



# Sequential nonlinear encoding: A low dimensional regression algorithm with application to EEG-based driving fatigue detection

M. Tabejamaat and H. Mohammadzade\*

*Department of Electrical Engineering, Sharif University of Technology, Tehran 11155-8639, Iran.*

Received 30 June 2019; received in revised form 26 November 2019; accepted 15 June 2020

## KEYWORDS

Brain computer interface;  
 Electroencephalography;  
 Fatigue detection;  
 Projection pursuit regression;  
 Information.

**Abstract.** Regression analysis of real-world data has not always been an easy task, especially when input vectors are presented in a very low dimensional space. EEG-based fatigue detection deals with low dimensional problems and plays a major role in reducing the risk of fatal accidents. We propose a kernel projection pursuit regression algorithm which is a two-step nonlinearity encoding algorithm tailored for such low dimensional problems such as fatigue detection. In this way, data nonlinearity can be investigated from two different perspectives: by first transforming the data into a high dimensional intermediate space and then, applying their spline estimations to the output variables allowing for hierarchical unfolding of data. Experimental results of the SEED VIS database illustrate the average RMSE values of 0.1080% and 0.1054%, respectively, for the temporal and posterior areas of the brain. Our method is validated by conducting some experiments on Parkinson's disease prediction, which further demonstrates the efficiency of our method. This paper proposes a novel regression algorithm to address the encoding problem of highly complex low dimensional data, which is usually encountered in bio-neurological prediction tasks like EEG-based driving fatigue detection.

© 2022 Sharif University of Technology. All rights reserved.

## 1. Introduction

Mental fatigue is a brain psycho-physiological state that is clinically characterized by increased production of cytokine or a decrease in the cortisol level. In addition, it is manifested as some forms of diminished mental alertness, listlessness, and even traces of languidness. Previous studies have shown that fatigue may generally occur in two different ways: active and

passive. Active fatigue refers to a situation where the brain is exposed to exhausting processing tasks (such as driving in a messy-busy road). In this way, it is triggered to release much dopamine in order to respond to the need for making decisions faster. However, the greater the dopamine secretion, the more metabolic wastes accumulated in the brain region. This eventually causes the brain to ask the frontal cortex for a temporarily shutdown. On the contrary, passive fatigue is a feeling of under stimulation that arises from conducting a monotonous task for a long time (such as driving in a straight and monotonous road), which is associated with the increased Cortisol level in the brain, leading to some forms of slow nerve conduction and reduced decision-making performance.

\*. *Corresponding author.*  
*E-mail addresses: tabejamaat4066@gmail.com (M. Tabejamaat); hoda@sharif.edu (H. Mohammadzade)*

Automatic detection of fatigue (vigilance) level has recently turned into a hot research topic with a wide variety of applications in the areas like medicine [1–4], transportation [5–8], and education [9,10]. For example, it can be used in monitoring brain functions during an awake surgery, evaluating anesthetic effects on the brain, detecting fatigue driving to establish an early warning or providing the information on shipping or insurance companies, or setting the ideal class length for students with different mental and physical capabilities. Among the applications, driving has received much more attention because it directly influences the lives of more people like drivers, passengers, pedestrians, and motor/pedal cyclists. In fact, fatigue is the starting point of a catastrophic chain that may ultimately lead to a fatal crash. First, it heavily lowers the driver's concentration, selective attention, and decision speed and next, it disrupts his eye-hand coordination and consequently, the control on steering wheel and pedals, resulting in eventual crash. Recent research shows that fatigue is responsible for 10% of road accidents all over the world [11]. Therefore, there is an urgent need for developing such a countermeasure as early warning fatigue driving systems. According to the types of symptoms, existing early warning Automatic Fatigue Detection (AFD) methods can be broadly categorized into five classes:

1. Questionnaire-based techniques that measure self-reported psychological qualities from the responses given to some situational questions [12–16]. The questionnaires are usually filled before and after a tedious task where the performance differences can reveal the extent of the fatigue state;
2. Video-based techniques that utilize such physical symptoms as yawning, pattern motion of eyelid, eye, and head as well as facial and eye expression [17–20];
3. Cognitive tasks that use the reaction time or the error rate of the responses to a set of visual stimuli [21–23];
4. Neurophysiological techniques such as electroencephalography (EEG) and electrooculography (EOG) [24,25];
5. Physiological techniques such as electrocardiography (ECG) [26–30] or Grip Force (GF) [31].

In comparison, EEG enjoys such advantages as noninvasive nature, low price, robustness to biased responses (unlike psychometric and cognitive methods), and simplicity, making it much more convenient to use than other rivals.

Most of the existing fatigue detection algorithms are generally established upon a classification strategy where each class is a qualitative measure of fatigue

state (e.g., sleep, drowsy, alert, near alert, etc.) [18]. More recently, few studies have addressed the problem of continuous fatigue detection (in the framework of a regression problem). However, there are two key issues with these methods including how to continuously annotate the EEG data and how to optimally project the samples into the continuous space of fatigue levels. For years, some traditional strategies including (1) manual annotation, (2) use of face expression, and (3) simultaneous cognitive tasks were available for continuous annotation of EEG data. However, these methods are either time-consuming and imprecise (like manual annotation and use of face expression) or require additional interactions of drivers that may distract their attention and cause safety issues (like cognitive tasks). It has recently been shown that eye tracking can be efficiently utilized for a fast continuous annotation of EEG data while it does not require any active cooperation and is quite synchronized with the variations of EEG signals. Most of the existing fatigue detection methods use Support Vector Regression (SVR) algorithm and report some very promising results.

Despite the advances, EEG-based fatigue detection methods still are subject to some major drawbacks including: (1) high computational cost due to the need for transforming the data into the frequency domain, (2) lack of any provision for dealing with the possible outliers of EEG data, and (3) being restricted in handling a large amount of data nonlinearity. In fact, these methods require their link (smooth) functions to be previously selected, restricting their ability to handle highly complex data.

This study proposes a novel double non-linearized pursuit regression algorithm that offers the ability to encode the nonlinearity of data in a sequential manner and is, therefore, appropriate for modeling such low dimensional problems as continuous fatigue detection. To this end, the data representing a linear combination of its elements in the high dimensional Hilbert space (termed feature map) are considered and then, are transformed into a continuous space of its output indices. In this way, nonlinearity can be encoded in two different manners: (a) mapping of the data into the Hilbert space and (b) nonlinear transformation of the feature maps into the output labels. In addition, a constraint is introduced on the projection weights of the pursuit regression algorithm so that it will be a convex representation of the Hilbert space variables. As feature extraction, we employ the Logarithmic Energy (LE) of several wavelet-based band limited epochs, which is performed in the time domain rather than using the conventional time-consuming spectral analysis of data.

The rest of this study is organized in the following. Section 2 provides a brief review of the related works. Section 3 presents the main idea of the Projection

Pursuit Regression (PPR). Section 4 introduces PPR for detecting fatigue states in the high-dimensional Hilbert space. Section 5 describes the evaluation of our method. Finally, Section 6 draws the conclusion.

## 2. Related works

This section provides a review of AFD algorithms using EEG signal, which is presented from the viewpoint of machine learning. However, there are also a variety of papers that have addressed the issue in psychological or medical terms which are beyond the scope of this paper.

### 2.1. Spectral analysis

The last 10 years have seen an explosion of scientific works for recognizing fatigue states from EEG data. Most of these works are based on the spectral analysis of band limited signals. For example, Shi et al. [32] utilized Differential Entropy (DE) to characterize 9 EEG signals from the brain occipital lobe. The vectors were then smoothed and represented by the Principal Component Analysis (PCA) [33]. Finally, the projection model was constructed by using the SVR algorithm. The researchers in [34] utilized the same feature extraction strategy as the one used in [32], but for 6 EEG signals from the occipital region. The Linear Coefficient Correlation (LCC) algorithm was then applied as a feature selection step before constructing the SVR model. In [35], Kirk et al. used a combination of PCA and Bicoherence Spectral Estimator (BSE) for extracting the features of  $O_1$ - $O_2$  EEG signals. In [36], weighted average of differential power spectra was calculated as the feature vector of each EEG epoch. Then, SVR was applied to map the vectors into the space of vigilance indices. Armanfard et al. [37] represented the input signals using the power and magnitude of its spectral patterns and employed the Minimum Redundancy-Maximum Relevance (MRMR) [38] strategy for selecting the most discriminative features. Finally, SVR was applied to map the data into the space of fatigue indices. The study in [39] utilized the spectral power and power ratios of band limited signals as the features of EEG data. Finally, Fisher scoring was used for predicting the vigilance level. In [40], Guo et al. represented the features using the weighted average of Power Spectral Density (PSD). Then, a modified Support Vector Machine (SVM) with the Particle Swarm Optimization (PSO) algorithm was used as the prediction model. Chen et al. [41] used the normalized amplitude of the spectral patterns as the features of EEG data. Then, the ratio of the normalized amplitude in the alpha band over the sum of the corresponding normalized amplitudes in the delta and theta bands was calculated as a description of fatigue level.

### 2.2. Connectivity analysis

Connections amongst areas of the nervous system are of interest to researchers applying behavioral neuroscience to problems like fatigue detection. For example, Cynthia et al. [42] utilized the Phase Locking Value (PLV) for studying the pattern of dependencies between multiple fatigue-induced EEG signals. Then, the strength, local efficiency, global efficiency, and clustering coefficient were extracted as the statistical features of the dependency pattern. Finally, K-nearest neighbor, SVM, and Multi-Layer Perception (MLP) were used for classifying the feature vectors into a set of predefined discrete states of vigilance. Dimitrakopoulos et al. [23] used the Generalized Partial Directed Coherence (GPDC) algorithm [43] to estimate the functional connectivity between the channels of EEG data. The connectivity values were directly used as features of each epoch. Then, Sequential Forward Floating Selection (SFFS) method [44] and SVM were applied as the feature extraction and classification algorithms, respectively.

### 2.3. Time-domain techniques

In these methods, fatigue indices are directly estimated based on the time domain information without the need for any transformation into the frequency domain. For example, Li et al. [45] utilized the Probabilistic Principal Component Analysis (PPCA) [46] for describing the band limited representation of the signals. In [47], mean, standard deviation, and power of EEG signals were introduced as the fatigue related attributes. Ning and Bronzino [48] investigated the Gaussianity of EEG signal distribution as a measure of fatigue states. Ouyang and Lu [49] extracted the mean and standard deviation of the Continuous Wavelet Transform (CWT) coefficients as the fatigue related features of EEG data. Then, the dimensionality of features was reduced by Random Forest (RF) algorithm. Finally, SVM was used for classification. Authors in [50] characterized the EEG data using the wavelet coefficients and then, used the Sparse Representation-based Classification (SRC) algorithm [51] for predicting the discrete states of vigilance.

However, these methods all suffer from the main disadvantage because they utilize a set of fixed link functions for encoding nonlinearity which restricts their ability to handle highly complex data. Moreover, their encoding procedure is restricted by their one-shot mapping strategy. That is, only one non-linear transformation is applied to each point of data. Unlike these methods, this paper aims to propose a sequential nonlinearity encoding technique that provides two major advantages: (1) investigating the non-linearity of data from two different perspectives (kernel mapping and spline projection) and (2) calculating the link functions rather than predefining them.

### 3. Projection Pursuit Regression (PPR)

PPR [52] is a nonparametric regression algorithm that utilizes one-dimensional smoothers (Subsection 3.1) to fit a regression model over a set of low dimensional feature spaces rather than high dimensional input variables. Let  $X \in R^{n \times m}$  denote a dictionary of input vectors  $x_i \in R^m$  and  $y$  be the corresponding dependent variables. The objective function of the model is defined as follows:

$$J = \min_{g_q, w_q} \sum_{i=1}^N \left\{ y_i - \sum_{q=1}^Q g_q(w_q^T x_i) \right\}^2, \quad (1)$$

where  $w_q$ ,  $q = 1, \dots, Q$ , is a coefficient vector (directional vector) that projects the input variables into the  $q$ th intermediate feature map  $f_q = Xw_q$ ,  $g_q$  is the corresponding smoothing function, and  $Q$  denotes the number of intermediate projections. Given that  $g_q$  is a twice-differentiable function, the model is able to characterize the nonlinearity of the intermediate variables. The smoothness of  $g$  determines the fidelity of the model to the training data. The smoother the function  $g$  is, the less accurate the model will be fitted, which may lead to some forms of an undesired bias. Conversely, with a too wiggly function, the model will be overfitted and it tends to become a linear estimate of the intermediate variables. This calls for a trade-off between the smoothness and fidelity of the solution which can be achieved using the Akaike information criterion, the cross-validation score, or the generalized cross-validation score. Note that, PPR does not imply any orthogonality restrictions between the directional vectors, which in turn allows for an interaction between the intermediate projections that may occur when the superposition principle does not hold for the responses of the regression model. For a comprehensive optimal solution, both the directional vectors (Eq. (2)) and smoothing functions (Eq. (4)) are estimated in an iterative manner. For this purpose, the data is first randomly projected into a set of one-dimensional representations. The functions  $g_q$ s are then calculated through the spline estimation of these intermediate projections to the output responses. Next, assuming a fixed smoothing function, the Gauss-Newton algorithm is utilized for estimating its corresponding directional vector [52]:

$$\begin{aligned} g_q(w_q^T x_i) &\approx g_q(w_{q,old}^T x_i) + g'_q(w_{q,old}^T x_i)(w_q - w_{q,old})^T x_i \\ &\Rightarrow \min_{w_q} \sum_{i=1}^N g'_q(w_{q,old}^T x_i)^2 \left[ \left( w_{q,old}^T x_i - \frac{y_i - \sum_{q=1}^Q g(w_{q,old}^T x_i)}{g'(w_{q,old}^T x_i)} \right) - w_q^T x_i \right]^2. \end{aligned} \quad (2)$$

PPR is closely related to different statistical and machine learning algorithms. With a single term, it resembles a Generalized Linear Model (GLM) with an inverse link function. Yet, with more terms at hand, it would be more similar to Single-Hidden-Layer (SHL) neural networks such as SHL Multi-Layer Perceptron (MLP) or SHL Extreme Learning Machine (ELM). Note that the use of multiple terms allows for combining more intermediate representations which further provides us with a more intricate model. However, despite the similarities, there are also some systematic differences in the optimization technique and free parameters of these networks. While MLP is usually trained using the Gradient Descent (GD) strategy, PPR has the advantage to use the Least Squares (LS) method that does not depend on any prerequisites like determining a learning rate. Unlike ELM that utilizes a non-tuned hidden layer, PPR uses the LS algorithm to adjust the transformation weights of the input variables to the intermediate representations. Moreover, unlike neural networks, to its advantage, PPR estimates its smoothing functions along with the projection weights and this is very relevant for dealing with a nonlinear data.

Notwithstanding, PPR is characterized by its own drawbacks, as well. For example, regarding a very complex dataset, a model can be fitted well only if the dimensionality of data is large enough to characterize the complexity of data distribution. On the other hand, the higher the dimensionality of data, PPR would require more intermediate terms to properly handle the resulting intricate patterns which, in turn, causes a serious computational problem.

#### 3.1. Smoothing spline

Data smoothing is a way to capture the main structure of data when the fine-scaled structure is considered to be rapid variation attributable to noise. The method used, namely the smoothing spline method, balances approximation errors with a derivative-based measure of overall smoothness for a dataset.

Let  $g$  be a smooth function defined at  $[a, b]$  so that its second derivative is available and the following minimization problem holds:

$$J_{sf} = \sum_{i=1}^N [y_i - g(\mathbf{o}_i)]^2 + \lambda \int_a^b [g''(\mathbf{o})]^2 d\mathbf{o}, \quad (3)$$

where the first term measures the fidelity of the fit to the data, while the second term regularizes its smoothness and curvature using the fixed parameter  $\lambda$ . Assuming that  $g$  is a natural cubic spline, it can be written as follows:

$$g(\mathbf{o}) = \sum_{j=1}^N B_j(\mathbf{o})\theta_j, \quad (4)$$

where  $B_j(x)$  is the  $j$ th basis function of  $g$  with the knots defined at  $x_1, x_2, \dots, x_N$ . Substituting Eq. (4) into Eq. (3), we get:

$$J_{sf} = (y - B\theta)^T(y - B\theta) + \lambda\theta^T\Omega\theta, \tag{5}$$

where  $B_{i,j} = B_j(\mathbf{o}_i)$  and  $\Omega_{i,j} = \int (B_i(\mathbf{o}))''(B_j(\mathbf{o}))''d\mathbf{o}$ . This equation can be finally solved through the LS algorithm as follows:

$$\begin{aligned} \hat{\theta} &= (B^T B + \lambda\Omega)^{-1} B^T y, \\ \hat{y} &= B\hat{\theta} = B(B^T B + \lambda\Omega)^{-1} B^T y. \end{aligned} \tag{6}$$

In this way, each of the  $Q$  smoothing functions of Eq. (1) can be easily estimated.

#### 4. Methodology

Nonlinear mapping of data is an appropriate tool for handling the complexity of data distribution in a low dimensional space, which is subtly used in PPR for analyzing multiple regression models. However, sometimes, data points have a rather complex structure to be modeled by nonlinear mapping of their far low dimensional representations in the intermediate space. Here, one solution is to utilize more intermediate terms to extend the dimensionality so that the nonlinear model can be well fitted to such extended intermediate representations. However, increasing the number of intermediate terms would also dramatically increase the computational cost and call for a heavy backfitting process. Note that the response variable in Eq. (1) is a non-weighted sum of the intermediate terms. Therefore, the use of more than single terms may introduce some bias in the training process. Such issues hinder the practical application of PPR for modeling a highly complex data, especially with low dimensional variables. As an alternative, we can project the input variables into a higher dimensional space rather than increasing the number of intermediate terms. In doing so, the complexity of data is simplified before transforming into the intermediate representation which further simplifies the structure of the fitted splines at the next phase. However, such a scheme not only increases the computational cost, yet requires a complicated feature engineering for representing input variables in a high dimensional space. To address these issues, we propose projecting the model into the Reproducing Kernel Hilbert Space (RKHS) so that it can implicitly operate in a high dimensional space without the need for calculating the new coordinates of data points.

Moreover, we propose introducing a constraint on the directional vectors  $w_q|_{q=1}^Q$  that significantly reduces the possibility of an overfitting and helps mitigate the singularity problem of the kernel matrices.

Let  $\phi$  be a function that nonlinearly maps an  $m$ -dimensional input data into an infinite-dimensional RKHS  $\mathcal{F}$  so that  $\phi : x_i \in R^m \rightarrow \phi(x_i) \in \mathcal{F}$  and the inner product  $\langle . \rangle$  on  $\mathcal{F}$  is delineated by a kernel function  $k(a, b) = \langle \phi(a), \phi(b) \rangle$ . Applying this function to the training set, we obtain:

$$\begin{aligned} X &= [x_1, x_2, \dots, x_N] \rightarrow \Phi(X) \\ &= [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]. \end{aligned}$$

According to Eq. (2), a novel objective function is introduced for modeling the projection pursuit in the high dimensional Hilbert space with an additional constraint on the weight vectors  $w_q$ .

$$\begin{aligned} J &= \min_{g_q, w_q} \sum_{i=1}^N \left\{ y_i - \sum_{q=1}^Q g_q(w_q^T \phi(x_i)) \right\} \\ \text{s.t.: } & \min \|w_q\|_2. \end{aligned} \tag{7}$$

Upon considering only one term and using the Lagrange multiplier theorem, the problem can be written in an unconstrained form:

$$J = \min_{g, w} \sum_{i=1}^N \{y_i - g(w^T \phi(x_i))\} + \mu|w|_2, \tag{8}$$

where  $\mu$  is a small regularization coefficient. Given the smooth function  $g$ , similar to Eq. (2), the Gauss–Newton algorithm can be used to minimize the equation over  $w$ . Therefore, the function can be rewritten as follows:

$$\begin{aligned} J &= \sum_{i=1}^N g'(w_{old}^T \phi(x_i))^2 \left[ w_{old}^T \phi(x_i) \right. \\ &\quad \left. + \frac{y_i - g(w_{old}^T \phi(x_i))}{g'(w_{old}^T \phi(x_i))} - w^T \phi(x_i) \right]^2 + \mu|w|_2. \end{aligned} \tag{9}$$

Let  $t_i = w^T \phi(x_i) + \frac{y_i - g(w^T \phi(x_i))}{g'(w^T \phi(x_i))}$  and  $a_i = (g'(w^T \phi(x_i)))^2$ . Rewriting Eq. (9) in a matrix form, we get the following expression:

$$\begin{aligned} J &= T^T A T - T^T A \Phi(X) w - w^T \Phi(X)^T A T \\ &\quad + w^T \Phi(X)^T A \Phi(X) w + \mu w^T w, \end{aligned} \tag{10}$$

where:

$$A = \begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & a_N \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}.$$

Taking partial derivatives with respect to  $w$  and setting the result equal to zero, we obtain:

$$\frac{dJ}{dw} = -\Phi(X)^T AT + \Phi(X)^T A \Phi(X)w + \mu w = 0. \quad (11)$$

After some algebraic manipulation and substituting the extended representation of  $T$  in the matrix form, we get:

$$w_n = (\Phi(X)^T A \Phi(X) + \mu I)^{-1} \Phi(X)^T A \left( \Phi(X)w + \frac{y - g(\Phi(X)w)}{g'(\Phi(X)w)} \right), \quad (12)$$

where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Also,  $I$  is the identity matrix. As the weight vector  $w$  lies in the span of  $\phi(x_1), \phi(x_2), \dots, \phi(x_N)$ , there exists a coefficient vector in the Hilbert space that satisfies the following:

$$w = \sum_{i=1}^N \alpha_i \phi(x_i) = \Phi(X)\alpha. \quad (13)$$

Substituting Eq. (13) into Eq. (12) and simplifying the equation, we obtain:

$$w_n = (\Phi\Phi^T A \Phi\Phi^T + \mu\Phi\Phi^T)^{-1} \Phi\Phi^T A \left( \Phi\Phi^T \alpha + \frac{y - g(\Phi\Phi^T \alpha)}{g'(\Phi\Phi^T \alpha)} \right). \quad (14)$$

As  $\mathcal{K}(X, X) = \Phi\Phi^T$  is the kernel function, the equation can be reformulated as follows:

$$w_n = (\mathcal{K}A\mathcal{K}^T + \mu\mathcal{K})^{-1} \times \mathcal{K}A \left( \mathcal{K}\alpha + \frac{y - g(\mathcal{K}\alpha)}{g'(\mathcal{K}\alpha)} \right). \quad (15)$$

For the smoothing functions, the input in the Hilbert space will be  $w^T \phi(X_i)$ . Therefore, the update equation for  $g$  in the new space can be formulated as follows:

$$\begin{aligned} g(\Phi(X)w) &= N(\Phi(X)w)\theta \\ &= N(\Phi(X)w)(N(\Phi w)^T N(\Phi w) \\ &\quad + \Omega_{N(\Phi w)})^{-1} N(\Phi w)^T y. \end{aligned} \quad (16)$$

It was previously assumed that  $w = \Phi\alpha$ . Therefore, Eq. (16) can be reformulated as:

$$\begin{aligned} g &= B(\Phi\Phi^T \alpha)(B(\Phi\Phi^T \alpha)^T B(\Phi\Phi^T \alpha) \\ &\quad + \Omega_{B(\Phi\Phi^T \alpha)})^{-1} B(\Phi\Phi^T \alpha)^T y. \end{aligned} \quad (17)$$

Given that  $\mathcal{K}(X, X) = \Phi\Phi^T$ , this equation is finally represented as follows:

---

**Input:**  $X = \{x_1, x_2, \dots, x_n\}$ ,  $y = \{y_1, y_2, \dots, y_n\}$   
**for**  $p=1:n, q=1:n$  **do**  
 $\mathcal{K}(p, q) = \|x_p - x_q\|$   
**end for**  
**for**  $i=1:M$  **do**  
 $\alpha_i \leftarrow$  Initialize the weight vectors  
 $\alpha_i \leftarrow \frac{\alpha_i}{\|\alpha_i\|}$   
 $g_i \leftarrow \{\mathcal{K}, \alpha_i, y\}$  Using equation (19)  
**end for**  
**for**  $i=1:M$  **do**  
 $pr \leftarrow pr + g(\mathcal{K}\alpha_i)$  Calculate the prediction  
**end for**  
 $err = \|y - pr\|$  Calculate the prediction error  
**while**  $err < tr$ ,  $itr < tr_{itr}$  **do**  
**for**  $i=1:M$  **do**  
 $\alpha^o \leftarrow$  remove  $\alpha_i$   
 $g^o \leftarrow$  remove  $g_i$   
**for**  $j=1:M-1$  **do**  
 $pr \leftarrow pr + g_j^o(\mathcal{K}\alpha_j^o)$   
**end for**  
 $y_{res} = y - pr$   
 $\alpha_i \leftarrow update(\mathcal{K}, y_{res}, \alpha_i, g_i)$  Update  $\alpha_i$  using equation (16)  
 $\alpha_i \leftarrow \frac{\alpha_i}{\|\alpha_i\|}$   
 $g_i \leftarrow \{\mathcal{K}, \alpha_i, y\}$  Using equation (19)  
**end for**  
**for**  $i=1:M$  **do**  
 $pr \leftarrow pr + g(\mathcal{K}\alpha_i)$  Calculate the prediction  
**end for**  
 $err = \|y - pr\|$  Calculate the prediction error  
**end while**

---

**Algorithm 1.** Sequential encoding of nonlinearity.

$$g = B(\mathcal{K}\alpha)(B(\mathcal{K}\alpha)^T B(\mathcal{K}\alpha) + \Omega_{B(\mathcal{K}\alpha)})^{-1} B(\mathcal{K}\alpha)^T y. \quad (18)$$

As can be seen, both the update equations for  $g$  and  $w$  only depend on the inner products of the transformed data points into the high dimensional space. In this way, we take advantages of a high dimensional space in discerning any structural nonlinearity of data points without the need for an actual transformation. The update procedure of our method is shown in Algorithm 1.

## 5. Narrow Band Spectral Analysis (NBSA) of band-limited signals for drowsiness detection

This section describes the role of the proposed method in detecting drowsiness levels upon analyzing EEG data. Let  $x_p(t)$  be a set of continuous EEG signals collected from  $P$  different channels. Given that EEG is susceptible to different sources of artifacts, data pre-processing is first required before extracting the features. For this purpose, a second-order Butterworth bandpass filter with cut-off frequencies of 1 and 75 Hz is applied to the raw EEG signals to reduce the influences of the transient noise. To achieve a real-time estimation scheme, any further preprocessing steps like Blind Source Separation (BSS) of data or statistical analysis for excluding its contaminated parts are ignored.

To extract features, the filtered signals are first partitioned into multiple disjunct  $m$ -second epochs. Each epoch is decomposed into multiple band-limited signals (for this purpose, we use the wavelet decomposition method proposed in [53]) and then, is characterized by LE defined as follows:

$$LE(p, q) = \log\left(\frac{1}{K} \sum_{t=1}^K x_p^q(t)^2\right), \quad (19)$$

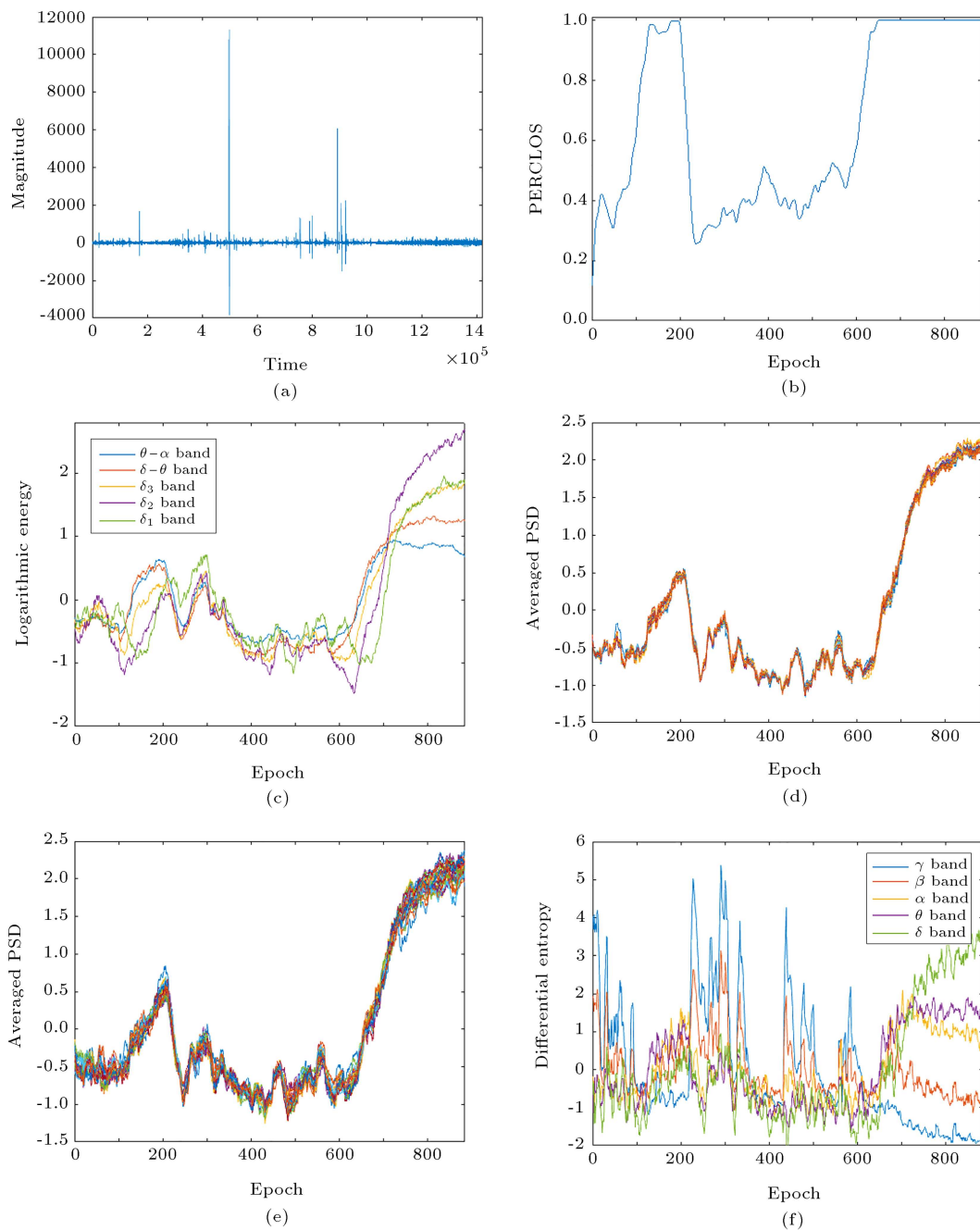
where  $x_p^q(t)$  is the limited-band signal within a frequency band  $q$  and  $K$  is the corresponding epoch length. This study uses the five conventional frequency bands of EEG signals including  $\delta_1$  (0.2–0.8 Hz),  $\delta_2$  (0.8–1.6 Hz),  $\delta_3$  (1.6–2.8 Hz),  $\theta_1$  (2.8–6.2 Hz), and  $\alpha_1$  (6.2–10 Hz).

In the following, we address the question of whether the narrow band frequency analysis of a band-limited signal is an appropriate tool for describing its fatigue-related characteristics. Mental fatigue manifests itself mostly as variations of the activity of some frequency bands including  $\alpha$ ,  $\theta$ , and  $\delta$ . Therefore, it is common to decompose EEG data first into multiple band-limited signals and then, extract the features of each band. The issue is that, most of the existing methods in the literature usually consider the Narrow Band Spectral Analysis (NBSA) as the most promising approach to characterizing the EEG epochs. However, this may not provide the most accurate and convenient way. In many studies, PSD is derived to scrutinize the narrow-band frequency characteristics where the strength of signal variations is computed as a function of different frequencies. Due to the noisy nature of the EEG data, authors usually tend to estimate their PSDs using spatiotemporal algorithms such as Short-Time Fourier Transform (STFT), AutoRegressive (AR) model, and Welch's technique. However, such schemes usually cause a heavy computational burden. Moreover, NBSA would only be effective when data includes some major cyclical components repeated with high frequencies, just like what occurs in experiencing epileptic seizures or triggering emotions. On the contrary, mental fatigue includes a wider range of neurological responses occurring over a broader range of frequencies where a far NBSA may lead to oversampling of the frequency elements. On the other hand, regardless of the main frequency origin of fatigue components, they can be detected easily using the magnitude of their variations compared to the patterns of the alertness regions. Figure 1 presents the features extracted by our method (LE) for one of the EEG channels compared to a set of the state-of-the-art frequency domain algorithms including Lin' method [36] using covariance-based Auto Regressive Power Spectral Density (ARPSD) algorithm [53], Welch's technique [54], and Dominant Frequency (DF) [55].

Motivated by the works in [56,57] in which the fatigue states are mainly attributed to the slower waves, in this paper utilizes five slow frequency bands, namely  $\delta_1$  (0.2 – 0.8 Hz),  $\delta_2$  (0.8 – 1.6 Hz),  $\delta_3$  (1.6 – 2.8 Hz),  $\delta - \theta$  (2.8 – 6.2 Hz), and  $\theta - \alpha$  (6.2 – 10 Hz) for all the techniques except for DF algorithm whose frequency bands are selected according to the emphatic recommendation of the original paper. Note that both LE and DE methods are employed to calculate the local energy of EEG signals. However, DE requires the appropriate size of PSD window and the overlapping percentage to be previously determined, which should be established according to the frequency range of input signals. To avoid such complexities, these values are directly taken from the original paper. For Welch's method, we divide each epoch into 2-secant segments with a 50% overlap. All the segments are extended to 256 pints and then, fed into the Welch's function. We use only the first 30 frequency bins (from 1 to 30 Hz) of each PSD as the features of the corresponding epoch. The spectral patterns are then converted into the logarithmic scale to make a more fairly distributed pattern called Logarithmic Spectral Density (LSD). Finally, the estimated densities of  $\theta - \alpha$ ,  $\delta - \theta$ ,  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$  bands are averaged to form the final feature vector. For ARPSD, we use a 4th order AR model and 30 frequency bins to represent a PSD vector. Like Welch's method, PSDs are converted into the logarithmic scale and averaged over different bands to form the final feature vector. For DF, we use fast Fourier transform with Hann window and consider the frequency with the largest average power as the DF feature of each segment. As can be seen, our method can uniquely characterize the frequency sub-bands of EEG data to explore different aspects of its neurological features that contribute differently to recognizing drowsiness states of different subjects according to their physical and behavioral characteristics. For example, the feature extracted from the  $\theta - \alpha$  band (marked in blue) is more consistent with the ground truth fatigue states shown in Figure 1(b). However, in the case of ARPSD and Welch's methods, averaging of different frequency bands causes the features to be biased toward the characteristics of the  $\delta$  band which corresponds to the purple curve in our method. In the case of DF, although different frequency sub-bands are individually characterized, a low signal-to-noise ratio and heavy computational burden hinder its application for detecting fatigue states in a real-world scenario.

## 6. Experimental results

This section is concerned with evaluating the performance of the proposed method in SEED VIS EEG database from two different perspectives: (1) analyzing the informativeness of the LE features for charac-



**Figure 1.** (a) Original EEG signal, (b) PERCLOS indices, the features extracted by (c) LE, (d) Lin' method using ARPSD (each color represents one bin of PSD vectors), (e) Lin' method using Welch's technique (each color represents one bin of PSD vectors), and (f) DF. As can be seen the  $\delta - \theta$  band of LE provides the most similar curve to the PERCLOS.

terizing fatigue-related features in EEG data; and (2) evaluating the regression performance of RKPPR compared with a set of linear and nonlinear regressors and investigating the overall performance of our method (LE+RKPPR) with a set of the state-of-the-art continuous fatigue estimation algorithms. Moreover, as a general-purpose framework, we evaluate the performance of our method in Parkinson's Disease (PD) prediction which, like fatigue detection, is a very low dimensional regression problem.

### 6.1. SEED VIS database

SEED VIS is a fatigue-related EEG database collected from 23 volunteers (12 women and 11 men with the average age of 23.3 and standard deviation of 1.14 years) during a simulated driving task. The simulator consists of a real vehicle without any motion platform, a video wall consisting of 9 rectangular flat monitors, and a software interface establishing the relationship between the wall and the vehicle. Unlike single seat driving simulators, here, the monotonous environment



of the vehicle helps induce much fatigue, leading to a more realistic driving condition. Moreover, steering wheel, electronic throttle, and brake pedal of the simulator as well as providing sufficient control on the speed, acceleration, and direction of the vehicle on the wall help volunteers experience a feeling like driving in a real environment. Volunteers are asked to avoid any caffeine, tobacco, or alcohol and get sufficient sleep per day before the experiment. None of them suffer from any vision problems and all are versed in how the system works at the pre-test phase. According to the circadian rhythm of the body, all the experiments were performed in the early afternoon or late night to induce maximum fatigue. Each experiment took about 2 hours during which the brain activity was uninterruptedly recorded by a NeuroScan system at a sampling rate of 1000 Hz. Due to previous findings on the key brain areas engaged in fatigue, SEED VIS only recorded the signals from the temporal and posteriors sites. Unlike traditional dual-task strategies, SEED VIS uses an eye tracking based technique for labeling the EEG data being simultaneously recorded on the system. For this purpose, volunteers were asked to wear SensoMotoric Instrument (SMI) eye tracking glasses when driving the vehicle. Each pupillary image taken by the glasses is processed to calculate the PERcentage value of eye CLOSure (PERCLOS) as follows:

$$PERCLOS = \frac{t_b + t_c}{t_b + t_f + t_s + t_c}, \quad (20)$$

where  $t_b$ ,  $t_c$ ,  $t_f$ , and  $t_s$  denote the durations of blinks, eye closure, fixations, and, saccades, respectively, during a recorded video. The PERCLOS values are finally smoothed by the moving average filter and considered as the fatigue indices. The indices range between 0 and 1, where 0 stands for the least drowsy level and 1 indicates the most drowsy state.

### 6.2. Analysis of informativeness measures

In this section, the similarity of LE features to the PERCLOS indices is examined without using any types of regression algorithms. For this purpose, a (0.2–75 Hz) second-order Butterworth filter is first applied to EEG data to remove its environmental artifacts and noises. The free-noise signals are then downsampled to 40 Hz. This efficiently reduces the computational complexity of further processing. Next, the data are decomposed into five frequency bands, each of which is then partitioned into 8-secant non-overlapping windows. For feature extraction, LEs are estimated for all the limited band signals of each window. The feature vectors are finally smoothed by the moving average filter with the window length of 200 s ( $8 \times 25$ , where 25 is the number of windows) to reduce the influence of the remaining noises. Finally, LE features extracted from all the frequency bands are concatenated to form the

feature vector of each window. Note that as the average feature vector of the previous window is available, only a total of 2 arithmetic operations (one subtraction and one addition) will be needed to update the average vector for the current window. Therefore, this method can be easily applied to online scenarios. Figure 2 shows the results of the decomposition for a typical segment of an EEG data.

As can be seen,  $\theta - \alpha$  and  $\delta - \theta$  rhythms provide more even distributions of energy in regions with approximately the same PERCLOS indices. Therefore, it is expected for them to be greatly similar to the PERCLOS values. To measure the similarity, the correlation coefficient is used as a quantitative measure of the linear relationship between the PERCLOS vector  $p$  and feature vector  $f$ , which is defined as follows:

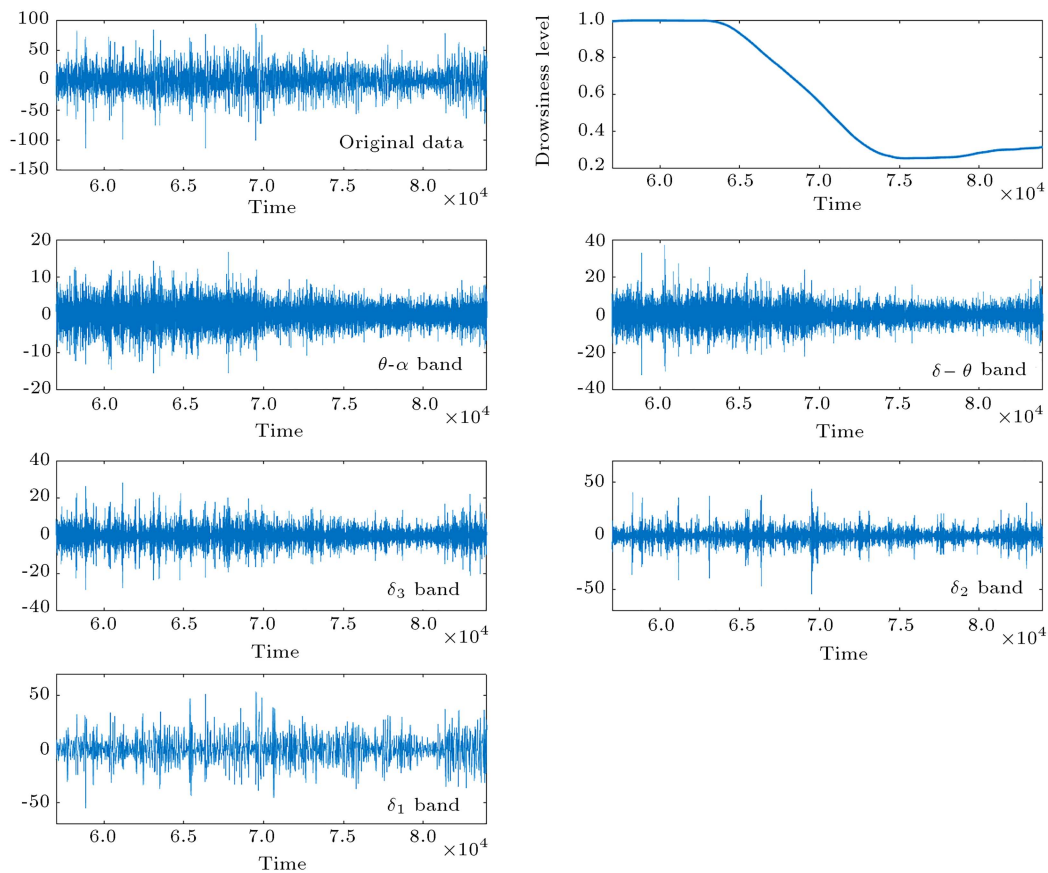
$$CC = \frac{\sum_{i=1}^N (p_i - \bar{p})(f_i - \bar{f})}{\sqrt{\sum_{i=1}^N (p_i - \bar{p})^2 \sum_{i=1}^N (f_i - \bar{f})^2}}, \quad (21)$$

where  $\bar{p}$  and  $\bar{f}$  denote the average of the PERCLOS  $p$  and the feature vector  $f$ , respectively, and  $N$  is the number of elements for each of the vectors and is equal to 885. Figure 3 shows the average CCs over different channels for each subject of the SEED VIS database.

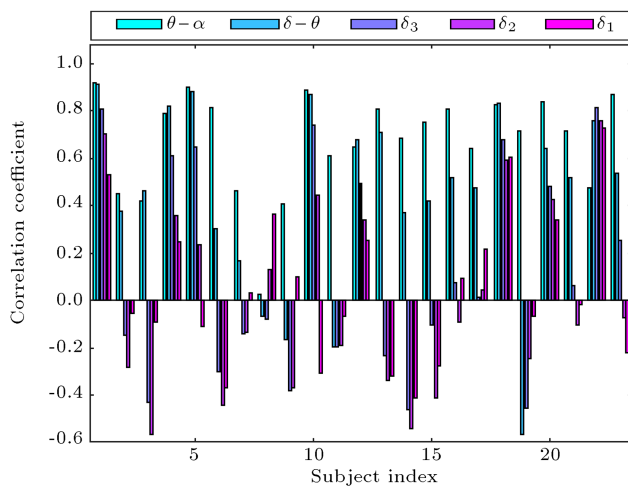
Note that a negative CC means that, an increase in vigilance level is directly associated with a decrease in the activity of the corresponding frequency band. It is clear that the features extracted from  $\gamma$  and  $\beta$  rhythms produce the highest CCs among the considered frequency bands which can be quite consistent with our previous finding on the distribution of energy.

From Figure 3, it can be found that fatigue reveals different frequency dependent behaviors across individuals. Therefore, it is impossible to derive a common brain pattern to characterize fatigue-related frequency activities. As a case in point, consider the 18th and 19th subjects. For the 18th user, the change of LE features in all the frequency bands follows the same trend as the PERCLOS indices. However, regarding the 19th user,  $\delta - \theta$ ,  $\delta_3$ ,  $\delta_2$ , and  $\delta_1$  activities are quite in reverse.

Additionally, it is interesting to study how the neural patterns of these individuals change over different states of vigilance. For this purpose, we define three different vigilance levels: awake ( $PERCLOS < 0.45$ ), tired ( $0.45 < PERCLOS < 0.75$ ), and drowsy ( $PERCLOS > 0.75$ ). For each level, the CCs are averaged over the points falling into the corresponding ranges. Then, the CC vector associated with the minimum value of the PERCLOS indices, which denotes the most alert pattern, is subtracted from each of the averaged CC vectors. The resulting differential coefficients are then mapped on a toy scalp map, which is referred to as the differential neural pattern.



**Figure 2.** Decomposition scheme for a typical segment of an EEG data.



**Figure 3.** Average CCs of LE features per frequency band for different subjects.

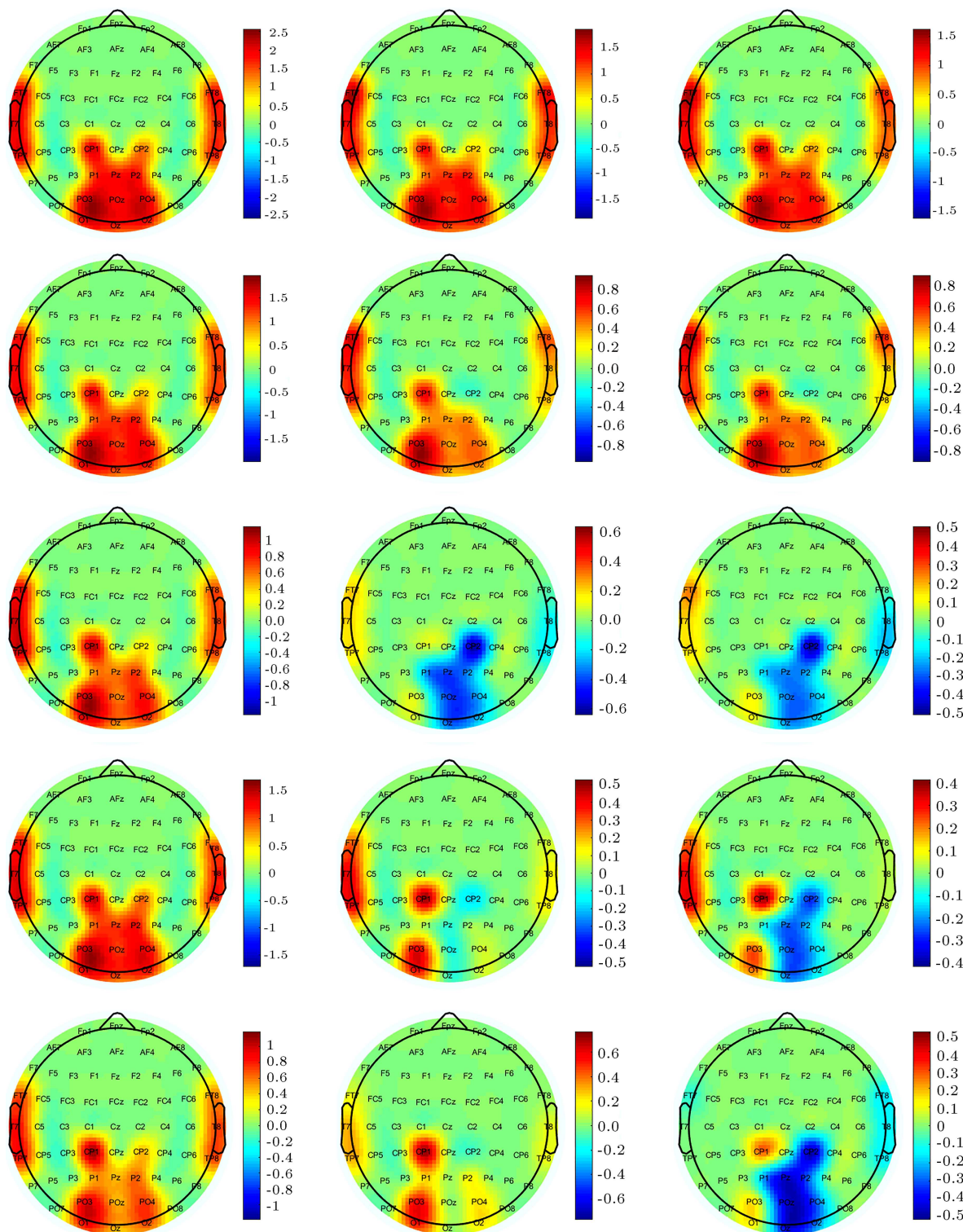
Figures 4 and 5 show the differential patterns for both of the subjects. As can be seen, there are significant differences between the neural patterns of different subjects. Therefore, a subject-specific training process would be necessary to estimate the vigilance patterns of each subject. Recent research has shown that transfer learning can be efficiently used for aligning the fatigue related features and generalizing the trained models.

However, to ensure greater efficiency, we focus on the subject-specific strategy and use it in the subsequent experiments.

### 6.3. Regression performance

This section evaluates the performance of our RKPPR algorithms compared to the PPR model as the baseline method. The EEG data and the protocol used in this experiment are the same as those in [25]. Accordingly, the entire data from one experiment is partitioned into 5 equal parts (80% used for training and the remaining 20% for testing) and a 5-fold cross-validation strategy is applied to measure the performance of the model. More information on the splitting protocol can be found in [25]. Due to the not-so-high dimension of the input variables and the findings of the previous research [58], only a single intermediate term is selected for both of the PPR and RKPPR models. The Euclidean distance is used as the kernel function of the PKPPR model [60]. The RMSE values per channels for both of the models are presented in Table 1. The labels in the first row indicate the name of the EEG channels.

As can be seen, on average, our method achieves 8.79% smaller RMSE than the baseline algorithm PPR. This superiority can be attributed to the nonlinear mapping to the Hilbert space and the use of the con-

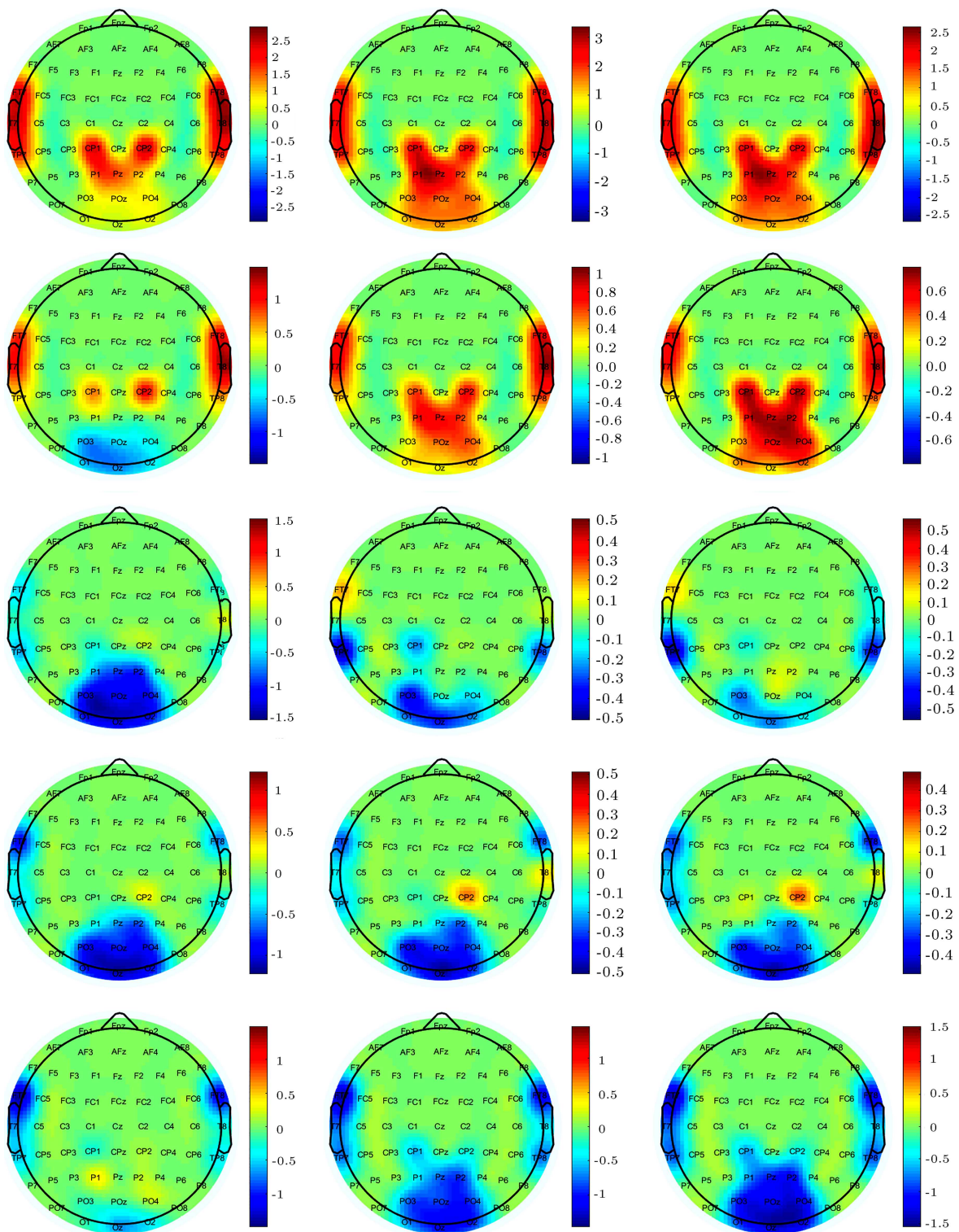


**Figure 4.** Differential neural patterns of drowsy (left column), tired (middle column), and awake (right column) states in five frequency bands;  $\delta_1$  (first row),  $\delta_2$  (second row),  $\delta_3$  (third row),  $\delta - \theta$  (fourth row),  $\theta - \alpha$  (last row) for the 18th subject.

**Table 1.** RMSE value of RKPPR and PPR on single channels for the SEED VIS database.

	FT7	FT8	T7	T8	TP7	TP8	CP1	CP2	P1	PZ	P2	PO3	POZ	PO4	O1	OZ	O2
PPR	0.1625	0.1506	0.1502	0.1450	0.1251	0.1316	0.1523	0.1456	0.1328	0.1274	0.1237	0.1141	0.1414	0.1225	0.1162	0.1408	0.1208
RKPPR	0.1386	0.1272	0.1320	0.1299	0.1192	0.1278	0.1363	0.1283	0.1237	0.1238	0.1177	0.1098	0.1238	0.1083	0.1151	0.1250	0.1130





**Figure 5.** Differential neural patterns of drowsy (left column), tired (middle column), and awake (right column) states in five frequency bands;  $\delta_1$  (first row),  $\delta_2$  (second row),  $\delta_3$  (third row),  $\delta - \theta$  (fourth row),  $\theta - \alpha$  (last row) for the 19th subject.

straint on the mapping weights, whose characteristics were described in detail in Section 3. The minimum value of RMSE is achieved in the channel PO3 in the posterior area. For RKPPR, this value is 0.1083 which is 5.08% lower than that achieved by PPR regressor.

The performance of these models is also evaluated with regard to the increased number of channels. The channels are added symmetrically, with the priority of adding the temporal channels. Despite the higher relevance of posteriors channels with vigilance, they

**Table 2.** RMSE value of RKPPR and PPR for cumulative structures of channels on the SEED VIS database.

Cumulative structure	PPR	RKPPR
FT7	0.1625	0.1395
FT7,8	0.1400	0.1228
FT7,8 + T7	0.1291	0.1181
FT7,8 + T7,8	0.1230	0.1143
FT7,8 + T7,8 + TP7	0.1228	0.1099
FT7,8 + T7,8 + TP7,8	0.1203	0.1080
FT7,8 + T7,8 + TP7,8 + CP1	0.1201	0.1084
FT7,8 + T7,8 + TP7,8 + CP1,2	0.1188	0.1068
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1	0.1126	0.1058
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z	0.1161	0.1052
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2	0.1173	0.1039
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3	0.1177	0.1013
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3,Z	0.1167	0.1033
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3,Z,4	0.1212	0.1026
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3,Z,4 + O1	0.1196	0.1010
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3,Z,4 + O1,Z	0.1189	0.1007
FT7,8 + T7,8 + TP7,8 + CP1,2 + P1,Z,2 + PO3,Z,4 + O1,Z,2	0.1181	0.1013

unfortunately cause great discomfort than the channels of the temporal area. Hence, the priority is given to the temporal channels. Table 2 shows the RMSE values for the cumulative structure of the channels.

For PPR, RMSE reaches a minimum value of 0.1126 and then, ascends when we keep adding the number of channels, indicating the weakness of PPR to integrate the information of different brain areas. In contrast, the minimum RMSE value for RKPPR, 0.1007 (10.72% smaller than that of PPR), is achieved based on the information of the first 16 channels. Therefore, one can infer that RKPPR outperforms PPR in that the former incorporates the fatigue patterns of different brain areas.

We also compare the performance of our RKPPR model with those of two well-known non-linear kernel-based regression algorithms called Kernel Support Vector Regression (KSVR) and GLM which have already been proven effective in real-world regression problems like analyzing the EEG data. For KSVR, we use the RBF kernel that has previously been successfully applied to fatigue data [25]. For GLM, two gamma and normal distributions are considered for the error in response variables. Moreover, we use the link functions of:

$$g_1(a) = a, \quad g_2(a) = \log(a), \quad g_3(a) = \log\left(\frac{a}{1-a}\right),$$

$$g_4(a) = \Phi^{-1}(a), \quad \text{and} \quad g_5(a) = \log(-\log(-(1-a))),$$

along with each error distribution. Table 3 shows the RMSE values obtained by the proposed method

compared with those obtained by the SVR, KSVR, and GLM algorithms.

It is clear that both KSVR and GLM algorithms exhibit high sensitivity to their adjustable parameters. Therefore, optimizing their performance would require quality assessment tests and manual inspection of the accuracy responses, which may hinder their application to real-world online situations. Finally, the overall performance of our method (LE+RKPPR) is compared with that of the single existing continuous method DE+SVR in the literature. The results are listed in Table 4. Note that these results have been achieved using the cumulative features of different brain areas rather than single channels used in Table 3. It is clear that the proposed method, even with a multichannel strategy that fulfils the need for seeking the best channel, significantly enhances the performance compared to DE+SVR.

#### 6.4. Prediction of PD progression

As a general purpose framework, the opportunity has arisen to apply RKPPR to any pragmatic regression problems. This section examines the performance of this method for prediction of PD. PD is the second most commonly diagnosed neurodegenerative disorder following Alzheimer's, currently affecting about six million people around the world. Due to the simultaneous effects on multiple brain areas, it usually tends to appear with a set of combined symptoms of motor and non-motor dysfunctions. At a clinical dissection, PD is usually characterized by two main neurochemical fact-

**Table 3.** The average RMSE value for different regression algorithms on LE features.

	GLM (normal)					GLM (gamma)					KSVR								
	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$\sigma = 1$	$\sigma = 2.5$	$\sigma = 5$	$\sigma = 7.5$	$\sigma = 10$	$\sigma = 12.5$	$\sigma = 15$	SVR	RKPPR
FT7	0.1515	0.1525	0.1519	0.1519	0.1489	0.1572	0.1527	0.1665	0.1667	0.1663	0.1428	0.1429	0.1483	0.1564	0.1644	0.1724	0.1793	0.1541	0.1386
FT8	0.1403	0.1421	0.1419	0.1418	0.1399	0.1420	0.1442	0.1643	0.1644	0.1679	0.1348	0.1303	0.1349	0.1418	0.1496	0.1572	0.1634	0.1411	0.1272
T7	0.1362	0.1392	0.1350	0.1352	0.1333	0.1472	0.1405	0.1545	0.1505	0.1491	0.1400	0.1367	0.1362	0.1444	0.1516	0.1578	0.1660	0.1401	0.1320
T8	0.1293	0.1362	0.1315	0.1314	0.1310	0.1341	0.1338	0.1415	0.1307	0.1411	0.1341	0.1306	0.1327	0.1402	0.1461	0.1523	0.1583	0.1323	0.1299
TP7	0.1180	0.1245	0.1146	0.1150	0.1149	0.1212	0.1231	0.1284	0.1385	0.1284	0.1278	0.1212	0.1212	0.1302	0.1370	0.1431	0.1497	0.1206	0.1192
TP8	0.1232	0.1318	0.1223	0.1223	0.1225	0.1266	0.1297	0.1197	0.1194	0.1203	0.1353	0.1249	0.1233	0.1300	0.1357	0.1450	0.1515	0.1245	0.1278
CPI	0.1439	0.1430	0.1419	0.1422	0.1401	0.1563	0.1456	0.1612	0.1611	0.1602	0.1435	0.1407	0.1470	0.1566	0.1614	0.1648	0.1678	0.1524	0.1363
CP2	0.1330	0.1343	0.1309	0.1311	0.1290	0.1415	0.1346	0.1333	0.1397	0.1333	0.1384	0.1310	0.1355	0.1427	0.1477	0.1514	0.1550	0.1381	0.1283
P1	0.1227	0.1246	0.1224	0.1226	0.1230	0.1342	0.1255	0.1369	0.1367	0.1282	0.1326	0.1274	0.1313	0.1386	0.1446	0.1496	0.1552	0.1303	0.1237
PZ	0.1235	0.1267	0.1226	0.1226	0.1230	0.1394	0.1262	0.1237	0.1246	0.1213	0.1281	0.1203	0.1235	0.1318	0.1381	0.1435	0.1488	0.1252	0.1238
P2	0.1161	0.1187	0.1146	0.1145	0.1136	0.1242	0.1202	0.1426	0.1423	0.1402	0.1270	0.1147	0.1191	0.1271	0.1343	0.1399	0.1453	0.1194	0.1177
PO3	0.1072	0.1101	0.1045	0.1047	0.1038	0.1255	0.1123	0.1091	0.1092	0.1086	0.1225	0.1080	0.1101	0.1196	0.1273	0.1353	0.1428	0.1087	0.1098
POZ	0.1303	0.1306	0.1276	0.1285	0.1270	0.1490	0.1351	0.1636	0.1707	0.1492	0.1308	0.1220	0.1266	0.1354	0.1429	0.1490	0.1552	0.1366	0.1238
PO4	0.1132	0.1146	0.1105	0.1105	0.1093	0.1230	0.1155	0.1191	0.1189	0.1128	0.1240	0.1113	0.1110	0.1171	0.1247	0.1320	0.1394	0.1148	0.1083
O1	0.1100	0.1143	0.1081	0.1081	0.1074	0.1298	0.1149	0.1114	0.1184	0.1107	0.1265	0.1126	0.1135	0.1237	0.1317	0.1397	0.1479	0.1109	0.1151
OZ	0.1234	0.1271	0.1220	0.1220	0.1205	0.1374	0.1271	0.1604	0.1651	0.1355	0.1284	0.1238	0.1276	0.1361	0.1444	0.1528	0.1606	0.1279	0.1250
O2	0.1139	0.1164	0.1119	0.1119	0.1110	0.1279	0.1180	0.1219	0.1217	0.1157	0.1260	0.1138	0.1132	0.1205	0.1290	0.1372	0.1445	0.1162	0.1130

**Table 4.** Overall performance of our method compared with the DE+SVR algorithm

	Posterior	Temporal
LE+RKPPR	0.1054	0.1080
DE+SVR	0.1429	0.1603
Improvement	26.24%	32.62%

ors: (1) massive loss of dopaminergic neurons in the substantia nigra which is manifested by some forms of reduced muscular control like speech disturbance, slow movement, postural instability, and resting tremor; and (2) the presence of  $\alpha$  synuclein aggregates within the cerebral cortex and limb system (linked to any impairments of memory, thinking, mood, and pain) as well as inferior temporal gyrus (responsible for any hallucinatory state). Although some genetic (e.g., gender, ethnicity, and heredity) and environmental (exposure to well water and pesticides as well as any imbalance of free radicals and antioxidants) agents have already been considered as the risk factors, the main cause of this disease has not been discovered yet. This is the reason why it cannot be detected by any single pathological biomarker. Currently, the most common diagnostic approach to PD is a clinical rating measure, called Unified Parkinson's Disease Rating Scale (UPDRS), which is a combined medical history and physical examination made up of three different parts: (1) evaluating the symptoms associated with the mentation, behavior, and mood of the patients, (2) examining the quality of performing daily routine automatized tasks, and (3) evaluating the motor deficits of limbs. Since the motor characteristics are of particular importance, the score of this section is presented either separately (motor-UPDRS) or in combination with the scores of the other parts (total-UPDRS). Despite its high accuracy, this measure requires to be administrated on a physical visit, which is a significant difficulty for many people with PD and one of the main factors of patients' delay which is significantly associated with the reduction of the number of receiving successful pharmaceutical therapies for alleviating the symptoms. Therefore, it is essential that a simple-to-use and noninvasive tool be devised that allows for early monitoring of patients beyond a clinical setting. Vocal impairment is one of the very first symptoms of PD that can be observed in about 90% of patients. From a broad perspective, this impairment can be channeled into four categories: (i) respiration: shortness of breath and symptoms of stridor arising from the diminished efficiency of respiratory muscles and an impaired chemoreceptor drive to breathing; (ii) phonation: impairment in producing individual phonetic segments deriving from any deficits in the vibration process of the vocal folds; (iii) articulation: impaired muscular control on articulatory organs such

as tongue, lip, and jaw causing some degrees of reduced speech intelligibility; and (iv) prosody: impairment of tonic properties of speech like rhythm and intonation. Despite being a more pragmatic test of speech impairment, respiration, articulation, and prosody measures suffer from a serious drawback, that is, they are heavily affected by linguistic components (e.g., intonation, tone, stress, and rhythm) and individual preferences.

This section addresses the telemonitoring problem of PD progression using RKPPR algorithm. For this purpose, each subject is requested to produce a sustained vowel at a normal pitch. Each signal is then represented by using multiple vocal attributes including cyclic (Jitter, Shimmer, Recurrence Period Density Entropy (RPDE), and Pitch Period Entropy (PPE)) and randomness related (Harmonics to Noise Ratio (HNR), Noise to Harmonics Ratio (NHR), and Detrended Fluctuation Analysis (DFA)) measures. Next, RKPPR is used to establish a mapping relationship between the vocal attributes and the UPDRS scores clinically estimated by a trained neurologist. For evaluation, we make use of Oxford Parkinson's Disease Telemonitoring (OPDT) database in which 5875 trials (sustained phonations of the vowel 'ahhh') from 42 patients (28 men and 14 women with the age ranging from 36 to 58 years) have been recorded by an Intel At-Home Testing Device at a frequency rate of 24 kHz and 16-bit resolution. None of the subjects were on any treatment at a six-month interval of the experiment. The UPDRS motor and total scores ranging from 0 to 108 and 0 to 176, respectively, where 0 indicates the best possible health state of subjects. Sixteen individual measurements of the above-mentioned attributes have been previously included in the database (Jitter(%), Jitter(Abs), Jitter: RAP, Jitter: PPQ5, Jitter: DDP, Shimmer, Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, Shimmer: APQ11, Shimmer: DDA, NHR, HNR, RPDE, DFA, PPE). To make a fairly even distribution, the attributes are converted into the logarithmic scale, which, in turn, allows for responding to the skewness of data towards large values.

The  $k$ -fold cross-validation strategy ( $k \in \{2, 3, 4, 5\}$ ) is utilized to evaluate the performance of the proposed method against two state-of-the-art regression algorithms, KSVR and GLM. For this purpose, subjects are divided into  $k$  nonoverlapping segments, where the models are fitted on one part and evaluated on the remaining ones. As shown in the previous section, RBF is utilized as the kernel function of KSVR. Moreover, gamma and normal distributions along with the four link functions of  $g_1$ ,  $g_2$ ,  $g_3$ , and  $g_4$  are considered to be the modeling assumptions of the GLM method. For all the algorithms, prediction is performed for both the motor and total scores of UPDRS. Tables 5 and 6 list the averaged RMSE value of the different algorithms on OPDT database. As can

**Table 5.** RMSE values for motor-UPDRS prediction using different regression algorithms.

	GLM (normal)				GLM (gamma)				KSVR							RKPPR
	$g_1$	$g_2$	$g_3$	$g_4$	$g_1$	$g_2$	$g_3$	$g_4$	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
2-fold	8.1293	8.1272	7.7146	7.0737	6.9380	6.9313	7.7711	7.6766	7.6937	7.7176	7.7101	7.7007	7.7090	7.7305	7.7305	6.7220
3-fold	8.1288	8.1274	7.8185	7.1919	7.0512	7.0323	7.8302	7.6878	7.7057	7.7309	7.7816	7.7082	7.7198	7.7426	7.7426	6.8307
4-fold	8.1292	8.1281	7.8705	7.3500	7.2345	7.2043	7.8525	7.6961	7.7132	7.7426	7.7404	7.7229	7.7315	7.7611	7.7611	6.9961
5-fold	8.1289	8.1279	7.9130	7.4483	7.3224	7.2957	7.8766	7.7078	7.7215	7.7427	7.8723	7.7402	7.7458	7.7650	7.7650	7.0644

**Table 6.** RMSE values for total-UPDRS prediction using different regression algorithms.

	GLM (normal)				GLM (gamma)				KSVR							RKPPR
	$g_1$	$g_2$	$g_3$	$g_4$	$g_1$	$g_2$	$g_3$	$g_4$	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
2-fold	10.7083	10.7065	10.1371	9.1506	8.9856	9.0327	10.3743	10.1771	10.2006	10.2245	10.2777	10.1995	10.2166	10.2376	10.2376	8.6652
3-fold	10.7081	10.7063	10.2658	9.3820	9.2174	9.2601	10.4391	10.1796	10.2047	10.2301	10.3081	10.1949	10.2149	10.2385	10.2385	8.8330
4-fold	10.7068	10.7055	10.3702	9.6046	9.4415	9.4571	10.4720	10.1986	10.2226	10.2463	10.3251	10.2249	10.2422	10.2639	10.2639	9.0803
5-fold	10.7078	10.7068	10.4244	9.7131	9.5812	9.5850	10.4764	10.2001	10.2223	10.2448	10.2937	10.2299	10.2432	10.2617	10.2617	9.1867

**Table 7.** RMSE values for motor-UPDRS prediction using the gender based regression strategy.

	GLM (normal)				GLM (gamma)				KSVR							RKPPR
	$g_1$	$g_2$	$g_3$	$g_4$	$g_1$	$g_2$	$g_3$	$g_4$	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
Men	7.8912	7.8895	7.4346	6.8642	6.7203	6.7405	7.6037	7.4841	7.4955	7.5108	7.7310	7.4960	7.5128	7.5334	7.5334	6.5517
Women	8.6173	8.6125	8.0897	6.7631	6.3463	6.2518	7.5897	7.3007	7.3594	7.4491	7.5544	7.3642	7.4657	7.6143	7.6143	5.9892

**Table 8.** RMSE values for total-UPDRS prediction using the gender based regression strategy.

	GLM (normal)				GLM (gamma)				KSVR							RKPPR
	$g_1$	$g_2$	$g_3$	$g_4$	$g_1$	$g_2$	$g_3$	$g_4$	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
Men	11.0112	11.0091	10.3031	9.3571	9.1981	9.2465	10.7526	10.4938	10.5147	10.5370	10.6275	10.5116	10.5356	10.5550	10.5550	8.9530
Women	9.8477	9.8460	9.4451	8.0048	7.5598	7.4818	8.6609	8.3788	8.4725	8.5826	8.8387	8.4273	8.5580	8.7023	8.7023	7.0268

be seen, our method achieves the best performance on both the motor and total sections. Moreover, all the algorithms have a better performance on motor score prediction than the total one. This supports the notion that as a motor sign, vocal attributes tend to be more closely connected to other motor dysfunctions than the non-motor symptoms.

We also evaluate how gender differences affect the prediction accuracy of the algorithms. For this purpose, two regression models are individually established for men and women, which allow for encoding the differences of their vocal folds into the regression process. Tables 7 and 8 show the RMSE values for predicting motor- and total- UPDRS measures using gender model-based algorithms and 2-fold cross-validation strategy. As can be seen, almost all the algorithms enjoy better performance than their corresponding mixed gender-based models. Specifically, in the motor section, gender-based RKPPR archives 2.5% and 10.9% lower RMSE values for men and women, respectively, than the corresponding mixed gender-based algorithm. However, for the total section, this value for the gender-based algorithm is 3.3% higher, but 18.9% lower than that of the mixed gender-based strategy.

In a follow-up experiment, we provide a comparison between our method and two well-known probabilistic regression algorithms including Bayesian Linear Regression (BLR) [62] and Gaussian Process (GP) [63]. Following the previous experiments, an evaluation is performed using the K-fold cross-validation strategy ( $K = 4$  for OPDT and  $K = 5$  for EEG VIS) while extracting the features as in Section 5. Table 9 lists the RMSE value for both of the databases. For the GP, five different covariance functions including exponential, squared exponential, Matern (with parameter 3/2), Matern (with parameter 5/2), and rational quadratic kernels are used that allow for a comprehensive analysis of this measurement. For EEG VIS, we proceed with the experiments along with the single P1 which has a moderate performance compared to the other channels provided in this database. In the case of the BLR, a normal-inverse-gamma conjugate model is used with a prior distribution of  $\beta|\sigma^2 \sim N_p(\mu, \sigma^2 V)$ , where  $p$  is the number of attributes. As can be seen, our method dominates all the rival algorithms in both of the databases. It is noteworthy that the performance of these algorithms heavily depends on the correct selection of the parameters as well as a proportionate



**Table 9.** RMSE values for motor-UPDRS prediction, a comparison with stochastic algorithms.

Method	OPDT database	SEED VIS database
RKPPR	6.9290	0.1237
GP (Squaredexponential)	7.1190	0.1717
GP (Exponential)	7.1420	0.1301
GP (Matern3/2)	7.2803	0.1463
GP (Matern5/2)	7.1866	0.1557
GP (Rationalquadratic)	7.1059	0.1369
BLR	7.6842	0.1266

distribution function, which is well eliminated by our proposed algorithm.

## 7. Conclusion

This paper proposed a novel regularized sequential projection pursuit algorithm to deal with low-dimensional regression problems like EEG-based fatigue detection. The key idea was to linearly embed a nonlinear mapping of the input variables into an intermediate space and then, to model the responses as a nonlinear function of the intermediate embedding. In this way, the nonlinearity can be encoded in two successive phases, first by projecting the input variables into a hidden space and then using their spline estimations to the output variables (leading to sequential encoding of the nonlinearity), which is quite proper for modeling a highly complex low dimensional structure. Moreover, this study proposed some constraints on the projection weights to avoid any overfitting of the intermediate space models. The performance of the proposed method was evaluated in the case of the publicly available SEED VIS database and an RMSE improvement of 8.79% was achieved in comparison to the conventional PPR algorithm. It is interesting that the price of such development is only the calculation of the similarity matrices that were considered by using the Euclidean distance, fulfilling the requirement for additional meta-parameters. The method performance was also evaluated for a Parkinson's Disease (PD) prediction task which further demonstrated the effectiveness of sequential nonlinearity encoding in dealing with very low dimensional regression problems.

## Acknowledgment

This work was partially supported by a grant from Cognitive Sciences and Technologies Council, and Research Office of the Sharif University of Technology.

## References

- Platte, K., Alleblas, C.C., Inthout, J., et al. "Measuring fatigue and stress in laparoscopic surgery: validity and reliability of the star-track test", *Minimally Invasive Therapy & Allied Technologies*, **1**, pp. 1–8 (2018).
- Ndaro, N.Z. and Wang, S.Y. "Effects of fatigue based on electroencephalography signal during laparoscopic surgical simulation", *Minimally Invasive Surgery*, **2018**, pp. 1–7 (2018).
- Amirian, I. "The impact of sleep deprivation on surgeons performance during night shifts", *Dan Med J*, **61**, pp. 1–14 (2014).
- Tuwairqi, K., Selter, J.H., and Sikder, S. "Assessment of surgeon fatigue by surgical simulators", *Open Access Surgery*, **8**, pp. 43–50 (2015).
- Dehais, F., Duprès, A., Di Flumeri, G., et al. "Monitoring pilots cognitive fatigue with engagement features in simulated and actual fNIRS-EEG passive BCI", *IEEE SMC* (2018).
- Zhang, L., Zhou, Q., Yin, Q., et al. "Assessment of pilots mental fatigue status with the eye movement features", In *International Conference on Applied Human Factors and Ergonomics*, Springer, pp. 146–155 (2018).
- Yan, R., Wu, C., and Wang, Y. "Exploration and evaluation of individual difference to driving fatigue for high-speed railway: a parametric SVM model based on multidimensional visual cue", *IET Intelligent Transport Systems*, **12**(6), pp. 504–512 (2018).
- Zhou, X., Yao, D., Zhu, M., et al. "Vigilance detection method for high-speed rail using wireless wearable EEG collection technology based on low-rank matrix decomposition", *IET Intelligent Transport Systems*, **12**(8), pp. 819–825 (2018).
- Liu, N.H., Chiang, C.Y., and Chu, H.C. "Recognizing the degree of human attention using EEG signals from mobile sensors", *Sensors*, **13**(8), pp. 10273–10286 (2013).
- Barwick, F., Arnett, P., and Slobounov, S. "EEG correlates of fatigue during administration of a neuropsychological test battery", *Clinical Neurophysiology*, **123**(2), pp. 278–284 (2012).
- NHTSA, Fatal Traffic Crash Data. <https://www.transportation.gov/briefing-room/nhtsa0117> (2017). [Online; accessed 06-October-2017].
- Hedlund, L., Gyllensten, A.L., and Hansson, L. "A psychometric study of the multidimensional fatigue

- inventory to assess fatigue in patients with schizophrenia spectrum disorders”, *Community Mental Health Journal*, **51**(3), pp. 377–382 (2015).
13. Meek, P.M., Nail, L.M., Barsevick, A., et al. “Psychometric testing of fatigue instruments for use with cancer patients”, *Nursing Research*, **49**(4), pp. 181–190 (2000).
  14. Lerdal, A., Wahl, A.K., Rustoen, T., et al. “Fatigue in the general population: a translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale”, *Scandinavian Journal of Public Health*, **33**(2), pp. 123–130 (2005).
  15. Whitehead, L. “The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures”, *Journal of Pain and Symptom Management*, **37**(1), pp. 107–128 (2009).
  16. Herlofson, K., Heijnen, C., Lange, J., et al. “Inflammation and fatigue in early, untreated Parkinson’s disease”, *Acta Neurologica Scandinavica*, **138**(5), pp. 394–399 (2018).
  17. Galarza, E.E., Egas, F.D., Silva, F.M., et al. “Real time driver drowsiness detection based on drivers face image behavior using a system of human-computer interaction implemented in a smart- phone”, In: *International Conference on Information Theoretic Security*, Springer, pp. 563–572 (2018).
  18. Jie, Z., Mahmoud, M., Stafford-Fraser, Q., et al. “Analysis of yawning behaviour in spontaneous expressions of drowsy drivers”, In: *Automatic Face & Gesture Recognition (FG 2018)*, 2018, 13th IEEE International Conference on IEEE, pp. 571–576 (2018).
  19. Kumar, A. and Patra, R. “Driver drowsiness monitoring system using visual behaviour and machine learning”, In: *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, IEEE, pp. 339–344 (2018).
  20. Sasikala, R., Suresh, S., Chandramohan, J., et al. “Driver drowsiness detection system using image processing technique by the human visual system”, *International Journal of Emerging Technologies in Engineering Research (IJETER)*, **6**(6), pp. 1–11 (2018).
  21. Liu, Y.T., Wu, S.L., Chou, K.P., et al. “Driving fatigue prediction with pre-event electroencephalography (EEG) via a recurrent fuzzy neural network”, In: *Fuzzy Systems (FUZZ-IEEE)*, 2016 IEEE International Conference on IEEE, pp. 2488–2494 (2016).
  22. McDonald, A.D., Lee, J.D., Schwarz, C., et al. “A contextual and temporal algorithm for driver drowsiness detection”, *Accident Analysis & Prevention*, **113**, pp. 25–37 (2018).
  23. Dimitrakopoulos, G.N., Kakkos, I., Dai, Z., et al. “Functional connectivity analysis of mental fatigue reveals different network topological alterations between driving and vigilance tasks”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(4), pp. 740–749 (2018).
  24. Jammes, B., Sharabty, H., and Esteve, D. “Automatic eeg analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions”, *Somnologie-schlafforschung und Schlafmedizin*, **12**(3), pp. 227–232 (2008).
  25. Zheng, W.L. and Lu, B.L. “A multimodal approach to estimating vigilance using eeg and forehead eeg”, *Journal of Neural Engineering*, **14**(2), 026017 (2017).
  26. Patel, M., Lal, S.K., Kavanagh, D., et al. “Applying neural network analysis on heart rate variability data to assess driver fatigue”, *Expert Systems with Applications*, **38**(6), pp. 7235–7242 (2011).
  27. Jung, S.J., Shin, H.S., and Chung, W.Y. “Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel”, *IET Intelligent Transport Systems*, **8**(1) pp. 43–50 (2014).
  28. Zhao, C., Zhao, M., Liu, J., et al. “Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator”, *Accident Analysis & Prevention*, **45**, pp. 83–90 (2012).
  29. Chowdhury, M.E., El Beheri, S.H., Albardawil, M.N., et al. “Driver drowsiness detection study using heart rate variability analysis in virtual reality environment”, In: *Qatar Foundation Annual Research Conference Proceedings*, 2018, HBKU Press Qatar, pp. 11–32 (2018).
  30. Lee, J., Kim, J., and Shin, M. “Correlation analysis between electrocardiography (ECG) and photoplethysmogram (PPG) data for drivers drowsiness detection using noise replacement method”, *Procedia Computer Science*, **116**, pp. 421–426 (2017).
  31. Chieh, T.C., Mustafa, M.M., Hussain, A., et al. “Driver fatigue detection using steering grip force”, In: *Research and Development, 2003. SCORED 2003. Proceedings. Student Conference on. IEEE*, pp. 45–48 (2003).
  32. Shi, L.C., Duan, R.N., and Lu, B.L. “A robust principal component analysis algorithm for EEG-based vigilance estimation”, In: *Engineering in Medicine and Biology Society (EMBC)*, 2013 35th Annual International Conference of the IEEE, IEEE, pp. 6623–6626 (2013).
  33. Mackiewicz, A. and Ratajczak, W. “Principal components analysis (PCA)”, *Computers and Geosciences*, **19**, pp. 303–342 (1993).
  34. Shi, L.C., Jiao, Y.Y., and Lu, B.L. “Differential entropy feature for EEG-based vigilance estimation”, In: *Engineering in Medicine and Biology Society (EMBC)*, 2013 35th Annual International Conference of the IEEE, IEEE, pp. 6627–6630 (2013).
  35. Kirk, B.P. and LaCourse, J.R. “Vigilance monitoring from the eeg power spectrum with a neural network”, In: *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, **3**, IEEE, pp. 1218–1219 (1997).

36. Lin, C.T., Chuang, C.H., Huang, C.S., et al. “Wireless and wearable EEG system for evaluating driver vigilance”, *IEEE Transactions on Biomedical Circuits and Systems*, **8**(2), pp. 165–176 (2014).
37. Armanfard, N., Komeili, M., Reilly, J.P., et al. “Vigilance lapse identification using sparse EEG electrode arrays”, In: *Electrical and Computer Engineering (CCECE), 2016 IEEE Canadian Conference on IEEE*, pp. 1–4 (2016).
38. Peng, H., Long, F., and Ding, C. “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(8), pp. 1226–1238 (2005).
39. Cao, L., Li, J., Sun, Y., et al. “Eeg-based vigilance analysis by using fisher score and pca algorithm”, In: *Progress in Informatics and Computing (PIC)*, **1**, IEEE, pp. 175–179 (2010).
40. Guo, Z., Pan, Y., Zhao, G., et al. “Detection of driver vigilance level using eeg signals and driving contexts”, *IEEE Transactions on Reliability*, **67**(1), pp. 370–380 (2018).
41. Chen, Y., Farrand, J., Tang, J., et al. “Relationship between amplitude of resting-state fNIRS global signal and EEG vigilance measures”, In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, IEEE, pp. 537–540 (2017).
42. Cynthia, A., Patricia, G., Nisrine, J., et al. “A new system for detecting fatigue and sleepiness using brain connectivity: EEG based estimation of fatigue, vigilance and sleepiness for drivers”, In: *Advances in Biomedical Engineering (ICABME), 2017 Fourth International Conference on IEEE*, pp. 1–4 (2017).
43. Baccalá, L.A., Sameshima, K., and Takahashi, D. “Generalized partial directed coherence”, In: *Digital Signal Processing, 2007 15th International Conference on IEEE*, pp. 163–166 (2007).
44. Pudil, P., Novovicova, J., and Kittler, J. “Floating search methods in feature selection”, *Pattern Recognition Letters*, **15**(11), pp. 1119–1125 (1994).
45. Li, M., Fu, J.W., and Lu, B.L. “Estimating vigilance in driving simulation using probabilistic PCA”, In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008, 30th Annual International Conference of the IEEE*, IEEE, pp. 5000–5003 (2008).
46. Tipping, M.E. and Bishop, C.M. “Probabilistic principal component analysis”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(3), pp. 611–622 (1999).
47. Mohammadpour, M. and Mozaffari, S. “Classification of EEG-based attention for brain computer interface”, In: *Intelligent Systems and Signal Processing (ICSPIS), 2017 3rd Iranian Conference on IEEE*, pp. 34–37 (2017).
48. Ning, T. and Bronzino, J.D. “Bispectral analysis of the rat eeg during various vigilance states”, *IEEE Transactions on Biomedical Engineering*, **36**(4), pp. 497–499 (1989).
49. Ouyang, T. and Lu, H.T. “Vigilance analysis based on continuous wavelet transform of eeg signals”, In: *Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on IEEE*, pp. 1–4 (2010).
50. Yu, H., Lu, H., Ouyang, T., et al. “Vigilance detection based on sparse representation of eeg”, In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, IEEE, pp. 2439–2442 (2010).
51. Wright, J., Yang, A.Y., Ganesh, A., et al. “Robust face recognition via sparse representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(2), pp. 210–227 (2009).
52. Friedman, J.H. and Stuetzle, W. “Projection pursuit regression”, *Journal of the American Statistical Association*, **76**(376), pp. 817–823 (1981).
53. Lago, P., Rocha, A., and Jones, N. “Covariance density estimation for autoregressive spectral modelling of point processes”, *Biological Cybernetics*, **61**(3), pp. 195–203 (1989).
54. Welch, P. “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms”, *IEEE Transactions on Audio and Electroacoustics*, **15**(2), pp. 70–73 (1967).
55. Shen, K.Q., Ong, C.J., Li, X.P., et al. “A feature selection method for multilevel mental fatigue EEG classification”, *IEEE Transactions on Biomedical Engineering*, **54**(7), pp. 1231–1237 (2007).
56. Plante, D.T., Goldstein, M.R., Cook, J.D., et al. “Effects of partial sleep deprivation on slow waves during non-rapid eye movement sleep: a high density EEG investigation”, *Clinical Neurophysiology*, **127**(2), pp. 1436–1444 (2016).
57. Neu, D., Mairesse, O., Verbanck, P., et al. “Non-rem sleep EEG power distribution in fatigue and sleepiness”, *Journal of Psychosomatic Research*, **76**(4), pp. 286–291 (2014).
58. Durocher, M., Chebana, F., and Ouarda, T.B. “A nonlinear approach to regional of hydrometeorology”, **16**(4), pp. 1561–1574 (2015).

## Biographies

**Mohsen Tabejamaat** was born in 1986 and received the BSc degree from Islamic Azad University, Iran. His research interests are in computational geometry, numerical simulations, visual analytic, data networking, semantic analysis, and bioinformatics.

**Hoda Mohammadzade** received her BSc degree from Amirkabir University of Technology (Tehran Polytechnic), Iran in 2004, the MSc degree from the University of Calgary, Canada in 2007, and the PhD degree from the University of Toronto, Canada in 2012, all in

Electrical Engineering. She is currently an Assistant Professor of Electrical Engineering at Sharif University of Technology, Tehran, Iran. Her research interests include machine learning, computer vision, signal and image processing, and bioinformatics.