Journal of Biomedical Informatics 58 (2015) 145–155

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Recognizing the intensity of strength training exercises with wearable sensors

Igor Pernek [a,*], Gregorij Kurillo [b], Gregor Stiglic [c,d], Ruzena Bajcsy [b]

[a] Pervasive Computing Applications, Research Studios Austria, Vienna, Austria
[b] EECS, University of California, Berkeley, CA, USA
[c] Faculty of Health Sciences, University of Maribor, Maribor, Slovenia
[d] Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

ARTICLE INFO

Article history:
Received 28 April 2015
Revised 29 July 2015
Accepted 27 September 2015
Available online 8 October 2015

Keywords:
Machine learning
Werable sensors
Strength training
Accelerometers

ABSTRACT

In this paper we propose a system based on a network of wearable accelerometers and an off-the-shelf smartphone to recognize the intensity of stationary activities, such as strength training exercises. The system uses a hierarchical algorithm, consisting of two layers of Support Vector Machines (SVMs), to first recognize the type of exercise being performed, followed by recognition of exercise intensity. The first layer uses a single SVM to recognize the type of the performed exercise. Based on the recognized type a corresponding intensity prediction SVM is selected on the second layer, specializing in intensity prediction for the recognized type of exercise. We evaluate the system for a set of upper-body exercises using different weight loads. Additionally, we compare the most important features for exercise and intensity recognition tasks and investigate how different sliding window combinations, sensor configurations and number of training subjects impact the algorithm performance. We perform all of the experiments for two different types of features to evaluate the feasibility of implementation on resource constrained hardware. The results show the algorithm is able to recognize exercise types with approximately 85% accuracy and 6% intensity prediction error. Furthermore, due to similar performance using different types of features, the algorithm offers potential for implementation on resource constrained hardware.

1. Introduction

Physical activity is an important component of a healthy life style. Evidence suggests that regular exercise participation results in improvements in the function of the cardiovascular system and the skeletal muscles [1] and significantly reduces the risk of developing different chronic diseases, such as hypertension, obesity, depression, and cardiovascular diseases [2–5].

Most health guidelines prescribe the recommended physical activity intake in terms of exercise duration and intensity [2]. While exercise duration can simply be measured with a stopwatch, exercise intensity is not as straightforward to capture. This is particularly true for stationary indoor exercises such as weightlifting and similar strength training activities that are unable to leverage the capabilities of the Global Positioning System (GPS) sensor. Speed and position, derived from the GPS data, can accurately represent the intensity of cardio training activities such as running or cycling, but are not suitable for activities performed in place such as upper body exercises. Capturing the intensity of such exercises is equally important, as studies have shown strength training is an important component of a balanced exercise regimen [6]. Additionally, exercise intensity awareness is important as under-training fails to deliver optimal training benefits [7], while over-training results in excessive exhaustion and consequently loss of exercise motivation [8]. Furthermore, appropriate exercise intensity is not only important for leisure activities. Studies have shown that in the medical rehabilitation environment, inappropriate exercise intensity could lead to injuries or even death in specific cases [9].

In this work, we introduce a hierarchical algorithm for detecting self-perceived intensity of strength training exercises. We define intensity as a derivative of Borg's rating of perceived exertion [10]. The algorithm uses two layers of supervised learning classifiers to first recognize the type of the exercise being performed and to detect the intensity of the exercise once its type has been recognized. The following allows us to perform intensity recognition more accurately and makes the algorithm easily extendible,

* Corresponding author.
E-mail addresses: igor.pernek@researchstudio.at (I. Pernek), gregorij@eecs.berkeley.edu (G. Kurillo), gregor.stiglic@um.si (G. Stiglic), bajcsy@eecs.berkeley.edu (R. Bajcsy).

http://dx.doi.org/10.1016/j.jbi.2015.09.020
1532-0464/

as new exercises can be included with minimal interference with the existing classifiers.

We evaluate the algorithm for different upper body exercises using a set of wearable accelerometers mounted on the upper body of participating subjects. However, due to the modular hierarchical approach, the algorithm could easily be extended to support arbitrary exercises. We choose to relay solely on accelerometers, as they are small and cheap and have already been validated for measuring activity intensity [11]. We investigate whether the sensors provide sufficient information to derive exercise intensity information for two distinct types of acceleration features, namely single-sensor (SS) and multi-sensor (MS) features. We define SS features as features calculated from a single sensor in real-time. Consequently, raw data does not need to be communicated between sensors, which improves the sensor autonomy. In contrast, we define MS features as those based on acceleration data from at least two distinct sensors, thus requiring raw data to be transferred either between sensor nodes or to a common gateway, such as a smartphone. The paper provides the following contributions:

- A hierarchical algorithm for intensity recognition of strength training exercises.
- Evaluation of the algorithm in terms of exercise type recognition accuracy and intensity prediction error for a set of upper body exercises.
- Comparison of the algorithm performance for two distinct groups of features.
- A study of using the algorithm with different sliding window configurations, sensor setups, and number of training subjects.

The rest of the paper is organized as follows. In Section 2 we briefly describe the related work. Section 3 contains a short overview of the system proposed. Section 4 outlines the algorithm and explains individual processing steps. We describe the evaluation protocol in Section 5 and provide results in Section 6. A short discussion of limitations of the proposed approach is presented in Section 7 before we conclude the paper in Section 8.

2. Related work

There have been several industrial and academic attempts investigating quantitative observations of stationary activities, such as strength training and rehabilitation exercises. Most of the existing approaches are based on video [12,13], garment [14,15], or wearable sensors. Since our approach uses wearable sensor, the rest of this chapter outlines some of the relevant work using wearable technology.

In [16] authors propose a Wireless Body Area Network of accelerometers to monitor biometric parameters while exercising. The sensors are positioned on the body of the person exercising and are able to capture the correctness of exercise repetitions. The authors define exercise correctness based on the body posture and execution speed during exercise. However, the proposed approach is very exercise specific and does not provide any evaluation for a broader set of exercises.

Chang et al. [17] propose a system for monitoring free weight exercises. The proposed solution, comprised of a smartphone and two wearable sensors, is able to recognize different exercises along with the number of repetitions performed. However, the system does not report any information on the quality and correctness of exercises being performed.

Similarly, myHealthAssistant [18] captures exercise repetition count using a smart phone and a set of wearable sensors. The authors leverage the modern smart phone processing capabilities to deploy an algorithm based on a trainable classifier. The algorithm is able to recognize exercise repetitions in real time with minimal impact on overall system's resources, but does not offer any guidance on exercise correctness.

In [19] the authors propose an algorithm for spotting upper body exercises and predicting the number of repetitions performed. The algorithm is able to perform user independent exercise recognition from a continuous stream of data provided by an off-the-shelf smartphone placed in a commercial arm holster. However, similarly as in [17,18] the proposed algorithm is not able to provide any information on exercise correctness.

In our previous work [20] we have proposed an approach that was able to advance the solutions described by performing not only exercise repetition counting, but also capturing exercise correctness for a range of different exercises. Exercise correctness recognition was performed by detecting individual repetition's start- and end-points and consequently recognizing the exercising tempo. Similar work has also been proposed by Spina et al. [21]. They have proposed a smartphone based motion rehabilitation system for individual exercising of chronic patients. The system is able to process motion sensor data online on the phone and provide real-time acoustic feedback regarding the exercise performance and quality. However, both approaches use a single sensor device and are thus constrained to simple exercises only. Additionally, only basic exercise correctness metrics, such as exercising tempo, were considered.

We advance the related work by using a network of wearable accelerometers with different groups of acceleration features and by predicting another exercise correctness metric, namely exercise intensity.

3. System overview

The prototype system consists of an off-the-shelf smartphone and five wearable sensors connected over bluetooth into a piconet local area network. In such setting, the smartphone serves as a hub responsible for receiving data from up to seven interconnected sensors and doing all the heavy processing. Additionally, such setting makes it possible to easily transfer data and processed results from the smartphone to remote locations and thus enables real-time supervision of the training by rehabilitation specialists or personal trainers.

Five wearable 3-axis accelerometers were body-mounted on different body locations of the participants. The sensors were sampling acceleration data with a sampling frequency of 30 Hz, which has turned out to be adequate for recognition of activities with similar motion dynamics [20,22]. Due to the most discriminative type of motion, the following sensor mounting locations were selected (as depicted on the left in Fig. 1):

- chest,
- left and right wrist, and
- left and right upper arm.

All the sensors were mounted using standard sports equipment, such as cotton wrist and elbow bands and a textile chest strap. Such installation of the sensors is low-cost and does not require any specific or hard to acquire equipment. Furthermore and most importantly, the placement of the sensors is not obstructing the exercise execution in any way.

4. Hierarchical algorithm

This section describes the hierarchical algorithm for predicting exercise intensity. The idea of the algorithm is to break the intensity recognition problem into two sequential tasks and solve each

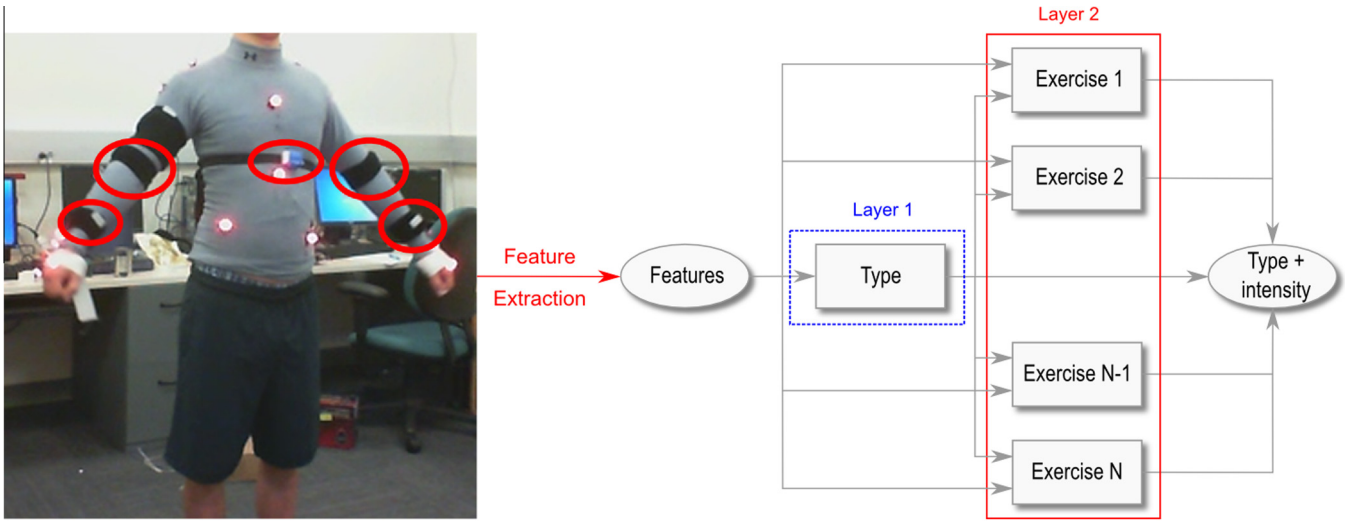

Fig. 1. Exercise recognition pipeline. Raw data is extracted from sensors depicted on the left. Features are extracted from raw data and passed to the hierarchical algorithm. Layer 1 classifier (dashed blue box) recognizes the type of exercise. Layer 2 classifiers (solid red box) are used to predict exercise intensity once its type has been recognized. At the end, both, exercise type and intensity are known. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

one independently. We design a two-step algorithm that first recognizes the type of the executed activity, followed by recognition of exercise intensity. The motivation for such an approach is based on the following observations:

1. different types of exercises are often reflected in different patterns of the acceleration signal, independent of the exercise intensity and
2. similar exercise intensities are reflected in a similar acceleration pattern, which is usually local and specific for a particular type of exercise.

This enables us to detect the intensity of various exercises using two layers of predictive models with different specializations. The first layer exploits the first observation and is used to recognize only the type of the performing exercise, ignoring all the intensity information. The second layer is based on the second observation and only recognizes the intensity of known exercise patterns. This allows us to detect exercise specific features and use them for intensity prediction. Fig. 2 provides visual motivation for the approach, depicting acceleration data for the low (first row) and high (second row) intensity lateral lift exercise. Acceleration data is shown for sensors placed on three different body locations represented in individual columns (chest, right wrist and elbow, respectively). A repetitive pattern pertaining to a series of repetitions is clearly visible. While in the scope of this paper we do not perform any exercise segmentation and repetition detection, an approach similar to the one we have proposed in [20] could be adapted to achieve this. Multisensor comparison of the same exercise performed with different intensities shows that while the signals in general look similar there are some obvious differences between high and low intensity exercises. In the case of the high intensity exercise there is much more motion registered by the chest sensor. This can be explained by the subject's upper body swinging while trying to perform the exercise with heavier weights. Furthermore, the acceleration amplitudes during the high intensity exercise are higher due to less controlled and consequently faster movement. Finally, the figure also shows that the high intensity signals are less stable, due to subjects not being able to perform those exercises as smooth as exercises with lower intensity. While similar properties can be observed for different exercises, they are not exactly the same, mostly varying in participating sensor channels, amplitudes, etc. Thus, different models are used to capture the specifics of individual exercises. Moreover, such approach makes the algorithm easily extendible, as new exercises can be added by retraining the first level and only adding the appropriate model to the second layer.

Fig. 1 depicts the exercise recognition pipeline and outlines the hierarchical structure of the algorithm. We used a supervised machine learning algorithm to implement the predictive models on different layers. More specifically, Support Vector Machines (SVM) were chosen due to their previous use in similar physical activity studies [19,23]. The following sections provide more information on the preprocessing steps, selected features and algorithm details.

4.1. Preprocessing

Two preprocessing steps are applied to the raw data to transform the acceleration information into a form suitable for predictive modeling. The steps used are:

- temporal alignment and
- uniform resampling.

In the temporal alignment step, we align each acceleration frame according to it's sampling time. Time synchronization is performed based on the arrival time of each frame to the smartphone. This means that we simply discard the internal sampling time information for each local sensor and consider only the time recorded by the central component of our network, the smartphone. This step does not influence the algorithm accuracy, since the latency of the bluetooth data transfer is negligible in relation to the dynamics of the movement we are trying to capture.

To decrease the computational complexity of the algorithm, temporally aligned data is then uniformly resampled with a sampling frequency of 25 Hz. Consequently, features calculated in the next step of the algorithm can easily be extracted without any temporal arithmetic. Preprocessing is performed independently for each of the three sensor channels, resulting in a total of 15 aligned streams of acceleration data.

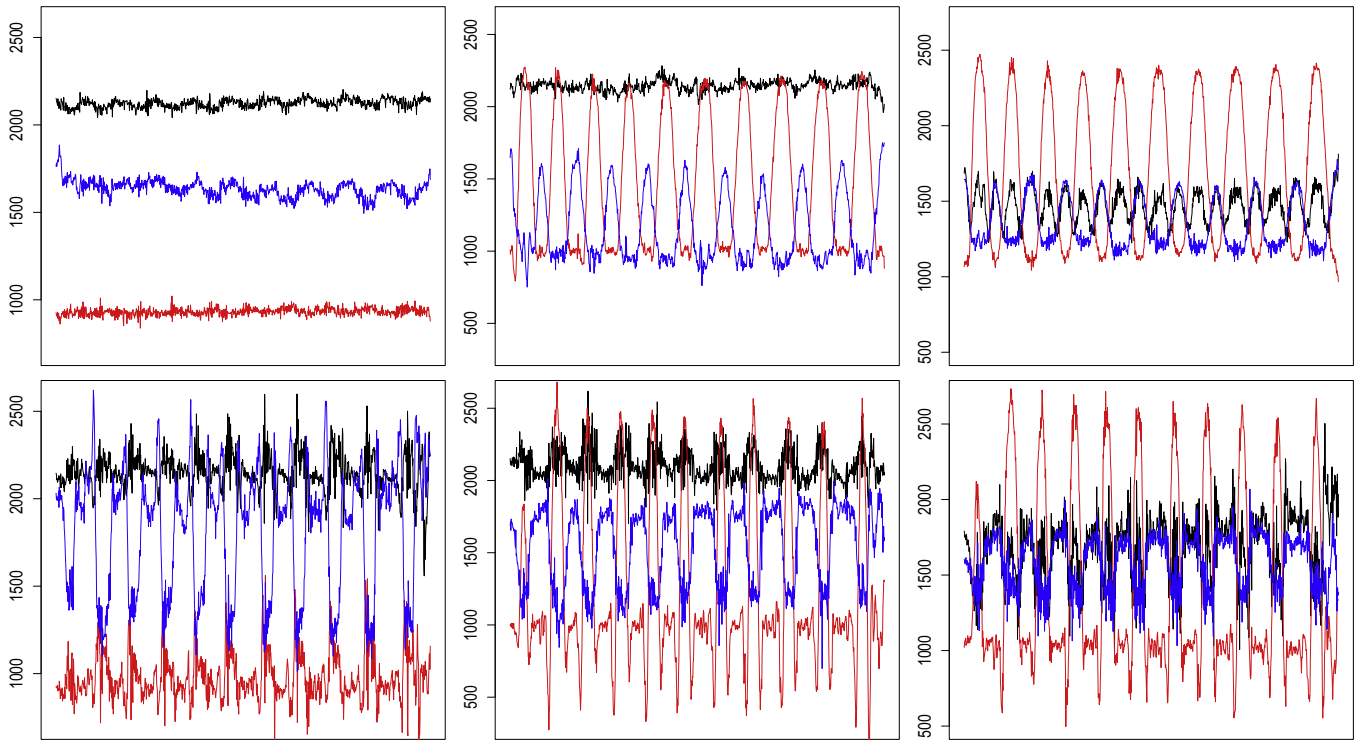

Fig. 2. Examples of acceleration signals for the LL exercise. In each column data from a particular sensor location is shown (left to right: chest, right wrist, right elbow). The first row represents sensor data of the lowest exercise intensity, the second row of the highest. Individual sensor channels are depicted with different colors (X: black, Y: red, Z: blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Feature engineering

Preprocessed data from the previous step is used to construct multiple aggregated features that are later on used to train and classify the type and intensity of the exercises. The features were calculated from raw acceleration data recorded in x, y, and z direction of individual sensor. Additionally, for each sensor, a magnitude (m) value was calculated $\left(m = \sqrt{x^2 + y^2 + z^2}\right)$. A sliding window approach is used to calculate features from the temporal stream of acceleration information. The sliding window parameters were chosen based on the results of a comparison study of four different sliding window sizes and two different overlaps. Feature extraction significantly reduces the dimensionality of the problem and allows more efficient implementation of the predictive models on each layer.

The calculated features are grouped into single-sensor (SS) and multi-sensor (MS) features, based on the number of sensors participating in the calculation of each individual feature. While MS features provide additional information about temporal interaction between different parts of the body, they significantly impact the autonomy of the system, as raw data has to be transferred over the network constantly. On the other hand SS features can be precalculated on each sensor node, which significantly decreases the network load and increases the autonomy of the system. Table 1 outlines the features used, along with the feature type information and instructions on how the features can be calculated. MS features type includes all the SS features along with the correlation information between different sensor nodes. While calculating correlation is much more expensive then calculating the other features, we hypothesized that it will allow us to capture more complex motion information, such as synchronicity of the upper extremities movement [24].

Table 1
Features used along with their types (SS: single-sensor, MS: multi-sensor) and formal definitions. Notation: N is the number of frames in the sliding window, i is the index of the individual frame, a and b represent individual sensor channels ($a, b \in \{Chest^s, Wrist^s_r, Elbow^s_r\}, r \in \{Left, Right\}, s \in \{x, y, z, m\}, a \neq b$).

Name	Type	Definition
Minimum	SS, MS	$min(a)$ = lowest a_i, $i = 1, 2, \ldots, N$
Maximum	SS, MS	$max(a)$ = highest a_i, $i = 1, 2, \ldots, N$
Range	SS, MS	$ran(a) = max(a) - min(a)$
Arithmetic mean	SS, MS	$mean(a) = \frac{1}{N}\sum_{i=1}^{N} a_i$
Standard deviation	SS, MS	$sd(a) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a_i - mean(a))^2}$
Root Mean Square	SS, MS	$rms(a) = \sqrt{\frac{1}{N}(a_1^2 + a_2^2 + \cdots + a_N^2)}$
Correlation	MS	$c(a, b) = \frac{\sum_{i=1}^{N}(a_i - mean(a))(b_i - mean(b))}{(N-1)sd(a)sd(b)}$

4.3. Classification

Calculated features were fed into two layers of SVM classifiers, the first doing type and the second doing intensity recognition. Following the recommendations on the SVM usage [25], the features were first scaled and centered. No feature selection was performed for statistical models on individual levels. Thus, all features calculated in the previous step of the algorithm were passed to SVM classifiers for exercise type and intensity recognition. Due to exercise types expressed as nominal groups and intensity expressed as a number denoting the rating of perceived exertion, the classifier at the first layer performed classification, while the classifiers at the second layer performed numerical regression. In addition to the six exercise classes, the first layer was also able to recognize the activity as non-exercise activity in case the subject was not exercising.

Due to the ability to detect non-linear relations, Gaussian Radial Basis Function (RBF) kernel was used for SVM classifiers on both layers. For each of the predictive models, parameter tuning of complexity (C) and radius of the RBF kernel (r) was performed using the grid search approach. The selected parameters $C = 2$ and $r = 0.01$ where used through all of the experiments.

5. Evaluation

Here we describe the experimental protocol used to evaluate the accuracy of the algorithm. We first, outline the demographics of the participants, explain the evaluation procedure and describe the sensors used in the experiment. Finally, we give a brief overview of the exercises used in the experiment along with their descriptions and reasons for inclusion.

5.1. Participants

To evaluate the algorithm 11 healthy individuals (3 female, 8 male; age: 27 ± 4.5 years) were recruited to perform exercises depicted in Fig. 3. Prior to performing the exercises all individuals have signed an informed consent explaining the protocol of the experiment. Moreover, an active IRB protocol for measurements of human activities with wearable sensors was in place.

Each individual performed ten repetitions of all exercises repeating the process with four different weights, which resulted in a total of 264 data tracks, 44 per each exercise. The load was selected according to the gender of the participant. Female subjects were asked to execute the first series without any load, followed by 3 lbs, 9 lbs and 15 lbs weights. Male subjects carried out the exercises using 3 lbs, 8 lbs, 15 lbs and 20 lbs weights. All participants were first shown an instructional video, demonstrating the correct execution of exercises with additional practical advice for optimal performance. Similar instructional videos, but shorter, were shown to the subjects before each individual exercise to remind them of proper execution.

During the experiments, the following data was collected for all subjects: (i) acceleration from five body mounted accelerometers (Shimmer Research, Dublin, Ireland), (ii) 3D motion using a motion capturing system Impulse (PhaseSpace Inc., San Leandro, CA), and (iii) video. Only acceleration data was used as an input to the proposed algorithm. The motion capturing system, consisting of 10 cameras in a circular configuration, was solely used as a reference for determining the start- and end-points for each series and repetition. Similarly, the video was used for annotation of individual activities with appropriate exercise type and load information. To synchronize different sensor streams each subject was asked to perform a clapping motion prior to starting the first exercise. This motion was then manually looked for in individual sensor streams and used to align the signals.

Furthermore, after each series of repetitions, participants were asked to self-assess the intensity of the exercise using the Borg's rating of perceived exertion (RPE) [10]. We have decided to use a subjective measure of intensity rather than an objective one (e.g. heart rate), as the available objective measures were not able to provide deterministic reference data. It is well known that sensor-observed heart rate changes are delayed with respect to actual intensity changes [26], however this lag is not constant and depends on different factors, such as age, gender and fitness level of the person. Due to all of our exercises being fairly short, such misalignment would impact a significant portion of our data. However, to make the Borg's RPE a bit more robust to operation by inexperienced user, an additional step was introduced. As described by the Eq. (1) Borg's RPE was further normalized for each

Fig. 3. Overview of exercises performed in the experiment. Each row contains a sequence of execution steps for an exercise. The exercise are depicted in the following order: biceps, triceps, front vertical lift, lateral vertical lift, rowing, overhead extension.

participant (index i), resulting in a number between 0 and 1, where 0 represented the lowest and 1 the highest perceived intensity. The normalization was performed to minimize the effects of participants having no prior experience with the Borg's scale and consequently using different values for e.g. the highest perceived intensity. Normalized RPE (NRPE) was then used as a ground truth for evaluating exercising intensity.

$$NRPE_i = \frac{RPE_i - min(RPE_i)}{max(RPE_i) - min(RPE_i)} \quad (1)$$

5.2. Exercises

A set of six upper body strength training exercises was chosen to evaluate the algorithm. Fig. 3 depicts the sequence of execution steps for each exercise. Due to the limited number of sensors being able to connect in a bluetooth piconet and the fact that in the past the upper body received much less attention in terms of movement quality evaluation [27], only upper body exercises were chosen.

All of the exercises with the exception of the triceps exercise were performed using both upper limbs. Although there are versions of the triceps exercise involving both arms, we decided to use the single arm version. The main reason for this was the participant's safety, as this exercise was quite strenuous to perform with the heavy load and involved movement around the participant's head. Thus, the participants were able to help themselves with the free hand if the exercise load became too heavy.

Deliberately some exercises with similar motion paths were chosen, to test the robustness of the algorithm against similar movement. An example of such exercises are front (FL) and lateral vertical lift (LL). Both exercises contain vertical movement, but differ in the orientation of the upper limbs while performing the exercise. Thus, the FL exercise is performed in front of the subject, while the LL exercise contains movement on the side of the subject's body.

As described in the previous section, all the exercises were performed with different loads, each load corresponding to a particular level of exercise intensity.

6. Results

The algorithm accuracy was evaluated in terms of exercise type and intensity recognition using the leave-test-out validation protocol. The leave-test-out is a robust cross-validation protocol that exploits the fact that the evaluation data was collected from different subjects. It does not split the data into the test and training set randomly, but rather uses data from some subjects for testing and data from others for training. Consequentially, test and training data never contain samples from the same subject, which makes the algorithm very robust to overfitting problems. In our case, the leave-one-user-out (LOUO) evaluation protocol was used, which means the algorithm was trained on data from 10 subjects and evaluated on the remaining subject. This was repeated until all the subjects were exactly once included in the test set. However, to facilitate easier comparison of the algorithm results, we additionally present the results for the standard 10-fold cross-validation (CV) evaluation protocol. In this case the dataset is divided into the test and training set using stratified splits to retain class distributions. We use 75% of data for training and the remainder for testing. CV evaluation protocol is less rigorous as data from the same subject can appear in both training and test set resulting in potentially overfit classifiers. To alleviate this, we use CV only to report general results in Table 4 and use the LOUO protocol for all other evaluations.

Furthermore, we evaluated the algorithm for two different types of features, as outlined in Table 1. For each feature type two numbers were reported, one expressing the accuracy of classifying the exercise types in one of the seven classes, the other expressing the error of intensity prediction. Thus, the first number was desired to be as close to 100 as possible (denoting the accuracy of 100%), while the lower second number meant a lower intensity recognition error (10 denoting 10% intensity prediction error).

6.1. Sliding window configurations

Selecting an optimal sliding window length and overlap importantly impacts the accuracy of the algorithm. Based on a review of similar studies [28–31,19,32], four different sliding window lengths (1 s, 2 s, 4 s and 6 s) and two different overlap sizes (25% and 50%) were evaluated to select a single sliding window parameter configuration to be used for all further algorithm evaluations.

Generating features from the same raw data stream with different sliding window parameters produces a different number of feature frames. E.g. using a 6 s sliding window with 50% overlap yields 3045 feature frames, while using a 1 s sliding window with 25% overlap generates 36,689 feature frames on the same data. Consequently, using all feature frames for each sliding window parameters combination favors the combinations with smaller window and overlap sizes, as using more feature frames for training the classifier produces more accurate prediction models. To alleviate this bias, we have selected a random subset of 2000 feature frames for each sliding window configuration and performed sliding window parameter evaluations against this subset of feature frames. However, all further evaluations of the algorithm were performed on the full dataset of features.

The accuracy of the algorithm for different sliding window configurations was evaluated for both, type recognition and intensity prediction. Table 2 reports the results as the difference between individual sliding window combinations. Exercise type recognition and intensity prediction error results are presented for two types of features. Furthermore, Table 3 provides a more detailed view of intensity prediction error for different types of exercises. In both tables 0 in each row denotes the sliding window combination with optimal exercise type recognition accuracy or intensity prediction error. The numbers in other columns represent the difference in % with respect to the optimal configuration.

The results clearly show that the 2 s window length with 50% overlap yields the highest exercise recognition accuracy for both, SS and MS features. In the case of intensity prediction error, the optimal sliding window parameters are split between 1 s (SS) and 2 s (MS) sliding window length with 50% overlap. However, detailed results in Table 3 show that for SS features the 2 s combination still produces the lowest intensity prediction error for 3 out of 6 exercises (B, FL, OE), while the 1 s combination results in the lowest intensity prediction error only for a single exercise (LL). Furthermore, using the same sliding window combination for exercise type recognition and intensity prediction has the advantage of generating features only once, which is especially important in resource limited wearable systems. Thus, all further experiments are performed with the 2 s sliding window length and 50% window overlap.

6.2. Algorithm accuracy

Table 4 presents the algorithm accuracy for 2 s sliding window length with 50% overlap. The results, presented for the LOUO and CV evaluation protocol, show that the algorithm performs similarly with both types of features. As LOUO evaluation results are obtained using a more rigorous validation procedure, we used those numbers to report the algorithm performance. However, to facilitate comparison with approaches using CV, we additionally provide CV results. As expected exercise type recognition accuracy

Table 2
Exercise type recognition accuracy and intensity prediction error (both in %) for different sliding window configurations. In each row 0 denotes the sliding window configuration with optimal accuracy and error, while other configurations are expressed as a relative difference with respect to the best one. Lower numbers for exercise type recognition represent lower accuracy, higher numbers for intensity prediction denote higher error.

		1 s		2 s		4 s		6 s	
		25%	50%	25%	50%	25%	50%	25%	50%
Type	SS	−3.2	−4.6	−4.5	**0**	−1.7	−3.9	−8.1	−6.3
	MS	−3	−3.9	−0.3	**0**	−2.6	−1.9	−3.5	−3.9
Intensity	SS	0.2	**0**	1.2	0.9	3.1	2.6	2.3	2.1
	MS	0.3	0.6	0.3	**0**	1.8	2.1	1.4	1.3

Table 3
Intensity prediction error (in %) for individual exercises and different sliding window configurations. In each row 0 denotes the sliding window configuration with the lowest intensity prediction error, while other configurations are expressed as a relative difference with respect to the best one. Higher number represents higher intensity prediction error.

		1 s		2 s		4 s		6 s	
		25%	50%	25%	50%	25%	50%	25%	50%
SS	B	1.7	0.6	0.8	**0**	1	1.3	1.7	1.8
	FL	0.4	0.4	0.5	**0**	0.5	1.4	1.8	2.2
	LL	0.6	**0**	1	0.4	0.5	0.2	0.3	0.5
	T	**0**	0.4	**0**	0.8	1.9	2.3	1.3	2.3
	R	0.9	0.9	0.8	1	**0**	0.3	0.3	0.6
	OE	0.4	0.4	0.4	**0**	0.2	0.4	0.1	0.1
MS	B	0.9	0.9	0.6	**0**	0.4	0.7	0.8	0.6
	FL	0.3	0.4	0.5	0.3	0.4	0.3	0.2	**0**
	LL	0.9	0.7	0.7	0.5	0.6	0.6	0.4	**0**
	T	0.1	0.2	0.5	**0**	0.8	1.3	1.5	2.2
	R	0.1	**0**	0.9	0.6	0.6	0.4	0.7	0.8
	OE	0.5	0.6	0.1	0.1	**0**	0	0.4	0.2

Table 4
Algorithm type and intensity recognition accuracy (mean ± standard deviation) for single (SS) and multi sensor (MS) features, calculated for the leave-one-user-out (LOUO) and cross validation (CV) evaluation protocol.

	Type accuracy (%)		Intensity error (%)	
	LOUO	CV	LOUO	CV
SS	84.2 ± 11.3%	92.4 ± 0.4%	6.6 ± 2.5%	1.2 ± 0.1%
MS	86.1 ± 8%	93.6 ± 0.5%	5.7 ± 0.6%	1.2 ± 0%

and intensity prediction error are slightly better for MS features, mostly due to additional information introduced by sensor to sensor correlations. However, the difference in type recognition accuracy is less than 2%, while the intensity prediction error is only 1% off, which makes the algorithm very useful for SS implementation on resource-constrained sensors.

We additionally evaluated exercise type recognition accuracy and intensity prediction error for individual exercises. Table 5 shows that the algorithm had the most problems recognizing the front (FL) and lateral lift (LL) exercises. The following was expected as those exercises were deliberately chosen to test the algorithm's robustness to exercises with similar movement. However, although the algorithm frequently confused those two exercises, it was still able to achieve 75% recognition accuracy for them. Moreover, the algorithm more frequently misclassified triceps (T) and FL exercises as non-exercise activity (NA). The following was due to those exercises being very strenuous to perform with heavy loads. Thus, the subjects took more time between individual repetitions while performing those exercises, making the algorithm infer they stopped exercising. However, the problem is a consequence of frame-based classification and could easily be alleviated by performing decisions on a higher level, taking into account sequences of subsequent frames. E.g. if a one second NA activity is detected in the middle of a five seconds exercise activity, it's reasonable to infer that the NA activity is misclassified. A majority voting approach could be used to further improve the type recognition accuracy and provide decisions based on the dominant exercise type in a sequence of frames.

Table 6 depicts individual exercise intensity recognition errors for both feature types. Similarly as with the overall intensity recognition, the mean intensity prediction error was only slightly lower for most of the exercises using MS features. This offers promising potential for developing resource constrained solutions, as the use of MS features does not significantly improve intensity recognition error for any exercise. The results show that the algorithm had the least trouble recognizing biceps (B), rowing (R) and overhead extension (OE) exercises, while similarly to exercise type recognition was more frequently struggling with front lift (FL),

Table 5
Column-wise normalized exercise type recognition confusion matrix for MS features (in %). Rows represent actual exercise labels, while columns show algorithm predictions. To improve the readability, the cells with the value 0 are left blank.

	NA	*B*	*T*	LL	FL	*R*	OE
NA	79.2	8	16.1	11.2	15.5	3.4	3.4
B	3	92					
T	6.5		83.9				
LL	3.7			75.4	7		
FL	3			13.5	77.5		
R	2.2					96.6	
OE	2.3						96.6

Table 6
Mean and standard deviation of intensity prediction error (in %) for individual exercises.

	B	FL	LL	*T*	*R*	OE
SS	3.8 ± 2	9.2 ± 6.7	7.2 ± 2.5	8.9 ± 8.3	5.5 ± 3.9	5.1 ± 3.6
MS	4.5 ± 2.3	6.7 ± 3.6	7 ± 2.4	7.1 ± 3	5.1 ± 3.1	3.9 ± 2.1

lateral lift (LL), and triceps (T) exercises. However, the mean error for FL, LL, and *T* was still under 10% and thus fairly close to the actual intensity the subjects perceived. Similarly as with type recognition, temporal aggregation can be used to additionally improve the results. Currently, the intensity recognition error is obtained for individual frames. However the intensity could significantly fluctuate in subsequent frames by going from easy (0.1) to hard (0.9) and then back to easy (0.2). Such kind of intensity fluctuation is not realistic and could be eliminated by introducing a kind of central measure of tendency calculation, such as median, mode or geometric mean, on a per exercise basis.

6.3. Feature importance

To depict the structure of individual statistical models, we here outline the most important features used for recognizing type and intensity of different exercises. Due to space constraints we do not provide the calculated importance scores for individual features, but simply list them in decreasing order. While feature importance scores allow detailed analysis of the degree a particular feature is more or less important than other, listing the features in relative order is sufficient to understand which sensors and statistical methods are contributing the most towards recognition of a particular exercise aspect. A filter-based feature importance approach was used for both, exercise and type recognition models. In case of exercise type recognition ROC curve analysis was performed. To estimate feature importance for exercise intensity prediction a loess smoother was fit between the outcome and the predictor and the R^2 statistic was calculated to derive the relative measure of feature importance. More detailed descriptions of the methods used for calculating feature importance are available in [33].

Table 7 presents the 10 most important features used for recognition of individual exercise types. The results clearly show that the most important statistical functions were maximum (*max*), root mean square (*rms*), and correlation (*c*) of different sensor streams. While the features derived from the chest worn sensor (*C*) were less frequently selected as being important for detection of exercise types, there was no significant difference in participation of features derived from left or right arm-worn sensors. The most important features consistently included correlations between elbow- and wrist-worn sensors. The results also show that features based on the magnitude (*m*) of individual sensor streams were less frequently included in the top features list and that the most informative features were based on single-axis motion of individual sensor.

Table 8 provides the lists of top 10 most important features used for predicting the intensity of individual exercises. There is a clear difference in comparison to the most important features used for recognizing exercise types. While correlation between different sensor streams represented an important property used for recognizing exercises, it was rarely highly ranked when predicting feature importance. Standard deviation, range, minimum and maximum were most frequently selected due to those features being the most discriminative when recognizing different acceleration amplitudes. Fig. 2 provides a visual intuition for this, showing that sensors during low intensity training produce significantly lower amplitudes. This is particularly obvious in the case of the chest mounted sensor (first column), where almost no acceleration variability was recorded during the low intensity exercise, while substantial movement on more than one axis was registered during high intensity workout. The results are also consistent with the fact that in Table 8 the top three most important features for the LL exercise were derived from the chest sensor. Furthermore, in contrast to exercise type recognition, magnitudes were frequently used for calculating top scoring features, most probably due to the fact that the direction of the movement vector was not as important as its size.

Table 7
A list of top 10 most important features (in decreasing order) used for recognizing different exercise types. Each variable is expressed as a combination of the statistical function (Table 1), sensor placement (*C*: chest, *W*: wrist, *E*: elbow; *R*: right, *L*: left) and acceleration channel (*m*: magnitude, *x*, *y*, *z*).

B	FL	LL	*T*	*R*	OE
$max(W_L^y)$	$max(E_R^y)$	$max(E_R^y)$	$max(E_R^y)$	$rms(W_R^y)$	$rms(W_R^y)$
$c(E_L^y, W_L^x)$	$c(W_L^y, E_R^y)$	$rms(E_L^m)$	$c(W_L^x, E_R^x)$	$rms(E_L^y)$	$max(W_L^y)$
$max(E_R^y)$	$rms(C^z)$	$rms(C^z)$	$c(E_L^y, W_L^y)$	$rms(W_L^m)$	$rms(E_L^y)$
$c(W_R^y, W_L^x)$	$max(C^m)$	$c(E_L^y, W_L^y)$	$rms(C^z)$	$rms(W_L^x)$	$rms(W_L^m)$
$rms(E_R^x)$	$min(W_L^m)$	$c(W_R^y, E_R^y)$	$rms(C^y)$	$c(E_L^y, W_L^x)$	$rms(W_L^x)$
$c(W_R^y, E_R^x)$	$c(W_L^x, E_R^x)$	$c(W_L^y, E_R^y)$	$c(W_R^y, E_R^y)$	$c(W_R^y, W_L^x)$	$rms(E_L^m)$
$c(W_L^x, E_R^y)$	$max(E_R^m)$	$rms(C^y)$	$rms(E_L^m)$	$rms(E_R^x)$	$rms(E_R^x)$
$rms(C^z)$	$ran(W_L^y)$	$ran(W_L^y)$	$c(W_L^y, E_R^y)$	$rms(C^z)$	$c(W_L^y, E_R^y)$
$c(W_L^y, E_R^x)$	$max(C^y)$	$rms(W_R^y)$	$rms(W_R^y)$	$c(W_R^y, E_R^x)$	$c(W_L^x, E_R^y)$
$c(E_L^y, E_R^x)$	$max(W_L^m)$	$rms(E_L^y)$	$rms(E_L^y)$	$max(C^m)$	$c(E_L^y, W_L^y)$

Table 8
A list of top 10 most important features (in decreasing order) used for predicting intensity of different exercise types. Each features is expressed as a combination of the statistical function (Table 1), sensor placement (*C*: chest, *W*: wrist, *E*: elbow; *R*: right, *L*: left) and acceleration channel (*m*: magnitude, *x*, *y*, *z*).

B	FL	LL	*T*	*R*	OE
$sd(C^m)$	$ran(C^y)$	$sd(C^m)$	$sd(C^m)$	$sd(C^m)$	$ran(C^y)$
$ran(C^z)$	$ran(C^z)$	$ran(C^x)$	$ran(C^z)$	$c(W_L^x, E_R^x)$	$max(C^y)$
$ran(C^x)$	$ran(C^m)$	$max(C^y)$	$max(C^y)$	$ran(C^y)$	$ran(C^x)$
$max(C^y)$	$sd(W_L^m)$	$ran(W_L^m)$	$ran(C^x)$	$ran(C^x)$	$c(W_L^x, E_R^x)$
$ran(W_L^z)$	$ran(E_L^x)$	$ran(E_R^x)$	$ran(W_L^x)$	$c(W_R^z, E_R^y)$	$sd(C^m)$
$sd(E_R^m)$	$ran(C^x)$	$min(C^m)$	$ran(E_R^x)$	$max(W_R^x)$	$ran(E_R^z)$
$min(E_R^m)$	$ran(E_L^m)$	$ran(W_L^x)$	$min(W_L^m)$	$c(W_R^z, E_R^z)$	$ran(C^z)$
$ran(E_R^x)$	$ran(E_L^y)$	$min(W_L^m)$	$c(E_L^m, E_R^m)$	$max(W_R^m)$	$max(E_R^z)$
$min(E_R^z)$	$ran(E_L^z)$	$c(W_L^z, E_R^z)$	$max(E_R^x)$	$ran(W_R^z)$	$c(W_L^x, E_R^y)$
$ran(W_L^x)$	$ran(W_L^z)$	$max(E_R^x)$	$c(E_L^y, W_L^y)$	$c(W_R^z, W_L^z)$	$rms(E_R^z)$

6.4. Different sensor combinations

Here we evaluate the algorithm accuracy for different sensor combinations to find the optimal trade-off between the number of sensors needed and the accuracy of the algorithm. Table 9 shows exercise type recognition and intensity prediction results for sensor configurations depicted in Fig. 4. The results are expressed as a distance to the reference sensor configuration, consisting of all the five sensors used in the experiment (S5). In case of exercise type recognition positive distance means accuracy higher than

Table 9
Algorithm results (in %) for different sensor combinations in terms of type recognition accuracy and intensity prediction error using single-sensor (SS) and multi-sensor (MS) features. The results are expressed as a difference between a particular sensor combination and the full sensor setup (S5).

	Type rec. accuracy		Intensity pred. error	
	SS	MS	SS	MS
S1 (Fig. 4a)	−9.4	−11.3	3	4
S2 (Fig. 4b)	−2.6	−1.8	0.5	2.2
S3 (Fig. 4c)	−0.9	0.8	0.5	1.3
S4 (Fig. 4d)	−1.2	0.3	0.8	0.9
S5 (Fig. 4e)	0	0	0	0
S6 (Fig. 4f)	−1.5	−1.3	7.8	7.1
S7 (Fig. 4g)	−3.1	−5.1	6.3	5.6
S8 (Fig. 4h)	−1.7	−0.5	6.7	5.8

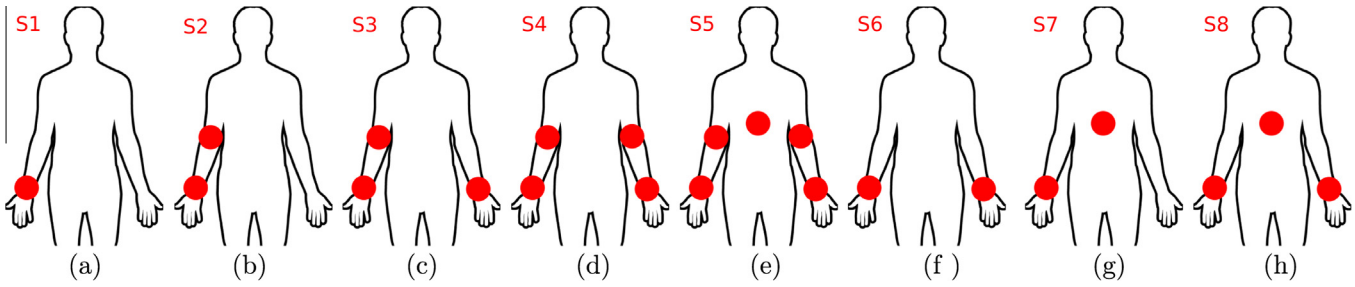

Fig. 4. Evaluated sensor combinations. Red spots depict sensor locations on a face-forward oriented body silhuete. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

those achieved by the reference configuration. In case of intensity prediction, the numbers represent prediction errors and should thus be lower and negative for better results.

The results show that the use of all five sensors achieves the lowest intensity prediction error for both, SS and MS features. However, in case of exercise type prediction, combination S3 and S4 actually outperform the reference combination for MS features. A possible reason for that could be the fact that a large number of extra MS features introduces noise in the exercise classification models and consequently degrades the results. Additionally, combinations S2, S3 and S4 yield much higher accuracy than combinations S1, S6, S7 and S8. This clearly show that single arm sensor setups are not capable of capturing full exercise dynamics and that the chest sensor location does not add any significant information and could safely be excluded without a serious impact on the algorithm accuracy.

6.5. Number of training subjects

Finally, we investigate how the number of subjects used for training the type recognition and intensity prediction models impact the algorithm accuracy. This is important as an algorithm that needs a large number of training subjects to train the supervised models would not be able to scale to a real world setting with a variety of different test subjects.

To evaluate how the number of training subjects impacts the algorithm, we train the algorithm's supervised models with different numbers of training subjects. We start by training the models with only one subject's data and continue by iteratively adding randomly chosen subjects, until there is only one subject left in the test set. Figs. 5 and 6 depict the results for different number of training subjects.

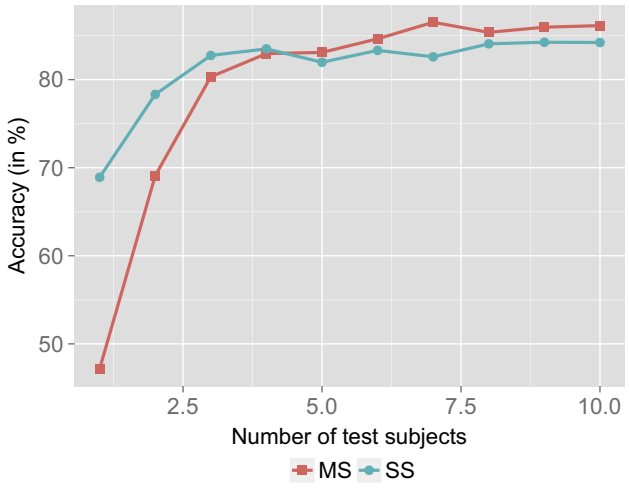

Fig. 5. Exercise type recognition accuracy (in %) for different numbers of training subjects.

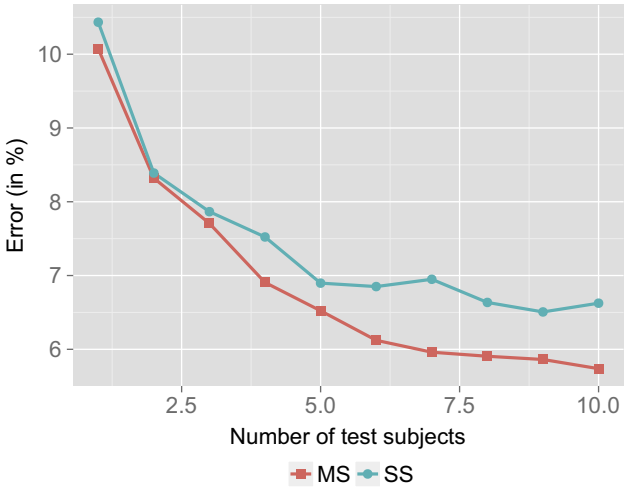

Fig. 6. Intensity prediction error (in %) for different numbers of training subjects.

The results show that exercise type recognition accuracy gets stable after the third subject has been added (Fig. 5). Adding additional subjects does not considerably improve exercise type recognition accuracy. On the other hand, intensity prediction error converges much later, after adding the sixth subject (Fig. 6). This is expected as intensity prediction is a harder task to perform. Additionally, the results reveal an interesting fact as by adding additional subjects to the test set, the exercise type recognition accuracy for both types of features converges, while in the case of intensity prediction, adding test subjects mostly benefits the MS feature group. The following means that for smaller training sets SS features perform similarly to MS features, which is promising for developing resource constrained sensor setups.

7. Discussion

In the scope of the paper a hierarchical algorithm for recognizing strength training exercise type and intensity was proposed. The results clearly show that wearable acceleration sensors present a promising technology for performing automated exercise assessment. In general, the algorithm was able to achieve 86% accuracy when performing exercise type detection and reported an average error of 6% when doing intensity prediction. It has to be pointed out that the error was not constant across different exercises, with some of them being easier to recognize than others. However, even in worst case scenario, the algorithm was still able to correctly rec-

ognize 75% of exercise types and commit only a 7% error when performing intensity prediction. In Section 6 we have already outlined some measures for improving the per exercise accuracy of the algorithm, such as performing majority voting over a range of consecutive classifier results to produce even more stable decisions. Furthermore, in the current implementation, only time domain features were used for performance evaluation of the algorithm. Calculating additional features in the frequency domain could potentially improve the algorithm results, particularly when predicting exercise intensity. Fig. 2 clearly shows that low intensity exercises (upper row) contain less high frequency components than high intensity ones (lower row). Thus, adding additional features encompassing frequency properties of the acceleration signal could further improve the results.

Although exercises selected for the experiments were targeting different muscle groups, all of them focused on the upper body. The reason for this was the limitation of the number of sensors imposed by the bluetooth piconet topology, described in Section 3. Instead of spreading the sensors across the participant's body, we decided to position them redundantly on the same body parts (e.g. wrist and elbow) to compare sensor importance for individual exercises. However, due to modular structure of the proposed algorithm, it is easy to accommodate for new exercise types. While the first layer of the recognition framework needs to be retrained, the second layer classifiers can stay intact and only need to be extended with intensity prediction models for additional exercises. Furthermore, results for evaluated exercises suggest that additional exercises including different muscle groups most probably would not degrade the overall algorithm accuracy, as exercises targeting different muscles follow different paths of motion and result in distinct movement dynamics. Existing results confirm this assumption, as exercises targeting similar muscles (FL, LL) produce lower accuracy scores as exercises encompassing different muscle groups (*B*, OE, *R*).

8. Conclusion

We have proposed an approach that uses a network of wearable sensors along with an off-the-shelf smartphone to recognize exercise intensity for a set of upper-body strength training exercises. To perform the recognition, we proposed a hierarchical algorithm that recognizes the exercise intensity in two sequential steps, by first recognizing the type of the exercise and recognizing the intensity of the exercise only once the type of the exercise is already known. We first evaluated the algorithm with different sliding window configurations. We have shown that using a sliding window of length 2 s with 50% overlap produces the best results. We than evaluated the algorithm for two groups of features with different calculation complexities in terms of exercise type recognition accuracy and intensity prediction error. The results showed, that the algorithm achieves high exercise type recognition accuracy (around 85%) and low intensity prediction error (around 6% of normalized Borg's RPE) for both feature groups. Furthermore, the results showed that high intensity exercises are harder to recognize as low intensity movement. Additionally, we evaluated the algorithm for different sensor combinations and with different number of training data. The results are promising, and show, that the algorithm produces comparable results with only a subset of sensors and training data used to train the prediction models.

In the future we plan to expand our work by including additional exercise quality parameters, such as exercise posture and dynamics. Additionally, we plan to include more rehabilitation specific exercises to study the appropriateness of the algorithm in the medical domain.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] E.T. Howley et al., Type of activity: resistance, aerobic and leisure versus occupational physical activity, Med. Sci. Sports Exercise 33 (6; SUPP) (2001) S364–S369.
[2] C.J. Hass, M.S. Feigenbaum, B.A. Franklin, Prescription of resistance training for healthy populations, Sports Med. 31 (14) (2001) 953–964.
[3] D.E. Warburton, C.W. Nicol, S.S. Bredin, Prescribing exercise as preventive therapy, Can. Med. Assoc. J. 174 (7) (2006) 961–974.
[4] B.K. Pedersen, B. Saltin, Evidence for prescribing exercise as therapy in chronic disease, Scand. J. Med. Sci. Sports 16 (S1) (2006) 3–63.
[5] A. Holzinger, M. Bruschi, W. Eder, On interactive data visualization of physiological low-cost-sensor data with focus on mental stress, Availability, Reliability, and Security in Information Systems and HCI, vol. 8127, Springer, Berlin Heidelberg, 2013, pp. 469–480.
[6] J.L. Alexander, The role of resistance exercise in weight loss, Strength Cond. J. 24 (1) (2002) 65–69.
[7] L.J. Micheli, Encyclopedia of Sports Medicine, Sage Publications, 2010.
[8] H. Kuipers, H. Keizer, Overtraining in elite athletes, Sports Med. 6 (2) (1988) 79–92.
[9] B. Celli, W.c. MacNee, A. Agusti, A. Anzueto, B. Berg, A. Buist, P. Calverley, N. Chavannes, T. Dillard, B. Fahy, Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper, Eur. Respir. J. 23 (6) (2004) 932–946.
[10] G. Borg, Psychophysical scaling with applications in physical work and the perception of exertion, Scand. J. Work Environ. Health (1990) 55–58.
[11] R. Guidoux, M. Duclos, G. Fleury, P. Lacomme, N. Lamaudière, P.-H. Manenq, L. Paris, L. Ren, S. Rousset, A smartphone-driven methodology for estimating physical activities and energy expenditure in free living conditions, J. Biomed. Inform. 52 (2014) 271–278.
[12] D. Rand, R. Kizony, U. Feintuch, N. Katz, N. Josman, P.L.T. Weiss, et al., Comparison of two VR platforms for rehabilitation: video capture versus HMD, Presence: Teleop. Virt. Environ. 14 (2) (2005) 147–160.
[13] J. Perkins, M. Pavel, J. Jimison, S. Scott, Gesture recognition for interactive exercise programs, in: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 1915–1917.
[14] T. Giorgino, P. Tormene, S. Quaglini, A multivariate time-warping based classifier for gesture recognition with wearable strain sensors, in: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 4903–4906.
[15] C. Mattmann, O. Amft, H. Harms, G. Troster, F. Clemens, Recognizing upper body postures using textile strain sensors, in: 11th IEEE International Symposium on Wearable Computers, 2007, pp. 29–36.
[16] S. Melzi, L. Borsani, M. Cesana, The virtual trainer: supervising movements through a wearable wireless sensor network, in: 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009, SECON Workshops' 09, IEEE, 2009, pp. 1–3.
[17] K.-H. Chang, M.Y. Chen, J. Canny, Tracking free-weight exercises, in: 9th International Conference on Ubiquitous Computing, UbiComp '07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 19–37.
[18] C. Seeger, A. Buchmann, K. Van Laerhoven, myHealthAssistant: a phone-based body sensor network that captures the wearer's exercises throughout the day, in: 6th International Conference on Body Area Networks, 2011, pp. 1–7.
[19] M. Muehlbauer, G. Bahle, P. Lukowicz, What can an arm holster worn smart phone do for activity recognition? in: 2011 15th Annual International Symposium on Wearable Computers (ISWC), 2011, pp. 79–82.
[20] I. Pernek, K.A. Hummel, P. Kokol, Exercise repetition detection for resistance training based on smartphones, Pers. Ubiquit. Comput. 17 (4) (2013) 771–782.
[21] G. Spina, G. Huang, A. Vaes, M. Spruit, O. Amft, COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback, in: ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13, ACM, New York, NY, USA, 2013, pp. 597–606.
[22] P. Siirtola, P. Laurinen, J. Roning, H. Kinnunen, Efficient accelerometer-based swimming exercise tracking, in: IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2011, pp. 156–161.
[23] N. Ravi, N. Dandekar, P. Mysore, M.L. Littman, Activity recognition from accelerometer data, in: 17th Conference on Innovative Applications of Artificial Intelligence, IAAI'05, vol. 3, AAAI Press, 2005, pp. 1541–1546.
[24] J.F. Knight, H.W. Bristow, S. Anastopoulou, C. Baber, A. Schwirtz, T.N. Arvanitis, Uses of accelerometer data collected from a wearable system, Pers. Ubiquit. Comput. 11 (2) (2007) 117–132.
[25] A. Ben-Hur, J. Weston, A users guide to support vector machines, in: Data Mining Techniques for the Life Sciences, Springer, 2010, pp. 223–239.
[26] L.H. Epstein, R.A. Paluch, L.E. Kalakanis, G.S. Goldfield, F.J. Cerny, J.N. Roemmich, How much activity do youth get? A quantitative review of heart-rate measured activity, Pediatrics 108 (3) (2001) e44.
[27] R. Lemoyne, C. Coroian, T. Mastroianni, W. Grundfest, Accelerometers for quantification of gait and movement disorders: a perspective review, J. Mech. Med. Biol. 8 (02) (2008) 137–152.

[28] W. Wu, S. Dasgupta, E.E. Ramirez, C. Peterson, G.J. Norman, Classification accuracies of physical activities using smartphone motion sensors, J. Med. Internet Res. 14 (5) (2012) e130.
[29] M. Ermes, J. Parkka, J. Mantyjarvi, I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, IEEE Trans. Inf. Technol. Biomed. 12 (1) (2008) 20–26.
[30] S.J. Preece, J.Y. Goulermas, L.P. Kenney, D. Howard, A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data, IEEE Trans. Biomed. Eng. 56 (3) (2009) 871–879.
[31] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: Pervasive Computing, Springer, 2004, pp. 1–17.
[32] N. Györbíró, A. Fábián, G. Hományi, An activity recognition system for mobile phones, Mob. Networks Appl. 14 (1) (2009) 82–91.
[33] M. Kuhn, Building predictive models in *R* using the caret package, J. Stat. Softw. 28 (5) (2008) 1–26.