Biomedical Engineering Letters manuscript No.
(will be inserted by the editor)

A Wearable Real-time Activity Tracker

Ulf Jensen · Heike Leutheuser · Steffen Hofmann · Benno Schuepferling · Gerald Suttner · Kristin Seiler · Johannes Kornhuber · Bjoern M. Eskofier

Received: date / Accepted: date

Abstract *Purpose* Exercise and physical activity is a driving force for mental health. Major challenges in the treatment of psychological diseases are accurate activity profiles and the adherence to exercise intervention programs. We present the development and validation of CHRONACT, a wearable real-time activity tracker based on inertial sensor data to support mental health.

Methods CHRONACT comprised a Human Activity Recognition (HAR) algorithm that determined activity levels based on their Metabolic Equivalent of Task (MET) with sensors on ankle and wrist. Special emphasis was put on wearability, real-time data analysis and runtime to be able to use the system as augmented feedback device. For the development, data of 47 healthy subjects performing clinical intervention program activities were collected to train different classification models. The most suitable model according to the accuracy and processing power tradeoff was selected for an embedded implementation on CHRONACT.

Results A validation trial (six subjects, 6 h of data) showed the accuracy of the system with a classification rate of 85.6%. The main source of error was identified in acyclic activities that contained activity bouts of neighboring classes. The runtime of the system was more than 7 days and continuous result logging was available for 39 hours.

Conclusions In future applications, the CHRONACT system can be used to create accurate and unobtrusive patient activity profiles. Furthermore, the system is ready to assess the effects of individual augmented feedback for exercise adherence.

Keywords activity tracking · wearable computing · met levels · accelerometers · mental health

U. Jensen, H. Leutheuser, S. Hofmann, B. Schuepferling, and B. M. Eskofier
Digital Sports Group
Pattern Recognition Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Haberstr.2
91058 Erlangen
Germany
Tel.: +49-9131-27297
Fax: +49-9131-302811
E-mail: ulf.jensen@fau.de

G. Suttner, K. Seiler, and J. Kornhuber
Department of Psychiatry and Psychotherapy
University Hospital Erlangen
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany

All authors declare that they have no conflict of interest in relation to the work presented in this article.

1 Introduction

Physical activity has various positive physiological effects. It is regarded as one of the most important factors to reduce the risk for cardiovascular diseases, diabetes and cancer [1]. Beside these preventive physiological effects, physical activity has also been proposed as preventive and interventional treatment to promote mental health. Research showed that exercise programs effectively lower depressive scores and indicated that similar effects to pharmaceutical and psychotherapeutic treatment can be achieved [2].

The effective structuring of exercise interventions in mental health is subject of current research. Studies have investigated what amount, type and intensity of activity is most beneficial [2] and how traditional treatment (e.g. pharmacotherapy) and exercise can be combined [3, 4]. However, these aspects have to be investigated for different kinds of psychological diseases, different severities and patients with chronic medical diseases. An accurate activity profile of a patient is needed for scientific assessment of intervention programs.

Thereby, one challenge is the objective measurement of activity. Controlled intervention programs (e.g. 30 min cycling) [2] as well as an uncontrolled setting [5] where patients received physical activity recommendations were described. An uncontrolled individual setting is even more interesting to prevent the concurrent investigation of social interaction and physical activity affects [6]. One issue of current activity profiling in an uncontrolled setting is the self-report bias [5]. Questionnaires or self-reports are filled out by the patient and therefore depend on their subjective assessment. An alternative measurement modality, accelerometry, is capable of providing an objective activity profile. However, sophisticated signal processing is needed to increase accuracy and customize for intervention activities [7].

A common issue of exercise prescription in clinical practice is exercise adherence [6, 8, 9]. It can be distinguished between adherence to the intervention program and adherence to continuing activity after the intervention. During intervention, depressive symptoms (e.g. fatigue, psychomotor retardation) are described to make it difficult to start exercising and patients need corresponding encouragement [3]. Receiving encouragement and performing prescribed exercises, patient can benefit from psychological mechanism (e.g. mastering, self-efficacy) as short-term effects [3, 10]. The encouraging effect of augmented feedback might be a valuable approach to increase exercise adherence and has not been investigated in mental health yet.

Human Activity Recognition (HAR) with inertial, body-worn sensors is capable of detecting a wide range of daily life and athletic activities. Beside the common detection of individual activities, algorithms that identify groups of similar activities (e.g. [11, 12]) or groups of activities with the same intensity level (e.g. [13–15]) were proposed. Furthermore, activities in a supervised and unsupervised setting were investigated [16]. An important challenge in system design design is the accuracy and processing power tradeoff [17, 18]. This tradeoff gets even more important if data analysis is executed in real-time to provide augmented feedback to the user. Computing relevant, understandable and accessible augmented feedback in real-time is a challenging task [19]. This is especially the case in the context of HAR systems for psychological patients that put specific demand on system characteristics like result precision, stigmatization, individualized feedback and runtime.

Fig. 1 Sensor positions on the left wrist and right ankle and the two components of the CHRONACT system.

This work describes the design and validation of a wearable system for activity tracking called CHRONACT (Fig. 1). The system collected acceleration data on the wrist and ankle and derived the activity level of the wearer. CHRONACT was designed for the use in mental health support and special emphasis was put on mobility and the possibility for augmented real-time feedback. Intended applications for CHRONACT are meaningful patient activity profiles and the investigation of augmented feedback on patient adherence.

2 Literature Review

There is a vast amount of studies on the effects of exercise on mental health. Studies vary in different aspects like type and severity of psychological disease, pharmaceutical treatment, intervention duration, population characteristics and follow-up strategy. We limited the literature overview to reviews and meta analyses of existing studies [2, 4, 6, 20–22]. In general, exercise can have beneficial but also detrimental effects on the mental state of patients and healthy individuals. Therefore, we further limited our review to results for the treatment of clinical cases of depression. In agreement, literature reports positive effects of exercise on the depression level of patients with mild or moderate stages of the disease that are comparable to other forms of treatment. Findings agree that exercise has a positive preventive effect and can effectively avoid a relapse of depression.

The programming of the exercise intervention is a current field of research but findings imply that all types of activity are beneficial in a specific way. Activity effectiveness is strongly coupled to the investigation of the biological processes that cause disease alleviation. One hypothesis is that activity activates specific brain areas and induces the release of neurotransmitters. With these processes, mood and exercise adherence is increased [6]. Further, it is hypothesized that regular exercise induces neurogenesis and angiogenesis which is beneficial for behavioral and cognitive function [9]. As biological phenomena are not completely understood, psychological mechanisms like mastering and distraction [3] as well as social interaction are hypothesized to cause positive effects on depression [6].

Brosse et al. underlined the need for objective physical activity monitors for investigating exercise interventions and enhancing exercise adherence [6]. An application example for such a system is a study in which the short-term effects of self-induced activity on the depression level were investigated [5]. Depression level was linked to the recently performed physical activity and exercise was found to be beneficial when performed in unsupervised and self-induced bouts. The authors underlined the need for objective activity and physiological measures and recommended that individuals with depression should be encouraged to engage in more physical activity in daily life.

Recently, activity monitors (RT3, Stayhealthy Inc., Monrovia, USA) were used in a mental health study to examine the relationship of psychiatric symptoms, cognitive functioning and physical activity [7]. The compliance of the patients to wear the accelerometer devices was sufficient but some research questions could not be completely addressed with the provided activity measure. The measure was described as summarized magnitude of the 3-D acceleration vector and did for example not detect the common activity of cycling. The authors concluded that additional effort is needed to motivate patients with mental health conditions to meet physical activity recommendations.

Physical activity and exercise monitoring for mental health support requires activity recognition. The emerging field of HAR has different applications in domains like healthcare, security, entertainment and sports [23,24]. It requires specific computational methods and faces a list of specific challenges [17]. Two aspects are important for augmented feedback in mental health support; the definition and diversity of physical activities and the accuracy and processing power tradeoff in system design. Specifically, the HAR system should be able to discriminate activities of interest with high accuracy while being able to run on a wearable system with limited computational resources and display results in real-time.

Augmented feedback and biofeedback applications proved their value in a wide range of healthcare applications. Balance and mobility training for elderly is one example in which wearable feedback systems are often implemented on low-power microcontrollers [25]. Another example is stroke rehabilitation to e.g. treat upper extremity spasticity, train hand function and the ability to perform activities of daily living [26]. Current research in the area of augmented feedback has recently been reviewed and summarized [19].

Embedded classification, classification algorithms on resource constraint hardware, is used for different data analysis tasks in the healthcare domain. One example is movement classification comprising posture determination, fall detection and energy expenditure estimation [27]. Another application is the analysis of the electrocardiogram (ECG) to detect heartbeats on an embedded device and reduce the communication effort to the current heart rate [28]. A general overview of signal processing techniques for embedded classification and more specifically on the accuracy and processing power tradeoff can be found in [29] and [30] respectively.

Often, HAR systems are used to classify specific activities based on accelerometer data recorded on different body locations [11]. However, the performed intensity level is often of greater interest than the specific activity [31]. A method to deduce the energy expenditure (EE) directly from the accelerometer data showed to be imprecise for sedentary activities and static exercises [32]. As an alternative, data can be classified in physical activity levels like low, medium and high [13]. Such a categorization can be realized with the Metabolic Equivalent of Task (MET, [33]). Compared to EE, MET has the advantage of being a relative measure and, therefore, being independent from size, weight and body composition of the individual. Ainsworth et al. created a "Compendium of Physical Activitiy" in which MET levels of over 800 activities were assigned [33]. They further described the three categories light-intensity ($<$ 2.9 MET), moderate-intensity (3.0 - 5.9 MET) and vigorous-intensity ($\geq$ 6.0 MET). Thus, they give a specification of how to group activities in intensity levels and can be used to label physical activity for HAR.

Grouping activities according to their intensity levels has been used in HAR systems to assess the physical activity distribution of an individual. Reiss and Stricker measured eight subjects performing 14 different activities [13]. They acquired data of one heart rate sensor and three 3-D accelerometer sensors, placed at the wrist, chest, and foot with a sampling rate of 100 Hz. The activities were grouped according to their MET levels into light, moderate and vigorous intensity. Their algorithm achieved a classification accuracy between 82.06% and 95.65% for different sensor combinations whereby all sensors achieved the best result.

Multiple commercial products for HAR are available. They can be categorized in medical devices (e.g. ActiGraph, RehaWatch), lifestyle products (e.g. FitBit Flex, Jawbone Up, Misfit Shine, Google Fit) and sports products (e.g. Polar Loop, Nike Fuel). These systems and devices have several drawbacks. First, systems calculate simple and generic measures (e.g. step count) that do not reflect all performed activities. Second, systems cannot calculate measures in real-time and therefore prohibit concurrent augmented feedback. Third, systems cannot be customized for individual feedback or do not comprise a display. Fourth, systems are easily identified as medical gear and patients might feel stigmatized.

The literature review revealed the need for objective HAR systems for mental health research and treatment. Augmented real-time feedback might thereby be a valuable method to motivate patients for more activity or adhere to exercise interventions. The real-time capability is a technically challenging task as computations have to be performed with the limited resources of a wearable system and results need to be available instantaneously. As augmented feedback is sensitive to different variables (e.g. timing, presentation, ...), full control is needed to be able to create and assess the effects of feedback.

This article presents and evaluates CHRONACT, a wearable real-time activity tracker for the use in mental health applications. CHRONACT is based on commercially available hardware that was previously used for biomedical applications and features a feedback display [34]. The HAR system classified activity levels rather than specific activities to address the definition and diversity of physical activities challenge. Our contributions to the field is an alternative real-time activity tracking algorithm with the underlying design methodology to implement the system on embedded hardware.

3 Methods

First, the two sensor system that we used are described. Second, the development workflow is introduced. Third, data collection is summarized. Fourth, the data analysis methodology to calculate MET levels from accelerometer data is described. The last part summarizes the methods to enhance the CHRONACT runtime.

3.1 Sensor Hardware

The sensors were positioned on the left wrist and the right foot on the medial side and proximal to the ankle to collect accelerometer data of the upper and the lower body part (Fig. 1) [11]. The wrist sensor was worn on the left side as this is the common position for a wrist-watch. Furthermore, feedback on the device is easily accessible with this position. We randomly selected the position on the right ankle as no side preference is described in literature. Two different hardware system were used for collecting training data (LOGGER) and the target real-time analysis system (CHRONACT).

3.1.1 LOGGER

Data was logged with sensor nodes from the Shimmer™ platform (Version 2R, Shimmer™, Dublin, Ireland) [35]. Each node was configured to collect 3-D acceleration with a range of ± 6 g. The sampling rate was 102.4 Hz and signals were calibrated with the manual provided by the manufacturer.

3.1.2 CHRONACT

The CHRONACT system comprised two eZ430-Chronos devices (Texas Instruments Inc., Dallas, USA). For inertial sensing, each device integrates a 3-D accelerometer supporting different ranges (± 2 g, ± 4 g, ± 8 g, ± 16 g) and a sampling rate up to 1000 Hz [36]. For user interaction, each device comprises a LCD display and four user buttons. The devices featured a CC430F6137 low-power microcontroller unit (MCU) with a proprietary CC1101 radio transceiver. Wireless communication is conducted using the SimpliciTI protocol in the 868 MHz band. The devices are powered by standard replaceable 3 V coin cell lithium batteries (CR2032) with a capacity of 240 mAh. The eZ430-Chronos devices were programmed with our custom firmware.

3.2 Development Workflow

The development workflow comprised four major steps (Fig. 2):

1. Training Data Collection
2. Data Analysis Algorithm Development
3. Mobile System Implementation
4. Mobile System Validation

We used different hardware for the collection of training data (LOGGER) and the wearable system (CHRONACT) to exploit the benefits of each system. The LOGGER system was capable of logging large amounts of data on the internal SD card. The internal rechargeable battery ran for hours in data collection mode. However, LOGGER did not feature a display and had a form factor that identified the sensor nodes as technical system.

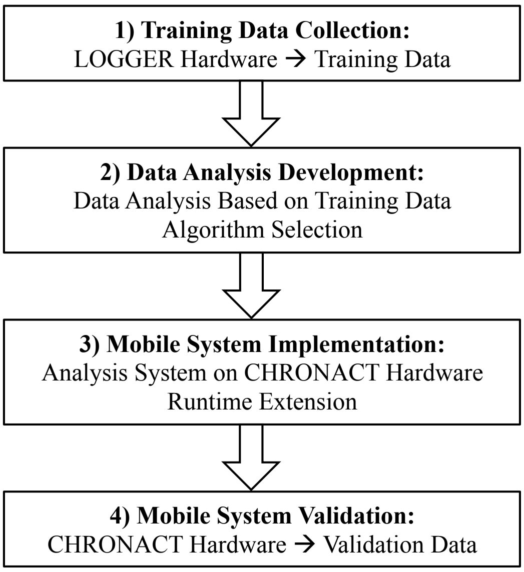

Fig. 2 Development workflow of the CHRONACT system.

The CHRONACT system was not capable of extensive logging due to an internal flash memory size of 8 KB and limited battery capacity. However, it had the benefits of being unobtrusive (small form factor), non-stigmatizing (looks like a sports watch) and including a display (augmented feedback ready). Consequently, training data was collected with the LOGGER system and the validation was performed with the final CHRONACT system.

LOGGER data were converted for the intended use with the CHRONACT system. Therefore, coordinate systems and data ranges were aligned. We also investigated the data characteristics of both system with a simultaneous recording trial (results not shown here) and registered only minor differences.

3.3 Data Collection

We collected data for training the data analysis system and for validating the final system.

3.3.1 Study Approval Statement

All subjects participating in the data collection gave written informed consent. Both studies were approved by the ethics committee of the medical faculty of the university under reference 106_13B.

3.3.2 Training Data

This dataset was collected with the LOGGER system and comprised two parts. First, we used the BaSA dataset that is freely available on the ActivityNet website [37, 38]. Second, we collected common activities that are performed as interventional treatment of depression patients to customize the activity tracker for this population. The exercises are either part of dedicated courses or are individually performed according to the physicians recommendation and patients preference. Both datasets were combined and used as training data set.

Data were assigned to three classes according to the MET level of the activity (LOW, MED, HIGH). The mapping was performed according to the activity categories described in [33]:

- LOW: < 2.9 MET
- MED: 3.0 - 5.9 MET
- HIGH: ≥ 6.0 MET

Furthermore, data in the LOW class was subdivided in static (LOW-STA) and active (LOW-ACT) (Tab. 1). This differentiation is a valuable information for the physicians to distinguish inactivity from light-intensity activities performed during occupational therapy.

Due to the high amount and range of activities, participants only completed parts of the activity list. However, each activity was performed by at least 5 subjects. Overall, 47 subjects (20 female) volunteered for collecting training data. They were recruited from age groups between 20 and 65, different fitness levels and varying height and weight. All subjects were healthy and performed each activity for around 1 to 2 min. Overall, the training dataset consisted of more than 12 hours of physical activity data and covered nearly all activities for patients during exercise intervention and in everyday life. Data of the classes LOW, MED and HIGH were balanced to avoid the preference of a specific class. Data of activities was not balanced as activity intensity instead of specific activities was classified.

3.3.3 Validation Data

This dataset was recorded with disjoint subjects, collected with the CHRONACT system and used to assess its performance. Therefore, six healthy subjects (3 female, age 25 to 52 years) were equipped with the system and performed a protocol of twelve different activities (Tab. 2). Each activity was conducted for 5 min. The classification results were computed in real-time, shown on the display (Fig. 3) and logged in flash memory. Overall, 6 h of validation data was recorded.

3.4 Data Analysis System

The data analysis system comprised a signal energy tracker and a pattern classification system. First, the

Table 1 Activities of the training data with classes LOW, MED and HIGH as well as static (LOW-STA) and active (LOW-ACT) activities in the LOW class.

Name	Description	Class
Lying	Lying on the couch	LOW-STA
Sitting	Sitting on a chair	
Standing	Standing still	
Gaming	Playing board games (sitting)	LOW-ACT
Painting	Painting canvas (standing)	
Mandala	Coloring mandalas (sitting)	
Sudoku	Solving sudokus (sitting)	
Strolling	Strolling around	
Wood	Woodworking (sitting)	
Badminton	Playing badminton	MED
Cycling	Riding a bike	
Descending	Descending stairs	
Dribbling	Ball dribbling and carrying (walking)	
Nordic (slow)	Slow nordic walking	
Paddling	Simulated paddling (standing)	
Stand aerobic	Aerobic (standing)	
Sit aerobic	Aerobic (sitting)	
Table tennis	Playing table tennis	
Table soccer	Playing tabletop soccer	
Throwing	Ball throwing and catching (standing)	
Walking	Walking with self-selected speed	
Ascending	Ascending stairs	HIGH
Basketball	Playing basketball	
Jogging	Jogging with self-selected speed	
Nordic (fast)	Fast nordic walking	
Passing	Ball passing while running	
Soccer	Playing soccer	
Sprinting	Different sprints in relays	
Volleyball	Playing volleyball	

Table 2 Activities of the validation data with classes LOW, MED and HIGH as well as static (LOW-STA) and active (LOW-ACT) activities in the LOW class.

Name	Description	Class
Lying	Lying on a mat	LOW-STA
Sitting	Sitting on a chair	
Standing	Standing still	
Mandala	Coloring mandalas (sitting)	LOW-ACT
Sudoku	Solving sudokus (sitting)	
Strolling	Strolling around	
Throwing	Ball throwing and catching (standing)	MED
Cycling	Riding a bike	
Nordic (slow)	Slow nordic walking	
Ball game	Basketball, Volleyball, Soccer (self-selected)	HIGH
Jogging	Jogging with self-selected speed	
Nordic (fast)	Fast nordic walking	

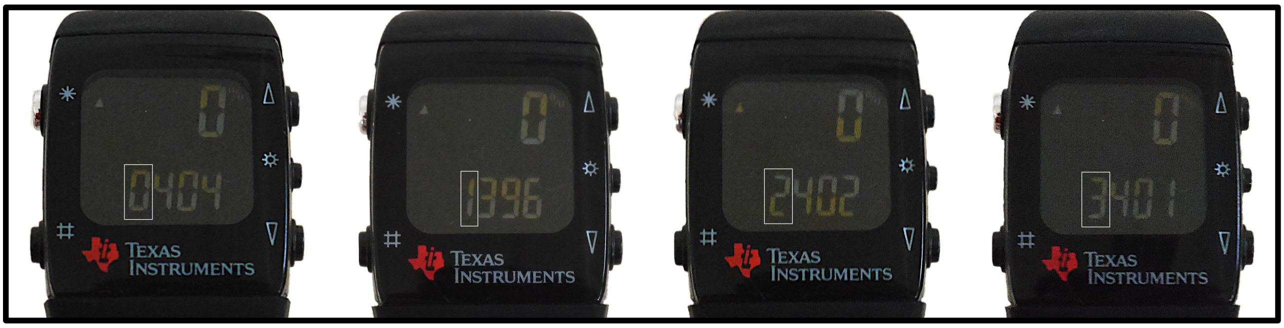

Fig. 3 CHRONACT systems displaying the classification results of **0** (LOW-STA), **1** (LOW-ACT), **2** (MED) and **3** (HIGH). Additionally, the devices show memory consumption (top right) and debug information for radio communication (bottom right).

signal energy tracking determined whether the person is moving. Second, the movement intensity was classified into the MET level classes LOW, MED and HIGH.

All data was processed in intervals of 5 s. This was defined as adequate granularity for the analysis of the physicians.

3.4.1 Signal Energy Tracker

This step determined whether the wearer was moving or not. Thus, it determined whether an interval belonged to the LOW-STA class or to one of the remaining classes. Therefore, we computed the signal difference

$$\widehat{f_{ij}}(t) = f_{ij}(t) - f_{ij}(t-1) \quad (1)$$

of each signal axis f_{ij} with $i \in I$, $j \in J$ where I denotes the two sensor position and J the three sensor axes. In case, one value of $\widehat{f_{ij}}$ was above the threshold, the interval was further processed, otherwise the LOW-STA label was assigned. The threshold for this distinction was set on the training dataset.

3.4.2 Pattern Classification

Each interval that was not classified as LOW-STA was further processed with a pattern classification system. It comprised a feature extraction and a classification step in the working phase and feature selection and classifier training in the training phase. Experiments were conducted with the Embedded Classification Software Toolbox (ECST[1]) [39]. The following paragraphs describe the processing steps in more detail.

Feature Extraction: We used a generic biosignal feature set comprising signal characteristics and statistical measures [40] (Tab. 3). Features were extracted in an update manner to balance the computational load. Thus, intermediate results were computed on every sample of the interval instead of computing the result when data of the complete interval is available. The computation of higher statistical measures is based on intermediate results for the statistical moments that are updated with every new sampling value [41, 42]. Compared to the standard computation, this approach has the advantage that less computational operations are needed at the end of an interval. The update extraction therefore fits well for the time-slice scheduling of the MCU. The eight features were computed for each axis of the two sensor positions so that overall 48 features were used for classification.

Classification: The "No Free Lunch" theorem states that a single classifier is not guaranteed to outperform all others in any classification problem [43]. Therefore, we compared several classifiers representing different classification paradigms. We limited the selection to four popular classifiers due to the high amount of training data and tested linear as well as non-linear classifiers.

- AdaBoost (AB): Non-linear ensemble of weak classifiers (decision stump) [44].
- Classification via Logistic Regression (CLR): Linear classifier using logistic class regression models [45].
- Naive Bayes (NB): Non-linear classifier based on the Bayes Rule [43].
- Nearest Neighbor (NN): Non-linear classifier using the euclidean distance [46].

Feature Selection: This step was used to reduce the dimensionality of the classification model. Furthermore, we were interested in a generalizable model as the classifier might be confronted with previously unseen activities. Feature selection is a suitable approach for these goals [43]. We used a linear forward selection with classification rate as score metric to determine a discriminative and reduced feature set for a specific classifier [43].

[1] http://www.tinyurl.com/ecstproject

Table 3 Set of eight features for each axis of the two sensor nodes.

Features	
Signal energy	Minimum
Maximum	Mean
Standard deviation	Variance
Skewness	Kurtosis

Classifier Training: All experiments for algorithm selection (classifier, sampling rate) were conducted on the training set and evaluated with a leave-subject-out cross-validation method (Fig. 2). In this approach, data of a single subject is used as test data while a model is trained on data of the remaining subjects. This procedure is repeated for all subjects. When feature selection was used, it was embedded in an inner cross-validation loop. Thus, an inner leave-subject-out cross-validation loop was performed to assess the performance of a candidate feature set.

Model Creation: The final classification model that was implemented on the CHRONACT hardware was created in a single training step. Therefore, the complete training data, the settings determined in the classifier selection step (classifier, sampling rate) and the selected features were used (Fig. 2). The accuracy of this model was evaluated with the disjoint validation data set.

3.5 Runtime Extensions

We used different methods to enhance the runtime of the final CHRONACT system. The goal was to provide a system that runs for an intervention period of at least five days reflecting the working week. However, power the saving opportunities described were exploited to full extend for maximum runtime. The device firmware was implemented with the IDE provided by the manufacturer (Code Composer Studio v5.5.0, Texas Instruments Inc., Dallas, USA).

3.5.1 Sampling Rate Reduction

A reduction of the sampling rate reduces the sampling and classification effort. However, due to the Nyquist theorem, the frequency content reduces with the sampling rate. We assessed the performance of reduced sampling rated by downsampling the training data and repeating the classifier training. We tested the sampling rates of 102.4 Hz, 51.2 Hz, 20.5 Hz and 10.2 Hz. All sampling rates could also be set in the CHRONACT system.

3.5.2 Arithmetic

The feature extraction and classification steps comprised floating point operations (FLOP) but the eZ430-Chronos hardware did not comprise a floating point unit (FPU). Compilers can convert FLOPs to fixed point arithmetic but FLOPs consume more machine cycles and therefore more power [47]. We therefore mapped the CHRONACT implementation to fixed point number representation for further processing [47, 48]. Raw data input was converted to integer numbers after sensing. The trained classification model was converted offline before integrating the required constants. We analyzed the range of variables and computations and decided to represent each variable as 32 bit integer with 16 bit each for the integral and fractional digits. The following computation were all conducted in fixed point arithmetic. A precision comparison revealed that this representation is sufficient for decision making as results slightly differed in the lowest decimal digits.

3.5.3 Radio Transmission

Sleeping mode management and radio transmission have a strong influence on the energy consumption of BSN nodes [30]. We therefore measured the power consumption in three modes (sleeping core, active core, active radio) to assess the highest potential for economization. Our multimeter measurement revealed values of less than 10 μA (sleeping core), 2.23 mA (active core) and 20.11 mA (active radio). Thus, we identified the radio transmission as main power consumer. Thus, we reduced radio transmission with a new operation mode that comprised feature level fusion and node synchronization.

First, data were fused on feature level. Features from each position were extracted on the corresponding device and fused on the wrist device. This device was then used to take the final classification decision and log the data. The data flow was unidirectional as the wrist device did not transmit payload to the ankle device.

Second, radio transmission was solely activated in a short window after the processing interval, thus every 5 s, to transmit the extracted features. Therefore, devices were required to synchronize clocks to activate radio transmission at the same time. We corrected the clock drift (e.g. crystal tolerances, temperature differences) with measuring the time differences of the active windows. The last five measurements were averaged and used to correct short-term clock drift after every processing interval with a granularity of 30 μs.

We measured the length of the radio active phase with feature level fusion and node synchronization. It

Table 4 Results of the algorithm selection experiments with the best performing classifier in bold face. Classifiers are abbreviated with AB (AdaBoost), CLR (Classification via Linear Regression), NB (Naive Bayes) and NN (Nearest Neighbor).

Sampling rate [Hz]	Classifier	Accuracy [%]
102.4	AB	83.2
	CLR	86.9
	NB	83.0
	NN	**88.9**
51.2	AB	80.0
	CLR	87.1
	NB	82.9
	NN	**88.4**
20.5	AB	84.1
	CLR	**87.1**
	NB	82.7
	NN	86.7
10.2	AB	80.8
	CLR	**86.6**
	NB	82.1
	NN	86.4

was 12.2 ms measured with varying sensor positions. This is beneficial for the power consumption but also for the robustness of the system as a minimum voltage of 2.0 V was required for correct functioning.

3.5.4 Data Logging

A bottleneck for the long-term monitoring by the system is the memory to save the activity profile. We analyzed the size of the source code and the amount of main memory used during execution. Our analysis revealed that 8 kB of flash memory and 2 kB of random access memory (RAM) were available for data logging. Our application used the free flash memory as well as parts of the RAM to log the activity profile in binary format.

4 Results

4.1 Algorithm Selection

For algorithm selection, classifiers and sampling rates were compared using the training data and the complete feature set (Tab. 4). The NN classifier performed best in the higher sampling rates (102.4 Hz, 51.2 Hz) and the CLR classifier performed best in the lower sampling rates (20.5 Hz, 10.2 Hz). For each sampling rate, an accuracy of above 86.5% was achieved by the best-performing classifier. When comparing sampling rates, the difference between the highest possible accuracy (102.4 Hz, NN, 88.9%) to the lowest accuracy (10.2 Hz, CLR, 86.6%) was 2.3%.

The latter classification system (10.2 Hz, CLR) was evaluated with feature selection and achieved an accuracy of 84.3% with 20 selected features. This corresponds to an accuracy difference of 5.6% between the highest possible accuracy (102.4 Hz, 48 features, NN, 88.9%) to the selected classification model (10.2 Hz, 20 features, CLR, 84.3%).

4.2 Mobile System Validation

The final CHRONACT system was configured to sample acceleration (range: ± 8 g) with 10.2 Hz and used a CLR model (20 features) to classify MET levels with a granularity of 5 s. It was evaluated for performance and system runtime.

4.2.1 Performance Evaluation

The results were obtained by comparing the classification results computed on the CHRONACT system with the reference of the validation study protocol.

First, we considered the three main classes LOW, MED and HIGH. The LOW class comprised data from LOW-STA and LOW-ACT. The CHRONACT system achieved a classification rate of 85.6%.

Second, we considered the four classes LOW-ACT, LOW-STA, MED and HIGH. The CHRONACT system achieved a classification rate of 69.1%.

Additionally, we computed the mean value and standard deviation of the classification result for each class in the four class setting (Tab. 5). The mean values showed a considerable difference between the four classes. Mean values were close to the actual label values for the LOW-STA, MED and HIGH class.

Table 5 Labeling, mean values (Mean) and standard deviation (Std) of the validation results.

Class	Label	Mean	Std
LOW-STA	0	0.01	0.08
LOW-ACT	1	0.55	0.80
MED	2	1.99	0.67
HIGH	3	2.91	0.29

4.2.2 System Runtime

We evaluated the runtime of the final CHRONACT system in a long-term trial. Being equipped with new batteries, the system ran for 7 days, after which we stopped the data analysis. During that time, the systems logged MET level data for the first 39 hours before the available memory ran out.

5 Discussion

5.1 Algorithm Selection

The experiments on the training data revealed that the activity level can be classified with moderate accuracy using accelerometer data from the wrist and the ankle. A comparison of different classifiers showed that NN and CLR performed equally well with CLR outperforming NN in the lower sampling rates. Furthermore, the performance negligibly decreased with decreasing sampling rate. Consequently, we selected a classification system comprising 10.2 Hz sampling rate and the CLR classifier according to the accuracy and processing power tradeoff of HAR systems. Besides the fact that CLR performed best on the training data, it is more suitable for an embedded implementation. The NN classifier requires to save the complete training data on the embedded device and furthermore requires to compute the distance to each training pattern during classification. Results from the ECST (not shown here) confirm this statement and revealed that NN has a considerably higher computational effort and memory demand if compared to CLR [39]. It is obvious that the lower sampling rates have beneficial effects on the runtime of the system as the number of computations for sampling, feature extraction and classification decrease with the sampling rate. The reduced feature set, obtained in the feature selection experiment, further reduced the computational effort and classification model size. Therefore, less features had beneficial effects on the system runtime with little loss of classification performance.

Reiss and Stricker [13] achieved an overall mean classification rate of 94.07% using three accelerometer sensors for the classification of activities grouped in three intensity classes and a sampling rate of 100 Hz. They investigated the accuracy of each single sensor, obtaining 90.47%, 86.47% and 88.08% for chest, dominant hand and foot sensor. These results are slightly better compared to the classification rates of the experiments with 102.4 Hz and two sensors (83.0% to 88.9%, Tab. 4). Reiss and Stricker acquired data from only 8 subjects performing 14 activities and thus analyzed a more homogenous dataset compared to our study. Furthermore, we designed the system for wearability and real-time processing. We assume that the higher amount of data as well as the accuracy and processing power tradeoff considerations are the main reason for the lower performance of our system. Our system might improve with an additional heart rate sensor, a more sophisticated feature set and more complex classification models.

We analyzed the misclassification of training data in more detail. The main source of error was imprecise labeling as subjects were not always able to perform the demanded activity throughout the complete interval. It was e.g. easy to collect cyclic activities like jogging where the same activity is continuously performed. However, when playing table tennis, subject were standing still when the opponent fetched the ball or had a chat when negotiating who serves next. In these situations, active and non-active phases alternated although the complete interval was labeled as active. Thus, the acyclic nature of many activities (e.g. ball games, strolling, painting) and the underlying imprecise labeling reduced the classification performance.

5.2 Performance Evaluation

The validation study with the CHRONACT system (85.6%) confirmed the expected classification performance (10.2 Hz, 20 features, CLR, 84.3%) for the three class setup. Misclassifications were mainly observed in strolling, ball aerobic and ball game. This confirms the findings on the test data that acyclic exercises comprise physical activity of different levels. Technically, these cases are not misclassifications but erroneous labels as we labeled the complete 5 min intervals instead of 5 s intervals and intensities with lower and higher intensity that the label were performed. For the physicians, the high granularity of 5 s is of interest as short activity bounds can be detected.

The performance in the four class setup was considerably lower. The reason for this decrease was the nature of the Mandala and Sudoku activities. All subjects were right handed and thus showed very litte activity in the left hand (sensor position) when performing these activities. Thus, LOW-ACT was misclassified to LOW-STA. This was expected as painting while sitting is very similar to just sitting. We anyhow included them as Mandala and Sudoku are common activities of the occupational therapy program and therefore performed very often. They provide edge cases that the presented methodology is not able to address. However, results are still valuable for the physicians when investigating activity profiles as there are phases when subjects grabbed a pencil or used their hand while chatting with other subjects. The system registered these phases as LOW-ACT and these activity bouts are visible in the activity profile. This additional distinction is of great value for the physicians as it helps them to distinguish phlegmatic patients and patients that participate in occupational therapy.

One possibility to cope with acyclic activities is averaging. Intermediate results (5 s intervals) can be combined to longer intervals that might provide a more valuable view on the activity profile of a patient. The mean values implied that classes can be well distinguished when averaged. This result is underlined with the small difference between label and mean value (Tab. 5). The mean value of the LOW-ACT shows its low precision. However, the LOW-ACT class can be well distinguished from the other classes although its mean values did not coincide with its label. Averaging the classification result can therefore level the misclassifications. In the results at hand, this was shown for misclassification due to the imprecise labeling.

5.3 System Runtime

The evaluation of the system runtime showed that the CHRONACT system is capable of supporting activity interventions of one week without charging or changing batteries. This is a valuable time period as exercise interventions are often conducted from Monday to Friday. A complete activity profile can currently only be stored for 39 h due to the limited amount of memory. This issue has to be addressed but preliminary results showed that the system is capable of logging ~100 h with data compression resulting in a granularity of 15 s. Nevertheless, the memory constraint only affects the activity profile logging and not the feedback capabilities.

5.4 General Discussion

The CHRONACT system is an activity tracker that was specifically designed for the use in mental health. The detection of activity levels was specifically designed for the exercises and activities that are prescribed and performed in clinical practice. One drawback of the system is that data for training and validation was collected with healthy subjects. On the one hand, this is legitimate as psychological patients do not show impairments like movement disorders in Parkinson's Disease. One the other hand, we have to expect slightly different movement profiles in all activity levels. Due to the high amount of training data and the variety in age, gender, fitness level, size and weight, we expect that the learned classification model generalizes good enough to cope with the slightly different movement profiles. This has to be proven in clinical trials.

The commercially available sensor hardware shows a high level of integration, as the wrist watch device limits stigmatization and the ankle device is even smaller and can be hidden underneath the socks. The runtime proved to be useful for exercise interventions of one week and the memory for activity profile logging showed to be the limiting factor. However, the CHRONACT system has to be regarded as a starting point for an even higher level of integration and longer runtimes. This can e.g. be achieved with data compression, results averaging and data logging on both devices (memory demand), miniaturization of sensor hardware (integration) and energy harvesting or batteries with higher capacity (runtime extension).

The results of the validation study were computed in real-time, displayed on the wrist device and logged to internal flash memory. Thus, results were instantly available and can be used in future augmented feedback applications. Still, aspects like feedback presentation and frequency need to be defined and integrated in the device software but the technical basis for such applications exists.

The objective classification of activity levels in daily life and exercise make it a useful and needed tool for mental health research [6]. An activity profile is instantly available during the next doctor's visit. As the results are available in real-time, the system can be used to investigate how activity records or feedback can enhance patient encouragement, short-term adherence to exercise programs and long-term adherence to physical activity.

6 Summary

This article introduced CHRONACT, a wearable real-time activity tracker for mental health. We described the development workflow and validated its functionality. The system implements a HAR and features optimizations to enable embedded implementation and enhance system runtime. In the validation study, it was capable of tracking the activity with a classification rate of 85.6%. The main source of error was identified as imprecise labeling of acyclic movements. Such a system can be used to create objective patient profiles of activity intensity. In future applications, CHRONACT can be used to assess the capabilities of short-term and long-term adherence to exercise interventions in mental health.

Acknowledgements This work was funded by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology and the European Fund for Regional Development. The authors thank all participants who volunteered in the study.

References

1. World Health Organization. Global Recommendations on Physical Activity for Health. *WHO Press*, pages 1–60, 2010.
2. C. D. Rethorst, B. M. Wipfli, and D. M. Landers. The Antidepressive Effects of Exercise. *Sport Med*, 39(6):491–511, 2009.
3. E. W. Martinsen. Benefits of Exercise for the Treatment of Depression. *Sport Med*, 9(6):380–389, 1990.
4. A. L. Dunn and J. S. Jewell. The Effect of Exercise on Mental Health. *Curr Sports Med Rep*, 9(4):202–207, 2010.
5. J. Mata, R. J. Thompson, S. M. Jaeggi, M. Buschkuehl, J. Jonides, and I.H. Gotlib. Walk on the Bright Side: Physical Activity and Affect in Major Depressive Disorder. *J Abnorm Psychol*, 121(2):297–308, 2012.
6. A. L. Brosse, E. S. Sheets, H. S. Lett, and J. A. Blumenthal. Exercise and the Treatment of Clinical Depression in Adults. *Sports Med*, 32(12):741–760, 2002.
7. G. J. Jerome, D. Rohm Young, A. Dalcin, J. Charleston, C. Anthony, J. Hayes, and G. L. Daumit. Physical Activity Levels of Persons with Mental Illness Attending Psychiatric Rehabilitation Programs. *Schizophr Res*, 108(1):252–257, 2009.
8. J. A. Blumenthal, P. J. Smith, and B. M. Hoffman. Is Exercise a Viable Treatment for Depression? *ACSMs Health Fit J*, 16(4):14–21, 2012.
9. E. M. Mello Portugal, T. Cevada, R. S. Monteiro-Junior, T. Teixeira Guimarães, E. da Cruz Rubini, E. Lattari, C. Blois, and A. Camaz Deslandes. Neuroscience of Exercise: From Neurobiology Mechanisms to Mental Health. *Neuropsychobiology*, 68(1):1–14, 2013.
10. J. S. Raglin. Exercise and Mental Health. *Sport Med*, 9(6):323–329, 1990.
11. H. Leutheuser, D. Schuldhaus, and B. M. Eskofier. Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. *PLOS ONE*, 8(10), 2013.
12. M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster. Classification of basic daily movements using a triaxial accelerometer. *Med Biol Eng Comput.*, 42:679–687, 2004.
13. A. Reiss and D. Stricker. Aerobic activity monitoring: towards a long-term approach. *Univ Access Inf Soc.*, 13:101–114, 2014.
14. A. Reiss. *Personalized Mobile Physical Activity Monitoring for Everyday Life*. PhD thesis, Technical University of Kaiserslauten, 2014.
15. A. Reiss, D. Stricker, and I. Lamprinos. An integrated mobile system for long-term aerobic activity monitoring and support in daily life. In *Proc: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 2021–2028, 2012.
16. M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE T Inf Technol B*, 12(1):20–26, 2008.
17. A. Bulling, U. Blanke, and B. Schiele. A Calculated Look at Fixed-point Arithmetic. *ACM Comput Surv*, 46(3):Article 33, 2014.
18. C. Wang, W. Lu, M. R. Narayanan, S. J. Redmond, and N. H. Lovell. Low-power Technologies for Wearable Telecare and Telehealth Systems: A Review. *Biomed Eng Lett*, 5(1):1–9, 2015.
19. R. Sigrist, G. Rauter, R. Riener, and P. Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychon Bull Rev*, 20(1):21–53, 2013.
20. P. Callaghan. Exercise: a Neglected Intervention in Mental Health Care? *J Psychiatr Ment Hlt*, 11(4):476–483, 2004.
21. C. Knöchel, V. Oertel-Knöchel, L. O'Dwyer, D. Prvulovic, G. Alves, B. Kollmann, and H. Hampel. Cognitive and Behavioural Effects of Physical Exercise in Psychiatric Patients. *Prog Neurobiol*, 96(1):46–68, 2012.
22. G. M. Cooney, K. Dwan, C. A. Greig, D. A. Lawlor, J. Rimer, F. R. Waugh, M. McMurdo, and G. E. Mead. Exercise for Depression. *Cochrane Database Syst Rev*, 9, 2013.
23. A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga. Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. In *Proc. of the International Conference on Architecture of Computing Systems (ARCS 2010)*, pages 1–10, Hannover, Germany, 2010.
24. O. D. Lara and M. A. Labrador. A Survey on Human Activity Recognition using Wearable Sensors. *Commun Surveys Tuts*, 15(3):1192–1209, 2013.
25. A. Zijlstra, M. Mancini, L. Chiari, and W. Zijlstra. Biofeedback for Training Balance and Mobility Tasks in Older Populations: A Systematic Review. *J Neuroeng Rehabil*, 7(1):Article 58, 2010.
26. M. Dogan-Aslan, G. F. Nakipoglu-Yüzer, A. Dogan, I. Karabay, and N. Özgirgin. The Effect of Electromyographic Biofeedback Treatment in Improving Upper Extremity Functioning of Patients with Hemiplegic Stroke. *J Stroke Cerebrovasc Dis*, 21(3):187–192, 2012.
27. D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. Implementation of a Real-time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. *IEEE Trans Inf Technol Biomed*, 10(1):156–167, 2006.
28. D. Albu, J. Lukkien, and R. Verhoeven. On-Node Processing of ECG Signals. In *Proc. of the 7th IEEE Consumer Communications and Networking Conference (CCNC)*, pages 1–5, Las Vegas, USA, 2010.
29. H. Ghasemzadeh, S. Ostadabbas, E. Guenterberg, and A. Pantelopoulos. Wireless Medical-Embedded Systems: A Review of Signal-Processing Techniques for Classification. *IEEE Sensors Journal*, 13(2):423–437, 2013.
30. Mark A. Hanson, H. C. Powell, A. T. Barth, and J. Lach. Application-Focused Energy-Fidelity Scalability for Wireless Motion-Based Health Assessment. *ACM TECS*, 11(S2):50:1–50:21, 2012.
31. S. J. Strath, L. A. Kaminsky, B. E. Ainsworth, U. Ekelund, P. S. Freedson, R. A. Gary, C. R. Richardson, D. T. Smith, and A. M. Swartz. Guide to the assessment of physical activity: Clinical and research applications - a scientific statement from the american heart association. *Circulation*, 128(20):2259–79, 2013.
32. C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A Triaxial Accelerometer and Portable Data Processing Unit for the Assessment of Daily Physical Activity. *IEEE Trans Biomed Eng*, 44(3):136–147, 1997.

33. B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. J. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon. 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Med Sci Sports Exerc*, 43(8):1575–1581, 2011.
34. I. N. Mohd Nordin, P. S. Chee, M. Mohd Addi, and F. K. Che Harun. EZ430-Chronos Watch as a Wireless Health Monitoring Device. In *Proc. of the International Conference on Biomedical Engineering (BIOMED 2011)*, pages 305–307, Kuala Lumpur, Malaysia, 2011.
35. A. Burns, B. R. Green, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca. Shimmer^{TM} - A Wireless Sensor Platform for Noninvasive Biomedical Research. *IEEE Sensors Journal*, 10(9):1527–34, 2010.
36. Texas Instruments Inc. Texas Instruments Wiki: EZ430-Chronos. `http://processors.wiki.ti.com/index.php/EZ430-Chronos`, 2014. [Online; accessed 30-Juli-2014].
37. H. Leutheuser, S. Doelfel, D. Schuldhaus, S. Reinfelder, and B. M. Eskofier. Performance Comparison of Two Step Segmentation Algorithms using Different Step Activities. In *Proc. of the International Conference on Wearable and Implantable Body Sensor Networks (BSN 2014)*, pages 143–148, Zurich, Switzerland, 2014.
38. Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). ActivityNet Benchmark Datasets: BaSA - Basic Step Activities. `http://www5.cs.fau.de/activitynet/benchmark-datasets/basa-basic-step-activities/`, 2014. [Online; accessed 30-Juli-2014].
39. M. Ring, U. Jensen, P. Kugler, and B. M. Eskofier. Software-based Performance and Complexity Analysis for the Design of Embedded Classification Systems. In *Proc. of the 21st International Conference on Pattern Recognition (ICPR 2012)*, pages 2266–2269, Tsukuba, Japan, 2012.
40. U. Jensen, M. Ring, and B. M. Eskofier. Generic Features for Biosignal Classification. In R. Byshko, T. Dahmen, M. Gratkowski, M. Gruber, J. Quintana, D. Saupe, M. Vieten, and A. Woll, editors, *KOPS Institutional Repository University of Konstanz*, pages 162–168, Konstanz, Germany, 2012.
41. D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms, 3rd Edition*. Addison-Wesley, 1998.
42. P. Pébay. Formulas for Robust, One-Pass Parallel Computation of Covariances and Arbitrary-Order Statistical Moments. Technical Report SAND2008-6212, Sandia National Laboratories, Livermore, USA, September 2008.
43. S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Fourth Edition*. Elsevier, 2009.
44. R. Polikar. Bootstrap-inspired Techniques in Computational Intelligence: Ensemble of Classifiers, Incremental Learning, Data Fusion and Missing Features. *IEEE Signal Process Mag*, 24(4):59–72, 2007.
45. I. H. Witten, E. Frank, and M. A. Hall. *Data Mining - Practical Machine Learning Tools and Techniques, 3rd Edition*, pages 124 – 131. Morgan Kaufman, 2011.
46. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition*. Wiley, 2001.
47. S. W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 2011.
48. R. Gordon. A Calculated Look at Fixed-point Arithmetic. *Embedded Systems Programming*, 11(4):72–79, 1998.