JOINT APPROXIMATE DIAGONALIZATION DIVERGENCE BASED SCHEME FOR EEG DROWSINESS DETECTION BRAIN COMPUTER INTERFACES

Tharun Kumar Reddy[⋆] *Yu-kai Wang* [†] *Chin-Teng Lin*[†] *Javier Andreu-Perez*[‡]

[⋆] Department of Electronics and Communication Engineering, Indian Institute Of Technology Roorkee
[†]University of Technology Sydney, Ultimo, NSW 2007, Australia
[‡]School of Computer Science and Electronic Engineering, University of Essex

ABSTRACT

Neurons usually converse through electrochemical signals and pooled neuronal firings feasibly be recorded on the scalp through the medium of electroencephalogram (EEG). EEG waveforms are recorded, analysed and categorized across directives concerning a Brain-Computer Interface (BCI). Deteriorated signal to noise ratio and non-stationarities stand as a paramount obstacle in steady decoding of EEG. Appearance of non-stationarities across EEG patterns notably upset the feature waveforms thus worsening the functioning of detection block and as a whole the Brain Computer Interface. Stationary Subspace schemes bring to light subspaces within which data distribution persists stably over time. Current work focuses on the development of a novel spatial transform based feature extraction scheme to address nonstationarity in EEG signals recorded against a drowsiness detection problem (a machine learning regression scenario). The presented approach: F-DIV-IT-JAD-WS derived features distinctly surpassed DivOVR-FuzzyCSP-WS based standard features across RMSE and CC performance criteria pair. We construe that the propounded feature derivation approach based on F-DIV-IT-JAD-WS will usher a significant attention in researchers who are developing algorithms for signal processing, specifically, for BCI regression scenarios.

Index Terms— Brain Computer Interface (BCI), Electroencephalogram (EEG), Stationary Subspace Analysis (SSA), Reaction Time (RT) prediction, Divergence One versus Rest Fuzzy Common Spatial Patterns Within Session (DivOVR-FuzzyCSP-WS), Common Spatial Patterns (Fuzzy CSP), Drowsiness

1. INTRODUCTION

Electroencephalogram (EEG), comprises neuro-physiological patterns collected non-invasively from scalp while being deployed to transform and decrypt the directives of a brain-computer interface (BCI). It is one of the most economical brain-imaging techniques in comparison to FMRI, MEG. Owing to the reasons of possessing large temporal resolution and being economical, it offers huge scope to decrypt the commands in a Brain-Computer Interface (BCI). BCI offers for specific control instruction communication linking the brain and an external device via brain signal activity. There are three major paradigms within BCI namely, the Motor Imagery (MI) [1], Steady State Visual Evoked Potentials (SSVEP) and P300 [2]. MI refers to imagining bringing about activity concerning a part of human anatomy as contrary to directing true body motion in stimuli based EEG experiments. MI builds on the evidence that brain activation is subject to reduction and switch on corresponding brain regions albeit in reality mobilizing a body part. The common spatial pattern (CSP) method [3] is a potential spatial transform which explores to locate a discerning subspace escalating one class variance albeit curtailing the another concurrently to categorize the movement gestures. Motor Imagery (MI) (eg: decoding thoughts of imagination of left or right hands) is a prominent BCI paradigm implemented often to command BCIs. A BCI pipeline is composed of preprocessing (data-preparation), feature extraction and classification or regression blocks. Motor imagery influences a portion of the neurons, yet transitions in several of such tasks can transform the corresponding surface electrode composite brain data. Because of the existence of significant aggregate of stimulus-irrelevant outputs, EEG waveforms carry noisy and are nonstationary thereby making BCI communication and command decoding to be a challenging task. Data preparation module comprises spatial transforms stage which are operated to enhance the signal to noise ratio (SNR) and markedly bring out features for MI states. Common Spatial Patterns (CSP) [3] approach exists to be an extensively applied block to carry on spatial transformations in a two class MI. In general, BCIs are constrained via substantial alterations based on individual's state, one as well as the other within and across sessions in addition to noise and artefacts. The vanilla CSP scheme by-no-means adapted along the line of nonstationarities [4], also performed poorly due to the above alterations. Several approaches (both unsupervised [5] and supervised (simultaneous streamlining of discriminability and stationarity) [6]) are proposed to extract stationary parts for EEG MI classification. Other works in this direction comprise of [7], where in spatial transforms are optimized using divergence. In addition, robustified subspace transforms are dealt with in [8]. CSP remains primarily designed towards a binary class problem. Methods akin to One-Versus-Rest (OVR) CSP [9], coupled binary classification followed by selective picking [1], CSP accompanied by spatial vector picking [10] and non-Euclidean frameworks [11]were outlined for feature extraction and multiclass classification. Stationarity augmentation schemes mentioned in [7] have not been adapted for machine learning problems outside binary classes. Recently, in [12, 13], a sound theoretical framework had been proposed with commensurate validation conducted for multiclass MI. In [14], dispersion within each class was used as a measure for quantifying the incorporation of non-stationarity and pruning the insignificant channels to enhance the performance of BCI. Recently, in [15] an approach based on non-stationarity was developed for EEG regression problems. Sleeplessness instigating Driver debility has been tagged by US NHTSA to be essential cause of many highway calamities [16]. It is a problem of more severe intensity than that of driving with distraction or driving consuming alcohol. Henceforth, there is a prompt need to undertake and put in place solutions for driver security. Uncertainties induced by factors alike artifacts, nonstationarities reduce the accuracy of EEG-based BCI systems [17]. Fuzzy logic based models address these issues efficiently as reported in earlier works [18, 15]. In addition to BCI, the application domains of fuzzy logic to address uncertainties expands to several areas like traffic life cycles [19] and networked control systems [20]. In this research paper, we further deploy fuzzy modeling approach for BCI. The machine learning problem for driver-drowsiness detection is an EEG based Reaction-Time prediction task [15, 21]. The schemes mentioned above in the introduction are pertaining to EEG MI classification. Yet, the nonstationarities aligned along EEG streams stand unaddressed for regression problems. Methods developed so far are for ML classification scenarios like for binary MI. EEG assisted Reaction-Time (RT) forecasting [22, 21, 23] stands reported to be an EEG signal processing and machine learning regression scenario. Different adaptations devoted to CSP alike DivOVR-FuzzyCSP-WS [15] and FuzzyCSP [21, 24] comprise the existing spatial transform associated feature derivation skeletons for regression. FuzzyCSP stands reposed alike divergence problem through calculation of filters optimizing a divergence cost [25]. The prediction rate of regression is to a great extent affected through the medium of non-stationarities within attribute (feature) admeasurements as demonstrated in [26, 27]. Thinking along similar lines, we hereby introduce non-stationarities with considerable success within the EEG regression.

In this paper, we demonstrate a bunch of contributions as notified underneath: (1) We broaden the divergence based study [7] for regression. (2) We came up with a new stationarity inclusive spatial transformation paradigm for feature computation within EEG regression. Established as such, we put forward F-Div-IT-JAD-WS, a stationarity encompassing spatial transform which is equally valid for two and more fuzzy classes. (3) The optimal transform so obtained retains stationarity within session as well as enhances the forecasting ability of regression.

This work is further organized as follows. Section 2 examines the EEG persistent attention task while duplicating actual drowsiness situation. Section 3 introduces a fundamental conceptualization comprising divergence stationed scheme for regression. Section 4 puts together divergence framework for combined streamlining of stationarity and forecasting power of model. Section 5 provides implementation specifics and a short analysis. Section 6 concludes the present paper and presents prospective directives.

2. DATA OF DRIVING SESSIONS

We elaborate driver demeanor and brain dynamics obtained from a 90-minute persistent-attention task (PAT) performed in a driving simulator. EEG signal trials with complementary trial-wise reaction time values from 11 subjects are listed in an individual session. The database has been already tested in several works [21, 24] as a benchmark for EEG regression. Brain signal data comprises time-series segments $\mathbf{X} \in \mathbb{R}^{C \times T}$ ((C, T) indicate total signal electrodes and Time samples respectively) with its complementary reaction time data (Y). Data description and experimental details can be found in section 1 of [28].

3. PROPOSED METHOD

In [7], authors proposed a KL divergence characterization concerning CSP for MI classification. For EEG machine learning regression, as presented in [15, 21, 24], Neural Networks [22] and LASSO [24] algorithms existed and implemented on characterizing attributes extracted from fuzzy CSP [24]. The frameworks proposed here for continuous variable prediction are analogous to well known classification problems and are incumbent against the EEG inference in the role of a time series. Grosse-Wentrup et al. [10] came up with Joint Approximate Diagonalization (JAD) framework coupled via information theoretic spatial transform selection guidelines to choose spatial transforms for classification.

3.1. Joint Approximate Diagonalization (JAD) and its Information theoretic interpretation (Motor Imagery for Multiple Classes)

Spatial filtering by Joint approximate diagonalization (JAD) scheme is a prevailing alternative to One Versus Rest- Common Spatial Patterns (OVR-CSP) (cf. [29]) for classification of MI with multiple classes. Assuming an EEG data of K different classes, CSP by JAD computes a spatial transform $\mathbf{Z} \in \mathbb{R}^{K \times K}$ that diagonalizes the individual class covariance matrices (cf. section 2 equation (4) of [12]). Further, an Information theoretic interpretation of the JAD framework for multiclass motor imagery can be found in the work of [12] (cf. section 2 equation (5)-(9)).

3.2. Fuzzy Divergence Information Theoretic Joint Approximate Diagonalization (F-DIV-IT-JAD)

Here, in this section, we propose Fuzzy Divergence Information Theoretic Joint Approximate Diagonalization for regression. We come up with a new objective function for regression using fuzzy covariance matrices.

If K denotes the number of fuzzy classes, assume $K \geq 2$. The important suppositions comprise: the conditional probability of every fuzzyclass is Normal distribution i.e. $\mathcal{N}(\mathbf{0}, \overline{\mathbf{\Sigma}}_1) \cdots \mathcal{N}(\mathbf{0}, \overline{\mathbf{\Sigma}}_K)$ considering all K fuzzy classes ($\overline{\mathbf{\Sigma}}_1 \cdots \overline{\mathbf{\Sigma}}_K$ are the fuzzy class covariance matrices).

We describe the subsequent preliminaries of normalized fuzzy covariance and KL divergences in the section 2 of the supplementary [28]. Readers are suggested to go through it before going forward in this formulation.

Gouy Paeiller et al.[30] has put forward an Information Theoretic (IT) explanation about JAD. As per the JAD scheme, we compute the matrix transform that together decomposes (via diagonalization) the class covariance matrices. This has been dealt with through minimization of the KL divergences across the pair: altered covariances (modified through transformation) with a corresponding diagonalized version.

If $\mathbf{X}$ is a matrix and $\mathbf{\Sigma}$ is a diagonal matrix, then by Pythagorean split one obtains:

$$D_{kl}(\mathbf{X} \parallel \Sigma) = D_{kl}(\mathbf{X} \parallel diag(\mathbf{X})) + D_{kl}(diag(\mathbf{X}) \parallel \Sigma) \tag{1}$$

Here, $diag(\mathbf{X})$ denotes a matrix whose diagonal values are equal to diagonal values of $\mathbf{X}$. To optimize $D_{kl}(\mathbf{X} \parallel \mathbf{\Sigma})$, the term $\mathbf{\Sigma}$ is to be same as $diag(\mathbf{X})$ (because KL divergence is greater than zero). Thus, retrieving the diagonal matrix $\mathbf{\Sigma}$ as $diag(\mathbf{X})$. The mathematical framework of IT-JAD is indicated below as

$$F(\mathbf{Z}) = \sum_{c=1}^{K} \mu_c D_{kl}(\mathbf{Z}^\intercal \overline{\mathbf{\Sigma_c}} \mathbf{Z} \parallel diag(\mathbf{Z}^\intercal \overline{\mathbf{\Sigma_c}} \mathbf{Z})) \tag{2}$$

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} F(\mathbf{Z}) \tag{3}$$

where K is # fuzzy classes (# different transforms combinedly decomposed using diagonalization). μ_c presents the fuzzy membership of belongingness of EEG trials to class c. The transform $\mathbf{Z}$ stands split into a multiplication of an orthogonal matrix and a whitening transform $\mathbf{Z}^\intercal = \mathbf{RW}$ [31]. We can edit (2) in terms of orthogonal transform $\mathbf{R}$ as indicated here:

$$J(\mathbf{R}) = \sum_{c=1}^{K} \mu_c D_{kl}(\mathbf{R}\mathbf{\Sigma_c}'' \mathbf{R}^\intercal \parallel diag(\mathbf{R}\mathbf{\Sigma_c}'' \mathbf{R}^\intercal)) \tag{4}$$

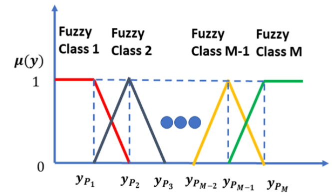

Fig. 1. Fuzzy membership distribution of output Reaction Time

$$\text{such that} \quad \mathbf{\Sigma_c}'' = \mathbf{W}\overline{\mathbf{\Sigma_c}}\mathbf{W}^\intercal$$
$$\sum_{c=1}^{K} \mathbf{W}\overline{\mathbf{\Sigma_c}}\mathbf{W}^\intercal = \mathbf{I} \tag{5}$$

The objective is to minimize $J(\mathbf{R})$ so that $\mathbf{R}\mathbf{\Sigma_c}''\mathbf{R}^\intercal$ is close to a diagonal representation.

Examine $\tilde{J}_1(\mathbf{R})$ which is the cardinal term in (2) and μ_1 be some constant.

$$\begin{aligned}
\tilde{J}_1(\mathbf{R}) &= D_{kl}(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal \parallel diag(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal)) \\
&= (log(det(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal)^{-1} diag(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal))) \\
&+ \underbrace{tr(diag(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal)^{-1}\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal)}_{\text{constant}} \\
&= -log(det(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal) + log(det(diag(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal))) + C
\end{aligned}$$

The derivative of $\tilde{J}_1(\mathbf{R})$ w.r.t square matrix R as it may be written from [32] as

$$\nabla_{\mathbf{R}}\tilde{J}_1(\mathbf{R}) = -2\mathbf{R}^{-1} + 2diag(\mathbf{R}\mathbf{\Sigma_1}''\mathbf{R}^\intercal)^{-1}\mathbf{R}\mathbf{\Sigma_1}'' \tag{6}$$

The gradient value in (6) could be reproduced to calculate the derivative of (4). The unified gradient can be further applied to optimize $\mathbf{R}$ on an orthogonal manifold. After $J(\mathbf{R})$ is optimized (convergence norms fulfilled), one can compute the spatial filters with $\mathbf{Z}^\intercal = \mathbf{RW}$. Further, the spatial transforms are sorted (i.e. column vectors of $\mathbf{Z}$) based on mutual information filter pruning condition [10]. We selected best $2 * K$ filters after the sorting, which are further used for Feature Extraction for regression.

4. REGRESSION FORTIFYING STATIONARITY

In previous works (for example: [7]), regularization terms are integrated into the objective for optimizing stationarity of EEG signal within session. In this section, we include stationarity into the JAD objective to deal with within session non-stationarities, while stationarity is optimized across every fuzzy class (appropriate definition of $G(\mathbf{Z})$). We work with an EEG based driving scenario while analyzing stationarity within session for every subject.

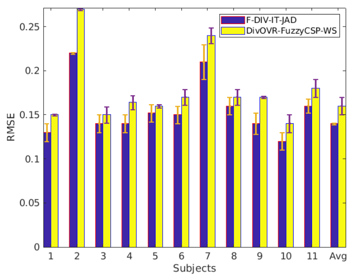

Fig. 2. RMSE of F-DIV-IT-JAD on 11 subjects.

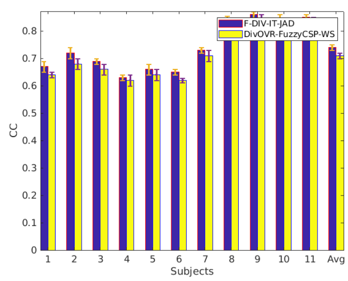

Fig. 3. CC of F-DIV-IT-JAD on 11 subjects.

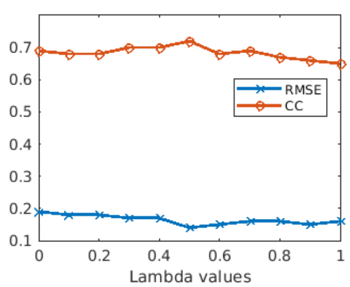

Fig. 4. RMSE and CC (both in seconds) concerning F-DIV-IT-JAD on 11 subjects (λ is a hyper-parameter).

$$G(\mathbf{Z}) = \lambda\Bigg(\frac{1}{KN}\sum_{c=1}^{K}\sum_{i=1}^{N}\mu_{i,c}D_{kl}(\mathbf{Z}^{\intercal}\overline{\mathbf{\Sigma_{i,c}}}\mathbf{Z} \parallel \mathbf{Z}^{\intercal}\overline{\mathbf{\Sigma_c}}\mathbf{Z})\Bigg) \tag{7}$$

Here, N presents # trials per individual fuzzy class. In (7), $\overline{\mathbf{\Sigma_{j,c}}}$ and $\overline{\mathbf{\Sigma_c}}$ indicates the trialwise and classwise covariances.

By means of this, we frame a combined objective cooperatively enhancing the pair prediction and stationarity.

$$\delta(\mathbf{Z}) = (1-\lambda)F(\mathbf{Z}) + \lambda G(\mathbf{Z}) \tag{8}$$

Here, λ prevails as a regularizing scaler. The optimization approach followed comprises subspace scheme with gradient descent against an orthogonal spatial structure (cf. algorithm 1 page 5 of [7]). In (8), the regularization term is added, as long as one is minimizing the pair, the prediction alongside the group stationarity. Optimal filter $\mathbf{Z}^*$ is obtained through as:

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} \delta(\mathbf{Z}) \tag{9}$$

We hereby propose an enhanced subspace approach for this optimization problem.

$$\mathbf{R}^* = \arg\min_{\mathbf{R}} (1-\lambda)F(\mathbf{R}) + \lambda G(\mathbf{I_d R}) \tag{10}$$

In (10), entire subspace has been retreaded for the combined approximate diagonalization description, nevertheless, optimizing stationarity requires selection of appropriate subspace by premultiplying $\mathbf{R}$ with $\mathbf{I_d R}$. In essence, we have obtained joint diagonalization objective together with stationarity invoking objective against the preliminary d elements peculiar to the filter. Above method is denoted as 'F-DIV-IT-JAD-WS' or plainly 'F-DIV-IT-JAD'. 'WS' stands for Within Session.

5. ANALYSIS RESULTS AND DISCUSSION

We apply 8-fold cross-validation to evaluate the prediction performance of the feature sets pertaining to different spatial filtering schemes $F-DIV-IT-JAD-WS$ and $DivOVR-FuzzyCSP-WS$ (baseline method introduced in [15]) respectively.

Using the spatial transform $\mathbf{Z}$ computed through several schemes of spatial filtering undertaken in the preceeding sections, the transformed EEG segment is derived:

$$\mathbf{X}^{''} = \mathbf{Z}\mathbf{X}^{k}. \tag{11}$$

In this place, $X^{''}{}_i$ represents i^{th} consecution pertaining to $\mathbf{X}^{''}$ $\mathbf{X}^k$ is the EEG data subsequent to transitioning by PREP [24]. Spatial transforms comprise the rows of matrix $\mathbf{Z}$, through which we compute the eventual features, $\mathbf{F}$= $\begin{bmatrix} F_1 \\ .. \\ F_{FM} \end{bmatrix}$ where each F_i is given by $\log_{10}\frac{\|X^{''}{}_i\|^2}{\sum_{j=1}^{FM}\|X^{''}{}_j\|^2}$ Integrating F-DIV-IT-JAD and DivOVR-FuzzyCSP-WS generated feature vector $\mathbf{F}$ through LASSO block resulted in the mean RMSE and CC presented in illustrations 2 and 3. Moreover, features enumerated from $DivOVR-FuzzyCSP-WS$ baseline and flowed as a consequence of LASSO regression directing to baseline mean performance values in illustrations 2 and 3.

Illustrations 2 and 3 indicate in particular $F-DIV-IT-JAD$ distinctly surmounted $DivOVR-FuzzyCSP-WS$ doubly with respect to mean value of performance measures.

Further, we indicate the utility pertaining to change in regularization parameter λ against performance measures (cf. plot 4). Illustrations 2 and 3 indicate that novel $F-DIV-IT-JAD$ scheme attained smaller RMSE and largest CC in contrast with standard $DivOVR-FuzzyCSP-WS$ [15].

The outputs are 0.14 and 0.74 seconds respectively. Detection pipeline presented so far can calculate the driver reaction time by an average RMSE of 0.14 seconds. By way of explanation, it is indicated that the deviation in the predicted driving range stands to be 3.9 m subject to a rate of 100 kmph. Also, $\lambda = 0.5$ directed us to the optima of mean performance measures.

Finally, Unpaired t-test disclosed probabilistically relevant ($p < 0.01$) performance for $F - DIV - IT - JAD$ over $DivOVR - FuzzyCSP - WS$.

6. CONCLUSION AND FUTURE WORK

Current manuscript contains description of a novel divergence based Joint Approximate Diagonalization for regression. A novel loss function for regression has been formulated using divergence, JAD and fuzzy covariances. Later, we also proposed a corresponding stationarity objective as a novel term. Finally, a conjoint optimization is conducted to generate optimal spatial filters.

In future, we shall look into the objectives of generalizing across session and across subjects conjointly with our proposed approach.

7. REFERENCES

[1] Johannes Müller-Gerking, Gert Pfurtscheller, and Henrik Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[2] Tharun Kumar Reddy, Rupam Biswas, Vipul Arora, Vinay Gupta, and Laxmidhar Behera, "Concealed Information Detection in P300 Using a Novel Semi-supervised Higher-Order Gated Recurrent Neural Networks," submitted to IEEE Transactions on Neural Networks and Learning Systems, 2020.

[3] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2007.

[4] Wojciech Wojcikiewicz, Carmen Vidaurre, and Motoaki Kawanabe, "Stationary common spatial patterns: towards robust classification of non-stationary eeg signals," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 577–580.

[5] Paul Von Bünau, Frank C Meinecke, Franz C Király, and Klaus-Robert Müller, "Finding stationary subspaces in multivariate time series," *Physical review letters*, vol. 103, no. 21, pp. 214101, 2009.

[6] Wojciech Samek, Klaus-Robert Müller, Motoaki Kawanabe, and Carmen Vidaurre, "Brain-computer interfacing in discriminative and stationary subspaces," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 2873–2876.

[7] Wojciech Samek, Motoaki Kawanabe, and Klaus-Robert Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.

[8] Wojciech Samek and Klaus-Robert Müller, "Tackling noise, artifacts and nonstationarity in BCI with robust divergences," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 2741–2745.

[9] Benjamin Blankertz, Guido Dornhege, Christin Schafer, Roman Krepki, Jens Kohlmorgen, K-R Muller, Volker Kunzmann, Florian Losch, and Gabriel Curio, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 127–131, 2003.

[10] Moritz Grosse-Wentrup and Martin Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.

[11] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[12] Satyam Kumar, Tharun Kumar Reddy, Vipul Arora, and Laxmidhar Behera, "Formulating divergence framework for multiclass motor imagery eeg brain computer interface," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1344–1348.

[13] Satyam Kumar, Tharun Kumar Reddy, and Laxmidhar Behera, "Divergence Framework for EEG based Multiclass Motor Imagery Brain Computer Interface," *arXiv preprint arXiv:1901.07457*, 2019.

[14] Khadijeh Sadatnejad, Aline Roc, Léa Pillette, Thibaut Monseigne, Fabien Lotte, et al., "Channel selection over riemannian manifold with non-stationarity consideration for brain-computer interface applications," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1364–1368.

[15] Tharun Kumar Reddy, Vipul Arora, Laxmidhar Behera, Yu-kai Wang, and Chin-Teng Lin, "Fuzzy divergence based analysis for eeg drowsiness detection brain computer interfaces," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2020, pp. 1–7.

[16] Katherine Hartman, Jennifer Strasser, and Cambridge Systematics, "Saving lives through advanced vehicle safety technology: intelligent vehicle initiative," Tech. Rep., United States. Federal Highway Administration, 2005.

[17] A. Ciaramella and P. Salma, "A neuro-fuzzy based approach for resting-state detection using a consumer-grade eeg," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–4.

[18] Yu-Ting Liu, Yang-Yin Lin, Shang-Lin Wu, Chun-Hsiang Chuang, and Chin-Teng 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, 2015.

[19] RP Alvarez Gil, Zsolt Csaba Johanyák, and Tamás Kovács, "Surrogate model based optimization of traffic lights cycles and green period ratios using microscopic simulation and fuzzy rule interpolation," *Int. J. Artif. Intell*, vol. 16, no. 1, pp. 20–40, 2018.

[20] Qing Lu, Peng Shi, Hak-Keung Lam, and Yuxin Zhao, "Interval type-2 fuzzy model predictive control of nonlinear networked control systems," *IEEE Transactions on Fuzzy systems*, vol. 23, no. 6, pp. 2317–2328, 2015.

[21] Tharun Kumar Reddy, Vipul Arora, Laxmidhar Behera, Yukai Wang, and CT 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 Transactions on Fuzzy Systems*, 2019.

[22] Tharun Kumar Reddy, Vipul Arora, Satyam Kumar, Laxmidhar Behera, Yu-Kai Wang, and Chin-Teng Lin, "Electroencephalogram based reaction time prediction with differential phase synchrony representations using co-operative multi-task deep neural networks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 5, pp. 369–379, 2019.

[23] Tharun Kumar Reddy, Rupam Biswas, Vipul Arora, Vinay Gupta, and Laxmidhar Behera, "EEG Reaction Time-prediction with fuzzy Independent phase locking value representations using Lagrangian based Deep Neural Networks," submitted to IEEE Transactions on Systems, Man and Cybernetics, 2020.

[24] Dongrui Wu, Jung-Tai King, Chun-Hsiang Chuang, Chin-Teng Lin, and Tzyy-Ping Jung, "Spatial filtering for EEG-based regression problems in brain–computer interface (BCI)," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 771–781, 2017.

[25] Wojciech Samek, Duncan Blythe, Klaus-Robert Müller, and Motoaki Kawanabe, "Robust spatial filtering with beta divergence," in *Advances in Neural Information Processing Systems*, 2013, pp. 1007–1015.

[26] Chun-Shu Wei, Yuan-Pin Lin, Yu-Te Wang, Chin-Teng Lin, and Tzyy-Ping Jung, "A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection," *NeuroImage*, vol. 174, pp. 407–419, 2018.

[27] Tharun Kumar Reddy, Vipul Arora, and Laxmidhar Behera, "HJB-equation-based optimal learning scheme for neural networks with applications in brain–computer interface," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018.

[28] "Overleaf URL supplementary material ieeefuzz 2021," https://www.overleaf.com/read/gzbqchrrhpvq, Accessed: 2021-05-23.

[29] Wei Wu, Xiaorong Gao, and Shangkai Gao, "One-versus-the-rest (ovr) algorithm: An extension of common spatial patterns (csp) algorithm to multi-class case," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 2387–2390.

[30] Cédric Gouy-Pailler, Marco Congedo, Clemens Brunner, Christian Jutten, and Gert Pfurtscheller, "Nonstationary brain source separation for multiclass motor imagery," *IEEE transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 469–478, 2009.

[31] Wojciech Samek, "On robust spatial filtering of EEG in nonstationary environments," *it-Information Technology*, vol. 58, no. 3, pp. 150–154, 2016.

[32] KB Petersen and MS Pedersen, "The matrix cookbook. technical university of denmark," *Technical Manual*, 2008.