



**HAL**  
open science

# Riemannian ElectroCardioGraphic signal classification

Aurélien Appriou, Fabien Lotte

► **To cite this version:**

Aurélien Appriou, Fabien Lotte. Riemannian ElectroCardioGraphic signal classification. LAWCN 2023 - IV Latin-American Workshop on Computational Neuroscience, Nov 2023, Envigado, Colombia. hal-04372193

**HAL Id: hal-04372193**

**<https://inria.hal.science/hal-04372193v1>**

Submitted on 4 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Riemannian ElectroCardioGraphic signal classification

Aurélien Appriou<sup>1\*</sup> and Fabien Lotte<sup>2</sup>

<sup>1\*</sup>Flit Sport, 20 rue Joseph Le Brix, Bordeaux, 33000, France.

<sup>2</sup>Inria Center at the University of Bordeaux / LaBRI, 200 avenue de la Vieille Tour, Talence, 33400, France.

\*Corresponding author(s). E-mail(s): [aurelien.appriou@flit-sport.fr](mailto:aurelien.appriou@flit-sport.fr);  
Contributing authors: [fabien.lotte@inria.fr](mailto:fabien.lotte@inria.fr);

## Abstract

Estimating mental states such as cognitive workload from ElectroCardioGraphic (ECG) signals is a key but challenging step for many fields such as ergonomics, physiological computing, medical diagnostics or sport training. So far, the most commonly used machine learning algorithms to do so are linear classifiers such as Support Vector Machines (SVMs), often resulting in modest classification accuracies. However, Riemannian Geometry-based Classifiers (RGC), and more particularly the Tangent Space Classifiers (TSC), have recently shown to lead to state-of-the-art performances for ElectroEncephaloGraphic (EEG) signals classification. However, RGCs have never been explored for classifying ECG signals. Therefore, in this paper we design the first Riemannian geometry-based TSC for ECG signals, evaluate it for classifying two levels of cognitive workload, i.e., low versus high workload, and compare results to the ones obtained using an algorithm that is commonly used in the literature: the SVM. Our results indicated that the proposed ECG-TSC significantly outperformed an ECG-SVM classifier (a commonly used algorithm in the literature) when using 6, 10, 20, 30 and 40-seconds time windows, suggesting an optimal time window length of 120 seconds (65.3% classification accuracy for the TSC, 57.8% for the SVM). Altogether, our results showed the value of RGCs to process ECG signals, opening the door to many other promising ECG classification applications.

**Keywords:** Mental Workload, Machine Learning, Classification, Riemannian geometry, Electrocardiographic signals

# 1 Introduction

Monitoring users' mental states such as cognitive workload or affective states, in real time, has recently aroused a particular interest from fields such as neuroergonomics [1], Human Computer Interaction (HCI) [2] or passive Brain-Computer Interfaces (BCI) [3, 4]. Subjective measures, e.g., the NASA-TLX questionnaire [5], as well as behavioural measures, e.g., reaction time or error rate, have been used for assessing these mental states for a while. However, new methods using physiological sensors have been recently developed in order to bypass the lack of objectivity and the discontinuity of measurements of such methods. This field, known as "physiological computing" aims at extracting real-time information about users' states from physiological data such as heart rate, electrodermal activity or breathing [6].

Among the commonly used physiological sensors, ElectroCardioGraphy (ECG) has proven to be a useful signal in order to estimate mental states such as cognitive workload [7] or emotions [8]. In this paper, we focus in particular on the estimation of cognitive workload: based on the Cognitive Load Theory (CTL) [9], the cognitive workload is defined as the amount of cognitive/working memory resources that are necessary to process the information. While this mental state can be measured subjectively using questionnaires or estimated from behavioural measures, it can also be estimated physiologically [10]. This can enable continuous and objective mental workload estimation without interrupting the user. Among the various physiological signals that can be relevant for workload estimation, e.g., electrodermal activity [11], electroencephalogram [3] or electrooculogram [12], ECG was shown to be well related to cognitive workload and to be the most commonly used physiological measure to estimate it [13].

## 1.1 ECG signal classification algorithms

If ECG has been widely used in the literature in order to study mental workload [13], the amount of studies that have attempted to estimate such a mental state using machine learning methods remains limited. Regarding ECG-based classification of mental workload levels, authors commonly focus on feature extraction methods, e.g., time domain and frequency domain-based features, combined with classical linear classifiers, e.g., Support Vector Machine (SVM) or Linear Discriminant Analysis (LDA). The time domain features include the descriptive statistics around the R-R interval (RRI), i.e., the interval between two successive heartbeats, or more precisely, the interval between two R peaks in the ECG signal [13], such as the standard deviation, the mean or the root mean square of these R-R intervals [14]. The frequency domain features are usually the power spectral densities in some frequency bands, e.g., the power spectral density of very low frequency band (0.003–0.040 Hz) to high frequency band (0.15–0.4 Hz), and power spectrum-based operations, e.g., the ratio, the normalization and/or the log transformation of various frequency band power. Additionally, non-linear features such as Poincaré plot, sample entropy or Shannon entropy can be used as well [15]. Concerning the classification methods that have been used for discriminating different levels of cognitive workload using ECG only, many machine learning algorithms have been used in the literature: SVM, LDA, K-Nearest Neighbor (KNN),

Decision Tree (DT), Gentle Boost (GB), Naive Bayes (NB), Artificial Neural Network (ANN) and Logistic Regression (LR) [16–18]. Note that other algorithms can be used in the literature in order to classify ECG features combined with features from other sensors (Electroencephalography - EEG, skin conductance, etc). For example, some studies combined ECG with EEG features to classify workload levels, e.g., [19]. However, it did not result in reliable improvements when compared to the classification performance obtained from a single physiological signal.

## 1.2 ECG-based Mental Workload classification

In [20], authors applied a NB classifier on ECG data from twenty participants and obtained 76% of classification accuracy for discriminating low from high workload, levels that have been induced with elementary cognitive tasks [21]. In the same study, they applied a SVM with a cross-validation scheme on 30 participants using a 4-classes paradigm. Results showed a 80% classification accuracy. Authors chose to induce these 4 levels of cognitive workload using a combination of driving and N-back tasks, i.e., driving only vs driving with 3 different N-back task levels. Hogervorst & Brouwer applied both an Elastic Net (61% classification accuracy) and an SVM (55% classification accuracy) for discriminating two levels of workload (low vs high) [19]. These workload levels have been induced on 14 participants using the N-back task, and 120s trials were fed to the algorithms. In [16], authors compared 5 classification algorithms in order to discriminate two levels of workload (low vs high) from 13 subjects using both driving and N-back tasks. The KNN significantly under-performed (71.5%) compared to ANN (74%), NB (74.1%) and LR (73.9%), but no difference was shown compare to DT (72.8%). In [22], authors applied a logistic regression in order to classify 2 levels of cognitive workload (low/high) and showed a average classification accuracy of 62.5% from 16 subjects. The high workload was induced by 6 different tasks, i.e., Go/No-go visual reaction, stroop, fast counting, speed run, visual forward digit span, and working memory tasks, whereas the low workload was based on a resting phase. Then, Tjolleng et al. designed an ANN in order to classify 3 states of workload, i.e., low, medium and high, induced using both simulated driving tasks and N-back tasks [23]. 15 subjects participated to the study whose results showed a classification accuracy reaching 82%. It is unfortunate though that this ANN was not compared to classical Machine Learning (ML) methods such as LDA or SVM. Parent et al. used a modified version of the N-back task, i.e., the Toulouse N-back task [24, 25], in order to classify mental workload from ECG signals into 3 classes [26], using 22 types of features (e.g., time domain or frequency domain features). Results showed a classification accuracy of 42% with a NB classifier onto 18 participants. In [17], authors compared 6 classification methods on Heart Rate Variability (HRV) features from 6 subjects in order to discriminate 2 levels of workload (low vs high) from 3-minutes sliding windows of ECG using 10-fold cross-validation with a subject-specific calibration. Results showed that on average, the SVM, the KNN and the GB obtained classification accuracies above 95%, outperforming results from LDA (52.27%), NB (84.99%) and DT (89.91%). Note that the workload levels were induced with an interaction with a robot, resulting in 10 minutes of ECG recordings when the subject does not perform any task and maintains a relaxed state, and 10 minutes when the subject

is asked to perform a complex task involving an interaction with the robot. In [18], authors focused on feature extraction in order to improve the classification accuracy in a two workload level paradigm, the levels being induced using the Multi-Attribute Task Battery (MATB-II, National Aeronautics and Space Administration (NASA), USA) [27]. They obtained 91% of classification accuracy by applying an SVM with a Radial Basis Function (RBF) kernel with cross-validation. Meteier et al. obtained 89% classification accuracy by applying a RF on ECG features in order to classify two levels of workload (low vs high) [28]. They used both a driving and an oral backward counting task [29] in order to induce different levels of workload from 90 participants.

In summary, most, if not all, of the features and machine learning algorithms cited in 1.1 were used to classify mental workload levels from ECG signals. Authors have used a number of workload inducing tasks such as N-back tasks [19], backward counting [28] or driving tasks [20], among others. Depending on the classifier used, the length of the time window of analysis and the mental workload inducing task, authors obtained accuracies between 55% to above 95% of accuracy, often around 75% on average, to discriminate two levels of workload, i.e., low versus high workload [16–20, 28]. Studies that compared various classifiers together, suggested that SVM and ANN usually give state-of-the-art accuracies, and often outperform other algorithms such as LDA, NB or kNN [16, 17]. Therefore, in this paper, we focus on the use of SVMs, both as a baseline state-of-the-art ECG classifier, and as a classifier used in the Riemannian Tangent Space of ECG covariance matrices, which we propose as a new classifier in this paper (see below for more details). We then compare this new Riemannian SVM ECG classifier to the classical use of SVM applied to euclidean ECG feature vectors.

### 1.3 Paper objectives

The accuracies obtained so far for classifying cognitive workload levels, mostly around 75% from long time windows (usually several minutes), revealed the need for more robust and accurate ECG classification algorithms in order to obtain trustable ECG-based cognitive load estimators. Therefore, we propose in this paper to explore a type of algorithms that proved very effective for EEG classification, but that has never been studied, to the best of our knowledge, for ECG classification: Riemannian Geometry-based Tangent Space Classifiers (TSC) [3, 30, 31]. Such family of algorithms represents EEG signals as Symmetric Positive Definite (SPD) matrices, typically covariance matrices, and then directly classify such matrices using Riemannian geometry. This last few years, Riemannian Geometry Classifiers (RGCs) have been shown to be the state-of-the-art EEG signal classification algorithms, winning the vast majority of international brain signal classification competitions [30, 32–34]. Their effectiveness is thought to be due to 1) their small numbers of parameters, which enables them to be trainable from little training data and 2) the invariance properties of the Riemannian distance (which is invariant to full rank linear transformations) used in RGCs, which enables RGCs to be naturally robust to some variabilities affecting EEG signals [30, 32, 33]. ECG and EEG signals share numerous properties, notably both are biomedical electrical time series, both suffer from various variabilities and classifiers of both signals usually need to be calibrated from little training data. Thus, it seems promising to explore RGCs for ECG signals classification as well. Therefore, in this

paper, we first propose a novel representation of ECG signals as SPD matrices. We then assess how well a Riemannian TSC could classify such ECG SPD matrices, and compared it to euclidean SVM ECG classifiers used as baselines. In this paper, we first present the workload data set used, then the ECG features we extracted as well as the machine learning algorithms employed. We used a subject-specific design - with each algorithm trained on data specific to each subject - and we then tested them on different data from the same subject - using a 5-fold cross-validation.

## 2 Materials and methods

### 2.1 Mental workload data set

The data set used for this study comes from [35], which conducted a workload inducing experiment on 22 subjects, while collecting physiological signals such as EEG (the initial focus of that paper) and ECG. Consent forms had been filled-in and signed by the subjects in the original paper study. Among those subjects, only 13 had usable ECG signals so we ran our ECG study on these 13 subjects. In the experiment, mental workload variations have been induced using N-back tasks. With such tasks, users saw a sequence of letters on screen, the letters being displayed one by one, every 2 seconds. For each letter the user had to indicate with a mouse click whether the displayed letter was the same one as the letter displayed N letters before. Users alternated between easy blocks (“low” workload label) with the 0-back task (the user had to identify whether the current letter was the letter ‘X’) and difficult blocks (“high” workload label) with the 2-back task (the user had to identify whether the current letter was the same letter as the one displayed 2 letters before) to reach a total of 24 blocks. Each block of 120 seconds was labeled with its workload level: “low” workload for 0-back trials and “high” workload for 2-back trials. Each block of 120 seconds was then split later on using 6, 10, 20, 30, 40, 60 or 120-second long time windows, resulting in the creation of 7 different data sets for our study. This splitting aimed at assessing the time window length influence on mental workload estimation quality from ECG signals. Signals from two ECG electrodes, 28 EEG electrodes, two facial ElectroMyoGram (EMG), four ElectroOculoGram (EOG), breath belt, pulse and a galvanic skin response sensor have been collected, all with a g.USBAmp (g.tec, Austria) and associated sensors. For our study, we kept ECG signals only, and assessed the performances of both classification algorithms (the proposed Riemannian TSC and the baseline euclidean SVMs) using a within-participant study with five-fold stratified Cross-Validation. This means that the data from each participant was divided into five parts: four parts were used for training the classifier and the fifth one for testing the resulting classifier for that participant. This process was repeated five times, with each part used exactly once as the testing set.

### 2.2 Feature extraction

In order to estimate workload levels from ECG signals, we investigated two different machine learning algorithms, i.e., an Euclidean SVM and a RGC-TSC, both using the same ECG features, i.e, time, frequency and nonlinear domain-based features

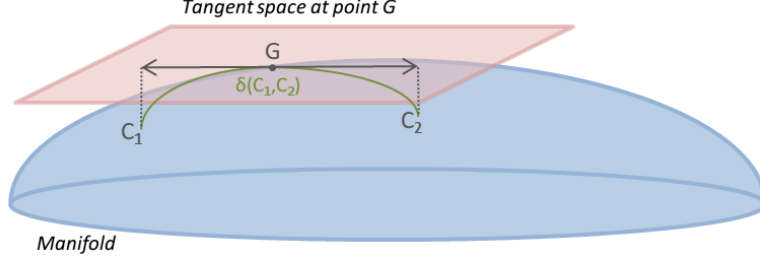
[36], represented classically as a feature vector for the Euclidean SVM, or as a newly proposed SPD matrix for the RGC-TSC. We thus first extracted the following 34 ECG features using neurokit [37]:

- *sdRR*: the standard deviation of the RRI [14].
- *meanRR*: the mean of the RRI [38].
- *RMSSD*: the Root Mean Square of the RRI [14].
- *CVSD*: the Coefficient of Variation of Successive Differences. This corresponds to the RMSSD divided by meanRR [39].
- *cvRR*: the RR coefficient of variation. This corresponds to the sdRR divided by the meanRR [39].
- *medianRR*: the median of the absolute values of the RRI's successive differences [38].
- *madRR*: the RRI's median absolute deviation [39].
- *mcvRR*: the RRI's median-based coefficient of variation. This corresponds to the ratio of madRR divided by medianRR [38].
- *RR50 or RR20*: the successive RRI's number of interval differences greater than 50ms or 20 ms, respectively [39].
- *pRR50 or pRR20*: the proportion derived by dividing RR50 (ou RR20) by the number of RRI [38].
- *triang*: the HRV triangular index measurement, i.e., plotting the integral of the ratio of RRI density histogram by its height [40].
- *Shannon\_h*: the Shannon entropy calculated from the class probabilities of the RRI density distribution [38].
- *VLF*: the HRV variance in the Very Low Frequency (0.003 to 0.04 Hz) [39].
- *LF*: the HRV variance in the Low Frequency (0.04 to 0.15 Hz) [38, 39].
- *HF*: the HRV variance in the High Frequency (0.15 to 0.40 Hz) [38, 39].
- *Total\_Power*: the total power of the density spectra [38].
- *LFHF*: the LF/HF ratio [38, 39].
- *LFn*: the normalized LF power. It can be calculated using the equation "LFn = LF/(LF+HF)" [39].
- *HFn*: the normalized HF power. It can be calculated using the equation "HFn = HF/(LF+HF)" [39].
- *LFp*: the LF/Total\_Power ratio [38].
- *HFp*: the HF/Total\_Power ratio [38].
- *DFA*: the Detrended Fluctuation Analysis (DFA) [41] of the heart rate raw signals.
- *Shannon*: the RRI's Shannon entropy [38].
- *sample\_entropy*: the RRI's sample entropy [42].
- *correlation\_Dimension*: RRI's correlation dimension [38].
- *entropy\_Multiscale*: the RRI's entropy multiscale [42].
- *entropy\_SVD*: the RRI's Singular Value Decomposition (SVD) entropy [38].
- *entropy\_Spectral\_VLF*: the RRI's Spectral Entropy (SE) over the VLF [38].
- *entropy\_Spectral\_LF*: the RRI's SE over the LF [38].
- *entropy\_Spectral\_HF*: the RRI's SE over the HF [38].
- *Fisher\_Info*: the RRI's Fisher information [41].
- *Lyapunov*: the RRI's Lyapunov exponent [41].

- *FD\_Petrosian*: RRI’s Petrosian’s Fractal Dimension [41].
- *FD\_Higushi*: the RRI Higushi’s Fractal Dimension [41].

### 2.3 Machine Learning algorithms

As mentioned earlier, we compared two ECG classification methods: a classical Euclidean SVM and the proposed Riemannian Geometry-based TSC. We describe how they are used below.



**Fig. 1** Schematic representation of a Riemannian manifold with matrix  $G$ , the Riemannian average of covariance matrices  $C_1$  and  $C_2$ . The tangent space to the Riemannian manifold at point  $G$  is represented in red.

#### 2.3.1 Euclidean Support Vector Machine

Among the 34 features described above, the ten most relevant ones for classification were selected using the minimal Redundancy Maximal Relevance (mRMR) feature selection algorithm [43], arranged as a vector and used to train and test a standard Euclidean SVM classifier. Note that we ran the SVM with two types of kernels, i.e., a linear kernel and a Radial Basis Function (RBF) kernel. We refer to them as  $SVM_{linear}$  and  $SVM_{RBF}$ , respectively. We used default hyperparameters ( $C = 1.0$ ) from scikit-learn [44].

#### 2.3.2 Riemannian Geometry Classifiers

As mentioned, RGCs proved to be very effective for EEG-based classification of mental states, reaching state-of-the-art performances [3, 30, 31, 33], but, to the best of our knowledge, have never been used for the classification of ECG signals yet. In this section, we first introduce the general concepts of RGCs as it is established for EEG classification, before presenting how we adapted a particular RGC, the TSC, to classify ECG.

##### *Riemannian Geometry general concepts*

In EEG Riemannian approaches, EEG trials are represented as covariance matrices - more specifically as symmetric positive definite (SPD) matrices - these matrices being manipulated with an appropriate geometry, the Riemannian geometry [30, 31]. With such a geometry, the *intrinsic* non-Euclidean distance between two SPD matrices, i.e.



the distance between two points (here  $\mathbf{C}_1$  and  $\mathbf{C}_2$ ) on the Riemannian *manifold*, can be estimated using the Riemannian distance  $\delta^2$ :

$$\delta^2(\mathbf{C}_1, \mathbf{C}_2) = \|\log(C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}})\|_F = \left(\sum_{i=1}^n \log^2 \lambda_i\right)^{1/2} \quad (1)$$

where  $\lambda_i$  are the eigenvalues of  $C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}}$  and  $\|\cdot\|_F$  the Frobenius norm. On the *manifold*, the set of tangent vectors to point  $\mathbf{G}$  defines the manifold tangent space at  $\mathbf{G}$ . Figure 1 shows the tangent space at point  $\mathbf{G}$ , which is the centroid (mean) of  $\mathbf{C}_1$  and  $\mathbf{C}_2$ . More generally, any SPD matrix  $\mathbf{C}_i$  can be projected onto the tangent space at point  $\mathbf{G}$  using:

$$\mathbf{S}_i = \text{Log}_{\mathbf{G}}(\mathbf{C}_i) = \mathbf{G}^{1/2} \text{logm}(\mathbf{G}^{-1/2} \mathbf{C}_i \mathbf{G}^{-1/2}) \mathbf{G}^{1/2} \quad (2)$$

$\mathbf{S}_i$  being the projection of  $\mathbf{C}_i$  onto the tangent plane, and  $\text{logm}(\cdot)$  denoting the logarithm of a matrix.

Multiple EEG RGCs are available, notably the Minimum Distance to the Mean with geodesic filtering (FgMDM) and the TSC classifiers [3, 30, 31]. In this article, we focus on the TSC as it proved to be robust and effective across multiple EEG classification tasks, including motor imagery, cognitive workload and affective state EEG-based classification [41].

### ***Tangent Space Classifier***

TSC projects SPD matrices  $\mathbf{C}_i$  onto the tangent space at point  $\mathbf{G}$  (the training data mean) using Eq. (2), to obtain matrices  $\mathbf{S}_i$ . Then, it uses any classifier such as LDA or SVM on the vectorized upper-triangular elements of the projected matrices [31]. In this paper, we used the exact same SVMs as for the classical ECG SVM classifiers (Section 2.3.1), with the same hyperparameters, for a fair comparison.

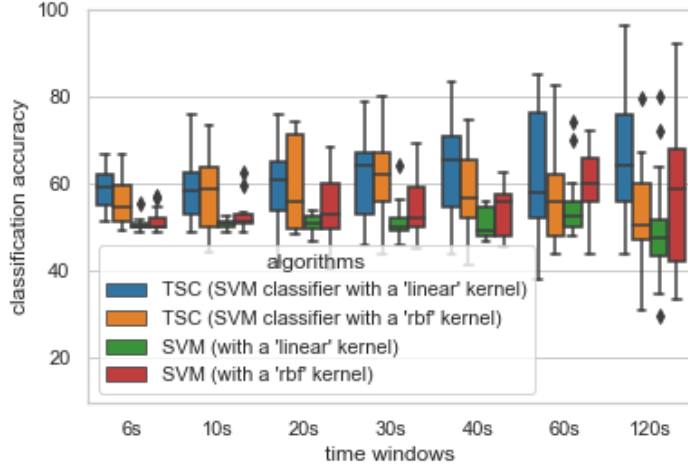
### ***TSC for ECG signals classification***

In our study, we revisited this EEG Riemannian classifier approach, i.e., the TSC, and adapted it to ECG signals. To do so, we first followed a two-step approach in order to represent ECG features as an SPD Matrix. First, we bandpass filtered the ECG into the 4 frequency bands described in Section 2.2, i.e., VLF (0.003 to 0.04 Hz), LF (0.04 to 0.15 Hz), HF (0.15 to 0.40 Hz), as well as the total power (TP, 0.003 to 40 Hz). We then computed the covariance matrix of these four frequency-specific ECG signals, leading to a  $4 \times 4$  covariance matrix. Second, we extended this  $4 \times 4$  matrix using 0s in order to obtain a  $10 \times 10$  matrix, the number 10 corresponding to the 4 frequency bands above in addition to the 6 most relevant features (selected using mRMR as for the classical Euclidean SVM) among the other features described above (see Section 2.2), i.e., the time, frequency and nonlinear domain-based features. We then added the values of the extracted features (except the ones already represented as covariance matrix, i.e., VLF, LF, HF and TP) onto the diagonal (from the 5th to the 10th diagonal element) of this  $10 \times 10$  matrix. This lead to a SPD matrix using 10

feature types, i.e., as many as the classical Euclidean SVM pipeline above. We finally use this new type of SPD matrix to represent ECG signals with the TSC classifier.

### 3 Results

The average classification accuracies obtained with each classifier and time window to classify low versus high mental workload from ECG are reported in Figures 1 and 2. As a reference, the statistical chance levels using [45] were estimated at 51.04% (480 trials), 51.33% (290 trials), 51.88% (145 trials), 52.31% (100 trials), 52.62% (75 trials), 53.23% (50 trials), as well as 54.46% (25 trials) for respectively the 6, 10, 20, 30, 40, 60 and 120-second long time windows, all for 13 subjects. We used a 2-way ANOVA with repeated measures to evaluate the performances of factor *time windows* according to factor *algorithms* ( $TSC_{SVM_{linear}}$  vs  $TSC_{SVM_{RBF}}$  vs  $SVM_{linear}$  vs  $SVM_{RBF}$ ). Note that we checked data sphericity and normality, and used Greenhouse-Geisser (GG) correction if needed.



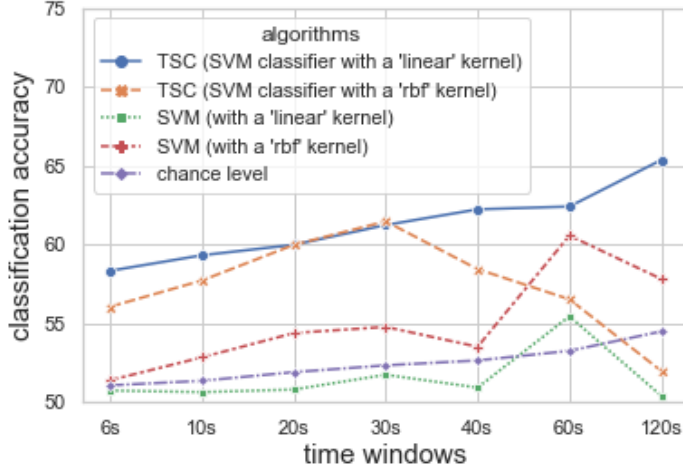
**Fig. 2** Classification accuracy for each time window and algorithm, i.e., (Euclidean) SVM with both 'linear' and 'RBF' kernels and (Riemannian) TSC with an SVM classifier using both 'linear' and 'RBF' kernels.

The ANOVA revealed a main effect of *algorithms* [ $GG(1,13)=11.17$ ,  $p=0.0008$ ] and for *algorithms\*time windows* [ $F(1,13)=1.776$ ,  $p=0.029$ ], but not for *time windows* [ $GG(1,13)=1.115$ ,  $p=0.350$ ]. Post-hoc analyses results using Student t-test with False Discovery Rate (FDR) corrections showed significant differences between  $TSC_{SVM_{linear}}$ ,  $TSC_{SVM_{RBF}}$ ,  $SVM_{linear}$  and  $SVM_{RBF}$  when using different times windows lengths. Here, we focus on the differences observed between the  $TSC_{SVM_{linear}}$  and  $SVM_{linear}$ , as well as the differences observed between the  $TSC_{SVM_{RBF}}$  and the  $SVM_{RBF}$ . Interestingly enough, the Riemannian  $TSC_{SVM_{linear}}$  outperformed

	TSC (SVM classifier with a 'linear' kernel)	TSC (SVM classifier with a 'rbf' kernel)	SVM (with a 'linear' kernel)	SVM (with a 'rbf' kernel)	chance level
<b>6s</b>	58.3 (5.0)	56.0 (5.4)	50.7 (1.6)	51.4 (2.8)	51.04
<b>10s</b>	59.3 (8.4)	57.7 (8.7)	50.6 (1.2)	52.8 (3.9)	51.33
<b>20s</b>	60.0 (10.5)	60.0 (11.0)	50.8 (2.0)	54.4 (7.1)	51.88
<b>30s</b>	61.2 (10.5)	61.4 (11.4)	51.7 (4.7)	54.8 (7.9)	52.31
<b>40s</b>	62.2 (11.7)	58.4 (10.1)	50.9 (3.4)	53.5 (6.3)	52.62
<b>60s</b>	62.4 (14.7)	56.5 (10.7)	55.4 (8.2)	60.5 (9.2)	53.23
<b>120s</b>	65.3 (15.6)	52.0 (13.9)	50.4 (14.2)	57.8 (19.5)	54.46

**Table 1** Mean classification accuracy (and standard deviation) for each algorithm, i.e., Euclidean Support Vector Machine (SVM) with both 'linear' and 'RBF' kernels and Riemannian Tangent Space Classifier (TSC) with a SVM classifier using both 'linear' and 'RBF' kernels, and for each time window length.

the Euclidean  $SVM_{linear}$  for the 6s [ $p \leq 0.05$ ], 10s [ $p \leq 0.05$ ], 20s [ $p \leq 0.05$ ], 30s [ $p \leq 0.05$ ] and 40s [ $p \leq 0.05$ ] time windows. The same post-hoc analyses showed that the Riemannian  $TSC_{SVM_{RBF}}$  also outperformed the Euclidean  $SVM_{RBF}$  for the 6s [ $p \leq 0.05$ ] time window. Finally, the comparison between the  $TSC_{SVM_{linear}}$  and the  $TSC_{SVM_{RBF}}$  showed a significant difference for the 120s [ $p \leq 0.05$ ] time window. The  $TSC_{SVM_{linear}}$  outperformed chance level on all time windows.



**Fig. 3** Classification accuracy for each time window and algorithm, i.e., Euclidean SVM with both 'linear' and 'RBF' kernels and Riemannian TSC with an SVM classifier using both 'linear' and 'RBF' kernels.

## 4 Discussion and Conclusion

In this paper, we studied an effective EEG classifier, the Riemannian Geometry TSC, and adapted it for classifying ECG, here to estimate different levels of cognitive workload. We notably proposed a new representation of ECG as an SPD matrix. We compared the results obtained by the proposed ECG-TSC to those of a widely used state-of-the-art ECG classifier, i.e., an Euclidean SVM, based on the same ECG features. We assessed such algorithms for multiple time window lengths, on an ECG data set from [35], where 0 and 2-back tasks have been used to induce low and high levels of workload.

First, the (Riemannian)  $TSC_{SVM_{linear}}$  obtained better performances than the (Euclidean)  $SVM_{linear}$  for all time windows, and even significantly outperformed this  $SVM_{linear}$  as well as the chance level when using 6, 10, 20, 30 and 40-second long time windows, with respectively 58.3%, 59.3%, 60.0%, 61.2% and 62.2% classification accuracy versus 50.7%, 50.6%, 50.8%, 51.7% and 50.9% for the  $SVM_{linear}$ .

The  $TSC_{SVMRBF}$  also outperformed the  $SVM_{RBF}$  with 56.0% classification accuracy versus 51.4%.

All together, these results highlight the potential of TSC for classifying ECG signals, here for estimating cognitive workload, which could prove useful for ergonomics, HCI or medical diagnostic among others. Another relevant result about the TSC itself is that, among the two types of studied kernels for the TSC-SVM, the 'linear' one obtained the best scores on most time windows, and significantly outperformed the 'RBF' kernel for the 120s time window (65.3% vs 52.0%). Finally, it is worth noting that the Riemannian TSC was able to train ECG classifiers from only a few dozens (with 120s or 60s long time windows) or a few hundreds (for the other time window lengths) training ECG trials, a typical number of training trials for ECG applications. Such a number of trials would be too small for training data hungry classification algorithms such as Deep Neural Networks.

Such results enable us to suggest guidelines about which algorithm to use for cognitive workload classification from ECG. First, the Riemannian Geometry TSC, that showed interesting results for EEG classification, also showed better classification accuracies than one of the state-of-the-art algorithm (SVM) for ECG classification. This suggest that this algorithm should be preferred to vector-based Euclidean linear classifiers such as the SVM in the future. Moreover, it be interesting to test the TSC for classifying other physiological signals such as electrodermal activity or breathing. Finally, we used here the TSC for classifying cognitive workload levels, but EEG-RGCs showed promising results for classifying other mental states such as epistemic curiosity or emotions [3, 46]. It would thus be interesting to study the use of TSC to classify epistemic curiosity or emotion levels from ECG as well. Altogether our results show that Riemannian geometry classifiers appear to be very promising for ECG signal classification as well, as such, should be explored deeper for that field as well.

## Acknowledgment

This work was supported by the European Research Council (grant ERC-2016-STG-714567)

## References

- [1] Parasuraman, R., Wilson, G.F.: Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors* **50**(3), 468–474 (2008)
- [2] Hewett, T., Baecker, R., Card, S., Gasen, J., Tremaine, M., Perlman, G., Strong, G., Verplank, W.: *Acm sigchi curricula for human-computer interaction*. WWW document (1992)
- [3] Appriou, A., Cichocki, A., Lotte, F.: Modern machine learning algorithms to classify cognitive and affective states from electroencephalography signals. *IEEE SMC Magazine*, 1–8 (2020)

- [4] Zander, T.O., Kothe, C.: Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J Neur Eng* **8** (2011)
- [5] Hart, S.: Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **50** (2006) <https://doi.org/10.1177/154193120605000909>
- [6] Fairclough, S.: Fundamentals of physiological computing. *Interacting with Computers* **21**, 133–145 (2009) <https://doi.org/10.1016/j.intcom.2008.10.011>
- [7] Fairclough, S., Houston, K.: A metabolic measure of mental effort. *Biological psychology* **66**, 177–90 (2004) <https://doi.org/10.1016/j.biopsycho.2003.10.001>
- [8] Hasnul, M., Ab.Aziz, N., Alelyani, S., Mohana, M., Abd Aziz, A.: Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors* **21**, 5015 (2021) <https://doi.org/10.3390/s21155015>
- [9] Sweller, J., Van Merriënboer, J.J.G., Paas, F.: Cognitive architecture and instructional design. *Educ. Psych. Rev.* **10**, 251 (1998) <https://doi.org/10.1023/a:1022193728205>
- [10] O'DONNELL, R.D.: Workload assessment methodology. cognitive processes and performance (1986)
- [11] Nourbakhsh, N., Chen, F., Wang, Y., Calvo, R.: Detecting users' cognitive load by galvanic skin response with affective interference. *ACM Transactions on Interactive Intelligent Systems* **7**, 1–20 (2017) <https://doi.org/10.1145/2960413>
- [12] Marquart, G., Cabrall, C., Winter, J.: Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing* (2015) <https://doi.org/10.1016/j.promfg.2015.07.783>
- [13] Charles, R., Nixon, J.: Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics* (2018) <https://doi.org/10.1016/j.apergo.2018.08.028>
- [14] Smith, A.-L., Owen, H., Reynolds, K.: Heart rate variability indices for very short-term (30 beat) analysis. part 1: survey and toolbox. *J Clin Monit Comput* **15** (2013)
- [15] Castaldo, R., Montesinos, L., Wan, T., Serban, A., Massaro, S., Pecchia, L.: Heart rate variability analysis and performance during a repeated mental workload task. (2017). [https://doi.org/10.1007/978-981-10-5122-7\\_18](https://doi.org/10.1007/978-981-10-5122-7_18)
- [16] Solovey, E., Zec, M., Garcia Perez, E., Reimer, B., Mehler, B.: Classifying driver

- workload using physiological and driving performance data: Two field studies. *Conference on Human Factors in Computing Systems - Proceedings* (2014) <https://doi.org/10.1145/2556288.2557068>
- [17] Shao, S., Wang, T., Wang, Y., Su, Y., Song, C., Yao, C.: Research of hrv as a measure of mental workload in human and dual-arm robot interaction. *Electronics* **9**, 2174 (2020) <https://doi.org/10.3390/electronics9122174>
- [18] Qu, H., Gao, X., Pang, L.: Classification of mental workload based on multiple features of ecg signals. *Informatics in Medicine Unlocked* **24**, 100575 (2021) <https://doi.org/10.1016/j.imu.2021.100575>
- [19] Hogervorst, M., Brouwer, A.-M., Erp, J.: Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience* **8**, 322 (2014) <https://doi.org/10.3389/fnins.2014.00322>
- [20] Haapalainen Ferreira, E., Kim, S., Forlizzi, J., Dey, A.: Psycho-physiological measures for assessing cognitive load, pp. 301–310 (2010). <https://doi.org/10.1145/1864349.1864395>
- [21] Stanton, N.: Human cognitive abilities: A survey of factor-analytic studies, by j. b. carroll. *Ergonomics* **38** (2007) <https://doi.org/10.1080/00140139508925174>
- [22] Zhang, H., Zhu, Y., Jayachandran, M., Guan, C.: Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine, vol. 2014 (2014). <https://doi.org/10.1109/EMBC.2014.6944250>
- [23] Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J., Park, S.: Classification of a driver’s cognitive workload levels using artificial neural network on ecg signals. *Applied Ergonomics* (2017) <https://doi.org/10.1016/j.apergo.2016.09.013>
- [24] Peysakhovich, V., Vachon, F., Dehais, F.: The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology* **112**, 40–45 (2017) <https://doi.org/10.1016/j.ijpsycho.2016.12.003>
- [25] Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., Causse, M.: Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology* **121** (2016) <https://doi.org/10.1016/j.biopsycho.2016.10.002>
- [26] Parent, M., Peysakhovich, V., Mandrick, K., Tremblay, S., Causse, M.: The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ecg and fnirs? *International Journal of Psychophysiology* **146** (2019) <https://doi.org/10.1016/j.ijpsycho.2019.09.005>

- [27] Qu, H., Shan, Y., Liu, Y., Pang, L., Fan, Z., Zhang, J., Wanyan, X.: Mental workload classification method based on eeg independent component features. *Applied Sciences* **10**, 3036 (2020) <https://doi.org/10.3390/app10093036>
- [28] Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Abou Khaled, O., Mugellini, E., Widmer, M., Sonderegger, A.: Classification of drivers' workload using physiological signals in conditional automation. *Frontiers in Psychology* **12**, 596038 (2021) <https://doi.org/10.3389/fpsyg.2021.596038>
- [29] Siegenthaler, E., Costela, F., McCamy, M., Di Stasi, L., Otero-Millan, J., Sonderegger, A., Groner, R., Macknik, S., Martinez-Conde, S.: Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *The European journal of neuroscience* **39**, 287–94 (2014) <https://doi.org/10.1111/ejn.12395>
- [30] Yger, F., Berar, M., Lotte, F.: Riemannian approaches in Brain-Computer Interfaces: a review. *IEEE TNSRE* **25**(10) (2016)
- [31] Barachant, A., Bonnet, S., Congedo, M., Jutten, C.: Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering* **59**(4), 920–928 (2012)
- [32] Congedo, M., Barachant, A., Bhatia, R.: Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces* **4**(3), 155–174 (2017)
- [33] Roy, R.N., Hinss, M.F., Darmet, L., Ladouce, S., Jahanpour, E.S., Somon, B., Xu, X., Drougard, N., Dehais, F., Lotte, F.: Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics*, 4
- [34] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for eeg-based brain-computer interfaces: a 10 year update. *Journal of neural engineering* **15**(3), 031005 (2018)
- [35] Mühl, C., Jeunet, C., Lotte, F.: EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience* **8**, 1–15 (2014)
- [36] Pham, T., Lau, Z.J., Chen, S., Makowski, D.: Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors* **21** (2021) <https://doi.org/10.3390/s21123998>
- [37] Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.H.A.: Neurokit2: A python toolbox for neurophysiological signal processing. *BRM* (2021) <https://doi.org/10.3758/s13428-020-01516-y>
- [38] Voss, A., Schroeder, R., Heitmann, A., Peters, A., Perz, S.: Short-term heart rate variability - Influence of gender and age in healthy subjects. *PLoS ONE* **10**(3),



1–33 (2015) <https://doi.org/10.1371/journal.pone.0118308>

- [39] Malik, M.: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **93**, 1043–1065 (1996)
- [40] Shaffer, ., Ginsberg, J.P.: An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in public health* **5**, 258 (2017) <https://doi.org/10.3389/fpubh.2017.00258>
- [41] Appriou, A., Pillette, L., Trocellier, D., Dutartre, D., Cichocki, A., Lotte, F.: Biopyc, an open-source python toolbox for offline electroencephalographic and physiological signals classification. *Sensors (Basel, Switzerland)* **21** (2021) <https://doi.org/10.3390/s21175740>
- [42] Tiwari, A., Albuquerque, I., Parent, M., Gagnon, J.F., Lafond, D., Tremblay, S., Falk, T.H.: Multi-scale heart beat entropy measures for mental workload assessment of ambulant users. *Entropy* **21**(8), 1–20 (2019) <https://doi.org/10.3390/e21080783>
- [43] Peng, H., Long, F., Ding, C.: Feature Selection Based On Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* **27** (2005)
- [44] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Dubourg, V.: Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12** (2011) <https://doi.org/https://dl.acm.org/citation.cfm?id=2078195>
- [45] Combrisson, E., Jerbi, K.: Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods* **250**, 126–136 (2015)
- [46] Appriou, A., Ceha, J., Pramij, S., Dutartre, D., Law, E., Oudeyer, P.-Y., Lotte, F.: Towards measuring states of epistemic curiosity through electroencephalographic signals. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 4006–4011 (2020). IEEE