Expert Systems With Applications 143 (2020) 112988

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Simultaneous feature selection and heterogeneity control for SVM classification: An application to mental workload assessment

Sebastián Maldonado[a,b,*], Julio López[c], Angel Jimenez-Molina[b,d], Hernán Lira[e]

[a] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes. Mons. Álvaro del Portillo 12455, Las Condes, Santiago, Chile
[b] Instituto Sistemas Complejos de Ingeniería (ISCI), Chile
[c] Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Ejército 441, Santiago, Chile
[d] Department of Industrial Engineering, University of Chile, Beauchef 851, Santiago, Chile
[e] School of Computing, KAIST, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

ARTICLE INFO

Article history:
Received 3 September 2018
Revised 4 September 2019
Accepted 26 September 2019
Available online 27 September 2019

Keywords:
Support vector machines
Feature selection
Heterogeneity control
Mental workload
Group penalty functions

ABSTRACT

In this study, an expert system is presented for analyzing the mental workload of interacting with a mobile phone while facing common daily tasks. Psychophysiological signals were collected from various devices, each characterized by a different cost and obtrusiveness. To deal with user-level signal data, a support vector machine-based feature selection approach is proposed. Given the limited person-level information available, our goal was to construct robust models by pooling population-level information across users (as a heterogeneity control). A single optimization problem that combines four objectives is proposed: model, margin maximization, feature selection, and heterogeneity control. The costs of using the devices were estimated, leading to a decision tool that allowed experiment designers to evaluate the marginal benefit of using a given device in terms of performance and its cost.

1. Introduction

People use electronic devices such as smartphones and tablets while performing daily routines, including talking, walking or listening to a lecture. Multitasking eventually leads to an increase in the person's productivity and quality of task performance, but currently, there is ample evidence that it might overwhelm users by increasing their mental workload (Jiménez-Molina & Ko., 2015; Lee et al., 2014; Okoshi et al., 2015; Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005; Samaha & Hawi, 2016; Smith & Dulay, 2014; Van Deursen, Bolle, Hegner, & Kommers, 2015; Wang, Wang, Gaskin, & Wang, 2015). The term "mental workload" refers to the amount of perceived mental effort induced by a particular task (Wickens, 2002). Therefore, having an intelligent system capable of continually assessing the mental workload in such conditions would permit, among other benefits, knowing the moments in which an interaction is possible and convenient and will not saturate the person's cognitive capacity.

Traditional approaches to assessing human mental workload have focused almost exclusively on subjective methods (Matthews, Reinerman-Jones, Barber, & Abich, 2015; Young, Brookhuis, Wickens, & Hancock, 2015), such as surveys and autoperception scales (Albers, 2011; Galy, Cariou, & Mélan, 2012). The most widespread example of this method is the NASA Task Load Index, which measures mental and physical performance as well as the effort and frustration of the user (Hart, 1988). These methods are limited by a reporting bias because they are applied after the person was engaged in the task. Additionally, real-time assessment of mental workload is not performed in these methods. Another approach to measuring mental workload consists of performance-based methods. Although the performance measures are objective and it is possible to measure them in real time, it has been shown that in certain cases they may not reflect the subtle changes in cognitive load during execution of the task. This approach has been recommended for tasks that induce changes in performance that are sufficient to be observed (Chi & Lin, 1997; Galy et al., 2012; Paas & Van Merriënboer, 1993).

Several expert systems have recently been proposed for assessing cognitive human behavior by means of psychophysiological responses (Bailey & Iqbal, 2008; Chen & Epps, 2013; Haapalainen, Kim, Forlizzi, & Dey, 2010; Ikehara & Crosby, 2005; Jiménez-Molina, Retamal, & Lira, 2018; Yoshida, Ohwada, Mizoguchi, & Iwasaki, 2014). For instance, peripheral psychophysiological signals captured from progressively smaller sensors, including an oximeter, electroencephalogram, electrodermal activity meter, thermometer, and pupillograph, among others, have been used to classify discrete

* Corresponding author.
E-mail addresses: smaldonado@uandes.cl (S. Maldonado), julio.lopez@udp.cl (J. López), ajimenez@dii.uchile.cl (A. Jimenez-Molina), lirahernan@kaist.ac.kr (H. Lira).

https://doi.org/10.1016/j.eswa.2019.112988
0957-4174/

levels of human mental workload in various settings (Haapalainen et al., 2010; Jiménez-Molina et al., 2018; Lo, Sehic, & Meijer, 2017). The obtained information consists of synchronous multi-device datasets. Moreover, several signals can be gathered from each device, and multiple features can be extracted from each signal, while each device has its own cost and obtrusiveness for the person. The advantage of using psychophysiological information is that it allows measuring objective indicators of mental workload because of the empirically demonstrated correlation between the reactions of the nervous system and psychological stimuli (Cacioppo, Tassinary, & Berntson, 2007).

Learning effective machine learning models to analyze these multidevice datasets is challenging. In fact, multiple objectives need to be integrated in the models. First, where possible, high-cost devices have to be avoided. Second, it is undesirable to use devices with high obtrusiveness for the person. Both would probably lead to losing features that could otherwise contribute to the prediction. In addition, since these studies of human behavior have a qualitative interest in the specific features that govern cognitive phenomena, it is important that the models incorporate feature selection during the classification task. All of these objectives must be balanced in personalized models that achieve a reasonable level of predictive performance.

Learning effective machine learning models to analyze these multidevice datasets is also challenging if multiple objectives are pursued. In our case, we seek accurate models while avoiding devices that are expensive and/or very obtrusive for the person. This can be achieved by incorporating feature selection during the classification task. The advantage of selecting the relevant sources of information is in fact twofold, since such studies of human behavior have a qualitative interest in the specific features that govern cognitive phenomena (Jiménez-Molina et al., 2018; Lo et al., 2017).

Some of the above challenges can be effectively approached using support vector machines (SVMs) (Cortes & Vapnik, 1995), which are well-known machine learning models with appealing advantages, such as providing a single optimum obtained via convex optimization, being flexible in allowing inclusion of additional objectives (Maldonado, Pérez, & Bravo, 2017), and offering a superior predictive performance (Cortes & Vapnik, 1995). Nevertheless, one issue that standard SVMs cannot handle is feature selection (Bradley & Mangasarian, 1998). An adequate selection allows a better interpretation of the process that generated the data, providing insights into human cognitive behavior (Guyon, Gunn, Nikravesh, & Zadeh, 2006). Additionally, using less information leads to a reduction in data collection costs (Guyon et al., 2006; Maldonado, Pérez et al., 2017). This last advantage is very important in the case study presented in this paper, since expensive devices, such as an eye tracker, are used to gather psychophysiological signals from the participants, and these are data used to determine if the task they are performing induces a *low* or *high* mental workload.

In this paper, a novel approach is proposed for simultaneous SVM classification and feature selection. Our contribution is twofold. The first is methodological: we propose a novel SVM approach for binary classification in which data are collected at an individual level, but not enough information is available for calibrating individual models properly. We propose a strategy in which all individual classifiers are constructed simultaneously, and information is pooled across individuals to improve predictive performance. This problem is related to panel data, which consist of observations obtained over multiple time periods for the same individuals. This problem has often been analyzed using econometric models, but few machine learning solutions have been proposed in the literature. Furthermore, our proposal performs group-level feature selection. Assuming that attributes stem from multiple devices, each generating several signals, we penalize the use of features at a group level using the infinity norm. If the model selects one attribute from a given group, additional attributes from the same source can be included at zero cost.

The second contribution is applied: we apply our proposal to a real-world problem of modeling mental workload based on psychophysiological signals, developing an expert system for analyzing mobile phone interactions and daily common tasks performed simultaneously. We assumed that a mobile phone task would be more demanding cognitively in the presence of interference, such as answering the experimenter's questions verbally. The psychophysiological signals were collected from four devices with different costs: an ECG, an eye tracker, a thermometer, and a pulse oximeter. In this study, we estimated the costs of using various devices from two perspectives: (1) the monetary cost of the device, and (2) the obtrusiveness for the participant in the experiments. These costs were incorporated explicitly in the modeling process.

The main goal of our expert system is to develop Pareto curves for using different devices, showing their respective predictive performance in terms of AUC and signal collection costs. This is a very useful tool for decision making, since it allows experiment designers not only to estimate the expected performance of a multidevice experiment for mental workload assessment but also to understand the trade-off between the cost of using a given device and its marginal benefit in terms of performance.

The remainder of this paper is structured as follows. In Section 2, background and previous studies of mental workload assessment are discussed, and developments in feature selection and SVM classification that are relevant to this study are presented. The proposed SVM approach is described in Section 3. In Section 4, experimental results for the case study are given. Finally, Section 5 provides the main conclusions of this study and discusses opportunities for future work.

2. Background and literature review

In this section, we first provide the theoretical basis and briefly discuss the background for the assessment of mental workload. Then, related work on mental workload classification based on psychophysiological signals is presented. Next, the classical SVM formulation for binary classification (Cortes & Vapnik, 1995) is described. Finally, related work on feature selection for SVM classification is discussed.

2.1. Mental workload background

Cognitive resources are assets used by humans to think, remember, make decisions, solve problems and coordinate movements (Wickens, 2002). Examples of cognitive resources include perception, attention, short- and long-term memory, and motor control, among others. According to Navon et al. (Navon & Gopher, 1979), there is a limited amount of these underlying resources in the human learning and information processing system. In addition, mental workload demands different amounts of these resources depending on the tasks in which the person is involved at the same time.

Multiple theories of human mental workload have been developed. One of them was proposed by John Sweller in the educational domain and states that a high mental workload results in additional demand for cognitive resources, which in turn decreases performance and processing efficiency, making learning more difficult. This theory distinguishes three categories or types of mental workload to determine which type is unnecessary. One of them is intrinsic mental workload, related to the complexity of the task; another is extraneous mental workload, related to the situation or context and to external factors such as time pressure and task organization; finally, germane mental workload is related to learning and the schematization of tasks, and is a desirable mental workload needed to automate the performance of similar tasks (Paas, Renkl, & Sweller, 2003; Sweller, 1988; 1994; Sweller, van Merrienboer, & Paas, 1998).

In general, excessive demand for cognitive resources can lead to distractions, increase errors, provoke stress and frustration, and reduce ability for mental planning, problem solving, and decision making. This can provoke a state of saturation, called cognitive resource depletion (Wickens, 2002). This overload implies that the brain cannot easily process new information, resulting in processing and execution errors. An example is the distraction caused by the arrival of notifications to the smartphone while the person is engaged in a different task. In this case, the person needs to divide his/her attention and expend a portion of his/her cognitive resources on the new stimulus.

A theory of human mental workload of special interest for this paper is Wickens's multiple resources model (MRM) (Wickens, 2002), which presents a comprehensive and quantitative method for the assessment of mental workload. This model shows empirical evidence of three cognitive dimensions that, depending on the type of tasks, may cause competition and interference among cognitive resources if they are simultaneously used. The first dimension is composed of three stages that are responsible for processing task information, known as the perceptual, central processing and response stages. The second dimension is composed of the cognitive resources that are demanded for each stage, including selective attention, perception, working memory and motor control. The third dimension contains the attributes that characterize cognitive resources and the way they are being used, such as input modality—visual, auditory or tactile—and processing code—spatial or verbal. The interference among attributes of simultaneously demanded resources increases the mental workload.

Three types of interference have been described (Wickens, 2002): by input modality, by processing stage and by processing code. The first occurs when two or more tasks use the same input modality. In this case, the interference is complete, which might cause serious errors in the execution of the task. The second type of interference occurs when two or more tasks use resources from the same stage. It can be complete or partial. In this case, it is possible to perform the tasks, but the performance is likely to be worse than under normal conditions. Finally, the third type of interference occurs when there is a conflict in the spatial or verbal processing code.

Oulasvirta et al. (2005) present evidence of competition due to the use of the cognitive resource of attention in smartphone interactions and daily tasks. The researchers propose a framework based on Wickens's MRM to evaluate competition and interference among cognitive resources. We use this approach to tag the levels of mental workload of combinations of smartphone interactions and daily tasks performed by the person.

2.2. Assessment of mental workload using psychophysiological signals

In recent years, several studies have addressed the assessment of mental workload using peripheral psychophysiological signals under diverse experimental conditions in which tasks with diverse levels of complexity had to be performed.

Bailey and Iqbal (2008) study the effect of interruptions on the performance of a person on particular tasks. The researchers observe that interruptions impact performance, but this impact is lower if the interruption occurs during a low mental workload period. Additionally, they assess mental workload by performing an experiment with three tasks and using pupil size (PS) as a psychophysiological measure. The study concludes that mental workload varies during the execution of tasks and that this variation is related to the distinct level of difficulty of the task.

Ikehara and Crosby (2005) use an eye tracker, a pressure sensor for the mouse, an electrodermal activity (EDA) sensor and a pulse oximeter. Several mathematical fractions are shown on a computer screen to the participants in the experiment, who have to select fractions that are less than 1/3. Experimenting with two levels of difficulty, the authors determine that the EDA signal and PS have the greatest statistical significance for assessing mental workload.

Shi, Ruiz, Taib, Choi, and Chen (2007) assess stress and arousal levels using an EDA signal. In the experiment, the subjects have to answer questions in three ways: using gestures and speaking, only speaking and only using gestures. The researchers show that the EDA signal significantly increases if the task is more complex. Similarly, Nourbakhsh, Wang, and Chen (2013) use an EDA signal to discriminate between the difficulty of eight arithmetic tasks with four levels of complexity.

Ryu and Myung (2005) evaluate the mental workload of arithmetic tasks with various levels of difficulty. The devices used are an electroencephalogram (EEG), an electrooculogram (EOG) and an electrocardiogram (ECG). The results indicate that the mental workload of an arithmetic task is accurately inferred using the suppression of the alpha rhythm—extracted from the EEG—and that of a tracking task is effectively measured using the number of eye blinks—extracted from the EOG—and heart rate variability (HRV)—extracted from the ECG.

Most recent studies focus on training classifiers with processed psychophysiological signal data to predict whether the mental workload on a specific task is high or low (Yoshida et al., 2014). For instance, Haapalainen et al. (2010) assess the mental workload of tasks such as solving problems related to visual perception and cognitive speed. By using an eye tracker, an EEG, an ECG, heat flow and heart rate (HR), the authors obtain 81.1% average accuracy in classifying the mental workload. Fritz, Begel, Müller, Yigit-Elliott, and Züger (2014) assess a computer code comprehension task. Obtaining data through an eye tracker, an EEG and an EDA device, they achieve an accuracy of 85% in classifying two levels of mental workload using an SVM model.

Chen and Epps (2013) propose an eye-based mental workload measurement system that analyzes three types of eye activity: pupillary response, blinks and eye movement (fixation and saccade). The experiment consists of solving arithmetic tasks while various images from the International Affective Picture System (IAPS) are shown. By training a Gaussian mixture model classifier, the authors obtain an average accuracy of 79% in classifying two levels of mental workload.

2.3. SVM background

Given a set of instances, each with a binary label, denoted by $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i \in \Re^n$ and $y_i \in \{-1, +1\}$ for $i = 1, \ldots, m$, the soft-margin SVM classifier (Cortes & Vapnik, 1995) finds a hyperplane of the form $\mathbf{w}^\top \mathbf{x} + b = 0$ by solving the following quadratic programming problem (QPP):

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0,\ \ i = 1, \ldots, m, \end{aligned} \tag{1}$$

where ξ_i is the soft-margin slack of the ith training point, and $C > 0$ is a regularization parameter.

The SVM approach has been extended to elicit consumer preferences by constructing individual utility functions in the context of choice-based conjoint analysis (Evgeniou, Boussios, & Zacharia, 2005; Evgeniou, Pontil, & Toubia, 2007; López, Maldonado, & Montoya, 2017). In choice-based conjoint studies, consumers choose between various fictitious product profiles in a questionnaire, generating (limited) information that is used to estimate utility functions (Evgeniou et al., 2005; Green & Rao, 1971). Conjoint analysis has important similarities with our case study, since the latter involves persons engaged in combinations of smartphone interactions and daily tasks, who therefore are generating individual data that can be used for modeling how their mental workload relates with the complexity of the combination.

Heterogeneity control for SVMs was first proposed by Evgeniou et al. (2005). The idea was to estimate all individual functions independently and to subsequently compute a population function as the average of all individual estimates. The final individual functions are a linear combination of the original functions and the population function. Later, Evgeniou et al. (2007) proposed LOG-Het, a single optimization problem that aimed at finding individual utility functions while penalizing their deviations from the population mean. LOG-HET used a logistic error function, and therefore, it could not be directly linked to an SVM. LOG-HET was adapted to an SVM approach by López et al. (2017) by replacing the logistic error function with the hinge loss. Finally, the idea of performing feature selection together with heterogeneity control was developed by Maldonado, Montoya, and López (2017), in which the ℓ_2 regularization was replaced by the ℓ_1-norm, and a backward elimination process was implemented.

2.4. Feature selection for SVMs

Feature selection is an important topic in artificial intelligence literature, since the research areas and application domains are vast, including business analytics (Seret, Maldonado, & Baesens, 2015), medical diagnosis (Shilaskar & Ghatol, 2013), and demand forecasting (Jiang, Chin, Wang, Qu, & Tsui, 2017). Although SVMs have important advantages in terms of predictive performance and computational efficiency, they cannot derive the importance of the variables automatically. Therefore, several extensions have been proposed in the literature, by either modifying the original SVM formulation presented in Eq. (1) or altering the structure of the method to assess feature relevance.

A well-known strategy for SVM classification is to penalize the use of features by replacing the Euclidean norm in Formulation (1) with a regularizer that encourages sparsity. The most common approach is the use of the LASSO penalty, or the ℓ_1-norm (Bradley & Mangasarian, 1998), which leads to the following formulation:

$$\begin{array}{lll} \min_{\mathbf{w},b,\xi} & \|\mathbf{w}\|_1 + C\sum_{i=1}^{m}\xi_i & \\ \text{s.t.} & y_i(\mathbf{w}^\top\mathbf{x}_i + b) \geq 1-\xi_i,\ \ \xi_i \geq 0, & i=1,\ldots,m, \end{array} \tag{2}$$

where $\|\mathbf{w}\|_1 = \sum_{i=1}^n |w_i|$ denotes the ℓ_1-norm of vector $\mathbf{w}\in\Re^n$.

Similarly, a *group penalty function* can be used as a regularization strategy. The idea is to penalize the use of a group of related variables together in such a way that sparsity is encouraged at a group level rather than by removing weights independently (Yuan & Lin, 2006). This strategy was originally developed for binary classification with categorical attributes with multiple levels, which are usually transformed into sets of dummy variables (Yuan & Lin, 2006). In such cases, it may be desirable to remove the full set of dummy variables (Yuan & Lin, 2006) rather than the levels individually. Feature selection can then be performed simultaneously at the variable level, jointly penalizing all the weights related to one attribute (Chapelle & Keerthi, 2008). This idea was extended to multiclass classification (Chapelle & Keerthi, 2008) and twin SVM classification (López & Maldonado, 2017).

The best-known group penalty is called *group-LASSO* (Yuan & Lin, 2006), which extends the idea of the LASSO penalty by penalizing the Euclidean norm of the weights related to a given group. Formally, suppose that each of n available attributes is assigned to one of J disjoint groups of size n_j, that is, $\{1,\ldots,n\} = \bigcup_{j=1}^J \mathcal{I}_j$, $|\mathcal{I}_j| = n_j$, and $\sum_{j=1}^J n_j = n$. A group structure for the weight vector can be defined as $\mathbf{w} = [\mathbf{w}^{(1)\top},\ldots,\mathbf{w}^{(J)\top}]^\top \in \Re^n$, with $\mathbf{w}^{(j)} \in \Re^{n_j}$ for $j=1,\ldots,J$. A generalized regularization term of the $\ell_{p,q}$-norm can then be formalized based on this group structure as follows (Maldonado, Bravo, López, & Pérez, 2017):

$$\|\mathbf{w}\|_{p,q} := \left(\sum_{j=1}^J \|\mathbf{w}^{(j)}\|_p^q\right)^{1/q}, \quad p,q>0, \tag{3}$$

where the ℓ_p-norm is applied for each group, and the ℓ_q-norm is subsequently applied for the resulting vector. Note that if $p=q$, the $\ell_{p,q}$-norm coincides with the ℓ_p-norm, that is, $\|\mathbf{w}\|_{p,p} = \|\mathbf{w}\|_p$. Taking into account the above norm, Yuan and Lin (2006) proposed the group-LASSO function as follows:

$$\Gamma(\mathbf{w}) = \sum_{j=1}^J \sqrt{n_j}||\mathbf{w}^{(j)}||_2. \tag{4}$$

Another penalty function is the $\ell_{\infty,1}$-norm, which has the following form:

$$\Gamma(\mathbf{w}) = \|\mathbf{w}\|_{\infty,1} := \sum_{j=1}^J ||\mathbf{w}^{(j)}||_\infty, \tag{5}$$

where $||\mathbf{w}^{(j)}||_\infty = \max_{l\in\mathcal{I}_j}\{|w_l|\}$. The $\ell_{\infty,1}$-norm penalty was originally developed, under the name F_∞-norm SVM (Zou & Yuan, 2008), for dealing with categorical variables in binary SVM classification. The advantage of the $\ell_{\infty,1}$-norm over the group LASSO function is that the resulting SVM formulation can be efficiently solved, since it can be cast into a linear programming problem. A set of slack variables $t_j = ||\mathbf{w}^{(j)}||_\infty$ can be introduced, and the problem then entails the minimization of these variables with the inclusion of constraints of the form $|w_l| \leq t_j$ for each $l\in\mathcal{I}_j$ and $j=1,\ldots,J$. For simplicity, we refer to the $\ell_{\infty,1}$-norm penalty as the ℓ_∞-norm in the remainder of this paper.

3. Proposed SVM approach for simultaneous feature selection and heterogeneity control

In this section, two novel SVM formulations are proposed. Assuming that several data points are generated by persons, the main idea is to construct linear SVM classifiers, one per person, based on four objectives:

- Generalization and complexity reduction: Following the reasoning behind SVM, we encourage complexity reduction by including a regularizer based on either the ℓ_2 or the ℓ_1-norm. The motivation behind this idea is to maximize the separation margin between the two classes, improving the model's generalization capabilities.
- Model fit: All artificial intelligence models need to be accurate. To guarantee an adequate classification performance, the model fit is maximized by including slack variables $\boldsymbol{\xi}$. The sum of these variables is minimized in a similar way to an SVM (using the hinge loss function).
- Cost-based feature selection: We include group penalty functions to encourage the use of few groups of related features. We assume that for each person, various devices are used to gather information, and several features can be collected from each device. Under this scheme, it is desirable to use as few devices as possible, taking into account that some devices are more expensive than others and that some tests can be more obtrusive than others. We suggest using the $\ell_{\infty,1}$-norm as the group penalty function.
- Heterogeneity control: Assuming that little per-person information is available, it is desirable to pool information across participants to improve generalization and reduce the risk of

overfitting that arises from the estimation of models with few samples. To this end, we minimize the differences between per-person weights and a population $\mathbf{w}_0$ estimated from the data of all persons.

This strategy avoids high deviations among persons, shrinking the weights to a population mean.

The combination of the last two objectives is the main methodological contribution of this paper. Note that these two objectives are related to the main goal of machine learning, i.e., to generalize new objects well. On the one hand, feature selection is important when few samples are available due to the *curse of the dimensionality*. Informally, the volume of the space increases so fast with dimensionality that the data points become sparse. Overfitting is a direct consequence of the curse of dimensionality, since using too many attributes leads to better in-sample classification, although it may not generalize well when new data are encountered (Guyon et al., 2006). On the other hand, heterogeneity control allows better generalization by forcing the parameters of the model to be less sensitive to the noise caused by small samples, aiming at balancing individual and general patterns. Although these two objectives pursue the same goal to some extent, they act in different directions; while feature selection shrinks the weights towards zero, heterogeneity control pulls them towards the population mean $\mathbf{w}_0$. In this context, the proposed formulations encourage the removal of a particular source of information in all the hyperplanes that are constructed when many components of $\mathbf{w}_0$ are close to zero. Nevertheless, it is important to set the trade-off parameters carefully.

First, we propose $\ell_2-\ell_\infty$-SVM-HET, which uses the Euclidean norm for the first and the fourth objectives (margin maximization and heterogeneity control). Next, $\ell_1-\ell_\infty$-SVM-HET is presented, which uses the ℓ_1-norm instead.

3.1. $\ell_2-\ell_\infty$-SVM-HET

Let us denote by $\mathbf{w}_k \in \Re^n$ the individual weight vectors for each person $k=1,\dots,K$. Given tuples of the form $\left(\mathbf{x}_{ik}, y_i^k\right)$, where $\mathbf{x}_{ik} \in \Re^n$ and $y_i^k \in \{+1,-1\}$ for $i=1,\dots,I_k$, the following formulation is proposed:

$$\begin{array}{ll} \min_{\substack{\mathbf{w}_k,\mathbf{w}_0,\\ b_k,\xi_i^k}} & \sum_{k=1}^K \left(\frac{1}{2}(\|\mathbf{w}_k\|^2+\theta\|\mathbf{w}_k-\mathbf{w}_0\|^2) \right. \\ & \left. \quad +\lambda \sum_{j=1}^{J} c_j ||\mathbf{w}_k^{(j)}||_\infty + \gamma \sum_{i=1}^{I_k}\xi_i^k \right) \\ \text{s.t.} & y_i^k(\mathbf{w}_k^\top \mathbf{x}_{ik}+b_k)\ge 1-\xi_i^k,\ \ \xi_i^k\ge 0, \\ & i=1,\dots,I_k,\ \ k=1,\dots,K, \end{array} \tag{6}$$

where θ, λ, and γ are positive trade-off parameters that can be set using cross-validation, and $\sum_{k=1}^K I_k = m$ is the total number of samples. Parameter $\mathbf{c}=(c_1,\dots,c_J)$ represents the relative cost for each data source and can be regarded as an incentive for the penalty function to prioritize the removal of expensive data sources. We recommend setting this parameter by estimating the cost of using a given group of variables relative to that of the others (see Section 4.7 for the derivation of $\mathbf{c}$ in our case study). The objective function in Formulation (6) represents the four objectives discussed previously. First, the Euclidean norm of weights $\mathbf{w}_k$ related to the individual classifiers is minimized (complexity reduction). The second term represents the squared differences between each individual function and the population vector $\mathbf{w}_0$, aiming at controlling heterogeneity. The third term represents the ℓ_∞-norm, as described in Eq. (5). The feature penalization is controlled by parameter λ. Finally, the last term in the objective function guarantees an adequate model fit by penalizing misclassified elements (using the hinge loss function).

To avoid using a nonsmooth function in Formulation (6), a set of auxiliary variables z_{kj} is introduced for each $k=1,\dots,K$, including the new constraints $|w_{kl}|\le z_{kj}$ for each $l\in\mathcal{I}_j$ and $j=1,\dots,J$. These modifications lead to the following QPP:

$$\begin{array}{ll} \min_{\substack{\mathbf{w}_k,\mathbf{w}_0,b_k,\\ \xi_i^k,\mathbf{z}_k}} & \sum_{k=1}^K \left(\frac{1}{2}(\|\mathbf{w}_k\|^2+\theta\|\mathbf{w}_k-\mathbf{w}_0\|^2)\right. \\ & \left. \quad +\lambda(\mathbf{c}*\mathbf{z}_k)^\top\mathbf{e}+\gamma\sum_{i=1}^{I_k}\xi_i^k\right) \\ \text{s.t.} & y_i^k(\mathbf{w}_k^\top\mathbf{x}_{ik}+b_k)\ge 1-\xi_i^k,\ \ \xi_i^k\ge 0, \\ & i=1,\dots,I_k,\ k=1,\dots,K, \\ & -z_{kj}\mathbf{e}\le \mathbf{w}_k^{(j)}\le z_{kj}\mathbf{e},\ \ j=1,\dots,J,\ \ k=1,\dots,K, \end{array} \tag{7}$$

where $\mathbf{z}_k=(z_{k1},\dots,z_{kJ})\in\Re^J$, the asterisk $(*)$ denotes the componentwise vector product operator, and $\mathbf{e}$ is a vector of ones of the appropriate dimension. The following remarks are important for an efficient implementation of our approach. The proofs for these remarks are presented as supplementary material.

Remark 3.1.

(a) Proposition 1 (see Appendix A in the online supplementary material) shows that the quadratic problem (7) is strictly convex. This result is key to our modeling approach, since it implies that our model guarantees a single optimal solution.

(b) Additionally, in Appendix A, it is shown that the following relation holds (see Eq. (A.15)):

$$\mathbf{w}_0=\frac{1}{K}\sum_{k=1}^K \mathbf{w}_k. \tag{8}$$

Hence, we can first compute the person-level weights $\mathbf{w}_k$ and subsequently obtain vector $\mathbf{w}_0$ by using Eq. (8).

(c) Note that if vector $\mathbf{w}_0$ is known a priori, Formulation (7) is separable, and we can solve the following problem:

$$\begin{array}{ll} \min_{\mathbf{w}_k,b_k,\xi_i^k,\mathbf{z}_k} & \frac{1}{2}(\|\mathbf{w}_k\|^2+\theta\|\mathbf{w}_k-\mathbf{w}_0\|^2)+\lambda(\mathbf{c}*\mathbf{z}_k)^\top\mathbf{e} \\ & \quad +\gamma\sum_{i=1}^{I_k}\xi_i^k \\ \text{s.t.} & y_i^k(\mathbf{w}_k^\top\mathbf{x}_{ik}+b_k)\ge 1-\xi_i^k,\ \ \xi_i^k\ge 0,\ \ i=1,\dots,I_k, \\ & -z_{kj}\mathbf{e}\le\mathbf{w}_k^{(j)}\le z_{kj}\mathbf{e},\ \ j=1,\dots,J, \end{array} \tag{9}$$

for each $k=1,\dots,K$.

Taking into account Remark 3.1(b) and (c), we propose a “divide and conquer” approach for Formulation (7) to compute the person-level weights $\mathbf{w}_k$. This approach is shown in Algorithm 1.

Algorithm 1 $\ell_2-\ell_\infty$-SVM-HET Algorithm.

1. For each $k=1,\dots,K$, $\mathbf{w}_k$ is obtained by solving

$$\begin{array}{l} \min\limits_{\mathbf{w}_k,b_k,\xi_i^k,\mathbf{z}_k} \ \ \frac{1}{2}\|\mathbf{w}_k\|^2+\lambda(\mathbf{c}*\mathbf{z}_k)^\top\mathbf{e}+\gamma\sum_{i=1}^{I_k}\xi_i^k \\ \quad \text{s.t.}\ y_i^k(\mathbf{w}_k^\top\mathbf{x}_{ik}+b_k)\ge 1-\xi_i^k,\ \ \xi_i^k\ge 0,\ \ i=1,\dots,I_k, \\ \quad\quad -z_{kj}\mathbf{e}\le\mathbf{w}_k^{(j)}\le z_{kj}\mathbf{e},\ \ j=1,\dots,J. \end{array} \tag{10}$$

2. Compute $\mathbf{w}_0$ via Eq. (8).
3. For each $k=1,\dots,K$, $\mathbf{w}_k$ is obtained by solving formulation (9).

Finally, the decision function for a new sample $\mathbf{x}$ is given by $f_k(\mathbf{x})=\mathbf{w}_k^\top\mathbf{x}+b_k$, for $k=1,\dots,K$.

3.2. $\ell_1-\ell_\infty$-SVM-HET

This approach is similar to $\ell_2-\ell_\infty$-SVM-HET, with the only difference being that the first two terms of the objective function in

Formulation (7) use the ℓ_1-norm instead of the ℓ_2-norm for complexity and heterogeneity reduction. Let us consider the following problem:

$$\begin{aligned}\min_{\substack{\mathbf{w}_k,\mathbf{w}_0,\\ b_k,\xi_i^k}} \quad & \sum_{k=1}^K \left(\|\mathbf{w}_k\|_1 + \theta\|\mathbf{w}_k - \mathbf{w}_0\|_1 \right. \\ & \left. +\lambda \sum_{j=1}^J c_j ||\mathbf{w}_k^{(j)}||_\infty + \gamma \sum_{i=1}^{I_k} \xi_i^k \right) \\ \text{s.t.} \quad & y_i^k(\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) \geq 1 - \xi_i^k, \ \ \xi_i^k \geq 0, \\ & i = 1,\ldots,I_k, \ \ k = 1,\ldots,K. \end{aligned} \tag{11}$$

To avoid using nonsmooth functions in Formulation (11), auxiliary vectors of nonnegative variables $\mathbf{u}_k$, $\mathbf{v}_k \in \Re^n$, $\mathbf{z}_k \in \Re^J$ are included for each $k = 1,\ldots,K$, leading to the following linear programming problem:

$$\begin{aligned}\min_{\substack{\mathbf{w}_k,\mathbf{w}_0,\\ \mathbf{u}_k,\mathbf{v}_k,\mathbf{z}_k,\\ b_k,\boldsymbol{\xi}^k}} \quad & \sum_{k=1}^K \left(\mathbf{u}_k^\top \mathbf{e} + \theta \mathbf{v}_k^\top \mathbf{e} + \lambda(\mathbf{c} * \mathbf{z}_k)^\top \mathbf{e} + \gamma\, {\boldsymbol{\xi}^k}^\top \mathbf{e} \right) \\ \text{s.t.} \quad & y_i^k(\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) \geq 1 - \xi_i^k, \ \xi_i^k \geq 0, \\ & i = 1,\ldots,I_k, \ k = 1,\ldots,K, \\ & -\mathbf{u}_k \leq \mathbf{w}_k \leq \mathbf{u}_k, \ \ k = 1,\ldots,K, \\ & -\mathbf{v}_k \leq \mathbf{w}_k - \mathbf{w}_0 \leq \mathbf{v}_k, \ \ k = 1,\ldots,K, \\ & -z_{kj}\mathbf{e} \leq \mathbf{w}_k^{(j)} \leq z_{kj}\mathbf{e}, \ \ j = 1,\ldots,J, \ \ k = 1,\ldots,K. \end{aligned} \tag{12}$$

Note that the decision rule for $\ell_1 - \ell_\infty$-SVM-HET is similar to that for $\ell_2 - \ell_\infty$-SVM-HET. Additionally, note that the proposed approach does not require a backward elimination step or any other pruning strategy for performing feature selection.

Analogously to $\ell_2 - \ell_\infty$-SVM-HET, we propose solving Formulation (12) in three steps, as shown in Algorithm 2.

Algorithm 2 $\ell_1 - \ell_\infty$-SVM-HET Algorithm.

1. For each $k = 1,\ldots,K$, $\mathbf{w}_k$ is obtained by solving

$$\begin{aligned}\min_{\mathbf{w}_k,b_k,\xi_i^k,\mathbf{z}_k} \quad & \|\mathbf{w}_k\|_1 + \lambda(\mathbf{c} * \mathbf{z}_k)^\top \mathbf{e} + \gamma \sum_{i=1}^{I_k} \xi_i^k \\ \text{s.t. } & y_i^k(\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) \geq 1 - \xi_i^k, \ \ \xi_i^k \geq 0, \ \ i = 1,\ldots,I_k, \\ & -z_{kj}\mathbf{e} \leq \mathbf{w}_k^{(j)} \leq z_{kj}\mathbf{e}, \ \ j = 1,\ldots,J. \end{aligned} \tag{13}$$

2. Compute $\mathbf{w}_0$ via Eq. (8).
3. For each $k = 1,\ldots,K$, $\mathbf{w}_k$ is obtained by solving

$$\begin{aligned}\min_{\substack{\mathbf{w}_k,\mathbf{u}_k,\\ \mathbf{v}_k,\mathbf{z}_k,b_k,\boldsymbol{\xi}^k}} \quad & \mathbf{u}_k^\top \mathbf{e} + \theta \mathbf{v}_k^\top \mathbf{e} + \lambda(\mathbf{c} * \mathbf{z}_k)^\top \mathbf{e} + \gamma\, {\boldsymbol{\xi}^k}^\top \mathbf{e} \\ \text{s.t.} \quad & y_i^k(\mathbf{w}_k^\top \mathbf{x}_{ik} + b_k) \geq 1 - \xi_i^k, \ \ \xi_i^k \geq 0, \ \ i = 1,\ldots,I_k, \\ & -\mathbf{u}_k \leq \mathbf{w}_k \leq \mathbf{u}_k, \\ & -\mathbf{v}_k \leq \mathbf{w}_k - \mathbf{w}_0 \leq \mathbf{v}_k, \\ & -z_{kj}\mathbf{e} \leq \mathbf{w}_k^{(j)} \leq z_{kj}\mathbf{e}, \ \ j = 1,\ldots,J. \end{aligned} \tag{14}$$

4. Case study

We apply the proposed approach described in Section 3 to a multidevice dataset obtained from a case study for classifying the mental workload of a set of participants through their psychophysiological signals. The tasks consisted of participants engaging in various combinations of smartphone interactions and daily common tasks simultaneously.

The goals of machine learning are to analyze (i) whether it is possible to infer mental workload from a multidevice dataset of psychophysiological signals as well as (ii) which signals explain mental workload best and which ones are truly needed for constructing accurate predictors, taking into account the cost of the devices used for capturing the psychophysiological signals. While the first objective requires predictive modeling, feature selection is extremely relevant to accomplishing the second.

4.1. Task design

Mental workload is manipulated through the cognitive interference phenomenon described in Wickens's MRM (see Section 2.1). In this case, it is important to note that a task will be more cognitively demanding if there are more cognitive interferences involved. Specifically, participants are required to perform seven tasks using certain applications on a smartphone while their psychophysiological signals are being measured. The first two smartphone interactions require the use of the Gmail application and are as follows: reading a simple e-mail and replying to a simple e-mail. The following five smartphone interactions require the use of the Gmail application and a suitable application for searching for certain information. These interactions are as follows: reading an e-mail with search instructions, opening the application to follow the search instructions, entering the search parameters, reading and analyzing the search results and replying to the e-mail with the results.

One of the most important applications of the MRM is assigning mental workload levels to two or more activities performed at the same time. This model can be used intuitively, as shown in Fig. 1. To this end, a score is assigned according to the cognitive resource demanded in the multitask scenario. Thus, if the cognitive resource is exclusive and therefore cannot be used simultaneously for two activities, the score is 1 (high); if the resource can be perfectly shared, the score is 0 (none), and a score of 0.5 (low) is chosen if the cognitive resource can be partially shared.

In this study, two treatments were designed to induce low and high mental workload. In the first treatment, participants must perform each of the smartphone interactions free of interruptions. According to the MRM, participants only need to control their personal space while performing the interactions. In this treatment, only one interference is caused by the simultaneous use of motor control resources and verbal code attributes. The second treatment requires performing all the interactions while simultaneously verbally answering questions of the experimenter and listening to music (in the native language of the participant) through a hearing aid. Thus, the participant should pay attention to, examine, analyze, and make sense of the questions while using auditory and visual input modalities and working memory. In this case, several partial interferences occur, since both the tasks and the verbal stimulus require motor control and working memory resources simultaneously (see Fig. 1).

4.1.1. Task difficulty validation

Considering the interference activities of Fig. 1 as an example, in the case of listening, there is a significant interference with the use of the cognitive resource of auditory perception, to which the value of 1 (high) is assigned. In addition, music in the native language creates a partial interference with the verbal perception resource, to which the value of 0.5 (low) is assigned; in turn, listening to music requires a high demand on the verbal information processing resource, which uses working memory, so the value of 1 (high) is assigned. For the resources not demanded by the task, the value of 0 (none) is assigned, since these resources remain fully available.

An additional validation was carried out once all the experimental sessions were finished. To this end, the NASA Task Load Index scores were used. For each task, the scores of the NASA Task Load Index for the two treatments were compared using a

Scenario 2		Perceptual Encoding								Central Processing		Responding	
		Selective Attention						Perception		Working Memory		Responsive (Motor Control)	
		Input Modality				Processing Code		Processing Code		Processing Code		Processing Code	
		Auditory	Focal visual	Ambient Visual	Touch	Spatial	Verbal	Spatial	Verbal	Spatial	Verbal	Spatial	Verbal
HCI Tasks	Typing Info		Low				Low		Low		Low	High	
	Searching from display		Low			Low	Low		Low		High		
	Deciding a path		Low								High		
	Reading		Low				Low		Low		High		
	Waiting for loading		High	High									
	Watching		High				Low		Low		Low		
	Analyzing										High		
	Listening	High					Low		Low		High		
Physical Tasks	Controlling Space	Low		Low		Low		Low				High	
	Talking	Low	Low	Low			Low		Low		High		High
	Analyzing										High		
	Listening	High					Low		Low		High		

Fig. 1. Example of a high mental workload treatment.

repeated measure ANOVA (RM-ANOVA). It was concluded that the difference between the mean of the scores was statistically significant for each task ($p < .001$), which showed that there were two levels of mental workload.

4.2. Participants

The experiment was performed with 50 engineering students from the University of Chile, including 33 men and 17 women with a mean age of 22.4 $\pm$ 2.8 years. They were recruited using the university news application. They did not suffer from cardiovascular diseases, had no vision impairments and were not being treated with medications that could affect their normal behavior. Additionally, to avoid bias, all of them were familiar with the use of smartphones. They were compensated CL$5,000 (approximately US$10) for their participation.

This study was approved by the Research Ethics Committee at the Faculty of Physical and Mathematical Sciences at the University of Chile. All participants read and signed an informed consent form.

4.3. Apparatus

HR, blood oxygen saturation (SpO_2) and temperature were collected from sensors provided by the Cooking Hacks Company, at a frequency of 50 Hz, using an e-health shield connected to an Arduino microcontroller that sent the data to a Linux server. The PS signal was collected at a rate of 120 Hz using an eye tracker developed by the Pupil Labs Company and equipped with two cameras for recording the eye and the vision of the participant. Since PS values are defined in pixels, a resolution setting of 320×240 pixels was used for the front camera, which was compatible with capturing 120 frames per second, as recommended by the manufacturer. This eye tracker also provided eye-gaze data points that were needed to identify the timestamps coinciding with saccades and blinks, which were useful in processing the PS data points, as shown in Section 4.5. HR was captured in beats per minute (bpm) and SpO_2 through the relation of the wavelength of blood with hemoglobin in the presence or absence of oxygen. The ECG, collected through electrodes provided by the Cooking Hacks Company, was captured in millivolts (mV) at a frequency of 110 Hz. Temperature was captured in degrees Celsius (°C).

4.4. Experimental protocol

As soon as each participant entered the laboratory, he/she was informed about the objectives of the experiment and was asked to complete a basic personal data questionnaire along with reading and signing of the informed consent form. The sensors were installed and began the data capture in the following order: the ECG electrodes on the chest, the pulse oximeter on the index finger of the nondominant hand, the thermometer on the ring finger of the same hand, and the eye tracker built into a pair of glasses. Then, the eye tracker was calibrated using the software provided by the manufacturer, and a testing activity was performed. Next, the participant was asked to close his/her eyes for 90 seconds to relax before starting the tasks. This was done to mitigate the modification of the behavior of the subjects—known as the Hawthorne effect (Parsons, 1974)—due to their awareness of being studied, followed by a resting state for two minutes. For each participant, the three tasks were presented randomly. In the beginning, the instructions for the correct completion of the tasks were given. After completing each task, the participant was asked to complete the NASA Task Load Index questionnaire (Hart, 1988). Subjects relaxed for 30 seconds between tasks. When a participant completed the

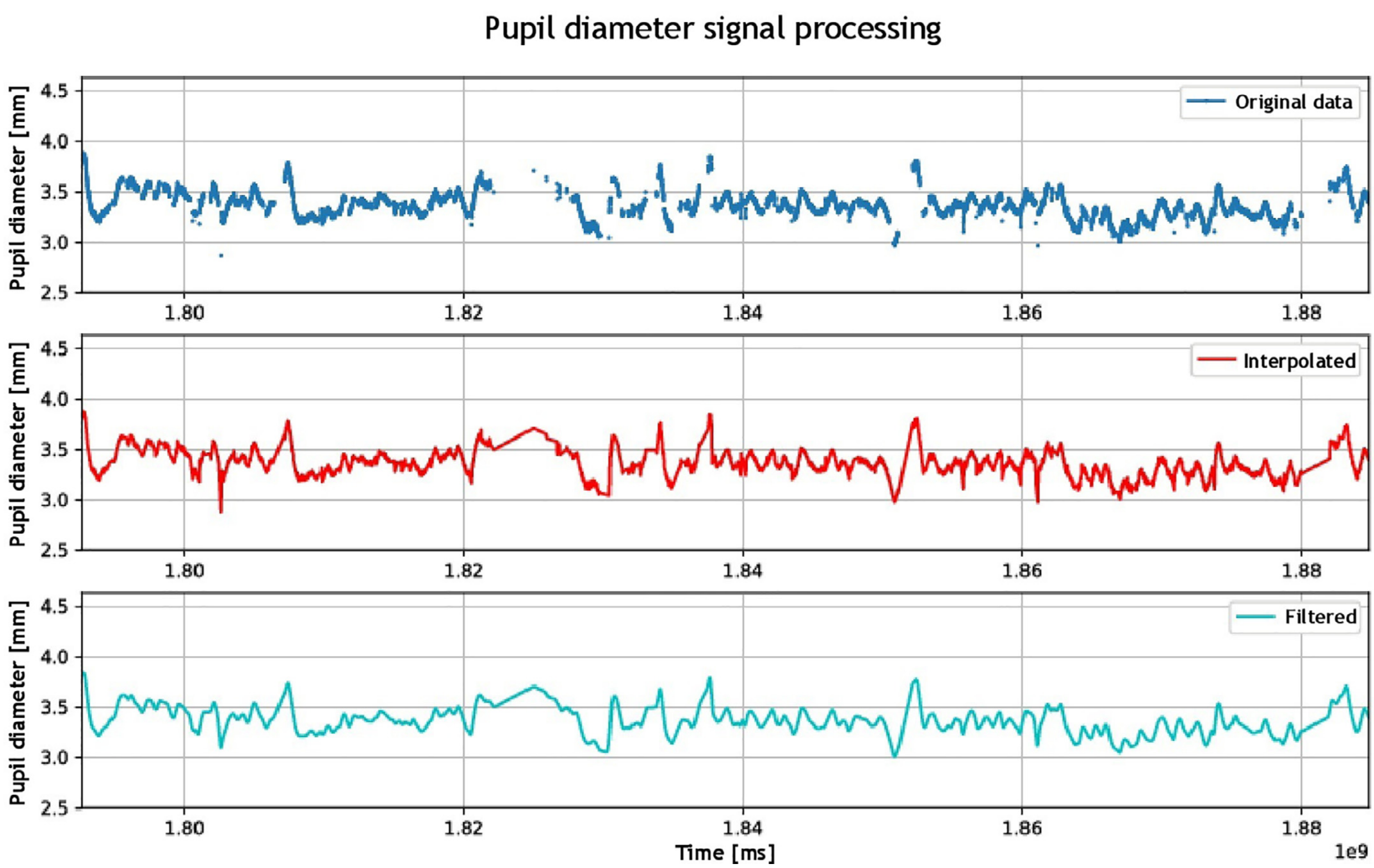


Fig. 2. Pupil diameter signal processing.

last questionnaire, the experiment was considered complete, and he/she was asked not to disclose the characteristics of the experiment to other people to avoid biasing other participants.

4.5. Signal processing

To use the psychophysiological signals, it is necessary to process various artifacts to obtain a noise-free signal suitable for use in machine learning models.

In the case of PS, the eye tracker and the Pupil Labs software provide a confidence level of the PS value obtained at each timestamp. This indicator is based on the adjustment of an ellipse using the Canny's edge detection algorithm, a filter that is based on the signal strength of neighboring pixels. If the value of the confidence level is above a certain threshold, the ellipse is reported as the true edge of the pupil. More details are available in Kassner, Patera, and Bulling (2014). Therefore, as the first step, we removed all the PS values with confidence levels of less than 60% as recommended by the manufacturer. This resulted in removal of 1.69% of the PS data points. The standard second step in processing of PS data is to remove and interpolate the PS values that coincide with the timestamps of the eye-gaze saccades and with the blinks—given that the eyelid blocks the infrared light of the eye tracker (Holmqvist et al., 2011). Since the time interval in which these events occur is a few milliseconds long, we performed a linear interpolation of the PS values for these timestamps. In the third step, a low-pass filter with a Blackman window with a cut-off frequency of 2 Hz was applied to eliminate high-frequency noise. These three steps are shown in Fig. 2.

For the ECG signal, we first performed the fast Fourier transform (FFT) to transform the samples from time domain to frequency domain and subsequently eliminate noisy records (high frequencies). To this end, we applied a 50 Hz Butterworth-type low-pass filter of the 8th order and then removed all the data points with higher frequencies. Finally, the inverse Fourier transform was performed to obtain a suitable signal (see Fig. 3).

Recent studies have proven that the discrete wavelet transform (DWT) is one of the most suitable tools for processing signals such as HRV. In fact, HRV was obtained from the ECG data points by identifying the maximum values in each cycle of the ECG wave (R points). This procedure was performed by using DWT, which provides a time-scale representation of a given signal, generated by dilation and translation of a function known as the discrete mother wavelet. Its structure is obtained by mother wavelets and scaling sequences deducted from one octave to the next by a two-scale difference equation (Addison, 2005). Specifically, signal $x_{j-1}[i]$ is passed through a low-pass filter and a high-pass filter with impulse responses $g[i]$ and $h[i]$ to produce the approximation coefficient $a_j[i]$ and the detailed coefficient $d_j[i]$, respectively, as shown in Eqs. (15) and (16).

$$a_j[i] = \sum_k x_j[k] g[2i-k], \quad j \geq 1, \tag{15}$$

$$d_j[i] = \sum_k x_j[k] h[2i-k], \quad j \geq 1. \tag{16}$$

We used the Daubechies-6 function with three grades of decompositions and eight levels, following the approach utilized in previous studies (Khandoker, Gubbi, & Palaniswami, 2009), since it is a suitable function for detecting variations in nonstationary signals (Haddad & Serdijn, 2009). The high-frequency components (coefficients d_1 and d_2) are removed from the DWT decomposition, so only d_3, d_4 and d_5 are used to reconstruct the signal. Next,

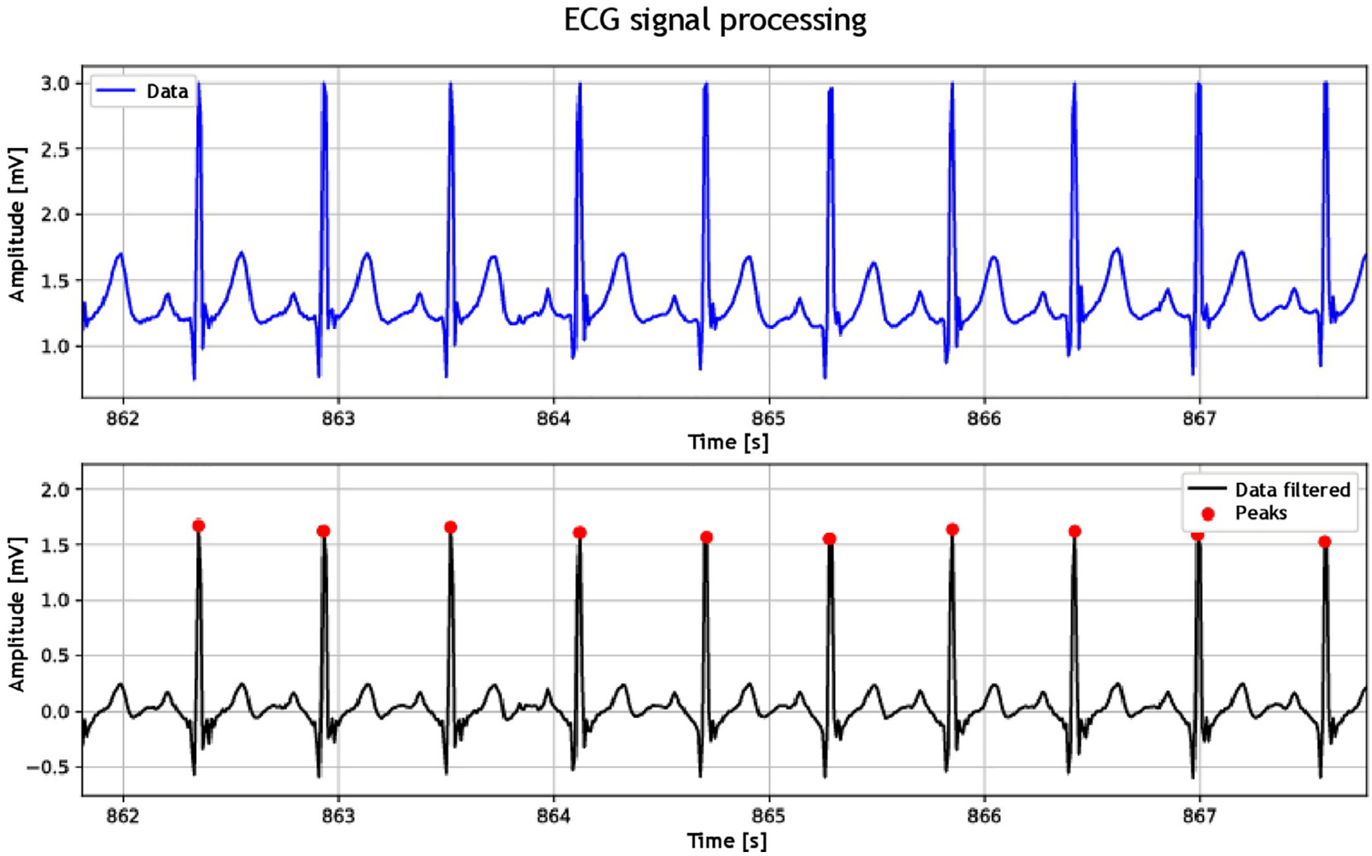


Fig. 3. Electrocardiogram signal processing.

its squared value is calculated to obtain the highest peaks, and an R-peak detection is performed. Finally, the time between each R point is calculated.

We applied a low-pass filter of 1 Hz to the temperature data points to reduce the noise and high variability between adjacent measurements (see Fig. 4).

For the HR and SpO2 signals, only missing values and outlier treatments were performed.

As a common treatment for all of the above signals, a normalization was performed using the baseline values obtained from the average of the signal data points at the last three seconds of the rest period of each participant. Then, the baseline values were subtracted from the signals obtained during the tasks' execution.

4.6. Feature extraction

The processed and standardized data of each task were divided into a set of data point segments, each with the duration of one second, according to the orders of magnitude of time window sizes frequently used in the literature on mental workload assessment with psychophysiological sensors (Fritz et al., 2014; Haapalainen et al., 2010; Ryu & Myung, 2005). Then, each one second long data point segment was labeled with the mental workload (low or high) induced by each treatment (previously validated by the score of the NASA Task Load Index). Next, several features were extracted from each of the data point segments based on the results of studies reviewed in Section 2, as described below.

Fritz et al. (2014), the mean, median and variance of the PS points in each segment were extracted as features. To highlight the sudden changes in pupillary signal slope, the first derivative $\delta[n]$ was also calculated, where $\delta[n] = \widehat{x}[n] - \widehat{x}[n-1]$.

Based on the study of Haapalainen et al. (2010), the median absolute deviation (MAD) of each ECG segment was obtained, as well as the mean, median and variance. The ECG MAD was obtained as $ECG_{MAD} = |ECG_i - median(ECG)|$.

Additionally, we calculated the mean, median and variance in each segment for each of the temperature, HR and SpO2 data series.

As recommended by Khandoker et al. (2009), we extracted the following features of HRV for each one-second-long segment:

- Mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{17}$$

where x_i corresponds to coefficients $a_j[i]$ and $d_j[i]$.

- Variance:

$$\sigma^2 = \sum_{i=1}^{N} (\bar{x} - x_i)^2. \tag{18}$$

- Energy of the approximate coefficients:

$$ENG_j = \sum_{i=1}^{M_j} |a_j[i]|^2. \tag{19}$$

- Entropy of the approximate coefficients:

$$ENT_j = -\sum_{i=1}^{M_j} p_j[i]\log_2 p_j[i], \tag{20}$$

with

$$p_j[i] = \frac{|a_j[i]|}{\sum_{k=1}^{M_j} |a_j[k]|}. \tag{21}$$

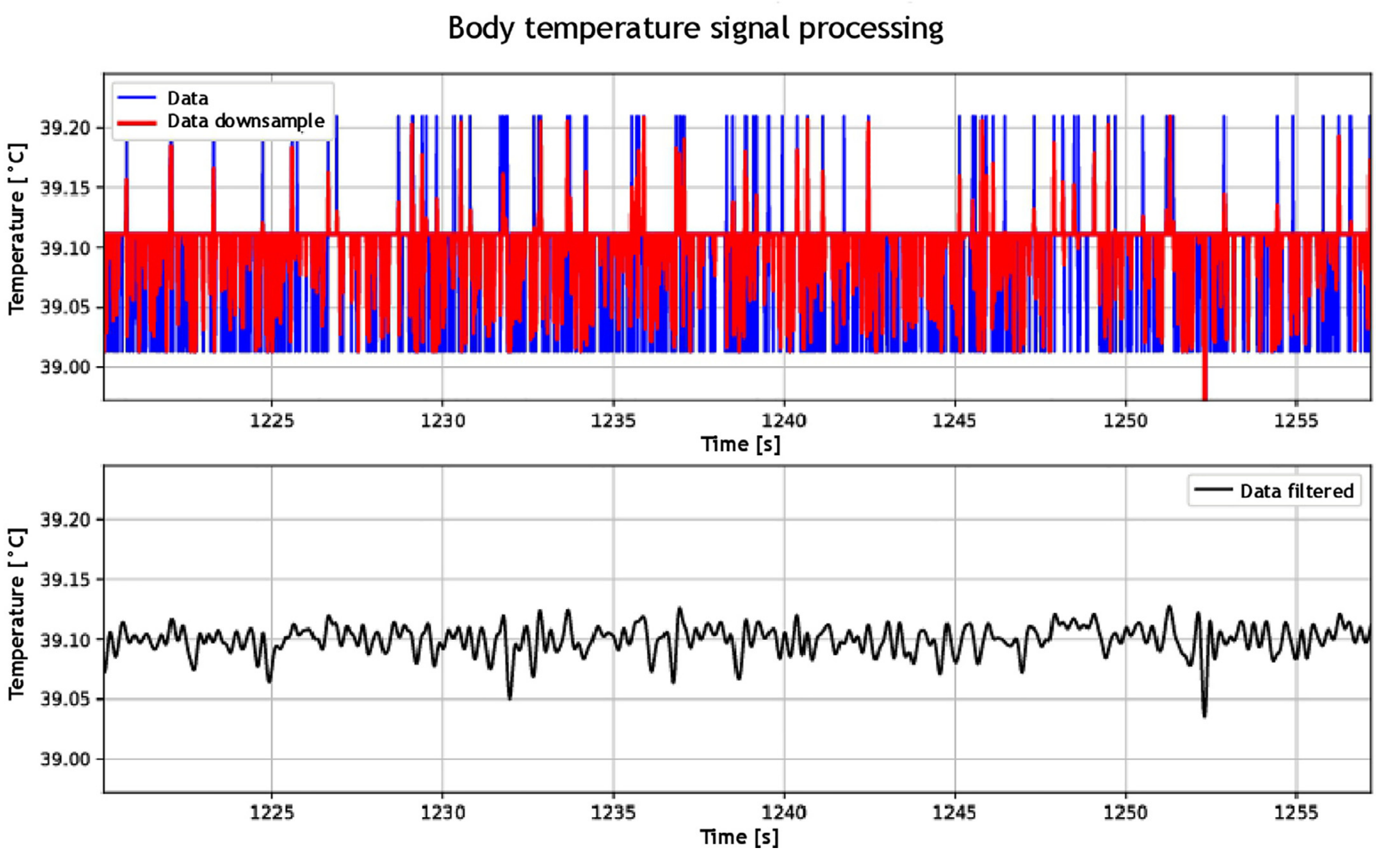


Fig. 4. Temperature signal processing.

4.7. Application of the proposed SVM approach

In this section, the performance of various machine learning methods applied to the features extracted from the multidevice dataset of the case study is explained and discussed. In particular, the proposed $\ell_2 - \ell_\infty$-SVM-HET and $\ell_1 - \ell_\infty$-SVM-HET methods are compared with the ℓ_2-SVM and ℓ_1-SVM methods.

Model validation is a very important step in SVM classification since it reduces the risk of overfitting (Tibshirani & Hastie, 2009). Based on Maldonado, Montoya et al. (2017), the following strategy was followed for model selection: hyperparameters C, λ, γ, and θ were set using a grid search with $\{2^{-7}, 2^{-6}, \ldots, 2^{0}, \ldots, 2^{6}, 2^{7}\}$ as possible values.

Regarding data normalization, all variables were scaled to [0,1]. For each person, we performed a leave-one-out (LOO) cross-validation (Tibshirani & Hastie, 2009) at the task level: each model was trained with $n-1$ tasks performed by a participant, leaving one out for testing. The process was repeated n times, where $n = 7$ was the number of tasks performed by each participant.

All experiments were performed using MATLAB R2016b. We used the LIBSVM (Chang & Lin, 2011) and LIBLINEAR (Fan, Chang, Hsieh, Wang, & Lin., 2008) toolboxes for the ℓ_2-SVM and ℓ_1-SVM methods, respectively. Our proposals and their variants were implemented on the CVX toolbox (Grant & Boyd, 2014), a multipurpose solver for convex optimization. To evaluate the performance of our method in terms of predictive accuracy and the cost of the solution, we first train all models using all available information and then perform a backward variable elimination process, removing the less relevant group of features and reevaluating the performance without retraining.

To compute the cost of the solution, we use the following values for the various devices that generate the five groups of signals (HR, SpO_2, temperature, PS, and HRV):

- SpO_2, HR, and temperature are the least expensive psychophysiological signals to acquire. As mentioned earlier, the first two signals are obtained using a pulse oximeter, while temperature is measured using a thermometer. These are also the least physically obtrusive instruments in this study because all of these sensors are placed on the fingers of the nondominant hand, as shown in Fig. 5. Therefore, we evaluate the relative cost of these signals to be 1. Note that although SpO_2 and HR are collected using the same device, they are considered two different data sources since they involve two different psychophysiological signals obtained for the participants. In other words, we regard the fixed cost of the device as marginal in comparison with the costs of providing a pulse oximeter device to the various participants.
- The cost of the ECG, which used electrodes placed on the chest to determine the HRV, is estimated to be 3 since its cost and obtrusiveness were considered to be medium.
- The glasses-shaped eye tracker, used to measure PS, is the most expensive device as well as the most physically obtrusive one for the participants in our experiments. Therefore, we evaluate the relative cost of this device to be 10, i.e., using this device is ten times as expensive as using the least expensive device.

Note that each of these “costs” is in reality a measure of the relative loss involved in using a given instrument and is inclusive of the monetary cost of buying and using a device, which is assumed to be fixed for all participants, and of the negative effects that the

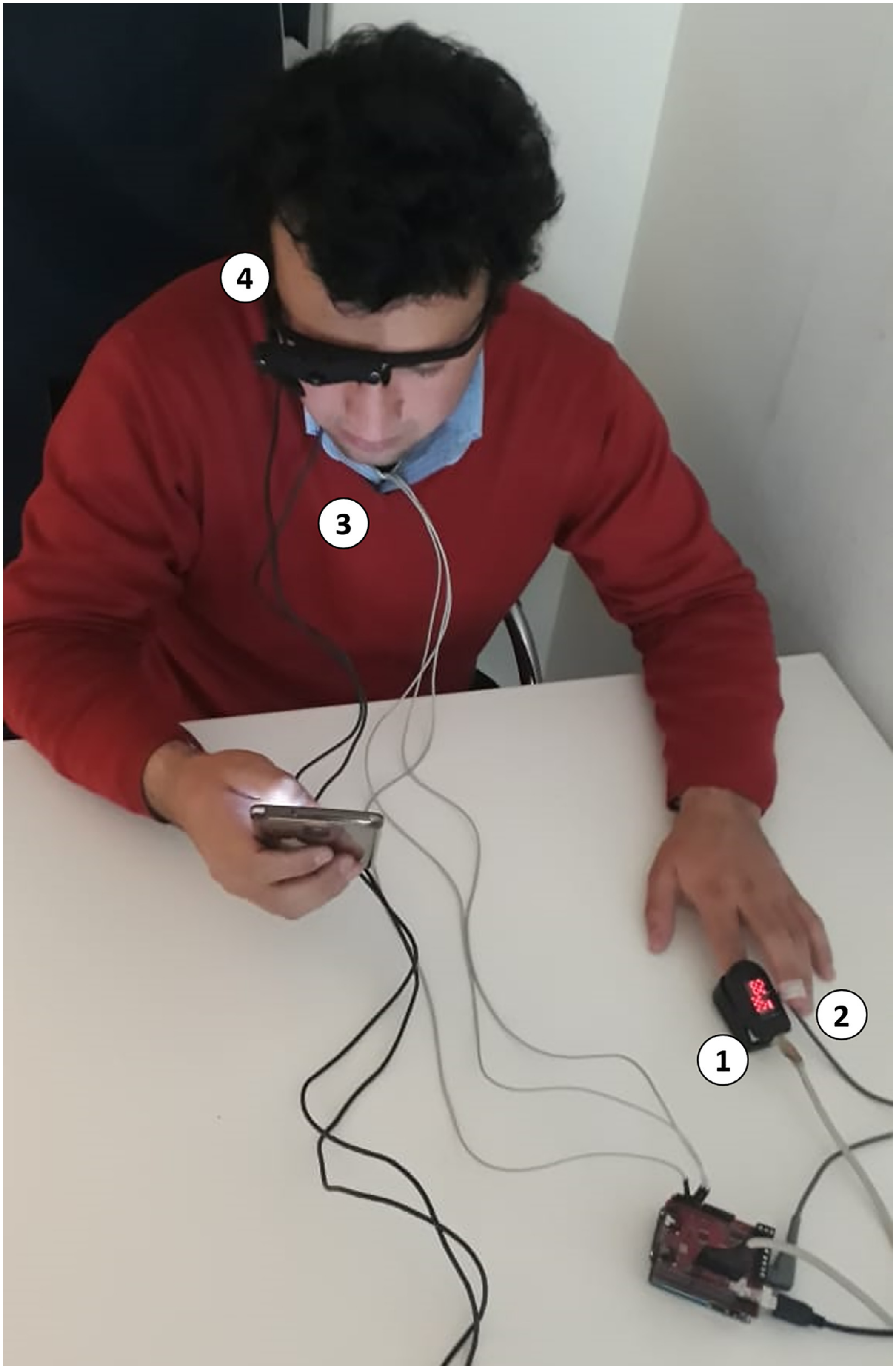
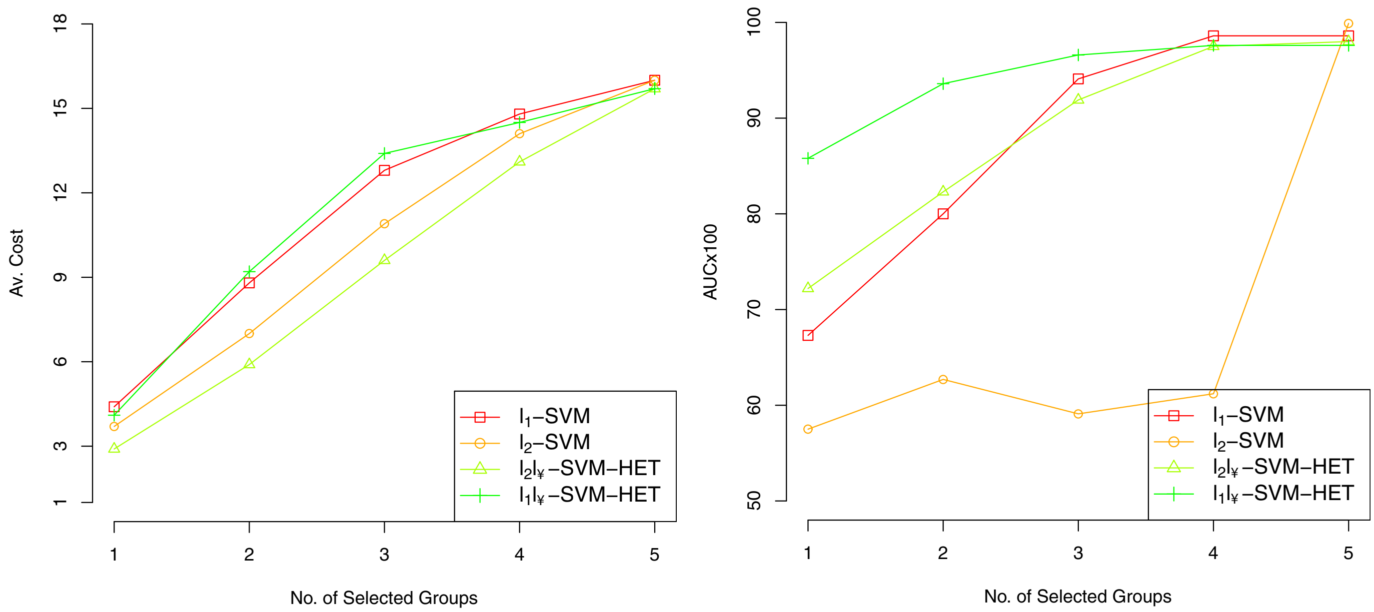


Fig. 5. Participant with the sensors performs the experiment. Sensors: (1) Pulse oximeter, (2) Skin temperature sensors, (3) ECG and (4) Eye tracker.



(a) Performance in terms of AUC. (b) Performance in terms of variable collection cost.

Fig. 6. Performance curves for the various SVM approaches.

Table 1
Performance summary for all approaches.

	AUCx100			Cost			Parameters		
	Mean ± STD	Max	Min	Mean ± STD	Max	Min	C	λ	θ
ℓ_2-SVM	68.1 ± 17.9	99.9	57.5	10.3 ± 5.1	16.0	4.4	2^6	–	–
ℓ_1-SVM	87.7 ± 13.7	98.6	67.3	11.4 ± 4.8	16.0	3.7	2^{-4}	–	–
$\ell_2-\ell_\infty$-SVM-HET	88.4 ± 11.0	98.0	72.2	9.4 ± 5.2	15.7	2.9	2^{-5}	2^4	2^5
$\ell_1-\ell_\infty$-SVM-HET	94.2 ± 5.0	97.6	85.8	11.4 ± 4.8	15.7	4.1	2^{-7}	2^{-4}	2^{-2}
ℓ_2-SVM-HET	81.7 ± 15.2	97.9	61.9	10.1 ± 5.0	15.7	3.5	2^{-5}	–	2^5
ℓ_1-SVM-HET	93.8 ± 5.5	97.6	84.6	11.5 ± 4.7	15.7	4.2	2^{-7}	–	2^{-2}
$\ell_2-\ell_\infty$-SVM	83.2 ± 13.7	93.7	61.8	9.2 ± 5.2	16.0	3.0	2^1	2^6	–
$\ell_1-\ell_\infty$-SVM	94.1 ± 7.4	99.6	81.9	10.5 ± 5.3	16.0	3.1	2^0	2^{-6}	–

device use causes to the individuals in terms of discomfort due to the physical obtrusiveness.

4.8. Results

Fig. 6 presents the performance in terms of AUCx100 (Fig. 6(a)) and the relative variable collection cost of the solution (Fig. 6(b)) for an increasing number of selected groups. The graphs are designed to be read backwards: the first solution for each method is that with five groups of features, which usually achieves the largest AUC. Then, the backward elimination process starts, reducing the accuracy yet also decreasing the costs. We look for an accurate solution at a low cost. These curves can be regarded as Pareto-optimal solutions in terms of balancing the predictive accuracy and signal acquisition costs.

Fig. 6 shows that all methods are approximately equally effective in terms of their predictive performance if five sources of signals are used; however, this scenario also entails the largest signal acquisition cost. After the first group of features is removed, the AUC values of most methods remain close to 1, with the exception of ℓ_2-SVM, which fails at constructing accurate models if these signals are removed. This result can be explained by the lack of a regularizer that encourages feature selection for ℓ_2-SVM. If one or two groups are used, it is observed that the two proposals are the best approaches in terms of AUC, and $\ell_1-\ell_\infty$-SVM-HET clearly outperforms all other methods. Considering variable acquisition costs, $\ell_2-\ell_\infty$-SVM-HET is the most effective method at prioritizing the source costs of signals, although all methods behave relatively similarly from the perspective of this measure.

Table 1 presents the best performance in terms of AUCx100 and signal acquisition costs for all methods (showing the average for all participants). For each method, the mean feature selection performance is presented with its corresponding standard deviation (MEAN ± STD), which is obtained by averaging the five results presented in Fig. 6. Additionally, the best and worst results in terms of AUC (MAX and MIN, respectively) are reported. Note that the worst performance in terms of AUC always results from using only one source of features, which is also the scenario with the minimum signal acquisition costs. Finally, the optimal hyperparameter configuration is also reported in Table 1 for all methods. For completeness, we also study the performance of our proposals if one of the two new objectives is not considered. That is, we set $\theta = 0$, i.e., no heterogeneity control is applied, leading to the $\ell_2-\ell_\infty$-SVM and $\ell_1-\ell_\infty$-SVM models, and we also set $C_1 = 0$, i.e., no grouped feature selection is performed, leading to the ℓ_2-SVM-HET and ℓ_1-SVM-HET methods.

Table 1 shows that the best overall performance is achieved with $\ell_1-\ell_\infty$-SVM-HET which outperforms all alternative approaches in terms of the average AUC, while the average costs are relatively similar. The other proposed method $\ell_1 - \ell_\infty$-SVM-HET achieves the second best average AUC with the lowest costs, also resulting in an interesting alternative for classification.

4.9. Discussion

A comparison between the use of the full proposal with the four objectives ($\ell_p - \ell_\infty$-SVM-HET with $p = \{1, 2\}$) and methods that consider three objectives (ℓ_p-SVM-HET and $\ell_p - \ell_\infty$-SVM) leads us to conclude that the balance between the former always outperforms the latter.

Similarly, the methods with three objectives are also an improvement over those that consider only two objectives. The most important improvement is achieved with the inclusion of the ℓ_∞-norm, followed by the use of heterogeneity control. Another important conclusion is that the ℓ_1-norm is more effective than the ℓ_2-norm when feature selection is performed. This result can be explained by the ℓ_1-norm being designed for this task, unlike the ℓ_2-norm, and being more compatible with the ℓ_∞-norm in its structure, since both strategies require the computation of the absolute values of the weights.

Another notable observation is that the ℓ_1-norm-based models are more computationally efficient than their ℓ_2-norm-based counterparts since the former can be cast into linear models, while the latter are QPPs. Linear models can usually be solved faster than quadratic models. Therefore, we recommend the $\ell_1 - \ell_\infty$-SVM-HET method for applications in which the balance between predictive performance and collection cost is important.

The preceding experiments imply that it is possible to classify mental workload with accuracy up to and indeed close to 100%, even if a subset of the psychophysiological signals is available from devices. That was the first objective of the study, as previously mentioned.

Considering the second objective, we use the backward elimination approach on the $\ell_1 - \ell_\infty$-SVM-HET method to obtain the relevance of each device. This method is chosen since it produces the best average performance in terms of AUC (see Table 1) and the best predictive performance (also in terms of AUC) using few sources of attributes (see Fig. 6(a)). All the remaining methods perform poorly if only one device is considered.

The feature relevance is obtained as follows: for each participant, the backward elimination approach leads to a ranking of devices based on the elimination order (the first removed device has a rank of 5, the second removed device a rank of 4, and so on). Then, an average rank is computed as the mean for all participants. Note that this procedure is only performed for illustrative purposes; our proposals do not require this backward elimination step, and the results reported in the previous section were obtained without backward elimination. This procedure leads to the following ranking in the decreasing order of importance:

1. HRV, measured with the ECG device, has an average rank of 1.44. This is the most important signal in this study according to the proposed model. The relative cost of this device was estimated to be 3.
2. PS has an average rank of 2.52. The relative cost of the device that captures this signal was estimated to be 10.
3. Temperature has an average rank of 3.20. The relative cost of the respective device was estimated to be 1.
4. HR has an average rank of 3.58. The relative cost of the respective device was estimated to be 1.
5. SpO_2 has an average rank of 4.22. The relative cost of the respective device (the same as that used to measure HR) was estimated to be 1.

The previous results imply that the importance of various devices varies with the participant. Nevertheless, HRV is usually in the first or the second position according to the rank for various participants, leading to an average of 1.44. Note that even though this approach is biased to favor the least expensive devices, the most expensive devices appear in the first and second places. The reason is that the validation procedure was performed using AUC as the main performance metric, and the previous ranking was obtained with the parameter configuration that maximized this measure. It is possible to increase the penalty parameter λ to encourage the least expensive devices to appear in the leading positions of the rank, although it will lead to a loss in performance.

Relation to other SVM classification methods

Various approaches to SVM-based feature selection are already available. We refer the reader to the book by Guyon et al. (2006). The SVM-RFE approach and other iterative search methods that it covers differ in regard to the feature selection methodology. The proposed technique is an embedded feature selection method, i.e., it directly obtains a variable subset that represents an attempt to improve classification performance simultaneously with minimizing dimensionality, and does not rank variables according to their contribution. Iterative search strategies, also known as wrapper methods, are usually more computationally demanding than embedded methods, such as that presented in this study.

As explained, a device generates several signals; therefore, feature selection cannot be performed at the variable level if the main goal is to reduce variable acquisition costs. Variables that belong to the same device must be penalized jointly to guarantee that the entire source of information is not taken into account in the modeling process (Yuan & Lin, 2006). Current research in SVM-based feature selection offers some alternatives for performing grouped variable selection, including the use of mixed-integer programming approaches (Maldonado, Pérez, Labbé, & Weber, 2014). However, the inclusion of binary variables leads to a more complex optimization scheme than that for a group penalty function for the weight vector (Maldonado et al., 2014).

The use of a group penalty function for cost-based feature selection was proposed in Maldonado, Bravo et al. (2017) for the task of credit scoring. However, this application does not deal with panel data as a series of signals; it only solves a single model for all the applicants. The estimation of the collection costs is also very different and involves only monetary costs while excluding factors such as the obtrusiveness of a given device from the perspective of the users.

An additional interesting challenge that arises in this study is the modeling of heterogeneity in participants' behaviors (Evgeniou et al., 2007). When per-participant models are estimated with little individual information, general patterns from all participants can be leveraged to construct robust predictors (Evgeniou et al., 2007; López et al., 2017). The cost-based model proposed in Maldonado, Bravo et al. (2017) does not control for the heterogeneity among various credit applicants.

Heterogeneity control has been studied in conjoint analysis (Green & Rao, 1971); however, this task is very different from binary classification; therefore, the models have a different structure. In López et al. (2017), an SVM model was proposed for deriving various utility functions of the respondents using a heterogeneity control scheme similar to that proposed in this paper. However, the above study's method does not consider feature selection. In Maldonado, Montoya et al. (2017), feature selection was included in the modeling process of conjoint analysis. However, both regularization and heterogeneity control strategies differ from those of our proposal. Furthermore, all attributes are assumed to have the same cost in both studies.

Finally, the current trend in physiological signal analysis involves the use of deep learning methods (Faust, Hagiwara, Hong, Lih, & Acharya, 2018). Deep neural networks have outperformed convex methods, such as SVMs, mostly because of their ability to perform automatic signal preprocessing (Faust et al., 2018). This is the main limitation of our approach since it requires transforming the signals to produce positive results. However, the main strengths of SVM classification are the ability to incorporate additional objectives, such as heterogeneity control and cost-based feature selection, and the absence of local minima via convex optimization. Due to using this multiobjective scheme, we are able to gain important insights into the application. Finally, black-box modeling, such as that performed by deep neural networks, is prone to overfitting, especially if few data points are available. Our goal is to construct user-level models even though limited information is available for each user. To the best of our knowledge, no strategy for heterogeneity control has been proposed for deep neural networks.

5. Conclusions and directions for future work

In this study, two novel SVM-based methods are presented for the task of feature selection and binary classification. Our proposals are designed for tasks in which (i) data are collected from various persons but the available individual information may be scarce and in which (ii) data are collected from various sources of attributes with heterogeneous acquisition costs. Considering the first issue, we extend the ideas of Evgeniou et al. (2007) and Maldonado, Montoya et al. (2017), who propose controlling heterogeneity in a conjoint analysis by minimizing the deviations between the individual functions and a general population pattern. Considering the second aspect, we follow the reasoning behind group penalty functions (Yuan & Lin, 2006) to design a strategy that penalizes the use of the entire data source to which the costs are attributed, rather than eliminating features independently.

From the expert systems' point of view, the proposed approach to simultaneous SVM classification and feature selection opens the possibility of developing an intelligent system capable of automatically assessing the mental workload of mobile users continuously, unobtrusively, and at low cost. Moreover, this assessment can facilitate the development of personal cognitive assistants to help users more effectively manage their multitasking while using mobile devices and performing daily routines. To our knowledge, the proposed approach and the chosen case study constitute the first attempt to bring together both the flexibility of SVM in simultaneous feature selection and the continuous assessment of mental workload.

The proposals were applied to a case study of assessing mental workload through psychophysiological signals. Our experiments demonstrate the advantages of the proposed approach which achieves a superior predictive performance compared with well-known SVM formulations if some devices (data sources) are removed. Additionally, our method achieves a positive performance even if a single device is used for each participant, leading to a reliable relevance ranking of various psychophysiological signals.

Some managerial insights into the application can be gained that could be useful for the design of expert systems based on psychophysiological signals from multiple devices. On the one hand, the feature ranking derived by the proposed method identifies signals that are redundant or that should be avoided due to their cost and obtrusiveness. For example, the analysis presented in the previous section suggests that the SpO_2 signals are almost irrelevant to the task. On the other hand, the Pareto curves shown in Fig. 6 allow modeling the trade-off of using expensive/obtrusive devices, such as the eye tracker, and the respective impact on performance.

Many other practical applications that involve assessing mental workload with psychophysiological signals are possible with the proposed approach. Examples include adapting web interfaces to users' changing mental workload (Jiménez-Molina et al., 2018), evaluating cognitive capabilities, delivering notifications (Bailey & Iqbal, 2008), and measuring drivers' alertness (Yoshida et al., 2014), among other tasks.

Furthermore, our proposal can be applied to tasks in several domains, for example, in medical diagnoses in which several signals are collected per patient. Respiratory diseases, such as obstructive sleep apnea syndrome (OSA), require adequate diagnosis data in the form of signals from various sources, such as a nasal airflow sensor or an ECG, and each test has a different cost and obtrusiveness (Ríos & Erazo, 2016). Our proposal can also be useful in business analytics, where the cost of acquiring data can be explicitly computed. Credit scoring is one example of an application for which several data sources are available, including potentially expensive data bought from credit bureaus that can be penalized in the modeling process (Maldonado, Pérez et al., 2017).

This study opens interesting possibilities for future development. As previously mentioned, our proposal can be applied in other contexts in which data from different sources are collected. Additionally, our approach can be extended to kernel methods. Although we believe that the use of the ℓ_∞-norm leads to a dual formulation to which the kernel trick cannot be applied, we in fact believe that our approach without feature selection (yet including heterogeneity control) can be extended as a kernel method (see the supplementary material for a detailed derivation of the dual form for $\ell_2 - \ell_\infty$-SVM-HET and a discussion of this topic). Finally, SVM has been adapted to functional data, which is an interesting approach if the covariates can be represented as a time series (Rossi & Villa, 2006). Our approach could be extended to functional SVM classification by penalizing the use of expensive functional data during SVM training.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Sebastián Maldonado: Conceptualization, Data curation, Formal analysis, Writing - original draft. **Julio López:** Conceptualization, Formal analysis, Writing - original draft. **Angel Jimenez-Molina:** Conceptualization, Data curation, Writing - original draft. **Hernán Lira:** Data curation, Formal analysis, Writing - original draft.

Acknowledgments

The first author was supported by FONDECYT project 1160738, the second author was supported by FONDECYT project 1160894, and the third author was supported by FONDEF project ID16I10222. The first and third author were supported by FONDECYT project 1181809. This research was partially funded by the Complex Engineering Systems Institute, ISCI (CONICYT PIA/BASAL AFB180003). The authors are grateful to the anonymous reviewers, who contributed to improving the quality of the original paper, and all the experiment participants.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eswa.2019.112988.

References

Addison, P. S. (2005). Wavelet transforms and the ecg: A review. *Physiological Measurement, 26*(5), R155.

Albers, M. (2011). Tapping as a measure of cognitive load and website usability. In *Proceedings of the 29th ACM international conference on design of communication sigdoc 2011* (pp. 25–32). ACM Press.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI), 14*(4), 21.

Bradley, P., & Mangasarian, O. (1998). Feature selection via concave minimization and support vector machines. In *Machine learning proceedings of the fifteenth international conference (ICML'98)* (pp. 82–90). San Francisco, California: Morgan Kaufmann.

Cacioppo, J., Tassinary, L., & Berntson, G. (2007). *The handbook of psychophysiology* (3rd). Cambridge University Press.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*. 27:1–27:27 Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chapelle, O., & Keerthi, S. (2008). Multi-class feature selection with support vector machines. In *Proceedings of the American statistical association, ASA, Denver, CO, USA, 3–7 August (on cd-rom).*.

Chen, S., & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine, 110*(2), 111–124.

Chi, C.-F., & Lin, F.-T. (1997). A new method for describing search patterns and quantifying visual load using eye movement data. *International Journal of Industrial Ergonomics, 19*(3), 249–257.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Evgeniou, T., Boussios, C., & Zacharia, G. (2005). Generalized robust conjoint estimation. *Marketing Science, 24*(3), 415–429.

Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling heterogeneity in conjoint estimation. *Marketing Science, 26*(6), 805–818.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin. , C.-J. (2008). Liblinear: A library for large linear classification. *Journal of Machine Learning Research, 9*, 1871–1874.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine, 161*, 1–13.

Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., & Züger, M. (2014). Using psycho-physiological measures to assess task difficulty in software development. In *Proceedings of the 36th international conference on software engineering* (pp. 402–413). ACM.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology, 83*(3), 269–275.

Grant, M., & Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research, 8*, 355–363.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2006). *Feature extraction, foundations and applications*. Springer, Berlin.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on ubiquitous computing* (pp. 301–310). ACM.

Haddad, S. A. P., & Serdijn, W. A. (2009). *Ultra low-power biomedical signal processing: An analog wavelet filter approach for pacemakers*. Springer Science & Business Media.

Hart, S. (1988). In P. A. Hancock, & N. Meshkati (Eds.), *Development of NASATLX (Task Load Index): results of empirical and theoretical research* (p. 139-183.). Human Mental Workload. Amsterdam, North-Holland.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Ikehara, C. S., & Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. *System sciences, 2005. HICSS'05. proceedings of the 38th annual Hawaii international conference on*. IEEE. 295a–295a

Jiang, S., Chin, K.-S., Wang, L., Qu, G., & Tsui, K. L. (2017). Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. *Expert Systems with Applications, 82*, 216–230.

Jiménez-Molina, A., & Ko. , I.-Y. (2015). Cognitive resource-aware unobtrusive service provisioning in ambient intelligence environments. *Journal of Ambient Intelligence and Smart Environments, 7*, 37–57.

Jiménez-Molina, A., Retamal, C., & Lira, H. (2018). Using psychophysiological sensors to assess mental workload during web browsing. *Sensors, 18*(2), 458.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication UbiComp '14 Adjunct* (pp. 1151–1160).

Khandoker, A. H., Gubbi, J., & Palaniswami, M. (2009). Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings. *IEEE Transactions on Information Technology in Biomedicine, 13*(6), 1057–1067.

Lee, U., Lee, J., Ko, M., Lee, C., Kim, Y., Yang, S., et al. (2014). Hooked on smartphones: An exploratory study on smartphone overuse among college students. In *Proceedings of the 32nd annual ACM conference on human factors in computing systems* (pp. 2327–2336). ACM.

Lo, J. C., Sehic, E., & Meijer, S. A. (2017). Measuring mental workload with low-cost and wearable sensors: Insights into the accuracy, obtrusiveness, and research usability of three instruments. *Journal of Cognitive Engineering and Decision Making, 11*(4), 323–336.

López, J., & Maldonado, S. (2017). Group-penalized feature selection and robust twin svm classification via second-order cone programming. *Neurocomputing, 235*, 112–121.

López, J., Maldonado, S., & Montoya, R. (2017). Simultaneous preference estimation and heterogeneity control for choice-based conjoint via support vector machines. *Journal of the Operational Research Society, 68*(11), 1323–1334.

Maldonado, S., Bravo, C., López, J., & Pérez, J. (2017). Integrated framework for profit-based feature selection and svm classification in credit scoring. *Decision Support Systems, 104*, 113–121.

Maldonado, S., Montoya, R., & López, J. (2017). Embedded heterogeneous feature selection for conjoint analysis: A svm approach using l1 penalty. *Applied Intelligence, 46*(4), 775–787.

Maldonado, S., Pérez, J., & Bravo, C. (2017). Cost-based feature selection for svm classification - An application in credit scoring. *European Journal of Operational Research, 261*(2), 656–665.

Maldonado, S., Pérez, J., Labbé, M., & Weber, R. (2014). Feature selection for support vector machines via mixed integer linear programming. *Information Sciences, 279*, 163–175.

Matthews, G., Reinerman-Jones, L., Barber, D., & Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors, 57*(1), 125–143.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, 214–255.

Nourbakhsh, N., Wang, Y., & Chen, F. (2013). Gsr and blink features for cognitive load classification. In *IFIP conference on human-computer interaction* (pp. 159–166). Springer.

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015). Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 475–486). ACM.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 919–928). ACM.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4.

Paas, F., & Van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737–743.

Parsons, H. M. (1974). What happened at hawthorne? *Science, 183*(4128), 922–932.

Ríos, S., & Erazo, L. (2016). An automatic apnea screening algorithm for children. *Expert Systems with Applications, 48*, 42–54.

Rossi, F., & Villa, N. (2006). Support vector machine for functional data classification. *Neurocomputing, 69*, 730–742.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics, 35*(11), 991–1009.

Samaha, M., & Hawi, N. S. (2016). Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. *Computers in Human Behavior, 57*, 321–325.

Seret, A., Maldonado, S., & Baesens, B. (2015). Identifying next relevant variables for segmentation by using feature selection approaches. *Expert Systems With Applications, 42*(15–16), 6255–6266.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (gsr) as an index of cognitive load. In *Chi'07 extended abstracts on human factors in computing systems* (pp. 2651–2656). ACM.

Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis : Evaluation for cardiovascular diseases. *Expert Systems with Applications, 40*(10), 4146–4153.

Smith, J., & Dulay, N. (2014). Ringlearn: Long-term mitigation of disruptive smartphone interruptions. In *Pervasive computing and communications workshops (percom workshops), 2014 IEEE international conference on* (pp. 27–35). IEEE.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tibshirani, R., & Hastie, T. (2009). *The Elements of Statistical Learning*. Springer.

Van Deursen, A. J., Bolle, C. L., Hegner, S. M., & Kommers, P. A. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in Human Behavior, 45*, 411–420.

Wang, J.-L., Wang, H.-Z., Gaskin, J., & Wang, L.-H. (2015). The role of stress and motivation in problematic smartphone use among college students. *Computers in Human Behavior, 53*, 181–188.
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177.
Yoshida, Y., Ohwada, H., Mizoguchi, F., & Iwasaki, H. (2014). Classifying cognitive load and driving situation with machine learning. *International Journal of Machine Learning and Computing, 4*(3), 210.
Young, M., Brookhuis, K., Wickens, C., & Hancock, P. (2015). State of science: Mental workload in ergonomics. *Ergonomics, 58*, 1–17.
Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B, 68*, 49–67.
Zou, H., & Yuan, M. (2008). The f-infinite norm support vector machine. *Statistica Sinica, 18*, 379–398.