Enhancing the Reliability of Wearable Cardiac Monitoring using Accelerometer Activity Data

1st Katri Karhinoja
Department of Computing
University of Turku
Turku, Finland
katri.a.karhinoja@utu.fi

2nd Tuukka Panula
Department of Computing
University of Turku
Turku, Finland

3rd Tuija Leinonen
Department of Computing
University of Turku
Turku, Finland

4th Antti Airola
Department of Computing
University of Turku
Turku, Finland

5th Sari Stenholm
Unit of Public Health
University of Turku
Turku, Finland

6th Matti Kaisti
Department of Computing
University of Turku
Turku, Finland
mkaist@utu.fi

***Abstract*—We developed a system for enhancing reliability of cardiac monitoring using activity recognition. The link between activity information and recorded cardiac information could be used to better incorporate physiological state in analysis of heart rate in free-living conditions. Our approach uses a machine learning model to predict the activity based on accelerometer data that is subsequently used to estimate cardiac monitoring reliability. We collected proof-of-concept data from eight healthy volunteers using accelerometers on wrist and on thigh and an electrocardiogram (ECG). The measurement protocol included eight activities (lying, sitting, standing, walking, jogging, walking stairs up and down and cycling). Each measurement was one minute long and the set was repeated 5-10 times per participant. In addition, three participants conducted outdoor "free-living" measurements which were used for testing. Time- and frequency-domain features were extracted and XGBoost model was trained. The model achieved macro-average AUC of 0.97 in both leave-one-subject-out and leave-half-subject-out cross validation. In the "free-living" test measurements the activities were correctly predicted with a mean accuracy of 81 %. Quality of the ECG was the highest for lying and the lowest for jogging and the quality metric had a strong correlation with beat detection estimation error.**

***Index Terms*—Activity prediction, cardiac monitoring, electrocardiogram, machine learning**

I. Introduction

Recognizing the type of activity during vital sign monitoring is important for two main reasons. First, vital signs are affected by the physiological state and the origin of the change is typically difficult to know e.g. the changes can be a result of a pathological change or a physical activity. Second, vital sign monitoring is highly prone to environmental noise and artifacts. Physical activity creates motion artifacts to the recordings and by knowing to which extent, reliability of the estimates can be assessed. This is especially true in free-living monitoring, where signal variations can stem from a range of physiological states, encompassing both normal and abnormal changes, as well as various types of artifacts effectively limiting the ability to specify the origin of the change.

There have been several studies concerning physical activity and vital sign monitoring. In a recent study, a publicly available human activity recognition dataset was gathered during normal daily living, using two accelerometers and an F1 score of 0.81 for activity recognition was achieved. [1]. Another study, with both children and adults as test subjects, tested sensor placement combinations (thigh, back, wrist) and found that thigh and back combination provided best recognition performance. [2] Physical activity prediction has also been studied as a part of Body Sensor Networks where self-organising maps were used to predict the activity from accelerometer data [3] and more recently, deep learning in real-time classification using low-power wearable was demonstrated. [4] Furthermore, several approaches have been developed to assess if the quality of the cardiac monitoring is acceptable for diagnosis or not [5], [6], [7].

The aim of this research was to study whether recognizing the type of activity and calculating quality index for ECG could improve the reliability of cardiac monitoring. We collected proof-of-concept accelerometer data during different types of activities in controlled conditions and trained a machine learning model for predicting the activity in test measurements that mimics free-living conditions. Concept of the model is shown in Fig. 1. The computational engine consists of a multiclass XGBoost machine learning model and quality index calculator. The engine predicts the activity for every second and calculates the quality index of the ECG for every ten second window.

II. Materials and methods

A. Data collection devices

Wearable data acquisition system for ECG and motion data collection consisted a wrist-worn main unit, a thigh-mounted sensor node and an Android mobile phone (Samsung

Supported by Research Council of Finland (CLISHEAT, project number 352893)

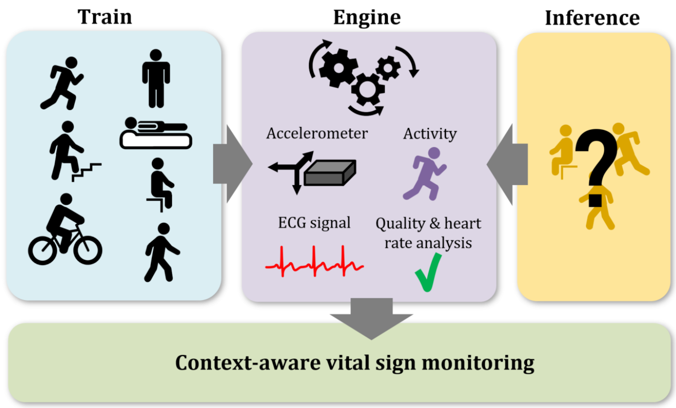

Fig. 1. Concept of the model. The model is trained with accelerometer data from different activities. The model predicts the physical activity and computes the heart rate and the quality of the ECG signal.

Galaxy S10, South Korea). The mobile phone has a GUI for interfacing with the wearable devices. The wrist unit houses a single channel ECG frontend (Maxim Integrated MAX30003, California), a 6-axis IMU (STMicroelectronics LSM6DSM, Switzerland) and a 1-Gbit flash memory IC (Cypress Semiconductor S70FS01GS, California). The thigh mounted device is equipped with the same 6-axis IMU. Both devices utilize a Bluetooth-ready nRF52 family (Nordic Semiconductor) microcontroller unit (MCU), specifically nRF52840 and nRF52832 in the wrist and thigh units respectively. The mobile phone and the wrist unit are connected via Bluetooth. The data from both devices are stored in the on-board flash memory as well as sent to the mobile phone. Both IMUs were sampled at 208 Hz and the ECG at 128 Hz. Data from the gyroscope was not used in this preliminary study. Fig. 2 shows the data acquisition system and the placement of the sensors.

B. Data acquisition

Activity measurements were taken from 8 healthy volunteers (2 females) with average age of 30.6 $\pm$ 4.2 years ranging from 26 to 40. The motion data was recorded using two motion sensors: one on wrist and one on thigh. Wrist was chosen as being one of the most common places for wearables. Measurements included eight activities: lying, sitting, standing, walking, jogging, walking stairs down and up, and cycling. Each activity was recorded for one minute (stairs for one floor) and each subject repeated the measurement cycle 5 to 10 times resulting in 30 to 65 minutes of data per participant. Apart from walking the stairs the activities were balanced in their recording time. In addition, three subjects carried out ”free-living” test measurements outdoors. The subjects were asked to walk, jog, sit, stand and walk stairs up and down during the 3-5 minute measurement. ECG was also recorded during the measurements. The cords of the ECG were taped to the body to reduce interference caused by the movement of the subject. All the measurements were performed according to the Declaration of Helsinki guidelines [8].

C. Data analysis

The measurements were analyzed using Python (version 3.9.7). The signals were filtered using 4th order Butterworth low-pass (accelerometer) and band-pass (ECG) filters with cut-off frequencies of 20 Hz and 0.5/40 Hz, respectively. The accelerometer signals were downsampled to 50 Hz and splitted to one second segments. For feature extraction, gravity and movement component of the accelerometer signal were computed by filtering the signal with a low-pass filter with a cut-off frequency of 1 Hz (gravity) and subtracting that from the raw signal (movement). The total acceleration of both sensors were calculated. Features were extracted from the accelerometer signal and from the gravity and movement component [1]. Extracted features are shown in table I.

XGBoost with default parameters (version 1.7.5) was used to predict the activity from the motion data. The model was evaluated using leave-one-subject-out cross validation (LOSOCV). This ensures that subject specific information does not leak from train to test data when evaluating the performance of the model. The model was also evaluated using leave-half-subject-out cross validation (LHSOCV) where half of the data from the left-out participant was included in the training set, and the other half in the test set. This mimics a personalised model training where the user's data can be included in the training data. Also, the model was tested using ”free-living” measurements to evaluate the performance of the model in an uncontrolled situation. The model was tested both with and without the data from the test subject in the training set. Accuracy and Area Under ROC curve (AUC) for one activity versus all activities were used as metrics. Also macro average of AUC values of all the activities was calculated (MA-AUC).

Quality index (QI) of the ECG was calculated for all activities. QI was computed for ten second segments by aligning heart cycles based on R peaks and computing a median waveform. The Pearson's correlation coefficient between the median and each of the heart cycle waveform was calculated and finally an average of those was taken. [7] The final QI of each activity is the mean of all segment QIs. R peak detection was conducted using pre-processing steps from Pan-Tompkins algorithm [9] and modified automatic multiscale-based peak detection (AMPD) algorithm for the peak detection [10]. Detected R peaks were post-processed to remove physiologically impossible peaks. The performance of the peak detection was evaluated by manually calculating false positive and negative detections.

III. Results and Discussion

Activity was correctly predicted using both sensors with a mean accuracy of 83.7 % using LOSOCV. Using the LHSOCV increased the mean accuracy to 87.4 %. MA-AUC was 0.97 with both cross validations. The AUC values were the lowest for sitting and walking stairs up (0.93 and 0.94, respectively). Testing of the ”free-living” measurements resulted in slightly lower mean accuracies and MA-AUC values (79.6 % and 0.93, respectively). In this experiment including data from the test

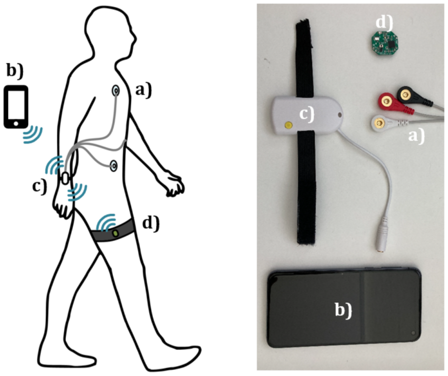

Fig. 2. Data acquisition system and the placement of the sensors: a) ECG electrodes, b) mobile phone, c) wrist sensor, d) thigh sensor.

TABLE I
FEATURES EXTRACTED FROM THE ACCELEROMETER SIGNAL.

Accelerometer signal
1^{st} and 2^{nd} degree polynomial fit of tilt (roll, pitch, yaw) Axis correlations: Between all axis and between total acceleration signals
Gravity component
Mean, median, standard deviation, coefficient of variation, 25^{th} percentile, 75^{th} percentile, minimum, maximum, number of zero crossings, number of turning points
Movement component
Skew, kurtosis, signal energy, frequency domain's magnitude's mean, frequency domain's magnitude's standard deviation, dominant frequency, magnitude of dominant frequency, spectral centroid, total signal power, number of zero crossings, number of turning points

subject in the training set did not meaningfully improve the results. The AUC values were also in this case lowest for walking stairs up (0.73 and 0.74). True and predicted activities during the "free-living" test measurement are shown in Fig. 3. All the results were lower when using only one sensor which is inline with previously reported results [11]. All the AUC values for different activities when using data from both sensors and the mean accuracies and MA-AUC values for all the experiments are shown in table II.

As can be seen from Fig. 3, most of the wrongly predicted activities were similar to the true activity. For example, walking was predicted as walking down or up the stairs and sitting down was predicted as lying, standing or cycling. Also, there were mostly only one or two wrong labels consecutively and hence post-processing the predictions could reduce wrong predictions. In addition, transitioning from one activity to another was labeled with the ongoing activity hence including event-type measurements to the dataset and labeling the transitions could improve the performance.

The training dataset was measured indoors whereas the "free-living" test set was measured outdoors. For example gait and pace of walking might differ indoors and outdoors. Expanding the dataset with measurements done outdoors could increase the performance.

As expected the quality of the ECG deteriorates with increasing movement being the lowest for jogging (0.73) and highest for lying (0.94). The quality of the ECG varied between the subjects which could be explained by differences in skin contact, shirt material and taping of the cords. Sensitivity and precision of R peak detection were also the lowest for jogging (0.87 and 0.97, respectively). Average QIs and sensitivity and precision of R peak detection for each activity are shown in table III. The QI during the "free-living" measurement is shown in Fig. 3. For this subject the quality of the ECG was good during all the activities.

The QI and sensitivity and precision of R peak detection both correlate well with the amount of movement in the activities. Activity prediction during vital sign monitoring could be used to help to evaluate the cause of alterations in vital signs or deteriorating quality of the vital sign signal.

TABLE II
CLASSIFICATION PERFORMANCE FOR EACH ACTIVITY USING BOTH SENSORS SEPARATELY AND TOGETHER.

AUC	LOSOCV	LHSOCV	free-living /wo	free-living /w
Lying	0.97	0.98		
Sitting	0.93	0.93	1.00	1.00
Standing	0.98	0.97	0.98	0.98
Walking	0.98	0.99	0.96	0.96
Jogging	1.00	1.00	0.99	0.98
Stairs down	0.97	0.98	0.92	0.92
Stairs up	0.94	0.94	0.73	0.74
Cycling	1.00	1.00		
MA-AUC / Mean accuracy (%)				
Thigh	0.93 / 69.69	0.96 / 75.09	0.88 / 57.19	0.89 / 55.34
Wrist	0.93 / 72.88	0.97 / 81.09	0.87 / 68.06	0.88 / 70.40
Both	0.97 / 83.71	0.97 / 87.36	0.93 / 79.62	0.93 / 80.65

AUC = area under ROC curve
MA-AUC = macro average of area under ROC curve
LOSOCV = leave-one-subject-out cross validation
LHSOCV = leave-half-subject-out cross validation
/wo = without data from the subject in the training set
/w = with data from the subject in the training set

REFERENCES

[1] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "Harth: A human activity recognition

TABLE III
ECG QUALITY INDICES (MEAN $\pm$ STANDARD DEVIATION) AND SENSITIVITY AND PRECISION OF R-PEAK DETECTION (MEAN [MIN, MAX]) FOR EACH ACTIVITY.

Activity	Quality	Sensitivity	Precision
Lying	0.94 ± 0.01	0.98 [0.91, 1]	1.00 [0.99, 1]
Sitting	0.94 ± 0.01	1.00 [0.94, 1]	1.00 [0.99, 1]
Standing	0.93 ± 0.02	1.00 [0.96, 1]	1.00 [0.96, 1]
Walking	0.86 ± 0.13	0.97 [0.67, 1]	0.99 [0.93, 1]
Jogging	0.73 ± 0.19	0.87 [0.61, 1]	0.97 [0.91, 1]
Stairs down	0.88 ± 0.15	0.98 [0.84, 1]	0.99 [0.94, 1]
Stairs up	0.87 ± 0.15	1.00 [0.95, 1]	1.00 [0.95, 1]
Cycling	0.92 ± 0.05	1.00 [0.95, 1]	1.00 [0.98, 1]

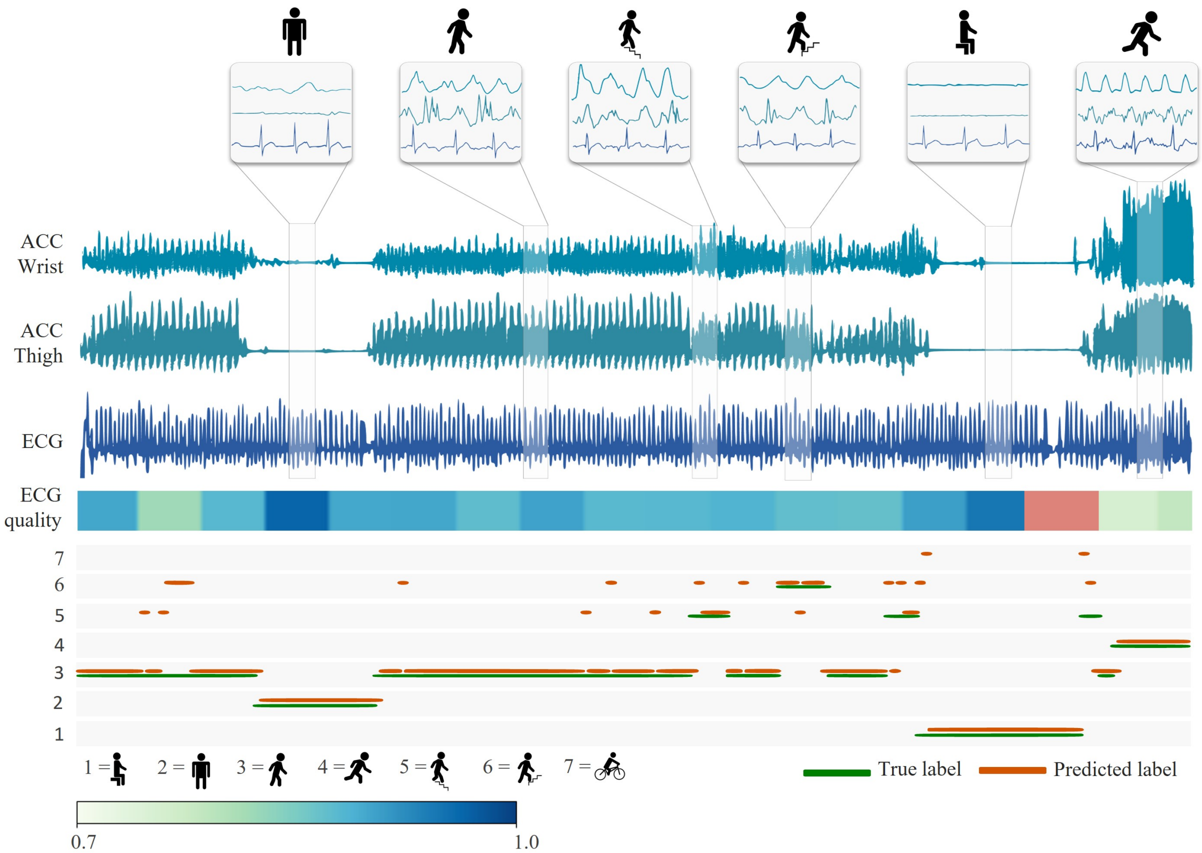

Fig. 3. Example of "free-living" test measurement and the outcome of the model. In the upper part of the figure there are signals from both accelerometers (wrist and thigh) and ECG, and zoomed in parts showing the signal during the activities. ECG quality illustrates the changes in the quality index of the ECG. The quality index could not be calculated from the part colored with red. The plot underneath shows the true and the predicted labels for the activities. The model was trained with the dataset including data from the subject.

dataset for machine learning," *Sensors*, vol. 21, no. 23, p. 7853, 2021.

[2] A. Narayanan, T. Stewart, and L. Mackay, "A dual-accelerometer system for detecting human movement in a free-living environment.," *Medicine and Science in Sports and Exercise*, vol. 52, no. 1, pp. 252–258, 2020.

[3] S. Thiemjarus, B. Lo, and G.-Z. Yang, "Context aware sensing — what's the significance?" *IET Seminar Digest*, vol. 2005, pp. 163–169, Nov. 2005. DOI: 10.1049/ic.2005.0799.

[4] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, IEEE, 2016, pp. 71–76.

[5] G. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiological measurement*, vol. 33, no. 9, p. 1419, 2012.

[6] J. Moeyersons, E. Smets, J. Morales, *et al.*, "Artefact detection and quality assessment of ambulatory ecg signals," *Computer methods and programs in biomedicine*, vol. 182, p. 105 050, 2019.

[7] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 832–838, 2015.

[8] W. M. Association *et al.*, "World medical association declaration of helsinki: Ethical principles for medical research involving human subjects," *Jama*, vol. 310, no. 20, pp. 2191–2194, 2013.

[9] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[10] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[11] I. Cleland, B. Kikhia, C. Nugent, *et al.*, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, 2013.