International Conference on Innovative Applications in Engineering and Information Technology(ICIAEIT-2017)

International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

A Random Forest based Classification Model for Human Activity Recognition

Prasnthi Mandha
Department of CS & SE,
College of Engineering,
Andhra University

Dr. G.LavanyaDevi
Assistant Professor
Dept of CSSE
AUCE(A)

Dr. S. Viziananda Row
Associate Professor
Dept of CSSE
AUCE(A)

Abstract

Human Activity Recognition is a promising area having potential to benefit the human society by developing assistive technologies in order to aid elderly, chronically ill and also for people with special needs. Accurate activity recognition is challenging because human activity is complex and highly diverse. Literature survey performed in this area has revealed data mining algorithms are employed for classification of activities. Hybrid mining techniques, Naive Bayes with SVM and C4.5 with Neural Network are proved to be efficient in classifying the accelerometers reading data. These datasets are having large set of instance with many continues values. Constructing a classifier that classify such data is still a challenging task. Random forest is known for achieving high accuracy in classification. It's robustness in classifying large datasetsis promising.This paper proposes a random forest based classification model for classifying/predicting the manner of exercises. Training data is preprocessed to attain consistency. Instances from training dataset are drawn in random for n samples, and n decision tree are constructed. Consequently, a random decision forest is constructed for classifying activates based accelerometers data values. To predict unlabeled exercise data, aggregation of n trees is performed. Experimental studies are conducted to study the activity recognition capability of the model, the results are compared with popular supervised classification techniques. It is observed that the proposed model outperformed the other classification techniques in comparative study. The designed classification model is limited to perform activity recognition in the context of weight lifting exercises. Human Activity recognition is can be applied to many real-life, human-centric problems.

I. INTRODUCTION

Human Activity Recognition (HAR) has emerged as a key research area in the last years and is gaining increasing attention by the pervasive computing research community, especially for the development of context-aware systems. There are many potential applications for HAR, like: elderly monitoring, life log systems for monitoring energy expenditure and for supporting weight-loss programs, and digital assistants for weight lifting exercises.

During the past decade, there has been an exceptional development of microelectronics and computer systems, enabling sensors and mobile devices with unprecedented characteristics. Their high computational power, small size, and low cost allow people to interact with the devices as part of their daily living. That was the genesis of *Ubiquitous Sensing*, an active research area with the main purpose of extracting knowledge from the data acquired by pervasive sensors.

The challenge specific to the design of HAR systems is to develop a clear understanding of the definition of the activities under investigation and their specific characteristics. This may seem trivial at first. But human activity is highly complex and diverse and an activity can be performed in many different ways, depending on different contexts, and for a multitude of reasons. Katz et al. developed the Activities of Daily Living (ADLs) index as a tool in elderly care. Providing a good initial taxonomy of activities, it served many researchers as an inspiration to recognize activities relevant to real-world applications. Other resources include the comprehensive compendium of physical activity. It groups physical activity in categories based on the metabolic equivalent. Another resource for activity definition is given by time use databases. These were assessed by the government to understand citizens' time use, and Partridge and Golleinvestigate the potential of this data repository for activity recognition systems. Besides providing prior probabilities for activities at a certain time of day or location, it provides a taxonomy that can serve as a good reference for activity recognition researchers.

While state-of-the-art systems achieve decent performance on many activity recognition tasks, research so far mainly focuses on recognizing "which" activity is being performed at a specific point in time. In contrast, only little work investigated means to extract qualitative information from sensor data that allow us to infer additional activity characteristics, such as the quality or correctness of executing an activity. For instance, while recognizing the task of brushing one's teeth is itself relevant and part of the ADL index, it may be even more relevant for a specific application to recognize whether this task is performed correctly. It is easy to see that such qualitative assessments are more challenging to perform automatically and have so far only been demonstrated for constrained settings, such as in sports. For general activities or physical behaviors, activity recognition research is still far from reaching a similar understanding. First, we have to learn what information about the activity is relevant for the potential application. Second, we need to identify the requirements to the recognition systems, to obtain the desired information about the activities. For example, for obtaining regularity of daily routines, it is not necessary to detect the activity, but using statistics based on clustering may be sufficient.

II.RELATED WORK

2.1 Recognition of Sports Activities

A large number of researchers have investigated means to provide computational support for sports activities. For example, Michahelles et al. investigated skiing and used an accelerometer to measure motion, force-sensing resistors to measure forces on the skier's feet and a gyroscope to measure rotation. Ermes et al. aimed to recognize several sports activities based on accelerometer and GPS data. In the weight lifting domain, Chang et al. used sensors in the athlete's gloves and waist to classify different exercises and count training repetitions. More recently, the Microsoft Kinect sensor has been used in research and uses a depth camera to extract a skeleton, which shows great potential for tracking sports activities unobtrusively.

2.2 Qualitative Assessment

While several works explored how to recognize activities only few addressed the problem of analyzing their quality. There has been work on using cameras for tracking spine and shoulders contours, in order to improve the safety and effective- ness of exercises for elder people. Moeller et al. used the sensors in a smartphone to monitor the quality of exercises performed on a balance board and provided appropriate feedback according to its analysis. Similarly, Wii Fit is a video game by Nintendo that uses a special balance board that measures the user's weight and center of balance to analyse yoga, strength, aerobics and balance exercises, providing feedback on the screen. With the objective of assessing the quality of activities Hammerla et al. used Principal Component Analysis to assess the efficiency of motion, but focused more on the algorithms rather than on the feedback. Strohrmann et al. used inertial measurement units installed on the users' foot and shin to analyse their running technique, but didn't provide feedback either.

2.3 Model-based Activity Recognition

Because sports exercises are often composed of well-defined movements, it is worth analysing approaches that leverage the capabilities of a model to analyse activities. For example, Zinnen et al. compare sensor-oriented approaches to model-based approaches in activity recognition. They proposed to extract a skeleton from accelerometer data and demonstrated that a model-based approach can increase the robustness of recognition results. In a related work, the same authors proposed a model-based approach using high-level primitives derived from a 3D human model. They broke the continuous data stream into short segments of interest in order to discover more distinctive features for Activity Recognition. Reiss et al. used a biomechanical model to estimate upper-body pose and recognize everyday and fitness activities. Finally, Beetz et al. used a model-based system to analyse football matches in which players were tracked by a receiver that triangulated microwave senders on their shin guards and on ball.

2.4 Quality in activity recognition

In order to discuss qualitative activity recognition we first need to define what we mean by the "quality of an activity". Although some works in activity recognition explored aspects of quality there is still no common understanding in the community as to what defines the quality of an activity and particularly what is "high" or "low" quality.

The term "quality" has been widely discussed in other fields, such as management research. The International Standards Association defines quality as the "degree to which a set of inherent characteristics fulfils requirements" and Crosby defines it as "conformance to specifications". What these definitions have in common is the fact that one starts with a product specification and a quality inspector measures the adherence of the final product to this specification. These definitions make it clear that in order to measure quality, a benchmark is needed to measure the quality of a product against, in this case its product specification. Adapting this idea to the qualitative activity recognition domain it becomes clear that if we can specify how an activity has to be performed we can measure the quality by comparing its execution against this specification.

From this, we define quality as the adherence of the execution of an activity to its specification. From this, we define a qualitative activity recognition system as a software arte-fact that observes the user's execution of an activity and compares it to a specification. Hence, even if there is not a single accepted way of performing an activity, if a manner of execution is specified, we can measure quality.

2.5 Qualitative activity recognition

Based on the definition of quality and qualitative activity recognition it is worth discussing which are its main aspects and challenges. Qualitative activity recognition differs from conventional activity recognition in a distinctive way. While the latter is concerned with recognizing which activity is performed, the former is concerned with assessing how (well) it is performed. Once an activity is specified, the system is able to detect mistakes and provide feedback to the user on how to correct these mistakes.

This directly raises three important questions. First, is it possible to detect mistakes in the execution of the activity. Traditional activity recognition has extensively explored how to classify different activities. Will these methods work as well for qualitative assessment of activities? The second question is how we specify activities. Two approaches are commonly used in activity recognition: a sensor-oriented approach, in which a classification algorithm is trained on the execution of activities and a model-oriented approach, in which activities are represented by a human skeleton model. The third is how to provide feedback in real-time to improve the quality of execution. Depending on how fast the system can make the assessment, the feedback will either be provided in real-time or as soon as the activity is completed. Real-time feedback has the advantage of al- lowing the user to correct his movements on the go, while an offline system might make use of more complex algorithms and provide useful information without distracting the user.

III. DATA COLLECTION AND PROCESSING

The data for this project come from Ugulino, Velloso, and Fuks's weight lifting exercises dataset.a study in which six participants were asked to perform barbell lifts correctly and incorrectly in five different ways.

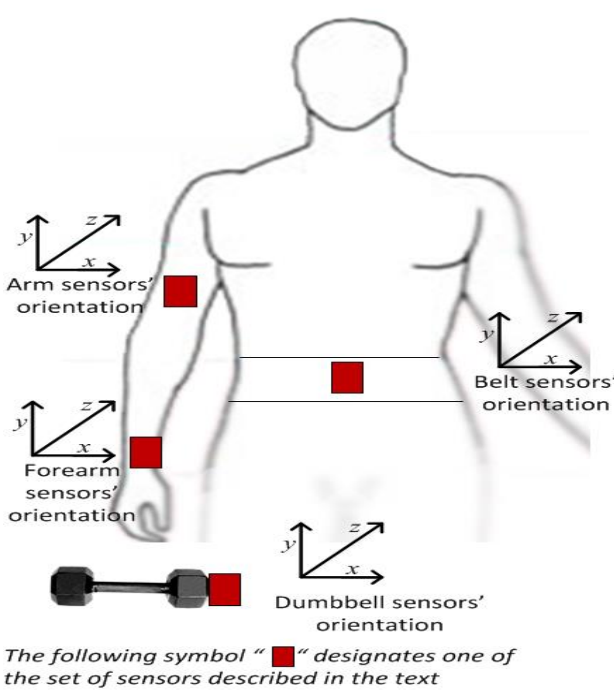

Fig 1. Accelerometers position on human body

Six young health participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different fashions: exactly according to the specification (Class A), throwing the elbows to the front (Class B), lifting the dumbbell only halfway (Class C), lowering the dumbbell only halfway (Class D) and throwing the hips to the front (Class E).

Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes. Participants were supervised by an experienced weight lifter to make sure the execution complied to the manner they were supposed to simulate. The exercises were performed by six male participants aged between 20-28 years, with little weight lifting experience. We made sure that all participants could easily simulate the mistakes in a safe and controlled manner by using a relatively light dumbbell (1.25kg).

Data pre-processing is one of the most important steps in the data mining process. It consists of filtering data, replacing the missing and outlier's values and extracting/selecting features. The first step was to load and process the training data. Values for the predictor variables were near to zero variance in the training data set and test data set .removing the variance are predictors having more than 50 percent null value in training data set and testing data set.

IV.ACTIVITY CLASSIFICATION USING KNN , CART AND RANDOM FOREST

Training and tuning HAR data using random forest technique and comparison with two other classification techniques. Random Forests (RF) consists of a combination of decision-trees. It improves the classification performance of a single-tree classifier by combining the bootstrap aggregating (bagging) method and randomization in the selection of partitioning data nodes in the construction of decision tree. The assignment of a new observation vector to a class is based on a majority vote of the different decisions provided by each tree constituting the forest. However, RF needs huge amount of labeled data to achieve good performances.

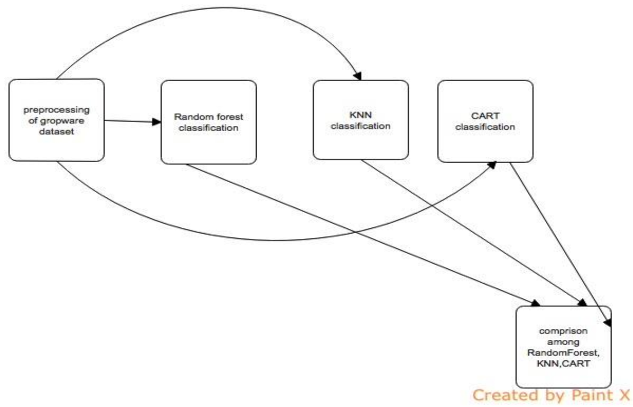

Fig 2. Data Flow Diagram

4.1 k-Nearest Neighbors

k-Nearest Neighbors (k-NN) is a supervised classification technique that can be seen as a direct classification method because it does not require a learning process. It just requires the storage of the whole data. To classify a new observation, the K-NN algorithm uses the principle of similarity (distance) between the training set and new observation to classify. The new observation is assigned to the most common class through a majority vote of its k nearest neighbors. The distance of the neighbors of an observation is calculated using a distance measurement called similarity function such as Euclidean distance. Moreover, one should note that when using the K-NN approach and a new sample is assigned to a class, the computation of distances (*i.e.*, the computation time) increases as a function of the existing examples in the dataset.

Foerster *et al.* were the first to apply the k-NN classification to differentiate between nine human activities using time-domain features obtained from three uni-axial accelerometers. In Foerster and Fahrenberg combined k-NN with a hierarchical decision approach to recognize nine activities using frequency-domain features. This approach has shown to be more efficient, in terms of classification accuracy, compared to the k-NN. Other studies based on k-NN for human activity recognition have also shown a high level of accuracy and satisfactory segmentation results.

4.2 Classification and Regression Tree

This algorithm classifies a sample according to groups of other samples with similar properties. During training, the training data is continuously divided into smaller subsets (tree nodes). When the divisions are finished, the samples are clustered together according to their properties. Testing samples are then evaluated against certain conditions in each node and propagated throughout the tree. When the sample reaches a leaf node, it is then assigned the class to which the samples in that node belong. In this paper, a binary tree with logical conditions was used. CARTs are still under extensive research and can be used even as part of larger algorithmic structures.

4.3 Random Forests

Random Forests (RF)consists of a combination of decision-trees. It improves the classification performance of a single tree classifier by combining the bootstrap aggregating (bagging) method and randomization in the selection of partitioning data nodes in the construction of decision tree. The assignment of a new observation vector to a class is based on a majority vote of the different decisions provided by each tree constituting the forest. However, RF needs huge amount of labeled data to achieve good performances.

In , the authors proposed a classification methodology to recognize, using acceleration data, different classes of motions, such as driving a car, being in a train, and walking, by comparing different machine learning techniques (Random Forests, SVM and Naive Bayes). The authors showed that Random Forest algorithm provides the highest average accuracy outperforming the SVMs and the Naive Bayes.

Implementation and comparison analysis done by using r programming ,with r packages.

V. EVALUATION AND RESULTS :

Evaluation : The accuracy measure is used to evaluate the classifiers performances. In fact, this metric measures the proportion of correctly classified examples. In the case of binary classification, the accuracy can be expressed as follows:

Accuracy = $\frac{Tp+Tn}{Tp+Tn+Fp+Fn}$

where T_n (true negatives) represents the correct classifications of negative examples, T_p (true positives) represents the correct classifications of positive examples. F_n (false negatives) and F_p (false positives) represent, respectively the positive examples incorrectly classified into the negative classes and the negative examples incorrectly classified into the positive classes. The accuracy measure does not take into account the unbalanced datasets. In this case, the accuracy is particularly biased to favor the majority classes. The following evaluation criteria are considered: average of the accuracy rate (R) and its standard deviation (std), F-measure, recall, precision and specificity.

The F-measure is defined as the combination of two criteria, the precision and the recall, which are defined as follows:

precision = $\frac{Tp}{Tp+Fp}$

recall= $\frac{Tp}{Tp+Fn}$

The specificity (SPC) is also used to evaluate the performances of the different algorithms and is calculated as follows:

specificity = $\frac{Tn}{Tn+Fp}$

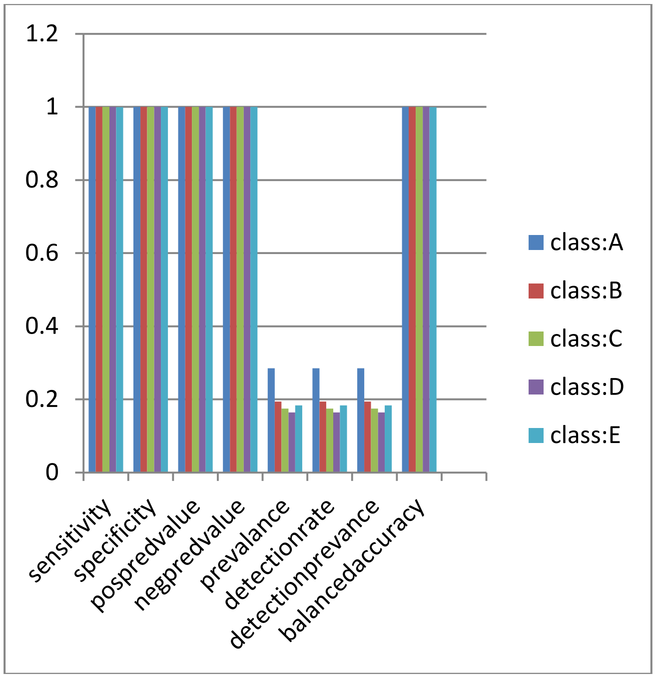

Fig 3. Confusion matrix chart for RF

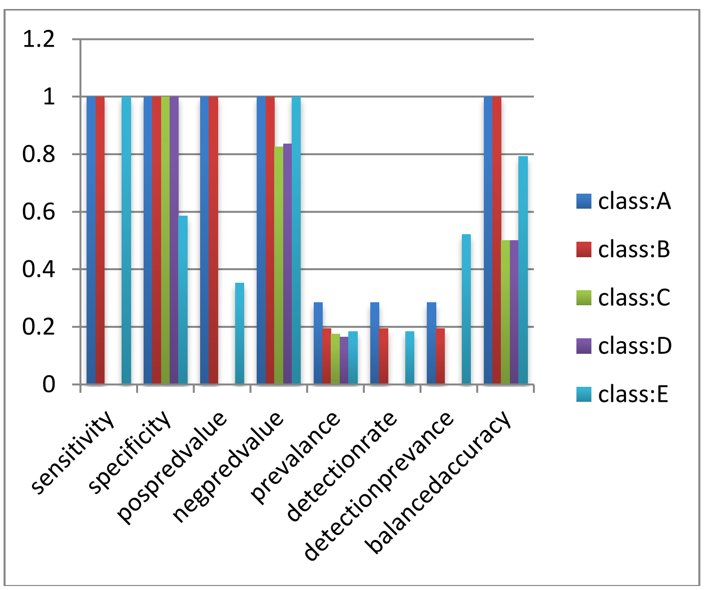

Fig 4. Confusion matrix chart for CART

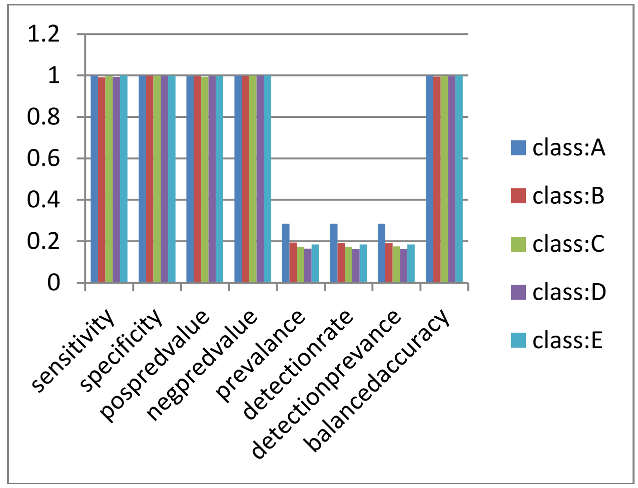

Fig 5. Confusion matrix chart for KNN

Error rate of CART,KNN and RF:

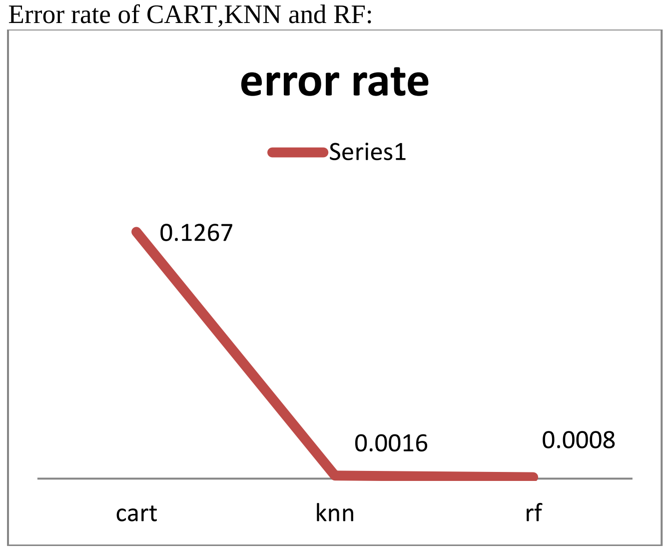

Fig 6. Error rate of KNN,CART,RF

Classification accuracy comparison:

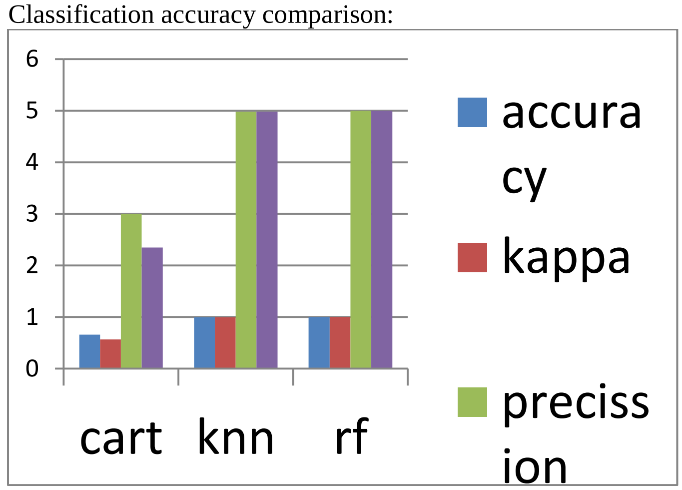

Fig 7. Classification accuracy

VI. CONCLUSIONS

The Random Tree model predicted over the test data-set with a 99.97% accuracy. KNN model predicted over the test data-set with a 99.59% accuracy. CART Tree model predicted over the test data-set with a 66.17% accuracy. This results demonstrates that Random Forest model was the correct choice to analyze the data. The weight lifting training data-set was used to create a model that predicted the way a subset performed the weight lifting exercise.The designed classification model is limited to perform activity recognition in the context of weight lifting exercises.

REFERENCES

[1] J. Perez, M. A. Labrador, and S. J. Barbeau, "G-sense: A scalable architecture for global sensing and monitoring," *IEEE Network*, vol. 24, pp. 57–64, 2010.

[2] Ulf Blanke and BerntSchiele, "Remember and transfer what you have learned recognizing composite activities based on activity spotting ", *Proceedings of ISWC*, pp. 1–8, 2010.

[3] K. Partridge and P. Golle, " On using existing time-use study data for ubiquitous computing applications", *Proceedings of UbiComp*, ACM, pp. 144–153, 2008.

[4] Bernd Tessendorf, Andreas Bulling, Daniel Roggen, Thomas Stiefmeier, Manuela Feilner, Peter Derleth, and Gerhard Tro¨ster. 2011a. Recognition of hearing needs from body and eye movements to improve hearing instruments. In Proceedings of the 9th International Conference on Pervasive Computing. Springer, 314–331.

[5] DOI:http://dx.doi.org/10.1007/978-3-642-21726-5_20 Bernd Tessendorf, Franz Gravenhorst, Bert Arnrich, and Gerhard Tro¨ster. 2011b. An IMU-based sensor network to continuously monitor rowing technique on the water. In *Proceedings of ISSNIP*.

[6] S. Katz, T. D. Downs, H. R. Cash, and R. C. Grotz. 1970. Progress in development of the index of ADL. The Gerontologist 10, 1 Part 1 (1970), 20.

[7] K. Partridge and P. Golle. 2008. On using existing time-use study data for ubiquitous computing applications. In *Proceedings of UbiComp*. ACM, 144–153.

[8] Matthias Kranz, Andreas Mo¨ller, Nils Hammerla, Stefan Diewald, Thomas Plo ¨tz, Patrick Olivier, and Luis Roalter. 2013. The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices. Pervasive and Mobile Computing 9, 2 (2013), 203–215.

[9] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. 2013b. Qualitative activity recognition of weight lifting exercises.In Proceedings of the 4th Augmented Human International Conference (AugmentedHuman 2013). 116–123. DOI:http://dx.doi.org/10.1145/2459236.2459256 .

[10] Taˆm Huynh, Mario Fritz, and BerntSchiele. 2008. Discovery of activity patterns using topic models. In *Proceedings of UbiComp*. 10–19.

[11] F. Michahelles and B. Schiele, "Sensing and monitoring professional skiers ". *IEEE Pervasive Computing*, 4(3), pp. 40–46, 2005.

a. M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. Information Technology in Biomedicine, IEEE Transactions on, 12(1):20 –26, jan. 2008.

[12] K.-h. Chang, M. Y. Chen, and J. Canny. Tracking free-weight exercises. In Proceedings of the 9th ACM International Conference on Ubiquitous computing, volume 4717, pages 19–37. Springer, 2007.

[13] T. Dutta. Evaluation of the kinectZ´ sensor for 3-d kinematic measurement in the workplace.Applied Ergonomics, 2011.

[14] T. C. Havens, G. L. Alexander, C. Abbott, J. M. Keller, M. Skubic, and M. Rantz. Contour tracking of human exercises. In Proceedings of Computational Intelligence for Visual Intelligence, pages 22–28, 2009.

[15] Moeller, L. Roalter, S. Diewald, M. Kranz, N. Hammerla, P. Olivier, and T. Ploetz. Gymskill: A personal trainer for physical exercises. In Proceedings of the International Conference on Pervasive Computing and Communications, 2012.

[16] N. Y. Hammerla, T. Ploetz, P. Andras, and P. Olivier. Assessing motor performance with pca. In Proceedings of the International Workshop on Frontiers in Activity Recognition using Pervasive Sensing, 2011.

[17] Strohrmann, H. Harms, G. Tro ster, S. Hensler, and R. Mu ller. Out of the lab and into the woods: Kinematic analysis in running using wearable sensors. In Proceedings of the 13th ACM International Conference on Ubiquitous Computing (UbiComp 2011). ACM, Sept. 2011.

[18] Zinnen, U. Blanke, and B. Schiele. An analysis of sensor-oriented vs. model-based activity recognition. In Wearable Computers, 2009.ISWC '09. International Symposium on, pages 93 –100, sept. 2009.

[19] Zinnen, C. Wojek, and B. Schiele. Multi activity recognition based on bodymodel-derived primitives. In Proceedings of the 4th International Symposium on Location and Context Awareness, LoCA '09, pages 1–18, Berlin, Heidelberg, 2009. Springer-Verlag.

[20] Reiss, G. Hendeby, G. Bleser, and D. Stricker. Activity recognition using biomechanical model based pose estimation. In Smart Sensing and Context, volume 6446 of Lecture Notes in Computer Science, pages 42–55. Springer, 2010.

[21] M. Beetz, B. Kirchlechner, and M. Lames. Computerized real-time analysis of football games. IEEE Pervasive Computing, 4(3):33–39, 2005.

[22] O. F. Standardisation. Iso 9000:2005 quality management systems – fundamentals and vocabulary. Management, 3, 2005.

[23] P. B. Crosby. Quality is Free. McGraw-Hill, 1979. M. A. Hall. Correlation-based Feature Subset Selection for Machine Learning. PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, Apr. 1999.

[24] L. B. Statistics and L. Breiman. Random forests. In Machine Learning, pages 5–32, 2001. L. Breiman. Bagging predictors. Machine Learning, 24:123–140, 1996. 10.1007/BF00058655.

[25] Duda, R.O.; Hart, P.E.; Stork, D.G. Pattern

Classification; John Wiley & Sons: Malden, MA, USA, 1999.

[26] Theodoridis, S.; Pikrakis, A.; Koutroumbas, K.; Cavouras, D. Introduction to Pattern Recognition: A Matlab Approach; Academic Press: Waltham, NA, USA, 2010.

[27] Trabelsi, D.; Mohammed, S.; Chamroukhi, F.; Oukhellou, L.; Amirat, Y. An unsupervised approach for automatic activity recognition based on hidden Markov model regression. IEEE Trans. Autom. Sci. Eng. 2013, 10, 829–835.

[28] Foerster, F.; Smeja, M.; Fahrenberg, J. Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring. Comput. Hum. Behav. 1999, 15, 571–583.

[29] Zhang, T.; Wang, J.; Xu, L.; Liu, P. Using Wearable Sensor and NMF Algorithm to Realize Ambulatory Fall Detection. In Advances in Natural Computation; Springer: Berlin, Germany, 2006; pp. 488–491.

[30] Chamroukhi, F.; Mohammed, S.; Trabelsi, D.; Oukhellou, L.; Amirat, Y. Joint segmentation of multivariate time series with hidden process regression for human activity recognition. Neurocomputing 2013, 120, 633–644.

[31] Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. Classification and Regression Trees; CRC press: Boca Raton, FL, USA, 1984.

[32] Velloso. E.: Bulling, A; Gellersen, H.; Ugulino, W.; Fuks. H. Qualitative Activity Recognition of Weight Lifting Exercises . Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13). Stuttgart. Germany: ACM SIGCHI, 20"! 3. Available at http://groupware.les.inf.puc-rio.br/work.jsf7p1 =11201

[33] Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiu, R.; Fuks, H. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. Proceedings of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012. In: Lecture Notes in Computer Science, pp. 52-61. Curitiba, PR: Springer Berlin / Heidelberg, 2012. ISBN 978-3-642-34458-9. DOI: 10.1007/978-3-642-34459-6_6.http://groupware.les.inf.puc-rio.br/work.jsf7p1=10335.