Random Forest-Based Approach for Physiological Functional Variable Selection: Towards Driver's Stress Level Classification

Neska El Haouij, Jean-Michel Poggi, Raja El Ghozi, Sylvie Sevestre-Ghalila, Mériem Jaïdane

Random Forest-Based Approach for Physiological Functional Variable Selection: Towards Driver's Stress Level Classification

Neska El Haouij[1,2,3], Jean-Michel Poggi[3,4], Raja Ghozi[1,2], Sylvie Sevestre-Ghalila[2,4] and Mériem Jaïdane[1,2]

[1] *Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, Unité Signaux et Systèmes, 1002 Tunis, Tunisia*

[2] *CEA-LinkLab, Telnet Innovation Labs, 2083 Ariana, Tunisia*

[3] *Université de Paris-Sud, Laboratoire de Mathématique d'Orsay, 91400 Orsay, France*

[4] *Université de Paris-Descartes, 75006 Paris, France*

elhaouij.nsk@gmail.com, Jean-Michel.Poggi@math.u-psud.fr, rjghozi@yahoo.com, Sylvie.SEVESTRE-GHALILA@cea.fr, meriem.jaidane@planet.tn

Keywords: Physiological signals, Functional data, Random forests, Recursive feature elimination, Wavelets, Grouped variable importance

Abstract: This paper is devoted to a statistical physiological functional variable selection for driver's stress level classification using random forests. Indeed, this study focuses on humans physiological changes, produced when driving in different urban routes, captured using portable sensors. Specifically, the electrodermal activity measured on two different locations: hand and foot, electromyogram, heart rate and respiration of ten driving experiments in three types of routes: rest area, city, and highway driving issued from *drivedb* database, available online on the PhysioNet website. Several studies were achieved on driver's stress level recognition using physiological signals. Classically, researchers extract expert-based features from physiological signals and select the most relevant ones for stress level recognition.

This work provides a random forest-based method for the selection of physiological functional variables in order to classify the driver's stress level. On the methodological side, the contributions of this work are to consider physiological signals as functional variables, decomposed on wavelet basis and to offer a procedure of variable selection. On the applied side, the proposed method provides a "blind" procedure of driver's stress level classification performing as the expert-based study in terms of misclassification rate. It offers moreover a ranking of physiological variables according to their importance in stress level classification. The obtained results suggest that electromyogram and heart rate signals are not very relevant when compared to the electrodermal and the respiration signals.

1 Introduction

This paper aims to provide a random forests-based method for the selection of physiological functional variables in order to classify the stress level experienced during real-world driving. For that, we present first the context of our work which concerns the affective computing aspects with a summary of the study introducing the physiological database *drivedb*. Then, methods on functional data, variable selection using random forests and grouped variables importance are addressed.

Stress level recognition while driving

Nowadays, the demand for affective computing is important especially in heath care and education. Although Bostrom (2005) reports that Picard which is the creator of the affective computing field (Kleine-Cosack (2006)) predicted that the first truly application of the affective computing to products will be in the automotive industry. Affective Computing is an interdisciplinary field combining computer science and engineering with cognitive science, physiology, psychology. According to Tao and Tan (2005), researches in this field built models based on sensors-captured data, and construct sensitive computing systems, able to perceive, to interpret human feeling and to provide smart responses to humans requests. Many research groups tried to provide solutions and tools to vehicles and roadway users in order to improve safety, efficiency and quality in the sector of transport. Smart et al. (2005) points out that according to the American Highway Traffic Safety Administration, high stress levels impact negatively drivers reactions especially in critical situations. It is one of the most prominent causes of vehicle accidents such as intoxication, fatigue and aggressive driving. In real world driving, human affective state monitoring can offer useful information to avoid traffic incidents and provide safe and comfortable driving.

With the increasing urbanization and technological advances, the new wearable and non-intrusive sensor technology, is not only providing real-time physiological monitoring, but also is enriching the tools for human affective and cognitive states monitoring. In particular, several studies have been reported the last years in the field of driver's stress monitoring. In this paper, base our analysis on the study of Healey and Picard (2005) where they presented a protocol of physiological data collection in real-world driving conditions in order to detect stress levels. Specifically, physiological signals such as Electrodermal Activity (EDA), Electrocardiogram (ECG), Electromyogram (EMG) and Respiration (RESP) were captured for 24 driving experiences. Features derived from non-overlapping segments of physiological signals taken from rest, highway and city of the driving experiences. The first analysis aiming to classify the stress levels allows to distinguish between the three levels of driver stress with an accuracy of 97%. The second analysis concerns the study of the correlation between extracted features from physiological signals and a stress levels metric created from the video tape. In this study, Healey and Picard (2005) reported that there is a correlation between driver's affective state quantified by the stress levels metric and the physiological signals, the highest correlation is with the EDA and HR. They have partially released their physiological database, labeled "*drivedb*", on-line on the PhysioNet website[1]. The data used in our work were extracted from the *drivedb* database which has a clear annotation of the several driving periods for each experience, allowing an easy exploitation of the information. Apart its availability on-line, various studies were based on this database which constitutes a main reference on stress level recognition in highway and city driving.

Variable Selection and functional data

The main issue of variable selection methods is their instability where a set of selected variables may change when perturbing the training sample. The most widely used solution to solve this instability consists in using bootstrap samples where a stable solution is obtained by aggregating selections achieved on several bootstrap subsets of the training data. Random forests algorithm, introduced by Breiman (2001), is one of these methods based on aggregating a large collection of tree-based estimators. These methods have good predictive performances in practice and they work well for high dimensional problems. Their power is shown in several studies summarized in Verikas et al. (2011). Moreover, random forests provide sseveral measures of the importance of the variables with respect to the prediction of the outcome variable. It has been shown that the permutation importance measure introduced by Breiman, is an efficient tool for selecting variables (Díaz-Uriarte and Alvarez de Andrés (2006); Genuer et al. (2010); Gregorutti et al. (2016)).

Functional Data Analysis (FDA) is a field that analyzes data generally indexed by time (see for example Ramsay

[1] http://physionet.org/

and Silverman (2005); Ferraty and Vieu (2006)). The standard approach in FDA consists in projecting the functional variables into a space spanned by a functional basis such as splines, wavelets, Fourier. Several regression and classification methods were the focus of studies in two situations: with one functional predictor and recently for several functional variables. Classification based on several possibly functional variables has also been considered using the CART algorithm for similar driving experiences in the study of Poggi and Tuleau (2007), using SVM in Yang et al. (2005) work. Variable selection using random forests was achieved in the study of Genuer et al. (2015). In our study, multiple FDA using random forests and the grouped variable importance measure proposed by Gregorutti et al. (2015) will be used.
The contribution of this study is twofold: on the methodological side, it takes advantage of the functional nature of the physiological data and offers a procedure of data processing and variable selection. On the applied side, the proposed method provides a blind (i.e. without prior information) procedure of driver's stress level classification that does not depend on the extraction of expert-based features of physiological signals. This allows automatic exploration of promising signals to be included in statistical models for driver's state recognition.

Paper Outline
This paper is organized as follows. After this introductory section, Section 2 is dedicated to the description of the database used in this study. Section 3 recalls a random forests-based procedure for functional variables selection. Section 4 presents results of the method on real world driving experiences. Finally Section 5 presents a discussion and some perspectives for future work.

2 Experimental protocol and data collection

Detecting stress in a timely fashion offers the opportunity to better understand the causes of this state and provides specialists with more relevant data in order to intervene in the best moment and monitor stress. Several studies have shown that physiological data recording and analysis can help in stress recognition during driving tasks. A brief list of works on the driver stress recognition is summarized in the Table 1. For each study, reference, objectives, statistical methods and physiological signals are listed. The first row of Table 1 corresponds to the study of Healey and Picard (2005) that provides the *drivedb* database. The second and the fourth rows correspond to two studies that used *drivedb* in the stress level identification.
Singh and Queyam (2013) reported that several studies have been done on driver's stress level recognition using physiological signals, but little work has been focused on feature selection and on studying the correlation between selected features and the driving route complexity. Moreover, all listed studies use expert-based features extracted from physiological signals.
In this section, the driving protocol will be first detailed, then the cohort will be described leading to the data acquisition, database construction and the model description.

2.1 Real-world driving protocol

In this study, we use signals from the *drivedb* database corresponding to the only 10 available driving experiences. The initial driving experiment was composed of 24 drives. The protocol is detailed in Fig.1.
Before each experiment, drivers were shown a map of driving route. Each driver was accompanied by an observer who sat in the rear seat in order to avoid interfering with the driver's natural behavior. This person has to ensure the event marking in the video records, the integrity of physiological signals and driver's question answering. The total time duration of each drive depends on traffic congestion, it varies from 50 minutes to 1 hour and half. All driving experiments began with a rest period which lasted over fifteen minutes on a garage. After rest the driver exited through a narrow ramp until reaching a congested city street. The drive through this main street is assumed to produce a high stress level due to its several stop and go traffic and faced unexpected hazards. The path continued from the city to an interrupted highway driving supposed to produce a medium stress state. After highway exit, the trip led driver to a turn around and re-enter the highway in the opposite direction. After this second highway driving segment, the driver re-entered the same main street in the city and exited to the garage. The drive finished by the same rest period as it began. Each participant was asked to seat in the garage,

Table 1 Summary of some studies achieved on driver's stress recognition

Reference	Objectives	Statistical Methods	Signals
Healey (2000)	Collect and analyze physiological data during real-world driving tasks to determine driver stress level	Linear Discriminant Analysis	ECG, EMG, EDA, RESP
Akbas (2011)	Identify the useful metrics indicating the stress level of drivers	Statistical mean comparison	ECG, EMG, EDA, RESP
Rigas et al. (2008)	Estimation of car drivers stress produced due to specific driving events	Dynamic Bayesian Network	EDA, ECG
Deng et al. (2012)	Features extraction and selection for stress level recognition	PCA and Classification algorithms	ECG, EMG, EDA, RESP

eyes closed and with the car in idle during each rest period. These two periods were used to create a low stress situation and to provide baseline measurements.

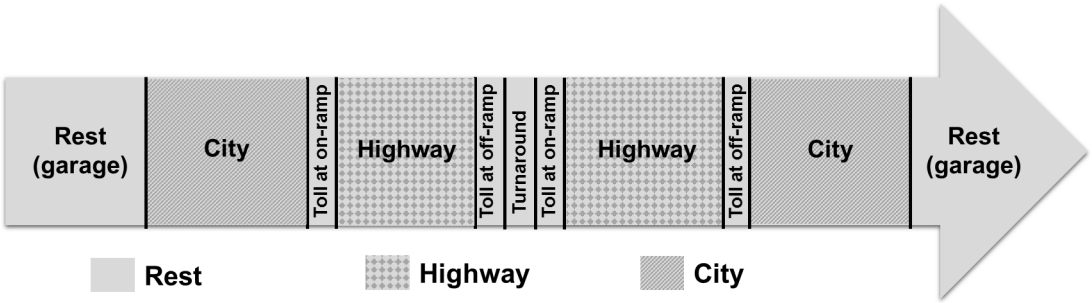


Fig. 1 Driver's path description of the each driving segment

The carried out protocol was consisted of a path which extended over 20 miles of open roads in the greater Boston area. Along suchpath, the driver was asked to follow a set of instructions in order to keep the drive uniform (e.g. respect speed limits, not to listen to radio). The path was chosen as a typical daily commute so reactions revealing stress would all be within the range of normal daily stress. Each drive mainly included periods of rest, highway and city driving assumed to produce respectively low, medium and high levels of stress. These assumptions were validated by two methods described in Healey and Picard (2005) article mainly using a driver questionnaire and a score derived from observable events and actions coded from video tape recorded during each drive.

2.2 Cohort description

We recall that for each of the drives, several physiological signals namely Electrodermal activity measured on two different placements: foot and hand, Electomygram signals, Respiration and Heart Rate from 10 different driving experiences were taken from open dataset available in PhysioBank (proposed by Goldberger et al. (2000)) titled *Stress Recognition in Automobile Drivers* labeled *drivedb*.
The set of data used in our work refer to 10 driving experiences achieved by 4 participants. The organization of the different driving experiences are detailed in Table 2. In the thesis of Healey (2000), participants profile is

listed. Participant M-3 contributed in the database by 2 experiences, he was an undergraduate male, having three years of driving experience, but he had not driven regularly for the three past years. The subject labeled M-4 did not find driving stressful, since he had over four years of driving experience. He was an undergraduate student. He had not driven a month previous to the experiment and the first drive in the experiment was his second drive in Boston area. F-8 was a female undergraduate having eight years of driving experience. She deviated from the itinerary on the first driving run. Finally, Ind 4 participated in the experiment by just one driving experience. Details concerning the Ind4 profile are not provided.

Table 2 Description of the different drives corresponding to 10 sets of signals.
Note that the label of the participant is composed by the gender (M:Male, F:Female)-driving experience (years), except the participant 4 labeled Ind 4, without information

Drive	Participant label	Date (mm-dd-yy)	Duration (hh:mm:ss)
1	M-3	07-28-99	1:24:15
2		08-04-99	1:20:46
3	M-4	07-15-99	1:28:38
4		08-05-99	1:21:11
5		08-13-99	1:10:52
6	F-8	08-02-99	1:21:16
7		08-05-99	1:21:13
8		08-06-99	1:23:04
9		08-09-99	1:17:38
10	Ind 4	07-16-99	1:04:57

2.3 Data Construction

2.3.1 Segments Extraction

All the physiological signals available on-line on the database, were stored at a sampling frequency $Fs = 15.5Hz$. From each drive, seven segments of 5 minutes were extracted as shown in the Fig. 2. For the standardization purpose and low stress level construction, segments from the last 5 minutes of the two rest period are taken in order to give participant enough time to relax from the previous task: putting the sensors for the first period and driving for the last rest period.
Two segments from the middle of highway driving periods and three from the middle of City driving are extracted in order to offer respectively medium and high stress levels. We took the segments from the middle of the periods in order to avoid both the memory effect and the anticipatory task. For the drive 5 and 10, the annotation of the last rest period is missing, thus the two segments couldn't be used in this study.
To sum up, the database extracted and used in this study is composed of 68 curves corresponding to 10 drives achieved by 4 participants. We finally note that inter-individual effect is not considered.

2.3.2 Physiological data preprocessing

The procedure of data standardization refers to data corrected to ensure that it can be comparable between individuals. For instance, when all EDAs are standardized, according to Dawson et al. (2007), the comparison between individuals differences is achieved according to only their physiological responsiveness unrelated to psychological processes such as skin thickness. There are several techniques for EDA standardization. The most common method is based on range correction proposed by Lykken et al. (1966). The procedure is based on computing the individual possible range of EDA and then expressing each value in terms of this range. The EDA standardization is achieved by centering the EDA using the minimum value during the first rest period, then reducing iusing the difference between the maximum of the signal during the entire day and minimum value during the first rest

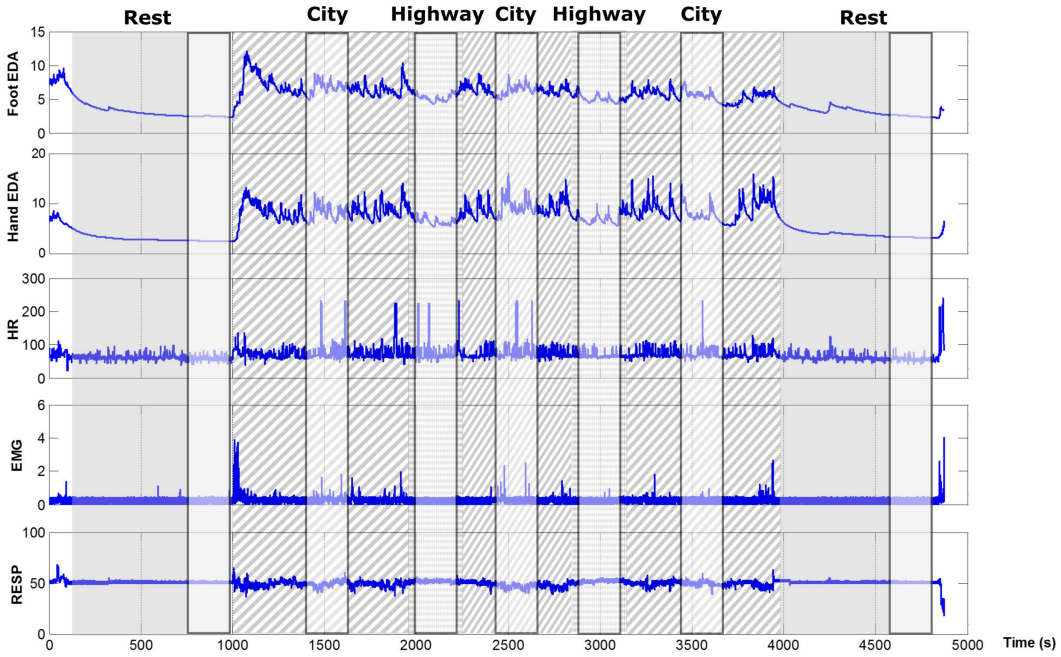


Fig. 2 Segments extraction from signals corresponding to the drive 7

period.
Lykken (1972) shows in his work that there is no significant difference when normalizing the HR using the range correction. Thus, we just center the HR by subtracting from each sample the mean value of the heart rate over the first rest period.
As to the EMG signals, they were first transformed using the absolute values transformation then centered by subtracting from each sample the mean value of the heart rate over the first rest period.
Finally, the Respiration signals were centered around the mean value of the respiration signal among the first rest period.

2.4 Data modeling and Model description

Let S be the vector of physiological signals used here as explanatory variables, we extract five minutes segments corresponding to the longest segment in the database. The choice of this long time allows the extracted segments to be informative and comparable over the window even if many variations in traffic conditions occurred on a second by second basis. Five minutes corresponds to 4650 samples based on the used sampling period $\Delta t = \frac{1}{F_s} = \frac{1}{15.5} = 0.065s$.
In order to project these segments into a wavelet basis (details presented in next section), we keep the first 2^{12} samples (a power of two to simplify the discrete wavelet decomposition) which corresponds to 4096 samples. $S = (S^1, ..., S^p) \in \mathcal{S}^p$ and $S^j = \{S^j(t) \in \mathbb{R}, t \in [0,T]\} \in \mathcal{S}$ which can be considered as functions of finite energy over $[0,T]$ where $T = 264.26s$ and $j = 1, ..., p$ *where* $p = 5$.
Let i be the index of the extracted segment, $i = 1, ..., N$ *where* $N = 68$.
For a given i, $S_i(t) = (S_i^1(t), ..., S_i^p(t))$ presents 5 physiological signals used as explanatory variables corresponding to the stress level y_i,

$$y_i = \left\{ \begin{array}{c} \text{H=High stress level} \\ \text{M=Medium stress level} \\ \text{L=Low stress level} \end{array} \right\}$$

Let us sketch the statistical model. Let $L = \{(S_1, Y_1), \ldots, (S_n, Y_n)\}$ be a learning set, consisting on n independent observations of the random vector (S, Y). The $(S_i(t), y_i)$ introduced previously are then considered as realization of unknown distribution of this sample. We aim to build an estimator of the Bayes classifier $f : \mathbb{R} \mapsto \mathcal{Y}$ minimizing

the classification error $P(Y \neq f(S))$. We will denote by $\hat{f}$ an estimator belonging to the fully non parametric set of models given by the random forests framework which will be introduced in the next section.

3 Multiple functional data analysis using variable importance

3.1 Functional data and wavelet representation

Let us consider a functions space noted $\mathcal{F}$ and Ω a probabilistic space. A functional random variable is a measurable application $S : \Omega \to \mathcal{F}$. $\mathcal{F}$ is constituted by functions defined on $[0,1]$ ($[0,T]$ equivalently in our specific situation) with values in $\mathbb{R}$.
Let us consider a sequence of functions ϕ and ψ obtained by translations and dilations of compactly supported scaling function $\tilde{\phi}$ and a compactly supported mother wavelet $\tilde{\psi}$. To build orthonormal wavelet basis, the discrete wavelet transform (DWT) uses dyadic scales and translations.
Thus, for any $j_0 \geq 0$,

$$\mathcal{B} = \left\{\phi_{j_0,k}, k = 0, \ldots, 2^{j_0} - 1\right\} \bigcup \left\{\psi_{j,k}, j \geq j_0, k = 0, \ldots, 2^j - 1\right\}$$

is an orthogonal basis of $L^2([0,1])$ where $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ and
$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$.
A wavelet transform of a function s of $L^2([0,1])$ is

$$s(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^{j_0}-1} d_{j,k}\psi_{j,k}(t) \tag{1}$$

with $c_{j,k} = \langle s, \phi_{j,k}\rangle_{L^2}$ and $d_{j,k} = \langle s, \psi_{j,k}\rangle_{L^2}$.
A multiresolution analysis of L^2, introduced by Mallat (1989), allows to describe the feature space as a set of nested subspaces V_j associated to the scale levels $j \in \mathbb{Z}$.
The first term in Equation (1) is the smooth approximation of s at level j_0 while the second term is the detail part of the wavelet representation.
For an input signal of length $N = 2^J$ and for the level of wavelet given by the size N of the sampling grid, a wavelet decomposition of s can also be given in a similar form as Equation 1. Thus,

$$\tilde{s}_J(t) = c_0\phi_{0,0}(t) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t)$$

where c_0 is the single scaling coefficient and $d_{j,k}$ denote the empirical wavelet coefficients derived from applying the DWT to the sampled values.

3.2 Random Forests and Variable Importance measure

Random Forests, introduced by Breiman (2001), are part of nonparametric statistical methods allowing to deal with classification and regression problems.
Let us consider $X = (X^1, \ldots, X^p), X \in \mathbb{R}^p$ the explanatory variables and Y a variable of interest. Y is a numerical response in the regression context and a label related to a class in the classification one. Let $L = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ be a learning set, consisting on n independent observations of the vector (X, Y). Random Forests is a technique that builds model providing estimators of either the Bayes classifier $f : \mathbb{R} \mapsto \mathcal{Y}$ minimizing the classification error $P(Y \neq f(X))$, in the classification problem, or the regression function s that verifies $Y = s(X) + \varepsilon$ with $E[\varepsilon|X] = 0$. See Hastie et al. (2001) for more information.
The Classification and Regression Trees (CART) algorithm, defined by Breiman et al. (1984), is a competitive technique to estimate f. A maximal tree is obtained by splitting the learning sample up to the last observations using splitting rules. At each node corresponding to a rule, data is divided into two parts with maximum of homogeneity. Maximal trees may be very complex therefore, they have to be optimized by choosing the right size of tree by cutting off insignificant nodes. This operation is known as the pruning and is achieved by minimizing a

penalized criterion. The problem with decision trees is their instability, indeed a small perturbation of the training sample can change the prediction values.
Thus, Breiman (2001) proposed the random forest as an improvement of the decision trees method. It consists of aggregating a set of random trees, built over n_{tree} bootstrap samples $L^1, \ldots, L^{n_{tree}}$ of the training set L.
At each node, a fixed number of variables is randomly picked to determine the splitting rule among them. The trees are not pruned so all the trees of the forest are maximal trees. The resulting learning rule is the aggregation of all of the estimators resulting from those trees, denoted by $\hat{f}_1, \ldots, \hat{f}_{n_{tree}}$. To make a prediction at a new point x, the aggregation consists of building

$$\hat{f}(x) = \begin{cases} \frac{1}{n_{tree}} \sum_{k=1}^{n_{tree}} \hat{f}_k(x) & \text{in regression} \\ \\ arg \underset{1 \leq c \leq C}{max} \sum_{k=1}^{n_{tree}} 1_{\hat{f}_k(x)=c} & \text{in classification} \end{cases}$$

We should introduce the Out-Of-Bag (OOB) sample that will be used in the definition of the variable importance measure. For each tree, an OOB sample corresponds to the set of observations that are not retained in the bootstrap samples.

3.2.1 Variable Importance Measure

According to Sauvé and Tuleau-Malot (2014) for example, the variable importance allows to assess the contribution of a variable to explain the phenomenon of interest. In the random forests model, the most used score of variable's importance is the increasing in mean of error of a tree when the values of the variable are randomly permuted in the OOB samples. The error of a tree is the Mean Square Error (MSE) for the regression and the misclassification rate for the classification. This score of variable importance is known as permutation variable importance. The RF permutation importance is shown to be a reliable indicator so we restrict our attention to this type of Variable Importance, henceforth, VI.
X^j is considered important if when breaking the link between X^j and Y, the prediction error increases. The prediction error of each tree $\hat{f}$ is evaluated among its OOB sample with the empirical estimator

$$\hat{R}(\hat{f}, \bar{L}) = \begin{cases} \frac{1}{|\bar{L}|} \sum_{i:(X_i,Y_i) \in \bar{L}} (Y_i - \hat{f}(X_i))^2 & \text{in regression} \\ \\ \frac{1}{|\bar{L}|} \sum_{i:(X_i,Y_i) \in \bar{L}} 1_{\hat{f}(X_i) \neq Y_i} & \text{in classification} \end{cases}$$

Let $\left\{ \bar{L}_k^j, k = 1, \ldots, n_{tree} \right\}$ refers to the set of OOB permuted samples resulting from random permutations of the values of the j-th variable in each out-of-bag subset.
The permutation importance measure is thus defined as the mean increase in the prediction error over all the trees

$$\hat{I}(X^j) = \frac{1}{n_{tree}} \sum_{k=1}^{n_{tree}} \left[\hat{R}(\hat{f}_k, \bar{L}_k^j) - \hat{R}(\hat{f}_k, \bar{L}_k) \right]$$

Several authors were interested in the numerical study of this VI (see Strobl and Zeileis (2008); Nicodemus et al. (2010); Auret and Aldrich (2011)). Some theoretical results are also available (see Louppe et al. (2013), Gregorutti et al. (2015)). Zhu et al. (2012) proved that $\hat{I}(X^j)$ converges to $I(X^j)$, where

$$I(X^j) = E[(Y - f(X^{(j)}))^2] - E[(Y - f(X))^2]$$

such that $X^{(j)} = (X^1, ..., X^{(j)}, ..., X^p)$ be the random vector where $X^{(j)}$ is an independent replication of X^j which is independent of Y and of all of the others predictors.

3.2.2 Grouped VI measure

Gregorutti et al. (2015) extends the permutation importance for a group of variables. In order to estimate the permutation importance of a group of variables denoted $\mathbf{X^J}$, let's consider for each $k \in \{1, ..., n_{tree}\}$, $\bar{\mathbf{L}}_k^J$ the permuted set of $\bar{\mathbf{L}}_k$ resulting by randomly permuting the group $\mathbf{X^J}$ in each OOB sample $\bar{\mathbf{L}}_k^j$. The permutation importance of $\mathbf{X^J}$ is estimated by

$$\hat{I}(\mathbf{X^J}) = \frac{1}{n_{tree}} \sum_{k=1}^{n_{tree}} \left[\hat{R}(\hat{f}_k, \bar{\mathbf{L}}_k^J) - \hat{R}(\hat{f}_k, \bar{\mathbf{L}}_k) \right]$$

3.3 Variable Selection using Random Forest-based Recursive Feature Elimination

In this study, Random Forests-based Recursive Feature Elimination (RF-RFE) is used. The RF-RFE algorithm, summarized in the Table 3, proposed by Gregorutti et al. (2015), was inspired from Guyon et al. (2002) introducing Recursive Feature Elimination algorithm for SVM (SVM-RFE).
At the first step, the dataset is randomly split into a training set containing two thirds of the data and a validation set containing the remaining one third. The procedure fits the model to all explanatory variables using Random Forests. Then, the variables are ranked using their importance measure. The grouped VI is computed only on the training set. The less important predictor is eliminated, the model is refit and the performance is assessed by a prediction error computed on the validation set. The variable ranking and elimination is repeated until no variable remains. The final model is chosen by minimizing the prediction error. It should be noted that at each iteration, the predictors importance is recomputed on the model composed by the reduced set of explanatory variables.

Table 3 Summary of the selection algorithm based on RF-RFE

1. Split the whole data L into a training set L_T containing $\frac{2}{3}$ of the data and a validation set L_V containing the remaining $\frac{1}{3}$. Set the subset of the variables $\mathcal{V}$ to the whole explanatory variables.
2. Fit a random forest model using L_T and considering the set of variables $\mathcal{V}$
3. Compute the VI measure (respectively the grouped VI measure)
4. Compute the error using the validation sample L_V
5. Eliminate the least important variable (resp. group of variables) and update $\mathcal{V}$.
6. Repeat 2-5 until no further variables (group of variables) remain
7. Select the variables (resp. the groups of variables) involved in the model minimizing the prediction error

In our case of functional variables, the selection is performed using the algorithm detailed in the Table 3 on two different type of groups. This allows to consider a group of variables as a whole, for example the group of the wavelet coefficients of a given signal, and to quantify its relative importance with respect to the other functional variables.

3.4 Our procedure: Variable selection using iterative RF-RFE

The proposed approach in this work aims to first eliminate the irrelevant physiological variables in the stress level classification task and then select among each kept variable the most relevant wavelet levels.
In this study, the number of variables is very large (20480), compared to the number of the observations (68), thus the procedure is not stable. In order to reduce the variability of the selection, the procedure is repeated 10 times.
Let us denote the procedure detailed at the subsection 3.3 by
RF-RFE $(Grp_1, ..., Grp_P)$ where

- $Grp_j = G(j)$ corresponding to wavelet coefficients grouped by physiological variables $j, j = 1, ..., p$.
- $Grp_j = G(w, k_r)$ corresponding to wavelet coefficients grouped by wavelet level $w, w = 1, ..., J, J = 13$ of the selected functional variable $k_r, r = 1, ..., R$

Remark: Of course the selection depends on the choice of δ and we propose here a classical trick: we look at the consecutive differences of the selection score. When starting from the highest values of the selection scores, we select variables just before the biggest gap or when starting from the lowest values of selection scores, eliminate variables just after the big gap. Many other automatic proposals are possible but are of the same flavor (for example the elbow criterion that allows to determine the number of factors in the PCA). In addition, if the number of variables is small, a visual inspection is also possible.

Table 4 Summary of the 3-steps proposed approach

Step1: Wavelet decomposition of the physiological functional variables

Step2: Physiological Functional variable elimination: Repeat 10 times:

1. RF-RFE $(G(1),...,G(p))$

2. Compute a selection score for each group $G(j)$ as following:

$$score(G(j)) = \sum_{m=1}^{M} occ(G(j), Q_m) \times [(p-m)+1] \tag{2}$$

where Q_m is the list of the m-th selected variables, $m = 1..M$ and $occ(G(j), Q_m)$ corresponds to the occurrence of the variable j in the list m.

3. Eliminate the less relevant functional variables (those of a selection score below a threshold δ)

Step3: Wavelet Levels Selection of the k_R selected functional variables: Repeat 10 times:

1. RF-RFE $(\{G(1,k_1),...,G(J,k_1),...,G(1,k_R),...,G(J,k_R)\})$

2. Compute a selection score for each group $G(w,k_R)$ as following:

$$score(G(w,k_r)) = \sum_{m=1}^{M\prime} occ(G(w,k_r), Q_m) \times [((J \times R)-m)+1] \tag{3}$$

where Q_m is the list of the m-th selected variables wavelet levels, $m = 1..M\prime$ and $occ(G(w,k_r), Q_m)$ corresponds to the occurrence of the wavelet level w of the kept functional variable k_r in the list m.

3. Eliminate the less relevant variables (those of a selection score below a threshold $\delta\prime$)

4 Variable selection results

In this section, results from the application of the different variable selection will be detailed. The objective of variable selection is first to eliminate physiological signals that do not contribute significantly in the stress level classification, then for the retained physiological variables, the most relevant wavelet levels will be selected.
When applying our procedure to the drivedb database, we perform at a first stage functional variables decomposition using the Haar wavelet which is considered as the simplest one. We recall thate the Haar wavelet's mother wavelet function $\psi(t)$ is given by:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the corresponding scaling function $\phi(t)$ is given by:

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

We pick 12 as the decomposition level which corresponds to the maximum level compatible with the $4096 = 2^{12}$ samples.
To achieve this work, we use the `R` software (Team, R Core (2016)), with the `randomForest` package proposed by Liaw and Wiener (2002) and `RFgroove` packages developed by Gregorutti, B. (2016).

4.1 Physiological functional variable elimination

Based on the *drivedb* database, Healey and Picard (2005) showed that the EDA first and the HR are most closely correlated with driver stress level.

In this subsection, we seek to eliminate the less important physiological signals contributing to the stress level recognition and compare the obtained results with those found in other studies.

4.1.1 Physiological variables importance

Grouped VI allows to characterize the importance of a functional variable in the model building. It offers the opportunity to consider only five informative VI instead of 20480 VI of scalar variables. Fig. 3 represents the boxplots of the grouped VI values computed for 100 runs when all the variables are included in the model. *Foot EDA* is the most important physiological variable with VI distribution around 4%. *RESP* is the second most important variable. The median value of the distribution of 100 VI is about 3%. The *Hand EDA* comes out important when compared to the *EMG* and *HR*. The distribution of its 100 VI is around 2%. The *Foot EDA* and the *Hand EDA* are among the most important variables which confirms Healey and Picard (2005) findings. The variables *EMG* and *HR* are found to have a distribution of 100 VI around 0.

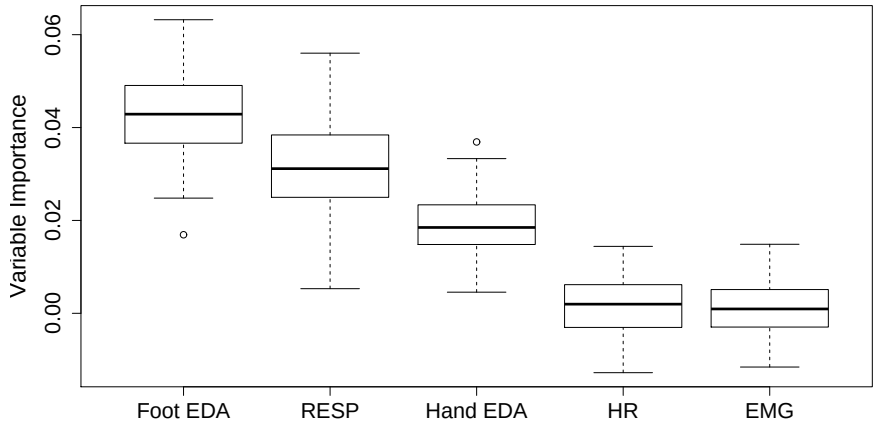


Fig. 3 Boxplots of grouped VI by physiological signals for 100 runs

4.1.2 A simple iteration of RF-RFE on the physiological functional variables

Let V1,...,V5 denote respectively the *Hand EDA*, *Foot EDA*, *HR*, *EMG* and *RESP*. Fig. 4 shows the misclassification rate computed on the validation set along the sequence of nested models of a simple iteration of the algorithm listed in the Table 3 . The selected model, based on the error minimization, is composed by two variables: *Foot EDA* and *RESP*. But it should be noted that when repeating the procedure with different learning and validation samples, the results of the selection change in terms of the number of variables making up the best model, the variables involved in the best model and even in the misclassification rate produced by the selection procedure.

Thus, an iterative procedure is needed in order to aggregate the information and provide the best model.

4.1.3 Iterative RF-RFE on the physiological functional variables

We apply the RF-RFE ten times to the wavelet coefficients grouped by functional variables. Each iteration offers a ranking of the 5 variables and the variables included in the selected model which offers the lowest misclassification rate. Table 5 shows the result of the 10 iterations of the RF-RFE algorithm on the five physiological variables. The shaded cells corresponds to the kept variables. For example, *Foot EDA*, *RESP* and *Hand EDA* are selected in the first iteration.

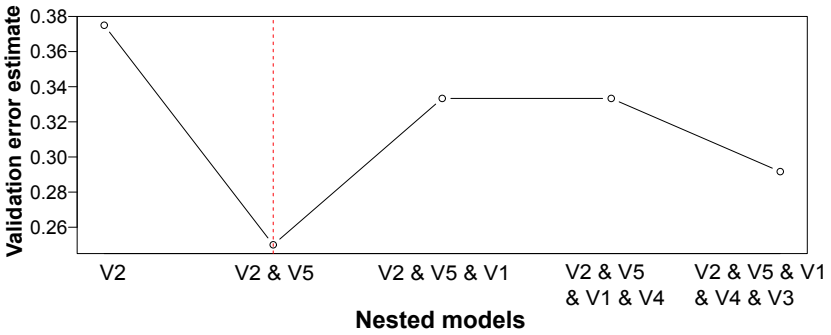


Fig. 4 Validation error for the nested models.
V1=*Hand EDA*, V2=*Foot EDA*, V3=*HR*, V4=*EMG* and V5=*RESP*

Based on these 10 iterations, we can conclude that *Foot EDA* is always included on the selected model. In addition, the *EMG* and the *HR* (except one time) are excluded from the list of the variables selected in the model with the minimum validation error. A variation of the number of selected variables and even in the order of the variables is notable. Thus we propose to use our approach in order to aggregate the information contained in these 10 iterations.

1	Foot EDA	RESP	Hand EDA	HR	EMG
2	HR	RESP	Hand EDA	Foot EDA	EMG
3	Foot EDA	Hand EDA	HR	EMG	RESP
4	Foot EDA	RESP	Hand EDA	EMG	HR
5	RESP	Foot EDA	Hand EDA	HR	EMG
6	Foot EDA	RESP	Hand EDA	EMG	HR
7	Foot EDA	Hand EDA	RESP	HR	EMG
8	Foot EDA	RESP	Hand EDA	HR	EMG
9	Foot EDA	Hand EDA	RESP	EMG	HR
10	Foot EDA	RESP	Hand EDA	HR	EMG

Table 5 Selected model for 10 runs of the RF-RFE algorithm. The shaded cells corresponds to the kept variables

Fig. 5 displays the selection score of the five physiological variables computed using equation (2). Firstly, the results confirm the order proposed by the VI measure and that both *EMG* and *HR* have low selection score. Secondly, it shows that *Foot EDA* is the most important variable in driver's stress level classification.
We note that the *Foot EDA* wasn't used in the study of Healey and Picard (2005). But the fact that both *Foot EDA* and *Hand EDA* appear in the best model confirms that the EDA is most closely correlated with driver stress level shown by Healey and Picard (2005) based on the *drivedb* database.
Classically, the EDA is measured on fingers or on the palms. But these two placements are not preferred in the real-world task performance. In the study achieved by Van Dooren et al. (2012) on the best placements of the EDA measure in terms of responsiveness and their similarity with the finger measures, foot and shoulders were shown being the most responsive and the measurements at the foot were the most similar with those of the finger .

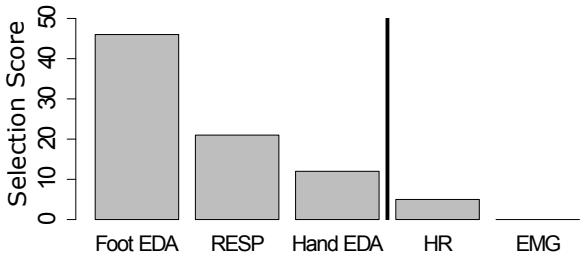


Fig. 5 Selection score of the five physiological variables for 100 runs

It should be noted that the ranking due to the selection score is close to the ranking given by the importance measures when all variables are taken in the model. The best model according to the proposed approach proposes to consider the three variables *Foot EDA*, *RESP* and *Hand EDA*. In the subsection 4.2, the wavelet levels of these selected three variables will be considered and the best model composed by the wavelet levels will be proposed.

4.2 Iterative RF-RFE on Wavelet levels

The grouped VI may increase with the number of variables in the group Gregorutti et al. (2015). Due to the fact that the number of variables in each group is different, we use the normalized version of the VI suggested by Gregorutti et al. (2015):

$$I_{nor}(\mathbf{X^J}) = \frac{1}{|J|} I(\mathbf{X^J})$$

We seek to determine, for the selected physiological functional data, the corresponding wavelet levels most able to predict the driver's stress level class. Thus, the wavelet coefficients are grouped by levels.

4.2.1 The global procedure

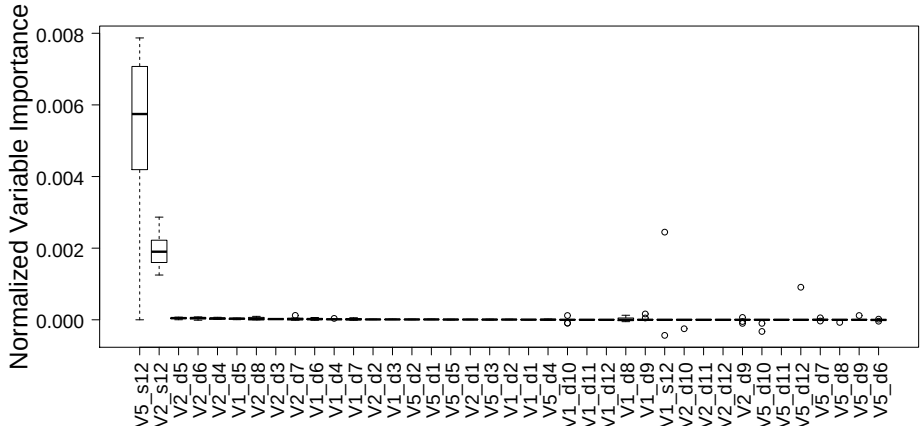
Wavelet levels importance

Let us recall that V1 stands for the *Hand EDA*, V2 the *Foot EDA* and V5 the *RESP*. Let us denote the approximation of the signal as $s12$ and the level i of the wavelet detail by di.

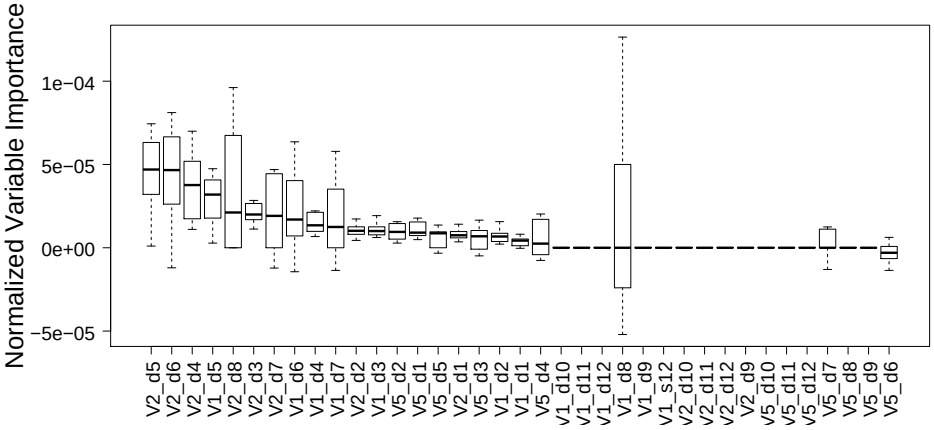
Fig. 6a represents the VI of the different levels of the three physiological functional variables. It shows that the approximation level ($s12$) of both *Foot EDA* and *RESP* are the most important levels in the stress level classification. The approximation level corresponds to $N/2^J$ coefficient thus one value. The variation of the distribution of the other wavelet levels is hidden, thus we consider a similar graph in the Fig. 6b, obtained by removing $V2\_s12$ and $V5\_s12$ and the outliers to better inspect the relative variation of small VI.
It is observed for the *Hand EDA* (V1), levels from 1 to 8 are the most important. Wavelet levels from 3 to 8 are important for the *Foot EDA* (V2). For the *RESP* (V5), levels 1 and 2 are the most important. In the next section, the selection score will be computed of all these wavelet levels in order to offer the possibility to select the best model.

Wavelet levels selection

When applying the RF-RFE algorithm on the groups composed by the wavelet levels of the three selected variables several times, the best model varies in terms of the number of the selected levels, the validation error and even the ranking of the different levels.
In order to reduce this variability, a selection score given by equation (3) allows to rank the different levels. The plot of the ranked variables by selection score is displayed on the Fig. 7. There are two ways to select the variables of the final model. If we look at the consecutive differences between the selection scores from the highest



(a) Grouped VI of the V1, V2 and V5 wavelet levels



(b) Grouped VI of the V1, V2 and V5 wavelet levels without V2_s12 and V5_s12

Fig. 6 Grouped VI of the Wavelet levels for 10 iterations

score and we cut on the biggest gap of this difference, we select $V2\_d4$ and $V5\_s12$. If we investigate the consecutive differences between the selection scores starting from the lowest score and we cut on the biggest gap of this difference, we select $V2\_d4, V5\_s12$ and $V2\_d3$. In this study we choose to keep three wavelet levels, especially since the third level is $V2\_d3$ which is almost the same as $V2\_d4$.

The wavelet level j corresponds to $\frac{N}{2^{j-1}}$ coefficients. For a wavelet level j, it consists of dividing the signals into $2^{(J-j)+1}$ segments and taking the center of each segment. For example, when the $s12$ is selected, this means that the approximation level corresponding to the coefficient located at the $t = \frac{2048}{F_s} = 132.13s$ is selected.
When comparing the selected levels and the VI of the different wavelet levels, we observe that the results are similar.

4.2.2 An additional individual information

In this subsection, we propose to study for a given physiological functional variable which wavelet levels are the most able to classify the driver's stress level. Thus, the selection of the wavelet levels is achieved independently for the *Hand EDA*, *Foot EDA* and the *RESP* respectively.
For that equation (3) is used to compute the selection score of the different wavelet levels of the *Hand EDA*. Fig.

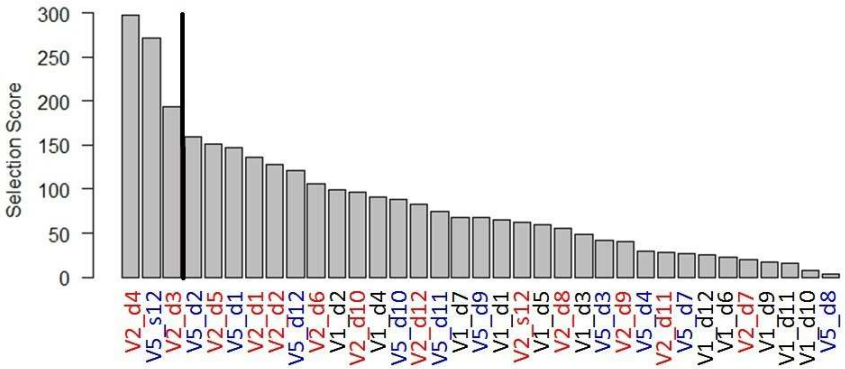


Fig. 7 Wavelet levels selection for 10 iterations

8 represents the *Hand EDA* wavelet levels ordered by the selection levels. When starting from the smallest value of the score and looking at the consecutive difference, levels 1 to 8 (except 7) are selected for the *Hand EDA*. This result was found when analyzing the VI of the *Hand EDA* levels.

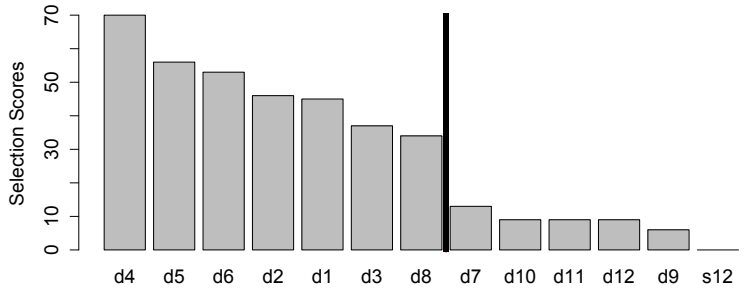


Fig. 8 Wavelet levels selection score of the Hand EDA for 10 iterations

We note that the ordered selection score of the *Foot EDA* wavelets is shown in Fig 9. It shows that the ranking of the wavelet levels of the *Foot EDA* $d4$, $d5$ then $d6$ is the same for the *Hand EDA* (see Fig. 8 for comparison). For this physiological predictor, two details are selected which are mainly the levels 4 and 5. The level 4 was selected in the best model that combine the different wavelet levels of the three physiological predictors.

Looking at the Fig. 10, we can observe that the first and the second details of the wavelet levels and the approximation are selected for the respiration (RESP) predictor. These findings confirms the result of the computed VI when considering the model composed by the wavelet levels of the three selected physiological variables.

4.3 Assessing the final selection

As detailed in Subsection 2.2, only four participants performed the real-world driving experience. This special form of the experiments (experiment repeated by one driver) is not taken into account in the model. Moreover, the general selection strategy applied was preliminary used to ensure the general applicability of the whole proposed procedure. As we have a prior knowledge of the cohort, we propose to use a kind of cross-validation reflecting this information, in order to assess the final model. Fig. 11 shows the different configurations considered.

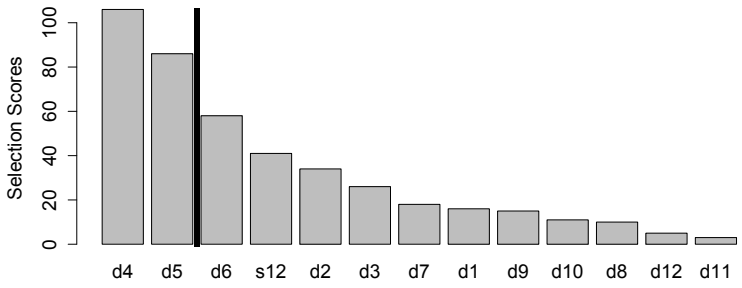


Fig. 9 Wavelet levels selection for the Foot EDA for 10 iterations

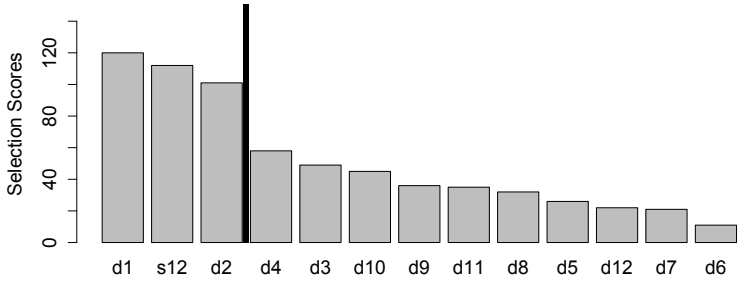


Fig. 10 Wavelet levels selection for the RESP for 10 iterations

As the fourth driver performs only one drive, we propose to keep always his signals in the learning set. The misclassification rate is computed on the three validation sets corresponding to the consideration of the driving experiences achieved by one driver.
When data related to the driver k, $k = 1$ to 3 are used in the validation sample, data of other drivers are used in the learning set.

Fig. 11 Learning and validation sets choice in the cross validation procedure

Each configuration offers a misclassification rate, and we take the mean of these rates as the procedure error leading to 0.17. This error is now to be compared with the final error given by the Expert-Based method detailed in the next subsection.

4.4 Comparative study of the method performance

In this subsection, the approach proposed by Healey and Picard (2005) based on physiological features extraction and stress level classification using Linear discriminant function will be performed on the database composed of the segments of *Hand EDA*, *Foot EDA* and *RESP*. The result of such analysis will then be compared to the proposed "blind" approach based on the Wavelet decomposition and Variable selection using RF-RFE, in terms of misclassification error.

Expert-Based features are extracted only from the three selected physiological signals: *Hand EDA*, *Foot EDA* and *RESP*.

- From the standardized *Hand EDA* and *Foot EDA*, the mean, the standard deviation and four features describing startles are extracted.
- The mean, the standard deviation and four spectral features were derived from the preprocessed Respiration signal.

The model is thus: $Y = f(S_1(t), ..., S_p(t))$, where $S_i(t)$ is summarized by its features. Then,

$$Y = \widetilde{f}(v_1^1, ..., v_{n_1}^1, ..., v_1^p, ..., v_{n_p}^p) + \varepsilon$$

After extracting the features, a linear discriminant analysis was performed and a leave-one-out cross validation was used. The misclassification rate is 0.22. So the Expert-Based method does not perform better than our method which does not incorporate any prior information.

5 Discussion

The proposed "blind" approach performs as the expert-based approach in terms of misclassification rate. This procedure offers moreover, additional information such as the physiological variables ranking according to their importance and the list of the relevant variables in stress level classification. The obtained results suggest that *EMG* and the *HR* are not very relevant when compared to the EDA and the respiration signals. This may help to investigate the list of physiological sensors that can be proposed to the smart vehicles designers, in order to determine the stress level.

In this work, the proposed selection procedure allows to first eliminate the irrelevant functional variables then to select, from the remaining functional variables, the important wavelet levels. This was achieved using the grouped variable importance measure and a wavelets projection-based approach. Several other groupings can be proposed for coefficients resulting from wavelet decompositions, especially when considering the physical variables characterizing the environment and the vehicle.

We used a procedure based on the RF-RFE, that consists to consider at each step only models composed by $p, p-1, ..., 1$ variables. Thus, not all variables combination are considered, which can perform model with lower error rate. Considering the five physiological variables combinations in our case remains feasible. But, in the case of more important number of variables this will be difficult.

In this study, the variable of interest is built according to the type of route crossed by the driver, thus it depends mainly on the hypothesis assuming that the stress level increases when driving in the city and decreases when the participant is at the rest. This hypothesis is sometimes not verified because we neglect the driver's current cognitive state, the effect of the state induced when he anticipates the situation or even he thinks about the past driving experience, called the memory effect. We need measurements characterizing the environment and the vehicle in order to include information about the environment complexity that can affect directly the stress level of the driver. Such measurements can contribute in the stress level construction, and bring objectivity to our variable of interest.

The fact of recognizing the low stress level as high one is more serious than predicting the high level as medium one, or the medium as high stress level. For a methodological perspective, the consideration of costs that penalize the high versus low confusion will be achieved.

REFERENCES

Akbas, A. (2011). Evaluation of the physiological data indicating the dynamic stress level of drivers. *Scientific research and essays*, 6(2):430–439.

Auret, L. and Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems*, 105(2):157–170.

Bostrom, J. (2005). Emotion-sensing pcs could feel your stress. *PC World*.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C.-J., and Olshen, R.-A. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

Dawson, M.-E., Schell, A.-S., Filion, D.-L., and Berntson, G.-G. (2007). The electrodermal system. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G., editors, *Handbook of Psychophysiology*, pages 157–181. Cambridge University Press, third edition. Cambridge Books Online.

Deng, Y., Wu, Z., Chu, C., and Yang, T. (2012). Evaluating feature selection for stress identification. In *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, pages 584–591.

Díaz-Uriarte, R. and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1):1–13.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Genuer, R., Poggi, J.-M., and Tuleau-Malot (2015). Vsurf: An r package for variable selection using random forests. *The R Journal*, 7(2):19–33.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225 – 2236.

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C.-K., and Stanley, H. (2000). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13)*.

Gregorutti, B., Michel, B., and Saint-Pierre, P. (2015). Grouped variable importance with random forests and application to multiple functional data analysis. *Computational Statistics and Data Analysis*, 90:15–35.

Gregorutti, B., Michel, B., and Saint-Pierre, P. (2016). Correlation and variable importance in random forests. *Statistics and Computing*, pages 1–20.

Gregorutti, B. (2016). Rfgroove: Importance measure and selection for groups of variables with random forests. *https://CRAN.R-project.org/package=RFgroove, R package version 1.1, (2016)*.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Healey, J.-A. (2000). *Wearable and automotive systems for affect recognition from physiology*. PhD thesis, MIT Dept. of Electrical Engineering and Computer Science.

Healey, J.-A. and Picard, R.-W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.

Kleine-Cosack, C. (2006). Recognition and simulation of emotions. *Seminar: Human-Robot Interaction, SoSe 2006, Fachbereich Informatik Universitat Dortmund*.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 26*, pages 431–439.

Lykken, D., Rose, R., Luther, B., and Maley, M. (1966). Correcting psychophysiological measures for individual differences in range. *Psychological Bulletin*, 66(6):481–484.

Lykken, D. T. (1972). Range correction applied to heart rate and to gsr data. *Psychophysiology*, 9(3):373–379.

Mallat, S.-G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Nicodemus, K.-K., Malley, J.-D., Strobl, C., and Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics*, 11(1):1–13.

Poggi, J.-M. and Tuleau, C. (2007). Classification of objectivization data using cart and wavelets. *Proceedings of the IASC 07, Aveiro, Portugal*, pages 1–8.

Ramsay, J.-O. and Silverman, B.-W. (2005). *Functional Data Analysis*. Springer-Verlag New York.

Rigas, G., Katsis, C., Bougia, P., and Fotiadis, D. (2008). A reasoning-based framework for car drivers stress prediction. In *Control and Automation, 2008 16th Mediterranean Conference on*, pages 627–632.

Sauvé, M. and Tuleau-Malot, C. (2014). Variable selection through CART. *ESAIM: Probability and Statistics*, 18:770–798.

Singh, M. and Queyam, A.-B. (2013). Stress detection in automobile drivers using physiological parameters: A review. *International Journal of Electronics Engineering*, 5(2):1–5.

Smart, R.-G., Cannon, E., Howard, A., Frise, P., and Mann, R.-E. (2005). Can we design cars to prevent road rage? *International Journal of Vehicle Information and Communication Systems*, 1(1-2):44–55.

Strobl, C. and Zeileis, A. (2008). Danger: High power! ? exploring the statistical properties of a test for random forest variable importance. *Proceedings of 18th International Conference on Computational Statistics*.

Tao, J. and Tan, T. (2005). *Affective Computing: A Review*, pages 981–995. Springer Berlin Heidelberg, Berlin, Heidelberg.

Team, R Core (2016). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. url:. *www. R-project. org*.

Van Dooren, M., De Vries, J.-J., and Janssen, J.-H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2):298 – 304.

Verikas, A., Gelzinis, A., and Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44(2):330–349.

Yang, K., Yoon, H., and Shahabi, C. (2005). A supervised feature subset selection technique for multivariate time series. *Proceedings of the Workshop on Feature Selection for Data Mining: Interfacing Machine Learning with Statistics*, pages 92–101.

Zhu, R., Zeng, D., and Kosorok, M.-R. (2012). Reinforcement learning trees. Technical report, University of North Carolina.