heinz, P. Lukowicz, D. Majoe, and J. Gutknecht, "Towards recognizing Tai Chi – an Initial Experiment Using Wearable Sensors," in *Applied Wearable Computing (IFAWC), 2006 3rd International Forum on*. VDE, 2006, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5758288

[5] A. Ahmadi, D. Rowlands, and D. A. James, "Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve," *Sports Technology*, vol. 2, no. 3-4, pp. 129–136, Feb 2010. [Online]. Available: http://dx.doi.org/10.1002/jst.112

[6] M. Bächlin, K. Förster, and G. Tröster, "Swimmaster: A wearable assistant for Swimmer," in *Proceedings of the 11th International Conference on Ubiquitous Computing*, ser. UbiComp '09. New York, NY, USA: ACM, 2009, pp. 215–224. [Online]. Available: http://doi.acm.org/10.1145/1620545.1620578

[7] C. Strohrmann, H. Harms, G. Tröster, S. Hensler, and R. Müller, "Out of the Lab and into the Woods: Kinematic Analysis in Running Using Wearable Sensors," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, ser. UbiComp '11. New York, NY, USA: ACM, 2011, pp. 119–122. [Online]. Available: http://doi.acm.org/10.1145/2030112.2030129

[8] C. Ladha, N. Y. Hammerla, P. Olivier, and T. Plötz, "ClimbAX: Skill Assessment for Climbing Enthusiasts," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 235–244. [Online]. Available: http://doi.acm.org/10.1145/2493432.2493492

[9] F. Michahelles and B. Schiele, "Sensing and monitoring professional skiers," *IEEE Pervasive Computing*, vol. 4, no. 3, pp. 40–45, July 2005.

[10] S. A. Rawashdeh, D. A. Rafeldt, and T. L. Uhl, "Wearable IMU for Shoulder Injury Prevention in Overhead Sports," *Sensors*, vol. 16, no. 11, p. 1847, 2016. [Online]. Available: http://dx.doi.org/10.3390/s16111847

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

[12] S. Del Din, S. Patel, C. Cobelli, and P. Bonato, "Estimating Fugl-Meyer clinical scores in stroke survivors using wearable sensors," *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 5839–5842, 2011.

[13] A. Derungs, J. Seiter, C. Schuster-Amft, and O. Amft, "Estimating physical ability of stroke patients without specific tests," in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ser. ISWC '15. New York, NY, USA: ACM, 2015, pp. 137–140. [Online]. Available: http://doi.acm.org/10.1145/2802083.2808412

[14] G. Spina, G. Huang, A. Vaes, M. Spruit, and O. Amft, "Copdtrainer: A Smartphone-based Motion Rehabilitation Training System with Real-time Acoustic Feedback," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 597–606. [Online]. Available: http://doi.acm.org/10.1145/2493432.2493454

[15] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011. [Online]. Available: https://doi.org/10.1137/090771806

[16] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural computation*, vol. 11, no. 2, pp. 443–482, 1999.

[17] J. H. Friedman, "Stochastic Gradient Boosting," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 367–378, Feb. 2002. [Online]. Available: http://dx.doi.org/10.1016/S0167-9473(01)00065-2

[18] S. C. A. Thomopoulos, R. Viswanathan, and D. C. Bougoulias, "Optimal Decision Fusion in Multiple Sensor Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-23, no. 5, pp. 644–653, Sept 1987.

[19] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[20] H. Drucker, "Improving regressors using boosting techniques," in *ICML*, vol. 97, 1997, pp. 107–115.

[21] D. Pardoe and P. Stone, "Boosting for Regression Transfer," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, J. Fürnkranz and T. Joachims, Eds. Haifa, Israel: Omnipress, June 2010, pp. 863–870. [Online]. Available: http://www.icml2010.org/papers/330.pdf

[22] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a Professional Framework for Multimodality Research," in *In Proceedings of the 5th International Conference on Language Resources and Evaluation*, ser. LREC 2006, 2006, pp. 1556–1559.

[23] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014. [Online]. Available: https://www.geosci-model-dev.net/7/1247/2014/

[24] K. Pearson, "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 187, pp. 253–318, 1896. [Online]. Available: http://rsta.royalsocietypublishing.org/content/187/253

[25] B. Eskofier, P. Kugler, D. Melzer, and P. 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*, May 2012, pp. 113–117.

[26] G. Wulf, M. Höss, and W. Prinz, "Instructions for Motor Learning: Differential Effects of Internal Versus External Focus of Attention," *Journal of Motor Behavior*, vol. 30, no. 2, pp. 169–179, 1998, pMID: 20037032. [Online]. Available: http://dx.doi.org/10.1080/00222899809601334

[27] G. Wulf, N. Mcconnel, M. Gärtner, and A. Schwarz, "Enhancing the Learning of Sport Skills Through External-Focus Feedback," *Journal of Motor Behavior*, vol. 34, no. 2, pp. 171–182, 2002, pMID: 12057890. [Online]. Available: http://dx.doi.org/10.1080/00222890209601939

[28] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "Unifying representations and large-scale whole-body motion databases for studying human motion," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 796–809, 2016.