EEG-Based Drowsiness Detection With Fuzzy Independent Phase-Locking Value Representations Using Lagrangian-Based Deep Neural Networks

Tharun Kumar Reddy, Vipul Arora, Vinay Gupta, Rupam Biswas, and Laxmidhar Behera, *Senior Member, IEEE*

***Abstract*—Passive electroencephalogram (EEG) brain–computer interfaces (BCI) have common usage in the area of Driver Drowsiness Detection. The approach presented herein identifies the cognitive state of the user while no mental action is required. Data recorded in EEG-based BCI experiments are generally noisy, nonstationary, and contaminated with artifacts that can deteriorate any analyzer's performance. Recently, common spatial patterns (CSPs) have been adapted with EEG state-space incorporating spatiospectral optimization using fuzzy time delay (FTD-CSSP). Temporal phase disparity sequence (TPDS) is used to measure synchrony between EEG signals. The output of Linear transforms operating on the TPDS constitute useful features for EEG regression problems. On similar lines, this article proposes spatiospectral optimized fuzzy-independent phase-locking value (SSO-FIPLV) representations (exploiting the spatiospectral information from TPDS) for EEG signals to monitor a user's cognitive states. Specifically, we analyze changes in EEG synchronization for a car driver as she/he drifts between alert and drowsy states. We use neural networks (NNs) for prediction. This article also proposes a cutting-edge method for training NN using the Euler–Lagrangian formulation. A stability proof is provided for the intended training approach alongside, and the performance is corroborated on the EEG reaction time prediction task, both within and across subjects, using a publicly available dataset. The NN trained by the proposed approach performs better than other competitive approaches in terms of minimizing root-mean-squared error and maximizing correlation coefficient. Channelwise feature importance in terms of average relevance values calculated from NN feature representations is visualized in the form of Topoplots using layerwise relevance propagation for regression.**

***Index Terms*—Brain–computer interface (BCI), electroencephalogram (EEG), fuzzy time-delay common spatiospectral patterns (FTD-CSSP), Lagrangian, neural networks (NNs), reaction time (RT).**

Manuscript received February 6, 2021; revised July 10, 2021; accepted September 11, 2021. Date of publication October 8, 2021; date of current version December 17, 2021. This article was recommended by Associate Editor C.-T. Lin. *(Corresponding author: Laxmidhar Behera.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Veterans General Hospital, Taipei, Taiwan under Term Accession No. OBI:0001928 ERO:0001337, and performed by recommendations in the Guide for the Committee of Laboratory Care and Use of the National Chiao Tung University, Taiwan.

Tharun Kumar Reddy is with the Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India (e-mail: tharun.reddy@ece.iitr.ac.in).

Vipul Arora, Vinay Gupta, Rupam Biswas, and Laxmidhar Behera are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India (e-mail: vipular@iitk.ac.in; guptak@iitk.ac.in; rupam@iitk.ac.in; lbehera@iitk.ac.in).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TSMC.2021.3113823.

Digital Object Identifier 10.1109/TSMC.2021.3113823

I. Introduction

Driver debility ending up in sleepiness has been noted to be an important factor liable for severe road casualties. In an article from U.S. NHTSA, it is established that drivers' fatigue concludes in 1550 demises, 71 000 wounded, and U.S. $12.5 billion losses in revenue annually [1]. In this era of deep learning, a lot of technologies are coming up which automatically detect drowsiness based on signals recorded from the body. Prior research in drowsiness systems [1] demonstrates that physiological parameters-based techniques (electroencephalogram (EEG), ECG, etc.) give more accurate results than others (vision-based sensors). Amidst such signals, EEG stands as the most steady marker of the cognitive state because it is closely connected to the performance of psychological and biological actions [2].

A wide range of methods is proposed in the literature that uses power spectrum-based analysis for drowsy state classification [3]–[6]. Within the same class of approaches, mutual information-based wavelets [7] and entropy-based features [8] constitute the alternate approaches using the EEG signal amplitude. Prior studies [9], [10] reveal that the EEG spectra in theta rhythm (4–7 Hz) and alpha rhythm (8–11 Hz) usually reflect the cognitive state and memory performance. Hence, EEG spectra in theta and alpha rhythms can be used to derive the drivers' alert models and detect their cognitive state. Beta band (β) conforms to the interval of 13–30 Hz. It conveys tension and anticipation and is usually displayed in both alert and anxious subjects. The fluctuations in β activity during the individual fatigues are still unclear [11]. Gamma band (γ) consists of frequencies above 30 Hz and usually does not have an impact on drowsiness detection [12].

Another class of approaches relies on the time-varying phase information recorded from multiple EEG signal channels. Phase-dependent methods are based on the correlation between individual signal channel pairs by studying the interconnection of the transitory phase across signals independent from their amplitudes [13]. In addition, such approaches are found useful in implementing psychotherapies [14]. Several studies have attempted to localize the drowsiness (or discrimination of drowsy and awake states) to certain brain regions [15], [16]. In other words, the working of phase synchrony is crucial for processing within a brain region as well as establishing information transfer between different brain regions [17]. Prior studies (cf. [18, Fig. 1] and [19, Fig. 2A]) in this area demonstrated that study of the six main sources (frontal, central, parietal, occipital, left motor area, and right motor area) in the brain capture the drowsiness phenomenon. In recent works [19], [20], 30 channels are used to cover all these regions [additionally (A1 and A2 references placed on the mastoid bones)] and this is a standard montage used in a lot of drowsiness research published across time [19], [20].

Phase synchrony refers to a study of an interplay between two EEG channels by only looking at the momentous phase disparity among signals independent of their magnitude [21]. In addition, the part played by synchronization is subject to specific frequencies. For example, synchrony in EEG theta waves is contrasted to active memory capacity [22], while high-frequency synchrony is instrumental to disseminate messages [23]. The phase-locking value (PLV) [24] explicitly quantifies frequency synchrony among the multichannel signals. PLV is nevertheless limited to compare frequency synchrony at a fixed time across trials. Since there is also a need to quantify trial-specific synchrony of the phase, researchers have come up with a per trial PLV [25]. Later, several works further modified per trial PLV [26]. Recently, a method was proposed in [27] that relies on features from the temporal phase disparity sequence (TPDS) for individual brain–computer interface trials. Furthermore, in [28], for the first time, authors proposed the differential phase synchrony (DPS) features for regression while combining TPDS with fuzzy common spatial filtering for regression-one versus rest (FCSPR-OVR). In all of these approaches, optimization is performed only across the spatial domain while not considering the optimization of spectral content. In a recent work, we proposed FTD CSSP-OVR [29], which performs joint spatiospectral filter optimization for EEG-regression. In this work, we propose novel spatiospectrally optimized fuzzy-independent PLV (SSOFIPLV) feature representations for drowsiness detection (SSO-FIPLV). We tested the SSO-FIPLV features for reaction time (RT) prediction in an EEG-based persistent attention task (PAT) [30].

Over the last decade, we have seen the contributions made in the field of neural networks (NNs) for many applications [31]. In particular, for EEG-based driver drowsiness detection application, NNs have received extensive attention, for instance, deep feedforward NNs (FNNs) [28], [32], [33], common Bayesian network [34], and fuzzy NN [35]. Recently, a novel complex network (CN) [36]-guided broad learning system is proposed to conceive an EEG data-driven drowsiness detection. The superiority of NN over various other models is the automated feature learning/feature extraction. In literature, training algorithms for the NNs are categorized into single-step and multistep algorithms depending upon the number of weight change steps per iteration. In recent times, Gradient Descent [BackPropagation (BP)], AdaGrad (ADG) [33], Rmsprop, Adam [37], etc., are the few examples of single-step learning algorithms for training NN. Several single-step weight update algorithms derived using recursive least squares [38], extended Kalman filtering [39], and Lyapunov stability theory [33] have been adapted from other fields, such as the optimal control theory and signal processing.

Multistep approaches [40] distinctly update only a subset of the weights at once. Output weight optimization–BP (OWO-BP), a multistep approach [40], updates the weights in the input layer while subsequently solving the linear equations for the weights in the output layer. Similar to many first-order algorithms, such as BP, LM, Adagrad etc., the multistep OWO-BP is susceptible to sluggish convergence and does not accommodate the affine invariance property [40]. Therefore, it is perceptible to input data mean and the selection of initial weights. However, we will focus on the single-step-based methods in this work. In [33] and [41], an update law has been derived to train the FNN by using the Hamilton–Jacobi–Bellman (HJB) framework. Dynamic programming is a prominent paradigm for optimal control of dynamical systems [42]. The learning process in NN can be treated as a dynamical system [31]. Dynamic optimization estimates a control input which is optimal, i.e., it leads to an optimal weight trajectory for the case of NN that optimizes a predefined objective. In this work, we hereby introduce a novel Lagrangian-based optimization approach while incorporating neural system dynamics for adapting FNN. Several supervised learning algorithms are explored in the literature for Drowsiness Detection. Support vector machine (SVM) is the most commonly used classifiers/regressors, which provided better accuracy and speed in most of the situations, but it is not suitable for large datasets with multiple subjects [31]. Also, recently, learning-based methods are proposed for drowsiness detection [43]. But, in the literature, sufficient testing or validation has not been done with diverse optimal learning methods for feedforward NN architectures, which we explored in this work. Interindividual variability is another big challenge that restricts the commercial use of physiological signal-based drowsiness detection systems, as generalization is the major problem. We demonstrated the generalizability of our method across subjects also indicating it in this regard in the supplementary material. This has been highlighted in green in the supplementary material. The error in RT prediction translates to the error in distance in detecting a drowsy driver. A new metric is formulated in this regard [denoted as error in estimated driving distance (EEDD)].

The central purposes leading to this research are as follows.

1) Toward establishing the efficacy in regard to a novel Lagrangian and neural system dynamics-based NN regressor for a PAT.
2) To represent the drivers' RT prediction as a learning task based on SSO-FIPLV features and adapting computationally intelligent models with the designated objectives.

The novelty and important contributions to be taken out of this research are as follows.

1) A unique Lagrangian and NN system dynamics-based learning scheme for NN regression problems.
2) Novel SSO-FIPLV features establish the efficacy of using phase-based EEG signal processing for the PAT.
3) Comprehensive tests (inclusive of contrast with advanced prediction models) are used with the goal of

substantiating the efficacy of the developed approach in an EEG PAT.

This work is constituted as follows. Section II contains the formulation of a novel NN weight update law. Section III describes the proposed SSO-FIPLV features. Section IV-D assesses the effectiveness of feature sets on the EEG PAT. Sections IV-E and IV-F describe the performance of Lagrangian-based learning DNN and all other key results with sufficient comparison to baseline approaches. Finally, conclusive assertions are indicated in Section V. Here, # is used to denote "number of."

II. Proposed Method

Consider an FNN, having O layers (hidden+output), with N_o neurons in the oth layer ($o \in \{0, 1, \ldots, O\}$, $o = 0$ for the input layer) and N_θ total number of tunable weights. For supervised learning, we are given N pairs of input and target output in a batch. Now, learning can be seen as a dynamic control problem, where the error $\mathbf{e} \in \mathbb{R}^{N_O N}$ between the estimated and the target outputs changes with a change $\mathbf{u} \in \mathbb{R}^{N_\theta}$ in FNN weights θ

$$\dot{\mathbf{e}} = -\mathbf{J}\mathbf{u}; \quad \mathbf{u} = \dot{\theta}. \tag{1}$$

Here, $\mathbf{e} = [\mathbf{e}_1^\mathsf{T}, \mathbf{e}_2^\mathsf{T}, \ldots, \mathbf{e}_N^\mathsf{T}]^\mathsf{T}$, and $\mathbf{J} = [\mathbf{J}_1^\mathsf{T}, \mathbf{J}_2^\mathsf{T}, \ldots, \mathbf{J}_N^\mathsf{T}]^\mathsf{T}$ is an $N_O N \times N_\theta$ matrix. The derivation of (1) and further explanation about different terms is provided in the supplementary material under Section III. The cost function in the discrete form shown in (2) is taken from [33]

$$V(\mathbf{e}_k, \mathbf{u}_k) = \sum_{l=k}^{K} L(\mathbf{e}_l, \mathbf{u}_l)\Delta t \tag{2}$$

$$\text{where, } L(\mathbf{e}_l, \mathbf{u}_l) = \frac{1}{2}\left(\mathbf{e}_l^\mathsf{T}\mathbf{Q}\mathbf{e}_l + \mathbf{u}_l^\mathsf{T}\mathbf{R}\mathbf{u}_l\right). \tag{3}$$

Here, k is the iteration index. To optimize the cost function defined in (2) while satisfying constraint (1), HJB formulation has been used in the past, and the weight update laws so derived have been employed [33], [44] to update NN weights during learning. But in this article, our approach is inspired from the Euler–Lagrangian formulation, which is commonly used for functional optimization. It is popularly used in the calculus of variation and in physics for action minimization/maximization. The constraint from (1) can be discretized using the Euler approximation and added to the cost function (2) to get the Lagrangian cost function

$$\mathcal{L}(\mathbf{e}_k, \mathbf{u}_k) = \sum_{l=k}^{K}\left(L(\mathbf{e}_l, \mathbf{u}_l)\Delta t + \lambda_l{}^\mathsf{T}(\mathbf{e}_l - \mathbf{e}_{l-1} + \mathbf{J}_l\mathbf{u}_l\Delta t)\right). \tag{4}$$

To derive the weight update law, we optimize $\mathcal{L}$ w.r.t. $\mathbf{e}_l$ and $\mathbf{u}_l$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{e}_l} = 0 \implies \mathbf{Q}\mathbf{e}_l\Delta t + \lambda_l - \lambda_{l+1} = 0 \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}_l} = 0 \implies \mathbf{R}\mathbf{u}_l + \mathbf{J}_l^\mathsf{T}\lambda_l = 0. \tag{6}$$

Eliminating λ's from (5) and (6), we get the optimal $\mathbf{u}$ as

$$\mathbf{u}_{l+1} = \mathbf{R}^{-1}\mathbf{J}_{l+1}^\mathsf{T}\left((\mathbf{J}_l^\mathsf{T})^\dagger\mathbf{R}\mathbf{u}_l - \mathbf{Q}\mathbf{e}_l\Delta t\right). \tag{7}$$

The FNN weights can be updated as

$$\theta_{k+1} = \theta_k + \eta\mathbf{u}_k \tag{8}$$

where $\eta \in \mathbb{R}$ is the learning rate and $\dagger$ denotes Moore–Penrose pseudo inverse. The optimal $\mathbf{u}_k$ has been derived above. The learning rate can also be optimized. Several methods have been proposed earlier for optimal learning rates for gradient descent-based learning schemes [33]. We use the adaptive gradients [AdaGrad (ADG)] method to update the learning rate. The final weight update law becomes

$$\theta_{k+1} = \theta_k + \frac{\eta}{\sqrt{\sum_{l=0}^{k}||\mathbf{u}_l||^2}}\mathbf{u}_k. \tag{9}$$

Input to state stability (ISS) of the proposed approach is provided in Section 3 of the supplementary material (Supplementary.pdf). The ISS criterion ensures that the update law is stable.

III. SSO-FIPLV Representations

Let $\mathbf{X}^r \in \mathbb{R}^{C\times T}, r \in \{0, 1, \ldots, N\}$ denote the rth EEG trial, C denotes the #channels, and T denotes the # time samples. The output variable (here, RT in each trial) denoted by Y^r or y_r is split into $K+1$ regions uniformly (with "K" fuzzy classes) [cf. Fig. 1(a)]. Any Y^r or y_r lies in a fuzzy class with a respective membership value $\mu(Y^r) \in [0, 1]$. In Fig. 1(a), we partition $[0, 100]$ into $K+1$ intervals where $P_1, P_2, \ldots, P_{K-1}, P_K$ denote percentile points with $(100K)/(K+1)$ standing for $(100K)/(K+1\text{th})$ percentile point.

A. Fuzzy Multiple Class Common Spatial Patterns

Common spatial patterns (CSPs) [45] is an important supervised technique to be applied for dealing with EEG-based machine learning problems. In the literature, three approaches are generally used to compute CSP filters: 1) one versus one; 2) one versus rest (OVR) [30]; and 3) joint approximate diagonalization (JAD) [46]. In this article, we demonstrate the optimization of the CSP algorithm for regression using JAD.

1) Joint Approximate Diagonalization Approach: Consider a regression problem, where the output variable is fuzzified using K fuzzy classes [cf. Fig. 1(a)]. We then compute an averaged signal $\overline{\mathbf{X}}_s$ for every fuzzy class s. Then, the covariance matrix corresponding to each fuzzy class $\overline{\Sigma}_s$ is determined

$$\overline{\mathbf{X}}_s = \frac{\sum_{r=1}^{N_s}\mu_{r,s}\mathbf{X}^r}{N_s} \quad s \in \{1, 2, \ldots, K\}. \tag{10}$$

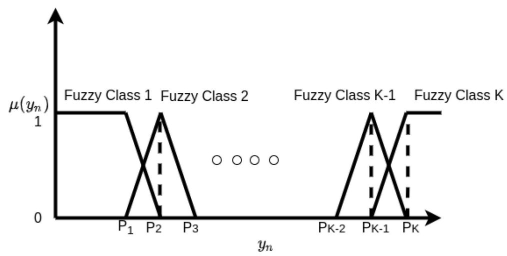
In (10), N_s denotes # trials in the sth fuzzy class

$$\overline{\Sigma}_\mathbf{s} = \overline{\mathbf{X}}^s\overline{\mathbf{X}}^{s\mathsf{T}} \quad s \in \{1, 2, \ldots, K\}. \tag{11}$$

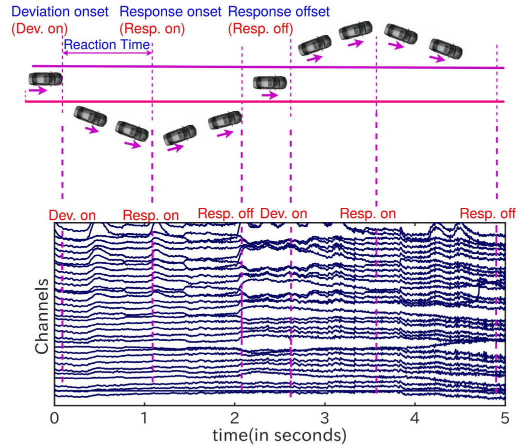
We normalize the class covariance matrices thus obtained above by using

$$\overline{\Sigma}_s = \frac{\overline{\Sigma}_s}{Tr(\overline{\Sigma}_s)} \quad \forall s \in \{1, 2, \ldots, K\}. \tag{12}$$

After computing the average and normalized covariance matrices for all K classes, a transformation matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$



(a)



(b)

Fig. 1. Plots of the experimental paradigm with fuzzy set-based analysis. (a) Fuzzified RT. (b) EEG signals with associated events. (EEG and behavior data were recorded simultaneously.)

needs to be found which concurrently diagonalizes them. Specifically, $\mathbf{W}$ must satisfy

$$\mathbf{W}\overline{\Sigma}_s\mathbf{W}^\mathsf{T} = \mathbf{D}_s \ \forall s \in \{1, \ldots, K\} \tag{13}$$

$$\sum_{s=1}^{K}\mathbf{D}_s = \mathbf{I}_{N\times N}. \tag{14}$$

Although, such kind of diagonalization can be obtained accurately for $N = 2$, only approximate solutions are existing for $N > 2$. Employing the weight matrix $\mathbf{W}$, we compute the transformed EEG signal

$$\mathbf{Z} = \mathbf{W}\mathbf{X}^r. \tag{15}$$

The spatial filters are the rows of matrix $\mathbf{W}$.

B. Fuzzy Time-Delay CSSP

In [47], we proposed a novel robust invariant CSP features extraction methodology by incorporating a fuzzy time-delay variable in the state-space model. We further demonstrated an improved version of (15) by introducing a generalizable state-space reconstruction with a fuzzy time delay

$$\mathbf{Z^r} = \int_{t'}\mu_{(t')}\mathbf{W}^{(t')} * (\delta^{t'}\mathbf{X^r})dt'. \tag{16}$$

Here, $\delta^{(t')}$ is the delay operator on the signal state space, $\mu_{t'}$ is the membership value of variable t', and $\mathbf{W}^{(t')}$ stands for the optimal fuzzy time-delay CSSP (FTDCSSP) transform

$$\delta^{(t')}(\mathbf{X}^r) = \mathbf{X}^{(r-t')}. \tag{17}$$

We consider the time delay term t' to follow an exponential membership $e^{-t'}$ for the reason that in system dynamics, higher order delays add marginally to the integral beyond a specific delay threshold. Equation (16) is thus reduced to

$$\mathbf{Z^r} \approx \sum_{t'=0}^{2}\mu_{(t')}\mathbf{W}^{(t')} * (\delta^{t'}\mathbf{X^r}). \tag{18}$$

Furthermore, (18) can be reduced to get

$$\mathbf{Z}^r = \left[\mathbf{W^{(0)}}\ \mathbf{W^{(1)}}\ \mathbf{W^{(2)}}\right]\begin{bmatrix}\mu_0\mathbf{X}^{(r)}\\ \mu_1\mathbf{X}^{(r-1)}\\ \mu_2\mathbf{X}^{(r-2)}\end{bmatrix}. \tag{19}$$

The composite vector $\begin{bmatrix}\mu_0\mathbf{X}^{(r)}\\ \mu_1\mathbf{X}^{(r-1)}\\ \mu_2\mathbf{X}^{(r-2)}\end{bmatrix}$ is indicated as the resultant EEG signal $\mathbf{X}^r$ in (15). Furthermore, fuzzy class covariance matrices ($K = 3$) $\overline{\Sigma}_\mathbf{1}, \overline{\Sigma}_\mathbf{2}, \overline{\Sigma}_\mathbf{3}$ are obtained from $\begin{bmatrix}\mu_0\mathbf{X}^{(k)}\\ \mu_1\mathbf{X}^{(k-1)}\\ \mu_2\mathbf{X}^{(k-2)}\end{bmatrix}$. Following the approach mentioned in steps (11)–(13), we generate a composite weight matrix $\left[\mathbf{W^{(0)}}\ \mathbf{W^{(1)}}\ \mathbf{W^{(2)}}\right]$ as the result of our optimization problem,. Each of the three matrices $\mathbf{W^{(0)}}$, $\mathbf{W^{(1)}}$, and $\mathbf{W^{(2)}}$ correspond to $\mu_0\mathbf{X}^{(k)}$, $\mu_1\mathbf{X}^{(k-1)}$, and $\mu_2\mathbf{X}^{(k-2)}$, respectively. Among the rows of the matrices $\mathbf{W^{(0)}}$, $\mathbf{W^{(1)}}$, and $\mathbf{W^{(2)}}$, the FTDCSSP filters are those filters which maximize the fuzzy mutual information criterion (22) mentioned in [29]. Before applying CSP filters [27], it is advised to pick at least two separate filters for individual class corresponding to maximum and minimum variance directions. Following this guideline, we have chosen $K = 3$ and $F = 2K = 6$ in the experiments. In (19), to calculate three weight matrices, we need to calculate $3 \times 2K = 18$ row vectors. The entire approach is summarized taking the shape of Algorithm 1.

C. Spatiospectral Optimized Fuzzy Independent Phase-Locking Value Representations

In this section, we propose the SSO-FIPLV obtained as an output of fuzzy spatiospectral filter optimization over TPDS. We describe the TPDS in Section III-C1. Subsequently, fuzzy spatiospectral optimization over TPDS is described in Section III-C2.

1) Temporal Phase Disparity Sequence: Consider two time series $r_1(t)$ and $r_2(t)$ whose phases are $\psi_1(t)$ and $\psi_2(t)$, respectively. The per-trial PLV (pPLV) [49] concerning each trial is

Algorithm 1 FTDCSSP

Input: EEG training signals $(X^{(r)}, Y^{(r)})\ \ r \in \{1, 2, ..N\}$
$\mathbf{X}^{(r)} \in \mathbb{R}^{C\times T}$
'L_s', # spatial filters per fuzzy class 's' (let $L_s = F$)
'K' is the # fuzzy classes
Output: Spatial filter matrices $\left[\mathbf{W^{(0)}}\ \mathbf{W^{(1)}}\ \mathbf{W^{(2)}}\right]$
EEG signals passed through PREP [48];
EEG signals processed through PREP are passed through band-pass filter in range [1-20] Hz;
Calculate the thresholds P_r for specific fuzzy classes (Fig. 1a);
Estimate $\overline{\mathbf{X}}_s$ using (10);
Estimate Co-variance $\overline{\Sigma}_{\mathbf{s}}$ for respective fuzzy class 's' with (11);
Adapt the covariances for respective classes through (12);
Calculate filter matrix $\mathbf{W}$ by (13);
Excerpt 'L_s' filters for respective fuzzy class as per equation (27);
Spatial filter matrix $\left[\mathbf{W^{(0)}}\ \ \mathbf{W^{(1)}}\ \ \mathbf{W^{(2)}}\right]$ is obtained, comprising $\sum_{i=1}^{K} L_s$ # rows;
return $\left[\mathbf{W^{(0)}}\ \ \mathbf{W^{(1)}}\ \ \mathbf{W^{(2)}}\right]$

determined by

$$pPLV = \left|\frac{1}{N_t}\sum_{i=1}^{N_t} e^{|\psi_1(i)-\psi_2(i)|}\right| \tag{20}$$

where N_t is the sample size per trial. The temporal phase $\psi(t)$ is calculated from the complex-valued time series (using the Hilbert transform). For a generic signal $r(t)$, the corresponding complex signal $z(t)$ is obtained as

$$z(t) = r(t) + i\tilde{r}(t) \tag{21}$$

$$\tilde{r}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{r(t')}{t-t'}dt'. \tag{22}$$

Here, $\tilde{r}(t)$ is the Hilbert transform of $r(t)$. The temporal phase $\psi(t)$ is then obtained as

$$\psi(t) = \arctan\left(\frac{\tilde{r}(t)}{r(t)}\right). \tag{23}$$

$pPLV \in [0, 1]$ and the corner values pertain to the instances of signal with no synchrony and full synchrony, respectively. TPDS $\Delta\psi(t)$ between a pair of distinct signals $r_1(t)$ and $r_2(t)$ is designated as

$$\Delta\psi(t) = |\psi_1(t) - \psi_2(t)|. \tag{24}$$

Kumar *et al.* [27] emphasized the concept of the variance of TPDS with pPLV. Later, a framework was proposed to calculate a linear transformation that enhances the variance of TPDS over a specific class whilst concurrently minimizing it over an auxiliary class. This scheme is analogous to CSP yet, in contrast, it largely uses knowledge of phase for two-class classification. Hence, extracting a similarity from the Fuzzy CSP for Regression, a novel scheme is devised in [28] which computes a linear transform on the TPDS in an aforesaid manner with variance optimization across fuzzy classes. Because TPDS is used to gauge co-instantaneity between EEG signals, we indicate the extracted representations as phase-synchrony (PS) features. The extensions to multiclass paradigms of CSP (OnevsOne, OnevsRest) are based on heuristics and a more principled approach uses JAD for CSP. But, CSP implemented using JAD is tantamount to independent component analysis (ICA) [46]. Motivated by these arguments, we propose a novel SSOFIPLV representation for the EEG RT regression problem.

2) Fuzzy Spatial Filter Optimization With TPDS (Fuzzy CSPR-OVR Applied Over TPDS): Fuzzy CSPR-OVR [30] widens the scope of CSP to solve regression problems by incorporating fuzzy sets. FTDCSSP [29] further improves the scope of fuzzy spatial filtering by incorporating spatiospectral content in filters. Multichannel EEG time series is subject to possess a subsided SNR, due to spatial blearing and spattering effects. With an objective to secure discriminative PS representations to classify across fuzzy classes, we estimate spatial filters to magnify the variance of the temporal phase with a specific fuzzy class and minimize over others. Reddy *et al.* [29] proposed an algorithm to optimize the spatial filters in order to maximize the resulting PS feature discriminative power. Mathematically, it can be written as

$$\mathbf{W}_u{}^* = \arg\max_{\mathbf{W}} \frac{Tr(\mathbf{W}^\intercal \chi_{\Delta\psi_u}\mathbf{W})}{Tr(\mathbf{W}^\intercal \sum_{v\neq u}\chi_{\Delta\psi_v}\mathbf{W})}. \tag{25}$$

Here, $\chi_{\Delta\psi_u}$ and $\chi_{\Delta\psi_v}$ are the covariance matrices of the TPDS for the fuzzy classes u and v. The column vectors of $\mathbf{W}$ are the spatial synchrony filters. An approach close to the FTDCSSP method is selected to extract features out of TPDS for each EEG trial filtered within a specific frequency spectrum. The computed representations are denoted as "DPS-CSPROVR" representations. In this work, we improve upon the feature representations by using SSO-FIPLV representations.

3) Filter Selection and Feature-Extraction for Proposed SSO-FIPLV Method (FTDCSSP Applied Over TPDS): FTDCSSP by JAD has been applied on the TPDS and further the obtained filters are to be selected using a criterion mentioned below. The eigenvalues corresponding to the JAD are given by

$$\lambda_s = \operatorname{diag}\left(\mathbf{W}^T\Sigma_s\mathbf{W}\right)\ \ s \in \{1, 2, \ldots, K\}. \tag{26}$$

Here, K denotes the number of fuzzy classes and λ_i stands for the vector of eigenvalues. Furthermore, we employ a filter selection criterion (27), where the eigenvalues obtained for each of the fuzzy classes are transformed and the L eigenvectors corresponding to the L largest eigenvalues are selected. It is possible that some eigenvectors may be repeated, so in that case, we have chosen the eigenvector corresponding to the next largest eigenvalue

$$\lambda_{i,j} = \max\{\lambda_{i,j}, \frac{\lambda_{i,j}}{1+\frac{(K-1)^2\lambda_{i,j}}{1-\lambda_{i,j}}}\}. \tag{27}$$

Furthermore, the so-obtained filters constitute the weight matrix $\mathbf{W}$. The TPD sequences collected in triads (19) are transformed by $\mathbf{W}$ to obtain a $6K \times T$ matrix from which we extract the final "SSO-FIPLV" features. $\mathbf{F}=\begin{bmatrix} F_1 \\ .. \\ F_{6K} \end{bmatrix}$ where F_i is given by $\log_{10}[(\|Z_i\|^2)/(\sum_{j=1}^{6K}\|Z_i\|^2)]$. We refer to them as independent phase-locking representations due to the reason that CSP by JAD is tantamount to ICA and since the transformations are computed on TPDS, which indicates the phase-locking level or synchronization between two waveforms, they are denoted as Phase-locking-independent value representations (PILV). Incorporating spatiospectral optimization for regression, the features so obtained constitute the "SSO-FIPLV."

IV. Experiments and Discussion

In this section, we evaluate the performance of the proposed approach with respect to the existing approaches for RT prediction on EEG data collected in a PAT [20]. We compare the performance of the proposed approach with different feature extraction techniques, regression models, and NN learning schemes.

A. PAT and EEG Data Preprocessing

The paradigm was designed to quantitatively measure the subject's RT to perturbations during a continuous driving task. A brief pictorial depiction can be found in Fig. 1(b), but the experimental description and preprocessing is described in detail in Section 1 of the supplementary material (Supplementary.pdf).

B. Feature Extraction

EEG signal further is prefiltered into a frequency range of [1, 20] Hz. Comprehensive reasons for selection of this band can be found in Section 5, page 4 of the supplementary material. Alpha and Theta power features [9], [10] can be extracted post spatial filtering (FS3) ($K = 3$, $F = 10$) and also without spatial filtering (FS1). F and K are two parameters whose appropriate values are set from [28] and [30] to ensure the correctness of comparison. We evaluate every combination of the feature set and regression approach using k-fold cross-validation ($k = 8$) and leave-one last session out cross-validation. There are two sets of subjects in the current study. One set of subjects have multiple sessions of data recorded at different times (denoted Subject set-2). We performed leave-one last session out cross-validation for these subjects. While for the other set of subjects (subjects 3, 5, 7, 12, 21, 24, 26, and 27) (denoted Subject set-1) with a single session of recorded data, 8-fold cross-validation is performed. The causality of the machine learning model during analysis could be maintained for this second set of subjects with multiple recorded sessions (Subject set-2) while it is not the case for the first set of subjects (Subject set-1). Each subject's sessionwise data can be found in Table 8 of the supplementary section.

1) Power in Theta and Alpha bands is extracted in dB from the prefiltered EEG trials to get the feature set FS1.
2) DPS features [28] are computed from the prefiltered EEG. We set $K = 3$ and $F = 21$ to get the feature set FS2.
3) Prefiltered EEG signal is passed through spatial filters via fuzzy CSPR-OVR. Furthermore, the power in Theta and Alpha bands are extracted in dB. We used $F = 10$ so that the filtered signal (30×1250) and the original signal (30×1250) have the same dimensions, ensuring a fair performance comparison. The feature set so obtained is denoted by FS3. Each feature vector has a size of 60×1.
4) The proposed SSO-FIPLV features are calculated from the prefiltered EEG trials. We set $K = 9$ and $F = 10$ again for the same reason stated above for FS3. So, the obtained feature is of size $(54 \times 1) = 54$ (SSO-FIPLV) and is indicated as FS4.

TABLE I
Evaluation Performance of FS1, FS2, FS3, and FS4

Feature-set	CC	RMSE	EEDD (in metres)
FS1	0.574	0.0801	2.225
FS2	0.676	0.0715	1.986
FS3	0.598	0.0795	2.208
FS4 (proposed)	**0.716**	**0.0615**	1.708

The features computed above are passed through LASSO for getting the predicted value of RT.

C. Evaluation Criterion

RMSE and CC comprise the criteria used for measuring the results of the regression. Consider N number of training samples, y_{d_i} denoting the actual RT for the ith example and y_i denoting the predicted RT. If RMSE is smaller, then our system is trained correctly for predicting the RT of drowsy drivers with minor errors. CC depicts the usefulness of the predicted features for RT prediction. Furthermore, to enhance practicability of the developed system, we introduce a criterion of EEDD (28) (a similar one is reported in the discussion of [50]) to detect a drowsy driver. We assume the average speed of the vehicle is 100 km/h

$$EEDD = \text{Average speed of vehicle} \times \text{RMSE}. \qquad (28)$$

D. Comparison of SSO-FIPLV Features With Other Feature Sets

The mean RMSE and CC computed after passing all features (elucidated in Section IV-B) through the LASSO regression block are displayed in Table I. For subject set-1, we have taken an average of all the performances for each of the 8-folds. While for subject set-2, the performance on the last session is recorded. Also, the mean performances of all the subjects for all the folds are determined. Both FS4 and FS2 achieved much smaller RMSEs and much larger CCs compared to FS3 and FS1. In general, FS4 demonstrated the best performance indicating that it is well suited for the RT regression task. In addition, the EEDD is the smallest for FS4 in comparison to all other features helping us to detect drowsy drivers within the smallest driving distance. An in-depth analysis of results and implementation technicalities examining the performance of the features is provided in Section V of the supplementary material (Supplementary.pdf).

TABLE II
(a) p-VALUES FOR PAIRED T-TESTS (LGN VERSUS NN LEARNING MODELS (HJB, ADG, AND BP) FOR $\{FS4; FS2\}$) (b) REGRESSION PERFORMANCE COMPARISON OF FS4 WITH LGN VERSUS OTHERS (SVM, RR, AND LASSO)

(a)

	FS2								FS4							
	RMSE				CC				RMSE				CC			
	LGN	HJB	ADG	BP	LGN	HJB	ADG	BP	LGN	HJB	ADG	BP	LGN	HJB	ADG	BP
LGN		0.01	0.04	**0.001**		0.004	0.03	0.01		0.02	0.05	**0.001**		0.002	0.04	0.02
HJB	0.03		0.04	0.01	0.02		**0.001**	0.04	0.02		0.03	0.04	0.04		**0.001**	0.01
ADG	0.03	0.03		0.002	**0.001**	0.01		0.03	0.02	0.01		0.001	**0.001**	0.02		0.04
BP	0.01	0.014	0.002		**0.001**	0.02	0.02		0.04	0.012	0.001		**0.002**	0.01	0.04	

(b)

Regression used	CC	RMSE
SVR	0.727	0.0810
RR	0.683	0.0885
LASSO	0.693	0.0882
LGN (proposed)	**0.767**	**0.0685**

TABLE III
(a) p-VALUES FOR POST-HOC PAIRED T-TESTS (DNN TRAINED WITH LGN AGAINST OTHER REGRESSION MODELS (SVR, LASSO, AND RR) FOR $\{FS4; FS2\}$ (b) REGRESSION PERFORMANCE OF FS4 WITH DIFFERENT NN REGRESSORS

(a)

	FS2						FS4					
	RMSE			CC			RMSE			CC		
	M	SVR	LASSO	M	SVR	LASSO	M	SVR	LASSO	M	SVR	LASSO
M		0.03	**0.001**		**0.001**	0.001		0.04	**0.002**		**0.002**	0.01
RR	0.03	0.04	0.02	0.03	0.01	**0.000**	0.02	0.01	0.02	0.04	0.01	**0.000**
SVR	0.02		**0.001**	**0.000**		0.01	0.03		**0.002**	**0.001**		0.04
LASSO	0.01	0.022	**0.002**	**0.000**	0.04	0.02	0.03	0.032	**0.002**	**0.000**	0.03	0.02

(b)

Learning algorithm	CC	RMSE
BP	0.670	0.0932
ADG	0.684	0.0937
HJB	0.735	0.0781
LGN (proposed)	**0.767**	**0.0685**

E. Comparison of NN Lagrangian (LGN) Regression With Other NN Learning Methods

In practice, three approaches—1) HJB [28]; 2) BP; and 3) ADG [33]—are generally used for updating NN weights. We compare the performance of our method with all of the above approaches. For comparison of learning methods, we use a network with four hidden layers having 200, 150, 120, and 45 hidden units, respectively. A similar network was validated in previous work [28]. The proposed SSO-FIPLV features are fed as input to the network, which is trained using BP, ADG, HJB, and the proposed Lagrangian-based learning in each of the cases, respectively. The dimension of the SSO-FIPLV feature is 54×1. Two hundred iterations are used for training. For the optimization of DNN learning, ADG is instantiated with learning rate $\{\eta = 0.1\}$, drop-out rate ($p = 0.5$), L_1 regularization parameter is 0.01, the size of minibatch is set to 8, and R is also fixed to identity while using the Lagrangian-based update law. SSO-FIPLV features (input features) are normalized to zero mean and unit variance. RMSE and CC are averaged considering all the subjects with 20 unique weight initializations for the deep network and is reported in Table III(b). The Lagrangian is an established paradigm from the mathematical optimization literature, allowing objective-only or bound-constrained optimizers to be deployed in settings with constraints. Considering the mean performance, from Table III(b), the Lagrangian-based DNN approach clearly outperformed HJB, ADG, and BP-based NN regression as far as RMSE and CC concern. The HJB-based method is a greedy approach for learning in NNs. The approach based on HJB performed superior to BP and ADG methods as visible from the Table III(b). We performed a simple hyperparameter tuning to determine the value of the initial learning rate, using a few common values, namely, 0.1 and 0.01. The initial value of $\mathbf{u}$ is chosen to be equal to that of the initial value of the HJB-based update law $\mathbf{u}^*$ [33]

$$\mathbf{u}^*(\mathbf{e}(t)) = \frac{\sqrt{\mathbf{e}^\top(t)\mathbf{e}(t)}}{||\mathbf{J}^\top\mathbf{e}(t)||}\,\mathbf{R}^{-1/2}\mathbf{J}^\top\mathbf{e}(t). \tag{29}$$

We noted that this initialization gave us a better performance than either random initialization or initialization using the backpropagation update law. The initial weights of the network are chosen using Glorot initialization [51]. The corresponding percentage improvements of the proposed deep network trained with LGN over the competing NN-based regression designs are displayed in the Figs. 5 and 6 in the supplementary material. For instance, the terms in legend `LGN/HJB` represent the improvements with Lagrangian-based learning over the HJB method. On average, LGN recorded a 12.35% smaller RMSE and a 4.35% larger CC than that of HJB. On average, the LGN method finished by a 26.55% lower RMSE, and a 14.44% larger CC than BP. Also, LGN achieved a 26.98% lower RMSE, along with a 12.09% larger CC than ADG. In Figs. 5 and 6 (from the supplementary material), on the y-axis, we refer to percent decrement in RMSE and percent increment in CC due to the proposed method over the other baseline methods. We note poor performance of RMSE for the 24th subject owing to smaller # informative EEG trials. Observations based on the RMSE and CC performance comparison, depicted in Figs. 5 and 6 (from the supplementary material), indicate significant improvements on training sets with fewer data points. Furthermore, an intersubject statistical analysis is demonstrated to examine different hypotheses on RMSE and CC. A two-way analysis of variance (ANOVA) is conducted for diverse kinds of regression approaches with the

TABLE IV
RESULTS COMPARISON BOTH WITHIN AND ACROSS NN METHODS. (a) p-VALUES OF TWO-WAY ANOVA : PROPOSED LGN VERSUS BP, ADG, AND HJB FOR $\{FS4; FS2\}$ (b) p-VALUES OF TWO-WAY ANOVA: PROPOSED LGN VERSUS SVR, RR, AND LASSO FOR $\{FS4; FS2\}$

(a)

	FS4		FS2	
	RMSE	CC	RMSE	CC
p-value	0.04	**0.001**	0.03	**0.001**

(b)

	FS4		FS2	
	RMSE	CC	RMSE	CC
p-value	0.025	**0.001**	0.02	**0.002**

goal to find out if the RMSE and CC deviations due to the discrepancies in regression methods used are statistically significant for features FS2 and FS4, while treating the subjects as a random factor. The results are demonstrated in Table IV(a), (p-value$<$0.05) which indicates that there exist statistically significant differences in RMSEs, and CCs for various regression models for features $\{FS4, FS2\}$ (Section IV-B). In other words, regression method selection creates a significant effect on the performance metrics RMSE and CC [p-value$<$0.05, cf. Table IV(a)].

Then, *post-hoc* nonparametric multiple comparison tests (paired t-tests in this case) are applied to find out if the difference between pairs of regressors is statistically significant, with the p-value corrected employing the false discovery rate (FDR) method [52]. The p-values are shown in Table II(a), where in most of the values point out probabilistic relevance. The proposed LGN method performed superior to the HJB, ADG and BP-based NN schemes for the chosen EEG drowsiness problem. Some of the advantages offered in particular include the nature of the update law which is iterative both in terms of the control input and current weights, unlike BP and HJB methods. It is also not dependent on the closed form expression for control input unlike the other update laws. The proposed method (LGN) works to optimize the integral of a functional over a curve using calculus of variations which leads to the Euler–Lagrange equations and corresponding control law for weight update. An alternative method explored in our prior work was based on Bellman's optimality principle, which leads to the HJB equations which when solved leads to optimal control law. Based on the literature,[1] each of these approaches offer advantages and disadvantages depending on the application, with numerous technical differences between them, but in the case for the regression problem when both are applicable, the LGN method demonstrated superior performance.

F. Comparison of NN Lagrangian (LGN) Regression With Other Regression Models

In order to describe the dynamic learning ability of the NN-based method, we contrast our LGN-based NN approach with the existing models, such as LASSO Regression, ridge regression (RR), and support vector regression (SVR) [53]. We have employed a Scikit-Learn SVR tool [54] and Scikit-Learn jointly to spurt the SVR models. We make use of grid-search to determine parameters C and γ, which are optimally obtained to be 2^5 and 2^2, with $\epsilon = 0.2$, respectively. For the cases of LASSO and RR, the adjustable parameter λ was selected by an inner 8-fold cross-validation [55] on the training dataset. On average, from Table II(b), the proposed LGN-based DNN approach clearly outperformed SVR, LASSO, and RR schemes in terms of RMSE and CC. The DNN is trained for 200 epochs with a learning rate of 0.1 and R is also fixed to identity while using Lagrangian-based update law. The respective percentage performance improvements of LGN over the other regression models are shown in Figs. 7 and 8 in the supplementary material. For instance, the terms in legend "LGN/SVR" represent the improvements with LGN over the SVR method. The notation of other terms in the legend is to be understood in a similar manner. On average, LGN recorded a 15.45% smaller RMSE and a 5.58% larger CC than SVR. On average, LGN had performed with a 22.35% smaller RMSE and a 10.71% larger CC than LASSO. Also, LGN had performed with a 22.69% smaller RMSE and a 12.28% larger CC than RR. We notice a drop in RMSE performance for subjects # 15, 22, and 23. This is again attributed to a limitation of designing subject adaptive DNN learning models with sufficiently elaborate neural architecture search and hyperparameter tuning. The reason for this is that the deep NN is not able to generalize effectively for certain subjects, which is either due to a mismatch between the number of parameters in the NN model and the volume of EEG trials collected for training for each subject. This will be dealt with separately in future work. For example, readers are referred to the paper on EEGNET [56], where authors also encountered this issue which they tried to deal with using separable and depthwise convolutions.

% CC performance is improved for LGN over all the baseline models (HJB, BP, SVR, RR, and LASSO). This is due to the effective SSO-FIPLV feature correlation with actual RT values. The proposed LGN method in certain situations although does not predict RT values accurately, its trajectory over accumulated time overlaps with actual RT values (referred to as 'matching trends'). This saturation at times leads to poor RMSE. Performance across subjects # 2, 7, 9, and 15 needs to be interpreted further while amassing more EEG sessions from these subjects, so that the deep network can be fine-tuned adequately for deducing better conclusions. Furthermore, an intersubject Statistical analysis is conducted to verify various hypotheses on RMSE and CC's. A two-way ANOVA is demonstrated for various kinds of learning-based regression approaches to find out if the RMSE and CC digressions due to the differences in learning-based regressors are statistically significant when subjects are considered as a random factor. The results presented in Table IV(a) (p-value$<$0.05) shows that a statistically significant difference exists in RMSEs, and CCs for different learning-based regression models when subjects are considered as a random factor. In other words,

[1] https://cutt.ly/mmY0tgc

performance metrics RMSE and CC depend on the selection of regression approach [p-value<0.05, cf. Table IV(a)]. Then, *post-hoc* nonparametric multiple comparison tests (paired t-tests in this case) are applied to find out if the difference between pairs of regressors is statistically significant, with the p-value corrected employing the FDR method [52]. The p-values are shown in Table III(a), where most of the values are statistically significant. Effect sizes for $p < 0.2$ are reported as partial eta squared, whose values can be benchmarked against Cohen's criteria [57] of small (0.01), medium (0.06), and large (0.14) effects, according to Richardson [58]. The p-values associated with effect sizes > 0.13 are bolded.

In the last phase of this work, we integrate the whole pipeline for regression, i.e., SSO-FIPLV features with LGN method for DNN regression. On average, this pipeline recorded the smallest RMSE and largest CC of all the DNN regression pipelines [cf. Table III(b)]. For all the above experiments, a standard DNN architecture is used which has been validated in a previous work by us [28]. In addition, we present a performance comparison of various combination of features and regression blocks in the Section 7 of the supplementary material. We obtain best performance for the combination of SSO-FIPLV features with LGN method for DNN regression. Layerwise relevance propagation (LRP) [59] explains for each input pattern, its importance in the network output decision. LRP analysis has been done for the NN trained with the proposed method and reported in Section 6 of the supplementary material (Supplementary.pdf). Channelwise feature importance in terms of average relevance values calculated from NN feature representations are visualized in the form of Topoplots. The drowsiness prediction system in Lin *et al.* [50] can estimate the driver RT in RMSE error to be 0.076 s on an average. In other words, it means that the EEDD is around 2 m under the constant speed 100 km/h. The average RMSE error in this manuscript is 0.00685 s translating to an EEDD of 1.708 m. This is better than that of Lin *et al.* [50].

V. Conclusion

In this work, we have explored the stability, convergence, and regression performance of the proposed LGN-based update law. Although, the proposed feature extraction method (SSO-FIPLV) is applied to EEG RT prediction task, the method is generalizable to various EEG Regression problems like depth decryption of cognitive processing [60], single-trial motor performance prediction [61], continuous estimation of movement path [62], etc. Based on the analysis of results, future work should include: the expansion of EEG trials to be collected for specific subject #'s (2, 7, 9, and 15) to be able to use deeper models for learning. Adaptation of the order of fuzzy time-delay approach across subjects is another important future work. In addition, the inclusion of regularization within the JAD framework for enhancing the scope of spatial filters generalization and second, transfer learning inclusion into the FTDCSSP framework to enhance generalization between subjects and within subjects across sessions are additional future works. A major development to be addressed in future work is to incorporate parallelism and implement the LGN update law for both convolutional and recurrent networks. Also, in addition to alpha and theta power features, gamma power features will be explored for drowsiness detection in future work. In the current experimental paradigm, as per the current design of road conditions, the response time for emergency stops has not been considered. In addition, driving at night is generally affected by the light on the road. These two concerns are limitations of the current study and will be taken care of in future work.

Acknowledgment

The authors are grateful to Prof. Dongrui Wu, Dr. Y. K. Wang, and anonymous reviewers for their valuable suggestions and guidance.

References

[1] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor applications and physiological features in drivers' drowsiness detection: A review," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3055–3067, Apr. 2018.

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

[3] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7169–7180, Dec. 2015.

[4] T. K. Reddy, V. Arora, L. Behera, Y.-K. Wang, and C.-T. Lin, "Fuzzy divergence based analysis for EEG drowsiness detection brain computer interfaces," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, 2020, pp. 1–7.

[5] T. K. Reddy, Y.-K. Wang, J. Andreu-Perez, and C.-T. Lin, "Joint approximate diagonalization divergence based scheme for EEG drowsiness detection brain computer interfaces," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, 2021, pp. 1–6.

[6] T. K. Reddy and L. Behera, "Driver drowsiness detection using intelligent BCI," submitted for publication.

[7] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[8] S. Kar, M. Bhagat, and A. Routray, "EEG signal analysis for the assessment and quantification of driver's fatigue," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 13, no. 5, pp. 297–306, 2010.

[9] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," *Brain Res. Rev.*, vol. 29, nos. 2–3, pp. 169–195, 1999.

[10] C.-T. Lin, C.-J. Chang, B.-S. Lin, S.-H. Hung, C.-F. Chao, and I.-J. Wang, "A real-time wireless brain–computer interface system for drowsiness detection," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 4, pp. 214–222, Aug. 2010.

[11] A. Craig, Y. Tran, N. Wijesuriya, and H. Nguyen, "Regional brain wave activity changes associated with fatigue," *Psychophysiology*, vol. 49, no. 4, pp. 574–582, 2012.

[12] A. Kandaswamy, V. Krishnaveni, S. Jayaraman, N. Malmurugan, and K. Ramadoss, "Removal of ocular artifacts from EEG–A survey," *IETE J. Res.*, vol. 51, no. 2, pp. 121–130, 2005.

[13] J. Sun, X. Hong, and S. Tong, "Phase synchronization analysis of EEG signals: An evaluation based on surrogate tests," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2254–2263, Aug. 2012.

[14] S. L. Koole and W. Tschacher, "Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance," *Front. Psychol.*, vol. 7, p. 862, Jun. 2016.

[15] X. Liu *et al.*, "Inter-hemispheric frontal alpha synchronization of event-related potentials reflects memory-induced mental fatigue," *Neurosci. Lett.*, vol. 653, pp. 326–331, Jul. 2017.

[16] 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, Jan. 2017.

[17] J. Qu, R. Wang, Y. Du, and J. Cao, "Synchronization study in ring-like and grid-like neuronal networks," *Cogn. Neurodyn.*, vol. 6, no. 1, pp. 21–31, 2012.

[18] C.-T. Lin, Y.-K. Wang, and S.-A. Chen, "A hierarchal classifier for identifying independent components," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2012, pp. 1–5.

[19] 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, May 2014.

[20] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, p. 19, Apr. 2019.

[21] P. S. Baboukani, G. Azemi, B. Boashash, P. Colditz, and A. Omidvarnia, "A novel multivariate phase synchrony measure: Application to multichannel newborn EEG analysis," *Digit. Signal Process.*, vol. 84, pp. 59–68, Jan. 2019.

[22] L. Payne and J. Kounios, "Coherent oscillatory networks supporting short-term memory retention," *Brain Res.*, vol. 1247, pp. 126–132, Jan. 2009.

[23] P. Fries, "A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence," *Trends Cogn. Sci.*, vol. 9, no. 10, pp. 474–480, 2005.

[24] P. Celka, "Statistical analysis of the phase-locking value," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 577–580, Sep. 2007.

[25] 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," *Int. J. Bifurcation Chaos*, vol. 10, no. 10, pp. 2429–2439, 2000.

[26] L. Santamaria and C. James, "On the existence of phase-synchronised states during motor imagery tasks," *Biomed. Signal Process. Control*, vol. 54, Sep. 2019, Art. no. 101630.

[27] S. Kumar, T. K. Reddy, and L. Behera, "EEG based motor imagery classification using instantaneous phase difference sequence," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2018, pp. 499–504.

[28] T. K. Reddy, V. Arora, S. Kumar, L. Behera, Y. Wang, and C. T. Lin, "Electroencephalogram based reaction time prediction with differential phase synchrony representations using co-operative multi-task deep neural networks," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 5, pp. 369–379, Oct. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8846113?source=authoralert

[29] T. K. Reddy, V. Arora, L. Behera, Y.-K. 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)," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 10, pp. 1943–1951, Oct. 2019.

[30] 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 Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 771–781, Apr. 2018.

[31] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[32] T. K. Reddy and L. Behera, "'Online eye state recognition from EEG data using deep architectures,"' in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2016, pp. 712–717.

[33] T. K. Reddy, V. Arora, and L. Behera, "HJB-equation-based optimal learning scheme for neural networks with applications in brain–computer interface," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 2, pp. 159–170, Apr. 2020.

[34] L. He, D. Hu, M. Wan, Y. Wen, K. M. Von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.

[35] D. Coyle, G. Prasad, and T. M. McGinnity, "Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain–computer interface," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1458–1471, Dec. 2009.

[36] Y. Yang, Z. Gao, Y. Li, Q. Cai, N. Marwan, and J. Kurths, "A complex network-based broad learning system for detecting driver fatigue from EEG signals," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 9, pp. 5800–5808, Sep. 2021.

[37] D. P. Kingma and J. B. Adam, "A method for stochastic optimization," 2014. [Online]. Available: arXiv:1412.6980.

[38] J. Bilski and L. Rutkowski, "A fast training algorithm for neural networks," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 6, pp. 749–753, Jun. 1998.

[39] Y. Iiguni, H. Sakai, and H. Tokumaru, "A real-time learning algorithm for a multilayered neural network based on the extended Kalman filter," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 959–966, Apr. 1992.

[40] M. D. Robinson, M. T. Manry, S. S. Malalur, and C. Yu, "Properties of a batch training algorithm for feedforward networks," *Neural Process. Lett.*, vol. 45, no. 3, pp. 841–854, 2017.

[41] V. Arora, L. Behera, T. Reddy, and A. P. Yadav, "HJB equation based learning scheme for neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 2298–2305.

[42] Q. Wei, F.-Y. Wang, D. Liu, and X. Yang, "Finite-approximation-error-based discrete-time iterative adaptive dynamic programming," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2820–2833, Dec. 2014.

[43] H.-S. Choi *et al.*, "Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography," *IEEE Access*, vol. 7, pp. 146390–146402, 2019.

[44] V. Arora, L. Behera, T. K. Reddy, and A. P. Yadav, "HJB equation based learning scheme for neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 2298–2305.

[45] B. Chen, Y. Li, J. Dong, N. Lu, and J. Qin, "Common spatial patterns based on the quantized minimum error entropy criterion," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4557–4568, Nov. 2020.

[46] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[47] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[48] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Front. Neuroinformat.*, vol. 9, p. 16, Jun. 2015.

[49] N. Caramia, F. Lotte, and S. Ramat, "Optimizing spatial filter pairs for EEG classification based on phase-synchronization," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)*, 2014, pp. 2049–2053.

[50] F.-C. Lin, L.-W. Ko, C.-H. Chuang, T.-P. Su, and C.-T. Lin, "Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 9, pp. 2044–2055, Sep. 2012.

[51] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. New York, NY, USA: Springer, 2012, pp. 437–478.

[52] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Stat. Soc. B Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.

[53] 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 Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.

[54] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[55] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Jul. 2010.

[56] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[57] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York, NY, USA: Academic, 2013.

[58] J. T. Richardson, "ETA squared and partial ETA squared as measures of effect size in educational research," *Educ. Res. Rev.*, vol. 6, no. 2, pp. 135–147, 2011.

[59] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017.

[60] I.-E. Nicolae, L. Acqualagna, and B. Blankertz, "Assessing the depth of cognitive processing as the basis for potential user-state adaptation-data set," *Front. Neurosci.*, vol. 11, p. 548, Oct. 2017.

[61] 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," *Front. Human Neurosci.*, vol. 10, p. 170, Apr. 2016.

[62] A. Úbeda, J. M. Azorín, R. Chavarriaga, and J. d. R. Millán, "Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, p. 9, 2017.

Tharun Kumar Reddy received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 2013, and the M.Tech.–Ph.D. degree in electrical engineering from IIT Kanpur in 2020.

He is an Assistant Professor with the Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, India. During the Ph.D. degree, he spent time in a research collaboration with the Centre for Artificial Intelligence, University of Technology Sydney, Sydney, NSW, Australia, and Brain Research Center, NCTU, Hsinchu, Taiwan. He represented India internationally as part of the IITK-TCS team which participated in the Amazon Picking Challenge held at Leipzig, Germany. He was a Tata Consultancy Services Research Fellow. His field of research includes machine learning for signal processing, deep neural networks, EEG signal processing, and brain–computer interfaces. His website is available at the link https://tharuniitk.github.io/.

Vipul Arora received the B.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology Kanpur (IIT Kanpur), Kanpur, India, in 2009 and 2014, respectively.

He was a Research Scientist with Amazon Alexa Team, Boston, MA, USA. He is currently an Assistant Professor with the Department of Electrical Engineering, IIT Kanpur. His research interests include music information retrieval, acoustic space modeling, and semantic signal processing.

Vinay Gupta received the B.Sc. and M.Sc. degrees in mathematics from the Indian Institute of Technology Kanpur, Kanpur, India, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering.

His research interests constitute machine learning, neural networks, physics, and EEG analysis.

Rupam Biswas is currently pursuing the B.Tech.–M.Tech. dual degree in electrical engineering with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, India.

His research interests constitute brain–computer interfaces, machine learning, assistive robotics, and biomedical engineering.

Laxmidhar Behera (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in engineering from the National Institute of Technology Rourkela, Rourkela, India, in 1988 and 1990, respectively, and the Ph.D. degree from the Department of Electrical Engineering, Indian Institute of Technology (IIT) Delhi, New Delhi, India, in 1996.

He is currently a Poonam and Prabhu Goel Chair Professor (H.A.G.) with the Department of Electrical Engineering, IIT Kanpur, Kanpur, India. He is also a Guest Professor with Hangzhou Dianzi University, Hangzhou, China. He was an Assistant Professor with BITS Pilani, Pilani, India, from 1995 to 1999 and received the postdoctoral studies from the GMD-German National Research Center for Information Technology, Sankt Augustin, Germany, from 2000 to 2001. From 2008 to 2010, he was a Reader with the Intelligent Systems Research Center, University of Ulster, Coleraine, U.K. He was also a Visiting Researcher/Professor with Fraunhofer-Gesellschaft, Munich, Germany, and ETH Zurich, Zurich, Switzerland. He has more than 200 papers to his credit published in refereed journals and conference proceedings. His research interests include intelligent control, robotics, semantic signal/music processing, neural networks, control of cyber–physical systems, and cognitive modeling.

Dr. Behera is a Fellow of the Indian National Academy of Engineering.