






Article

# Optimization of Physical Activity Recognition for Real-Time Wearable Systems: Effect of Window Length, Sampling Frequency and Number of Features

Ardo Allik <sup>1,\*</sup> , Kristjan Pilt <sup>1</sup> , Deniss Karai <sup>1</sup> , Ivo Fridolin <sup>1</sup> , Mairo Leier <sup>2</sup>  
and Gert Jervan <sup>2</sup> 

<sup>1</sup> Department of Health Technologies, Tallinn University of Technology, Ehitajate tee 5, 12616 Tallinn, Estonia; kristjan.pilt@taltech.ee (K.P.); deniss.karai@taltech.ee (D.K.); ivo.fridolin@taltech.ee (I.F.)

<sup>2</sup> Department of Computer Systems, Tallinn University of Technology, Ehitajate tee 5, 12616 Tallinn, Estonia; mairo.leier@taltech.ee (M.L.); gert.jervan@taltech.ee (G.J.)

\* Correspondence: ardo.allik@taltech.ee

Received: 20 October 2019; Accepted: 7 November 2019; Published: 12 November 2019



**Abstract:** The aim of this study was to develop an optimized physical activity classifier for real-time wearable systems with the focus on reducing the requirements on device power consumption and memory buffer. Classification parameters evaluated in this study were the sampling frequency of the acceleration signal, window length of the classification fragment, and the number of classification features, found with different feature selection methods. For parameter evaluation, a decision tree classifier was created based on the acceleration signals recorded during tests, where 25 healthy test subjects performed various physical activities. Overall average F1-score achieved in this study was about 0.90. Similar F1-scores were achieved with the evaluated window lengths of 5 s ( $0.92 \pm 0.02$ ) and 3 s ( $0.91 \pm 0.02$ ), while classification performance with 1 s were lower ( $0.87 \pm 0.02$ ). Tested sampling frequencies of 50 Hz, 25 Hz, and 13 Hz had similar results with most classified activity types, with an exception of outdoor cycling, where differences were significant. Using forward sequential feature selection enabled the decreasing of the number of features from initial 110 features to about 12 features without lowering the classification performance. The results of this study have been used for developing more efficient real-time physical activity classifiers.

**Keywords:** accelerometer; activity classification; activity trackers; machine learning; wearable systems

## 1. Introduction

It is important to propagate active lifestyle, since routine physical activity has been found to have multiple benefits, such as preventing chronic diseases and increasing psychological well-being [1,2], while prolonged inactivity has been shown to lead to an increase of chronic diseases and obesity [1,3]. Advancement of technology has brought a surge of popularity for many activity trackers in the form of mobile phone apps or wearable systems. With these devices, users are able to keep track of their training schedule, exercises and lost calories [4]. Since this makes training more interactive and allows users to have better overview of their progress, then it often motivates the users to have a more active lifestyle and lose weight over sustained periods [5–7].

Wearable systems are used to conveniently measure, collect and analyze the user's psychological data. This requires wearables to be small and unobtrusive, which in turn puts significant demand on reducing power consumption of the system [8]. This is also significant for real-time physical activity classification, which can be used in wearables for online activity recognition by allowing automatic recognition of the activities the user is performing [9,10]. Real-time activity recognition provides

valuable information for improving online feedback of the activity trackers or for providing extra safety by monitoring the status of the users working in high-risk environments [11].

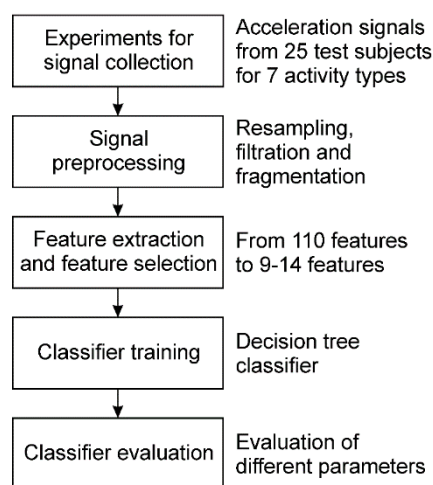
Power consumption required for physical activity classification is determined by multiple different components. Some of these components are based on the processing of the acceleration values, such as sampling rate of the signal and filtering [12]. Other elements are based on classification mechanics, such as classification window length, feature calculation, and the used machine learning algorithm. While studies have explored classification mechanics such as training times of different physical activity classification algorithms [13,14], they do not provide valuable information for real-time classification, since classifier training can be done previously on a desktop computer and later implemented into the wearable system. For classification systems working in real time, it is important to focus on processing time of the calculations the system has to do online [13,15].

In an earlier study, our group explored how different accelerometer sampling frequencies, classification window lengths, and the number of correlating features affect the classifier performance [16]. Few studies before have evaluated how different window lengths (commonly chosen between 1.5 s [17] and 5 s [13]) affect physical activity classification performance [15,18], but the lack of gold standard in physical activity classification makes it difficult to compare these results [19]. It has been stated that frequencies above 20 Hz cannot be expected to arise from voluntary movement [20], but comparable performance has been reported while using lower sampling frequencies [12,21]. Various methods have been used for feature selection, such as the ReliefF algorithm [22], principal component analysis [13], or information gain [15], but not in connection with window length and sampling frequency.

The aim of this study was to create an optimized physical activity classifier that would be suitable for implementation on real-time wearable systems. The focus was on testing various sampling frequencies, window lengths and number of features in order to reduce the power consumption, and to decrease the required memory buffer without compromising classification performance. Other classification elements were chosen based on the results of other studies with emphasis on high classification performance and low power consumption.

## 2. Materials and Methods

Physical activity classification often uses machine learning methods, where the classification is usually based on acceleration signals. Overview of the steps taken to create and evaluate the classifier used in this study are shown in Figure 1.



**Figure 1.** A summary of methods used in the study.

### 2.1. Instrumentation

Acceleration signals were measured with Shimmer3 (from here on Shimmer) sensor platform (Shimmer Research, Dublin, Ireland). While sensor fusion between accelerometers and gyroscopes has shown to increase classification performance in some studies [23], then others have found that gyroscope information does not contribute to activity recognition performance [22]. Due to the emphasis on designing physical activity classifier with low power consumption, gyroscope data were disregarded in this study.

The Shimmer sensor system has two built-in triaxial accelerometers: low noise accelerometer with the dynamic range of  $\pm 2$  g and a wide range accelerometer with the dynamic range switchable between  $\pm 2$  g to  $\pm 16$  g (where 1 g equals to about  $9.81 \text{ m/s}^2$ ). Since acceleration values during human motion surpass  $\pm 2$  g [24], the data from wide range accelerometer was used with the dynamic range set to  $\pm 16$  g. The wide range accelerometer uses STMicroelectronics LSM303AHTR sensor (Geneva, Switzerland), which has a numeric resolution of 16-bit. Acceleration was measured with a sampling rate of 512 Hz.

### 2.2. Study Group

The study was approved by the Tallinn Medical Research Ethics Committee. The main study group consisted of 25 healthy 21–45 year old test subjects (with an exception of one 57-year-old male), of whom 13 were male and 12 female. Average age was  $32.0 \pm 8.8$  years (median 30.0) for the whole group,  $32.8 \pm 10.0$  years (median 30.0) for males, and  $31.0 \pm 7.7$  years (median 30.0) for females. A separate study group was used to measure the signals of outdoor cycling. This group consisted of 5 males with an average age of  $38.4 \pm 5.3$  years (median 37.0).

### 2.3. Test Overview and Recorded Signals

Test subjects performed various physical activities during which acceleration signals were measured and recorded using the Shimmer sensor system. The sensor was located on the left wrist for feasibility of implementing the results in an activity tracker worn on the wrist. Even though using multiple sensors has been shown to increase the classification performance [25,26], having a wearable system with only one sensor is more comfortable and convenient for the user.

Each test subject conducted activities based on a precise schedule, where each activity was carried out for a fixed amount of time, shown in Table 1. For classification, these activities were grouped into different activity types, shown in Table 2. Indoor activities were divided into three different parts, during which each activity was performed for 3 min, with the exception of lying down, which lasted 4 min. There were short pauses between each activity, which were later discarded from the signals.

**Table 1.** Conducted activities and their duration in minutes.

Indoor Test 1	Indoor Test 2	Indoor Test 3 (% Shows Angle)	Outdoor Test
Walking (3)	Sitting on chair (3)	Walking (3 km/h) (3)	Cycling (14)
Running (3)	Lying on bed (4)	Walking (5 km/h) (3)	Cycling uphill (4)
Walking upstairs (3)	Typing on computer (3)	Walking (3 km/h, 10%) (3)	Cycling downhill (1)
Walking downstairs (3)	Folding clothes (3)	Walking (5 km/h, 10%) (3)	
	Cleaning surface (3)	Running (6 km/h) (3)	
		Running (10 km/h) (3)	
		Running (12 km/h) (3)	
		Running (6 km/h, 10%) (3)	

**Table 2.** Classified activity types.

Activity Type	Activities Concluded
Static	Lying, sitting, standing
Low Intensity	Typing on computer
Moderate Intensity	Folding clothes
Rhythmical Intensity	Cleaning a surface with a towel
Walking	Walking in a corridor, walking on a treadmill, walking upstairs, walking downstairs
Running	Running in a corridor, running on a treadmill
Outdoor Cycling	Cycling outdoors on different terrains

In the first part, test subjects walked in a corridor, ran in the corridor, walked upstairs, and walked downstairs. Altogether, a total of 12 min of acceleration signals were used from this part.

The second part consisted of sitting on a chair, lying on a bed, typing on a computer while sitting, standing, folding clothes while standing, and cleaning a surface while standing. A total of 19 min of signals were used from the second part.

The third indoor part consisted of walking on a treadmill at different speeds and angles (3 km/h, 5 km/h, 3 km/h with uphill angle 10%, 5 km/h with uphill angle 10%) and running on treadmill at different speeds and angles (6 km/h, 10 km/h, 12 km/h, 6 km/h with 10% uphill angle). A total of 24 min of signals were used from this part.

Outdoor cycling signals were recorded separately with a different study group. These signals consist of 14 min of cycling on a plain road, 4 min of cycling uphill, and 1 min of cycling downhill.

#### 2.4. Resampling and Sampling Frequency

As an aim of this study, it was tested how different sampling frequencies affect the classification results. Lowering the sampling frequency,  $f_s$ , decreases the number of samples in the classification fragment,  $s_f$ , which is calculated as follows:

$$s_f = f_s \cdot w_f, \quad (1)$$

where  $w_f$  is the window length of a fragment given in seconds.

To test different sampling frequencies, the signals that were initially recorded with a sampling frequency of 512 Hz were later resampled using a MATLAB function `resample` (R2016b, MathWorks, Natick, MA, USA). This function applies interpolation and decimation in order to achieve the desired sampling rate. In case of interpolation, the function inserts points with 0-values between each of the original samples of the signal, after which the signal is low-pass filtered at half of the desired sampling rate. To obtain the final result, decimation is applied by selecting samples from the filtered output [27]. The sampling frequencies of 50 Hz, 25 Hz, and 13 Hz were chosen for evaluating the effects of different sampling frequencies on classifier performance.

#### 2.5. Filtering

Following resampling, filtering was applied to separate the recorded acceleration signals into static and dynamic components for physical activity classification. The static component in the acceleration signal is mostly affected by gravity and captures the posture information, while the dynamic component is based on motion and captures the human movement information.

In this study, the static component was found using a third order low-pass Butterworth infinite impulse response (IIR) filter. The passband and stopband edge frequencies and ripples were 0.1 Hz and 0.5 Hz, and 1 dB and 20 dB, respectively. The dynamic component was found by subtracting the static component from the original signal by taking into account the group delay of the low pass filter.

## 2.6. Fragmentation and Window Length

For classifier training, acceleration signals were fragmented into shorter consecutive fragments. Before fragmentation, the short pauses in the signals between different conducted activities were removed and only signals recorded during activities listed in Table 2 were kept. While some studies opt for an overlap between windows to increase the classification performance, in this study, no overlap was used to keep the computational power minimal.

In a system with a physical activity classifier working in real time, the window length determines the delay of the system, since each classification is done after signals have been collected for a whole window. The number of samples in the fragment is determined by both the sampling frequency and the window length according to Equation (1).

To evaluate how different window lengths affect the classifier performance, the window lengths of 5 s, 3 s, and 1 s were chosen, which are near the values usually used for physical activity classification in previous studies [13,17].

## 2.7. Feature Extraction

When using machine learning methods for physical activity classification, the classifier training is done based on features that are extracted from signal fragments. The feature set has to capture specific and diverse information of posture and human motion to allow precise activity classification. The initial set of 110 features used in this study were mostly adopted from previous studies by other researchers: (1) 60 various time-domain features from [28]; (2) 10 body posture related, 6 motion shape related features and 6 motion periodicity related features from [15]; (3) 24 various time-domain features from [22]; and (4) 9 separately added additional features.

Only time-domain features were chosen in this study in order to keep computing power minimal. While activity recognition studies have also used frequency-domain and wavelet transform features, the transforms needed to calculate these features would require extra resources. Additionally, it has been found that time-domain features give comparable results to other feature types [29].

## 2.8. Feature Selection

Another major aim of this study was to analyze how different number of features affects physical activity classification and what is the minimal number of features to use without compromising classification performance. For that, two different feature selection schemes were used to optimize the feature set.

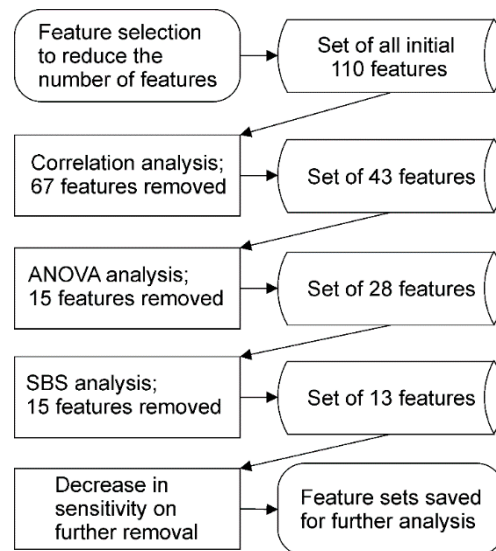
One scheme was based on various methods that were used successively (Figure 2). This scheme used the features extracted with sampling frequency of 50 Hz and window length of 3 s and the achieved optimized feature set was later used with other frequency and window length combinations.

First, correlating features were removed based on a large correlation matrix that showed each feature's correlation coefficient with other features. From feature pairs or groups with a very high correlation (correlation coefficient larger than 0.9 or lower than  $-0.9$ ), only the simpler features in terms of computational power requirements and complexity were kept. By using this method, 67 features were removed from the initial set, and a new subset of 43 features was formed. This method and the results have also been described in the previous study done by the authors [28].

Further feature optimization was done with one-way analysis of variance (ANOVA). The purpose of one-way ANOVA is to determine whether data from several groups of a factor have a common mean. ANOVA was used in this work to find out which features did not differentiate between any of the activities and thus did not provide any useful information for activity classification. Based on ANOVA results, 15 features were removed that were found not to affect classifier performance, and a new subset of 28 features was formed.

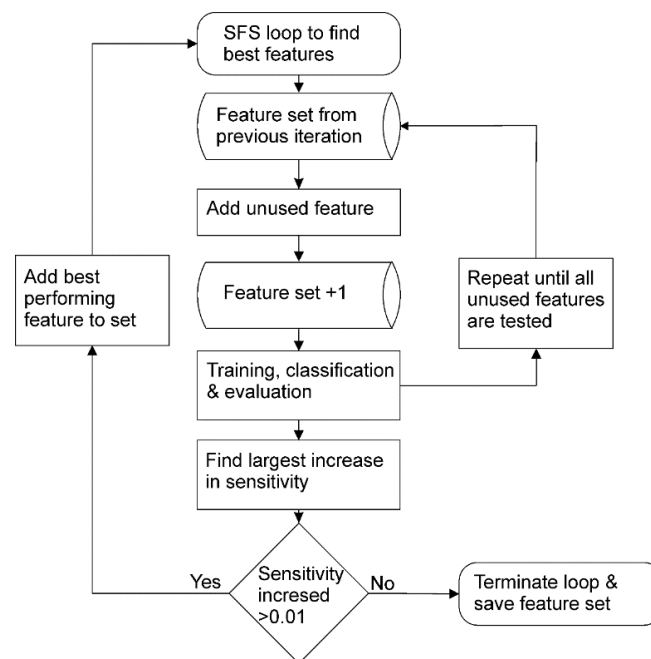
Finally, a sequential backward selection (SBS) procedure was repeated, where each feature was again removed one-by-one (those calculated similarly over all axes were removed together), and the

feature that decreased the classifier performance the least was removed. After removing features this way, the classifier performance was still persistent with 13 features used. Further removal of features showed a decrease in activity classification sensitivities.



**Figure 2.** First feature selection scheme using correlation analysis, ANOVA, and backwards sequential feature selection with the number of features removed in each step.

The second feature selection scheme used in this study was a sequential forward selection (SFS) method similar to the last steps used in the first scheme (Figure 3). In this method, features were added one-by-one by conducting physical activity classification with each feature and, for every iteration, the best feature was kept. Features were added until the overall average classification sensitivity did not improve by more than 0.001. This method was completed for every sampling frequency and window length combination, and was used to compare the results of the first method.



**Figure 3.** Forward sequential feature selection (SFS) method used in the second feature selection scheme.

### 2.9. Classifier Training

A machine learning based decision tree classification algorithm was chosen, which has been previously used in real-time physical activity classification and proposed as the most suitable in terms of performance and computational power needed for real-time classification [15,30]. The classifier was trained based on training data using MATLAB's function `fitctree`, which returns a fitted binary classification decision tree based on the input variables.

### 2.10. Classifier Evaluation

The classifier performance was evaluated using a leave-one-out cross-validation scheme where each test subject's signals were classified with a classifier that was trained using the signals from all the other test subjects. This method has been previously used in other physical activity classification studies to reduce overfitting errors [29,31].

Sensitivity (also called recall or true positive rate) was chosen as a statistical measure to evaluate classification performance during feature selection. Sensitivity shows the proportion of true positives classified (True\_positives) in relation to correct or real ones (Real\_positives), i.e., true positives that are correctly identified [32], and it is calculated as follows:

$$\text{Sensitivity} = \frac{\text{True\_positives}}{\text{Real\_positives}} = \frac{\text{True\_positives}}{\text{True\_positives} + \text{False\_Negatives}} \quad (2)$$

Classification results were evaluated using F1-score (also called F-score or F-measure), which is calculated as harmonic mean of precision and sensitivity [27], using the following formulas:

$$\text{Precision} = \frac{\text{True\_positives}}{\text{Predicted\_positives}} = \frac{\text{True\_positives}}{\text{True\_positives} + \text{False\_positives}}, \quad (3)$$

$$\text{F1-score} = \frac{2 \cdot \text{Sensitivity} \cdot \text{Precision}}{\text{Sensitivity} + \text{Precision}}. \quad (4)$$

While evaluating the results with different window lengths, sampling frequencies and number of features, F1-scores were calculated separately for each activity type. Additionally, an average F1-score for different parameter combinations was found as a means of the activity type F1-scores.

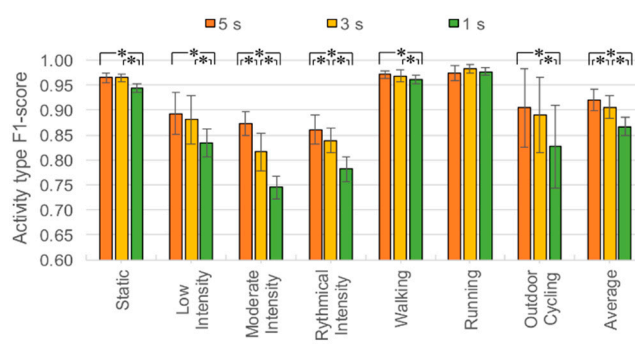
A paired *t*-test ( $p < 0.05$ ) was used to find statistical differences between the classification F1-scores of different activity types and averages while using different window lengths and sampling frequencies.

## 3. Results

### 3.1. Classifier Performance with Different Window Lengths

An overall average classification F1-score of about 0.90 was achieved for the physical activity classifier in this study, depending on the used window length, sampling frequency, feature set, and classified activity type. To evaluate how each of these parameters affected the classifier individually, classifier F1-scores were averaged over other parameters.

Figure 4 shows the classification F1-score of activity types for the different window lengths when averaged over different sampling frequencies (50 Hz, 25 Hz, 13 Hz) and feature sets (110 features, 43 features, 28 features, 13 features, and SFS feature set). The classifier had better performance with the average F1-score over 0.9 classifying static, walking and running activity types. Window lengths of 5 s and 3 s had similar results with the average F1-scores of  $0.92 \pm 0.02$  and  $0.91 \pm 0.02$ , while the result with 1 s was  $0.87 \pm 0.02$ .

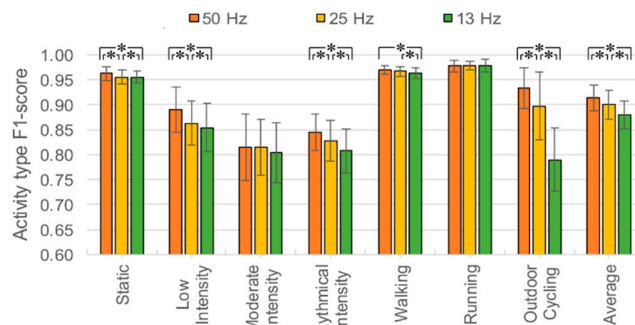


**Figure 4.** F1-scores of different activity types (mean ± SD (Standard deviation)) averaged over sampling frequencies and feature sets using different window lengths. Asterisks show significant statistical difference between different values of the window length ( $p < 0.05$ ).

Statistically significant differences (marked with an asterisk in Figure 4) were found in moderate intensity and rhythmical intensity activity types between window lengths of 5 s and 3 s. Window length of 1 s had a statistical difference classifying every activity type other than running compared to both 5 s and 3 s window length.

### 3.2. Classifier Performance with Different Sampling Frequencies

To compare the results with different sampling frequencies, F1-scores were averaged over different window lengths and feature sets (Figure 5). Overall, the classifier had similar average F1-score with 50 Hz ( $0.92 \pm 0.02$ ) and 25 Hz ( $0.91 \pm 0.02$ ), while the average F1-score with 13 Hz was lower ( $0.87 \pm 0.02$ ).



**Figure 5.** F1-score of different activity types (mean ± SD) averaged over window lengths and feature sets using different sampling frequencies. Asterisks show a significant statistical difference between different values of the sampling frequency ( $p < 0.05$ ).

Statistically significant differences between different sampling frequencies (marked with an asterisk in Figure 5) were found for most activity types with the exceptions of moderate intensity and running.

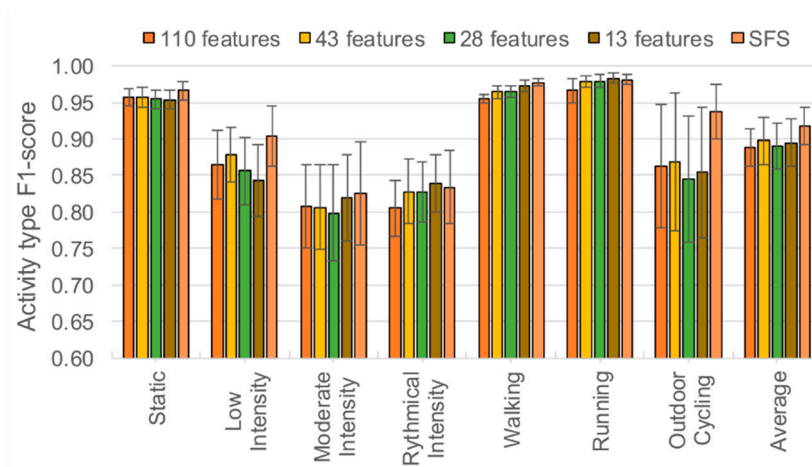
Very large differences in classification performance were noted while classifying outdoor cycling, where the F1-score was  $0.93 \pm 0.04$  with 50 Hz,  $0.90 \pm 0.07$  with 25 Hz and  $0.79 \pm 0.06$  with 13 Hz.

### 3.3. Classifier Performance with Different Feature Sets

To evaluate how the feature selection methods and the number of features used for classification affect the classifier performance, the results were averaged over different sampling frequencies and window lengths while using different feature sets (Figure 6). The feature sets of 110 features, 43 features, 28 features and 13 features, achieved with the first feature selection scheme, had similar average F1-scores between 0.89 and 0.90. The SFS feature set had a slightly higher average F1-score of  $0.92 \pm 0.03$ .

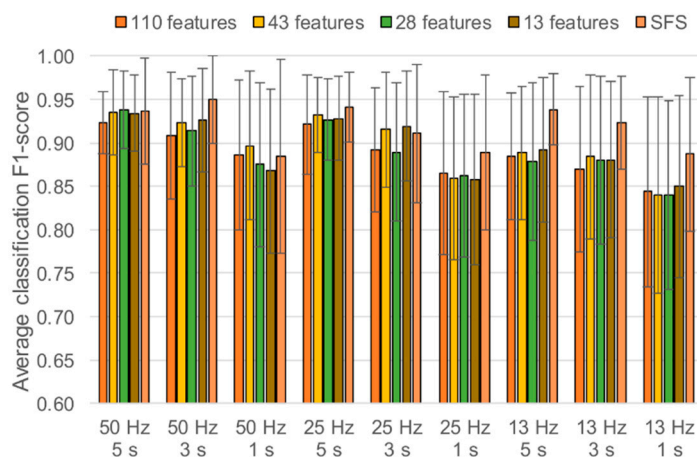


The SFS feature set had a major increase in performance compared to other feature sets classifying outdoor cycling ( $0.94 \pm 0.04$  compared to an average of  $0.86 \pm 0.09$  with other sets) and a slight increase in classifying low intensity activity type ( $0.90 \pm 0.04$  compared to an average of  $0.86 \pm 0.04$ ).



**Figure 6.** F1-scores of different activity types (mean  $\pm$  SD) averaged over window lengths and sampling frequencies using different feature sets.

Since both classification window length and sampling frequency of the acceleration signal affect the number of samples in classification fragments, it is important to evaluate their combined effect on classification performance. Figure 7 shows the average classification F1-scores with different feature sets using different combinations of sampling frequencies and window lengths. The SD values were large, since the results were averaged over different activity types with different F1-scores.



**Figure 7.** F1-scores (mean  $\pm$  SD) averaged over all activities using different feature sets, window lengths and sampling frequencies.

The average F1-scores of all the combinations of sampling frequencies and window lengths were similar to all of the feature sets of the first feature selection scheme. The classification performance was better with combinations that had more samples per classification fragment, with the highest average of  $0.93 \pm 0.05$  achieved with the combination of 50 Hz and 5 s. The results with the combinations that had either 1 s window length or sampling frequency of 13 Hz were lower compared to other combinations with most feature sets.

Compared to the feature sets of the first feature selection scheme, the SFS method used in the second scheme had higher performance with most window length and sampling frequency combinations. This difference was very noticeable with 13 Hz sampling frequency. The number of features used in

SFS feature sets was between 9 and 14 (Table 3), being remarkably lower than the number of features in most of the feature sets achieved with the first feature selection scheme.

**Table 3.** Number of features in sequential feature selection (SFS) feature sets with different sampling frequencies and window lengths.

Sampling Frequency and Window Length Combination	Number of Features in the SFS Feature Set
50 Hz, 5 s	12
50 Hz, 3 s	12
50 Hz, 1 s	9
25 Hz, 5 s	11
25 Hz, 3 s	12
25 Hz, 1 s	12
13 Hz, 5 s	11
13 Hz, 3 s	11
13 Hz, 1 s	14

### 3.4. Best Parameter Combination for Different Activity Types

While the results of this study generalized the effect of different sampling frequencies, window lengths, and number of features over various activity types, then it might also be useful to know the best combination for each activity type separately. Table 4 shows the parameter combination the highest F1 score for each classified activity type. The values are shown separately for both feature reduction schemes in order to compare the differences.

**Table 4.** Parameter combination with highest F1-score for different activity types and the average for both feature reduction schemes.

Activity Type	Window Length (s)	Sampling Frequency (Hz)	Number of Features	F1 Score
Static	5	25	110	0.97
	3	50	12 (SFS)	0.98
Low Intensity	5	13	110	0.93
	3	50	12 (SFS)	0.97
Moderate Intensity	5	50	110	0.90
	5	13	11 (SFS)	0.91
Rhythmical intensity	5	50	13	0.90
	5	25	12 (SFS)	0.89
Walking	3	50	43	0.98
	3	50	12 (SFS)	0.98
Running	3	25	13	0.99
	3	50	12 (SFS)	0.99
Outdoor Cycling	5	50	43	0.97
	3	50	12 (SFS)	0.98
Average	5	50	28	0.94
	3	50	12 (SFS)	0.95

## 4. Discussion

In this study it was analyzed for the first time how different window length, sampling frequency, and feature set combinations affect the performance of physical recognition based on decision tree classifiers in order to optimize the classifier for real-time wearable systems. The results of this study have been implemented into a smart work-wear prototype [11]. The main findings were: (1) classification F1-scores with window lengths of 5 s and 3 s were similar, while results with 1 s were lower; (2) all sampling frequencies performed similarly for most activity types, with an exception of

outdoor cycling; (3) similar or better results were achieved with the feature sets with 9 to 14 features, achieved with either feature reduction scheme, compared to the initial full feature set of 110 features.

The window lengths of 5 s, 3 s and 1 s were used in this study to analyze how different window lengths affect the performance of physical activity classifier. F1-scores of walking, running and low intensity activity types were similar to all window lengths, while the differences with moderate intensity, rhythmical intensity, and outdoor cycling were larger. Even though window lengths between 3 s and 1 s have been found to be suitable for other studies (2.56 s in [22], 2 s in [26], 1.5 s in [17], 1 s in [18]), in this study, the classifier performance had a larger drop when decreasing the classifier window down to 1 s, while window lengths of 5 s and 3 s had similar results. The window length of 1 s had statistically significant differences with both 3 s and 5 s window lengths while classifying static, moderate intensity rhythmical intensity and outdoor cycling activity types. This could be caused by 1 s window length not being long enough to capture the movement of the body during activities where one period of movement exceeds the window length.

Different sampling frequencies of 50 Hz, 25 Hz, and 13 Hz were used to investigate how sampling frequency affects classification performance. For most classified activity types, no statistical differences were found between tested sampling frequencies, but there were large differences while classifying outdoor cycling. Previously, it had been found that frequencies above 20 Hz cannot be expected to arise from voluntary human movement, where the accelerometer is not in contact with vibrating external sources [20]. It is likely that the 13 Hz sampling frequency was not high enough to capture the vibration during outdoor cycling.

A total of 110 features were extracted from acceleration signals for physical activity classification. To reduce and optimize the number of features, two different feature selection schemes were used in this study. While the first scheme used different consecutive methods to reduce the number of features, the second scheme used forward SFS where features were added one-by-one. The first feature selection scheme enabled the reduction of the feature set from 110 features to 13 features without decreasing the classifier performance. It is possible that the feature set with 13 features was overfit for the conditions used in this study and would perform worse in other conditions.

Compared to the feature sets of the first feature selection scheme, the SFS method used in the second scheme had higher performance with most window length and sampling frequency combinations. This difference was very noticeable when using the sampling frequency of 13 Hz. The number of features used in SFS feature sets were between 9 and 14 (Table 3). The large differences in average F1-scores shown in Figure 7 between SFS feature set and other feature sets while using sampling rates of 25 Hz and 13 Hz were mostly affected by outdoor cycling. Unlike other feature sets, the SFS feature set had a high F1-score while classifying outdoor cycling with all sampling frequency and window length combinations. The highest average classification F1 score was achieved with a parameter combination with SFS feature set (3 s window length, 50 Hz sampling frequency, 12 features), which also had the best performance while classifying static, low intensity, walking and outdoor cycling activity types (Table 4).

It was predictable that the SFS method would provide better results, since the SFS method chose the best features to maximize the classification sensitivity separately for each window length and sampling frequency combination, while, with the first scheme, features were selected based on one sampling frequency and window length combination. The SFS method proved to be a simple comparison method for more comprehensive feature selection and showed that the effect of features depends on different classifier parameters, of which sampling frequency and window length were tested in this study.

Despite the recent advances in deep learning based activity recognition, which reduces the dependency on hand-crafted feature sets and thus could outperform more traditional machine learning methods, it is still far from being used in online mobile systems due to excessive computational power it requires [33]. Thus, the methods and results of this study provide useful information to other

researchers for designing and implementing state-of-the-art physical activity recognition for real-time wearable systems.

## 5. Conclusions

This study evaluates the effects of sampling frequency of the acceleration signal, window length of the classification fragment, and number of features on classifier performance. The methods were chosen in order to reduce the requirements on computational power and available memory and are suitable for implementing physical activity classification in real-time systems.

We acknowledge some limitations in our approach that could be improved on in the future studies. First, sampling frequency and window length values evaluated in this study were chosen as a representative of the values used in other studies (low value, mid-range value, high value), but the optimum value could be somewhere between or even out of the explored range. It would be possible to classify larger numbers of different activity types and the acceleration signals should be measured under normal daily living conditions, which would allow for better physical activity classification during everyday life. The results could be evaluated with other machine learning algorithms that are used for physical activity classification, such as support-vector machines, Bayesian networks, and k-nearest neighbor algorithms, in order to see if there are any differences in the effects of the explored parameters.

**Author Contributions:** Conceptualization, A.A., K.P., and I.F.; formal analysis, A.A. and K.P.; methodology, A.A., K.P., and I.F.; investigation, A.A., K.P., D.K., and M.L.; data curation, D.K. and M.L.; writing—original draft preparation, A.A.; writing—review and editing, K.P., I.F., and G.J.; visualization, A.A.; validation, M.L.; supervision, I.F. and G.J.

**Funding:** The research was funded partly by the Estonian Ministry of Education and Research under institutional research financing IUTs 19-1 and 19-2, and by Estonian Centre of Excellence in IT (EXCITE) funded by European Regional Development Fund.

**Acknowledgments:** The authors wish to thank Siiri Mägi and Karl Erlenheim for assistance during the experiments, medical doctors Mari Meren and Ave Nagelmann from the Department of Pulmonology, North Estonia Medical Centre Foundations, Tallinn, Estonia for providing the environment for the study, and also those subjects who so kindly participated in the experiments.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyzes, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Dunstan, D.W.; Howard, B.; Healy, G.N.; Owen, N. Too much sitting—A health hazard. *Diabetes Res. Clin. Pract.* **2012**, *97*, 368–376. [[CrossRef](#)] [[PubMed](#)]
2. World Health Organization (WHO). *Obesity: Preventing and Managing the Global Epidemic*; WHO Technical Report Series 894; World Health Organization (WHO): Geneva, Switzerland, 2000.
3. Warburton, D.E.R.; Nicol, C.W.; Bredin, S.S.D. Health benefits of physical activity: The evidence. *Can. Med. Assoc. J.* **2014**, *174*, 801–809. [[CrossRef](#)] [[PubMed](#)]
4. Evenson, K.R.; Goto, M.M.; Furberg, R.D. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int. J. Behav. Nutr. Phys. Act.* **2015**, *12*, 1–22. [[CrossRef](#)] [[PubMed](#)]
5. Coughlin, S.; Stewart, J. Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies. *J. Environ. Health Sci.* **2016**, *2*, 1–6. [[CrossRef](#)] [[PubMed](#)]
6. Maher, C.; Ryan, J.; Ambrosi, C.; Edney, S. Users' experiences of wearable activity trackers: A cross-sectional study. *BMC Public Health* **2017**, *17*, 880. [[CrossRef](#)] [[PubMed](#)]
7. Middelweerd, A.; Mollie, J.S.; Wal, C.N.; Brug, J.; Velde, S.J. Apps to promote physical activity among adults: A review and content analysis. *Int. J. Behav. Nutr. Phys. Act.* **2014**, *11*, 1–9. [[CrossRef](#)] [[PubMed](#)]
8. Senevirante, S.; Hu, Y.; Nguyen, T.; Lan, G.; Khalifa, S.; Thilakarathna, K.; Hassan, M.; Senevirante, A. A Survey of Wearable Devices and Challenges. *IEEE Commun. Surv. Tut.* **2017**, *4*, 2573–2620. [[CrossRef](#)]
9. Lee, K.; Kwan, M.P. Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results. *Comput. Environ. Urban Syst.* **2018**, *67*, 124–131. [[CrossRef](#)]

10. Wannenburg, J.; Malekian, R. Physical Activity Recognition from Smartphone Accelerometer Data for User Context Awareness Sensing. *IEEE Trans. Syst. Man. Cybern. Syst.* **2017**, *47*, 3142–3149. [[CrossRef](#)]
11. Leier, M.; Pilt, K.; Allik, A.; Karai, D.; Jervan, G.; Fridolin, I. Fall detection and activity recognition system for usage in smart work-wear. In Proceedings of the 16th Biennial Baltic Electronics Conference, Tallinn, Estonia, 8–10 October 2018.
12. Yan, Z.; Subbaraju, V.; Chakraborty, D.; Misra, A.; Aberer, K. Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach. In Proceedings of the 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 17–24.
13. Altun, K.; Barshan, B.; Tuncel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.* **2010**, *43*, 3605–3620. [[CrossRef](#)]
14. Feng, Z.; Mo, L.; Li, M. A Random Forest-based ensemble method for activity recognition. In Proceedings of the 37th Annual International Conference of the IEEE-EMBC, Milan, Italy, 25–29 August 2015; pp. 5074–5077.
15. Tapia, E.M. Using Machine Learning for Real-time Activity Recognition and Estimation of Energy Expenditure. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, June 2008.
16. Allik, A.; Pilt, K.; Karai, D.; Fridolin, I.; Leier, M.; Jervan, G. Activity classification for real-time wearable systems: Effect of window length, sampling frequency and number of features on classifier performance. In Proceedings of the IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia, 4–8 December 2016; pp. 460–464.
17. Aktaruzzaman, M.; Scarabottolo, N.; Sassi, R. Parametric estimation of sample entropy for physical activity recognition. In Proceedings of the 37th Annual International Conference of the IEEE-EMBC, Milan, Italy, 25–29 August 2015; pp. 470–473.
18. Bulling, A.; Blanke, U.; Schiele, B. A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. *ACM Comput. Surv.* **2014**, *46*, 33. [[CrossRef](#)]
19. Awais, M.; Mellone, S.; Chiari, L. Physical activity classification meets daily life: Review on existing methodologies and open challenges. In Proceedings of the 37th Annual International Conference of the IEEE-EMBC, Milan, Italy, 25–29 August 2015; pp. 5050–5053.
20. Bouten, C.V.C.; Koekkoek, K.; Verduin, M.; Kodde, R.; Janssen, J.D. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans. Biomed. Eng.* **1997**, *44*, 136–147. [[CrossRef](#)] [[PubMed](#)]
21. Lee, J.; Kim, J. Energy-Efficient Real-Time Human Activity Recognition on Smart Mobile Devices. *Mob. Inf. Syst.* **2016**, *2016*, 2316757. [[CrossRef](#)]
22. Moncada-Torres, A.; Leuenberger, K.; Gonzenbach, R.; Luft, A.; Gassert, R. Activity classification based on inertial and barometric pressure sensors at different anatomical locations. *Physiol. Meas.* **2014**, *35*, 1245–1263. [[CrossRef](#)] [[PubMed](#)]
23. Wang, Z.; Yang, Z.; Dong, T. A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital Signs in Real Time. *Sensors* **2017**, *17*, 341. [[CrossRef](#)] [[PubMed](#)]
24. Chuang, F.C.; Yang, Y.T.C.; Wang, J.S. Accelerometer-based Energy Expenditure Estimation Methods and Performance Comparison. In Proceedings of the 2nd International Conference on Advances in Computer Science and Engineering (CSE 2013), Los Angeles, CA, USA, 1–2 July 2013; pp. 99–103.
25. Chowdhury, A.K.; Tjondronegoro, D.; Chandran, V.; Trost, S.G. Physical Activity Recognition Using Posterior-Adapted Class-Based Fusion of Multiaccelerometer Data. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 678–685. [[CrossRef](#)] [[PubMed](#)]
26. Loh, D.; Lee, T.J.; Zihajehzadeh, S.; Hoskinson, R.; Park, E.J. Fitness activity classification by using multiclass support vector machines on head-worn sensors. In Proceedings of the 37th Annual International Conference of the IEEE-EMBC, Milan, Italy, 25–29 August 2015; pp. 502–505.
27. Rajamani, K.; Lai, Y.-S.; Furrow, C.W. An efficient algorithm for sample rate conversion from CD to DAT. *IEEE Signal Process. Lett.* **2000**, *7*, 288–290. [[CrossRef](#)]
28. Liu, S.; Gao, R.X.; Freedson, P.S. Computational methods for estimating energy expenditure in human physical activities. *Med. Sci. Sports Exerc.* **2012**, *44*, 2138–2146. [[CrossRef](#)] [[PubMed](#)]
29. Preece, S.J.; Goulermas, J.Y.; Kenney, L.P.J.; Howard, D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 871–879. [[CrossRef](#)] [[PubMed](#)]

30. Altini, M.; Penders, J.; Amft, O. Energy Expenditure Estimation Using Wearable Sensors: A New Methodology for Activity-Specific Models. In Proceedings of the Wireless Health, San Diego, CA, USA, 23–25 October 2012.
31. Bao, L.; Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. *Pervasive Comput.* **2004**, *3001*, 1–17.
32. Powers, D.M.W. Evaluation: From precision, recall and F-factor to ROC, informedness, markedness and correlation. *J. Mach. Learn. Tech.* **2011**, *2*, 37–63.
33. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).