The Impact of Data Reduction on Wearable-Based Human Activity Recognition

Hosein Nourani and Emad Shihab
Dept. of Computer Science & Software Engineering
Concordia University
Montreal, Canada
{h_nouran, eshihab}@encs.concordia.ca

Omid Sarbishei
Research and Development (R&D)
Motsai Inc.
Montreal, Canada
o.sarbishei@motsai.com

***Abstract*—One crucial step toward improving any pattern recognition model is refining the data (feature extraction) and simplifying it (feature selection) for the classifier. In this paper, we investigate the impact of feature reduction on the performance of HAR. We collected step data from two subjects and answer research questions related to the impact of feature reduction in terms of performance, generalizability and varying classifiers. Our findings indicate feature reduction can reduce the number of features by close to 90%, while only having an impact of 1-2% in model performance. Moreover, we find that feature reduction can impact the generalizability of HAR models. Lastly, we find that feature reduction does not have a major impact on most classifiers examined. Our results are useful for designers of HAR systems to help them optimize their models while ensuring high performance.**

***Index Terms*—Data Reduction, Feature Selection, Human Activity Recognition, Wearable, IMU, Inertial Sensors, Neural Network, Linear Regression**

I. Introduction

Pervasive sensors that can be conveniently worn and ubiquitously used aiming at a wide range of potential applications including various individual's health monitoring, rehabilitation, and intelligent assistance [1]. These sensors have begun to reduce in size, become more accurate, and more popular (e.g., Smart phones and Wearables) [2], leading to extensive research that improves algorithms inferring meaningful knowledge from sensor data. In recent years, many studies in Human Activity Recognition (HAR) using Wearables have been carried out that provides a promising performance [3] [4], [5] [6].

The process of HAR, in the well-known form, contains five main phases [4], [5] including 1)*Data collection*, 2)*Segmentation*, 3)*Feature extraction*, 4)*Feature selection*, 5)*Prediction*. First, the data is acquired by motion sensors in form of data streams. Next, these streams will be segmented using the time windows technique (e.g., a window with the length of 5 seconds shifting every 200ms). In the third, the useful features are extracted using Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), or any other feature extraction schemes. At this step, a feature may be considered as *relevant* if the classifier improves performance using it; and, conversely, *redundant* if it does not improve the performance of the classifier [1]. Then, finding a set of features containing the Maximum Relevant and Minimum Redundant features (mRMR) is the main goal in *feature selection* phase. This data is used to train a classifier; Then, it is utilized in predicting one or more specific activities.

Previous work introduced a wide range of features that could be beneficial for HAR. Time/frequency domain features, auto correlation features, histogram bins are some examples of such features. One simple solution is to give all features to the classifier, whether they are relevant or not. However, collecting and calculating features comes at a computational cost. Particularly in the case of HAR, which typically is done on resource constrained Wearables. Moreover, using more features than needed could lead to many unwanted side effects including lower model performance, overfitting and higher cost and execution time [7]–[9]. Therefore, we need a procedure to refine and intelligently select the best features.

The goal of this paper is to examine the impact of feature reduction on wearable-based HAR. Specifically, we aim at examining the trade-off between feature reduction and model performance (RQ1), model generalizability (RQ2) and different classifiers (RQ3). We perform our experiments using step data collected using the Neblina system-on-module chip. Generally, our data contains more than 2,000 steps from two different subjects. We extracted a total of 119 different features from the Neblina, which were used to examine the impact of feature reduction on HAR.

Our findings showed that feature reduction can reduce the number of features by close to 90%, while only having an impact of 1-2% in model performance. Feature reduction can impact the performance of the general models (i.e., that are cross-subject), however, which subject a model is trained on does matter. Feature reduction does not have a considerably impact on most classifiers examined.

The rest of the paper is organized as follows. The state of the art data reduction methods for HAR are presented in Section II. Section III sets up our case study, providing details about the dataset, feature extraction & selection and classifiers used. Section IV presents our results. Section V discusses the relation between features and sensors. Section VI concludes the paper.

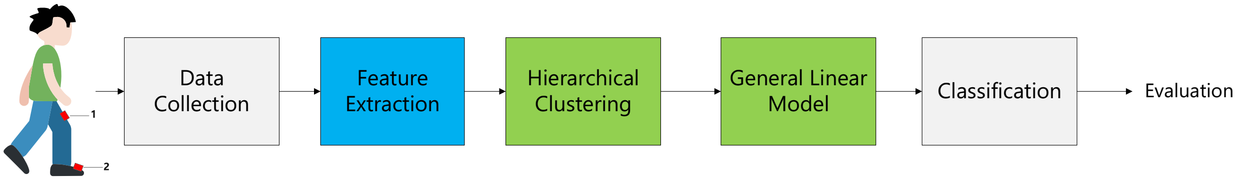

Fig. 1: Main approach for Human Activity Recognition. Two red squares show sensor positions on the thigh and foot of the subject. The green blocks show the feature selection phase.

II. Related Works

Comprehensive reviews about the subject of feature reduction are available in the literature. In [4], the accuracy of 293 classifiers are evaluated using 14 datasets involving accelerometer sensor data. Similarly, the approach in [10] uses principal component analysis (PCA) to feed the classifiers. Since the dataset contains recording data under different setups, using PCA, they extract those features that are independent from x/y/z axes and consequently independent from sensor orientation. Then, it lets them to treat identically with different datasets. They figured out the ensemble methods of KNN provides the best recognition rate and Decision Tree (DT) provides the worst. They also showed, on average, the best and worst positions for attaching sensor are right thigh and left lower arm, respectively. Similarly, Shoaib et al. [6] performed an experiment with 10 subjects and 5 sensor positions to show impact of sensor positions on activity recognition. Their results are also confirmed that right pocket (upper thigh) and wrist are respectively the best and worst positions. Comparing different featuresets, they also concluded that selecting the best sensor (accelerometer vs gyroscope) to achieve best performance depends on body position, activity type, and classifier.

The authors in [11] extracted three feature-sets including time-domain, frequency domain, and wavelet-domain statistics. They employed an ensemble selection on five feature selection methods and showed that their best results is achieved using time domain features. Zhang et al. [12] extracted some self-designed features called physical features and showed that these features have more contributions rather time-domain features to the recognition system. Introducing a multi-layer classifier, they also show that different feature-sets are appropriate for different activities.

More approaches on feature selection methods such as recursive feature selection [7], correlation based feature selection (CFS) [13], Independent Component Analysis (ICA) [14], and Local Discriminant Analysis (LDA) [15], targeting HAR, also exist in the literature.

As mentioned above, existing works mostly employ different forms of feature selection to find best performance of their model. In this work, we investigate feature selection attributes independently and in relation with the whole model. For feature selection method introduced in this work, the most relevant previous work is from Ienco et al. [16], who similarly divides the process into two stages and uses a hierarchical clustering followed by a wrapper method. They show that their method (is not for HAR) on various datasets outperforms filter and wrapper methods. Furthermore, using the dendrogram of features provided by hierarchical clustering gives a semantic view of feature space. However, they do not explain how much their method can reduce the size of feature set and how it effects on the generality of the models. In this work, we address these aspects as well as we will have a deeper view on advantages of data reduction on HAR pipeline..

III. Study Setup

Our main approach follows a typical HAR classification solution, as shown in Figure 1. The approach is composed of five main phases described in the following:

A. Data Collection

In order to examine the effectiveness of our feature selection method on data reduction, we first need to collect data of certain activities. Specifically, we selected walking in flat steps, ascending up stairs and descending down stairs [17] as our target activities. We repeat the experiments over two subjects and over two sensor positions.

Senors. We leveraged Motsai's Neblina system-on-module (SoM) solution. Neblina is a customizable module that is equipped with a tri-axial gyroscope, accelerometer, and magnetometer in conjunction with a 32-bit processor and 2X256KB of flash memory [18]. The data come from *Neblina* composes of the following features:

Acceleration data (ax,ay,az). Gyroscope data(gx/gy/gz). Magnetometer data (mx/my/mz). Force data, i.e., the acceleration vector minus gravity (fx/fy/fz). Euler Angle data (yaw/roll/pitch). We also calculate cosine for angle of roll and pitch and call them roll2 and pitch2. Time stamp (time). Step type (st) which contains the labels corresponding to each step type. The collected data has been recorded on Neblina, and is later downloaded to a Windows machine. The sampling rate was set to 50Hz [1]. The total steps in each trial is given by Table I.

Experiments. We collected data from two male participants between the ages of 25 and 30. Similar to prior studies [12], we attached the sensor to the thigh [19] and foot of each participant and perform the trials at various indoor and outdoor locations without supervision. The sensor was strapped by an elastic belt on the front of right thigh. Two sensor positions are shown in Figure 1.

TABLE I: Total number of steps based on subjects, activity types, and sensor positions.

Sensor Position	**Thigh**			**Foot**		
Activity Type	Up	Down	Flat	Up	Down	Flat
Subject A	249	228	420	206	219	386
Subject B	265	250	770	222	242	504

B. Feature Extraction and Selection

To divide the stream into segments corresponding to each step, we leveraged a pedometer [20]. Since the time-domain features have already been shown to be quite effective for HAR as opposed to frequency-domain and wavelet features [6], we have chosen the following time-domain features for our analysis, namely mean, median, variance, standard deviation, root mean square, mean absolute deviation and median absolute deviation. Then, we have 17X7=119 feature-dimension values for each step span.

Feature Selection. There are many different techniques that can be applied to select features. These methods generally are divided into three major categories including a) filter methods, b) wrapper methods, and c) embedded methods [21]. In this work, we employ an embedded method that is a heuristic approach orientated toward the notion of minimizing redundancy and maximizing relevancy (mRMR). Explicitly, our method benefits from two inexpensive blocks (green blocks in Figure 1) to find an optimum set of features. Firstly, it filters highly correlated features out. Next, it ranks them, using GLM, on a basis of how much features are statistically significant and takes the top ones. These blocks are explained as follow:

Hierarchical Clustering block. The main goal of this block is to find those features that have the minimum redundancy between them. In other words, it aims at discovering a set of lowest correlated features. To achieve that, we use hierarchical clustering (HC) [22] method, and measure the Spearman correlation coefficient through all features. Then, we need to split the tree into number of clusters. For this reason, we have to define a cut-off line in the dendrogram. In this work, as in [23], we put the cut-off line on 0.7 and it returns 15 clusters. It means that the correlation among clusters are between -0.3 and +0.3. At the second step, we get a representative feature from each cluster. To aim this, using the ability of features in predicting each other, we employ *Goodman and Kruskal* [24] algorithm to measure which feature is more appropriate to be kept to represent others. Then, we take those representatives that are better in predicting other features in a certain cluster. Using this method, we end up with 91 features (out of 118).

General Linear Model block The main goal of this block is to measure the features in terms of their contribution in predicting response. With aim of this goal, we train a linear model feeding features received from prior block. Then, using p-value (< 0.05) we take the statistically significant features in the trained model. Taking features with p-value less than significance level, we will have more certain candidates to be fed with the classifier at the following phase.

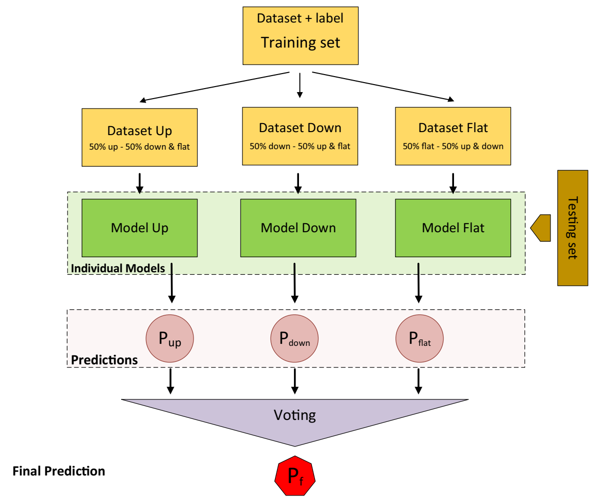

Fig. 2: Ensemble Strategy. Training m models to detect M activities. In voting block, the best prediction gets selected.

C. Classification Model

In this section we explain the mechanism of our classifier. More classifiers are also taken into consideration in RQ3. Janidarmian et al. [5] and [25] applied an ensemble of classifiers on HAR and showed that it outperforms other conventional classifiers in dealing with more difficult problems.

Following the state-of-the-art, we used an ensemble of GLMs in this work. Intuitively, instead of training each classifier for all classes, we assign each class (step type) to one classifier. Therefore, to predict 3 step types, we train 3 models (individual models in Figure 2). Then, using a voting classifier which is an ensemble of classifiers method, we combine their results into one final decision. In this work, we choose the class with the highest score through the voting block. For example, for certain input data, if individual models predict as following: $P_{up} = 0.8, P_{down} = 0.5, P_{flat} = 0.2$, the voting block infers "Up" as the final prediction. To improve the classification, during training, each individual model has been provided with an exclusive dataset biased toward a certain step type. However, during testing phase, only one dataset supplies all classifiers.

D. Performance Evaluation

Adopting a 10-folds cross validation strategy, we divide our data into ten folds; Nine folds to train the model and one left to test it. One step will be considered as *unknown* if the highest score is gained by more than one individual model (e.g., $P_{up} = P_{down}$). Since our test set contains no unknown labelled data, any unknown step prediction will be considered as false negative (FN). With the same rational, we will not have true negative (TN) result in our evaluation. To identify true positive (TP), we consider all correct step type predictions. And to identify false positive (FP), we consider all incorrect step type predictions. Using the predicted value for each step, we are able to calculate precision ($\frac{TP}{Tp+FP}$ and recall

($\frac{TP}{TP+FN}$). We then present our results using Accuracy, F-measure, and Mis-Classification to evaluate the performance.
Accuracy: Measures the rate of correctly classified step and non-steps types over all steps. It is calculated as $\frac{TP+TN}{TP+FP+TN+FN}$
F-measure (F1): Presents the harmonic mean between precision and recall. It is calculated as $2 * \frac{precision*recall}{precision+recall}$
Mis-Classification (MC): Measures the rate of incorrectly classified steps and non-steps over all of steps.

IV. Case Study Results

In this section we present the results of our experiments that answer our research questions.

A. *RQ1- How much does our feature reduction impact performance?*

As motivated earlier, most wearable devices that are used for HAR are resource constrained. Hence, reducing the amount of features (and consequently data to be collected) may improve the power consumption, latency, memory usage, etc. of these wearable devices. At the same time, the general belief is that reducing the number of features fed to a classifier also negatively impacts accuracy [11]. The focus of this question is - what is the tradeoff between the amount of data we can reduce and the performance impact.

Similar to prior work, we use accuracy, F1-measure and the misclassification rate [26] to measure the impact of performance and use the number of features used in the model to measure the data savings. To compare the two setups, we conduct one experiment using all of the features available to us and then repeat the same experiment using our reduced set of features. The differences in performance and data savings between the two experiments are then reported.

Table II shows the results of our experiments for the two subjects, A and B (results of the full model are shown in parenthesis). In each Table, the first line is result of the ensemble model (considering all steps) followed by results of individual models, i.e., step up, down and flat. From the Table II **we see that we are able to reduce the number of features by 92% ±1% (from 119 to 8) features, which impacts the performance of models for both subjects by approximately 1 - 2%**. Interestingly, the flat walking model works better after the feature reduction for both subjects. The step up model has the lowest accuracy (97%, which is still quite high) among the models examined. Comparing the results of two subjects, at the same level of performance, the total number of features for subject A is 30% lower than subject B (8 vs. 12 features), which indicates that the reduction may be subject specific. Either way, for both subjects though the reduction in features is significant.

B. *RQ2- How does the feature reduction impact the generalizability of the model?*

As we have seen from the results of RQ1, different individuals do not perform the same HAR activity in the same way [5].

TABLE II: Impact of data reduction on performance of model. The numbers in parenthesis are results of base-line model (using 119 features). The model ”All Steps” means that it can classify all three step types.

(Subject A)

Model	No. Features	Accuracy	F1	MC
All Steps	(119)**8**	(0.99)**0.98**	(1.00)**0.99**	(0.01)**0.02**
Step Up	(119)8	(0.97)0.95	(0.95)0.93	(0.03)0.03
Step Down	(119)**7**	(0.99)0.99	(0.99)0.99	(0.0)0.01
Flat walking	(119)9	(1.00)1.00	(1.0)1.00	(0.0)0.01

(Subject B)

Model	No. Features	Accuracy	F1	MC
All Steps	(119)**12**	(0.99)**0.98**	(1.00)**0.99**	(0.01)**0.02**
Step Up	(119)**11**	(0.97)0.99	(0.95)0.99	(0.03)0.03
Step Down	(119)13	(0.99)0.99	(0.99)0.99	(0.00)0.01
Flat walking	(119)12	(1.00)1.00	(1.00)0.99	(0.00) 0.01

The pattern of doing the activity depends on many factors, including the physical body of the subject, his/her level of fatigue, experiences, and so on. Consequently, a model trained on one subject may not be applicable to another subject [6], [27]. In our case, we are interested in examining the impact of the feature reduction on the generality of the model.
Cross-subject validation uses the data from one subject to train the model, then tests the model on data from another (independent subject). In this paper, as our target is to examine the impacts of feature reduction, hence, similar to the case of RQ1, we repeat the cross-subject validation twice, once with all features that are available to us and once with the reduced set of features.

Table III shows results for both experiments (results of the full model are shown in parenthesis). In the top Table, we train on data from subject A and test on subject B's data and vice versa for the Table on the bottom. First, we see that the feature reduction does decrease performance, however, its performance is comparable. Second, we notice that although the model trained on subject B's data has 3 more features (less data reduction) than the model trained using subject A's data (15 against 12), it does not provide a higher accuracy (81% vs. 80%). **Overall, we conclude that although feature reduction does impact the overall performance when evaluated across subjects, the impact is not significant. That said, again, which subject you train and test on does impact the results.**

C. *RQ3- How does feature reduction impact different classifiers?*

In most related work, the authors evaluate their feature selection method using different classifiers to identify the best model. However, different classifiers are affected by feature reduction differently. Prior work examined various different classifiers and showed that they deal with feature dimensionality differently [28]. However, their setting was

TABLE III: Cross-subject validation results on two subjects. A vs B means testing model of subject A on data of subject B.

Subject A v.s. Subject B

Model	No. Features	Accuracy	F1	MC
All Steps	(119)**12**	(0.93)**0.80**	(0.97)**0.89**	(0.07)**0.20**
Step Up	(119)15	(0.93)0.86	(0.90)0.81	(0.03)0.09
Step Down	(119)**10**	(0.97)0.87	(0.96)0.82	(0.03)0.17
Flat walking	(119)12	(0.97)0.86	(0.96)0.75	(0.15)0.35

Subject B v.s. Subject A

Model	No. Features	Accuracy	F1	MC
All Steps	(119)**15**	(0.96)**0.81**	(0.98)**0.89**	(0.04)**0.19**
Step Up	(119)13	(0.95)0.86	(0.92)0.79	(0.05)0.24
Step Down	(119)**12**	(0.97)0.88	(0.96)0.83	(0.02)0.08
Flat walking	(119)20	(0.99)0.84	(0.98)0.73	(0.05)0.26

slightly different since they used PCA, which may reduce dimensionality, however it is not guaranteed to reduce the number of needed features since one PC may be a combination of many features.

Therefore, in this RQ, we investigate the impact of feature reduction on 6 of the most common classifiers used in HAR. Again, we build a model using all of the features available to use and compare that with a model built using the reduced feature set. We merge the data from both, subject A and B to perform this analysis. We mostly used the default parameter settings for the various models, except for the Neural Network model, in which we used a configuration that was recommended in earlier work [29]. The NN model used a 5-layer network utilizing two drop-out layers and three dense fully connected layers. Layers use rectified linear (ReLU) activation functions except for a Softmax activation on the one-hot output layer.

Table IV shows the results of our experiment. As we can see from the Table, the models perform very well, with and without feature reduction. In terms of F1-measure, GLM, NN and RF slightly outperform the SVM, KNN and BT models. That said, all models do not seem to be impacted much by the feature reduction. In general, the Random Forest model seems to perform the best overall, and for that model, the feature reduction only impacts the F1-measure by 1%. **Overall, we see that most models are quite robust to the feature reduction.**

V. Discussion and Future Work

Limitations. One of our contributions is introducing a feature selection method that showed a significant result in reducing data size. However, this method may not provide the best results as we did not compare its result with any other conventional feature selection methods. For this reason, more validation of feature selection method is important future work. In addition, as showed in RQ2, the result might be affected by certain subject. A wider range of activities beside more number of subjects are required to decrease the impact of subjects. So, using big enough datasets like [30] will be considered in our future works.

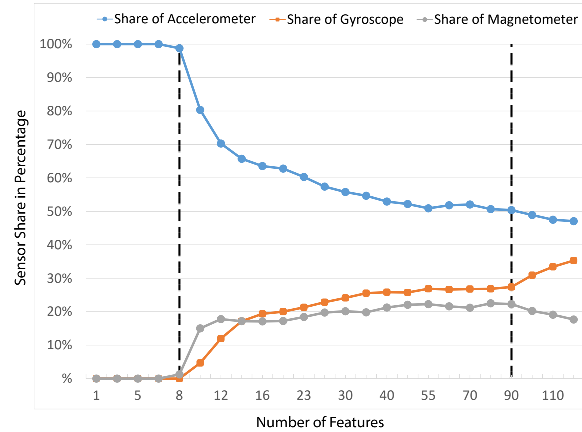

Fig. 3: contribution of three sensors over different sizes of featureset

TABLE IV: Impact of data reduction on six classifiers including SVM, GLM, NN, KNN, Random Forest, and Boosted Tree. The result of base-line model is written in parenthesise behind the number.

Model	N. features	Accuracy	F1	MC
GLM	(119)12	(0.99)**0.98**	(1.00)0.99	(0.01)0.02
SVM	(119)12	(0.98)0.97	(0.99)0.98	(0.02)0.03
NN	(119)12	(0.99)**0.98**	(0.99)0.99	(0.01)0.02
KNN	(119)12	(0.98)0.96	(0.99)0.98	(0.02)0.04
Random Forest	(119)12	(0.99)**0.99**	(1.00)**0.99**	(0.01)**0.01**
Boosted Tree	(119)12	(0.98)0.96	(0.99)0.98	(0.02)0.04

Features vs. Sensors. As we have shown, feature reduction is a viable way to help save the resources of wearables used in HAR. However, there is a key distinction between the features and sensors that are used to derive these features. Although reducing the number of features used will help save computation resources, etc., a real gain can be obtained if we are able to reduce the number of sensors that are on in these Wearables. Therefore, we ran an experiment to determine which sensors provided the most contributing features. The experiment was performed on data from both subjects, and included all steps in our dataset.

Figure 3 shows the share of each sensor vs. the total number of features for each of the three sensors on the Neblina, namely accelerometer, gyroscope and magnetometer. We observe from the figure that the accelerometer contributes the highest percentage of features, generally making up close to 50% of the features at any given point. On the other hand, the gyroscope and magnetometer, have similar contributions, which does not exceed 40%.

These results indicate that for HAR, we have the potential to not only reduce features, but perhaps do some sensor optimizations to maximize savings of wearable devices. Such optimizations are beyond the scope of this paper, however, we plan to develop such methods and examine the effectiveness in the future.

VI. Conclusion

In this paper, we have investigated the impact of feature reduction on the performance of HAR. We collected step data using the Neblina system-on-module solution from two subjects and have answered three research questions related to the impact of feature reduction in terms of performance, generalizability and varying classifiers. Our findings indicate that feature reduction can have a significant reduction in using resources while achieving comparable results to a full model. Our main findings are:

- Feature reduction can reduce the number of features by close to 90%, while only having an impact of 1-2% in model performance.
- Feature reduction can impact the performance of the general models (i.e., that are cross-subject), however, which subject a model is trained on does matter.
- Feature reduction does not have a major impact on most classifiers examined.

Our analysis also have showed that the accelerometer contributes most of the features used in HAR models. In the future, we will be introducing methods that can optimize sensor operation in order to maximize the resource savings of Wearables.

References

[1] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research," *ASU feature selection repository*, pp. 1–28, 2010.

[2] S. Sprager and M. B. Juric, "Inertial sensor-based gait recognition: A review," *Sensors*, vol. 15, no. 9, pp. 22 089–22 127, 2015.

[3] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[4] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17 472–17 500, 2013. [Online]. Available: http://www.mdpi.com/1424-8220/13/12/17472

[5] M. Janidarmian, A. Roshan Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors*, vol. 17, no. 3, p. 529, 2017.

[6] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10 146–10 176, 2014. [Online]. Available: http://www.mdpi.com/1424-8220/14/6/10146

[7] T. T. Nguyen, D. Fernandez, Q. T. Nguyen, and E. Bagheri, "Location-aware human activity recognition," in *International Conference on Advanced Data Mining and Applications*. Springer, 2017, pp. 821–835.

[8] C. M. Bishop, "Pattern recognition and machine learning (information science and statistics) springer-verlag new york," *Inc. Secaucus, NJ, USA*, 2006.

[9] S. Mujahid, G. Sierra, R. Abdalkareem, E. Shihab, and W. Shang, "Examining user complaints of wearable apps: a case study on android wear," in *Proceedings of the 4th International Conference on Mobile Software Engineering and Systems*. IEEE Press, 2017, pp. 96–99.

[10] C. Y. Yong, R. Sudirman, N. H. Mahmood, and K. M. Chew, "Human hand movement analysis using principle component analysis classifier," in *Applied Mechanics and Materials*, vol. 284. Trans Tech Publ, 2013, pp. 3126–3130.

[11] Ç. B. Erdaş, I. Atasoy, K. Açıcı, and H. Oğul, "Integrating features for accelerometer-based activity recognition," *Procedia Computer Science*, vol. 98, pp. 522–527, 2016.

[12] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proceedings of the 6th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011, pp. 92–98.

[13] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*. IEEE, 2006, pp. 4–pp.

[14] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 747–752.

[15] H. Ghasemzadeh, V. Loseu, E. Guenterberg, and R. Jafari, "Sport training using body sensor networks: A statistical approach to measure wrist rotation for golf swing," in *Proceedings of the Fourth International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 2.

[16] D. Ienco and R. Meo, "Exploration and reduction of the feature space by hierarchical clustering," in *Proceedings of the 2008 SIAM International Conference on Data Mining*. SIAM, 2008, pp. 577–587.

[17] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[18] "Promotion - motsai documentation," http://docs.motsai.com/Neblina/Neblina_Development_Kit/ProMotion/index, (Accessed on 11/13/2017).

[19] K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications," *Computer Animation and Virtual Worlds*, vol. 15, no. 2, pp. 79–94, 2004. [Online]. Available: http://dx.doi.org/10.1002/cav.2

[20] S. Jayalath, N. Abhayasinghe, and I. Murray, "A gyroscope based accurate pedometer algorithm," in *International Conference on Indoor Positioning and Indoor Navigation*, vol. 28, 2013, p. 31st.

[21] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[22] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview, ii," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 6, pp. e1219–n/a, 2017, e1219. [Online]. Available: http://dx.doi.org/10.1002/widm.1219

[23] C. H. Park, "A feature selection method using hierarchical clustering," in *Mining Intelligence and Knowledge Exploration*. Springer, 2013, pp. 1–6.

[24] J. A. Davis, "A partial coefficient for goodman and kruskal's gamma," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 189–193, 1967.

[25] N. C. Oza and K. Tumer, "Classifier ensembles: Select real-world applications," *Information Fusion*, vol. 9, no. 1, pp. 4–20, 2008.

[26] K. Deng, "Omega: On-line memory-based general purpose system classifier," Ph.D. dissertation, Carnegie Mellon University, 1998.

[27] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "Recofit: using a wearable sensor to find, recognize, and count repetitive exercises," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 3225–3234.

[28] M. Nabian, "A comparative study on machine learning classification models for activity recognition," *Journal of Information Technology & Software Engineering*, vol. 7, no. 04, pp. 4–8, 2017.

[29] T. J. O'Shea, J. Corgan, and T. C. Clancy, *Convolutional Radio Modulation Recognition Networks*. Cham: Springer International Publishing, 2016, pp. 213–226. [Online]. Available: https://doi.org/10.1007/978-3-319-44188-7_16

[30] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones." in *ESANN*, 2013.