Electroencephalogram based Reaction Time prediction with Differential Phase Synchrony representations using Co-operative Multi-Task Deep Neural Networks

Tharun Kumar Reddy, Vipul Arora, Satyam Kumar and Laxmidhar Behera, *Senior Member, IEEE*, Yu-kai Wang, *Member, IEEE* and Chin-Teng Lin, *Fellow, IEEE*

***Abstract*—Driver drowsiness is receiving a lot of deliberation as it is a major cause of traffic accidents. This paper proposes a method which utilizes the Fuzzy Common Spatial Patterns (CSP) optimized Differential Phase Synchrony representations (DPS) to inspect Electroencephalogram (EEG) synchronization changes from the alert state to the drowsy state. EEG based reaction time prediction and drowsiness detection are formulated as primary and ancillary problems in the context of multitask learning. Statistical analysis results suggest that our method can be used to distinguish between alert and drowsy state of mind. The proposed Multi-Task DeepNet (MTDNN) performs superior to the baseline regression schemes like Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (LASSO), Ridge Regression (RR), kNN and ANFIS (Adaptive Neuro Fuzzy Inference Scheme) in terms of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation Coefficient (CC) metrics. In particular, the best performing multitask network, $MTDNN_5$ recorded a 15.49% smaller RMSE, a 27.15% smaller MAPE and a 10.13% larger CC value than SVR.**

***Index Terms*—Brain Computer Interface (BCI), Deep Neural Network (DNN), One Versus Rest (OVR), Reaction Time (RT), Multi-Task Learning, Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation Coefficient (CC)**

I. Introduction

Drowsiness (fatigue), also mentioned as sleepiness, signifies "the tendency to fall asleep". A major shift leading to the deployment of drowsiness detection systems in vehicles is largely attributed to the WHO report in 2013, which stated that almost almost 6% of world's road accidents are caused by drivers in drowsy state [1]. Some recent developments in drowsiness detection research are discussed below.

Among various physiological signals, EEG is one of the most reliable indicators because it is in close conjunction with mental and physical activities [2]. Several methods proposed in literature for detecting fatigue by EEG can be categorized into amplitude and phase based approaches. At first, we summarize the amplitude based approaches. Budi et al. [3] used EEG spectral feature segments to analyze four algorithms for detecting fatigue. They demonstrated that the ratio of the total spectra power in theta and alpha bands to the power in beta band witnessed a greater upsurge over time with drowsiness phenomenon. Wei et al. [4] defined a term, the Level of Session Generalizability (LSG) through a novel Transfer Learning (TL) based method. Their method utilizes a subject's pilot data to select ancillary data from other subjects to enhance the performance of an EEG based BCI for drowsiness detection. Authors in [5] propose regression with Random Forest on multiband power features providing a highly accurate fatigue index using only three electrode positions. [6] contains two novel driving fatigue prediction metrics. First integrated fatigue metric is based on power spectrum density analysis with subject specific channel selection and second metric is based on sample entropy analysis from 'O1h' and 'O2h' electrodes. A. Saha et al. [7] used motor planning phase to detect cognitive failure in driving using type-2 fuzzy classifiers.

The phase based approaches for detecting drowsiness are discussed next. Phase based analysis has also been demonstrated to be the cynosure of functional neural connectivity inference [8]. Phase Synchronization (PS) or phase based analysis can detect the spatial lateralization of drowsiness phenomenon. This approach studies the interplay between signal pairs through the functional relationships of instantaneous phase among the signals independent of their amplitudes. Thus, the notion of PS [9] is fundamental for neuronal information processing within the brain region as well as communication between different brain regions. Lachaux et al. [10] proposed Phase Locking Value (PLV) statistic to quantify the frequency specific synchronization between the neural signals. However, since PLV is a temporal measure of synchrony across the trials, it is not suitable for trial wise analysis. To measure trial wise phase synchrony information, a new statistical single trial PLV was formulated [11]. Caramia et al. [12] proposed a modified single trial PLV but it did not account for lead lag behavior between EEG signals. Kumar et al. [13] proposed to extract features based on Instantaneous Phase Difference (IPD) for trial wise analysis. Further, [14] explored Mean Phase Coherence (MPC) with large and local scale synchrony for fatigue detection. In this work, authors

Tharun Kumar Reddy, Vipul Arora, Satyam Kumar and Laxmidhar Behera are with the Department of Electrical Engineering, Indian Institute of Technology, Kanpur, Kanpur 208016, India (e-mails: tharun@iitk.ac.in, vipular.iitk@gmail.com, satyamk@iitk.ac.in lbehera@iitk.ac.in)

Y.-K. Wang and C.-T. Lin are with the Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, Univer- sity of Technology Sydney, Ultimo, NSW 2007, Australia (e-mails: yukai.wang@uts.edu.au, chintenglin@gmail.com)

adapt the DPS representations for regression which are an outcome of integration of IPD with fuzzy CSPR-OVR [15] framework. Authors further validate the proposed DPS features on the reaction time dataset from an EEG based lane-keeping task.

Thus, a large amount of literature already deals with signal processing for BCIs based on EEG classification problems, but research on EEG regression problems is mostly neglected till now. Some of the prominent EEG based regression problems are: estimation of continuous workload levels [16], Reaction Time (RT) for the EEG-tracked SVIPT [17] and RT prediction for the EEG based PVT and lane-keeping tasks [18].

After the EEG signal is obtained, the regression process involves several steps: 1) Signal pre-processing to enhance the Signal to noise ratio (SNR). Filters in frequency realm such as low-pass filters, band-pass filters, band-stop filters, and spatial filters such as independent component analysis (ICA), SPoC [19], CSP and fuzzy CSSP [20] are frequently used here. 2) Feature representations to shape relevant predictors, e.g., Riemannian Geometry features [18] and EEG power band features [15]. 3) Regressors to estimate the continuous valued variable, e.g., Ordinary Linear Regression (OLR) [21], Ridge Regression [22], LASSO [15, 18], K-nearest neighbors (kNN) [15], fuzzy neural networks [23], Transfer Learning [4, 24, 25], active learning [26], domain adaptation [27], multiview learning [28], multitask learning [29] etc.

In the multitask learning setup, there are multiple tasks, each of which is a general task such as supervised task, unsupervised task, semi-supervised task etc. A handful of these tasks, or a portion of them are related to each other. Cooperative training among these tasks can lead to a greater performance enhancement compared to training the tasks one at a time [30]. In BCIs, *a task* is usually considered a unique recording session, either for an individual or multiple subjects [31]. In [32], authors demonstrated the effectiveness of multitask learning for classification of motor imagery trials. The multitask learning approach proved to be robust against misclassification in different experimental conditions. Furthermore, multitask learning technique efficiently estimates spatial filters for classification of motor imagery in subjects with no prior training data.

RT denotes the time period between the onset of the lane deviation and the onset of the response and is used as an objective measure of the drowsiness (DS) level during each lane departure event [33]. Since EEG drowsiness estimation and EEG reaction time prediction are two related problems, where approximate solution to either of them helps to solve the other, Hence, multitask learning can be applied in this scenario.

The primary objectives of this study are :

- To demonstrate the utility of phase based feature representations for EEG based drowsiness detection.
- To model the drivers' drowsiness detection as a multitask learning problem based on DPS-FCSPR-OVR feature representations and train intelligent models for the proposed tasks.

The major contributions of this work are:

- Novel DPS-FCSPR-OVR representations are proposed to demonstrate the utility of phase based EEG representations for EEG based RT prediction.
- A novel MTDNN framework with a supervised pre-training and fine tuning steps is proposed.
- Extensive experiments (including comparison with advanced regression models) are carried out to confirm the effectiveness of the proposed method in EEG based RT prediction.

The novelty of the proposed method is highlighted as follows:

- The utility of phase based feature representations is very scarcely studied in literature for EEG based drowsiness detection problem. For addressing this issue, the DPS-FCSPR-OVR representations are adopted to train intelligent models.
- In the multitask BCI literature, the concept of task is usually limited to either a subject, session etc. For the first time, we extend the notion of BCI task to address two BCI problems: drowsiness detection and RT prediction.

This paper is organized as follows: Section II presents the formulation of proposed DPS-FCSPR-OVR representations. Section III evaluates their performance on the EEG lane keeping task. Section IV leads us to the proposed co-operative DNN based multitask approach (MTDNN). Section V evaluates the proposed MTDNN approach with several baseline DeepNets and other regression schemes. Finally, discussion and concluding remarks are provided in sections VI and VII respectively.

II. Differential Phase Synchrony(DPS) Representations

A. Fuzzy CSPR-OVR

Let $\mathbf{X}_n \in \mathbb{R}^{C\times T}$ $n \in \{1, 2, .., N\}$ denote the n^{th} EEG trial, where C denotes number of channels and T denotes number of time samples per trial. Trial $\mathbf{X}_n$ constitutes a band pass filtered signal whose mean is removed from each of the channels. Using the concept of fuzzy sets, we define fuzzy classes (assume M fuzzy classes) to generalize to regression problems. Next, the interval range $[0, 100]$ is used to generate $M + 1$ regions and let us denote the boundary partition points by $\{y_{p_m}\}$, $m = \{1, .., M\}$. To mention

$$P_m = \frac{100 \cdot m}{M+1}, \quad m = 1, ..., M \tag{1}$$

Each y_{p_m} is the P_m percentile value of the training set of RTs. Next, we define M fuzzy classes and categorize training RT values into one of the M fuzzy classes, in a manner analogous to M classes in a multiclass classification scenario. For each fuzzy class, a y_n can belong to it at a membership degree $\in [0, 1]$. Further, we compute an average covariance matrix for each fuzzy class as:

$$\tilde{\mathbf{\Sigma}}_m = \frac{\sum_{n=1}^{N} \mu_m(y_n)\mathbf{X}_n\mathbf{X}_n{}^{\intercal}}{\sum_{n=1}^{N} \mu_m(y_n)}, \quad m = 1, ..., M \tag{2}$$

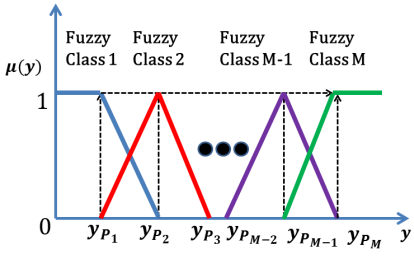

Fig. 1: 'M' fuzzy classes for training RT values using triangular fuzzy membership

where $\mu_m(y_n)$ is the membership degree of y_n in fuzzy class m.

-We now mention the One-Versus-Rest (OVR) CSP to extend the CSP from binary classification to M classes. For class m, OVR-CSP finds a matrix $\mathbf{W}_m^* \in \mathbb{R}^{C\times F}$, where F is number of spatial filters to maximize variance of class m against rest.

$$\mathbf{W}_m^* = \arg\max_{\mathbf{W}} \frac{\mathrm{Tr}(\mathbf{W}^\intercal \tilde{\mathbf{\Sigma}}_m \mathbf{W})}{\mathrm{Tr}(\mathbf{W}^\intercal \sum_{j\neq m} \tilde{\mathbf{\Sigma}}_j \mathbf{W})} \tag{3}$$

$\tilde{\mathbf{\Sigma}}_m$ is the mean covariance matrix of trials in class m. $\mathbf{W}_m^*$ is the concatenation of the F eigenvectors associated with the F largest eigenvalues of the matrix $(\sum_{i\neq m} \tilde{\mathbf{\Sigma}}_i)^{-1}\tilde{\mathbf{\Sigma}}_m$. We concatenate the obtained F filters for each of the M classes to obtain $\mathbf{W}^* = [\mathbf{W_1^*}, ...\mathbf{W_M^*}] \in \mathbb{R}^{C\times MF}$. Then, one can compute a spatially transformed trial by $\mathbf{X}_n^{'} = \mathbf{W}^{*\intercal}\mathbf{X}_n, \quad n = 1, .., N$.

B. Phase Locking Value

PLV is a statistic used to investigate synchronization of neural activity from EEG data and expresses a transient measure of connectivity. Any cognitive task results due to combination of various functional areas distributed over different regions of the brain. The task induced coupling between these areas is interpreted as synchronization of neural activity. PLV and its variants are common measures of phase synchronization.

Consider two signals $x_1(t)$ and $x_2(t)$ whose instantaneous phases are $\psi_1(t)$ and $\psi_2(t)$. In accordance with [12] the Single trial PLV (sPLV) for a given trial can be defined as follows:

$$sPLV = \left| \frac{1}{N_s} \sum_{t=1}^{N_s} e^{|\psi_1(t)-\psi_2(t)|} \right| \tag{4}$$

where N_s is the number of samples in the trial. The instantaneous phase $\psi(t)$ can be obtained using the analytic signal calculated from Hilbert transform. For any arbitrary signal $x(t)$ the analytic signal $z(t)$ is given as

$$z(t) = x(t) + i\tilde{x}(t) \tag{5}$$

$$\tilde{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{6}$$

where $\tilde{x}(t)$ is the Hilbert transform of $x(t)$. The instantaneous phase $\psi(t)$ is then calculated using

$$\psi(t) = \arctan\left(\frac{\tilde{x}(t)}{x(t)}\right) \tag{7}$$

The sPLV holds a value between 0 and 1 with extremas corresponding to the cases of signal being unsynchronized and completely synchronized respectively.

C. Differential-Phase Synchrony representations

Instantaneous Phase Difference (IPD) sequence $\Delta\psi(t)$ between a pair of distinct signals $s_1(t)$ and $s_2(t)$ is defined as

$$\Delta\psi(t) = \left|\psi_1(t) - \psi_2(t)\right| \tag{8}$$

Authors in [13] coupled the notion of variance of instantaneous phasors with sPLV. Furthermore, they formulated a framework to estimate a linear transform that maximizes the variance of instantaneous phasors across one class while simultaneously minimizing it across the other class. The proposed framework is similar to CSP but, in contrary, it explicitly uses phase information for binary classification. Thus, drawing an analogy from the regression CSP algorithm (Fuzzy CSPR-OVR), a novel framework is formulated to find a linear transform on the IPD sequence in such a way that it maximizes the variance across the temporal phase differences of each fuzzy class and simultaneously minimizes the variance of phases across the other fuzzy class. As the IPD sequence is used to measure synchrony between EEG signals, we refer to the extracted features as DPS representations.

1) Fuzzy Spatial Filter optimization: Fuzzy CSPR-OVR [15] extends multiclass CSP to regression problems using fuzzy sets. Indeed, as mentioned before, the EEG signals from individual channels are prone to have a low SNR, due to spatial blurring and smearing effects. In order to obtain more discriminative DPS features for EEG classification of fuzzy classes, it thus seems prudent to compute spatial filters which maximize the variance of instantaneous phase across a particular fuzzy class and minimize across rest of them. We therefore propose an algorithm to optimize the spatial filters in order to maximize the resulting DPS feature discriminative power. Mathematically, it can be written as

$$\mathbf{W}_i{}^* = \arg\max_{\mathbf{W}} \frac{\mathrm{Tr}(\mathbf{W}^\intercal \mathbf{\Sigma}_{\Delta\psi_i} \mathbf{W})}{\mathrm{Tr}(\mathbf{W}^\intercal \sum_{j\neq i} \mathbf{\Sigma}_{\Delta\psi_j} \mathbf{W})} \tag{9}$$

where $\mathbf{\Sigma}_{\Delta\psi_i}$ and $\mathbf{\Sigma}_{\Delta\psi_j}$ are the covariance matrices of the IPD sequence for the fuzzy classes i and j. The column vectors of $\mathbf{W}$ are the spatial synchrony filters. A scheme similar to fuzzyCSP algorithm is employed for feature extraction from IPD sequence of an EEG trial in a given frequency range. The obtained features are coined as 'DPS' repersentations.

III. Evaluation of Differential-Phase Synchrony representations For Reaction Time Prediction

The proposed feature representations are evaluated and compared to other phase based baseline methods on an EEG based reaction time (RT) prediction in a lane-keeping task.

A. Lane-Keeping Task

The EEG signals were recorded from 30 active electrode sites which were placed according to modified international 10-20 electrode montage system.

The Institutional Review Board of the Veterans General Hospital, Taipei, Taiwan, approved the study. A total of 12 university students (average age 22.4, standard deviation 1.6) from the National Chiao Tung University (NCTU) in Taiwan volunteered to support the data-collection efforts over a five-month period to study EEG correlates of attention and performance changes under specific conditions of real-world drowsiness [34].

Simulated driving experiments were conducted on a virtual reality (VR)-based dynamic driving simulator. A real car frame was mounted on a six degree-of-freedom Stewart motion platform which moved in sync with the driving scene during "motion" sessions. The motion platform was inactive during "motionless" sessions. The VR driving scene simulated nighttime cruising (100 km/h) on a straight highway (two lanes in each direction) without other traffic. The computer program generated a random perturbation (deviation onset), and the car started to drift to the left of the right of the cruising lane with equal probability. Following each deviation, subjects were required to steer the car back to the cruising lane as quickly as possible using the steering wheel (response onset), and hold on the wheel after the car returned to the approximate center of the cruising lane (response offset). A lane departure trial is defined as consisting of three events, deviation onset, response onset, and response offset. The next lane-departure trial randomly occurs about 5 to 10 sec after response offset in the current trial. The subject's reaction time (RT) to each lane departure trial is defined as the interval between deviation onset and response onset. If the subject does not respond promptly within 2.5 (1.5) sec, the vehicle will hit the left (right) roadside without a crash and continue to move forward against the curb event the subject completely ceases to respond. No intervention was made when the subject fell asleep and stopped responding. After reaching the lapse period, subjects resumed the task voluntarily and steered the car back to the cruising position at the earliest.

The goal is to predict RT using a 5-s EEG trial immediately before it.

B. EEG Pre-processing

- At first, raw EEG data was passed through standard pre-processing pipeline (PREP) of EEGLAB to increase the signal to noise ratio, it comprises mainly of three operations [35][36].
  - Removing line noise.
  - Determining and removing robust reference signal.
  - Interpolating the bad channels.
- Further, the data was downsampled to 250 Hz.
- Then, the data was epoched to 5 sec trials, i.e. if the lane deviation is starting at time 't' then the EEG data from $[t-5, t]$ is used to predict the RT. Each EEG trial is of size 30×1250.
- Outliers in the RT values are removed by ignoring the EEG trials with RT values greater than sum of mean and three times the standard deviation.
- Thus, the obtained trials are filtered by a $[1, 20]$ Hz finite impulse response band-pass filter.
- The obtained data is then fed through the appropriate spatial filters.

TABLE I: Number of Trials (Ntrials) and Mean RT in the dataset

Subject	1	2	3	4	5	6
Ntrials	628	568	653	350	362	361
MeanRT	0.897	1.668	0.925	0.881	1.070	1.145
Subject	7	8	9	10	11	12
Ntrials	239	484	550	641	190	143
MeanRT	2.020	1.541	1.001	0.704	1.076	2.677

C. RT Pre-processing

The RT values for 12 subjects are pre-processed in a way similar to that of the paper (cf. section IV D of [15]). The data collected from subject 12 is erroneous with data recording anomalies and is removed from further analysis. This is because, a large number of response times were longer than 5 seconds, which are highly absurd in practice. The final distribution of RTs obtained after pre-processing are shown in Fig.2.

D. Feature Evaluation

8-fold cross-validation is used to compute the regression performance for each possible fusion of feature set and regression method. Following feature sets are extracted for each EEG trial.

1) Theta and Alpha powerband features are extracted from the band-pass filtered EEG trials. We computed the average power spectral density in the Theta band (4-8 Hz) and Alpha band (8-13 Hz) for each channel using method, and converted these $30 \times 2 = 60$ band powers to dBs as our features (denoted as '*FS1*').
2) Differential Phase Synchrony features (DPS) are extracted from the band-pass filtered EEG trials. We used 3 fuzzy sets ($K = 3$) for the RTs, and 21 spatial filters ($F = 21$) for each fuzzy class. A vector of size $(63 \times 1) = 63$ constitutes the feature vector (DPS). It is denoted as '*FS2*'.
3) Theta and Alpha powerband features extracted from EEG trials filtered by fuzzy CSPR-OVR. This procedure was almost identical to that of '*FS1*', except that the band-pass filtered EEG trials were also spatially filtered by fuzzy CSPR-OVR before the powerband features

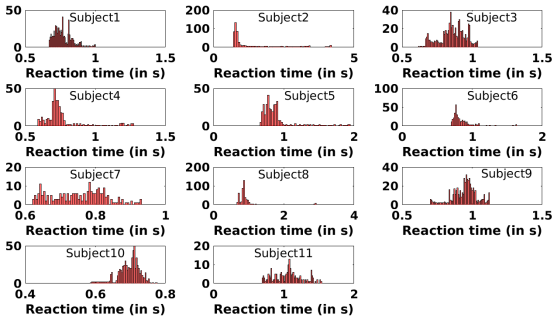

Fig. 2: Distribution of RT values

TABLE II: Regression performance of *FS1*, *FS2* and *FS3*

Feature-set	CC	RMSE	MAPE
FS1	0.501	0.0861	20.56
FS2 (proposed)	0.684	0.0726	18.02
FS3	0.519	0.0849	19.30
FS4 (proposed)	**0.719**	**0.0695**	**16.43**

were computed. We used 3 fuzzy sets for the RTs, and 10 [1] spatial filters for each fuzzy class, so that the spatially filtered EEG trials is of the dimension 30×1250, and the resultant feature '*FS3*' has 60×1 dimensions.

4) Differential Phase Synchrony features (DPS) extracted from the EEG trials pre-filtered by fuzzy CSPR-OVR. We used 3 fuzzy sets ($K = 3$) for the RTs, and 10 spatial filters ($F = 10$) for each fuzzy class again for a fair comparison. A vector of size $(63 \times 1) = 63$ constitutes the feature vector (DPS-CSPR-OVR). It is denoted as '*FS4*'.

All the feature sets obtained above are passed through LASSO regressor to obtain final reaction time value.

E. Performance Metrics

RMSE, CC and MAPE are the metrics in use for judging the regression performance. Assume, there are N training points, y_{di} represents the true reaction time value of the i^{th} data point and y_i represents the predicted reaction time value.

F. Regression Results

The average RMSEs, CCs and MAPEs of LASSO using the four feature sets (explained in section III-D) are shown in the Table II. For each subject, 8-fold cross validation has been used to partition the feature set into training and validation sets. The performance is averaged across all the 8-folds. Also, the average performance across all the subjects is reported. Here, in general 'FS4' recorded the best performance, and both FS4 and FS2 achieved much smaller RMSEs, MAPEs and much larger CCs than FS3 and FS1, suggesting that our extension of FS4 from supervised classification to supervised regression can indeed improve the regression performance.

[1]We used 10 spatial filters here so that the filtered signals had almost the same dimensionality as the original signals, which ensured fair performance comparison

In conclusion, FS4 had better regression performance than FS2, FS3 and FS1.

More detailed performance analysis of results and implementation details comparing performance of the features in section III-D are presented in a supplementary file (DPS-fuzzyCSPR-OVR.pdf).

IV. Multi-task Deep Neural Networks

Wei et al., [37] treated drowsiness detection as a classification problem by formulating set of thresholds on reaction time values. The ancillary task in the proposed multitask method addresses drowsiness detection as a classification problem. Two tasks in the name of primary and ancillary are in use for the experiment. EEG based RT prediction is the primary task and drowsy state classification problem is considered as an ancillary problem. As far as the ancillary task is concerned, EEG trials with RT shorter than $1.5 \times (alertRT)$ are categorized as 'Alert' trials, whereas those with RT longer than $2.5 \times (alertRT)$ are taken to be as 'Lapse' trials indicating 'drowsy' phase. In addition, those EEG trials with RT shorter than $2.5 \times (alertRT)$ but longer than $1.5 \times (alertRT)$ are categorized as 'Semi-alert' trials. The $alertRT$ was individually estimated for each subject as suggested in [4]. Primary task is accomplished using DeepNet-2 and 3, while the ancillary task is accomplished using DeepNet-1 and 3.

A. Pre-training

Pre-training is used to avoid the learning algorithm to get stuck in a local optimum. This is especially true while training a deep model in the situation of a scarce training data. In the present work, we propose a supervised pre-training approach. Fig.3 shows how the proposed Deep network-based method in DeepNet-1 and DeepNet-2 incorporates the label information. The supervised pre-training consists of two steps. Firstly, the DeepNet-1 was trained to predict the three levels of drowsy states, namely drowsy, transition and awake. Keeping the hidden layers of DeepNet-1 intact (with the pre-trained weights), an output layer consisting of a single output node is added to construct DeepNet-2, which is then further pre-trained to predict the reaction time values. Note that the regression layer of DeepNet-2 was initialized with random weights because this top layer is different from DeepNet-1. It is easier for Deep network to learn the three class problem than an infinite class classification problem (i.e. regression). Nonetheless, the weights of DeepNet-2 are tuned-up based on DeepNet-1. It follows the acumen of a meaningful human learning process: simple to complex tasks. The understanding of learning simpler tasks be able to benefit the learning for complex tasks.

B. Multi-task Objective Function

DeepNet-1 uses the cross entropy loss-function. DeepNet-2 uses the mean-squared error loss-function. In DeepNet-1, softmax layer is the output layer, while in DeepNet-2, a sigmoid unit is present in the output. Further, the output layer of DeepNet-3 is a composed of both softmax and sigmoid

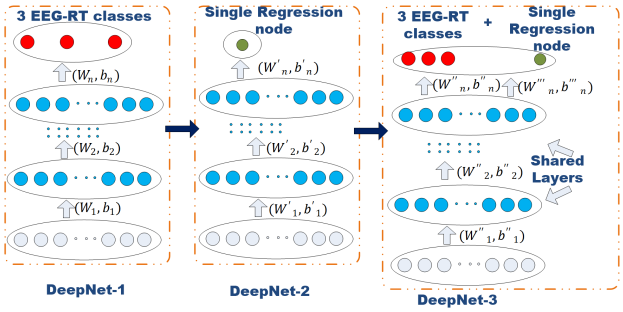

Fig. 3: Proposed Multitask-DeepNet (MTDNN) approach

units, where the network is jointly trained for classification and regression. Accordingly, DeepNet-3 uses a combination of two losses:

$$-\lambda \sum_{i=1}^{M}\sum_{j=1}^{N}(d_{i,j}\log(p_{i,j})) + \mu \sum_{i=1}^{M}(y_{d_i} - y_i)^2 \qquad (10)$$

In (10), $\lambda+\mu = 1$, M is the number of mini-batch samples, N is the number of classes, $d_{i,j}$ represents the desired probability and $p_{i,j}$ denotes the predicted probability of the j^{th} sample for the i^{th} class. We choose $\lambda = \mu = 0.5$ in our experiments.

V. Experiments and Results

A. Experiments on DPS-FCSPR-OVR features from EEG RT dataset

The experiments are conducted on reaction time dataset for EEG Lane-keeping task. The dataset preprocessing is provided in section III-B. $K = 3$ and $F = 21$ are obtained as an optimal choice for the number of fuzzy classes and the number of filters per each fuzzy class. DPS-FCSPR-OVR features are extracted from pre-processed EEG RT dataset. The dimension of the spatial filtered feature vector is 63×1. 63 dimensional feature vector is used to train DeepNet-1 with classification labels obtained by the procedure mentioned at the start of section IV. Number of iterations used in the training of DeepNets-{1,2,3} are {200, 200, 150} respectively. Supervised pre-training steps used larger number of iterations than the fine-tuning step. Adagrad optimizer is used with the initial learning rate set to $\{0.1, 0.5\}$ for supervised pre-training DeepNet-{1,2} and joint fine-tuning DeepNet-3, respectively. We tried to avoid overfitting during training by employing drop-out (drop-out probability $p = 0.5$) and L_1 regularization ($\lambda = 0.01$). Size of Mini-batch is set to 8. DPS-FCSPR-OVR features (input features) are normalized to zero mean and unit variance. CC, RMSE and MAPE are the metrics used for regression performance comparison, while precision, recall and F1 score are the metrics used for classification performance analysis.

B. Comparison among MTDNNs with different configurations

A comparison using RMSE, CC and MAPE is being made for the multi-task deepnets with different configurations of hidden layers and number of nodes in each hidden layer. Table III contains the mean values of each of the metrics with

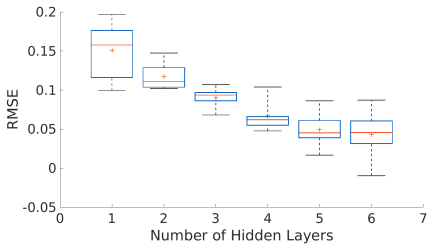

Fig. 4: Effect of changing model architecture on RMSE

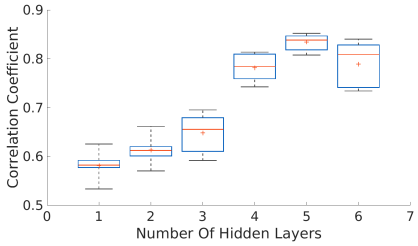

Fig. 5: Effect of changing model architecture on CC

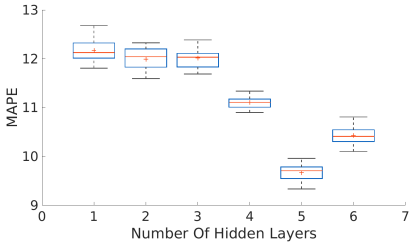

Fig. 6: Effect of changing model architecture on MAPE

varying number of hidden nodes and hidden layers. In figs. 4–6, an effect of the depth of the neural nets on the performance is studied. Box plots show the distribution of RMSE, CC and MAPE associated with 20 different initializations of the weights of Deepnets-1,2&3 (highest and lowest quartiles in box, plus outliers beyond highest and lowest quantiles). Mean RMSE value decreased or remained constant with increasing number of hidden layers except for a increment on transition from 5 to 6 hidden layers. A similar trend has been observed for MAPE values except that there is an increment in the MAPE value on transition from 2 to 3 hidden layers. CC strictly increased with the number of hidden layers, but it started to decrease at the transition from 5 to 6 hidden layers. More number of random initialization trials are needed to make a more accurate inference. In Table III, multitask deep neural networks with $\{N_n = [150 \quad 120 \quad 80], N_l = 3\}$ and $\{N_n = [200 \quad 150 \quad 120 \quad 45], N_l = 4\}$ are denoted as $MTDNN_4$ and $MTDNN_5$ respectively.

C. Comparison among MTDNNs and other regression models

To further establish the powerful representational capability of the DNN-based multi-task model, a comparison with various traditional regression models like LASSO Regularized Regression, Ridge Regression (RR), Support Vector regression

TABLE III: Regression metrics for different architectures of MTDNN. N_l: number of hidden layers; N_n: number of hidden nodes.

N_l	1	2	2
N_n	[45]	[80 50]	[100 60]
RMSE	0.15	0.11	0.09
CC	0.58	0.61	0.66
MAPE	12.2	12.1	11.9
N_l	3	4	5
N_n	[150 120 80]	[200 150 120 45]	[240 170 130 75 30]
RMSE	0.06	**0.05**	0.06
CC	0.78	**0.83**	0.81
MAPE	11.2	**9.7**	10.5

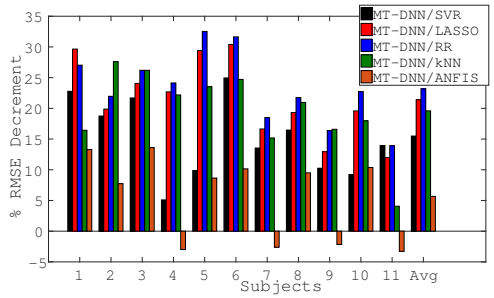

Fig. 7: Percent improvement (decrement) in RMSE with proposed MTDNN over other methods

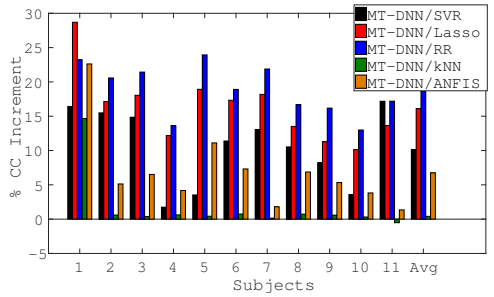

Fig. 8: Percent improvement (increment) in CC with proposed MTDNN over other methods

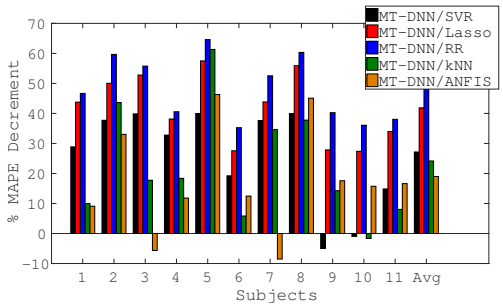

Fig. 9: Percent improvement (decrement) in MAPE with proposed MTDNN over other methods

(SVR) [38], ANFIS (Adaptive Neuro Fuzzy Inference) [2]and

[2]http://yarpiz.com/301/ypfz101-nonlinear-regression-using-anfis

TABLE IV: Comparison of regression metric for MTDNNs over standard regression models

Model	$MTDNN_5$	SVR	$MTDNN_4$	
RMSE	0.0500	0.0600	0.0600	
CC	0.830	0.778	0.810	
MAPE	9.70	11.34	11.20	
Model	RR	k-NN	ANFIS	$LASSO$
RMSE	0.0700	0.0676	0.0570	0.0695
CC	0.690	0.758	0.782	0.719
MAPE	17.48	12.73	11.92	16.43

TABLE V: Two-way ANOVA results MTDNN versus others

	RMSE	CC	MAPE
$p-value$	0.015	**0.000**	0.002

kNN (k-nearest neighbours) is investigated. LASSO and Ridge regression models are implemented in Matlab R2016. We utilized scikit-learn SVR tool [39] and scikit-learn respectively to run the SVR and kNN simulations. Based on the simulations, we observe that: a) SVR is poor at dealing with large scale training datasets due to the computational complexity and memory storage; b) with a high input feature dimension, the generalization ability of SVR is very limited [40]. High computational resources are associated with solving quadratic programming problem amid the optimization of the cost function in the training stage of the SVM. Additionally, a requirement for a large memory to store the kernel products also limits the pertinence of using the SVMs on very large data sets. Grid search technique is employed to determine parameters C and γ, which are set as 2^5 and 2^2, respectively. ϵ is set to 0.2. For the two regression methods: LASSO and RR, the adjustable parameter λ was optimized by an inner 8-fold cross-validation [41] on the training dataset. We employ $k = 5$ with inverse distance weighting for kNN regression. Feature vector of 63 dimensions is used to train the ANFIS regressor with 5 fuzzy partitions for each dimension. Gaussian and linear memberships are chosen for input and output memberships respectively.

On an average, from Table IV, the proposed $MTDNN_4$ and $MTDNN_5$ approaches clearly outperformed SVR, LASSO, RR, kNN and ANFIS schemes in-terms of RMSE, CC and MAPE. In all these experiments, we used $MTDNN_5$ (from Table III) as the representative MTDNN model. Each of the MTDNN is trained for 100 epochs with a learning rate of 0.1 using Adagrad.

The respective percentage performance improvements of MTDNN over the other regression models are shown in the figs. 7–9. For instance, the terms in legend `'MTDNN/SVR'` represent the improvements with MTDNN over the SVR method. The notation of other terms in the legend is to be understood in a similar manner. On an average, MTDNN recorded a 15.49% smaller RMSE, a 27.15% smaller MAPE and a 10.13% larger CC than SVR. On an average, MTDNN had performed with a 21.41% smaller RMSE, a 41.81% smaller MAPE and a 16.11% larger CC than LASSO. Also, MTDNN had performed with a 23.18% smaller RMSE, a

TABLE VI: Non-parametric multiple comparison tests based on paired t-tests (MTDNNs versus other regression models) $MTDNN$ denoted by M

	RMSE				CC				MAPE			
	M	SVR	LASSO	ANFIS	M	SVR	LASSO	ANFIS	M	SVR	LASSO	ANFIS
M		0.02	**0.001**	0.01		**0.000**	0.002	0.042		0.01	0.01	0.01
RR	0.03	0.04	0.02	0.04	0.03	0.01	**0.000**	0.01	**0.000**	0.01	0.01	0.03
SVR	0.02		**0.001**	0.02	**0.000**		0.02	0.03	0.01		**0.000**	0.023
kNN	0.02	0.024	**0.001**	**0.000**	**0.000**	0.004	0.02	0.045	0.02	0.01	0.034	**0.000**

TABLE VII: Regression metrics for MTDNNs and corresponding VDNs with same hidden layer configuration. N_l : number of hidden layers, N_n : number of hidden nodes

Model	$MTDNN_4$	VDN_4	$MTDNN_5$	VDN_5
N_l	3	3	4	4
N_n	[150 120 80]	[150 120 80]	[200 150 120 45]	[200 150 120 45]
RMSE	0.06	0.08	**0.05**	0.07
CC	**0.78**	0.64	0.83	0.69
MAPE	11.2	14.8	**9.7**	17.5

TABLE VIII: Two-way ANOVA results $MTDNN_5$ versus VDNs

	RMSE	CC	MAPE
$p-value$	0.005	**0.000**	**0.000**

18.67% smaller MAPE and a 48.17% larger CC than RR. In a similar manner, on an average, MTDNN outperformed ANFIS and kNN approaches with a 5.66% and 19.59 % smaller RMSE, 18.98% and 24.16% smaller MAPE, 6.76% and 0.5% larger CC values respectively. Performance across subjects 6,7 and 10 needs to be analyzed further while collecting more EEG sessions from these subjects, so that the $MTDNN$ can be trained sufficiently for better inference.

Further, Statistical analysis is performed to test several hypotheses on MAPE, RMSE and CC's across the subjects. A two-way Analysis of Variance (ANOVA) is performed for different types of regression schemes with the objective to figure out if the RMSE, CC and MAPE variations due to the changes in regressors are statistically significant, with the subjects treated as a random factor. The results are shown in Table V, ($p-value < 0.05$) which indicates that there were statistically significant differences in RMSEs, CCs and MAPEs among different regression methods with subjects as a random factor. In other words, there is a significant main effect of regression method on the performance metrics RMSE, CC, MAPE ($p-value < 0.05$, cf. Table V).

Then, post-hoc non-parametric multiple comparison tests (paired t-tests in this case) are conducted to find out if the difference between pair of regressors is statistically significant, with the $p-value$ corrected employing the False Discovery Rate method [42]. The p-values are shown in Table VI, where in most of the values are statistically significant. The bolded ones in Table VI are extremely statistically significant[3].

[3] $p-value \leq 0.001$

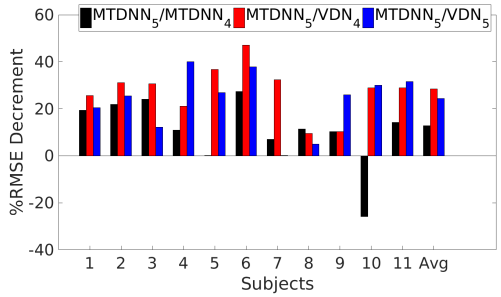

Fig. 10: Percent improvement (decrement) in RMSE with proposed MTDNN over other DNNs

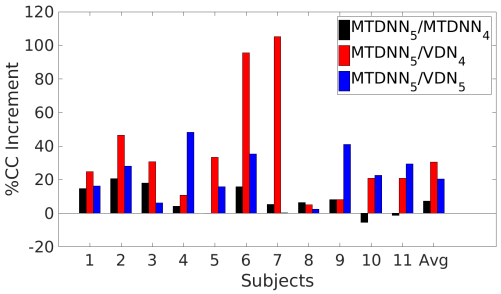

Fig. 11: Percent improvement (increment) in CC with proposed MTDNN over other DNNs

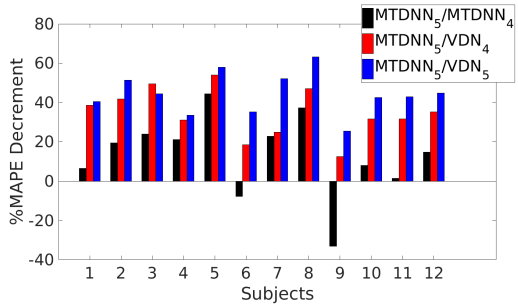

Fig. 12: Percent improvement (decrement) in MAPE with proposed MTDNN over other DNNs

D. Comparison among DNNs and MTDNNs

In this section, we compare the performance of the Multitask DNNs ($MTDNN_4$ and $MTDNN_5$) with vanilla DNNs (4 hidden layer architecture of $MTDNN_5$ (VDN_5) and 3 hidden layer architecture of $MTDNN_4$ (VDN_4) respectively). Table VII presents a systematic performance comparison on the basis of several regression metrics. Each of the MTDNNs and vanilla DNNs (VDN_4 and VDN_5) are trained for 100

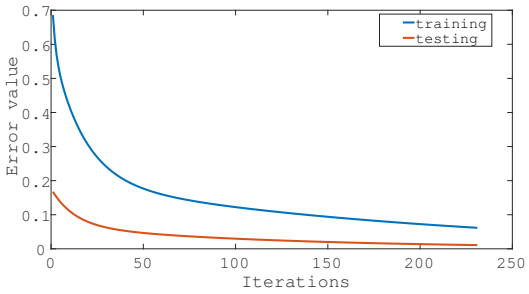

Fig. 13: Average Training and generalization error for proposed MTDNN

TABLE IX: Non-parametric multiple comparison tests based on paired t-tests ($MTDNN$s versus other DNNs), $MTDNN$ and VDN denoted by M and V resp.

	RMSE			CC			MAPE		
	M_5	V_4	V_5	M_5	V_4	V_5	M_5	V_4	V_5
M_5		0.02	**0.001**		0.01	**0.000**		**0.000**	0.01
V_4	0.02	0.04	0.02	0.01	0.01	**0.000**	**0.000**	0.01	0.01
V_5	**0.001**	0.02	**0.001**	**0.000**	**0.000**	0.02	0.01	0.01	**0.000**

epochs with a learning rate of 0.1 using Adagrad. We circumvent overfitting during training by using drop-out (drop-out probability $p = 0.5$) and L_1 regularization ($\lambda = 0.01$). Fig. 13 presents the average generalization ability of the proposed $MTDNN$ approach for a stable [43][44] learning rate η. One can infer that, on an average, the network is not overfitting. The respective percentage performance improvements of MTDNNs over the conventional or vanilla DNN regression models are shown in the figs. 10–12. For example, the terms in legend $MTDNN_5/VDN_5$ represent the improvements with Multi-task DNN ($MTDNN_5$) over the vanilla DNN (VDN_5) method. The notation of other terms in the legend is to be understood in a similar manner. On an average, $MTDNN_5$ recorded a 12.88% smaller RMSE, a 7.28% larger CC and a 14.83% smaller MAPE than $MTDNN_4$. On an average, $MTDNN_5$ had performed with a 28.53% smaller RMSE, a 35.20% smaller MAPE and a 30.42% larger CC than VDN_4. Also, $MTDNN_5$ had performed with a 24.4% smaller RMSE, a 44.70% smaller MAPE and a 20.51% larger CC than VDN_5.

Further, Statistical analysis is performed to test several hypotheses on MAPE, RMSE and CC's across the subjects. Firstly, a two-way Analysis of Variance (ANOVA) is performed for various MTDNNs and VDN's with the intent to assess if the RMSE, CC and MAPE discrepancies due to the changes in the factor ($MTDNN/VDN$) are statistically significant, with the subjects set as a random factor. The results are shown in Table VIII, ($p - value < 0.05$) which indicates that there were statistically significant differences in RMSEs, CCs and MAPEs among MTDNNs and VDNs with subjects regarded as a random factor. In other words, there is a significant main effect of the type of DNNs (Vanilla or Multitask) on the performance metrics RMSE, CC, MAPE ($p - value < 0.05$, cf. Table VIII).

In addition, post-hoc non-parametric multiple comparison tests (paired t-tests in this case) are conducted to find out

TABLE X: Classification (auxiliary task) performance of proposed method, averaged over all subjects

	$MTDNN_4$			$MTDNN_5$		
Class	Alert	Semi-alert	Lapse	Alert	Semi-alert	Lapse
Precision	0.91	0.90	0.94	0.90	0.92	0.92
Recall	0.88	0.89	0.91	0.91	0.93	0.94
F1	0.89	0.89	0.92	0.90	0.92	0.93

if the difference between multitask and vanilla modules is statistically significant, with the $p-value$ corrected employing the False Discovery Rate method [42]. The p-values are shown in Table IX, where in the values are statistically significant. In all the pairwise comparisons $MTDNN_5/VDN_4$, $MTDNN_5/VDN_5$ etc., the results are statistically significant. The bolded ones in Table IX are extremely statistically significant [4].

E. Classification results for the ancillary task for the proposed MTDNNs

Table X contains the classification performance of the proposed MTDNNs for the ancillary task (drowsiness classification). Since this is an ancillary task, we have not provided a detailed comparison with baseline methods.

VI. Discussion

In this work, we propose the Differential Phase Synchrony (DPS) representations for regression. Moreover, we integrate it with fuzzy-CSPOVR framework (DPS-FCSPR-OVR). We further validate the proposed features on the RT dataset from an EEG based lane-keeping task. 8-fold cross validation is used for all the experiments. Statistical analysis and regression results validate the effectiveness of proposed DPS-FCSPR-OVR representations. An interesting direction of future work is to integrate DPS features with Regularised Fuzzy-CSP and SF-BCSP. In future, we seek to integrate IPD sequence with CSP through JAD (Joint Approximate Diagonalization) framework. In addition, MTDNN based regressor is proposed to predict the reaction time with an ancillary DNN trained for drowsiness prediction dealt as a classification problem. Multitask learning enforces regularization hence reduces the risk of overfitting, and hence an improved performance. Furthermore, as both of tasks are related tasks and complimentary to each other, multitask learning bolster the performance in comparison to either of the tasks. A computational complexity analysis of proposed approach is included in a supplementary file for the readers. Due to the non-availability of a public dataset for regression, In our research, we majorly focus on our driving data, instead of the other data.

VII. Conclusion

In this paper, by drawing an analogy with the regression CSP algorithm (Fuzzy-CSPOVR), a novel framework is formulated to find a linear transform on the IPD sequence in such a way that it maximizes the variance across the temporal phase

[4] $p - value \leq 0.001$

differences of each fuzzy class and simultaneously minimizes the variance of phases across the other fuzzy class. The IPD sequence is used to quantify synchrony between EEG signals. We refer to the extracted features as Differential-phase synchrony (DPS) representations. Proposed DPS feature framework performed superior to powerband features interms of better RMSE, Correlation Coefficient(CC) and MAPE values when both DPS and powerband features are passed through LASSO. In addition, a Multitask DeepNet (MTDNN) approach is proposed, where, a three step procedure is used to train on both drowsiness classification and RT prediction tasks. A comparison is made among different configurations of MTDNNs. Also, MTDNNs are compared with single task DNNs, SVR (Support Vector Regression), LASSO, RR (Ridge Regression), kNN and ANFIS. The proposed MTDNN ($MTDNN_5$) exhibited a 15.49% smaller RMSE, a 27.15% smaller MAPE and a 10.13% larger CC than SVR. The proposed approach achieved the best performance especially suggesting its utility in BCI applications of online adaptation with small training data. Also, the present study is a first step towards realizing large scale Multi-task learning BCIs. The two most prominent pre-training approaches for DNN's are the RBM [45] and stacked auto-encoder [46] algorithms. But, both of the above algorithms are unsupervised. Semi-supervised multi-task approaches can be more advantageous and will be explored in a future work.

VIII. Acknowledgement

This work was supported in part by the Australian Research Council (ARC) under discovery grant DP180100670 and DP180100656 and also in part by the Ministry of Human Resource Development (MHRD), Government of India under the project 'MHRD-EE-2016150'. Research was also sponsored in part by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022 and W911NF-10-D-0002/TO 0023. First author is supported by the project TCS/CS/2011191A.

References

[1] T. Toroyan, M. M. Peden, and K. Iaych, "WHO launches second global status report on road safety," 2013.

[2] S. K. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biological psychology*, vol. 55, no. 3, pp. 173–194, 2001.

[3] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2352–2359, 2009.

[4] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, T.-P. Jung, N. Bigdely-Shamlo, and C.-T. Lin, "Selective Transfer Learning for EEG-based drowsiness detection," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3229–3232.

[5] G. N. Dimitrakopoulos, I. Kakkos, N. V. Thakor, A. Bezerianos, and Y. Sun, "A Mental Fatigue index based on regression using Multiband EEG features with application in simulated driving," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017, pp. 3220–3223.

[6] H. Wang, A. Dragomir, N. I. Abbasi, J. Li, N. V. Thakor, and A. Bezerianos, "A novel real-time driving fatigue detection system based on wireless dry EEG," *Cognitive Neurodynamics*, pp. 1–12, 2018.

[7] A. Saha, A. Konar, and A. K. Nagar, "EEG Analysis for Cognitive Failure Detection in Driving Using Type-2 Fuzzy Classifiers," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 6, pp. 437–453, 2017.

[8] A. Pikovsky, M. Rosenblum, J. Kurths, and J. Kurths, *Synchronization: a universal concept in nonlinear sciences*. Cambridge university press, 2003, vol. 12.

[9] J. Qu, R. Wang, C. Yan, and Y. Du, "Oscillations and synchrony in a cortical neural network," *Cognitive neurodynamics*, vol. 8, no. 2, pp. 157–166, 2014.

[10] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring Phase Synchrony in Brain Signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[11] J.-P. Lachaux, E. Rodriguez, M. Le Van Quyen, A. Lutz, J. Martinerie, and F. J. Varela, "Studying single-trials of Phase Synchronous activity in the Brain," *International Journal of Bifurcation and Chaos*, vol. 10, no. 10, pp. 2429–2439, 2000.

[12] N. Caramia, F. Lotte, and S. Ramat, "Optimizing Spatial Filter pairs for EEG Classification based on Phase-Synchronization," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2049–2053.

[13] S. Kumar, T. K. Reddy, and L. Behera, "EEG based Motor imagery classification using Instantaneous Phase Difference Sequence," in *IEEE International Conference on Systems, Man, and Cybernetics*, no. CONF, 2018.

[14] W. Kong, Z. Zhou, B. Jiang, F. Babiloni, and G. Borghini, "Assessment of Driving Fatigue based on Intra/Inter-region Phase Synchronization," *Neurocomputing*, vol. 219, pp. 474–482, 2017.

[15] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial Filtering for EEG-based Regression problems in Brain-Computer Interface (BCI)," *IEEE Transactions on Fuzzy Systems*, 2017.

[16] J. Frey, M. Daniel, J. Castet, M. Hachet, and F. Lotte, "Framework for Electroencephalography-based evaluation of user experience," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 2283–2294.

[17] A. Meinel, S. Castaño-Candamil, J. Reis, and M. Tangermann, "Pre-trial EEG-based single-trial motor performance prediction to enhance neuroergonomics for a hand force task," *Frontiers in human neuroscience*, vol. 10, p. 170, 2016.

[18] D. Wu, B. J. Lance, V. J. Lawhern, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-based User Reaction Time Estimation using Riemannian Geometry Features," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2157–2168, 2017.

[19] A. Meinel, F. Lotte, and M. Tangermann, "Tikhonov Regularization Enhances EEG-based Spatial Filtering for Single Trial Regression," in *International Brain-Computer Interface Conference*, 2017.

[20] T. K. Reddy, V. Arora, L. Behera, Y. Wang, and C. T. Lin, "Multi-class Fuzzy time-delay Common Spatio-Spectral Patterns with fuzzy Information theoretic optimization for EEG-Based Regression Problems in Brain-Computer Interface (BCI)," 2018, submitted to IEEE Transactions on Fuzzy Systems.

[21] R. Frömer, M. Maier, and R. Abdel Rahman, "Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models," *Frontiers in Neuroscience*, vol. 12, p. 48, 2018.

[22] I. Sturm, S. Dähne, B. Blankertz, and G. Curio, "Multivariate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli," *PloS one*, vol. 10, no. 10, p. e0141281, 2015.

[23] Y.-T. Liu, Y.-Y. Lin, S.-L. Wu, C.-H. Chuang, and C.-T. Lin, "Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 2, pp. 347–360, 2016.

[24] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from EEG signals using Online Weighted Adaptation Regularization for Regression (OwARR)," *IEEE Transactions on Fuzzy Systems*, 2016.

[25] L. Xie, Z. Deng, P. Xu, K.-S. Choi, and S. Wang, "Generalized Hidden-Mapping Transductive Transfer Learning for Recognition of Epileptic Electroencephalogram Signals," *IEEE Transactions on Cybernetics*, 2018.

[26] D. Wu, "Pool-Based Sequential Active Learning for Regression," *arXiv preprint arXiv:1805.04735*, 2018.

[27] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Granular Fuzzy Regression Domain Adaptation in Takagi-Sugeno Fuzzy Models," *IEEE Transactions on Fuzzy Systems*, 2017.

[28] S. Sun, X. Xie, and C. Dong, "Multiview Learning With Generalized Eigenvalue Proximal Support Vector Machines," *IEEE Transactions on Cybernetics*, 2018.

[29] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer Learning in Brain-Computer Interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[30] Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," *arXiv preprint arXiv:1707.08114*, 2017.

[31] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A Review of Classification Algorithms for EEG-based Brain–Computer Interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[32] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask Learning for Brain-Computer Interfaces," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 17–24.

[33] C.-H. Chuang, L.-W. Ko, T.-P. Jung, and C.-T. Lin, "Kinesthesia in a sustained-attention driving task," *Neuroimage*, vol. 91, pp. 187–202, 2014.

[34] S. Kerick, C. Chuang, J. King, T. Jung, J. Brooks, B. Files, K. McDowell, and C. Lin, "Inter-and Intra-individual variations in sleep, subjective fatigue, and vigilance task performance of students in their real-world environments over extended periods," 2016.

[35] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis," *Frontiers in neuroinformatics*, vol. 9, p. 16, 2015.

[36] http://vislab.github.io/EEG-Clean-Tools/.

[37] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C. Lin, and T.-P. Jung, "Transfer Learning with Large-Scale data in Brain-Computer Interfaces," in *IEEE Annual International Conference of the Engineering in Medicine and Biology Society*. IEEE, 2016.

[38] M. Sánchez-Fernández, M. de Prado-Cumplido, J. Arenas-García, and F. Pérez-Cruz, "SVM Multi-Regression for nonlinear channel estimation in Multiple-Input Multiple-Output systems," *IEEE transactions on signal processing*, vol. 52, no. 8, pp. 2298–2307, 2004.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[40] I. Kondofersky and F. J. Theis, "TREVOR HASTIE, ROBERT TIBSHIRANI, AND MARTIN WAINWRIGHT. Statistical Learning with Sparsity: The Lasso and Generalizations. Boca Raton: CRC Press." *Biometrics*, vol. 74, no. 2, pp. 769–769, 2018.

[41] G. C. Cawley and N. L. Talbot, "On Over-Fitting in Model Selection and subsequent Selection Bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.

[42] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300, 1995.

[43] O. Bousquet and A. Elisseeff, "Stability and Generalization," *Journal of machine learning research*, vol. 2, no. Mar, pp. 499–526, 2002.

[44] T. K. Reddy, V. Arora, and L. Behera, "HJB-Equation-Based Optimal Learning Scheme for Neural Networks With Applications in Brain–Computer Interface," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–12, 2018.

[45] T. K. Reddy and L. Behera, "Online Eye State recognition from EEG data using Deep architectures," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 000 712–000 717.

[46] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic Facial Expression Recognition system using Deep Network-based Data Fusion," *IEEE transactions on cybernetics*, 2016.