Insights from two Studies on AI-based Learning in Strength Training

Bastian Dänekas^{1,*}, Tanja Döring¹, Tjorven Schnack¹, Georg Volkmar¹, Robert Porzel¹ and Rainer Malaka¹

¹University of Bremen, Bibliotheksstraße 1, Bremen, 28359 Bremen, Germany

Abstract

AI-based exercise execution recognition is a current topic in computer science in sports. Through different learning algorithms systems can be build, which do give feedback about the correct and wrong execution of an athlete in regard to a specific exercise. We built two exercise execution systems in two separate studies. While one was built using supervised learning and investigating the exercise push-ups, the other was created through unsupervised learning methods for the exercise of military press. Both systems were able to detect exercise execution very well for individual persons, while correct recognition rates for the whole population of participants was worse. These two studies revealed two main challenges, which are not solely solvable in the area of AI. However, HCI researchers are be able to address those challenges and to develop future inventions. This paper opens up the design space for future HCI research in AI-based exercise execution systems, where athletes will greatly benefit from.

Keywords

Strength Training, Challenges of AI, Computer Science in Sports, Supervised Learning, Unsupervised Learning, Trends in HCI and Sports

1. Introduction

Strength training is an essential component for a sustainable healthy life. Strength training is divided into different exercises, which can either be done with or without weights. If exercises are performed incorrectly, joint injuries can occur in the short or long term, which makes future training impossible and therefore has a negative effect on peoples health [1, 2, 3]. In order to prevent incorrect execution, it is recommended to consult a trainer about the correct execution. The trainer can then correct misalignments during the performance of an exercise and thus help to ensure correct execution. However, training lessons with a fitness coach are quite expensive and not every training session can be monitored by a coach.

Furthermore, technology in the area of training is constantly evolving. Cameras can be used to record and analyze motion patterns. In addition, every commercially available smartphone now contains sensors that can measure not only the position of a person on the globe but also

 ^{0000-0002-4240-0761 (}B. Dänekas); 0000-0001-8648-340X (T. Döring); 0000-0002-3421-7932 (T. Schnack); 0000-0001-8602-9442 (G. Volkmar); 0000-0002-7686-2921 (R. Porzel); 0000-0001-6463-4828 (R. Malaka)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)

NTSPORT'22: New Trends in HCI and Sports Workshop at MobileHCI'22, October 1, 2022 *Corresponding author.

[△] daenekba@uni-bremen.de (B. Dänekas); tanja.doering@uni-bremen.de (T. Döring);

tjorven.schnack@googlemail.com (T. Schnack); gvolkmar@uni-bremen.de (G. Volkmar); porzel@tzi.de (R. Porzel); malaka@tzi.de (R. Malaka)



Figure 1: Push-Up (left side) and military press (right side) repetitions as examples for AI-based exercise execution recognition systems.

acceleration and rotation at a high frequency. Resulting data can be used to provide positive feedback on exercise performance beyond a static representation of correct exercise performance by training classifiers.

However, such technical applications are often very limited by the data they are trained on. In this paper, we present insights from two studies on AI-based learning in strength training. In these studies, we evaluated two different training exercises, which can be seen in figure 1, by different classification or clustering methods and identified challenges, which will be presented in this paper. These challenges indicate future opportunities and possibilities that can be solved and explored by HCI designers.

2. Related Work

The recognition of certain movement patterns has been the focus of research for quite some time. The goals of automatic motion detection are various. For example, steps can be counted to measure the activity level, which is a good indication for overall morbidity and fitness [4, 5, 6]. Another area of application would be the detection of whether an elderly person falls and an ambulance has to be called [7, 8]. Also feasible are new possibilities of controlling video games, which allow specific interactions based on certain movement patterns. This can also be transferred to HCI, since motion pattern recognition makes new natural ways of interaction possible.

Another area of this recognition is sport and training. In the professional field, motion sequences of a player can be analyzed and improved in terms of efficiency and effectiveness

	Study 1 (S1)	Study 2 (S2)	
Learning Method	Classification (Sup. Learning)	Clustering (Unsup. Learning)	
Hardware Used	IMU	2D Camera	
Strength Exercise	Push-Up	Military Press	
# of Participants	5	18	
# of Recorded Repetitions	278	909	
# of Established Execution Classes	6	5	
# of Trainers for Labeling	2	-	

 Table 1

 Overview on the two conducted studies on AI-based recognition methods in strength training

(IMeasureU¹, MyoMotion², XSens³). Moreover, beginners, who are new to a sport, can learn the movement sequences better through such analyzed data. This detection is also helpful for the physical rehabilitation of patients who have to perform certain exercises correctly.

The most widely used, and very usable in mobile sports settings, are motion detection algorithms using inertial measurement units (IMUs) or cameras. IMUs are a combination of three accelerometers, gyroscopes and magnetometers in three-dimensional space. These sensors are nowadays present in all kinds of mobile devices such as smartphones or smartwatches. Cameras, both 2D and 3D depth cameras, are also suitable for capturing motion data. Smartphone cameras, which have both a high resolution and a front and back camera, are especially suitable for mobile environments.

Research in the domain of evaluation for AI-based exercise execution exists in both camerabased detection [9, 10, 11, 12] and IMU-based detection [13, 14, 15, 16]. Several factors come into play when weighing the hardware, which are dependent on the context. While captured camera images are easier to understand and interpret, and correct exercise execution is easier to validate, lighting and occlusion issues can occur. IMUs on the other hand are not susceptible to occlusion or lighting influences, but suffer from problems in the understandability of the captured data, can be mounted incorrectly, which has a major impact on the captured data, or suffer from long-term drift problems due to the temperature increase of the sensors.

3. Two Studies on AI-based Learning in Strength Training

In order to explore AI-based recognition methods in strength training, we conducted two studies. These studies differ e.g. in their learning method, hardware used to record the data, number of participants for recording. An overview of these aspects can be seen in table 1.

In Study S1, we investigated whether an IMU in a smartphone is suitable to be used to build a good classifier for push-ups. We developed a total of five types of error classes from qualitative interviews with a total of four trainers in addition to the correct execution class. These error classes included a misalignment of elbows, two misalignments of the hands in relation to the

¹https://imeasureu.com/, accessed 05.06.2020

²https://www.velamed.com/produkte/3d-inertialsensor-system, accessed 02.06.2022

³https://www.xsens.com/inertial-sensor-modules, accessed 09.06.2022

body, instability of the trunk during execution, and a class for other errors. After recording the data with a total of five subjects and 278 total recorded push-up repetitions, they were classified by two trainers so that supervised classification methods could be applied. Not all error classes were apparent in the recorded data set. Only the following classes remained: correct execution (KA), position of hands hands is to high in relation to the body (OS), instability of the trunk (IR), the combination of OS and IR (OSIR) and other errors (S). The data was recorded with a frequency of 103Hz as prior research suggested [16, 15]. The recorded data set was very unbalanced in the distribution of exercise executions for each participant. An example is given in table 2.

Table 2

Distributions of the respective classes per participant according to the second trainer labeling each execution

	P1	P2	P3	P4	P5	\sum
KA	28	0	0	0	0	28
OS	0	0	31	0	0	31
IR	26	49	0	0	0	75
OSIR	0	0	20	0	63	83
S	11	0	0	50	0	61
\sum	65	49	51	50	63	278

Ten signals were recorded for the classification process: x-, y-, and z-axis of the accelerometer and gyroscope in addition to yaw, pitch, roll and magnitude of acceleration. For each of these signals, ten statistical features were calculated: arithic mean, standard deviation, skewness, kurtosis, maximum, minimum, range, first quartile and third quartile. Each individual repetition data of a push-up done in a set was segmented by hand. Five different classification algorithms were used to train the data on: Random Forests, Extra Trees, Support Vector Machine, Logistic Regression and K-Nearest Neighbors. The classifiers were evaluated using a 10-fold-crossvalidation.

The multi-class classifiers gave good results with an accuracy of up to 87.44% (Random Forest), while the binary classifiers achieved an accuracy of up to 90.34% (Support Vector Machine). Classifiers that did not distinguish by exercise execution but by athlete achieved an accuracy of up to 99.64% (Extra Trees).

The second study S2 investigated whether clustering algorithms from unsupervised learning are suitable to recognize execution types of the strength training exercise military press. Based on existing literature, a total of five classes were collected, which were divided into the correct execution, an execution with too wide a grip, an execution with too small a range of motion (ROM), the push press execution class in which the legs are also bent and straightened, and the olympic execution class in which the spine is hyperextended to move more weight. A total of 909 repetitions of military press across 18 participants were recorded and skeletal points were calculated using OpenPose [17]. V-measure [18] is a metric that compares a clustering to the class of data points. The value is the harmonic mean of homogeneity and completeness. Homogeneity is a measure of whether a cluster has only data points of the same class associated with it. Completeness is a measure of whether all data points of a class are assigned to the same cluster. Values range from 0 to 1, with higher being better. A V-measure of 1 means that the

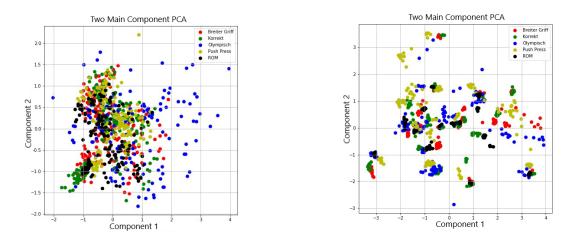


Figure 2: Representation of the skeletal data from 19 persons with normalized (left side) and without normalized (right side) pose coordinates after a PCA with two principal components.

clustering fully matches the actual assignment. While the V-Measure of individual subjects yielded very good results with an average value of 0.935, the V-Measure value for a clustering with all 18 participants resulted in only 0.024. The result after a principle component analysis with 2 main components showing the non differentiable clusters can be seen in figure 3.

Both studies revealed challenges that cannot be attributed solely to an unfavorable choice of learning algorithms and methods. In the following, we briefly summarize the studies and discuss the difficulties of classifying sports exercises.

4. Challenges in Classifying Sports Exercises

4.1. No Al for All

In both studies, classifiers who rated person-dependent exercise execution achieved much better results than classifiers who rated exercise execution for the entire population of participants. While this phenomenon in S1 can be explained by the unequal distribution of different exercise execution classes across participants, the same effect could also be observed for S2. In S2, a normalization procedure for camera-based systems was used to compensate for the different anatomical features such as upper-body-lower-body ratio, for example [19]. Based on the results of these two studies, it seems apparent that learning algorithms designed to evaluate and determine execution types of a strength exercise perform better when the data comes from a single individual.

The reason for this observation could be the individual training level and anatomy of an athlete. The difference between incorrect and correct execution in strength training can be minimal. Even a small change in the angle of the upper arms to the floor during a push-up can be decisive for which muscles and which joints are stressed to what extent. In addition, some exercises are easier for certain groups of people than others due to their anatomy. The range of

motion (ROM) of a movement varies greatly from person to person. This observation therefore additionally raises the question of whether there is a "right" or "wrong" in strength training exercises at all, or whether fine nuances do indeed depend much more on the individual person.

4.2. Inconsistency in Trainer Evaluations

Two trainers labeled the 278 push-ups recorded in S1. One trainer held a B license according to the German Olympic Sports Federation (DOSB), while the other was authorized to evaluate strength training exercises through a dual study program in fitness economics. Despite this background, both trainers achieved a very low level of agreement in the classification. Only for nine repetitions both coaches fully agreed. Partial agreement could be found for 137 repetitions, while the classification completely contradicted each other for the remaining 132 repetitions. The rating occurred without the two trainers knowing of the other labeling. Due to this disagreeing labeling, only the labels of the second trainer were used. Since in many well-known studies on the classification of weight training exercises one to two trainers simultaneously label the recorded data [16, 15], it remains questionable how good these classifiers really are - assuming that other trainers would have evaluated the exercises differently. This observation does also raise the question, for what reason the inconsistency of the trainer feedback arised and if the small nuances between right and wrong in individual exercise execution are also a cause for it.

5. Opportunities for HCI

These two observations, which are not unique to these two studies, suggest several implications. First, the classifiers could be optimized by AI methods and tailoring to the individual athlete. Furthermore, each learning method could ensure that a diverse distribution of different exercise executions is represented and reduce the uncertainty of the system.

However, an optimization of these methods will always have issues, even if they will be represented in small numbers. This is exactly where methods from HCI can help. The uncertainty of a system, which is caused by the disagreement of trainers, could be reflected by the system as feedback. If an athlete feels restricted in movement due to an injury or, conversely, has greater mobility due to muscle length training, the system should be able to reflect this exactly.

By using mobile hardware, such as the smartphone, individual virtual companions could be created through the clever use of gamification and the use of avatars, which grow with the athlete. Instead of a one-for-all solution in AI, HCI can use long-term studies to monitor and re-evaluate an athlete's development and needs in the area of strength training. By combining key elements from AI and HCI, new ways of providing accurate and useful feedback to a strength athlete are enabled.

6. Conclusion

Two studies in the application area of AI-based exercise execution recognition, revealed a total of two challenges for futures HCI researchers. The first challenge is the individualization of AI-based recognition applications. Through mobile hardware and new innovations, personalized

digital exercise trainer can adapt and reflect the progress of an individual athlete. Depending on an athletes state of mobility and strength, some ways to execute an exercise are more beneficial and therefore "right" than the same execution for another athlete. The third challenge is the uncertainty of AI-based training algorithms. Even though, those algorithms will get more accurate when applying a more fitting data set and algorithm, some uncertainty and false evaluations will still remain. Instead of just confronting the user with an uncertain result, HCI researchers can develop methods on how to reflect this uncertainty in the corresponding feedback given. We showed through this paper, that the space for future collaborations for AI and HCI methods in sports is big and solutions are needed. Moreover, we hope that multiple ideas will be discussed and researched, so that future users can benefit from mobile and AI-based strength training exercise execution applications.

References

- R. W. Westermann, M. Giblin, A. Vaske, K. Grosso, B. R. Wolf, Evaluation of men's and women's gymnastics injuries: a 10-year observational study, Sports Health 7 (2015) 161–165.
- [2] P. T. Hak, E. Hodzovic, B. Hickey, The nature and prevalence of injury during crossfit training., Journal of strength and conditioning research (2013).
- [3] S. Kaiser, T. Engeroff, D. Niederer, H. Wurm, L. Vogt, W. Banzer, The epidemiological profile of calisthenics athletes., German Journal of Sports Medicine/Deutsche Zeitschrift fur Sportmedizin 69 (2018).
- [4] E. A. Bakker, Y. A. Hartman, M. T. Hopman, N. D. Hopkins, L. E. Graves, D. W. Dunstan, G. N. Healy, T. M. Eijsvogels, D. H. Thijssen, Validity and reliability of subjective methods to assess sedentary behaviour in adults: a systematic review and meta-analysis, International Journal of Behavioral Nutrition and Physical Activity 17 (2020) 1–31.
- [5] L. Laranjo, D. Ding, B. Heleno, B. Kocaballi, J. C. Quiroz, H. L. Tong, B. Chahwan, A. L. Neves, E. Gabarron, K. P. Dao, et al., Do smartphone applications and activity trackers increase physical activity in adults? systematic review, meta-analysis and metaregression, British journal of sports medicine 55 (2021) 422–432.
- [6] A. Brajdic, R. Harle, Walk detection and step counting on unconstrained smartphones, in: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, 2013, pp. 225–234.
- [7] A. Bourke, J. O'brien, G. Lyons, Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm, Gait & posture 26 (2007) 194–199.
- [8] M. Kangas, A. Konttila, P. Lindgren, I. Winblad, T. Jämsä, Comparison of low-complexity fall detection algorithms for body attached accelerometers, Gait & posture 28 (2008) 285–291.
- [9] E. Velloso, A. Bulling, H. Gellersen, W. Ugulino, H. Fuks, Qualitative activity recognition of weight lifting exercises, in: Proceedings of the 4th Augmented Human International Conference, 2013, pp. 116–123.
- [10] S. Asteriadis, A. Chatzitofis, D. Zarpalas, D. S. Alexiadis, P. Daras, Estimating human motion from multiple kinect sensors, in: Proceedings of the 6th international conference

on computer vision/computer graphics collaboration techniques and applications, 2013, pp. 1–6.

- [11] A. Kitsikidis, K. Dimitropoulos, S. Douka, N. Grammalidis, Dance analysis using multiple kinect sensors, in: 2014 International Conference on Computer Vision Theory and Applications (VISAPP), volume 2, IEEE, 2014, pp. 789–795.
- [12] S. Kaenchan, P. Mongkolnam, B. Watanapa, S. Sathienpong, Automatic multiple kinect cameras setting for simple walking posture analysis, in: 2013 international computer science and engineering conference (ICSEC), IEEE, 2013, pp. 245–249.
- [13] A. Yurtman, B. Barshan, Detection and evaluation of physical therapy exercises by dynamic time warping using wearable motion sensor units, in: Information Sciences and Systems 2013, Springer, 2013, pp. 305–314.
- [14] O. Giggins, D. Kelly, B. Caulfield, Evaluating rehabilitation exercise performance using a single inertial measurement unit, in: 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, IEEE, 2013, pp. 49–56.
- [15] O. M. Giggins, K. T. Sweeney, B. Caulfield, Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study, Journal of neuroengineering and rehabilitation 11 (2014) 158.
- [16] D. Whelan, M. O'Reilly, T. Ward, E. Delahunt, B. Caulfield, Evaluating performance of the single leg squat exercise with a single inertial measurement unit, in: Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques, 2015, pp. 144–147.
- [17] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7291–7299.
- [18] A. Rosenberg, J. Hirschberg, V-measure: A conditional entropy-based external cluster evaluation measure, in: Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL), 2007, pp. 410–420.
- [19] S. Chen, R. R. Yang, Pose trainer: correcting exercise posture using pose estimation, arXiv preprint arXiv:2006.11718 (2020).