Topographic mapping for quality inspection and intelligent filtering of smart-bracelet data

Davide Bacciu · Gioele Bertoncini ·
Davide Morelli

Abstract Wrist-worn wearable devices equipped with heart activity sensors can provide valuable data that can be used for preventative health. However, hearth activity analysis from these devices suffers from noise introduced by motion artifacts. Methods traditionally used to remove outliers based on motion data can yield to discarding clean data, if some movement was present, and accepting noisy data, i.e. subject was still but the sensor was misplaced. This work shows that Self Organizing Maps (SOMs) can be used to effectively accept or reject sections of heart data collected from unreliable devices, such as wrist worn devices. In particular, the proposed SOM-based filter can accept a larger amount of measurements (less false negatives) with an higher overall quality with respect to methods solely based on statistical analysis of motion data. We provide an empirical analysis on real world wearable data, comprising heart and motion data of users. We show how topographic mapping can help identifying and interpreting patterns in the sensor data and help relating them to an assessment of user state. More importantly, our experimental results shows the proposed approach is able to retain almost twice the amount of data while keeping samples with an error that is an order of magnitude lower with respect to a filter based on accelerometric data.

D. Bacciu
University of Pisa
Tel.: +39-050-2212749
Fax: +39-050-2212726
E-mail: bacciu@di.unipi.it

G. Bertoncini
University of Pisa

D. Morelli
Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, OX2 6DP, United Kingdom, and
Biobeats Group LTD, Hatfields, London, England, SE1 9PG, United Kingdom

Keywords self-organizing maps · visualization · inspection · sensor data analysis

1 Introduction

Wearable sensing devices are becoming ubiquitous[1], in particular when considering smart-bracelets and smart-watches and especially within the context of fitness tracking and healthcare/well-being monitoring. These applications typically leverage two main information types, that are heartbeat and motion information, collected in a semi-continuous fashion throughout the day, where the monitoring rate is mostly the result of an energy-effectiveness tradeoff (the latter being application dependant).

Acquisition of heart-related and motion-related information is often mediated by sensing devices and operating conditions characterized by low signal-to-noise ratios [28], resulting in a collection of data that is deeply affected by artifacts. On the side of the heart-related data, artifacts are mostly the result of displacements of the sensor on the skin of the user as a result of daily or sport activities. When it comes to motion data, noisy measurements are mostly inherent in the nature of the inertial sensors, which capture accelerometric recordings whether they are relevant to a user activity or not.

Such a condition has motivated a whole body of research [18,28,29,35] aiming that pre-processing, adjusting and filtering data from wearable sensors, with the purpose of recording only the relevant and trustworthy measurements, while discarding noisy ones and artifacts. Being able to dismiss irrelevant data is not only an application-driven requirement aimed at performing monitoring and analytical procedures only on clean data. Rather, it is primarily a system-driven requirement, aimed at containing the use of the limited storage resources on-board the device and at preserving battery lifetime, by reducing the amount of unnecessary data transmitted by the wearable to the cloud or the device hosting the monitoring application. Overall, the noisy nature of sensed data together with stringent energy-saving requirements typically leads to extremely aggressive data filtering policies which typically discard every measurement taken under motion readings that are non-negligible [28]. As a result of this, the monitoring application is often deprived of valuable data, especially when sensed during the execution of daily activities which are associate to moderate-to-high motion.

Motivated by such an application scenario, in this paper, we explore the use of Self Organizing Maps (SOMs) [21] as an effective means to inspect wearable sensor data comprising both motion and heart-related information and to assess its quality. Following up of the result of this first analysis, we put forward a novel intelligent filter leveraging the topographic map trained on the sensor measurements and using auxiliary information to label the units of the map as being responsive to either noisy or trustworthy measurements.

---

[1] https://www.statista.com/outlook/319/100/wearables/worldwide, accessed 20th October 2020

By this simple approach, it is possible to determine whether or not to discard future measurements based on which unit of the map they are projected to.

Our work considers a real-world setting and associated data coming from the continuous passive monitoring of users during their daily activities, in a fully uncontrolled environment. In this context, we confront our simple SOM-based filtering approach with a consolidated motion-based filter [28], showing a consistent increase in effectiveness and performance, assessed both in terms of an higher number of retained sample and higher quality of the preserved data. Our analysis also shows that the SOM-based intelligent predictor is characterised by a good generalization across subjects, which is a rare result when dealing with the adaptive processing of biosignals, which are characterised by non-negligible and (mostly) physiological inter-subject differences. Finally, while SOMs are a widely popular and consolidated approach for exploratory data analysis, we believe this to be the first work in which it is proposed and highlighted their effectiveness in filtering heart-beat and accelerometric data.

The remainder of the paper is organized as follows: Section 2 provides a definition of the problem and the background on wearable-based monitoring. Section 3 describes the methodology used in the paper to realize the filter and analyse the data, providing a brief introduction to SOM for the sake of paper self-containment. Section 4 discusses our use case, providing results as concerns visual inspection of sensor measurements and and empirical validation of the proposed filtering approach; Section 5 concludes the paper.

2 Background and Problem Definition

2.1 Problem Definition

Wearable fitness monitors and smart-watches have become widely adopted. Many of those wearable devices have a photoplethysmography (PPG) sensor that can measure the users heart rate semi-continuously, 24/7.

In theory PPG devices allow the estimation of Inter Beat Intervals (IBI), i.e. the duration of every heartbeat. IBI data can be used to estimate Heart Rate Variability (HRV) [15,26], that is influenced by the activity of the Autonomous Nervous System (ANS) [27]. In turn, HRV can be used to estimate the users well-being and health [22,2,1], their recovery [16], and even to detect the onset of health deterioration [23].

However, PPG sensors on wrist-worn wearable devices are heavily affected by motion artifacts [28]. As a result of this, most fitness trackers only report the average heart rate (HR), i.e. the number of heartbeats in one minute. This calculation is resilient to outliers and noise, because HR is an average over one minute of data, equivalent to applying a low pass filter. This operation filters noise introduced by motion artifacts (that adds high frequency components to the signal).

Since the algorithms to pre-process PPG data have been defined in medical research, they could use clean signals captured in highly controlled environ-

ment (i.e. measurements in hospitals from clinical-grade PPG monitors). The algorithms used to remove outliers from IBI time series have been designed for a setting where most of the data is precise and reliable, with only sporadic noise. Most approaches identify outliers by using the distribution of the data, e.g. removing IBI points that differ by more than a fixed percentage from the average of the preceding IBI data [18]. This approach is only applicable if most of the IBI data is reliable.

Before fitness trackers, people would have their HRV measured only when they performed some medical exam. Wrist-worn wearable devices have introduced a different paradigm to HRV analysis. Instead of few data points collected in controlled environment, we now have access to a large amount of noisy data.

As previously discussed, existing noise removal algorithms are difficult to apply to this data because the assumption that most of the IBI data is reliable does not hold in this setting. There is therefore the need to find new algorithms that can automatically estimate when IBI from noisy PPG data should be considered clean enough to allow HRV estimations.

2.2 Biosignal Data Filtering

Previous work we have conducted found that statistics on accelerometer data can be used to estimate the quality of IBI data, and the error propagation to HRV features can be estimated [28,29]. However, in practice, most of the data collected during the day would be unusable. Outlier detection and interpolation techniques have been introduced in literature to handle occasional ectopic beats. Applying such techniques to frequent motion artifacts would make the HRV look closer to what is expected [28] therefore apparently more acceptable, but in practice they would also alter the original signal by adding information that was not present in the original data, and after this processing step it would not be possible to distinguish between true and arbitrarily reconstructed information.

As previously discussed, HRV processing literature is mainly concerned with occasional ectopic beats. Frequent outliers introduced by motion artifacts have been studied only recently, once the adoption of commercial grade wrist worn wearable devices produced a large amount of data, but with much lower quality (due to noisy sensors and uncontrolled environment) than the traditional HRV datasets, usually composed of ECG data collected in controlled environments, i.e. hospitals. Typical algorithms to detect and handle sporadic outliers identify extraneous beats by excluding heartbeats with duration outside physiologically plausible range, i.e. less than 250 ms or more than 1800 ms. Other algorithms also assume that heartbeats should have a duration similar to preceding heartbeats, excluding heartbeats that differ from the mean of the previous heartbeats by more than a predefined amount relative to the standard deviation of the previous heartbeats. Once outliers have been identified, they are dealt with in two ways: by deleting them, or by attempt-

ing the reconstruction of the original signal. The simplest way of handling the inconsistent RR-intervals provided in literature is to delete them [35]. In this approach, the abnormal heartbeats are simply removed from the list of heartbeats without replacement. This approach does not introduce extraneous information, but reduces the overall length of the heartbeats list, with potential implications for the HRV features extracted from the heartbeats [35]. More frequently, removed heartbeats are replaced with interpolation methods, such as zero-degree, linear, quadratic, or cubic spline interpolation [17]. Reconstructing the heartbeats list has the desirable property of preserving the original overall duration, but it also introduces extraneous information, visible in the HRV power spectrum [29]

2.3 Self Organizing Maps in Biosignal Processing

In this study, we outline a novel approach to IBI data filtering, based on Self Organising Maps (SOM) [21], that yield to more precise results, with lower error on HRV features, and allow to retain more datapoints. Being able to exploit the data collected from fitness trackers for HRV analysis could allow continuous passive monitoring of the health of the population, with a significant positive impact on their health status, enabling preventative health interventions to be activated on the onset of diseases instead of waiting for serious symptoms to develop

While we are not aware on previous works as concerns the use of SOMs to realize an intelligent filter for raw sensor data, their use on biosignal processing is evident from the literature. In [11] it is is presented a closely related work where SOMs are used to process Electrocardiography (ECG) features to partition normal from anomalous beats. In this work, however SOMs operate on engineered time domain features extracted from the ECG data rather than on the raw biosignals. Similarly, [10] has leveraged SOM as a means to preprocess and partition ECG features by telling apart normal beats from different classes of anomalous heartbeats, focusing in particular on an interactive and explorative approach to the data. The application of SOM on ECG data are numerous but they typically share the common trait of feeding the map with engineered features, e.g. a decomposition of the signal in Hermite functions [25], rather than inputting the raw biosignals. SOMs have also been used to provide interpretable models of the respiratory signals by considering the neurons in the map as identifiers of specific internal states of the dynamical system generating the respiratory timeseries [12]. A similar approach has been taken by [8,24] where they have been used to identify and visualize the different mental states from engineered features extracted from Electroencephalography data. SOMs have also been used in stress detection tasks [14] where they are fed with features from skin conductance and ECG data and the resulting maps are clustered by a Gaussian Mixture Model to identify SOM units that can be associated with relax phases or stress phases, in a fully unsupervised way.

3 Methodology

This study centers on analysing sensor data from two modalities, namely hearth-beat and accelerometric information with a two-fold aim. First we target an explorative analysis of the data, aiming to characterize prototypical sensor profiles which can be interpreted as proxies of the state and activities performed by the user wearing the smart-bracelet. The second aim is related to identifying which of such prototypical profiles can be associated with noisy or erroneous measurements to drive the implementation of an intelligent data filter. Our approach to the problem relies on topographic maps, whose use on explorative data analysis is widely consolidated [37, 19, 36, 20]. Their application to smart filtering is, instead, less known but, as it will become clear throughout our empirical analysis, no less effective.

Self-organizing Maps (SOMs) are undoubtedly the most popular family of neural-based approaches to topographic mapping. Their founding concepts were introduced in the seminal paper by Kohonen [21], mostly targeting a computational model of biological processes of neural soft-competition and self-organization in the motor-sensory cortex. SOMs have since then evolved as an effective computational methodology for adaptive data exploration, visualization and clustering.

In our work, we leverage a standard SOM model for vectorial data, comprising a single-layer of neurons organized into a grid-like lattice which associates to each unit u_i, where $i \in C$ indexes the neurons, a position on the grid $I(u_i) = (r_i, c_i)$, where r_i and c_i are row and column indexes, respectively. The SOM takes as input a vector $\mathbf{x}_j \in \mathbb{R}^d$ which, for the purposes of this work, is a 24-dimensional vector organized as follows (j-th index omitted to avoid cluttering)

$$\mathbf{x} = [h_1, \ldots, h_{12}, a_1, \ldots, a_{12}] \tag{1}$$

where h_l and a_l denote, respectively, the heart-beat and accelerometer value for the l-th time window of the $\mathbf{x}$ measurement. The 24-dimensional vector $\mathbf{x}$ collects the heart-beats detected by the sensor on 12 successive windows of measurements, each lasting 10 seconds, followed by the corresponding 12 measurements collected by the accelerometer sensor on the same time windows.

Each neuron u_i is associated to a prototype vector $\mathbf{w}_i \in \mathbb{R}^d$, also referred to as codebook, defined in data space which characterizes the preferred input stimuli for the neuron. Whenever a sample $\mathbf{x}_j$ is presented to the network, the Best Matching Unit (BMU) u_* is determined according to a similarity measure between a sample and each weight vector which, for our purposes, leverages the Euclidean distance

$$u_* = \arg \min_{u_i \in \{u_1, \ldots, u_C\}} \|\mathbf{w}_i - \mathbf{x}_j\|_2. \tag{2}$$

During map usage, the BMU is the one where the input sample is projected to for analysis and visualization purposes, i.e. an input $\mathbf{x}_j$ is projected on the lattice in position $I(u_*)$. During training, the activation of the BMU is propagated to its neighbors on the lattice to implement a biologically inspired

competition mechanisms. The neighbourhood can have different shapes. In our work we use the standard radial basis function

$$N(u_*, u_i)^t = \exp\left(-\frac{\|I(u_*) - I(u_i)\|_2}{2(\sigma^t)^2}\right) \tag{3}$$

where $\sigma^t = \sigma^0 \exp(-\frac{t}{\tau})$ is an exponentially decaying neighborhood radius, with σ^0 being the initial radius and τ a time-constant determining the slope of the decaying weight. The learning phase adapts the prototype vectors to become representative of a group of input training vectors following a competitive soft-update rule

$$\mathbf{w}_i^{t+1} = \mathbf{w}_i^t + \nu^t N(u_*, u_i)^t \left(\mathbf{w}_i^t - \mathbf{x}_j\right) \tag{4}$$

where ν^t is a time decaying learning rate (typically with an exponential schedule). The rationale of (4) is to move the prototype weight vectors $\mathbf{w}_i$ of both the winner unit and of its neighbors closer to the current input sample. By these means nearby neurons tend to become responsive to similar stimuli, hence realizing topographical organization of the neural lattice. In a discrete setting, the learning rule in (4) can also be seen as the result of a stochastic gradient descent on the loss [33]

$$\mathcal{L} = \frac{1}{2M} \sum_{j=1}^{M} \sum_{i=1}^{C} N(u_j^*, u_i) \|\mathbf{w}_i - \mathbf{x}_j\|^2 \tag{5}$$

where u_j^* is the BMU for data point $\mathbf{x}_j$.

The choice of founding our approach on a SOMs is based on the following considerations

– SOMs are a simple model which allows for a straightforward interpretation of the results and of the neuron responses.
– Our data comprises sensor measurements whose noise can be more favourably managed through a flexible non-parametric neural approach, rather than with parametric models requiring the identification of accurate noise distributions, such as in Generative Topographic Mapping [6].
– SOMs are a consolidated model with robust and well validated implementations and provide a variety of visualization tools and methodologies to represent and interpret the results of the trained map (e.g. the unified distance matrix).
– SOMs can be implemented efficiently, with contained memory fingerprint and without resorting to heavy numerical computing libraries. These aspects are of paramount importance as they allow to implement the intelligent filter to run embedded on the bracelet device, which is a key requirement in order to save battery and communication costs.

We believe that, altogether, the characteristics listed above make SOM an effective methodology for tasks related to quality assessment and filtering of noisy, and potentially erroneous, sensor measurements. For our purposes,

the analysis of the data samples from the application described in Section 2 proceeds according to the the following. First, we train the SOM with data samples arranged as in (1) and we inspect the resulting unit prototypes to identify and visualize the prototypical measurement patterns present in our training data (e.g. presence of spiked accelerometer activations at specific time slots). As an additional level of analysis we consider the availability of samples $\mathbf{x}_j$ and associated explanatory information y_j. The term y_j is a scalar providing aggregated statistics related to $\mathbf{x}_j$. For the purposes of our analysis, we consider different types of explanatory information for $\mathbf{x}_j$, including

- the average activation of the accellerometers across the 12 time-points in $\mathbf{x}_j$;
- the average number of hearbeats across the 12 time-points in $\mathbf{x}_j$;
- a measure of heartbeat estimation error obtained by confronting the smart-bracelet readings in $\mathbf{x}_j$ with the aligned readings from an external device (i.e. chest strap).

For this latter analysis, we feed the trained SOM with samples $\mathbf{x}_j$ and assign them to their BMU. Based on such an assignment, for a full sample batch, we can compute aggregated statistics of the explanatory information associated to each neuron

$$\hat{y}(u_i) = \frac{\sum_{j=1}^{N} bmu(u_i, \mathbf{x}_j) y_j}{\sum_{j=1}^{N} bmu(u_i, \mathbf{x}_j)}$$

where $bmu(u_i, \mathbf{x}_j)$ is an indicator function returning 1 if u_i is the BMU for $\mathbf{x}_j$ and 0 otherwise. The $\hat{y}(u_i)$ term can be computed for each unit and visualized on the lattice to provide insights, e.g., on the areas of the map characterized by higher heartbeat estimation errors. We will show a practical application of this analysis to our scenario in Section 4.

Sensor-sourced information has an inherent sequential nature which represents the evolution of the measurement in time. In this respect, SOM has long since been extended to a recurrent approach to deal with input information that is of sequential nature, such as in the Temporal Kohonen Map [9] and the Recurrent Self-Organizing Map [38]. Similarly generative topographic maps have found extension to deal with timeseries data analysis and visualization [5,30]. Both neural and generative maps have been extended further to deal with more articulated structured data types, comprising trees and directed acyclic graphs, such as in the SOM for Structured Data [13], Recursive SOM [39] and the GTM for Structured Data [3,4]. For the sake of our application, we are restricted to consider information of a vectorial nature, as the sensor measurements collected by the smart-bracelet come bundled in fixed length vectors comprising sensor readings for 12 time samples. Nonetheless, the analysis presented in the following can be replicated for variable length sequences of measurements, by adopting any of the temporal SOM extensions referenced above.

4 Results and Discussion

The section discusses the application of our SOM-based approach to quality inspection and filtering of heartbeat and accellerometer data, using publicly available samples collected as part of two independent experiments. The first dataset comprises smart-bracelet data from activities of daily living. The second dataset [34] comprises smart-bracelet data aligned with chest-strap measurements. The former is used to train the SOM on real-world noisy data, while we exploit the resulting topographic mapping for interpretation purposes. The latter serves to provide a qualitative and quantitative assessment of the performance of our SOM-based filtering procedure, leveraging chest-strap measurements as ground-truth heartbeat information. We compare the effectiveness of our SOM-based filtering against a standard non-adaptive filtering policy.

4.1 Datasets and Experimental Setup

Our analysis considered datasets from two experiments, where participants were equipped with a wrist-worn wearable device to collect IBI data through the Photoplethysmogram (PPG) sensor. The wearable device is a BioBeam (BioBeats Ltd)[2], a customization of a LifeSense Band 2 with firmware modified to collect IBI. The band is a commercial grade fitness monitor with (with CE marking).

- The first dataset (D1) contains an extraction of IBI data collected on a group of 273 users through wristband sensor [32]. The participants used the BioBeam for four weeks. The data gathering protocol turns on the IBI sensor for 2 minutes every 30 minutes. In this window of time, the device provides also a measure of the motion as the standard deviation of the magnitude of the axes of the accelerometer, calculated as described in [28]. All participants provided informed, electronic consent prior to their enrollment in the study. Data from this study, including the preregistration protocol, are available on the Open Science Framework website [31]. This study was approved by an Institutional Ethics Committee at the University of Exeter (UEBS Research Ethics Committee, ethics application number: eUEBS002252).
- The second dataset (D2) is a publicly available dataset [34], that contains IBI and motion measurements, collected using the BioBeam device and the protocol described above, from a group of 22 people (with no overlap to D1), with the IBI sensor active continuously for 24 hours. In addition D2 contains a ground truth for the IBI time series collected with an ECG chest strap (Polar H7)[3].

---

[2] `https://fccid.io/OU9LS417-F01/User-Manual/User-Manual-part-2-3373513`

[3] `https://fccid.io/INWX0/User-Manual/H7-user-manual-1615156`

After the preprocessing phase described in Section 4.2 the number of data points in D1 and D2 are 312316 and 3796 respectively.

In the following subsections, we provide a multi-faceted analysis of the datasets above. First, in Section 4.2, we give a qualitative interpretation of the sample by visualizing prototypical behaviours and aggregated statistics of the sensor data on the SOM topographic map. At this stage of the analysis, we have considered different configurations of the map, using neuron grids of hexagonal topology with different dimensions: 8x8, 16x16, 56x56, 96x96. The models have been trained applying the online update rule, described in (4), for 100 epochs with linear decreasing learning rate $\nu \in [0.05, 0.01]$ and initial radius of the neighborhood function $\sigma^0 = 2/3 \cdot d_m$, where d_m refers to the diameter of the grid, during training the radius is linearly decreased from σ_0 to 0. The analysis in the following (including model training and visualization functions) have been implemented using the `kohonen` R package [41] [40]. For the sake of conciseness, we report only a sample of results corresponding to the 16x16 map, leaving the other visualizations to the appendix. All models were trained with data from D1. The choice of a 16x16 map has been motivated both by considerations as regards quality of the visualization and by avoiding an excess of sparsely populated neurons, i.e. neurons which are BMU for few samples (see the results in Appendix for the larger maps).

In Section 4.3, we shift our attention to using the SOM map as a predictor of the quality of sensor measurement data. In particular, we assess whether specific neurons in the map and their associated prototype (representing a characteristic behaviour of sensed data in the measurement window) can be used as proxy of the accuracy of the heartbeat estimated by the smart-bracelet. To this end, we label each neuron in the map as discussed in Section 3, using as auxiliary information the error between wristband and chest strap measurements averaged on all data samples projected on the unit, and we provide a visualization of this information on the map. Projected data samples are taken from D2 while the map is the 16x16 SOM obtained from the previous step and trained on D1 data only. To compute the error with respect to the ground truth, we used the following metrics, considering a segment $\mathbf{t} = [t_1, ..., t_n]$ of intervals between successive heartbeats (IBI):

– Average value of the IBI:

$$\mathrm{ANN}(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^{n} t_i$$

– Standard deviation of the IBI [27]:

$$\mathrm{SDNN}(\mathbf{t}) = \sqrt{\frac{\sum_{i=1}^{n} (t_i - \mathrm{ANN}(\mathbf{t}))^2}{n}}$$

– RMSSD: Root mean square of the differences between successive IBI [27]:

$$\mathrm{RMSSD}(\mathbf{t}) = \sqrt{\frac{\sum_{j=1}^{n-1} d_j^2}{n-1}}$$

where $d_j = t_{i+1} - t_i$.

Finally, in Section 4.4, we provide an assessment of the predictive value of a filter built on the top of the error maps discussed above, confronting it with a standard data filtering policy based on predefined thresholds on the accelerometer measurements. For this purpose, we again consider the D2 data with chest-strap as a ground truth and the 16x16 map fit on D1.

4.2 Visualization and Inspection

In order to train the SOM, data have been transformed into 24-dimensional vectors described in (1). In particular the first 12 components h_i are the number of the heartbeats detected by the sensor on successive windows of 10 seconds and the latter 12 components a_i are motion-related measurements collected on the same time windows. As anticipated, we have fitted maps of different size, but in the following we analyse the results restricted to the 16x16 map, which is the best suited for interpretability and visualization purposes (the reader is referred to the appendix for further visualizations). Fig. 1 shows the codebooks of the 16x16 SOM trained on D1 data: this visualization allows to appreciate different activation and intensity patterns in the two sensor modalities. By looking at such patterns, it is possible to straightforwardly discern measurements that can be associated to different activities being performed by the subject. For instance, the unit marked with a red `A` in Fig. 1 seems prototypical of measurements pertaining physical activities, given the contemporaneous high cardiac activity and motion activation (we will see in the following how such behaviour is not the result of noisy measurements due to wrist-band movement). Instead, neurons in regions of the map marked as `B` and `C` are characteristic of measurements with low accelerometer readings and different heartbeat patterns. In particular, unit `C` characterizes by very low heartbeats across all components, paired with low accelerometer value, and can be thus interpreted as responding to measurements taken during sleep cycles. Neurons from the `B` region, on the other hand, capture more peculiar heartbeat patterns of mild arousal which is not backed-up by physical activation (as measured by accelerometers). Hence they might have an interesting interpretation as proxies for measurements taken during an emotional state denoted by stress or anxiety of the subject.

The detailed visualization of the prototype map can be complemented with aggregated statistics, as discussed in Section 3. In particular, Fig. 2a and 2b show the distribution of average heartbeat counts and motion values across the 12 time measurements computed with respect to the prototypes components of each neuron. These heatmaps provide a snapshot view of the regions of the map characterized by different degrees of physical and cardiac activity, where the top-right area of the map is the region of interest characterizing measurements with possible links with intense emotional states. Fig. 2c completes the picture by providing an insight on the number of hits per neuron, i.e. the amount of samples projected on each unit of the map. From this view one can note how,
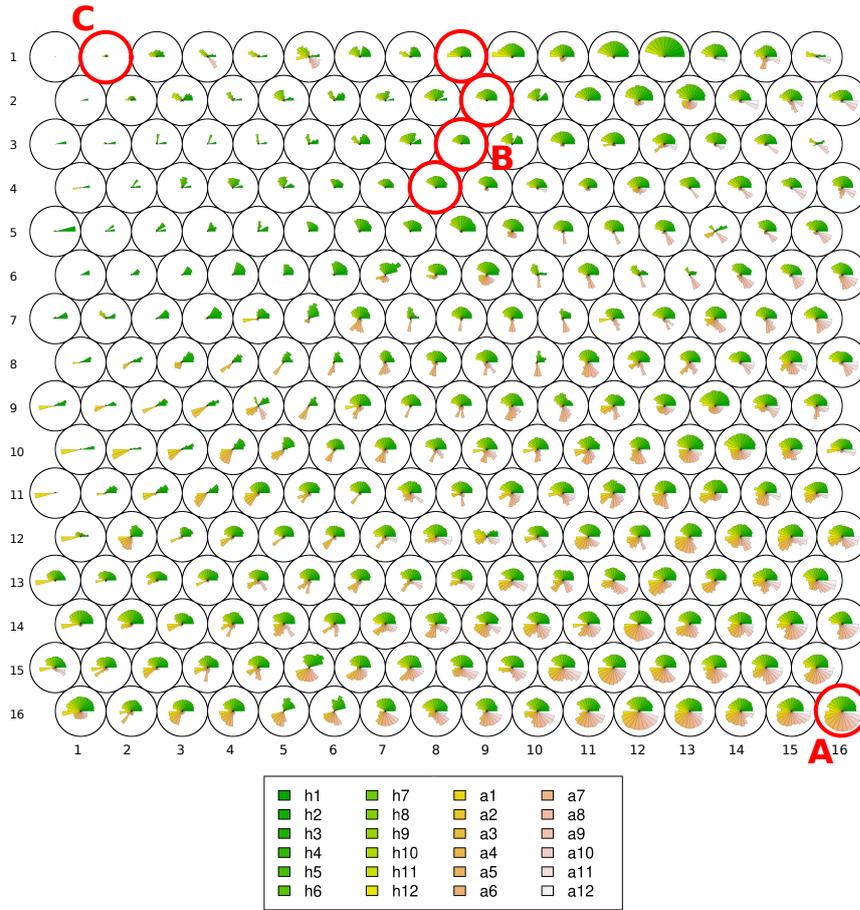
Fig. 1: Codebook map [40, 41] displaying unit prototypes for the 16x16 map trained on D1 data (row and column numbers are the neuron indexes on the grid). For each unit (circle) we report an histogram representing the value of the 12 heartbeat features (upper-half of the histogram in the plot, marked as h in the legend) and 12 accelerometer features (bottom-half of the histogram in the plot, marked as a in the legend) of its prototype vector. The numbers by the h and a letters in the legend identify one of the 12 time windows of sensor readings within each packet sent by the device. For instance, $h1$ and $a1$ are the heart-beat counts and associated accelerometer readings for the first time window.

unsurprisingly, neurons characterized by low activity values tend to be highly popular, since these are the measurements collected during sleep-time. More interestingly, one can note how the area marked as B in Fig. 1 and associated to intense emotions, also contains neurons with a large number of hits.

(a) Average on heartbeat components

(b) Average on motion components

(c) Neuron hits count
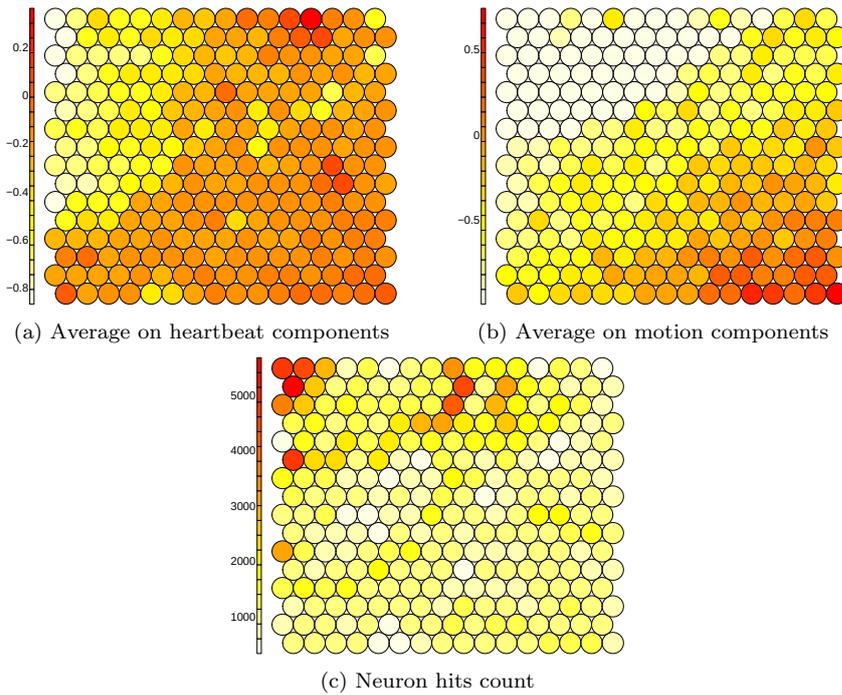
Fig. 2: Heatmaps displaying aggregated statistics of the 16x16 SOM trained on the D1 data.

4.3 Assessment of measurement quality

Having analysed how the SOM can identify and distinguish different and interesting patterns in the data collected by the wrist-band sensors, we now turn our attention to investigate the quality of such measurements. In particular, we are interested in assessing whether the activity patterns are the results of *physiological* activity or if, instead, they are due to noise and artifacts during data acquisition. For instance, one common assumption in the monitoring of daily living activities is that smart-bracelet heartbeat measurements cannot be trusted when they are associated to moments of high accelerometer activity (e.g. because of possible alterations in the measurements captured by the PPG sensor introduced by the wristband movement) [35,17].

In order to do this, we have set up an experiment leveraging D2 data and associating to each measurement in this dataset a scalar measure of discrepancy between the cardiac activity, as estimated using the wristband, and the one estimated by the chest-strap (here considered as ground truth). In particular, we have computed the Root-Mean-Square Error (RMSE) between the wristband and chest-strap estimates of the three scores introduced in Section 4.1, that are ANN, SDNN and RMSSD. Based on this information, we have computed the distribution of the error for each map unit, again considering

the average error for the measurements projected in each neuron. Note that the SOM model is the same discussed Section 4.2 and trained solely on D1 data, while D2 samples used for projection purposes only. Fig. 3 shows the unit-specific errors for the map. Gray-shaded units are those that never won the competition for any of the samples in D2, that is no example in D2 is assigned to a grey unit. Fig. 3a clearly points out an area of the map with few units are characterized by the highest errors with respect to the ANN metric. The surprising insight delivered by the error heatmaps is that the units responding to high accelerometer readings (bottom-right on the maps) are not characterized by substantially higher estimation errors. This suggests that such highly valuable samples (because they are related to physical activity in the subjects) need not to be discarded a-priori (as previously suggested by common practice).

(a) RMSE on ANN

(b) RMSE on SDNN

(c) RMSE on RMSSD
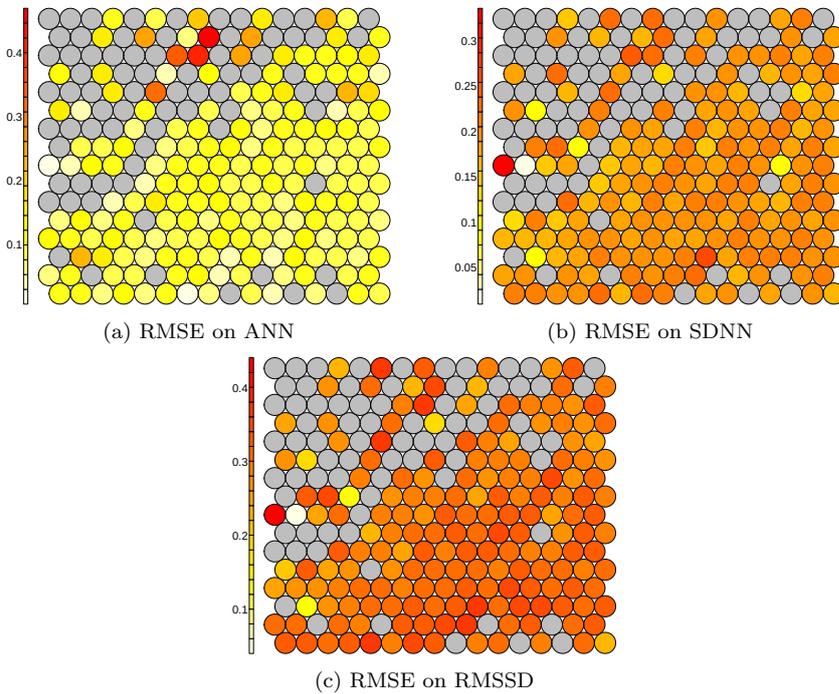
Fig. 3: Distribution of cardiac activity estimation errors on the map, computed with respect to ANN, SDNN and RMSSD scores.

To assess whether the results found on the maps in Fig. 3 could be generalized to new data, we have built a predictive nearest-neighbor-based setting. We have split D2 data in four partitions to perform a 4-fold cross validation. For each fold $k \in 1, 2, 3, 4$, we take 3 splits to compose the training data (TR_k) and the remaining split as validation (VL_k). Given a training partition TR_k

		ANN	SDNN	RMSSD
S1	Mean	0.0745	0.0473	0.0691
	Standard deviation	0.0014	0.0003	0.0006
S2	Mean	0.0854	0.0492	0.0722
	Standard deviation	0.0186	0.0085	0.0106

Table 1: Mean and standard deviation of the validation RMSE on the 4-fold for the three metrics and the SOM-based predictor.

we project its samples on the SOM previously trained on D1 and we label each unit in the map with the corresponding average $RMSE_k$, as described above. For each training partition TR_k, we estimate the accuracy of the corresponding predictor by projecting the VL_k samples on the map, associating to each VL_k example the $RMSE_k$ of the unit it was projected to. As an aggregated performance value, we have computed the validation RMSE between the actual cardiac-estimation error (known from D2) and the one predicted by the map labelling (for each validation set VL_k). Table 1 reports the mean and the standard deviation of the validation error across the four folds. We have applied the same procedure to the centroids produced by running the kmeans algorithm with 256 clusters on D1 data, to provide reference results on the task. Table 2 reports the corresponding validation RMSE.

We have used two different strategies to assemble the cross-validation splits. The first strategy (S1) splits the data in four random samples, each comprising 25% of the original D2 data. The second strategy (S2) creates the cross-fold partitions on a subject level. Each partition (of the 4-fold) contains all data from subjects that only belong to that partition. This ensures that the validation partition at iteration k (i.e. VL_k) contains data from subject unseen in TR_k, to provide an measure of the out-of-subject generalization. Also in S2, subjects are split in such a way that each fold again contains a number of samples that roughly corresponds to 25% of the original D2 data.

The validation errors for both setups are reported in Table 1 and Table 2 for SOM and kmeans predictors, respectively. These results show for both methods an excellent generalization of the quality assessment predictions in all three metrics and across both validation setting. When comparing the SOM-based and kmeans-based predictors, the former shows a better performance on the ANN score, in both the S1 and S2 settings. Overall the two methods tend to behave similarly in this predictive setting, which is not surprising given the tight relationships existing between the two models [7]. On the other hand, we believe that SOM offers some additional advantages over a kmeans-based approach, that are related with a better interpretability, which is a key factor for our explorative analysis in Section 4.2.

4.4 Comparative assessment

Once we have determined that the quality assessment of the SOM generalizes well to unseen data, to conclude our analysis, we consider its practical

		ANN	SDNN	RMSSD
S1	Mean	0.0763	0.0464	0.0680
	Standard deviation	0.0023	0.0011	0.0017
S2	Mean	0.0875	0.0491	0.0719
	Standard deviation	0.0159	0.0089	0.0112

Table 2: Mean and standard deviation of the validation RMSE on the 4-fold for the three metrics and the kmeans-based predictor.

application as an intelligent filter to discard noisy/unreliable measurements. We define a rejection threshold that determines if a neuron is unreliable, and hence all measurements mapped to it. Based on domain knowledge on IBI data, we have determined that measurements for which the ANN RMSE exceeds a $\theta = 10\%$ threshold are excessively noisy and can be filtered out. Hence, we label all neurons as being associated to reliable/unreliable measurements according to such threshold. By this means, we are able to assess the reliability of a new sample by feeding it to the network and associating it with the label of the BMU.

We have compared our method, referred to as f_{som} in the following, with a filter (f_a) from literature [28] which determines sample rejection based on motion-related features. We assess f_{som} and f_a in terms of both quantity and quality of the retained data. In particular, we measure the proportion of discarded samples with respect to the full data and we compute the quality of retained measurements as the ANN RMSE between wristband and chest-strap on the preserved samples. These performance metrics are computed as average values on the four D2 validation splits under the S1 and S2 cross-validation setups discussed in the previous section.

The results in Table 3 show that f_{som} is advantageous in both terms, being able to retain almost twice the amount of data retained by f_a while achieving an RMSE that is an order of magnitude lower. As a final step of verification, we have applied our f_{som} filter to the data retained by the f_a filter. By projecting these data on our map, one can appreciate how f_a, despite having rejected more samples, retains a significant amount of measurements that are labelled as noisy by our method. In particular, when considering out of subjects data in D2, 79% of the samples that were considered reliable by f_a are projected by the map on units that have an ANN percentage greater than the threshold θ. All in all, these results confirm that the proposed intelligent filter can be valuable for practical use, as it allows to retain more, and more accurate, data for later usage in human state monitoring applications leveraging noisy sensor data from wearable devices.

5 Conclusions

We have introduced what we believe to be the first application of SOMs to quality inspection and intelligent filtering of wearable sensor data. Despite being one of the longest standing neural models, SOM still supports effective

			ANN-RMSE	% of discarded samples
S1	f_a	Mean	0.2570	97.3
		Standard deviation	0.0146	0.9
	f_{som}	Mean	0.0409	52.8
		Standard deviation	0.0011	4.4
S2	f_a	Mean	0.1848	97.3
		Standard deviation	0.1101	0.6
	f_{som}	Mean	0.0411	48.9
		Standard deviation	0.0037	11.24

Table 3: Performance comparison between the SOM-based f_{som} and motion-based f_a filters: results are assessed in terms of percentage of discarded samples and average RMSE on ANN for the retained data.

exploratory analysis and visual inspection of heterogeneous and noisy data, providing valuable insights into the data patterns, especially when used in combination with heatmaps of aggregated statistics associated to input samples. In this work, we have focused on a specific case study supported by real-world data from a personal well-being application, comprising both heartbeat and accelerometer data. Throughout our SOM-based analysis, we have identified various prototypical patterns that can be associated to relevant physical and mental states of the human subjects involved in the study. More importantly, we have put forward a predictive use of the SOM that allows assessing the reliability of a sensor measurement inputted to the network. We have shown how the resulting SOM-based filter can achieve consistent improvements with respect to a widely-used motion-based filter, increasing both the quality and the quantity of the retained data. While in the present work we have focused on a vectorial representation of the sensor measurements, it would be interesting to assess how recurrent extensions of the SOM can be applied to a sequential representation of the sensory information. On a more practical side, it would be interesting to leverage the contained complexity of the SOM to explore a filter implementation embedded on the wearable device. Such a solution might prove advantageous in terms of energy-efficiency, allowing the device to transmit only those measurements that are marked as reliable by the on-board filter.

Funding

Conflict of interest

D. Bacciu has received funding from Biobeats Group Ltd. D. Morelli is affiliated to Biobeats Group Ltd.

Appendix

The appendix complements the visual analysis in Section 4.2. In Figure 4, we provide another example of a 16x16 map generated with a different random initialization to show reproducibily of the results in Figure 1. One can clearly appreciate the coherence of the outcomes in the two maps, which appear one the mirror version of the other with respect to the vertical axis.
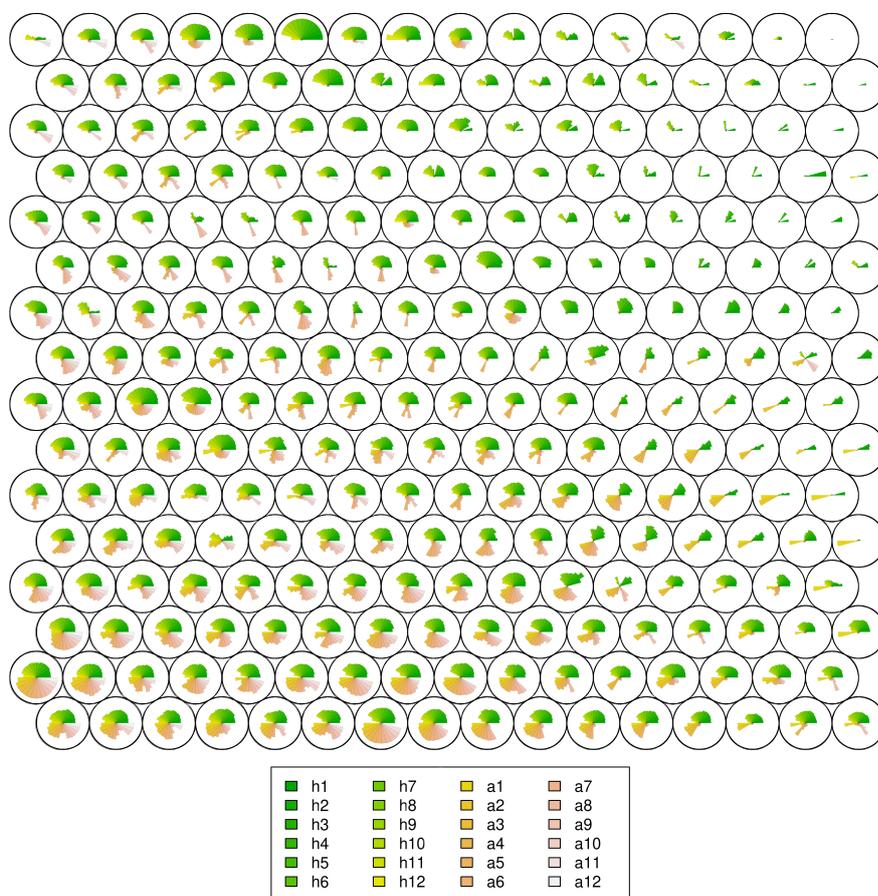
Fig. 4: Additional codebook map displaying unit prototypes for the 16x16 map trained on D1 data with different random initialization.

The remainder of the plots provides maps for other grid sizes explored in our empirical analysis, i.e. 8x8, 56x56 and 96x96. While for the former we provide both codebook plots and heatmaps, for the latter two maps we omit codebooks because of their limited readability.
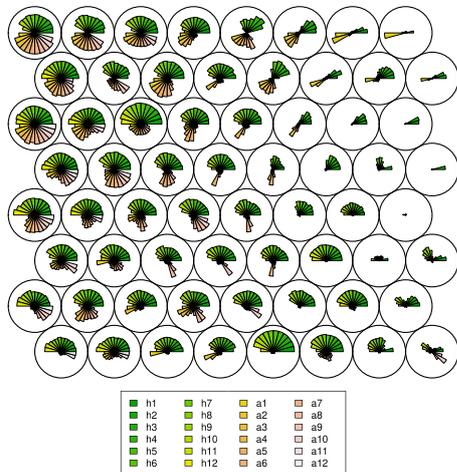
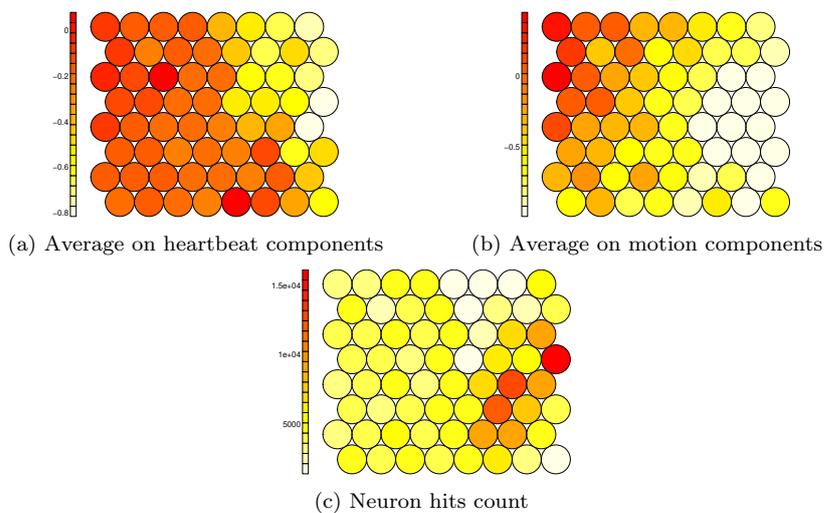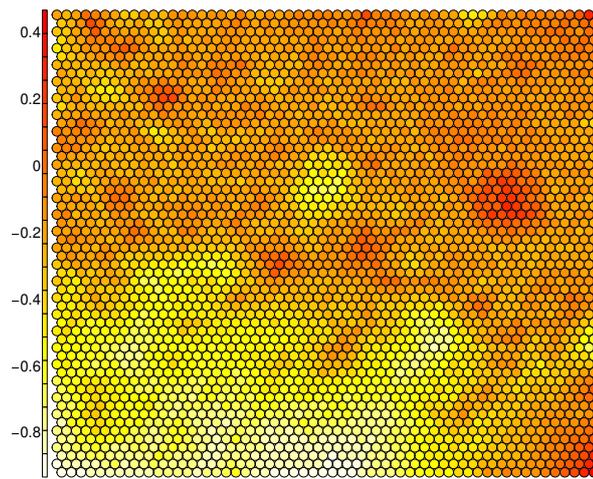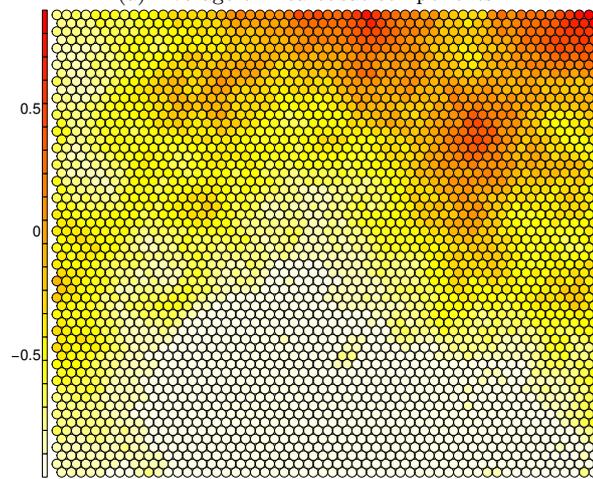Fig. 5: Codebook map displaying unit prototypes for the 8x8 map.

(a) Average on heartbeat components

(b) Average on motion components
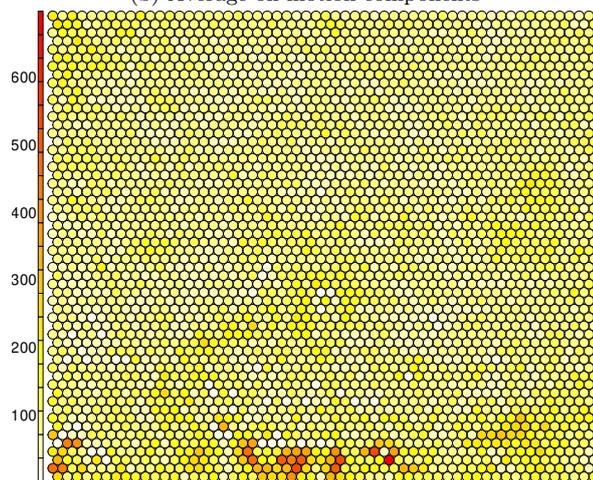

(c) Neuron hits count

Fig. 6: Heatmaps displaying aggregated statistics of the 8x8 SOM.
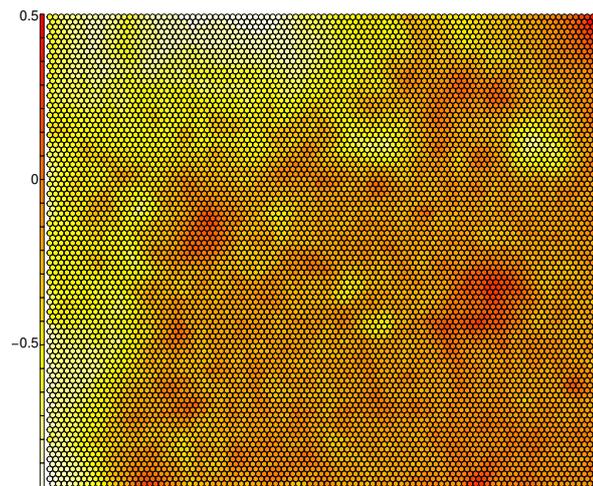
(a) Average on heartbeat components
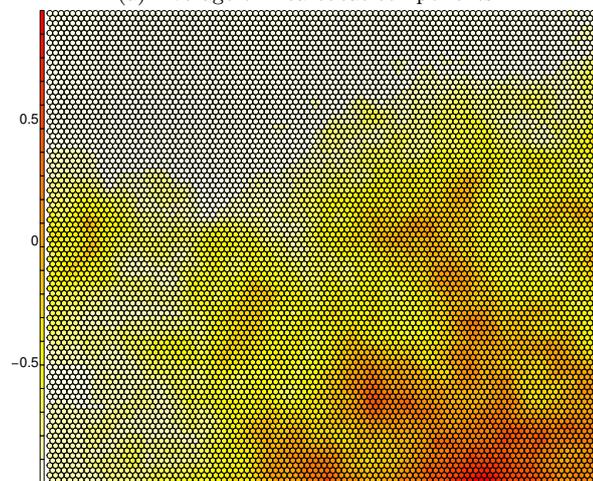

(b) Average on motion components
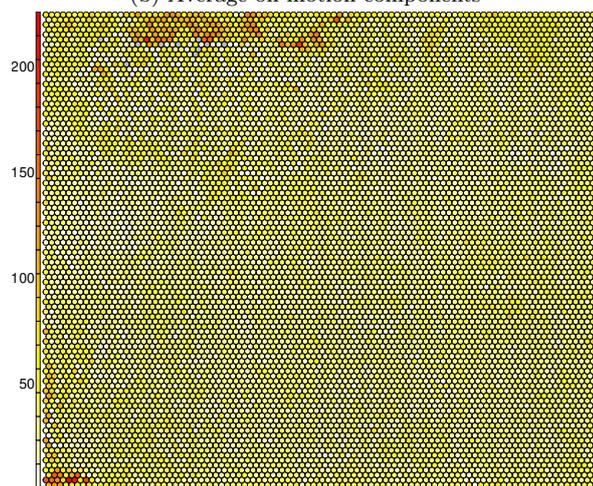

(c) Neuron hits count

Fig. 7: Heatmaps displaying aggregated statistics of the 56x56 SOM.

(a) Average on heartbeat components

(b) Average on motion components

(c) Neuron hits count

Fig. 8: Heatmaps displaying aggregated statistics of the 96x96 SOM .

References

1. Bacciu, D., Colombo, M., Morelli, D., Plans, D.: Elm preference learning for physiological data. proceedings of the european symposium on artificial neural networks, computational intelligence and machine learning. (esann'17). In: Proceedings of the European Conference on Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017), pp. 99–104 (2017)
2. Bacciu, D., Colombo, M., Morelli, D., Plans, D.: Randomized neural networks for preference learning with physiological data. Neurocomputing **298**, 9 – 20 (2018). DOI https://doi.org/10.1016/j.neucom.2017.11.070. URL http://www.sciencedirect.com/science/article/pii/S092523121830211X
3. Bacciu, D., Micheli, A., Sperduti, A.: Compositional generative mapping of structured data. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2010)
4. Bacciu, D., Micheli, A., Sperduti, A.: Compositional generative mapping for tree-structured data—part ii: Topographic projection model. IEEE Transactions on Neural Networks and Learning Systems **24**(2), 231–247 (2013)
5. Bishop, C., Hinton, G., Strachan, I.: Gtm through time. In: Artificial Neural Networks, Fifth International Conference on (Conf. Publ. No. 440), pp. 111–116. IET (1997)
6. Bishop, C., Svensen, M., Williams, C.: Gtm: The generative topographic mapping. Neural Computation **10**, 215–234 (1997). DOI 10.1162/089976698300017953
7. Bodt, E., Verleysen, M., Cottrell, M.: Kohonen maps versus vector quantization for data analysis. In: Proceedings of the European Conference on Neural Networks, Computational Intelligence and Machine Learning (ESANN 1997) (1997)
8. Bueno, L., Bastos Filho, T.F.: A self-organizing maps classifier structure for brain computer interfaces. Research on Biomedical Engineering **31**, 232 – 240 (2015)
9. Chappell, G.J., Taylor, J.G.: The temporal kohønen map. Neural Networks **6**(3), 441 – 445 (1993). DOI https://doi.org/10.1016/0893-6080(93)90011-K. URL http://www.sciencedirect.com/science/article/pii/089360809390011K
10. Gacek, A.: Preprocessing and analysis of ecg signals – a self-organizing maps approach. Expert Systems with Applications **38**(7), 9008 – 9013 (2011). DOI https://doi.org/10.1016/j.eswa.2011.01.119. URL http://www.sciencedirect.com/science/article/pii/S0957417411001394
11. Ghosal, P., Sarkar, D., Kundu, S., Roy, S., Sinha, A., Ganguli, S.: Ecg beat quality assessment using self organizing map. In: 2017 4th International Conference on Opto-Electronics and Applied Optics (Optronix), pp. 1–5 (2017). DOI 10.1109/OPTRONIX.2017.8349994
12. González-Obregón, C., Horowitz, R.: Self organizing maps in respiratory signals classification. In: C. Müller-Karger, S. Wong, A. La Cruz (eds.) IV Latin American Congress on Biomedical Engineering 2007, Bioengineering Solutions for Latin America Health, pp. 988–991. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
13. Hagenbuchner, M., Sperduti, A., Ah Chung Tsoi: A self-organizing map for adaptive processing of structured data. IEEE Transactions on Neural Networks **14**(3), 491–505 (2003)
14. Huysmans, D., Smets, E., De Raedt, W., Van Hoof, C., Bogaerts, K., Van Diest, I., Helic, D.: Unsupervised learning for mental stress detection exploration of self-organizing maps. pp. 26–35 (2018). DOI 10.5220/0006541100260035
15. Jeyhani, V., Mahdiani, S., Peltokangas, M., Vehkaoja, A.: Comparison of hrv parameters derived from photoplethysmography and electrocardiography signals. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5952–5955 (2015)
16. Kaikkonen, P., Hynynen, E., Mann, T., Rusko, H., Nummela, A.: Can HRV be used to evaluate training load in constant load exercises? European Journal of Applied Physiology **108**(3), 435–442 (2010). DOI 10.1007/s00421-009-1240-1. URL https://doi.org/10.1007/s00421-009-1240-1
17. Kamath, M., Fallen, E.: Correction of the heart rate variability signal for ectopics and missing beats. in heart rate variability. ed M Malik and AJ Camm (1995)

18. Karlsson, M., Hörnsten, R., Rydberg, A., Wiklund, U.: Automatic filtering of outliers in RR intervals before analysis of heart rate variability in Holter recordings: a comparison with carefully edited data. BioMedical Engineering OnLine **11**(1), 2 (2012). DOI 10.1186/1475-925X-11-2. URL `https://doi.org/10.1186/1475-925X-11-2`

19. Kaski, S., Kohonen, T.: Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. In: Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets, pp. 498–507. World Scientific (1996)

20. Knopf, G.K., Sangole, A.: Scientific data visualization using three-dimensional self-organizing feature maps. In: 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236), vol. 2, pp. 759–764 vol.2 (2001). DOI 10.1109/ICSMC.2001.973006

21. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biological Cybernetics **43**(1), 59–69 (1982). DOI 10.1007/BF00337288. URL `http://dx.doi.org/10.1007/BF00337288`

22. Kristal-Boneh, E., Raifel, M., Froom, P., Ribak, J.: Heart rate variability in health and disease. Scandinavian Journal of Work, Environment & Health **21**(2), 85–95 (1995). URL `http://www.jstor.org/stable/40966336`

23. Kudaiberdieva, G., Görenek, B., Timuralp, B.: Heart rate variability as a predictor of sudden cardiac death. Anatolian Journal of Cardiology/Anadolu Kardiyoloji Dergisi **7** (2007)

24. Kuzovkin, I., Tretyakov, K., Uusberg, A., Vicente, R.: Mental state space visualization for interactive modeling of personalized BCI control strategies. Journal of Neural Engineering **17**(1), 016059 (2020). DOI 10.1088/1741-2552/ab6d0b. URL `https://doi.org/10.1088%2F1741-2552%2Fab6d0b`

25. Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., Sornmo, L.: Clustering ecg complexes using hermite functions and self-organizing maps. IEEE Transactions on Biomedical Engineering **47**(7), 838–848 (2000). DOI 10.1109/10.846677

26. Lin, W.H., Wu, D., Li, C., Zhang, H., Zhang, Y.T.: Comparison of heart rate variability from ppg with that from ecg. In: The International Conference on Health Informatics, pp. 213–215. Springer (2014)

27. Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal **17**(3), 354–381 (1996). DOI 10.1093/oxfordjournals.eurheartj.a014868. URL `https://doi.org/10.1093/oxfordjournals.eurheartj.a014868`

28. Morelli, D., Bartoloni, L., Colombo, M., Plans, D., Clifton, D.A.: Profiling the propagation of error from ppg to hrv features in a wearable physiological-monitoring device. Healthcare Technology Letters **5**(2), 59–64 (2018)

29. Morelli, D., Rossi, A., Cairo, M., Clifton, D.A.: Analysis of the impact of interpolation methods of missing rr-intervals caused by motion artifacts on hrv features estimations. Sensors **19**(14), 3163 (2019)

30. Olier, I., Vellido, A.: Advances in clustering and visualization of time series using gtm through time. Neural Networks **21**(7), 904 – 913 (2008). DOI https://doi.org/10.1016/j.neunet.2008.05.013. URL `http://www.sciencedirect.com/science/article/pii/S0893608008001196`

31. Plans, D., Morelli, D., Murphy, J., Ponzo, S., Kawadler, J.: Biobase rct 1 (2019). URL `osf.io/2zd45`

32. Ponzo, S., Morelli, D., Kawadler, J.M., Hemmings, N.R., Bird, G., Plans, D.: Efficacy of the digital therapeutic mobile app biobase to reduce stress and improve mental well-being among university students: Randomized controlled trial. JMIR Mhealth Uhealth **8**(4), e17767 (2020). DOI 10.2196/17767

33. Ritter, H., Martinetz, T., Schulten, K.: Neural Computation and Self-Organizing Maps; An Introduction, 1st edn. Addison-Wesley Longman Publishing Co., Inc., USA (1992)

34. Rossi, A., Pozzo, E.D., Menicagli, D., Tremolanti, C., Priami, C., Sirbu, A., Clifton, D., Martini, C., Morelli, D.: Multilevel monitoring of activity and sleep in healthy people. PhysioNet (2020)

35. Salo, M.A., Huikuri, H.V., Seppanen, T.: Ectopic beats in heart rate variability analysis: effects of editing on time and frequency domain measures. Annals of noninvasive electrocardiology **6**(1), 5–17 (2001)

36. Ultsch, A.: Self-organizing neural networks for visualisation and classification. In: O. Opitz, B. Lausen, R. Klar (eds.) Information and Classification, pp. 307–313. Springer Berlin Heidelberg, Berlin, Heidelberg (1993)

37. Ultsch, A., Siemon, H.P.: Kohonen's self organizing feature maps for exploratory data analysis. In: B. Widrow, B. Angeniol (eds.) Proceedings of the International Neural Network Conference (INNC-90), Paris, France, July 9–13, 1990 1. Dordrecht, Netherlands, vol. 1, pp. 305–308. Kluwer Academic Press, Dordrecht, Netherlands (1990)

38. Varsta, M., Heikkonen, J., Lampinen, J., Millan, J.d.R.: Temporal kohonen map and the recurrent self-organizing map: Analytical and experimental comparison. Neural Processing Letters **13**, 237–251 (2001). DOI 10.1023/A:1011353011837

39. Voegtlin, T.: Recursive self-organizing maps. Neural Networks **15**(8), 979 – 991 (2002). DOI https://doi.org/10.1016/S0893-6080(02)00072-2. URL `http://www.sciencedirect.com/science/article/pii/S0893608002000722`

40. Wehrens, R., Buydens, L.M.C.: Self- and super-organizing maps in R: The kohonen package. Journal of Statistical Software **21**(5), 1–19 (2007). DOI 10.18637/jss.v021.i05

41. Wehrens, R., Kruisselbrink, J.: Flexible self-organizing maps in kohonen 3.0. Journal of Statistical Software **87**(7), 1–18 (2018). DOI 10.18637/jss.v087.i07