Sequential Nonlinear Encoding; A Low Dimensional Regression Algorithm with Application to EEG based Driving Fatigue Detection

Mohsen Tabejamaat^a, Hoda Mohammadzade^{a,*}

^aDepartment of Electrical Engineering, Sharif University of Technology, Tehran 11155-8639, Iran.

Abstract

Background: 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 is one such low dimensional problems, playing a major role in reducing the risk of fatal accidents.

New Method: We introduce a novel Regularized Kernel Projection Pursuit Regression method which is a two-step nonlinearity encoding algorithm tailored for such very low dimensional problems as fatigue detection. This way, the data nonlinearity can be investigated from two different perspectives, first by transforming the data into a high dimensional intermediate space and then by using their spline estimations to the output variables which allows for a hierarchical unfolding of data.

Result: Experimental results on the SEED VIS database shows an average RMSE value of 0.1080% and 0.1054% respectively for the temporal and posterior areas of the brain. Our method is also validated by conducting some experiments on Parkinson's disease prediction which further demonstrate the efficiency of our method.

Conclusion: This paper proposes a novel regression algorithm to address the encoding problem of a highly complex low dimensional data, which is usually encountered in bio-neurological prediction tasks like EEG based driving fatigue detection.

Keywords: Brain Computer Interface, Electroencephalography, Fatigue Detection, Projection Pursuit Regression, Information

^{*}Corresponding Author

 $Email\ addresses:\ {\tt tabejamaat40660gmail.com}\ ({\rm Mohsen\ Tabejamaat}),\ {\tt hoda@sharif.edu}\ ({\rm Hoda\ Mohammadzade})$

1. Introduction

Mental fatigue is a brain psycho-physiological state which is clinically characterized by an increased production of cytokine or a decrease in the cortisol level and manifests as some forms of diminished mental alertness, listlessness, and even traces of languidness. Previous studies have shown that fatigue can generally occur in two different ways namely active and passive. Active fatigue refers to a situation where the brain is exposed by too much processing tasks (such as driving in a messy-busy road). This way, it would be triggered to release more and more dopamine in order to response to the need for making decisions faster and faster. However, the more the dopamine secretion, the more metabolic wastes would accumulate in the brain region. This eventually causes the brain to ask the frontal cortex for a temporarily shut down. On the contrary, passive fatigue is the feeling of understimulation, arising from doing a monotonous task for a long time (such as driving in a straight and monotonous road) which is associated with an increased Cortisol level in the brain, leading to some forms of slow nerve conduction and therefore reduced decision making performance.

Automatic detection of fatigue (vigilance) level has recently turned to be a hot research topic with a wide variety of applications in the areas like medicine [1, 2, 3, 4], transportation [5, 6, 7, 8], and education [9, 10]. For example, it can be used in the monitoring of the brain functions during an awake surgery, evaluating anesthetic effects on the brain, detecting fatigue driving to establish an early warning or providing the information for shipping or insurance companies, or setting ideal class length for students with different mental and physical capabilities. Among the applications, driving has aroused much more attention because it directly influences much more people's lives like drivers, passengers, pedestrians, and motor/pedal cyclists. In fact, fatigue is the starting point of a catastrophic chain which may ultimately lead to a fatal crash. First, it heavily lowers driver's concentration, then, his selective attention and decision speed and next, declines his eye hand coordination and consequently the control on steering wheel and pedals, and eventually crash!. Recent research shows that fatigue is responsible for 10% of road accidents all over the world [11]. Therefore, there is an urgent for developing such 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; (a) questionnaire based techniques that measure self-reported psychological qualities from the responses given to some situational questions [12, 13, 14, 15, 16]. The questionnaires are usually filled before and after a tedious task where the performance differences can reveal the extent of the fatigue state. (b) 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, 18, 19, 20]. (c) cognitive tasks that use the reaction time or the error rate of the responses to a set of visual stimuli [21, 22, 23]. (d) neurophysiological techniques such as electroencephalography (EEG) and electrooculography (EOG) [24, 25], and (e) physiological techniques such as electrocardiography (ECG) [26, 27, 28, 29, 30] or Grip Force (GF) [31]. In comparison, EEG has such advantages as non invasive nature, low price, robustness to biased responses (unlike psychometric and cognitive methods), and simplicity which make it much more convenient to use than the other rivals.

Most of the existing fatigue detection algorithms are generally established upon a classification strategy where each class is a qualitative measure of a 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 have been available for continuous annotation of EEG data: (1) manual annotation, (2) the use of face expression, and (3)by the help of simultaneous cognitive tasks. However, these methods are either time consuming and imprecise (like manual annotation and the use of face expression), or require additional interactions from drivers which 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 suffer from 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 which restricts their ability for handling highly complex data.

In this paper, we propose a novel double non-linearized pursuit regression algorithm which offers the ability to encode the nonlinearity of data in a sequential manner and is therefore appropriate for modelling such low dimensional problems as continuous fatigue detection. For this purpose, data is considered to be represented as a linear combination of its elements in the high dimensional Hilbert space (termed feature map) and then transformed to the continuous space of its output indices. This way, nonlinearity can be encoded in two different manners; first by mapping the data into the Hilbert space, and then by nonlinear transforming of the feature maps into the output labels. In addition, we introduce a constraint 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 propose to utilize the logarithmic energy 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 the paper is organized as follows. A brief review on the related works is provided in Section 2. Section 3 presents the main idea of the projection pursuit regression. Section 4 introduces Regularized Projection Pursuit Regression for detecting fatigue states in the high-dimensional Hilbert space. The evaluation of our method is described in Section 5. Finally, Section 6 draws the conclusion of our paper.

2. Related Works

This section provides a review on automatic fatigue detection algorithms using EEG signal which is presented from the view point of machine learning. However, there are also a variety of papers which address the issue from the psychological or medical points of view that 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 the Differential Entropy (DE) to characterise 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 Support Vector Regression (SVR) algorithm. The work 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 work 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 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 analysing

Connections amongst areas of the nervous system are of interest to researchers applying behavioural 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 Support Vector Machine (SVM) were respectively applied as the feature extraction and classification algorithms.

2.3. Time-domain techniques

In these methods, fatigue indices are directly estimated from the time domain information without the need for any transformation into the frequency domain. For example, Li et al. [45] utilized the probabilistic principle 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. Nine et al. [48] investigated the Gaussianity of EEG signal distribution as a measure of fatigue states. Ouyang et al. [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 the 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 a main disadvantage because they utilize a set of fixed link functions for encoding of 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 on each point of data. Unlike these methods, this paper aims to propose a sequential nonlinearity encoding technique which 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

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

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

Where w_q ; q = 1, ..., Q is a coefficient vector (directional vector) that projects the input variables into the qth intermediate feature map $f_q = X w_q$, g_q is the corresponding smoothing function, and Q denotes the number of intermediate projections. As g_q is a twice-differentiable function, the model is able to characterize the nonlinearity of the intermediate variables. The smoothness of q determines the fidelity of the model to the training data. The smoother the function q 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 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 crossvalidation 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 (equation (2)) and smoothing functions (equation (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 by using 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]:

$$g_{q}(w_{q}^{T}x_{i}) \approx g_{q}(w_{q,old}^{T}x_{i}) + g_{q}'(w_{q,old}^{T}x_{i})(w_{q} - w_{q,old})^{T}x_{i}$$

$$\Rightarrow \min_{w_{q}} \sum_{i=1}^{N} g_{q}'(w_{q,old}^{T}x_{i})^{2}[(w_{q,old}^{T}x_{i} - \frac{y_{i} - \sum_{q=1}^{Q}g(w_{q,old}^{T}x_{i})}{g'(w_{q,old}^{T}x_{i})}) - w_{q}^{T}x_{i}]^{2}$$
(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, it would be more similar to Single-Hidden-Layer (SHL) neural networks such as SHL MultiLayer 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 which 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, different from neural networks, PPR has another advantage to estimate its smoothing functions along with the projection weights that is very relevant for dealing with a nonlinear data.

Notwithstanding, PPR has its own drawbacks as well. For example, regarding a very complex data, 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 resultant 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 we used, the smoothing spline method, balances approximation errors with a derivative based measure of overall smoothness for a data set.

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

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

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

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

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

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

where $B_{i,j} = B_j(\mathfrak{o}_i)$, and $\Omega_{i,j} = \int (B_i(\mathfrak{o}))''(B_j(\mathfrak{o}))'' d\mathfrak{o}$. This equation can be finally solved by using the Least Squares (LS) algorithm;

$$\hat{\theta} = (B^T B + \lambda \Omega)^{-1} B^T y$$

$$\hat{y} = B\hat{\theta} = B(B^T B + \lambda \Omega)^{-1} B^T y$$
(6)

This way, each of the Q smoothing functions of equation (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 analysing multiple regression models. However, sometimes, data points have a rather complex structure to be modeled by the 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 on such extended intermediate representations. However, increasing the number of intermediate terms, would also dramatically increase the computational cost and calls for a heavy backfitting process. Note that, the response variable in equation (1) is a non-weighted sum of the intermediate terms. Therefore, the use of more than single term 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. This way, the complexity of data is simplified before transforming into the intermediate representation which further simplifies the structure of the fitted splines in 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 to project 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 to introduce 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 ϕ be a function that nonlinearly maps an *m*-dimensional input data into an infinite-dimensional RKHS \mathcal{F} so that $\phi : x_i \in \mathbb{R}^m \to \phi(x_i) \in \mathcal{F}$ 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 on the training set, we obtain:

$$X = [x_1, x_2, ..., x_N] \rightarrow \Phi(X) = [\phi(x_1), \phi(x_2), ..., \phi(x_N)]$$

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

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

s.t. $\min \|w_q\|_2$ (7)

Considering just one term and using the Lagrange multiplier theorem, the problem can be written in an unconstraint form:

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

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

$$J = \sum_{i=1}^{N} g'(w_{old}^{T}\phi(x_{i}))^{2} [w_{old}^{T}\phi(x_{i}) + \frac{y_{i} - g(w_{old}^{T}\phi(x_{i}))}{g'(w_{old}^{T}\phi(x_{i})))} - w^{T}\phi(x_{i})]^{2} + \mu \|w\|_{2}$$
(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 Equation (9) in a matrix form, we get the following expression:

$$J = T^T A T - T^T A \Phi(X) w - w^T \Phi(X)^T A T + w^T \Phi(X)^T A \Phi(X) w + \mu w^T w$$
(10)

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

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

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

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

$$w_{n} = (\Phi(X)^{T} A \Phi(X) + \mu I)^{-1} \Phi(X)^{T} A(\Phi(X)w) + \frac{y - g(\Phi(X)w)}{g'(\Phi(X)w)})$$
(12)

where $y = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{vmatrix}$, and *I* is identity matrix. As the weight vector *w* lie in the

span of $\phi(x_1), \phi(x_2), ..., \phi(x_N)$; there exists a coefficient vector in the Hilbert space which satisfies the following equation:

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

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

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

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

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

$$g(\Phi(X)w) = N(\Phi(X)w)\theta$$

= $N(\Phi(X)w) (N(\Phi w)^T N(\Phi w) + \Omega_{N(\Phi w)})^{-1} N(\Phi w)^T y$ (16)

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

$$g = B(\Phi\Phi^{T}\alpha) (B(\Phi\Phi^{T}\alpha)^{T} B(\Phi\Phi^{T}\alpha) + \Omega_{B(\Phi\Phi^{T}\alpha)})^{-1} B(\Phi\Phi^{T}\alpha)^{T} y$$
(17)

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

$$g = B(\mathcal{K}\alpha) \left(B(\mathcal{K}\alpha)^T B(\mathcal{K}\alpha) + \Omega_{B(\mathcal{K}\alpha)} \right)^{-1} B(\mathcal{K}\alpha)^T y$$
(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. This way, we take advantages of a high dimensional space in discerning any non-linearity structure of data points without the need for an actual transformation. The update procedure of our method has been shown in Algorithm 1.

5. Narrow Band Spectral Analysis of Band-limited Signals for drowsiness detection

In this section, we describe our method for detecting drowsiness levels through analysing EEG data. Let $x_p(t)$ be a set of continuous EEG signals collected from P different channels. As EEG is susceptible to different sources of artifacts, a data pre-processing is first required before extracting the features. For this purpose, a second order Butterworth bandpass filter with cuttoff frequencies of 1 and 75Hz is applied on the raw EEG signals to reduce the influences of the transient noise. To achieve a real-time estimation scheme, we ignore any further preprocessing steps like Blind Source Separation (BSS) of data or statistical analysis for excluding its contaminated parts.

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 [?]), and then characterized by the Logarithmic Energy (LE) which is defined as follows:

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

Algorithm 1 Sequential Encoding of Nonlinearity

Input: $X = \{x_1, x_2, ..., x_n\}$, $y = \{y_1, y_2, ..., y_n\}$ for p=1:n,q=1:n do $\mathcal{K}(p,q) = \|x_p - x_q\|$ end for for i=1:M do $\alpha_i \leftarrow \text{Initialize the weight vectors}$ $\alpha_i \leftarrow \frac{\alpha_i}{\|\alpha_i\|}$ $g_i \leftarrow \{\mathcal{K}, \alpha_i, y\}$ 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 $\alpha^{\mathbf{o}} \leftarrow \text{remove } \alpha_i$ $\mathbf{g^o} \leftarrow \text{remove } g_i$ for j=1:M-1 do $pr \leftarrow pr + g^o_j(\mathcal{K}\alpha^o_j)$ end for $y_{res} = y - pr$ $\alpha_i \leftarrow update(\mathcal{K}, y_{res}, \alpha_i, g_i \text{ Update } \alpha_i \text{ using equation (16)}$ $\alpha_i \leftarrow \tfrac{\alpha_i}{\|\alpha_i\|}$ $g_i \leftarrow \{\mathcal{K}, \alpha_i, y\}$ 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

where $x_p^q(t)$ is the limited-band signal within a frequency band q and K is the corresponding epoch length. In this study, we use the five conventional frequency bands of EEG signals including δ_1 (0.2-0.8 Hz), δ_2 (0.8-1.6 Hz), δ_3 (1.6-2.8 Hz), θ_1 (2.8-6.2 Hz), and α_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 fatiguerelated characteristics. Mental fatigue manifests itself mostly as variations in the activity of some frequency bands including α , θ , and δ . Therefore, it is usual to first decompose EEG data 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 for 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 the 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 also Welch's technique. However, such schemes usually causes 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 which occur over a broader range of frequencies where a far narrow band spectral analysis may lead to oversampling of the frequency elements. On the other hand, regardless the main frequency origin of fatigue components, they can be easily detected using the magnitude of their variations compared to the patterns of the alertness regions. Figure 1 illustrates 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 AutoRegressive PSD (ARPSD) algorithm [53], and 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 we utilize five slow frequency bands namely $\delta_1(0.2-0.8)Hz, \ \delta_2(0.8-1.6)Hz, \ \text{and} \ \delta_3(1.6-2.8)Hz, \ \delta-\theta(2.8-6.2)Hz, \ \text{and}$ $\theta - \alpha (6.2 - 10) Hz$ 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 their 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-second 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 $\theta - \alpha$, $\delta - \theta$, δ_1 , δ_2 , and δ_3 bands are averaged to form the final feature vector. For ARPSD, we use a 4th order autoregressive (AR) model and also 30 frequency bins to represent a PSD vector. Like welch's method, PSDs are converted to 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 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 in recognizing drowsiness states of different subjects according to their physical and behavioural 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, regarding ARPSD and Welch's methods, averaging over different frequency bands causes the features to be biased toward the characteristics of the δ band which corresponds to the purple curve in our method. In the case of DF, although different frequency sub-bands are individually characterized, it's low signal-to-noise ratio and heavy computational burden hinder its application for detecting fatigue states in a real-world scenario.

[Figure 1 about here.]

6. Experimental Results

In this section, we evaluate the performance of our method on SEED VIS EEG database from two different perspectives; (1) analysing the informativeness of the LE features for characterizing fatigue related features in EEG data, (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 for Parkinson's disease 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 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 more fatigue leading a more realistic driving condition. Moreover, steering wheel, electronic throttle and brake pedal of the simulator, providing sufficient control on the speed, acceleration and direction of the vehicle on the wall, all help the volunteers experience a feeling like driving in a real environment. Volunteers were asked to avoid any caffeine,

tobacco, or alcohol and have sufficient sleep a day before the experiment. None of them suffer from any vision problems and all were versed in how the system works via a pre-test phase. According to the circadian rhythm of body, all the experiments were performed in the early afternoon or late night to induce the maximum amount of 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 the previous findings on the key brain areas engaged in fatigue, SEED VIS has only recorded the signals from the temporal and posteriors sites. Unlike traditional dual-task strategies, SEED VIS uses an eye tracking based technique for labelling the EEG data being simultaneously recorded on the system. For this purpose, volunteers were asked to wear a SensoMotoric Instruments (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);

$$PERCLOS = \frac{t_b + t_c}{t_b + t_f + t_s + t_c} \tag{20}$$

where t_b , t_c , t_f , and t_c respectively denote the durations of blinks, eye closure, fixations, and, saccades 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 to 1, where 0 stands for the least drowsiness level and 1 indicates the most drowsiness state.

6.2. Analysis of Informativeness Measures

In this section, we examine the similarity of LE features to the PERCLOS indices without using any kinds of regression algorithms. For this purpose, a [0.2-75] Hz second order Butterworth filter is first applied on 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 the further processing. Next, the data is decomposed into the five frequency bands and each band is then partitioned into 8 second 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 200s (8×25) , 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, thus only a total of 2 arithmetic operations (one substraction and one addition) will be needed to update the average vector for the current window. Therefore, this method can be easily applied for online scenarios. Fig. 2 show the results of the decomposition for a typical segment of an EEG data.

[Figure 2 about here.]

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

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

where \overline{p} and \overline{f} respectively denote the average of the PERCLOS p and the feature vector f, 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 γ and β rhythms produce the highest CCs among the considered frequency bands which is quite consistent with our previous finding on the distribution of energy.

[Figure 3 about here.]

From Fig. 3, it can also be found that the 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 follow the same trend as the PERCLOS indices. However, regarding the 19th user, the $\delta - \theta$, δ_3 , δ_2 . and δ_1 activities are quite revers.

Additionally, it is interesting to study how the neural patterns of these individuals changes over different states of vigilance. For this purpose, we define three different vigilance levels named; 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 differential neural pattern. Figures 4 and 5 show the differential patterns for both 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, because of much better efficiency, we focus on the subject-specific strategy and use it in the subsequent experiments.

[Figure 4 about here.]

[Figure 5 about here.]

6.3. Regression Performance

In this section, we evaluate 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 is the same as the work in [25]. Accordingly, the entire data from one experiment is partitioned into 5 equal parts (80% is used for training and the remaining 20% is used for testing) and 5-folds cross validation strategy is conducted 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 also the finding of the previous research [58], only a single intermediate term is selected for both the PPR and RKPPR models. We use the Euclidean distance as the kernel function of the PKPPR model [?]. The mean RMSE values per channels for both the models are presented in Table 1. The labels in the first row indicate the name of the EEG channels.

[Table 1 about here.]

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 also the use of the constraint on the mapping weights whose characteristics were described in detail in Section III. The minimum value of RMSE is achieved by the channel PO3 in the posteriors area. For RKPPR, this value is 0.1083 which is 5.08% lower than that achieved by PPR regressor. The performance of these models are also evaluated with regard to increasing the 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 unfortunately cause more discomfort than the channels of the temporal area. That is way the priority is given to the temporal channels. Table 2 shows the RMSE values for the cumulative structure of the channels.

[Table 2 about here.]

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 areas of the brain. In contrary, the minimum RMSE value for RKPPR, 0.1007 (10.72% smaller than that of PPR), is achieved by using the information of the first 16 channels. Therefore, one can infer that RKPPR outperforms the PPR in terms of incorporating the fatigue patterns of the different brain areas.

We also compare the performance of our RKPPR model with two well-known non-linear kernel based regression algorithms named Kernel Support Vector Regression (KSVR) and Generalized Linear Model (GLM) which have already been proven to be effective in real-world regression problems like analysing the EEG data. For KSVR, we use the RBF kernel which 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 $g_1(a) = a$, $g_2(a) = log(a)$, $g_3(a) = log(\frac{a}{1-a})$, $g_4(a) = \Phi^{-1}(a)$, and $g_5(a) = log(-log(-(1-a)))$ along with each the error distribution. Table 3 shows the RMSE values for our method compared with those of the SVR, KSVR, and GLM algorithms.

[Table 3 about here.]

It is clear that, both the KSVR and GLM algorithms show a high sensitivity towards their adjustable parameters. Therefore, optimizing their performance would require the quality assessment tests and manual inspection of the accuracy responses which hinders their application for real-world online situations. Finally, we compare the overall performance of our method (LE+RKPPR) with the single existing continuous method in the literature DE+SVR. The results are listed in Table 4. Note that, these results has been achieved using the cumulative features of different brain areas and not the single channels used in Table 3. We can see that our method even with multichannel strategy (which fulfils the need for a search for the best channel) significantly improves the performance compared to DE+SVR.

[Table 4 about here.]

6.4. Prediction of Parkinson's disease progression

As a general-purpose framework, RKPPR offers the opportunity to be applied on any pragmatic regression problems. In this section, we examine the performance of this method for Parkinson's disease (PD) prediction. 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 tend to be presented with a set of combined symptoms of motor- and non-motor dysfunctions. At a clinical dissection, it is usually characterized by using two main neurochemical factors: (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 as well as resting tremor, and (2) the presence of α 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 imbalancement of free radicals and antioxidants) agents are already considered to be the risk factors, the main cause of this disease has not been discovered yet. This causes it can not be detected by any single pathological biomarker. Currently, the most common diagnostic way of PD is by 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 is of particular importance, score of this section are presented either separately (motor-UPDRS) or in a combination with the scores of the other parts (total-UPDRS). Despite the high accuracy, this measure requires to be administrated on a physical visit, which is a significant difficulty for many people with Parkinsons disease and one of the main factors of patient's delay that is significantly associated with a decrease in getting a successful pharmaceutical therapies for alleviating the symptoms. Therefore, it is essential to devise a simple-to-use and noninvasive tool 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 which causes some degrees of reduced speech intelligibility. (iv) prosody: impairment in tonic properties of speech like rhythm and intonation. Despite being a more pragmatic test of speech impairment, respiration, articulation, and prosody measures all suffer from a serious drawback that they are heavily affected by linguistic components (e.g., intonation, tone, stress, and rhythm) and individual preferences.

In this section, we address 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 for the six-month interval of the experiment. The UPDRS motor and total scores respectively range between 0 to 108 and 0 to 176 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 more fairly even distribution, we convert the attributes into the logarithmic scale that 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 evalu-

ate the performance of our 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 the one part and evaluated on the remaining ones. As 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_5 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. Table 5 and 6 list the average RMSE value of the different algorithms on OPDT database. As can 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.

[Table 5 about here.]

[Table 6 about here.]

We also evaluate how gender differences affects the prediction accuracy of the algorithms. For this purpose, two regression models are individually established for men and women which allows 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 have a better performance than their corresponding mixed gender based models. Specifically, in motor section, gender based RKPPR respectively archives 2.5% and 10.9% lower RMSE value for men and women than the corresponding mixed gender based algorithm. However, for total section, this value for gender based algorithm is respectively 3.3% higher but 18.9% lower than that of mixed gender based strategy.

[Table 7 about here.]

[Table 8 about here.]

In a follow up experiment, we provide a comparison between our method and two well known probabilistic regression algorithms including Bayesian Linear Regression (BLR) [?] and Gaussian Process (GP) [?]. Following the previous experiments, we perform the evaluation 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 the databases. For the Gaussian Process, we use five different covariance functions including exponential, squared exponential, Matern (with parameter 3/2), Matern (with parameter 5/2), and rational quadratic kernels that allows for a comprehensive analysis of this measurement. For EEG VIS, we proceed the experiments with the single P1 which has a moderate performance compared to the other channels provided in this database. As to the BLR, we use a normal-inverse-gamma conjugate model with a prior distribution of $\beta \mid \sigma^2 \sim N_p(\mu, \sigma^2 V)$, where p is the number of attributes. As can be seen, our method dominates all the rival algorithms in both the databases. It is noteworthy that the performance of these algorithms heavily depend on the correct selection of the parameters as well as a proportionate distribution function, which is well eliminated by our proposed algorithm.

[Table 9 about here.]

7. Conclusion

This paper proposed a novel regularized sequential projection pursuit algorithm for 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 model the responses as a nonlinear function of the intermediate embedding. 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 a sequential encoding of the nonlinearity), which is quite appropriate to model a highly complex low dimensional structure. Moreover, we proposed to constraint the projection weights to avoid any overfitting for the models of the intermediate space. The performance of our method was evaluated on the publicly available SEED VIS database and showed an RMSE improvement of 8.79% over the conventional PPR algorithm. It is interesting that the price of such development is only the calculation of the similarity matrices which are performed by using the Euclidean distance, fulfilling the requirement to any additional meta-parameters. The performance of our method was also evaluated on a PD prediction task which further demonstrated the effectiveness of sequential nonlinearity encoding for 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 Technol ogy.

4 Declarations of interest: none

5 Biography

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

10

Hoda Mohammadzade received her BSc degree from Amirkabir University of Technology (Tehran Polytechnic), Iran, in 2004, the MSc degree from the ¹³ University of Calgary, Canada, in 2007, and the PhD degree from the University
¹⁴ of Toronto, Canada, in 2012, all in electrical engineering. She is currently an
¹⁵ assistant professor of electrical engineering at Sharif University of Technology,
¹⁶ Tehran, Iran. Her research interests include machine learning, computer vision,
¹⁷ signal and image processing, and bioinformatics.

18 References

- Platte, K., Alleblas, C.C., Inthout, J., Nieboer, T.E.. Measuring fatigue and stress in laparoscopic surgery: validity and reliability of the star-track test. *Minimally Invasive Therapy & Allied Technologies* pp. 1–8 (2018).
- Ndaro, N.Z., Wang, S.Y.. Effects of fatigue based on electroencephalography signal during laparoscopic surgical simulation. *Minimally invasive surgery* 2018 (2018).
- Amirian, I.. The impact of sleep deprivation on surgeons performance
 during night shifts. *Dan Med J* 61 pp. B4912 (2014).
- Tuwairqi, K., Selter, J.H., Sikder, S.. Assessment of surgeon fatigue by surgical simulators. *Open Access Surgery* 8 pp. 43–50 (2015).
- 5. Dehais, F., Duprès, A., Di Flumeri, G., Verdière, K.J., Borghini, G.,
 Babiloni, F., Roy, R.N.. Monitoring pilots cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid
 fnirs-eeg passive bci. *IEEE SMC* (2018).
- 6. Zhang, L., Zhou, Q., Yin, Q., Liu, Z.. 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).
- 7. Yan, R., Wu, C., 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* (2018).
- 8. Zhou, X., Yao, D., Zhu, M., Zhang, X., Qi, L., Pan, H., Zhu, X.,
 Wang, Y., Zhang, Z.. Vigilance detection method for high-speed rail
 using wireless wearable eeg collection technology based on low-rank matrix
 decomposition. *IET Intelligent Transport Systems* (2018).
- 9. Liu, N.H., Chiang, C.Y., Chu, H.C.. Recognizing the degree of human attention using eeg signals from mobile sensors. Sensors 13(8) pp. 10273-10286 (2013).
- 47 10. Barwick, F., Arnett, P., Slobounov, S.. Eeg correlates of fatigue during
 48 administration of a neuropsychological test battery. *Clinical Neurophysiol-*49 ogy 123(2) pp. 278–284 (2012).

- 11. NHTSA, . Fatal Traffic Crash Data. https://www.transportation.gov/
 briefing-room/nhtsa0117 (2017). [Online; accessed 06-October-2017].
- Hedlund, L., Gyllensten, A.L., 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).
- Meek, P.M., Nail, L.M., Barsevick, A., Schwartz, A.L., Stephen, S.,
 Whitmer, K., Beck, S.L., Jones, L.S., Walker, B.L.. 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., Hanestad, B.R., Moum, T.. Fatigue in the general population: a translation and test of the psychometric
 properties of the norwegian version of the fatigue severity scale. *Scandina- vian journal of public health* 33(2) pp. 123–130 (2005).
- ⁶⁴ 15. Whitehead, L.. The measurement of fatigue in chronic illness: a systematic
 ⁶⁵ review of unidimensional and multidimensional fatigue measures. Journal
 ⁶⁶ of pain and symptom management 37(1) pp. 107–128 (2009).
- Herlofson, K., Heijnen, C., Lange, J., Alves, G., Tysnes, O.B., Friedman,
 J., Fagundes, C., Inflammation and fatigue in early, untreated parkinson's disease. Acta Neurologica Scandinavica (2018).
- 17. Galarza, E.E., Egas, F.D., Silva, F.M., Velasco, P.M., Galarza, E.D..
 Real time driver drowsiness detection based on drivers face image behavior
 using a system of human computer interaction implemented in a smartphone. In: International Conference on Information Theoretic Security.
 Springer pp. 563–572 (2018).
- 18. Jie, Z., Mahmoud, M., Stafford-Fraser, Q., Robinson, P., Dias, E.,
 Skrypchuk, L.. 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).
- ⁷⁹ 19. Kumar, A., Patra, R.. Driver drowsiness monitoring system using visual
 ⁸⁰ behaviour and machine learning. In: 2018 IEEE Symposium on Computer
 ⁸¹ Applications & Industrial Electronics (ISCAIE). IEEE pp. 339–344 (2018).
- 20. Sasikala, R., Suresh, S., Chandramohan, J., Valanrajkumar, M.. Driver
 drowsiness detection system using image processing technique by the hu man visual system. International Journal of Emerging Technologies in
 Engineering Research (IJETER) 6(6) pp. 1–11 (2018).
- ⁸⁶ 21. Liu, Y.T., Wu, S.L., Chou, K.P., Lin, Y.Y., Lu, J., Zhang, G.,
 ⁸⁷ Lin, W.C., Lin, C.T.. Driving fatigue prediction with pre-event elec ⁸⁸ troencephalography (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., Brown, T.L.. 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., Wang, H., Sgarbas, K.,
 Thakor, N., Bezerianos, A., Sun, Y.. Functional connectivity analysis
 of mental fatigue reveals different network topological alterations between
 driving and vigilance tasks. *IEEE transactions on neural systems and re- habilitation engineering* 26(4) pp. 740–749 (2018).
- ⁹⁹ 24. Jammes, B., Sharabty, H., Esteve, D.. Automatic eog analysis: A first
 ¹⁰⁰ step toward automatic drowsiness scoring during wake-sleep transitions.
 ¹⁰¹ Somnologie-schlafforschung und Schlafmedizin 12(3) pp. 227–232 (2008).
- 25. Zheng, W.L., Lu, B.L.. A multimodal approach to estimating vigilance
 using eeg and forehead eog. *Journal of neural engineering* 14(2) pp. 026017
 (2017).
- Patel, M., Lal, S.K., Kavanagh, D., Rossiter, P.. 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., 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., Zheng, C.. Electroencephalogram and
 electrocardiograph assessment of mental fatigue in a driving simulator. Ac cident Analysis & Prevention 45 pp. 83–90 (2012).
- 29. Chowdhury, M.E., El Beheri, S.H., Albardawil, M.N., Moustafa, A.K.M.N., Halabi, O., Kiranyaz, M.S.. Driver drowsiness detection study using heart rate variability analysis in virtual reality environment. In: *Qatar Foundation Annual Research Conference Proceedings*; vol. 2018. HBKU Press Qatar pp. 11–32 (2018).
- 30. Lee, J., Kim, J., 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).
- Chieh, T.C., Mustafa, M.M., Hussain, A., Zahedi, E., Majlis, B.Y..
 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., 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., Ratajczak, W.. Principal components analysis (pca).
 Computers and Geosciences 19 pp. 303–342 (1993).
- 34. Shi, L.C., Jiao, Y.Y., Lu, B.L.. Differential entropy feature for eegbased vigilance estimation. In: *Engineering in Medicine and Biology So- ciety (EMBC), 2013 35th Annual International Conference of the IEEE*.
 IEEE pp. 6627–6630 (2013).
- 35. Kirk, B.P., 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; vol. 3. IEEE pp. 1218–1219 (1997).
- 36. Lin, C.T., Chuang, C.H., Huang, C.S., Tsai, S.F., Lu, S.W., Chen,
 Y.H., Ko, L.W., 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., Pino, L.. 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., 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., Zhu, H., Yan, C.. Eeg-based vigilance analysis
 by using fisher score and pca algorithm. In: *Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on*; vol. 1. IEEE
 pp. 175–179 (2010).
- 40. Guo, Z., Pan, Y., Zhao, G., Cao, S., Zhang, J.. 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., Chen, Y., O'Keeffe, J., Shou, G.,
 Ding, L., Yuan, H.. 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., Mahmoud, H., Sandy, R.. 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., Takahashi, D.. Generalized partial directed coherence. In: *Digital Signal Processing*, 2007 15th International *Conference on*. IEEE pp. 163–166 (2007).
- 44. Pudil, P., Novovičová, J., Kittler, J.. Floating search methods in feature selection. *Pattern recognition letters* 15(11) pp. 1119–1125 (1994).
- 45. Li, M., Fu, J.W., 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., 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., 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., 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., 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., Liu, H., Lu, B.L.. 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).
- ¹⁹⁵ 51. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.. Robust
 ¹⁹⁶ face recognition via sparse representation. *IEEE transactions on pattern* ¹⁹⁷ analysis and machine intelligence 31(2) pp. 210–227 (2009).
- 52. Friedman, J.H., Stuetzle, W.. Projection pursuit regression. Journal of
 the American statistical Association 76(376) pp. 817–823 (1981).
- 53. Lago, P., Rocha, A., Jones, N.. Covariance density estimation for autore gressive 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., Hui, Z., Wilder-Smith, E.P., 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., Smith, R., Riedner, B.A.,
 Rumble, M.E., Jelenchick, L., Roth, A., Tononi, G., Benca, R.M., 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., Linkowski, P., Le Bon, O..
 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., Ouarda, T.B.. A nonlinear approach to regional flood frequency analysis using projection pursuit regression. *Journal*of Hydrometeorology 16(4) pp. 1561–1574 (2015).

221 List of Figures

222	1	(a) Original EEG signal, (b) PERCLOS indices, the features ex-	
223		tracted by (c) LE, (d) Lin' method using ARPSD (each color rep-	
224		resents one bin of PSD vectors), (e) Lin' method using Welch's	
225		technique (each color represents one bin of PSD vectors), and (f)	
226		DF. As can be seen the $\delta - \theta$ band of LE provides the most similar	
227		curve to the PERCLOS	28
228	2	Decomposition scheme for a typical segment of an EEG data	29
229	3	Average CCs of LE features per frequency band for different sub-	
230		jects.	30
231	4	Differential neural patterns of drowsy (left column), tired (middle	
232		column), and awake (right column) states in five frequency bands;	
233		δ_1 (first row), δ_2 (second row), δ_3 (third row), $\delta - \theta$ (fourth row),	
234		$\theta - \alpha$ (last row) for the 18th subject	31
235	5	Differential neural patterns of drowsy (left column), tired (middle	
236		column), and awake (right column) states in five frequency bands;	
237		δ_1 (first row), δ_2 (second row), δ_3 (third row), $\delta - \theta$ (fourth row),	
238		$\theta - \alpha$ (last row) for the 19th subject	32



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



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.



Figure 4: Differential neural patterns of drowsy (left column), tired (middle column), and awake (right column) states in five frequency bands; δ_1 (first row), δ_2 (second row), δ_3 (third row), $\delta - \theta$ (fourth row), $\theta - \alpha$ (last row) for the 18th subject.



Figure 5: Differential neural patterns of drowsy (left column), tired (middle column), and awake (right column) states in five frequency bands; δ_1 (first row), δ_2 (second row), δ_3 (third row), $\delta - \theta$ (fourth row), $\theta - \alpha$ (last row) for the 19th subject.

239 List of Tables

240	1	RMSE value of RKPPR and PPR on single channels for the	
241		SEED VIS database	34
242	2	RMSE value of RKPPR and PPR for cumulative structures of	
243		channels on the SEED VIS database	35
244	3	The average RMSE value for different regression algorithms on	
245		LE features	36
246	4	Overall performance of our method compared with the DE+SVR	
247		algorithm	37
248	5	RMSE values for motor-UPDRS prediction using different regres-	
249		sion algorithms.	38
250	6	RMSE values for total-UPDRS prediction using different regres-	
251		sion algorithms.	39
252	7	RMSE values for motor-UPDRS prediction using the gender based	
253		regression strategy.	40
254	8	RMSE values for total-UPDRS prediction using the gender based	
255		regression strategy.	41
256	9	RMSE values for motor-UPDRS prediction, a comparison with	
257		stochastic algorithms.	42

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

	FT7	FT8	Τ7	T8	TP7	TP8	CP1	CP2	P1	PZ	P2	PO3	POZ	PO4	01	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

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

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

	GLM(normal) GLM(gamma)								KSVR SVI						SVR	RKPPR				
	g_1	g_2	g_3	g_4	g_5	g_1	g_2	g_3	g_4	g_5	$\sigma = 1$	$l\sigma =$	= 2.5	$\sigma = 5 a$	$\sigma = 7.5$	$\delta\sigma = 10$	$\sigma = 12.5$	$5\sigma = 15$		
$\overline{\mathrm{FT7}}$	0.15150	0.15250	.15190	.15190	.1489	0.15720	.15270	.16650	.16670	.1663	0.1428	80.1	1429 (0.1483	0.1564	0.1644	0.1724	0.1793	0.1541	0.1386
FT8	0.1403	0.14210	.14190	.14180	.1399	0.14200	.14420	.16430	.16440	.1679	0.1348	80.1	1303 ().1349	0.1418	0.1496	0.1572	0.1634	0.1411	0.1272
T7	0.13620	0.13920	.13500	.13520	.1333	0.14720	.14050	.15450	.15050	.1491	0.1400	00.1	1367 (0.1362	0.1444	0.1516	0.1578	0.1660	0.1401	0.1320
T8	0.12930	0.13620	.13150	.13140	.1310	0.13410	.13380	.14150	.13070	.1411	0.134	10.1	1306 (0.1327	0.1402	0.1461	0.1523	0.1583	0.1323	0.1299
TP7	0.11800	0.12450	.11460	.11500	.1149	0.12120	.12310	.12840	.13850	.1284	0.1278	80.1	1212 (0.1212	0.1302	0.1370	0.1431	0.1497	0.1206	0.1192
TP8	0.12320	0.13180	.12230	.12230	.1225	0.12660	.12970	.11970	.11940	.1203	0.1353	30.1	1249 ().1233	0.1300	0.1357	0.1450	0.1515	0.1245	0.1278
CP1	0.14390	0.14300	.14190	.14220	.1401	0.15630	.14560	.16120	.16110	.1602	0.143!	50.1	1407 ().1470	0.1566	0.1614	0.1648	0.1678	0.1524	0.1363
CP2	0.13300	0.13430	.13090	.13110	.1290	0.14150	.13460	.13330	.13970	.1333	0.1384	40.1	1310 (0.1355	0.1427	0.1477	0.1514	0.1550	0.1381	0.1283
P1	0.12270	0.12460	.12240	.12260	.1230	0.13420	.12550	.13690	.13670	.1282	0.1320	60.1	1274 ().1313	0.1386	0.1446	0.1496	0.1552	0.1303	0.1237
\mathbf{PZ}	0.12350	0.12670	.12260	.12260	.1230	0.13940	.12620	.12370	.12460	.1213	0.1281	10.1	1203 (0.1235	0.1318	0.1381	0.1435	0.1488	0.1252	0.1238
P2	0.1161	0.11870	.11460	.11450	.1136	0.12420	.12020	.14260	.14230	.1402	0.1270	0 0.1	1147 ().1191	0.1271	0.1343	0.1399	0.1453	0.1194	0.1177
PO3	0.10720	0.11010	.10450	.10470	.1038	0.12550	.11230	.10910	.10920	.1086	0.122!	50.1	1080 ().1101	0.1196	0.1273	0.1353	0.1428	0.1087	0.1098
POZ	20.13030	0.13060	.12760	.12850	.1270	0.14900	.13510	.16360	.17070	.1492	0.1308	80.1	1220 (0.1266	0.1354	0.1429	0.1490	0.1552	0.1366	0.1238
PO4	0.11320	0.11460	.11050	.11050	.1093	0.12300	.11550	.11910	.11890	.1128	0.1240	0 0.1	1113 ().1110	0.1171	0.1247	0.1320	0.1394	0.1148	0.1083
O1	0.11000	0.11430	.10810	.10810	.1074	0.12980	.11490	.11140	.11840	.1107	0.126!	50.1	1126 ().1135	0.1237	0.1317	0.1397	0.1479	0.1109	0.1151
ΟZ	0.12340	0.12710	.12200	.12200	.1205	0.13740	.12710	.16040	.16510	.1355	0.1284	40.1	1238 (0.1276	0.1361	0.1444	0.1528	0.1606	0.1279	0.1250
O2	0.11390	0.11640	.11190	.11190	.1110	0.12790	.11800	.12190	.12170	.1157	0.1260	0 0.1	1138 ().1132	0.1205	0.1290	0.1372	0.1445	0.1162	0.1130

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

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

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

	(GLM(r	ormal)	(GLM(g	amma)	KSVR								
	g_1	g_2	g_3	$\mid g_5$	g_1	$ g_2 \rangle$	g_3	$\left \begin{array}{c} g_{5} \end{array} \right $	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$		
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								
	g_1	g_2	g_3	g_5	g_1	g_2	g_3	$\int g_5$	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$			
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.

	$\operatorname{GLM}(\operatorname{normal})$)	(GLM(g	gamma)	KSVR							RKPPR
	g_1	g_2	g_3	g_5	g_1	g_2	g_3	g_5	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
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										
	g_1	g_2	g_3	g_5	g_1	g_2	g_3	g_5	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 10$	
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

Method	OPDT database	SEED VIS database
RKPPR	6.9290	0.1237
GP(Squared exponential)	7.1190	0.1717
GP(Exponential)	7.1420	0.1301
GP(Matern 3/2)	7.2803	0.1463
GP(Matern 5/2)	7.1866	0.1557
GP(Rationalquadratic)	7.1059	0.1369
BLR	7.6842	0.1266

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