TOWARDS ADAPTIVE BRAIN-COMPUTER INTERFACES: STATISTICAL INFERENCE FOR MENTAL STATE RECOGNITION

by

MASTANEH TORKAMANI AZAR

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Sabancı University

August 2020

TOWARDS ADAPTIVE BRAIN-COMPUTER INTERFACES:
STATISTICAL INFERENCE FOR MENTAL STATE RECOGNITION

APPROVED BY

Assoc. Prof. Müjdat ÇETİN ....
(Dissertation Supervisor)

Prof. Selim BALCISOY ....
(Dissertation Co-supervisor)

Assoc. Prof. Serap AYDIN..

Asst. Prof. Sinan YILDIRIM ....

Assoc. Prof. Kemal KILIÇ ....

Assoc. Prof. Devrim ÜNAY ....

DATE OF APPROVAL: ......11 August 2020..............................

ABSTRACT

TOWARDS ADAPTIVE BRAIN-COMPUTER INTERFACES:
STATISTICAL INFERENCE FOR MENTAL STATE RECOGNITION

MASTANEH TORKAMANI AZAR

Electronics Engineering, Ph.D. Dissertation, August 2020

Dissertation Supervisor: Assoc. Prof. Mujdat CETIN

Dissertation Co-Supervisor: Prof. Selim BALCISOY

Keywords: Brain-computer interfaces, adaptive systems, electroencephalography, sensorimotor rhythms, motor imagery, spatio-spectral features, phase connectivity, mental state recognition, cognition, sustained attention, vigilance, SART, statistical signal processing, statistical inference, deep learning, convolutional neural networks, Bayesian models, changepoint detection.

Brain-computer interface (BCI) systems aim to establish direct communication channels between the brain and external devices. The primary motivation is to enable patients with limited or no muscular control, including amyotrophic lateral sclerosis (ALS) and stroke patients, to use computers or other devices by automatically interpreting their intent based on the measured brain electrical activity. Furthermore, enabling healthy individuals to use BCI systems as an additional communication channel in certain human computer interaction systems is also a current topic of interest.

Current experimental BCI systems are trained in a supervised fashion and then evaluated during test sessions. With increasing demands for daily and long-term use of BCIs in real-life applications such as in semi-autonomous cars, BCIs have been tested on longer sessions in which researchers have observed considerably lower performance of trained systems. This is believed to be caused by the nonstationary nature of the electroencephalographic (EEG) signals. As a result, semi-supervised adaptation of BCI systems based on test data has emerged as a new research domain. One of the main reasons underlying the nonstationarity of signals involves changes in the users' cognitive states such as the cognitive load, alertness, attention, fatigue, boredom, and motivation. However, dynamically extracting information about such cognitive states from EEG signals and using that to improve the performance of BCI systems is currently an open research problem.

In this thesis, we tackle the highly complex problem of estimating the level of alertness and vigilance of users during execution of cognitive tasks. To identify the neural, EEG-based correlates of long-term task and response time consistency, we devise a series of experiments running the sustained attention to response task (SART). After proposing

a novel adaptive scoring scheme for vigilance, we provide new evidence on the close relationship between intrinsic resting and task-related brain networks and develop models to predict consistency in tonic performance and response time using neural networks and feature relevance analysis from spatio-spectral features of resting-state EEG signals. Next, focusing on the imminent goal of predicting low and high vigilance intervals, we propose fully automated systems based on convolutional neural networks (CNNs) using phase locking value features as successful pre-trial predictors of phasic vigilance and performance consistency. In all of these contributions, we consider the personal vigilance traits and individual psychophysiological differences for modeling and detecting the extremely alert and drowsy trials in long and monotonous experiments, and enrich the literature with the evidence on spatio-spectro-temporal correlates of vigilant and consistent behavior.

We then utilize Bayesian changepoint models for sequential inference and detection of instants at which continuous vigilance levels of users enter a new phase. We demonstrate the success of our online and offline vigilance models in detecting changepoints from both the SART datasets collected in our lab and driving datasets that contain vigilance labels. Finally and as the highlight of this thesis, we hypothesize that the underlying vigilance levels affect users' reaction time and thus the ability to focus and engage in motor imagery BCI paradigms. We then introduce an adaptive alertness-aware MI classification system for motor imagery BCI that uses a series of novel unsupervised learning schemes for labeling trial vigilance levels during training and test sessions, and leads to a method with full adaptation in both feature extraction and training of its classifier parameters. Three different versions of this adaptive classification approach are introduced that are trained differently on trials labeled with low vigilance levels by our various vigilance clustering schemes. We report improvements in the overall test accuracy of adaptive versions with respect to the original, non-adaptive baseline for our own SPIS MI-BCI dataset and the BCI Competition IV Dataset 2a. A number of datasets collected in our BCI laboratory are uploaded to a public repository at https://github.com/mastaneht.

ÖZET

UYARLANABILIR BEYIN-BILGISAYAR ARAYÜZLERINE DOĞRU: ZIHINSEL DURUM TANIMA IÇIN İSTATISTIKSEL ÇIKARIM

MASTANEH TORKAMANI AZAR

Elektronik Mühendisliği, Doktora Tezi, Ağustos 2020

Tez Danışmanı: Assoc. Prof. Müjdat ÇETİN

Tez Eş-danışmanı: Prof. Selim BALCISOY

Anahtar Kelimeler: Beyin-bilgisayar arayüzleri, uyarlanabilir sistemler, elektroensefalografi, sensorimotor ritimler, motor hareketlerin zihinde canlandirilmasi, uzamsal-izgesel öznitelikler, faz baglantisalligi, zihinsel durum tanima, bilis, sürekli dikkat, uyaniklik, SART, istatistiksel sinyal isleme, istatistiksel çikarim, derin ögrenme, evrisimli sinir aglari, Bayes modelleri, degisim noktasi tespiti.

Beyin-bilgisayar arayüzü (BBA) sistemleri, beyin ile harici cihazlar arasında doğrudan iletişim kanalları kurmayı amaçlamaktadır. Bu arayüzleri inşa etmek için birincil motivasyon inme ve amyotrofik lateral skleroz (ALS) gibi, kas kontrolü sınırlı olan veya hiç olmayan hastaların, ölçülen beyin elektriksel aktivitelerine dayalı biçimde, niyetlerini otomatik olarak yorumlayarak, bilgisayarları veya diğer cihazları kullanmalarını sağlamaktır. Ayrıca, günümüzde sağlıklı bireylerin BBA sistemlerini ek bir iletişim kanalı olarak, belirli insan bilgisayar etkileşim sistemlerinde, kullanmalarını sağlamak da büyük bir ilgi çekmektedir.

Mevcut deneysel BBA sistemleri gözetimli bir şekilde eğitilip daha sonra test oturumu verilerinde değerlendirmektedir. BBA'ların günlük ve uzun vadeli, örneğin yarı otonom arabalarda, kullanımına yönelik artan taleplerle, bu tür sistemler daha uzun zamanlı oturumlarda test edilmiştir, ve bu bağlamda araştırmacılar eğitimli sistemlerin başarımlarının önemli ölçüde düştüğünü gözlemlemişler. Bunun elektroensefalografik (EEG) sinyallerin durağan olmayan doğasından kaynaklandığına inanılmaktadır. Bunun sonucunda, test oturumları sırasında bu tür değişikliklere uyum sağlayan, yarı gözetimli öğrenme ile uyarlanabilir BBA'ların tasarlanması yeni bir araştırma alanı olarak ortaya çıkmıştır. Bu sinyallerin durağan olmamasının temel nedenlerinden biri, kullanıcıların bilişsel yük, uyanıklık, dikkat, yorgunluk, can sıkıntısı ve motivasyon gibi bilişsel durumlarındaki değişikliklerdir. Ancak, EEG sinyallerinden bu tür bilişsel durumlar hakkındaki bilgileri dinamik olarak çıkarmak ve bunu BBA sistemlerinin başarımlarını iyileştirmek için kullanmak önemli ve hâlâ çözülememiş zor bir araştırma sorunudur.

Biz bu tezde çok karmaşık bir sorun olan, bilişsel görevlerin yürütülmesi sırasında kul-

lanıcıların uyanıklık ve dikkat düzeyini tahmin etmeyi ele alıyoruz. Uzun vadeli görev ve tepki süresi tutarlılıklarının nöral, EEG tabanlı ilintilerini belirlemek için tepki görevine sürekli dikkat (SART) testine dayalı bir dizi deney tasarlıyoruz. Uyanıklık için yeni bir uyarlanabilir puanlama şeması önerdikten sonra, içsel dinlenme ve görevle ilgili beyin ağları arasındaki yakın ilişki hakkında yeni kanıtlar sağlıyor ve sinir ağları ve dinlenme durumu EEG sinyallerinin uzamsal-izgesel öznitelikleri üzerinde alaka analizi kullanarak tonsal başarım ve tepki süresindeki tutarlılığı öngörmek için modeller geliştiriyoruz. Daha sonra, düşük ve yüksek uyanıklık aralıklarını öngörmek hedefine odaklanıp, evresel uyanıklığın ve başarım tutarlılığının başarılı öngörücüleri olarak evre kilitleme değeri özniteliklerini kullanan, evrişimli sinir ağlarına (CNN'ler) dayalı tam otomatik sistemler öneriyoruz. Bu katkılarımızın tümünde, uzun ve monoton deneylerdeki aşırı uyanık ve uykulu aralıkları modellemek ve tespit etmek için kişisel uyanıklık özniteliklerini ve bireysel psikofizyolojik farklılıkları dikkate alıyoruz, ve literatürü, uyanık ve tutarlı davranışın uzamsal-izgesel-zamansal ilintilerine dair kanıtlarla zenginleştiriyoruz.
Ardından, kullanıcıların sürekli uyanıklık seviyelerinin yeni bir aşamaya girdiği anların sıralı çıkarımı ve tespiti için değişim noktası modellerini kullaniyoruz. Çevrimiçi ve çevrimdışı uyanıklık modellerimizin, hem laboratuvarımızda toplanan SART veri kümelerinde hem de uyanıklık etiketleri içeren sürüş veri kümelerinde değişim noktalarını başarılı olarak tespit etmesini gösteriyoruz. Sonunda, bu tezin en öne çıkan katkısı olarak, altta yatan uyanıklık seviyelerinin kullanıcıların tepki verme süresini ve dolayısıyla BBA motor hareketlerini zihinde canlandırmaya odaklanma kabiliyetini etkilediğini varsayıyoruz. Daha sonra, eğitim ve test oturumları sırasında aralıkların uyanıklık seviyelerini etiketlemek için bir dizi yeni gözetimsiz öğrenme şeması kullanan ve hem öznitelik çıkarımı hem de sınıflandırıcı parametrelerinin eğitiminde tam uyarlanma özelliğine sahip bir yönteme yol açan Hayali Motor Hareketleri Tabanlı BBA için bir Uyarlanabilir Uyarılılığa dayalı Sınıflandırmayı sunuyoruz. Bu uyarlanabilir sınıflandırma yaklaşımının, çeşitli uyanıklık kümeleme şemalarımız tarafından düşük uyanıklık seviyeleriyle etiketlenmiş aralıklarda farklı şekilde eğitilmiş üç farklı versiyonu tanıtılıyor. Sonuç olarak, kendi SPIS MI-BCI veri kümemiz ve BCI Competitıon IV 2a veri kümesi için orijinal, uyarlanabilir olmayan temele göre uyarlanabilir versiyonların genel test doğruluğundaki gelişmeleri rapor ediyoruz. BBA laboratuvarımızda toplanan birkaç veri kümesi, şu adreste halka açık bir depoya yüklenmiştir https://github.com/mastaneht.

ACKNOWLEDGEMENTS

From the moment I was admitted to Sabanci University, I have been blessed for being guided and supported by my supervisor Prof. Müjdat Çetin whose visions and mentorship have gone beyond shaping the project to which the current thesis belongs and resulted in utmost personal and professional growth for me. It has been an absolute honor to introduce myself as his student and hear nothing but appreciation and acknowledgment of his shining personality and his achievements in signal processing and statistical inference. I am especially and deeply grateful for having worked with a true leader who cared for all of his students and made sure they had a safe and encouraging environment to share their opinions. Prof. Çetin has this sheer ability of explaining the most complicated problems in an amazing way, and that has immensely contributed to my love and passion for teaching. I hope the achievements of this work are up to his standards and expectations.

In addition, I am greatly indebted and honored for my collaboration with Prof. Serap Aydın. I appreciate our ongoing discussions as well as our common interests and passion for mathematical modeling of human cognition and neural disorders. Prof. Aydın has constantly brought new perspectives to our project and I look forward to our continued collaboration. I also appreciate and acknowledge long discussions with Prof. Sinan Yıldırım and his special talent in modeling the most complicated biomedical problems in an amazingly simple way, and for delivering the incredibly insightful course on Bayesian statistics and Markov Chain Monte Carlo.

I sincerely thank my dissertation jury members Prof. Selim Balcısoy, Prof. Kemal Kılıç, and Prof. Devrim Ünay for sharing their knowledge throughout these years and for their careful evaluation of my work. Their highly insightful feedback and suggestions have helped to improve the quality and applicability of this dissertation.

Our work on this TÜBİTAK project[1] would not have been completed without the constant collaboration and brainstorming with two dear postdoctoral colleagues, Dr. Sumeyra Demir Kanik and Dr. Aysa Jafari Farmand. Thank you both for helping to shape the research and experimental directions and for all the sleepless nights of discussion, data analysis, and writing. At times we were immensely puzzled by the uncertainty of experimental design and complexities of modeling the human brain, but I think we can now take a deep breath and claim that we did our part although we now have even more ideas to work on and will be in touch for the years to come.

My interest in biomedical engineering and biosignal processing for the greater goal of neurorehabilitation has been shaped by several knowledgeable professors who provided amazing opportunities for me to learn and teach in academia and understand the real needs of individuals affected by neuromuscular disorders. I would like to remember and

[1]This work has been partially supported by the Graduate School of Engineering and Natural Sciences, Sabanci University, and by the grant 116 E 086 from the Scientific and Technological Research Council of Turkey (TÜBİTAK).

thank Prof. Ken Yoshida, Prof. Ali Okatan, Prof. Hakan Ekmekci, Prof. Mehmet Çelik, and Prof. Ali Bülent Uşaklı in this regard. I also thank Prof. Yuki Kaneko for all the encouraging and inspiring discussions.

I want to convey my gratitude to my friends from the BCI group and the larger SPIS laboratory and VPALAB at Sabancı University who have colored my life, supported my work, and cheered me up throughout these years from close and afar. I wish the very best for each and every one of them in their personal and professional lives. I also appreciate all the efforts of the academic and administrative staff at Sabanci University for providing a calm and safe environment to live, learn, and work, and specially thank Mr. Osman Rahmi Fıçıcı, Ms. Banu Akıncı, and Mr. Daniel Lee Calvey for their friendly assistance. Collecting data for our experiments would not have been possible without the interest and devotion of several students, staff, and faculty around the campus, and I thank them all on behalf of myself and my colleagues.

Last but not least, I want to express my love, gratitude, and appreciation to my parents for their many sacrifices that one only comprehends as the years go by, and for going above and beyond to make sure my siblings and I received the education and found the vision that was spectacular in every sense. Thank you for your care and patience and for all the supportive discussions on anything brain related. My biggest hugs go to my sister Goldaneh and my brother Sahand. I know I have been away for way too long, but thinking of your numerous success stories gives me happiness and fills my heart with joy all the time. Finally, to my grandparents here and in heaven: Thank you for the life-long memories and prayers.

Istanbul, August 2020

To the BME community
for the common goals and challenges

Contents

List of Figures

List of Tables

1 Introduction

The last three decades have seen a considerable amount of research on enabling individuals suffering from stroke, Parkinson's disease, and Amyotrophic Lateral Sclerosis (ALS) with the power to gain control of external devices. In this context, systems known as brain-computer interfaces (BCIs) have been developed to provide these users with the means for non-muscular communication and control through interpretation of their brain electrical activity. As shown in Figure 1.1, BCIs are generally designed to record brain signals and extract correlates of intentional control from the central nervous system, and to provide real-time feedback in the form of detected mental actions to patients, their caregivers, and their medical teams. A common BCI specially has to include signal processing and machine learning components to classify features that distinguish between brain rhythms activated during the tasks of interest.
Historically, BCIs have been meant to accomplish one of the following goals: (a) to replace lost functions and skills, as in the case of automatic spellers and word decoders, (b) to restore impaired skills, as in the case of stimulating neural pathways for brain-controlled orthopedics assisting patients with walking or grasping objects, (c) to speed up the rehabilitation process by, for example, stimulating motor cortex through execution of motor imagery tasks, and (d) to enhance the quality of user's experience during interaction with brain-computer and human-computer interfaces by brain/mental state monitoring for detecting correlations of mental workload variations, onset of fatigue, decline of motivation, and lapses of attention or vigilance [1].

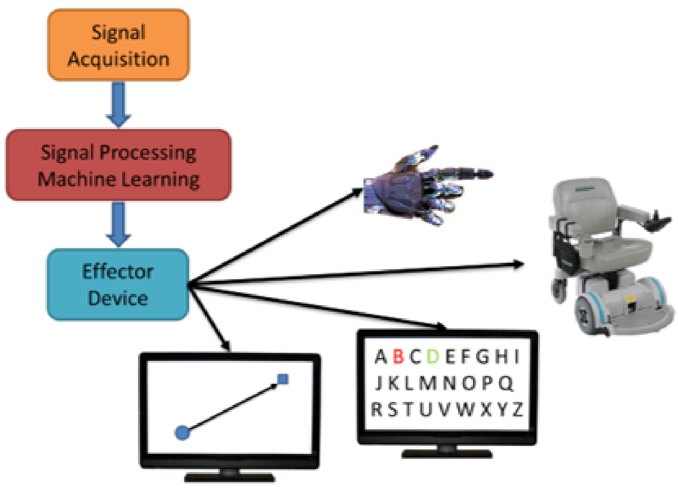


Figure 1.1: The major blocks of a Brain-Computer Interface.

Figure 1.2: A user attending a motor imagery session in the SPIS BCI laboratory.

The last goal, i.e., detecting the underlying mental state has important implications for both patients and healthy subjects. One of the highly used BCI paradigms is motor imagery (MI) in which users have to imagine the movement of a limb when prompted by a cue – in the case of synchronous BCIs – or at arbitrary time points and at their own will in asynchronous BCIs. Common instructions include imagining movements (quick rotation or flexion/extension) of the left hand versus right hand in two-class MI, or to imagine movements of the left and right hands, feet, and tongue in four-class MI. Such imaginations, even when not accompanied by a physical movement, activate brain regions related to motor execution and speed up the recovery of gait and lost movements in a number of neuromuscular disorders [2]. Furthermore, healthy athletes use MI as an assistive and complementary mode for training prior to competitions to improve their ability for modulation of sensorimotor rhythms (SMR) [3]. Interestingly, practicing motor skills through motor imagination is reported as an effective learning tool that improves the surgical performance of medical and surgical trainees [4].

In terms of signal acquisition, BCIs may acquire data invasively through electrodes implanted in the cerebral cortex using the electro-corticography (ECoG) from extremely locked-in patients, or noninvasively and through scalp sensors using electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), or positron emission tomography (PET) technologies. Among these different modalities, EEG recordings provide a relatively less expensive, more reliable, and more robust basis for information extraction and command execution. In this context, each user attends a number of calibration sessions for supervised training of systems' classifiers, and then participates in test sessions where the trained classifier should detect the user's intentions such as the spelled words or direction of imagined movements without any further instruction. Figure 1.2 demonstrates a user with a wired EEG headset attending to a motor imagery visual interface inside the Faraday's case of the Signal Processing and Information Systems (SPIS) BCI laboratory at Sabanci University.

To increase the robustness of BCI systems for long-term use, as needed for locked-in patients or those undergoing long rehabilitation sessions, participants of clinical trials and cognitive studies are invited to attend multiple test sessions or report their experience during daily BCI use. However, classical classifiers show a decline in performance as the duration of the test session increases. Such reductions in correct classification rates are caused by several factors, including high nonstationarity in the brain electrical activity which changes or shifts the learned statistical distributions of EEG signals between trials or across sessions [5], and results in ambiguity in perception of user's intended tasks [6]. Designing traditional or deep learning-based semi-supervised BCIs that adapt to such variations during the test sessions is still an open challenge [7]–[9].

The inherent nonstationarity in cortical activities is likely to be caused by three main factors: (a) Occurrence of physiological events, such as sleep spindles, epileptic spikes, and high frequency activities due to psychological disorders that affect the spatio-spectro-temporal features and statistical distributions of EEG signals [10], (b) Non-cortical sources of disturbance and artifacts such as ocular and muscular movements, cardiac activity, and instrumentation noise [11], and (c) Variations in the users' cognitive states such as the loss of motivation, increase in fatigue and boredom, fluctuations in the cognitive workload due to varying task difficulties, and lapses in sustained attention or reduction in alertness during execution of daily tasks demanding a certain level of engagement [12], [13]. Losing interest and feeling drowsy during long BCI sessions in unstimulating lab settings in a common problem that affects the perception of visual stimuli, results in asking oneself "Did a cue occur?" or "Was the cue pointing to the left or right?", and deters the ability to concentrate on the actual imagination of moving or rotating the limbs. To be more precise, a drowsy user either takes longer to start imagining the limb movement and feels unable to decrease their SMR to the degree that the classifier can distinguish it from a different class or resting state, or completely misses the cue and the upcoming trial. Therefore, dynamically extracting information about such cognitive states from EEG data and using this information to improve the performance of BCI systems is currently an open and extremely challenging research problem to which this dissertation has attempted to answer.

Interested in the ability to maintain attention over a long period of time in response to infrequent but important stimuli, we focus on sustained attention or vigilance as the cognitive variable of choice to explore in this dissertation. We first discuss inferential methods for estimation of attention levels during the execution of a long Sustained Attention to Response Task (SART) using a variety of spatio-spectral features from EEG signals recorded *before* the task execution during the resting state of the brain, and during the actual execution but before observation of visual stimuli. The flow of one trial of SART can be seen in Figure 3.2 in which digit 3 is the infrequent target and the rest of digits constitute the frequent non-target distractors. Next, we present a novel BCI system that extracts information about the current vigilance level during test sessions, extracts the corresponding MI features from EEG signals adapted to that vigilance level, and performs MI classification. In this context, our proposed Adaptive Alertness-Aware MI Classification falls in the area of cognitive computing in which systems learn to interact with humans and adapt to the context and environmental variations through learning from huge amount of data and acting upon their predictions and inferences. Adaptation of BCI systems to changes in personal or environmental factors is one of the topics on the agenda of BCI research, also foreseen in the 2020 Horizon roadmap for Brain/Neural-Computer Interaction Horizon [1]. However, making this update based on the users' sustained attention level has not yet been fully achieved, and the development of such "neuro-adaptive" systems based on continuous assessment of attention level is an important contribution of this dissertation.

In this thesis we have aimed to develop new collective and sequential inference techniques based on deep learning architectures to estimate the level of sustained attention from EEG data during SART tests as our ground truth, labeled data. Second, we aimed to implement a system for estimating the sustained attention level during a complex BCI task such as the motor imagery paradigm. This inferential model combines the perceived intention from the users' EEG data – similar to active BCIs – and the neural correlates of reduced attention – as in passive BCIs – from the aforementioned learned features. Finally, we incorporated the machine learning and inference algorithms developed in the previous phases to develop a neuro-adaptive BCI classification system that tackles the challenging task of updating the BCI classifier based on the estimated level of attention lapses. To the best of our knowledge, adaptive BCIs based on objective and unsupervised inference of sustained attention and other cognitive states as side variables of BCIs that report improved classification accuracy do not currently exist in the literature.

1.1 Recent Work on Adaptive Brain-Computer Interfaces

EEG-based BCIs enable communication by interpreting the user's intent based on measured brain electrical activity. Such interpretation is usually performed by supervised classifiers constructed during training sessions. However, the fact that static classifiers are not robust to shifts in the EEG feature space from one session to the next [5], from the training/calibration session to the test/evaluation/feedback session, and to the changes in cognitive states of users, has generated interest in adapting BCI classifiers in supervised, semi-supervised, and unsupervised manners [7], [14]. The first two options, however, require access to additional labeled data that is hard to obtain objectively. In the past two decades, the BCI community has recognized this need and attempted to develop online learning and classifier adaptation methods [14]–[17]. In one of the major works on BCI adaptation, Vidaurre and Blankertz [18] divided BCI users to three categories: Users for whom a classifier can be trained and run in real-time to provide feedback with acceptable accuracy, those for whom the trained classifier needs to be updated to be successfully used in feedback sessions due to changes in learned features because of various sources of nonstationarity, and people for whom even the training phase fails and results in the

chance level accuracy. For the third group, the classifier cannot either detect any sensorimotor rhythms over the motor region, or no distinguishable activity is detected in the left and right cortices. Updating of classifiers and co-adaptation learning have been proposed as a possible solution for users in the second and third groups [18], [19].
In the context of covert adaptation for BCIs, classifiers can be updated with supervised methods using only the labeled data, in a semi-supervised manner with both labeled and unlabeled data, and in an unsupervised approach with only unlabeled data. Supervised methods for updating the covariance matrix based on subject-independent and subject-specific as well as unsupervised adaptation with subject-specific features were utilized by the Berlin group on the three aforementioned types of users [18]. Semi-supervised versions of linear discriminant analysis (LDA) are frequently utilized, assuming that class conditional attributes are variables with normal distribution [7], [20]. Online experiments have shown that these approaches, through adaptation to the sensorimotor modulation patterns, perform better than non-adaptive methods by reducing the training time and resulting in classifiers that can be applied to more than one user [18], [20]. Semi-supervised learning with self-labeled data has been studied by our group in the context of P300 spellers and motor imagery experiments as well [7]. These methods are easier in the context of synchronous BCIs compared to self-paced, asynchronous BCIs [21]. Utilized methods involve rotating the LDA hyperplane through adapting to EEG features, or shifting this hyperplane in parallel to the initial plane to minimize the classifier's time-normalized false positive rate. Error-related potentials (ErrP) have been also used to adapt the BCI systems [22]. Efforts to increase the reliability of MI-based BCIs usually focus on three main approaches:

1. Improving the machine learning and signal processing algorithms for increasing the classification accuracy. These efforts include but are not limited to, common spatial pattern (CSP) filtering, filter bank CSP (FBCSP) [23], Laplacian filtering [24], Riemannian geometry-based classifiers [25] and their variations, deep and shallow CNN-based architectures [26], [27], transfer learning and domain adaptation methods for reducing the calibration time [17], and FBCSP followed by adaptive ensemble learning [5] or neuro-fuzzy classifiers [28].

2. Training the users to better control their sensorimotor rhythms (SMR) while presenting feedback through visual, audio, or tactile modalities or even learning companions [29]. This training should also focus on a combination of personal traits and habits since a variety of psychological, cognitive, physiological, and technology-related factors as well as spatial and attention-related abilities affect the usability and reliability of BCIs in general and MI-based BCIs in particular [30], [31].

3. Co-adaptation of users and machines/BCIs: Recent studies have shown that extreme rates of machine/classifier adaptation slow down human learning [32], so a balance or personalization has to be reached between the adaptation and re-training of algorithms based on users' individual traits. [19].

1.2 Machine Learning for Mental State Recognition in BCIs: A Challenging Problem

In this thesis, we pay special attention to the fact that changes in cognitive states such as alertness and vigilance during test sessions lead to variations in EEG patterns of the user and deteriorate the calibrated classification and interpretation rates of BCI systems [33]. It has been shown that increased cognitive load, induced by presenting visual distractors during the execution of MI BCI, could significantly predict reduction in BCI performance of users whose undisturbed accuracy was below 75% [34]. This finding further supports our work that was started by a long experiment of Go/NoGo or target/non-target selection demonstrated in SART. A wide variety of studies, including those published by the author of this thesis and her co-authors, have been concerned with psychological tests and assessments of drivers' and operators' vigilance, fatigue, and drowsiness and have introduced features to characterize those states under different experimental protocols [35], [36]. For these reasons, a large body of literature supporting our mindset arises from studies on driver vigilance assessment and sleep state classifications. However, although there have been advertisements on use of cognitive computing outside lab settings, many of these studies continue to be conducted inside the controlled lab environments. In this section, we present the most important arguments for the challenging nature of mental state recognition using BCIs in these confined conditions. Some of these challenges may apply to classification and regression tasks in the context of other medical imaging and signal processing tasks as well. Still, these arguments are supported by our own experience during data acquisition and data analysis stages of this work.

1. Small datasets due to the limited number of participants: Without having access to medical/clinical data in a hospital setting which requires completion of certain guidelines, collecting neurophysiological and BCI datasets in a lab setting using healthy participants is a challenging task that requires carefully designed approaches for participant recruitment. Before consent forms are signed, experimenters need to attract the volunteers' trusts and assure them of the safety and privacy of collected data. The duration of headset setup in the case of traditional and gel-based electrodes and the need for cleaning the hair after the experiment further complicates the procedure of participant recruitment.

2. Limited number of trials in each experimental session: In motor imagery datasets of BCI Competition IV organized by the Graz BCI group [37], each trial lasts for 8 seconds. As described in the experimental setup of Chapter 6, we reimplemented the visual paradigm of this dataset and reduced it to 6 seconds; thus, a 30-minute session only provides 300 trials which are 1) scarce when compared to an image classification task that could contain millions of images, and 2) highly variable in terms of alertness levels as comprehended through facial video recordings of participants and their post-experiment narrations. The temporal inconsistencies and

non-stationarity prohibits common data permutation schemes, and can be resolved considering the solutions proposed for cross-validation of block-wise neuroimaging data [38].

3. Curse of dimensionality in neurophysiological datasets [39]: In the case of classifying imagination of left or right hands, up to 10 electrodes placed over the sensorimotor cortex have deemed essential for classifying the motor imagery activity. When it comes to characterization of sustained attention and vigilance through spatial networks and connectivity analysis, fMRI has been a tool of choice in clinical settings that has helped to identify attention networks in the frontal and parietal lobes as well as their bidirectional interactions in improved levels of sustained attention [40]. Studies wishing to reproduce those spatial links using surface EEG electrodes thus naturally employ larger number of electrodes across the whole scalp to utilize source separation methods and characterize spatially-wide electrical neuronal networks in the cortex. Thus, extracting multiple spectral, temporal, and spatial features from collections of at least 64 electrodes has been a common practice. Regardless of the classification or regression algorithm of choice, any dataset with high dimensions and low number of trials is susceptible to overfitting. Thus, learning the key spatio-spectral features to obtain acceptable detection rates in intra- and inter-subject classification schemes is an ongoing challenge.

4. Artifact contamination: The amplitude of noninvasively recorded EEG signals is in the order of microvolts and results in a poor signal-to-noise ratio (SNR) due to their contamination with power line noise, weak electrode contact with the head, and current drifts [41] which all have non-cortical sources. Furthermore, artifacts induced by muscle movements that are divided into electromyograms (EMG) from face and neck muscles, electrocardiogram (ECG) from cardiac activities, and electrooculogram (EOG) from horizontal and vertical eye movements are highly visible in raw EEG recordings if the data acquisition system does not apply any artifact rejection technique. Studies utilize online and offline solutions to monitor temporal features and omit trials contaminated with heavy artifacts. However, due to our already small number of trials, we do not have the liberty of discarding those trials and prefer to utilize artifact reduction techniques that span simple temporal and statistical features as well as more complicated techniques such as spectral-filtering, source separation, adaptive fuzzy networks, and the like [42].

5. Ground truths for cognitive states: Obtaining the ground truth for vigilance levels and other invisible cognitive states – thoughts and affective events that do not necessarily result in actions and movements [43] is a challenging task [44]. Unlike image, video, and emotion classification datasets that are widely annotated, tagging neurophysiological datasets in terms of alertness, frustration, or boredom is extremely challenging. Pausing the experiments to collect subjective answers for such states, although practiced by a few groups [33], [45], severely disrupts the natural flow of

cognitive tasks [46] while resulting in highly subjective and biased evaluations [47] that ignore the immediate cognitive reactions to the stimuli [48]. Lack of objective ground truth results in an unfair disadvantage in cognitive monitoring since datasets on epileptic seizures or sleep stages are generally annotated by clinicians [49], [50]. For these reasons, a scheme was suggested for scoring vigilance levels based on the occurrence of sleep spindles in resting-state EEG recordings [51] which is not completely useful during demanding cognitive tasks due to increase in similar brain activities.

6. Noisy labels: In machine learning, noise refers to the mislabeling of trials and samples. Besides difficulties in obtaining valid ground truth for cognitive states in the first phase of this work, we were facing a more critical problem in the second phase when we turned to identifying changes in vigilance levels in the background while the participants were focused on motor imaginary tasks. Was the classifier unsuccessful because the user was fighting drowsiness despite trying hard to execute the instructed task? Was the user alert but merely unable or not trained enough to control and desynchronize their brain rhythms in the left (right) cortex while imagining their right (left) hand movements? Were they even concentrated enough on the task or were they day dreaming or frustrated because of the long duration of the task and being confined in their seat?

These limitations and challenges necessitate the need for introducing new solutions that are valid across the datasets of the majority, if not all, of participants considering their personal patterns of attention and concentration maintenance during cognitive tasks execution. Most importantly, the proposed methods should be able to decode the obscured and intended mental command from these small, high-dimensional datasets with noisy labels. In Section 1.3, we briefly introduce the contributions of this thesis and invite the reader to respective methods in each chapter for a detailed description of our proposed solutions for the aforementioned challenges on mental state recognition for BCIs.

1.3 Thesis Contributions

The contributions of this thesis can be described as follows:

1. We present a summary of findings on neural and behavioral correlates of task execution and attention decline during a series of SART, a standardized test battery used by the behavioral neuroscience community. We introduce a novel and adaptive cumulative vigilance score (CVS), shown in Figure 1.3 for two different participants of our 105-minute experimental sessions. Interested in explaining the neural and behavioral correlates of such diverse personal differences in maintaining a consistent performance or falling asleep and regaining alertness, we demonstrate how the intrinsic activity of the brain while the user is still at rest and completely disengaged from demanding cognitive loads of visual perception and memory tasks can

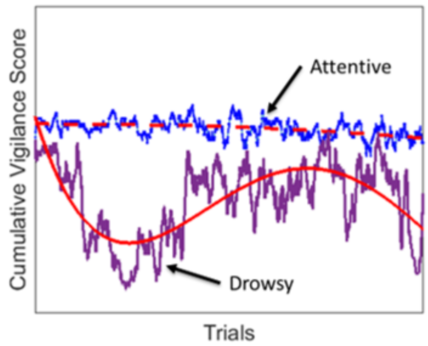


Figure 1.3: CVS curves of two different SART participants demonstrate highly different individual vigilance patterns.

predict the average and variability of performance scores and response time in a long task with infrequent targets and frequent visual distractors. The first part of this work exploits models used from 168-dimensional band-power features. Besides the multivariate regression approach and use of deep networks, we add to the literature by findings on the roles of beta and gamma oscillations in human attention and impulsiveness. To the best of our knowledge, deep architectures using resting-state features for regression of sustained attention objective measures either during SART execution or as side variables of BCIs do not currently exist in the literature. This study has been published in [52].

2. In the second part of this work, we focus on brain connectivity and interactions among pre-frontal regions involved in high-level cognitive tasks, with attention networks distributed across the frontal and parietal cortex, and visual cortex in the occipital region. Using multivariate pattern analysis, this work demonstrates more accurate prediction models built from intrinsic resting-state networks of the brain computed using multi-spectral matrices of pairwise phase synchrony indices. This work builds on and extends our results in [52] with the use of more advanced features capturing spatial interaction patterns.

3. We develop an inference technique based on deep neural networks to estimate the level of sustained attention from EEG data during SART sessions. We determine which spatial and spectral features of EEG signals, when recorded up to one second before occurrence of visual stimuli, are best regressors of cross-correlated models that predict the average block-wise CVS and response time for all users from phase synchrony indices of participants. This work has been published in [36].

4. Using pre-trial band-power and phase locking values (PLV), we investigate which spatial, spectral, and temporal features correlate with and distinguish between periods of low and high vigilance for each participant alone. Unlike several studies that claim classification of attention versus no-attention states, these periods are based on objective performance measures and not intentionally pre-designed to include periods of high and low cognitive loads. We demonstrate the superiority of

PLV features when paired with a customized convolutional neural network (CNN) architecture for our task.

5. Focusing on changepoint detection as a sequential inference technique in the context of time-series modeling, we demonstrate that online and offline Bayesian changepoint algorithms can detect onsets of vigilance variations from both the objective performance curves as well as the EEG-based neural correlates of attention variations identified in the previous chapters. The novel fusion of band-power vigilance predictors with Bayesian changepoint detection (BCPD) methods demonstrates the success of unsupervised inference models in cognitive state monitoring and has utmost potential for alarming clinicians, educators, BCI experimenters, and operators of critical systems about onsets of attention decline.

6. Proposing that the vigilance level affects reaction time and thus the performance in an MI task, we present applicability of predicting high and low vigilance levels, obtained from a cumulative MI classification accuracy, using a series of pre-trial band-power features. This work provides a solution to the problem of estimating the sustained attention level during a complex MI BCI task that combines the perceived intention from the users' EEG data –similar to active BCIs– and the neural correlates of reduced attention –as in passive BCIs. This study has been published in [53].

7. Finally, utilizing the machine learning and inference algorithms developed in the previous chapters, we present a neuro-adaptive BCI classification system that, during continuous estimation of attention levels of BCI users, performs covert adaptation by updating the classifiers' parameters. This latest contribution involves a fully adaptive BCI classification framework that infers vigilance-related EEG features from a continuous time-series and applies an unsupervised scheme to report a discrete-valued vigilance level which will then be used by the CSP and LDA classifier to adaptively extract motor imagery features and perform classification tailored to that vigilance level.

1.4 Thesis Organization

In this section, we present a brief overview of the topics and organizations of the rest of this thesis.

1.4.1 Chapter 2: Background

Chapter 2 presents an overview of electroencephalography and its related modalities most commonly used in the context of EEG-based BCI systems, lays the neurological foundations for sustained attention and motor imagination, and presents an overview of state-of-the-art techniques for cognitive state inference and adaptation in BCI systems.

1.4.2 Chapter 3: Multivariate Regression Models for Vigilance Prediction from Resting-State Spatio-Spectral Features

Chapter 3 introduces sustained attention to response task (SART) as our fundamental paradigm for objective evaluations of within- and between-subject vigilance variations. It then presents details of calculating a novel and objective cumulative vigilance score (CVS), calculated from error counts and response time, which is *adapted* to the users' personal reaction patterns. Finally, two proposed cross-validated multivariate regression and neural network models are presented that predict the mean and variability of performance score and response time during the long, 105-minute experiments from spatio-spectral features of resting-state EEG networks.

1.4.3 Chapter 4: Deep Neural Networks for Vigilance Prediction from Pre-Trial Spatio-Spectral Features

Chapter 4 presents two pieces of work for (a) regression for block-wise, continuous-valued performance scores and (b) classification for trial-wise, discrete-valued vigilance levels using the band-power and phase synchrony features obtained from pre-trial EEG signals during the execution of long SART. Conducted using deep and convolutional neural networks with traditional classifiers serving as their baseline models, these experiments put special emphasis on predicting highly different behavioral traits of vigilance and alertness habits, as reflected in individual CVS curves, from customized temporal and spatio-spectral features collected in immediate short time periods.

1.4.4 Chapter 5: Bayesian Models for Changepoint Detection in Vigilance Time-Series

Focusing on the sequential inference methodologies from time-series data, Chapter 5 presents details of online and offline Bayesian changepoint detection (BCPD) algorithms that are validated with a vigilance dataset labeled with eye-closure curves. We demonstrate the superiority of online changepoint detection from band-power EEG features which is very promising for detection of transitions in datasets without vigilance or drowsiness labels. In the context of EEG datasets, changepoint models have been either applied on sleep recordings or epileptic seizure detection, and this novel application provides a promising result for real-time detection of vigilance transitions in an unsupervised approach.

1.4.5 Chapter 6: Adaptive Alertness-Aware Classification for Motor Imagery-based Brain-Computer Interfaces

Building on behavioral and neural correlates of vigilance variations developed and discussed in the previous chapters, Chapter 6 presents two novel pieces of work that evaluate the effects of users' vigilance on the tonic performance of MI-based BCIs. Concerned with detecting signs of attention decline, the first work proposes and demonstrates the effectiveness of a variety of pre-trial spatio-spectral alertness features in predicting MI classification performance with vigilance information inferred from a newly constructed cumulative classification score. In the second work, focusing on the pre-trial EEG correlates of attention variations, an unsupervised scheme is proposed to cluster trial-wise vigilance features in the training and test sessions. This new information is used in the context of a novel alertness-aware adaptive classification approach. Results indicate the improved inference of the two-class motor imagery performance in our own dataset as well as a commonly used public dataset from BCI Competition IV.

1.4.6 Chapter 7: Contributions and Future Work

Chapter 7 summarizes the contributions of the previous chapters and discusses planned extensions. Suggestions are also presented for future work on mathematical models of cognitive inference and alertness-aware adaptive BCIs considering the limitations of and ideas inspired by the conducted studies.

2 Background

In this chapter, we set the foundation for introducing neuro-adaptive brain-computer interface (BCI) systems that continuously detect lapses in sustained attention during the execution of psychological tasks and active BCI experiments and adapt the system to variations in the user's cognitive states. Performing the aforementioned updates for BCI systems considering the level of sustained attention and through inferential techniques has not been fully comprehended or studied yet. The visions behind this approach and proposed methods rely on findings and theories developed in behavioral and neurophysiological studies on working memory and cognitive workload during goal-directed task execution, spectro-temporal features of perception and attention lapses, arguments on impulsivity and inattentiveness in attention deficit/hyperactivity disorder, event-related potentials in response to visual stimuli, vigilance and maintenance of sustained attention in simulated driving environments, feature selection, feature extraction, and pattern recognition procedures for brain imaging and EEG signals, and statistical and inferential methods utilized for brain state estimation in active BCI systems – especially the P300 spellers and motor imagery decoders. These topics will be gradually discussed in this and subsequent chapters based on their use in our ongoing and future work.

This chapter starts with a brief presentation of neurophysiological signals and their spectral, temporal, and spatial correlates before briefly discussing the EEG-based BCIs. It will then continue with definitions of sustained attention and vigilance. This chapter will concluded by a literature review on assessment and inference of cognitive states in general and attention and vigilance in the context of active and passive BCIs and methods used in development of adaptive BCI systems.

2.1 Neurophysiological Signals for Brain-Computer Interfacing

BCI systems communicate with the outside world using brain commands generated in response to external and internal stimuli. External stimuli such as sound, vision, heat, pressure, and smell stimulate sensory pathways used to make high-level decisions in the cortical area. These decision-making processes are then converted into brain signals with the intentional, voluntary control or the motivation to perform motor activities. Online and offline BCI systems measure, record, process and transmit these signals to computer applications, electromechanical devices, and intelligent systems. Experiments are usually carried out in a controlled environment where users are instructed to focus on audiovi-

sual cues and respond to certain alphanumeric elements, graphical templates in various forms, or to be unresponsive but alert. Data obtained in offline experiments are labeled with intended activities, and acquired signals are used to construct probabilistic models or train classifiers with machine learning algorithms so that the trained models can be used to classify test signals. These systems are evaluated in the active and passive BCI categories. Imaginary motor movement experiments, P300 spellers [10], [54]–[56], and wheelchair control for disabled patients are examples of active BCIs, generally focusing on rehabilitation of people with neuromuscular disorders such as the stroke, ALS, and Parkinson. These systems are in contrast to passive BCIs that are based on involuntary and arbitrary brain signals [57]. Passive BCIs aim for modeling background brain activities, recognizing mental states [58], performing daily volunteer activities in a specific way, and increasing the quality of brain-computer interaction for patients and healthy users alike [1], [59].
A challenging aspect of these experiments is their long duration. While traditional psychology tests are usually conducted for a short time interval and in a controlled environment, BCI tests are carried out for long hours or multiple number of short sessions and relatively uncomfortable conditions [60]. The experiment duration, number of trials, attractiveness of experiments, and the degree of individuals' engagement influence the quality of the recorded signals and the accuracy of the classifier, and there are large experimental evidence that accuracy declines are caused by variations in the learned underlying statistical distributions of each class features. Repeated experiments, particularly where the stimulus is presented at a certain level, can lead to users' habituation. What's more, long breaks in tests with target and non-target stimuli can cause a person's attention to diminish. For these reasons, developing passive BCIs and neuro-adaptive systems have gained considerable attention from the BCI community.

2.1.1 Electroencephalography in Brain-Computer Interfacing

Neural oscillations are defined as the rhythmic fluctuations of electrical activity generated by tissues in the central nervous system. Having a spatially multi-scale nature, these spikes occur at the synaptic and transmembrane potentials at frequencies in the order of Gigahertz before their distributed activities are sensed at the intracellular and local extracellular fields. At an even higher spatial level, the electrical activity in the cerebral cortex is known as the electrocorticogram (ECoG) or intracranial electroencephalogram (iEEG). In this level, each location's cortical activities represent local synchrony while different cortical sources are relatively independent. However, when signals pass the scalp and reach the skin, EEG signals are projected as point processes, with temporal resolutions in the order of milliseconds. A general assumption is that multi-channel EEG sensors (electrodes) record a linear combination of EEG source signals with negligible propagation delays which could be decomposed through blind source separation (BSS) algorithms. Independent component analysis (ICA) is one of these methods that attempts to reconstruct statistically independent source signals by minimizing their redundancy or

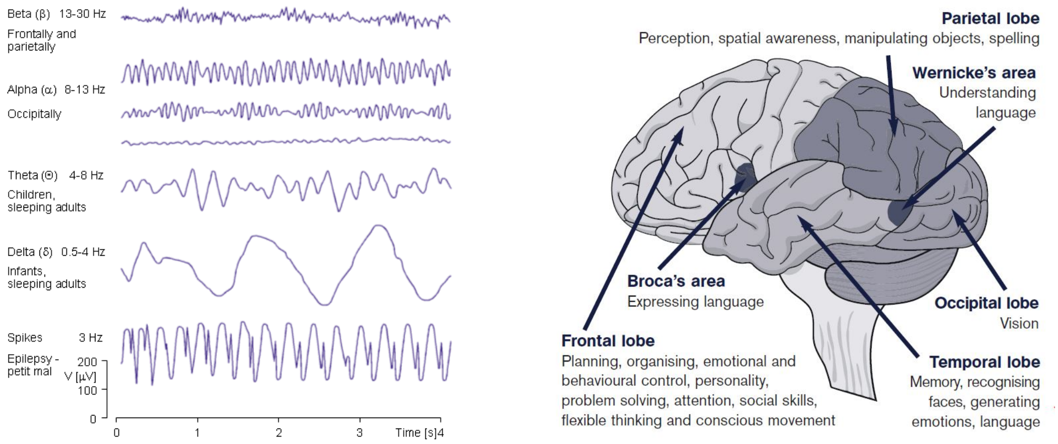


Figure 2.1: (Left) Sample EEG oscillations [61], (right) brain lobes and functions of main cortical regions, picture from Headway Thames Valley.

mutual information.

In the spectral domain, many studies utilize the fixed-brain rhythms known as the delta (0 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 30 Hz), and gamma (30 – 70 Hz) bands. The left plot in Figure 2.1 demonstrates general shapes of these oscillations while the right plot shows the main brain lobes and high-level cognitive functions with which they are associated. Delta rhythms are slow, high-amplitude waves and, in adults, they are mostly visible over the frontal cortex during sleep. Cortical locations for theta frequency can be quite distributed, but frontal theta is known to be a sign of drowsiness and idling during resting states and a correlate of high engagement during task performance. Alpha activity, usually divided into lower and upper sub-bands, is mostly visible in the occipital lobe during relaxed states and eyes closed, but spreads towards the parietal, temporal, and frontal cortex as the relaxation level increases. The mu rhythm shares the same spectral band as alpha and is visible over the sensorimotor cortex and modulated by movements, intentions, and imagery activities. Beta band is more observed over frontal and sensorimotor cortex and is a correlate of alertness, active concentration, and tension. Finally, high frequency gamma activity is generally focused over the somatosensory cortical areas.

A number of EEG oscillations used in BCI research are the sensorimotor rhythms (SMRs), event-related potentials (ERPs), steady state visual evoked potentials (SSVEP), and slow cortical potentials (SCPs). Here, we introduce the first two oscillations. SMRs are a group of EEG waves including the mu rhythms generally observed over the motor cortex. It is long established that imagining limb movements, also known as motor imagery, reduces the power of mu rhythms. This fact is largely used in BCI experiments as a correlate of intentional control and for rehabilitation of individuals suffering from motor disorders in stroke and other neuromuscular damages to the central nervous system [62]. In these paradigms utilizing the motor imagery concept, participants are instructed to imagine moving or pushing their left and right hands or feet, to which a reduction in the mu power will take place in the opposite brain hemisphere during movement preparation and increase after the movement has taken place. A screenshot of the online motor imagery interface developed by the SPIS BCI group is presented in Figure 2.2. Participants have to imagine rotating their hands or performing flexion/extension movements according to the demonstrated arrow, and after a real-time acquisition and feature extraction of EEG signals, the trained classifier should (1) detect the intended movement, and (2) provide visual feedback to the user by moving the ball one position to the left or right according to the detected movement.

Figure 2.2: The ball-and-arrow paradigm developed by the SPIS BCI group for online motor imagery experiments.

In the temporal domain, ERPs are known to be a class of potentials displaying stable time relationships to a definable reference event. In order to extract ERPs, the experiment interface saves time stamps corresponding to occurrence of stimuli in a trigger signal which is then saved together with the EEG channel data. Once EEG signals are preprocessed and filtered, specific segments before and after the stimuli (event) onset are extracted. The extracted, event-locked waveform is known as an epoch. ERP components are labeled with respect to their polarity – positive or negative – and position (delay) with respect to the stimuli onset within the waveform. However, different experimental modalities can stimulate similarly labeled components that bear no functional relationships with each other [63].

The most famous EPR component is P300 or P3b occurring around 300 ms post-stimuli onset in the parietal cortex following a frontally maximal component called P3a. P300 is generally evoked as a result of attending to an unexpected, surprising, and infrequent target within a flow of more frequent and non-target stimuli. In our BCI experiments, this component is mostly utilized in letter/word speller paradigms highly useful for individuals suffering from post-stroke or age-related aphasia, i.e., disability in speech production because of damage to the language-related tissues in the left brain hemisphere. Known as the P300 Speller and shown in Figure 2.3, the interface for this paradigm consists of a matrix of alpha-numeric characters whose rows and columns flash continuously. The participant is instructed to focus on the character s/he intends to spell by counting or keeping track of flashes in the row and column containing that character. Since the interface repeats these flashes, averaging the waveforms in the post-stimuli onset and in the window between 250 ms to 450 ms can lead to the detection of P300 components.

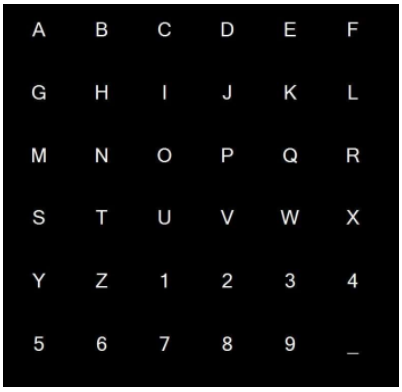


Figure 2.3: The alpha-numetric matrix of the P300 speller interface used in the SPIS BCI experiments.

2.1.2 EEG Signal Acquisition

Due to the location and size of their tissues of origin, ECoG signals are recorded through surgically-mounted, invasive microelectrodes. However, EEG signals can be acquired non-invasively using electrodes mounted on the skin. Although such recordings are affected by low signal-to-noise ratios (SNR) compared to invasive recordings, their electrodes are easier to mount and less expensive for clinical and experimental use. The high input impedance caused by skin and hair can be further reduced if active electrodes with pre-amplifiers are utilized and a conductive gel is applied between the sensors and the scalp. In medical settings that use EEG recordings for diagnosis of sleep disorders and localization of epileptic spikes, a dry gel is applied to the electrodes before directly placing them over the scalp. In our wired BCI setup based on Biosemi devices (Biosemi Inc., Amsterdam, The Netherlands), electrodes are mounted inside caps worn over the head and are connected to an amplifier which delivers the signals to the recording station.
In our experiments with the Biosemi EEG acquisition system, 64 electrodes are set up according to the International 10-10 Electrode Placement System demonstrated in Figure 2.4. This systems labels each electrode with respect to its underlying cerebral cortex area. Number 10 in its name imply that distances between adjacent sensors are 10% of the distance from right to left or front to back of the skull.

2.1.3 Signal Processing for EEG-based Feature Extraction

In general, the majority of machine learning and pattern recognition studies use the following list of features acquired from EEG signals: absolute and relative power ratios from the Fast Fourier Transform (FFT) and Power Spectral Density (PSD), short-time Fourier transform (STFT) spectral images, event-locked measures such as ERPs, event-related spectral perturbations (ERSP) where an event could be the onset of a cue, stimuli, or response/click, decomposition coefficients from multi-scale analysis such as wavelets, filter banks, clustering measures, distance metrics between spatio-spectral maps, Common Spatial Patter (CSP) filtering where after band-pass filtering, simultaneous diagonalization of sample spatial covariance matrices for two classes is performed, entropy

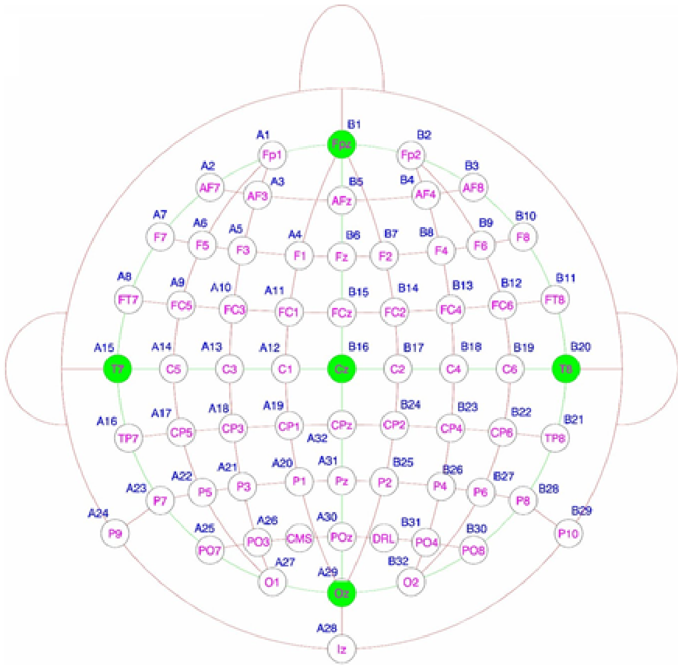


Figure 2.4: Biosemi's 64-electrode montage following the International 10-10 Electrode Placement System.

measures, cross-correlation and autocorrelation measures, and functional connectivity measures such as the phase locking value (PLV).

2.1.3.1 Individual Alpha Frequency

Although the majority of BCI experiments utilize the fixed-interval brain rhythms, papers focusing on attention and cognition emphasize the importance of defining these bands individually for each participant for several reasons [64]–[67]. First, the alpha wave's frequency increases from early childhood to adulthood and decreases afterwards with age or age-related neurological disorders; thus, it is considered a stable neurophysiological trait in adults. Second, analysis of resting-state recordings has shown that the individual alpha frequency (IAF) does not change after cognitive training interventions. Furthermore, the dominant alpha frequency is higher in individuals with better memory performance compared to age-matched controls and positively correlated with the speed of information processing. Females also seem to have higher mean alpha frequencies than males. In some individuals, the lower alpha band shifts towards the theta frequencies which necessitates customization of frequency intervals. More importantly, the narrowband analyses of frequencies within 7 – 12 Hz have revealed their different responses to attentional demands during increasing and decreasing mental engagement and transition between sleeping and waking states.

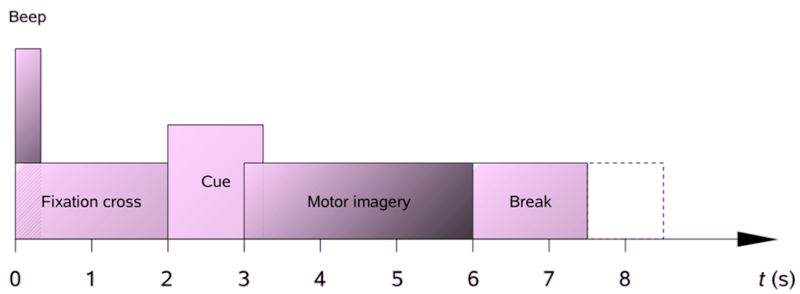


Figure 2.5: The timing flow of 8-second trials in BCI Competition IV - Dataset 2a [71].

2.1.3.2 Common Spatial Pattern Filtering

Common spatial pattern (CSP) filtering is a highly utilized method for feature extraction specially in two-class motor imagery BCI systems. CSP designs spatial filters to maximize the difference between two class variances of the filtered data [28]. Consider the bandpass filtered EEG data for class labels $i = 1,2$ as matrix $U_i \in R^{C\times N}$ where C indicates the number of channels and N is the number of data samples after downsampling per channel. CSP applies simultaneous diagonalization of the average class covariance matrices to design the transmission matrix W. The EEG data are thus projected as $Z = W\,U$. The first and last rows of Z yield the maximum variances for one class and the minimum for the other one, respectively. Log-variance of the projected EEG signals, $S = \log(\mathrm{Var}(Z))$, are utilized to calculate the MI class discriminating features. Experimenters often choose the first and last three or five rows of Z for feature extraction. CSP is usually followed by Linear Discriminant Analysis (LDA) or other binary classifiers.
In motor imagery experiments where a cue or arrow is shown on the screen in the beginning of each trial as shown in Figure 2.5, participants are instructed to start imagination of a hand or foot movement at a certain time. However, the exact timing at which the SMR decreases and desynchronizes with respect to the resting state and can be used for extraction of best and most representative EEG-based MI features by a usually binary classifier is an and challenging topic [68]–[70]. In 6.4.1.5, we propose a flexible time interval feature extraction approach for adaptive MI classification.

2.2 Sustained Attention

In its most general form, attention is defined as "a mental process that deals with the distribution of one's limited capacity among many stimuli in the environment" [72], and attentional processes are defined as processes that control allocation of human cognitive resources. This allocation directly affects the quality of learning and communicating when a large amount of information is available in the environment [60]. As shown in Figure 2.1, attention is one of high level cognitive functions administered by the frontal cortex.
Psychological studies usually focus on the following forms of attention: a) selective attention, or the ability to direct sensory processes to a specific stimulus despite the presence

of distracting information, b) divided attention, or the ability to concentrate on and analyze more than one stimulus in one modality or various stimuli in different modalities at a given time, and c) visual attention, or the mechanism that determines which information can be extracted from inside the visual field.

In the context of studies on selective and divided attention, the use of visual, visuospatial, and auditory stimuli has raised questions on the role of modalities and specific channel pathways [73], [74]. However, when considering alertness during brain-computer and human-computer interaction, an important question is whether attention, or better to put, inattentiveness and mental fatigue can be characterized as mental processes independent of the task modality and stimuli type. An answer indeed exists for this question based on the concept of sustained attention: "The state of maintaining attention over time to continuously monitor a situation for detecting usually infrequent but still significant events" [75], [76], or “the state of being alert or wakeful over time” [77]. The key idea in these definitions is that although the focus of attention could lapse momentarily, sustained attention enables the individual to “re-focus” on the stimuli or task in hand after the distraction. Sustained attention performance is closely linked to the activation of the prefrontal and parietal cortex regions, and is thought to be a component of the top-down attentional process [78]. During this procedure, cortical sensory and related information processing mechanisms such as distractor attenuation are improved.

Several studies of our interest, especially in psychology and neuroscience use the term “vigilance” to describe the same phenomena. In psychiatry, however, vigilance may be referred to attention to potential dangers or the state of being alertly watchful. And, in clinical neuropsychology, vigilance is interpreted as the arousal levels on the sleep-awake spectrum [79]. In this work, we use the first interpretation of vigilance and assume it is similar to sustained attention and tonic alertness. Vigilance has been shown to be a resource-demanding task as it requires constant attention and resource utilization [80]. It is also known to be affected by psychophysics [79], motivation [81], [82], drowsiness, rest, and sleep deprivation [83]–[85], impairment of sleep-wake cycles [79], and brain injury [86].

Some of the task-oriented tests for assessment of sustained attention can be listed as follows: Conners continuous performance tests (CPT and CPT 3) [87]–[89] and their variations with visual and audio stimuli [72], Flanker Inhibitory Control & Attention task [77], Macworth Clock Test [90], and the Sustained Attention to Response Task [66], [86], [91]–[93]. These experiments and many others in cognitive psychology are generally used with EEG and fMRI recordings [92], [94], [95] as well as eye movements and psychophysiological measures such as cerebral blood flow velocity (CBFV) [96].

2.3 Mental State Inference and Neuro-Adaptive BCIs

Fatigue, stress, task engagement, distractibility, and similar variables used to measure the cognitive state of a person are called side variables of the BCI systems as they are not outputs of a traditional BCI experiment. In this context, passive BCIs aim to measure and predict levels of motivation, task involvement, attention, and mental workload while the user is attending to resource-demanding tasks such as spelling words by counting the number of flashes, generating motor imagery commands, performing mental arithmetic operations, and the like [97]. When the goal is to improve the quality of human-machine interaction, these systems can reply on signals involuntarily generated by the brain instead of waiting for the user to generate signals based on voluntary commands [98]. In this section, we first present a summary of studies concerned with the role of attention – but not necessarily sustained attention – on the users' performance in BCI sessions, and review a number of methods using probabilistic graphical models for detection of mental states in general and vigilance in particular. Finally, we look at attempts for developing systems that utilize incoming, real-time information about the user's performance to update their predicting models.

2.3.1 Attention Measurement in the Context of Active BCIs

When performance decline for a trained/calibrated model is observed in an online session or during data analysis, it is important to understand the cause for the model's failure in correctly predicting the class of user generated signals. This shortcoming could be due to the participant's inability in performing the instructed spatial or imaginary movements, for example, the need for re-updating the classifier due to variations in signals' underlying statistics, the increasing fatigue and boredom and lack of motivation due to long recording sessions, or poor interface design and bad relationship or fear of the BCI technology in the participant [18], [31].

A number of research groups have studied the relationships mental states and psychological traits between mood, motivation, mastery confidence, control beliefs or optimism in the experiment's outcome, self-efficacy, fear of the BCI technology, attention and memory span, learning styles (active versus reflective), and imagination abilities on the performance in BCI sessions. Lotte *et al* from Inria have investigated the neural correlates of cognitive and psychological factors and their roles on motor imagery and mental rotation BCI performance [30]. Based on evidence from studies on changes in attention focus and the resource-model theory, they suggested monitoring variations in band powers and attention networks as predictors of SMR-based BCI performance [31].

A few studies have looked at the predictive power of gamma band on motor-imagery BCI outcome [99], [100]. One study by Myrden and Chau on BCI performance and mental states focused on attention, frustration, and fatigue during execution of user-specific tasks, and reported no effect of attention on performance [33]. Although their work was interesting in terms of task selection due to fusing the participants' choices with calibration classification accuracies, their method of asking participants to use sliders for reporting their mental states during the task seemed to be at fault. This issue is discussed more on the following section.

2.3.2 Passive BCIs and Probabilistic Models for Mental State Recognition

The attractive aspect of probabilistic graphical models (PGM) is that they enable efficient modeling and integration of complex temporal and spatial relationships that can be used for implementation of efficient learning and inference algorithms. In the EEG-based BCI problems, labels can be letters to be written through a P300 speller, motor imagery left or right movements, or levels of vigilance and fatigue that are encountered in the process of making such movements. In this way, the extrinsic dynamics of the sequential EEG data are nicely captured. Hierarchical hidden Markov models (HMMs), usually with two layers, have been proposed for learning intermediate states of the mental states inferred from EEG sequences [101]. In these models, the EEG features are assumed to be independent at each time point conditioned on the underlying states. It was also shown that latent-dynamic conditional random fields (LDCRFs) better represent the asynchronous BCI data for execution of motor imagery tasks [102]. In our own group, algorithms based on HMMs and hidden conditional random field (HCRF) have been developed for the BCI experiments on imaginary motor movements [102]–[105] and P300 speller based systems [56], [106].

One of the contributions of this thesis is to use dynamic modeling for inferring sustained attention levels from EEG signals. Dynamic models of of sustained attention and other cognitive states either during long sustained attention execution or as side variables of BCIs are very rarely used in the literature. However, several studies on fatigue and vigilance recognition for drive or flight simulators have developed time-series models and HMMs to detect and estimate nonstationarities in the EEG data. These studies use the correlates of vigilance loss such as reduction in the number of eye blinks, ratio of eye closure during a specific interval (PERCLOS), variations in the speech signal features, slowness of response and reaction time, or a fusion of these features to *label* the electrophysiological recordings [107], [108]. Fusion-based classifiers utilizing features from EEG recording as well as from physiological markers such as ECG signals were shown to increase the classification accuracy and detection speed in mental workload assessment with flight simulators [109]. Similar features were also used for driver fatigue estimation through a dynamic Bayesian network [107]. Interestingly, similar to observation regarding SART correlates, such extracted features have been stable during one week after initial assessment.

Besides these objective metrics, experimenters may ask users to provide subjective assessments of their own sleepiness levels before and after the task [110] or rate the fatigue and sleepiness levels throughout the experiment [33]. Others may divide the experimental sessions into high vigilance, sleepy, and low alertness intervals based on subjectively assigned thresholds of the aforementioned labeling parameters. A number of other studies have used the averaged error rates to obtain a measure for categorization of vigilance or fatigue levels [111], [112]

More recently, Zheng *et al.* developed a multimodal approach for vigilance estimation using EEG and forehead oculogram (EOC) [35]. They applied ICA to obtain the components for eye blink activities and computed differential entropy from the temporal and posterior EEG channels band-passed from 1 to 50 Hz. They manually thresholded the PERCLOS values for data annotation to obtain three levels of awake, tired, and drowsy, and developed support vector regression models. To characterize temporal variations of vigilance estimation, they applied continuous CRF (CCRF) and continuous conditional neural field (CCNF) models [113]. Their results indicated higher theta and alpha and lower gamma activities at temporal and parietal locations during the drowsy-labeled trials compared to the awake state.

2.3.3 Adaptive BCI Systems in the Literature

As mentioned earlier, long experiments with monotonic and steady audiovisual stimuli increase the boredom in participants and creates idle phases in the cortical networks and reduces alertness. The fact that the static classifiers are not robust to changes in the EEG feature space from one session to the next, from the training/calibration session to the test/feedback session, and to the changes in cognitive states of users, has raised the issue of adapting BCI classifiers in various ways. In the past decade, the BCI community has noticed and attempted to address online learning and classifier adaptation methods [16]. Such methods are called covert adaptation techniques. BCI classifiers can be updated with supervised methods using only the labeled data, in a semi-supervised manner with both labeled and unlabeled data, and in an unsupervised approach with only unlabeled data. Supervised methods for updating the covariance matrix based on subject-independent and subject-specific as well as unsupervised adaptation with subject-specific features were utilized by the Berlin group on the three aforementioned types of users [18]. Semi-supervised versions of LDA are frequently utilized, assuming that class conditional attributes are variables with normal distribution [20]. Online experiments have shown that these approaches, through adaptation to the sensorimotor modulation patterns, perform better than non-adaptive methods by reducing the training time and resulting in classifiers that can be applied to more than one user [18], [20]. Semi-supervised learning with self-labeled data has been studied by our group in the context of P300 spellers and motor imagery experiments as well [7]. These methods are easier in the context of synchronous BCIs compared to self-paced, asynchronous BCIs [21]. Utilized methods comprised of rotating the LDA hyperplane through adapting to EEG features, or shifting this hyperplane in parallel to the initial plane to minimize the classifier's time-normalized false positive rate. Error-based potentials (ErrP) have been also used to adapt the BCI systems [22].

Overt adaptation techniques, on the other hand, attempt to update the experiment interface and experiment flow to decrease the participant's boredom and enhance the interaction outcome. In a study measuring the level of participation while reading a paragraph, whenever a reduction in the engagement level was measured through a wireless EEG device, a video was shown about the same paragraph [114]. In another work on designing adaptive agents for education, participants were monitored while listening to a story, and different types of cues were offered by the robotic agent when low attention levels were detected. It has been shown that with training sessions that use wakeful alertness levels, the classifiers' accuracy significantly increases in experiments focusing on attention, concentration, and control [115]. More recently, Zander and his colleagues presented an EEG-based cognitive agent that determines uncertainty in a pilots's path and provides assistance to when the uncertainty arises [116].

3 Multivariate Regression Models for Vigilance Prediction from Resting-State Spatio-Spectral Features

In recent years, analyzing dynamics of the brain during its resting states has gained momentum as earlier studies have demonstrated that these 'intrinsic' activities are shared between the resting oscillations and task-activated networks and could be used as the baseline for assessment of cognitive and sensory functions during task engagement [117]–[119]. Functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) measurements have been used to gain more knowledge about the dynamics of connections among the sensorimotor, auditory, visual, and attentional networks while the person is at rest [120]. Entropy-based measures from the resting brain activity have helped to characterize different levels of cognitive impairment in neurodegenerative diseases [121]. A variety of spatio-spectral, temporal, statistical, and connectivity-based models have been proposed to predict the ability to regulate sensorimotor rhythms (SMRs) [122], identify correlates of motor adaptation learning [123], predict response latency in an attention task for Parkinson's patients and healthy controls [124], characterize the correlates of improved engagement in short Go/NoGo and visual selection tasks [125], [126], predict the users' performance in decision-making situations [127], and, in a work by the author of thesis, develop regression models to identify inter-subject predictors of speed and performance maintenance in long sustained attention tasks [52].

The aforementioned applications of resting-state characterization for predicting performance, task engagement, and response latency have important implications for BCIs that rely on mental selection of target stimuli or repetitive mental imagery and arithmetic calculations, monitoring safety and precision of critical operators and long-haul drivers, and simulations of drivers' vigilance. Vigilance or sustained attention is defined as the ability to maintain alertness in detecting infrequent but important stimuli over long periods of time while blocking the distracting events [78]. In this chapter, we report the results of a work published in the IEEE Journal of Biomedical and Health Informatics (JBHI) in which common predictors of cumulative vigilance scores, response time, and their variabilities for participants of a long attention task are characterized through a multivariate pattern analysis (MVPA) and deep neural networks using band-power features of resting-state EEG [52]. To simulate a scenario in which participants have to tend to a repetitive task while paying attention to the occurrence of less frequent targets, this study conducts a session of fixed-sequence Sustained Attention to Response Task (SART) that lasts up to 105 minutes [93]. While some participants exhibit long periods of sleepiness, others are

able to maintain their performance levels and overcome extreme fatigue while attending to the repetitive and idling conditions of the experiment. This study shows the potential of modeling non-stationarity of resting-state brain dynamics in decoding changes in the underlying mental states such as fatigue, drowsiness, and vigilance that would improve the reliability of neurorehabilitation systems and assessment of operators' performance in critical control tasks [59].
Subsequently, we use a similar MVPA on the phase domain, functional brain connectivity of resting-state EEG to obtain correlates of long-term task engagement. A synchrony measure called the phase-locking value (PLV) [128], [129] is used to extract symmetric, inter-trial phase variability between paired channel signals from seven different frequency bands. For each band and each resting condition with eyes open and closed, PLV is averaged to extract the phase synchorny index (PSI) or phase lag index (PLI) [130] as a potential predictor for the corresponding condition. PLV is robust to fluctuations in signal amplitudes and presence of common sources such as the reference electrode and volume conduction [131], [132], and is known to have low computational cost and high accuracy in representing the connectivity networks from resting EEG and fMRI data [133]. By using the averaged PLV from each of the eyes-open and closed resting-state conditions, this study identifies the common predictors of errors in target and non-target detection, long-term task performance, response time, and their variabilities from resting-state EEG dynamics in a group of participants. Regression results demonstrate that reductions in short-range and regional synchronizations within the posterior cortex predict higher vigilance scores and faster responses while the variability of CVS and response time are generally predicted by frontal inter-hemispheric or fronto-parietal connections. To the best of our knowledge, this is the first time that phase correlates of all discrete and continuous performance measures of long-term SART are discussed in a unified manner.

3.1 Motivation

Cognitive and affective state monitoring has found numerous applications in evaluating memory functions and learning abilities [114], [134], assessing operators' performance in critical environments [135], and providing information to users after evaluation of their controlled and automatic cognitive processes [136], [137]. More recently, it has become a subject of interest in the development of BCIs with neurfeedback to improve symptoms of attention-deficit hyperactivity disorder (ADHD) [138]. Such monitoring can involve the utilization of a variety of neurophysiological biomarkers to determine lapses in attention or vigilance, onset of drowsiness and sleep spindles, or changes in the mental workload under different task difficulty levels [44]. Detecting these mental states can increase the accuracy of human-computer interactions since changes in these cognitive and affective states lead to nonstationarities in the brain electrical activity and cause challenges for automated intent inference [6]. Specifically, the inability to maintain attention and inhibit distracting mental processes is one of the important factors behind

lower-than-expected accuracy of BCI systems [31]. Monitoring the attention level is, among other uses, critical in development of adaptive and assistive BCI technologies for neuromuscular rehabilitation and performance assessment of operators in monotonous and critical tasks such as air traffic control and long-haul driving [44].
Building on well-established assumptions on the close relationship between brain networks during rest and task execution, we propose that the high resolution EEG features recorded while the brain is in the wakeful and alert state can be indicators of task sustainability in a long fixed-sequence SART session. This is especially valuable since few related studies on vigilance estimation, except for the driving simulator experiments, record data for over an hour [139], [140], and fewer studies have presented concise predictors of task-induced vigilance variability from spectro-spatial and specially brain connectivity features of the resting-state EEG. This analysis can be used to predict the stability of an operator's performance *prior* to task execution and to adjust the interface environment and the frequency and type of stimuli parameters in P300 word spellers or motor imagery experiments [56], [105] during calibration sessions. To evaluate the readiness of the resting brain, i.e., its ability to monitor sources of information and to detect, process, discriminate, and respond effectively to them [78], [141], we use multivariate pattern analysis (MVPA), a method mainly applied on fMRI to study the connectivity of distributed brain networks involved in task-related activities [142]. MPVA has been previously used to identify the intrinsic and pre-trial correlates of motor learning in an EEG-based experiment [123].
In this study, we analyze the statistical relationships between objective task performance measures for attention, obtained during a long SART session, and resting state EEG data recorded immediately before the session. Our technical contributions are multi-fold:

1. A novel cumulative vigilance score (CVS) is calculated from error counts and response time (RT) of correct non-target trials, and is *adapted* to the users' reaction time from the initial 50 seconds when they are still highly attentive. We emphasize modeling of RT and CVS variability in addition to their average values as indicators of performance consistency or stability for sustaining attention and motor execution.

2. Various neural networks are trained and cross-validated with EO and EC band-power ratio and pairwise phase synchrony indices. The relative rankings of obtained weights uncover associations between the performance measures and narrow-band, resting-state spatio-spectral features.

3. Multivariate regression models are developed using a thorough feature relevance analysis that demonstrates the effectiveness of small feature subsets from resting-state networks in predicting the overall task-related performance measures.

The rest of this chapter is organized as follows. Section 3.2 introduces SART as our fundamental paradigm for objective evaluations of within- and between-subject vigilance variations, and discusses recent studies on prediction of cognitive and SMR function from the resting-state network. Sections 3.3 and 3.4 then present methods and results of calculating a novel and objective cumulative vigilance score (CVS) and two proposed cross-validated multivariate regression and neural network models for prediction of mean and variability of performance measures. Finally, a detailed discussion on implications of our findings and comparisons with the state- of-the-art method are presented in Section .

3.2 Related Work

3.2.1 Sustained Attention to Response Task

Sustained attention, also known as vigilance or tonic alertness in psychology, is the ability to maintain attention to detect infrequent but important stimuli – signals – over a long period of time while blocking the distracting events – noise, and is characterized by the activation of right prefrontal and parietal regions [78], [143]. In clinical settings, sustained attention is generally quantified by the number of errors and reaction delays while participants are attending to monotonic paradigms such as the Continuous Performance Test (CPT) or Sustained Attention to Response Task (SART) [144], [145]. SART consists of multiple instances of digits from 1 to 9 shown on the screen. These digits could appear consecutively and in a predictable order – thus the fixed-SART paradigm or $SART_{fixed}$, or in a random and unpredictable sequence – hence the term random-SART or $SART_{random}$. The user has to respond to the occurrence of the more frequent, non-target digits by clicking with a mouse or pressing a button while inhibiting their responses to the less frequently observed target, usually digit 3. In clinical assessment of ADHD in children and adults where tests of sustained attention are administered, balanced, conservative, or liberal response styles are also taken into consideration. Furthermore, the variabilities of error and response time, defined as the ratios of standard deviation to the mean, are computed to analyze the ability to maintain executive attention levels needed for information processing [88].

Running either in multiple short blocks or in a long session until signs of fatigue and lapses of attention occur, SART is a type of Go/NoGo experiment assumed to measure the failures of sustained attention through number of errors and response time by demanding automatic and habitual responses to highly frequent distractors in Go trials while inhibiting responses to the infrequent target stimuli. Robertson *et al.* demonstrated the individual performance in SART is stable over time [91]. This experiment was subsequently modified and used by others in neuroscience as it was shown to have a good sensitivity for discriminating error rates between healthy participants and patients suffering from traumatic brain injury (TBI) [86], [146].

Considering the specific instructions given prior to the experiment and the fact that all trials – digits – occur similarly and without any specific cue, performing SART requires an endogenous task control. This is in contrast with exogenous tasks, such as in supervised motor imagery or P300 speller experiments, in which a visual or auditory cue is given to alert the user. Endogenous tasks are considered to be more challenging since the users themselves need to notice and keep track of the stimuli sequence [147]. For this reason, remembering to act as initially instructed could be correlated with higher 'prospective' memory performance, especially in the random-SART paradigm. Still, it has been also suggested that the sequence of 9-1-2 in fixed-SART paradigms behaves as a cue for occurrence of the target trial and enables the brain to re-focus on the instructed action, thus making the task more similar to the common supervised experiments tested in BCI settings.

3.2.2 Resting-State Networks and Brain Connectivity

The resting-state brain activity has been often used as the baseline for activations occurring during subsequent cognitive and sensorimotor functions. Barry *et al.* [117] suggested to consider the eyes-closed (EC) resting-state EEG as the arousal baseline for tasks not involving any visual stimuli, and the eyes-open (EO) recordings as the activation baseline for other experiments with a visual fixation. It has also been shown that functional brain networks used during cognitive tasks are continuously active during the resting state as well [148]. In addition, resting-state oscillatory dynamics reflecting the intrinsic activity of the brain are shared by task activated networks [118] and associated with performance measures in experiments on sensorimotor rhythms (SMR), motor adaptation learning, and attention-related tasks [122], [123], [125], [126]. Cole et al. [119] also showed the information flow in resting networks, estimated from functional magnetic resonance imaging (fMRI), could predict the cognitive task-evoked activations.
In the context of attentional networks, it has been shown that right fronto-parietal regions are stimulated in simple sustained attention tasks with activations associated with increased engagement [149], [150] and deactivations associated with increased mental fatigue and declining performance [151]. Such inter-regional connections as well as alpha desynchronization over parietal channels are critical for coordination of brain subregions related to the attentional processes [152], [153]. However, only a few studies have characterized changes in the EEG-based functional connectivity and phase coherence measures during the alert states prior to long task engagement sessions. Extracting such correlates is challenging due to the lower spatial resolution of surface EEG signals compared to functional neuroimaging recordings. Furthermore, several high-level cognitive and motor functions such as target/non-target recognition, motor execution and inhibition, hyperactivity, variations in allocations of attentional processes, and transitions in alert-fatigued states are likely to be observed in a long experiment, just like the uni-modal paradigm to be described later in this paper [66], [154], [155]. Studies with similar purpose, often in the context of long simulated driving experiments, look into task-related recordings rather than the pre-task resting-state sessions. Sun *et al.* [151] used the first and last 5 minutes of a 20-minute recording during execution of the psychomotor vigilance test (PVT) to identify correlates of alert and fatigued execution from low alpha activity of 26 regions of interest. Kong *et al.* [156] administered two driving simulation sessions con-

sisting of conditions with and without external stimuli which could challenge participants or drive them into drowsiness, and obtained intra- and inter-region coherence from the computed mean phase coherence (MPC) for each electrode pair. More recently, Wang *et al.* obtained a 96.76% accuracy using beta-band phase lag index (PLI) for classification of extremely alert versus fatigue states from the first and last 5 minutes of a 90 minute driving task [157].

3.2.3 Regression Models and Importance of Objective Labeling

Developing regression models for such scenarios requires ground truth labels corresponding to several attention states. However, it is challenging to obtain the ground truth for vigilance levels and other invisible cognitive states – thoughts and affective events that do not necessarily result in actions and movements [43], [44]. Several studies on vigilance assessment from simulated driving sessions label trials by visually inspecting the participants' facial features or assuming they are maximally awake and alert in the beginning of a given task and sleepy towards the end [140], [157]. However, our experimental results show that humans exhibit large differences in temporal transitions between their alert and sleepy states [36]. Other protocols pause the experiment flow and ask participants to rate their own cognitive and physiological states using discrete or continuous scales [33]. These subjective evaluations are prone to high bias and experimental errors [47]. Furthermore, momentary pauses disrupt the natural tonic levels of sustained attention in otherwise fatigued individuals [46]. Finally, self-reported ratings ignore the immediate reactions to the stimuli and require reflective thinking and decision making [48] while the parameters used to assess these cognitive variables should not be affected by delay and consequent memory lapses. In more objective assessments, the average number of errors provides a continuous measure for classification of vigilance patterns [111]. A number of other studies on fatigue and vigilance recognition rely on a fusion of EEG and electrooculogram (EOG) and variations in physiological events such as eye closure intervals, circadian rhythm, speech signal features, and face orientation [35], [107], [108], [158] which require extra processing modules. Thus, an automatic method for quantifying the ground truth of vigilance levels through objective measures, such as the error rates and

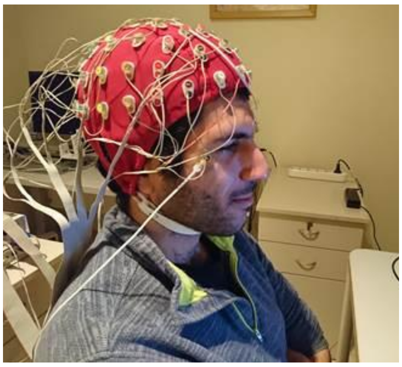


Figure 3.1: A user wearing the 64-channel Biosemi headset (Biosemi Inc., Amsterdam, the Netherlands) and 3 surface EOG electrodes.

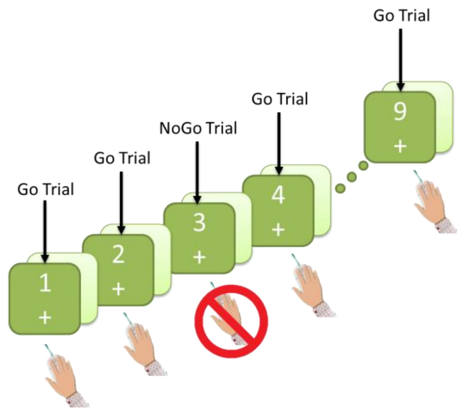


Figure 3.2: One sequence of fixed-SART-varying-ISI. Digit display: 250 ms, response interval: 300 ms, ISI $\sim$ U(400,1000) ms.

response time, deems essential.

3.3 Methods

Ten healthy volunteers, six females and four males, with the average age of 30.25 $\pm$ 6.95 (min: 22, max: 45.5) attended the fixed-sequence SART sessions. Participants were right-handed, had normal or corrected-to-normal vision, and were not under any drowsiness-inducing medications. All participants were naïve to BCI experiments in general and all but two to the SART protocol in particular. Participants provided signed informed consents in accordance with the Sabanci University Research Ethics Council guidelines, and received monetary compensation upon experiment completion.

3.3.1 EEG Acquisition and SART Procedure

Data collection was performed in a dimly lit EEG room within a Faraday cage. Participants were comfortably seated in a chair 20 cm away from a 17-inch LCD monitor. Monopolar EEG activity was collected via 64 Ag/AgCl active electrodes mounted according to the 10-10 International Electrode Placement System as shown in Figure 3.1. Experiments were conducted in the early afternoon hours to induce drowsiness in the already idled brain networks [35].

Participants completed a 2.5-minute resting session with eyes open followed by a 2.5-minute resting session with eyes closed. Before each resting-state session, the interface prompted participants to perform a specific mental multiplication to ensure they were in alert and wakeful states prior to the task. After a practice session with one sequence of random-SART paradigm [92], 12 blocks of fixed-SART with varying inter-stimulus intervals (ISI) were executed. Randomized ISIs eliminate any chance of participants becoming habituated by the stimulus timing and reduce the occurrence probability of automatic clicks [93]. Each block lasted for 8:04-8:20 minutes and consisted of 25 sequences of digits 1 to 9 appearing sequentially on the screen [93] with different font sizes to remove the chance of habitualization. Blocks were separated by a 5-s relaxation period. The goal

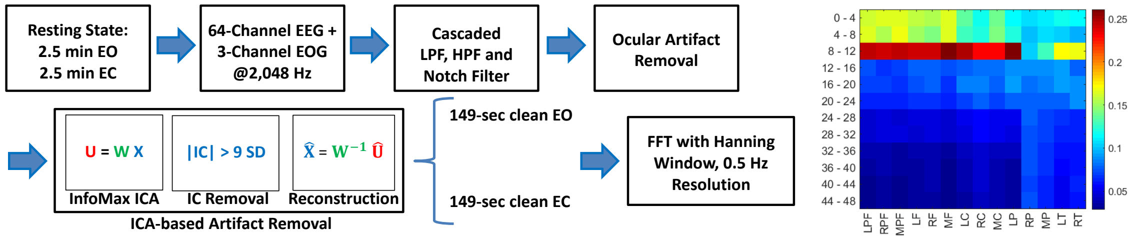


Figure 3.3: Automated pipeline for preprocessing and feature extraction from resting-state EEG. Signals are band-pass filtered in 1-70 Hz. Ocular artifacts are removed with the linear method of [104]. Independent components of logistic Infomax [161] from EEGLAB [162] are z-score standardized before artifact rejection. The heat map demonstrates ratios of BP features from the EO session of participant S10 for the left, midline, and right pre-frontal (LPF, MPF, and RPF), frontal (LF, MF, and RF), central (LC, MC, and RC), parietal (LP, MP, and RP), and left and right temporal (LT and RT) ROIs.

was to press the left mouse button once and as soon as any digit appeared on the screen except for the digit 3, in which case responses should be withheld. Digit 3 was thus the target – NoGo trial – while the other eight digits were non targets or Go trials. Figure 3.2 shows one sequence of this experiment. A full session of this SART paradigm would last for 2,700 trials and between 100 to 105 minutes.

3.3.2 Band-Power Feature Extraction

The automated preprocessing and feature extraction steps applied offline on the EO and EC signals are described in Figure 3.3. Since our preliminary analysis showed participants had different levels of band-powers (BP) in the EO and EC states and the actual SART sessions, the ratios of the 12 non-overlapping band powers up to 48 Hz are computed from the magnitudes of fast Fourier transform (FFT) coefficients for each trial. To efficiently study spatial variations in cortical activities, EEG electrodes are grouped into 14 regions of interest (ROIs) as mentioned in Figure 3.3. These 14×12 features are hereinafter referred to as the BP-ROI feature set.

Defining the BP-ROI feature set has a few advantages. First, comparing these ratios across participants enables us to analyze individual differences in a unified manner. Second, considering narrow bands especially for beta oscillations used in the upcoming regression models allows us to account for individual traits in the modulation of different frequencies and to better analyze the opposite roles of lower beta frequencies as indicators of fast idleness, middle beta oscillations which appear during high engagement and alertness [159], and faster beta activities which reflect signs of existing anxiety [160].

3.3.3 Phase Synchrony Feature Extraction

To compare differences in phase synchronization, resting-state signals are divided into epochs downsampled to 512 Hz with 923 samples. This specific length is selected since the digit-locked epochs of this experiment analyzed in a different study on pre-trial PLV analysis had a length of 1.90 ms equivalent to 923 samples [36], also see Section 4.3.1.1. Epoched data, hereafter known as the resting-state trials, are then band-pass filtered in the following frequency ranges: alpha (8-12 Hz), lower beta-1 (12-16 Hz), lower beta-2 (16-20 Hz), mid-beta (20-24 Hz), upper beta (24-28 Hz), wide-band beta (12-28 Hz), and wide-band gamma (31-60 Hz). For each epoch, the Hilbert transform of band-passed signals is computed [163] using the following formula

$$x_{HT}(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(t')}{t-t'}dt' \tag{3.1}$$

where $x_{HT}(t)$ is the convolution of the real-valued $x(t)$ with $1/\pi t$. $\Phi_i(t,n)$, the inverse tangent of $x_{HT}(t)/x(t)$ for electrode i, is then extracted as the instantaneous phase of transformed signals for electrode i at time t and trial $n, n \in 1,2,\ldots,N$.

Next, the instantaneous phase difference for each electrode pair (i,j) is computed at each time bin from

$$\Phi_{ij}(t,n) = \Phi_i(t,n) - \Phi_j(t,n), \tag{3.2}$$

where $i = 1,2,\ldots,63$ and $j = 2,3,\ldots,64$ in our 64-channel setup. Finally, PLV matrices are computed for all the unique electrode pairs by averaging phase differences across all epochs of similar conditions, i.e.,

$$PLV_{ij}(t) = \frac{1}{N}\left|\sum_{n=1}^{N} e^{j\Phi_{ij}(t,n)}\right|. \tag{3.3}$$

Thus, if the signals measured at electrodes i and j have a small phase difference at each time bin t and all trials, or are in maximal synchrony with each other, their PLV is close to 1. But if their signals have large difference in instantaneous phases or show random pairwise phase variations, their PLV will be closer to 0.

In epoch-based analysis, PLV values are usually locked and computed to the stimuli onsets; however, since the resting-state signals are not locked to any event onset, we average $PLV_{ij}^{EO}(t)$ and $PLV_{ij}^{EC}(t)$ across all the time bins of our downsampled epochs to obtain the phase synchrony index (PSI) for EO and EC conditions. The PSI values from the 2,016 unique electrode pairs of the aforementioned seven frequency bands are collectively called the PSI features.

3.3.4 Cumulative Vigilance Score

Figure 3.4 presents the pipeline for trial-based and cumulative measure calculations. The experimental interface detects the occurrence of commission errors (CE) during NoGo trials, omission errors (OE) during Go trials, and double clicks. Trial RT is defined as the latency of each click with respect to the digit onset, and hit response time (HRT) is defined as the RT for correctly performed Go trials. Since RT variations indicate the inability to maintain vigilance during long attention tasks and tests of attention deficits [88], variability of the overall HRT is calculated as the ratio of the standard deviation

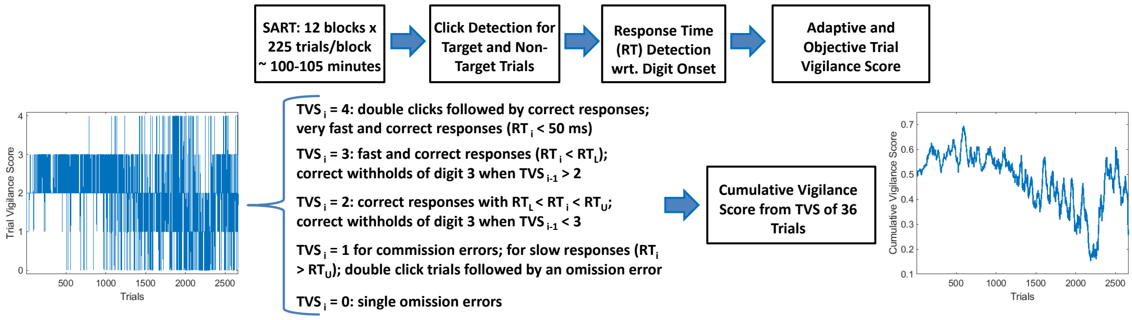


Figure 3.4: Pipeline for detection of trial-wise events and calculating the adaptive and objective Trial Vigilance Score (TVS) and Cumulative Vigilance Score (CVS). RT at each trial is compared with RT_L = 250 ms and RT_U = mean + 2 SD of RT from the first 27 trials.

(SD) to the averaged HRT [36]. This pipeline does not omit trials with RT below a certain threshold as done in [126] to enable the analysis of fast reactions as a natural occurrence in the response traits.

3.3.4.1 Adaptive Vigilance Labels

As shown in Figure 3.4, the adaptive 5-level Trial Vigilance Score (TVS) is proposed as an objective measure for labeling sustained attention without interrupting the users. To avoid penalizing participants with conservative and slow responses, the upper threshold is adjusted to accommodate for each person's response style assuming fast reactions in the first 27 trials (or 50 s) before occurrence of fatigue signs. TVS considers correct response commission and inhibition while rewarding consistency in correct and fast performance (levels 2 to 4), and penalizing inconsistencies when double clicks are performed and subsequent trials are missed (level 1). Double click events usually occur prior to OEs when a participant misses the natural flow of trials and automatically clicks due to being in a low vigilant state. Less frequently, these events take place when a user is in a high vigilant state and expects the next digit, but mistakenly clicks due to the varying duration of the ISI while still managing to respond correctly to the next digit. This novel labeling strategy thus provides a useful and adaptive measure for assessment of vigilance maintenance.
To reflect the tonic user performance, Cumulative Vigilance Score (CVS) at each trial is obtained by calculating the average TVS from 36 preceding trials – lasting for 4 sequences or 73 seconds, and normalizing the result between 0 and 1. Subsequently, the averaged CVS (CVSmean) and reaction time (HRTmean), and their variabilities (CVSvar and HRTvar) as indicators of failure in performance stability and attention sustainability are extracted.

3.3.5 Feature Selection and Visualization with Neural Networks

The literature contains several discussions on correlations among the performance measures with channel-wise BP features. Due to the small size of our dataset and sensitivity to the data of individual participants, we investigate the use of neural networks (NNs) with multiple hidden units for developing the aforementioned regression models. Zheng *et al.* [164] had investigated the critical frequencies and electrodes from trained deep belief networks (DBNs) for emotion classification. They noticed beta and gamma features had received higher average weights in the trained networks across all participants, and saw an improvement in classification accuracy using the differential entropy of all bands with reduced electrode sets. Our analysis will open the path for analyzing learned weights for feature reduction during BCI-based vigilance estimation.

Focusing on the BP-ROI features with 14 (ROIs)$\times$12 (bands), eight schemes are analyzed to predict the four continuous performance measures separately from the EO and EC states. Feature matrices X_{EO} and $X_{EC} \in R^{168\times N}$, N being the number of participants, are separately fed to NNs. Ten dataset permutations are executed, each consisting of leave-one-subject-out cross-validations (LOO-CV) among 10 participants for the CVSmean, HRTmean, and HRTvar measures, and 9 participants for CVSvar. S04 is removed from the CVSvar experiments since their score is more than 2 standard deviations larger than the average CVS variability. Feature matrices are standardized before being fed to the NNs that consisted of an input layer, one fully connected (FC) layer with hidden perceptrons, a rectified linear unit (ReLU) activation layer, and an output regression layer with the mean-squared-error loss function.

Experiments are run with different numbers of hidden units in the FC layer. Each network is initially tested with 40 hidden units with 1,000 epochs and a mini-batch size of 8. To tackle the overfitting problem, a validation patience scheme is utilized to stop the training if the validation loss did not improve after one epoch. Using Adam [165], a method for adaptive moment estimation as the optimization algorithm, a grid search is performed to optimize each loss function for 15 learning rates and 15 ℓ_2 regularization coefficients logarithmically increased within the $[10^{-5}, 10^{-1}]$ and [0.01,10] intervals, respectively. For each combination of the learning rate and ℓ_2 parameters, the network performance is assessed in each fold with the root-mean-squared error (RMSE) between the true and predicted outputs. The LOO-CV estimator is obtained by minimizing the average error across all folds and permutations [38], i.e.,

$$err_{LOO-CV}(lr,\lambda) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(y_n - f(X_n|D\backslash D_n))^2 \tag{3.4}$$

Here, M is the number of permutations, N is the number of validation folds, D and D_n denote the original sample set and the validation set from the n-th run, y_n and $X_n \in R^{168\times 1}$ represent the true label and feature vector for sample n, $f(X_n|D\backslash D_n)$ is the estimated output by the neural network, and lr and λ are the learning rate and the ℓ_2 regularization coefficients, respectively.

To study which features are given higher priority during the training and, subsequently, to perform supervised feature selection, the input weights of the first FC layer obtained during validation from the optimal pair of our hyperparameters are summed over all the U hidden units and averaged for all the N folds and M permutations. In other words,

$$\bar{W}_j(lr^*,\lambda^*) = \frac{1}{M}\cdot\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{i=1}^{U} W_{ij}(lr^*,\lambda^*), \tag{3.5}$$

where (lr^*,λ^*) represents the hyperparameters that minimized err_{LOO-CV}, and W_{ij} is the weight associated with unit i on the first hidden layer and feature j of the input vector. The averaged weight vector $\bar{W}_j \in R^{168\times 1}$ is subsequently visualized for feature selection.

3.3.6 Feature Relevance Analysis for Multivariate Prediction of SART Performance Measures

Sixteen prediction schemes are designed to predict the four continuous dependent variables from the BP-ROI and PSI datasets from EO and EC states. Single linear regression (SLR) models are developed for the initial feature selection: Each of the 168 BP-ROI features are standardized and individually entered in the model as an independent variable. Similarly, PSIs from each of the 2,016 unique electrode pairs from alpha, four beta sub-bands, wide-band beta, and wide-band gamma are individually entered in the model as the independent variables. After performing a leave-one-subject-out cross-validation (LOO-CV) across all participants, the R^2, adjusted R^2, RMSE, Pearson's linear correlation coefficient (r), and its corresponding p-value are calculated from the true and predicted outputs of these SLR models. Prior to these analyses, participant S04 was detected as an outlier and removed from the CVSvar experiments since the variability of their CVS scores was more than 2 standard deviations (SD) larger than the average CVS variability of all participants.
Once the n features whose individual regression models resulted in prediction correlations significant at the 0.1 level for each performance measure and feature set were detected, the possibility of predicting these measure from a group of such significant features is investigated. The MVPA is utilized to identify the most predictive feature subset, and multivariate linear regression (MLR) models are trained from all the 2^n-1 non-empty subsets [123]. To test the significance of each model, a null hypothesis of having no correlation between the true and predicted outputs is defined, and features are permuted 500 times across all participants. This results in obtaining the p-value for each prediction model having achieved a correlation higher than the original feature assignment. Finally, for each feature modality, subset, and performance measure, models with the most outstanding goodness-of-fit metrics are reported in Sections 3.4.3 and 3.4.5

3.4 Results

3.4.1 Behavioral Results

To visualize inter-individual differences in attention maintenance using a single a single time-series, Figure 3.5 shows the CVS curves for four participants: S03 with their balanced response style and consistent performance (CVsmean=0.48, CVSvar=0.05), S04

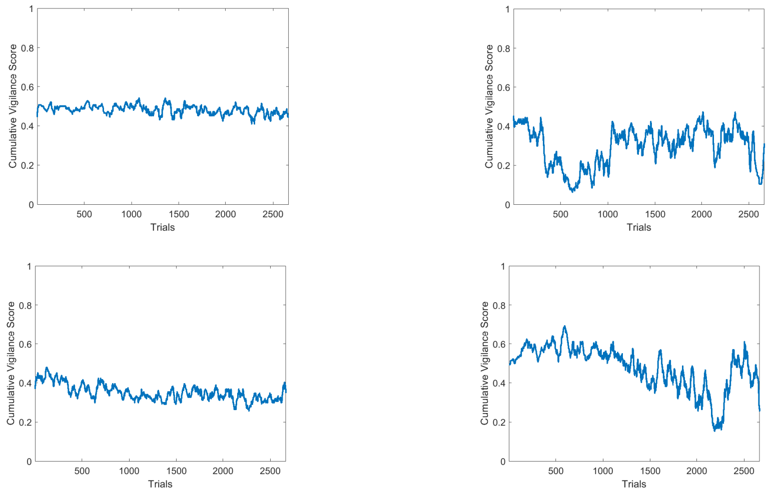


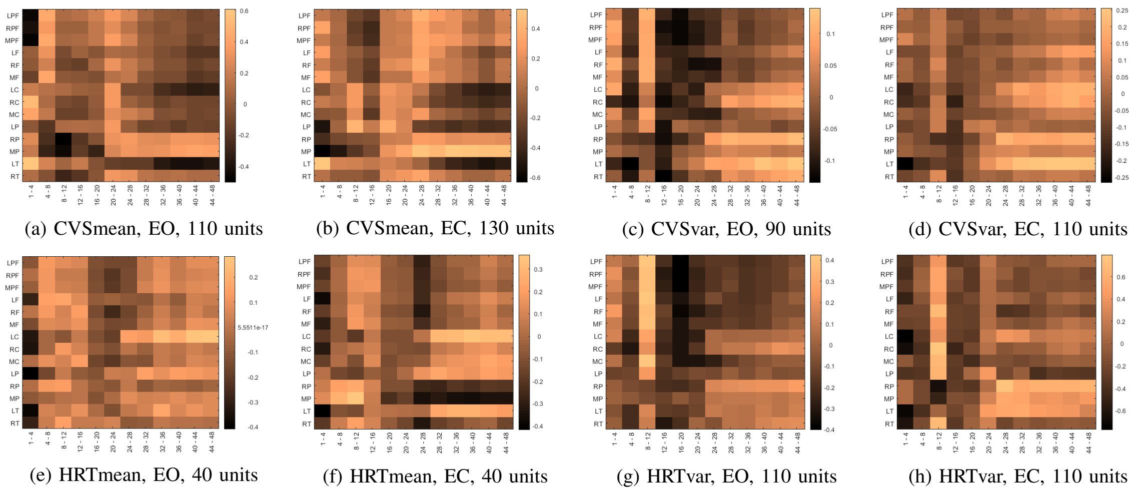
Figure 3.5: CVS curves for four participants (S03, S04, S06, and S10) demonstrating different patterns of maintaining tonic attention.

Table 3.1: Correlations among the overall behavioral measures of the fixed-sequence SART. *N*=10. *: *p*<0.05.

	CE%	OE%	CVSmean	CVSvar	HRTmean
OE%	0.80*				
CVSmean	-0.47	-0.68*			
CVSvar	0.90*	0.91*	-0.69*		
HRTmean	0.38	0.34	-0.88*	0.51	
HRTvar	0.80*	0.46	-0.17	0.61	0.24

who fell asleep very early in the experiment and recovered later (CEmean=0.30, CVS-var=0.33), S06 with their slow and conservative responses and a gradual attention decline in the second half of the experiment (HRTmean=583 ms), and S10 with an excellent performance in the beginning but the highest HRTvar (0.90) due to the extreme fatigue and drowsiness in blocks 9 and 10 before slightly recovering. Thus, contrary to a number of long experiments on vigilance estimation that divide the experiment intervals into three periods of high, middle, and low vigilance [140], these plots demonstrate vigilance levels can decline at any point during the experiment and be followed by a relative recovery. In fact, the majority of participants reported their alertness increased after a short, involuntary nap, indicating the brain's ability to regain its alertness after a period of idling.

Table 3.1 demonstrates correlations among the objective behavioral measures. Obtained *p*-values are corrected at the 0.05 level using the False Discovery Rate (FDR) method [166]. The number of CEs and OEs had a strong, positive linear association at 0.80. Furthermore, the average of overall CVS had a significant correlation with the number of missed trials while its variability was equally correlated with the percentage of both errors. Although HRTmean did not demonstrate any significant association with the number of errors, meaning that *fast or slow responses did not necessarily imply wrong responses*, it did have a strong and negative correlation with the average CVS. The variability of HRT was strongly associated with CE% while the variability of CVS and HRT fell short of being significantly at p<0.05. Therefore, it seems informative to analyze the variabilities of



(a) CVSmean, EO, 110 units (b) CVSmean, EC, 130 units (c) CVSvar, EO, 90 units (d) CVSvar, EC, 110 units

(e) HRTmean, EO, 40 units (f) HRTmean, EC, 40 units (g) HRTvar, EO, 110 units (h) HRTvar, EC, 110 units

Figure 3.6: Heatmaps for 168-d weight vectors averaged across 10 runs of one-fully-connected layer NNs with various numbers of hidden units resulting in the minimum CV error. Captions indicate the resting state and number of hidden units.

CVS and response time in more depth due to their large correlations with the number of errors.

3.4.2 Band-Power Feature Detection using Neural Networks

LOO-CV was performed to train one-layer NNs with normalized BP-ROI features for prediction of the four overall performance measures. The pair of best regularization coefficient and learning rate for each network was obtained and W_{ij}, the weight associated with unit i and feature j, was averaged over all the units to obtain $\bar{W}_j$. Figure 3.6 presents heat maps for these weights averaged for all the folds and runs. The light and dark cells, respectively, denote positive and negative signs of these spatiospectral features in the CV MLR models.

The obtained weights demonstrate pre-frontal delta predicts lower CVS mean – due to higher CEs and OEs – and more inconsistent CVS during EO while increase in the left temporal delta has the opposite effect. Pre-frontal delta predicts faster responses and more RT variability during EC – a sign of hyperactivity and impulsivity observed due to keeping extremely fast responses in the analysis. Frontal and central alpha from EO are correlates of more variability in both scores and response time while pre-frontal and frontal theta predicts slower responses as well as better and more consistent CVS.

Heat maps also show beta-1 oscillations (12-16 Hz) are generally similar to alpha in predicting slower responses and lower average CVS scores. But for predicting lower variability (more consistency) in HRT and CVS, they are similar to the 16-24 Hz oscillations from the frontal region. During EO and EC, larger ratios of gamma from right and especially midline parietal regions are predictors of faster responses and higher CVSmean, and during the EC, predictors of shorter response time in average. The left temporal (upper) gamma, on the other hand, is similar to the central gamma in predicting lower scores

Table 3.2: LOO-CV-based feature relevance analysis for MLR to predict the mean and variability of CVS and HRT from EO and EC BP-ROI features. Statistical measures are reported for the best models of subset sizes with the highest adjusted R^2, highest r, or lowest RMSE. If more than one subset satisfied these conditions, all of the best subsets are displayed. ***: p <0.001, **: p <0.01, *: p <0.05.

	BP-ROI Features, EO						BP-ROI Features, EC					
Measure	No. Features	No. Subsets	R^2	Adj. R^2	RMSE	Corr. Rho	No. Features	No. Subsets	R^2	Adj. R^2	RMSE	Corr. Rho
CVSmean	3	20	0.91	0.87	0.02	0.96***	2	10	0.83	0.78	0.03	0.92***
CVSvar	1	4	0.35	0.25	0.06	0.62**						
	2	6	0.42	0.23	0.05	0.69**	2	21	0.68	0.57	0.04	0.83***
HRTmean	8	495	1.00	1.00	0.39 ms	1.00***	3	20	0.83	0.74	37.77 ms	0.92***
HRTvar	2	28	0.36	0.18	0.15	0.64*	2	3	0.14	-0.11	0.17	0.49*

and less performance consistency from both EO and EC, and slower responses and HRT variability from EO features.

3.4.3 Band-Power Feature Relevance Analysis and Regression

Table 3.2 presents the goodness-of-fits from cross-validated predictions for the best subsets in each performance measure-resting state scheme. Figure 3.7 shows scatter plots of the true and predicted values for these best models as well as the topographic plots of significant predictors. MVPA results indicate that feature relevance analysis is efficient in obtaining a small number of features for highly significant prediction models, and that their signs are consistent with those from the averaged NN weights in Figure 3.6.

3.4.4 Phase Synchrony Correlates of Performance Measures

Topographic plots of Figures 3.8 and 3.9 demonstrate statistically significant linear correlations between the overall performance measures and the resting-state PSIs from different frequency bands during the EO and EC sessions, respectively, with red and blue lines indicating significantly positive and negative correlations at the 0.05 level. CEs and OEs are also analyzed here to enable better differentiation between the correlates of commission and inhibition errors. A two-way analysis of variance (ANOVA) reveals significant effects of frequency bands and electrode pairs on the correlations of EO and EC PSIs with all the six performance measures, $p < 0.001$. Two-sample t-tests show that, except for the omission error correlations in which significant differences exist between the average values –and not patterns– of correlations with lower-beta bands, $p < 0.05$, no significant differences are found among the mean correlations of lower-beta sub-bands with mid-beta, $p > 0.2$, and with mid-beta and upper beta, $p > 0.5$, during the EO sessions. For the EC condition, two-sample t-tests show no significant differences between the mean correlations of lower beta bands, $p > 0.3$. However, significant differences are observed between correlations of the mid-beta and upper beta bands, $p < 0.05$, except for the HRT mean and HRT variability. We thus group similar beta sub-bands together, calculate their average PSIs, and obtain their correlations with the overall performance measures.

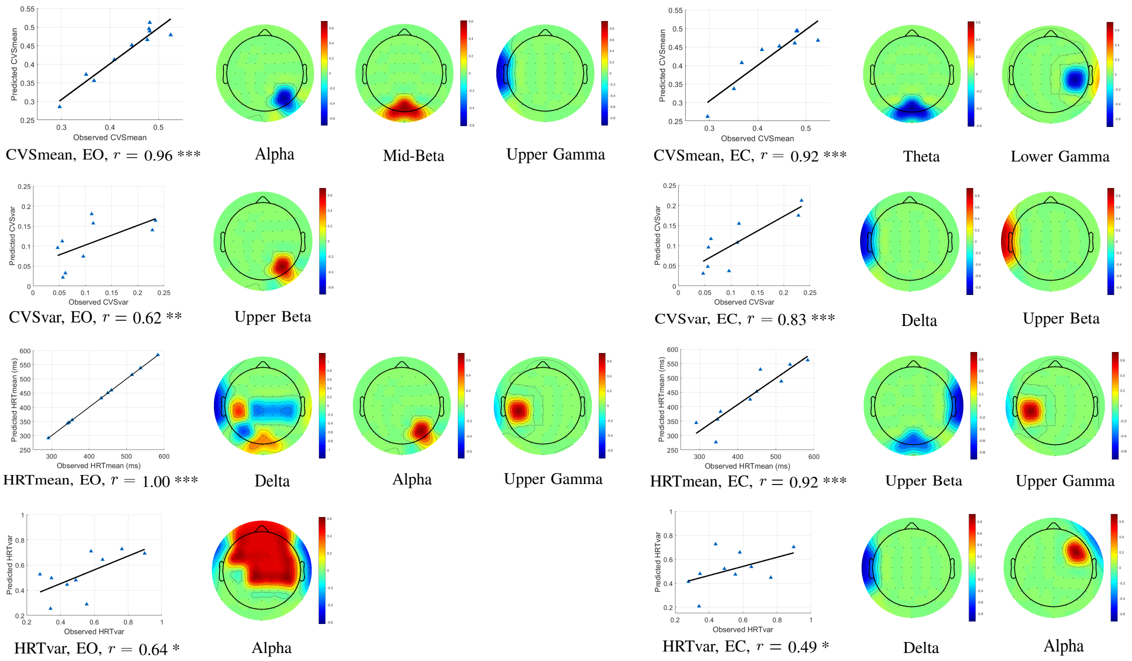
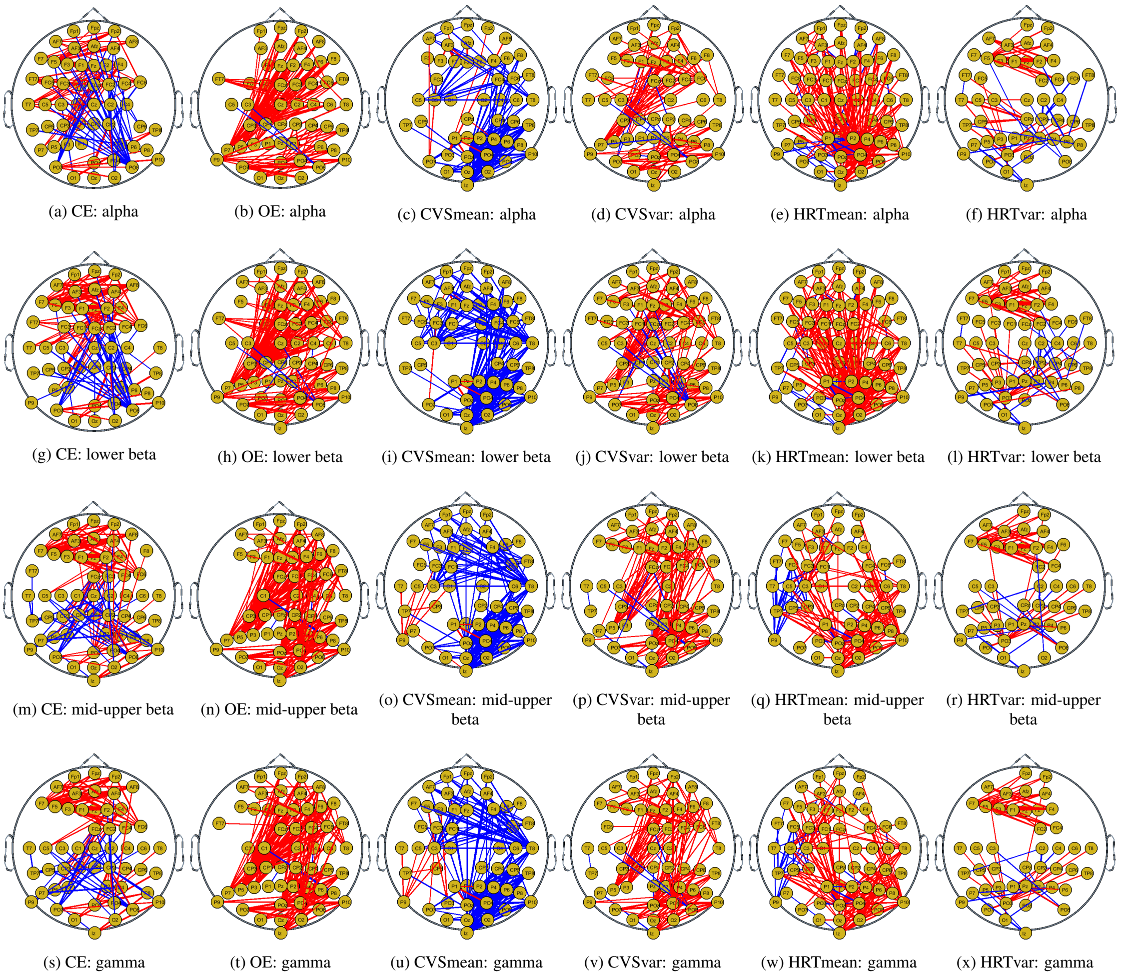


Figure 3.7: Scatter plots for predicted-vs-true performance measures from LOO-CV MLR models of Table 3.2 with the highest adjusted R^2. Polar map distributions show weights of significant BP-ROI predictors.

As can be seen from these plots, clear differences exist in the correlation patterns involving channels from the right and left posterior regions. Increase in the synchrony of right parieto-occipital channels with left frontal, central, and temporal regions are significant correlates of fewer commission errors in all bands of EO and EC states, fewer errors of omission –scarce in EO, all but the upper beta in the EC state, and less variability of response time from mid-beta to upper gamma during the EC recordings. Thus, long-range connections should be established between the fronto-parietal networks for information to effectively flow between the visual cortex and regions controlling the attentional processes. This pattern could be linked to the right lateralization of sustained attention processes during modulation of selective responses to target stimuli [150]–[153]. This is while an increase in synchrony within the right parietal-occipital cortex and long-range synchronization between the left posterior with frontal channels are significant correlates of more commission and omission errors.

Furthermore, an increase in the inter-hemispheric synchrony in the frontal and pre-frontal regions and decrease in phase coherency of parieto-occipital with midline and opposite frontal (as well as fronto-central and centro-temporal) during EO sessions are correlated with more commission and omission errors. The former can be linked to the deactivation and idling of cortical networks in the frontal lobe while the latter is a sign of decreased information flow between the visual cortex and attention regions in the front. The patterns of increased inter-hemispheric synchrony are very dense in EO recordings.

Consequently, reductions in synchrony of interhemispheric frontal alpha to mid-beta and reductions in their synchrony within the right and midline parietal to occipital are asso-



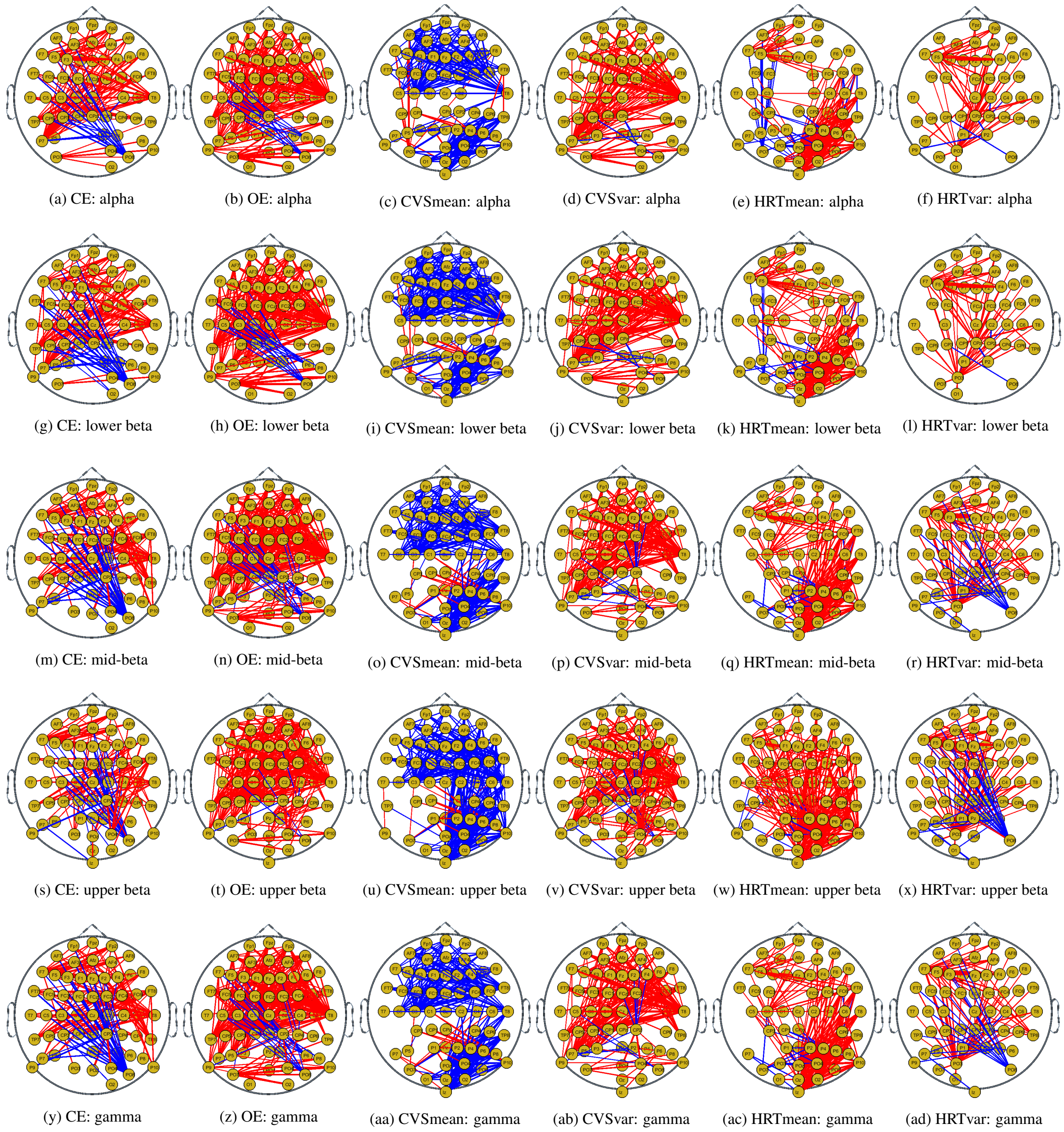
(a) CE: alpha (b) OE: alpha (c) CVSmean: alpha (d) CVSvar: alpha (e) HRTmean: alpha (f) HRTvar: alpha

(g) CE: lower beta (h) OE: lower beta (i) CVSmean: lower beta (j) CVSvar: lower beta (k) HRTmean: lower beta (l) HRTvar: lower beta

(m) CE: mid-upper beta (n) OE: mid-upper beta (o) CVSmean: mid-upper beta (p) CVSvar: mid-upper beta (q) HRTmean: mid-upper beta (r) HRTvar: mid-upper beta

(s) CE: gamma (t) OE: gamma (u) CVSmean: gamma (v) CVSvar: gamma (w) HRTmean: gamma (x) HRTvar: gamma

Figure 3.8: Significant correlations of PSIs during the EO resting-state with six overall performance measures, $p < 0.05$. Red and blue lines demonstrate the significantly positive and negative correlations, respectively. Line widths are proportional to the absolute value of correlation coefficients. Rows from top to bottom: Alpha (8-12 Hz), lower beta (12-20 Hz), mid- and upper beta (20-28 Hz), and gamma (31-60 Hz). Columns: CE%, OE%, average CVS, variability of CVS, average HRT, and variability of HRT.



(a) CE: alpha (b) OE: alpha (c) CVSmean: alpha (d) CVSvar: alpha (e) HRTmean: alpha (f) HRTvar: alpha

(g) CE: lower beta (h) OE: lower beta (i) CVSmean: lower beta (j) CVSvar: lower beta (k) HRTmean: lower beta (l) HRTvar: lower beta

(m) CE: mid-beta (n) OE: mid-beta (o) CVSmean: mid-beta (p) CVSvar: mid-beta (q) HRTmean: mid-beta (r) HRTvar: mid-beta

(s) CE: upper beta (t) OE: upper beta (u) CVSmean: upper beta (v) CVSvar: upper beta (w) HRTmean: upper beta (x) HRTvar: upper beta

(y) CE: gamma (z) OE: gamma (aa) CVSmean: gamma (ab) CVSvar: gamma (ac) HRTmean: gamma (ad) HRTvar: gamma

Figure 3.9: Significant correlations of PSIs during the EC resting-state with six overall performance measures, $p < 0.05$. Red and blue lines demonstrate the significantly positive and negative correlations, respectively. Line widths are proportional to the absolute value of correlation coefficients. Rows from top to bottom: Alpha (8-12 Hz), lower beta (12-20 Hz), mid-beta (20-24 Hz), upper beta (24-8 Hz), and gamma (31-60 Hz). Columns: CE%, OE%, average CVS, variability of CVS, average HRT, and variability of HRT.

Table 3.3: LOO-CV-based feature relevance analysis for MLR to predict the mean and variability of CVS and HRT from EO and EC PSI features. For the n initially selected features for each performance measure and each feature set, all the 2^n-1 non-empty subsets were individually analyzed. Statistical measures are reported for the best models of subset sizes with the highest adjusted R^2, highest correlation r, or lowest RMSE (ms for HRTmean). If more than one subset satisfied these conditions, all of the best subsets are displayed. **: $p < 0.001$, *: $p < 0.01$.

Measure	PSI Features, EO						PSI Features, EC					
Measure	No. Features	No. Subsets	R^2	Adj. R^2	RMSE	Corr. Rho	No. Features	No. Subsets	R^2	Adj. R^2	RMSE	Corr. Rho
CVSmean	2	36	0.90	0.87	0.02	0.95**	8	45	1.00	0.97	0.00	1.00*
	5	126	0.92	0.83	0.02	0.97**						
CVSvar	3	84	0.99	0.99	0.01	1.00**	1	8	0.93	0.92	0.02	0.96**
							7	8	0.98	0.87	0.01	0.99*
HRTmean	4	126	0.95	0.92	19.48	0.98**	7	120	0.99	0.97	7.88	1.00*
HRTvar	8	165	1.00	1.00	0.00	1.00**	2	45	0.92	0.89	0.05	0.96
							6	210	0.96	0.87	0.04	0.98**

ciated with better average CVS. Connectivity plots of EC PSIs with CVSmean are more dense and significant that the ones from EO. For the upper beta during EC, reductions in the synchrony of midline parietal with frontal and central channels show significant associations as well. For EO gamma band, increase in the synchrony of left parietal with left fronto-central also correlates with higher mean CVS. When considering the synchrony measures in correlation with CVS variability, almost opposite patterns of those mentioned for average CVS are found to be significant. However, especially during the EC conditions, higher synchrony between the left parietal channels with the frontal cortex is a significant correlate of more vigilance inconsistency.

Faster responses are correlated with smaller synchrony of 8 to 28 Hz sub-bands within the right and midline occipital and parietal cortex, between two hemispheres' centro-parietal channels, and between midline occipital with right central and parietal cortex during the EC sessions. For the pre-task EO recordings, smaller alpha and lower beta synchrony between POz with right and left frontal cortex, and higher gamma synchrony between the left centro-parietal channels with right and left frontal and fronto-central cortex are significant correlates of shorter response time.

Finally, correlation patterns of PSIs with response time variability are much more scattered than other performance measures: For the alpha to mid-beta bands, reduction in the frontal inter-hemispheric synchrony during both eyes open and closed sessions, reduction in their synchrony from PO3 and CPz with the right and left frontal and central cortex, and increase in their mid-beta and upper beta synchrony from right parieto-occipital (PO8) with the left frontal channels during the EC sessions are correlates of more consistent response speeds.

3.4.5 Phase Synchrony Feature Relevance Analysis and Regression

Considering these linear associations, the resting-state phase synchrony features are used to predict the mean and variation of the cumulative vigilance score and response time in long SART. As explained in Section 3.3.6, single linear LOO-CV models are developed

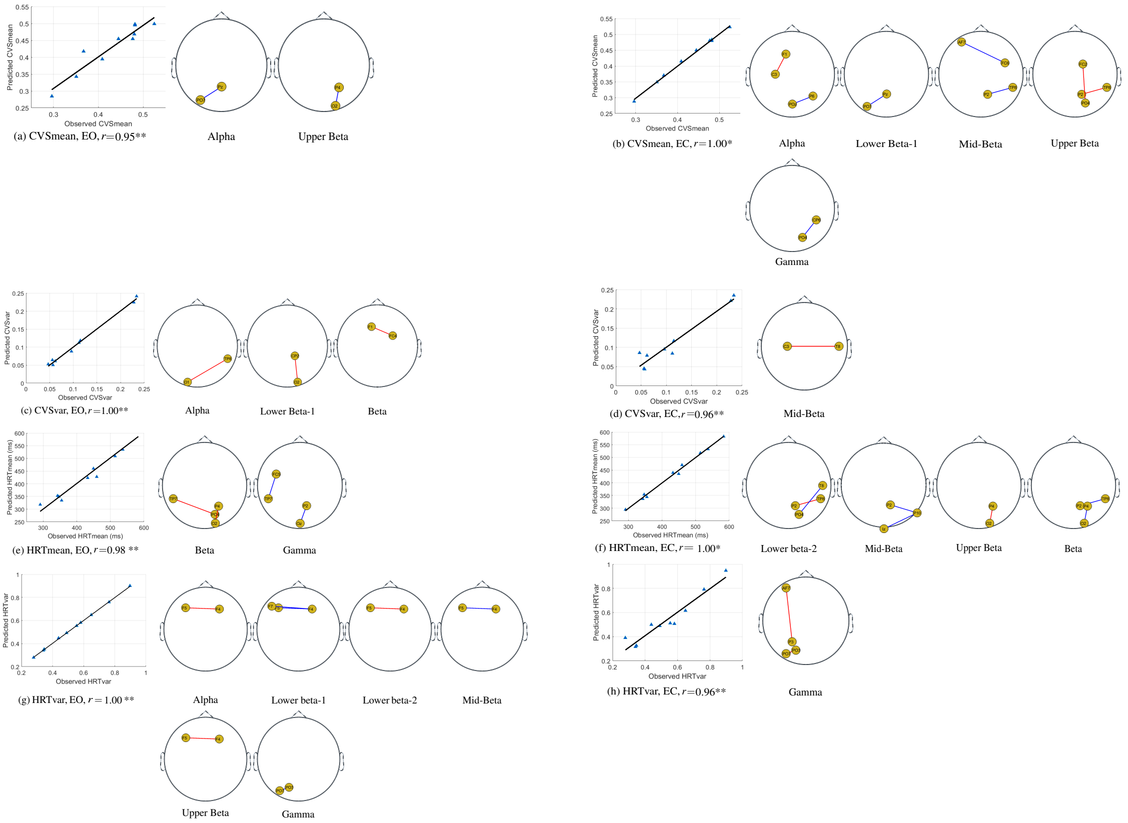


Figure 3.10: Scatter plots for predicted versus true performance measures from the LOO-CV multivariate linear regression models reported in Table 3.3 with the highest adjusted R^2. The connectivity plots demonstrate the significant PSI features decomposed in electrode pairs from different frequency bands. Red and blue links denote the positive and negative estimated weights, respectively, in the cross-validated regression models.

and, based on the statistical significance of the obtained models, all the non-empty subsets of the selected features are used for developing multivariate linear regression models. All features are standardized before entering the models. For each of the eight pairs of performance measure and resting-state conditions, Table 3.3 demonstrates the goodness-of-fit measures –from the cross-validation predictions versus true values – for the best subsets of the entire subset sizes.

Figure 3.10 visualizes the significant channel pairs whose phase synchrony results in significant predictors for the continuous performance measures. Red and blue links indicate the positive and negative estimated weights in the cross-validated regression models, respectively. Also demonstrated are the scatter plots of the true and predicted values for the best PSI-based models in Table 3.3 which demonstrate the superiority of phase synchrony features in predicting the overall performance measures in comparison to the performance of BP-ROI features reported in Table 3.2. In other words, models built using the pairwise synchrony features result in improved adjusted R^2 and lower RMSEs compared to the ones built from BP-ROI features. The only exception is the model of average response time from eyes-open BP-ROI features which resulted in a surprisingly low RMSE of 0.39 ms compared to the 19.48 ms of the EO PSI set. This performance is especially of interest since, in the case of HRT variability, the EO and EC band-power features failed to result in significant regression models. Furthermore, from significant predictors shown in Figure 3.10, it can be understood that PSIs from a maximum of four channel pairs can accurately and significantly predict the CVSmean and HRTmean (from eyes-open PSIs), CVSvar (from both eyes-open and closed PSIs), and HRTvar (from eyes-closed features). Thus, it is safe to claim that reductions in short-range or regional synchronizations are significant predictors of higher scores and faster responses, while inter-hemispheric and long-range synchronies are better predictors for the variability or inconsistency in performance and response time of this long, uni-mode SART paradigm.

3.5 Discussion

3.5.1 Labeling and CVS Validation

The proposed CVS is based on several well-established observations regarding the type and number of errors, length of RT, and their variability. CE, OE, and RT parameters have been used as the quantitative measures of sustained attention and response inhibition failure for healthy and patient groups [66], [92], [93], [144], [154]. In [167], SART accuracy in the proposed Gaussian linear mixture model (GLMM) is built as a function of RT mean and variance and a variety of psychological and physiological factors. The labeling strategy in [111] is based on the number of mistakes during a driving experiment. Finally, in [168], good and poor driving performance are determined based on RT thresholding.

In our paradigm, recorded videos were carefully assessed to validate the proposed CVS, and it was observed that intervals representing slow and sharp declines in the CVS curves indeed matched video frames with increased eye closures. Furthermore, the periods corresponding to global minima in the CVS curves for two participants (S04 and S10) matched their extreme head tilts and deep sleep intervals.

3.5.2 Roles of Delta and Theta Ratios

Higher ratios of EO delta from the left frontal and temporal regions predict faster responses and less RT variability, and point to the role of delta oscillations in improving the “Go stimulus-responses” [126] and suppression of the irrelevant stimuli. In a much shorter task of auditory Go/NoGo and after rejection of erroneous and extremely fast trials, increase in task-related delta (1-3 Hz) with respect to the EO condition was correlated with higher OEs and more standard deviation of RT [126]. The latter finding should not be interpreted as being different from our prediction results as we wanted to model such impulsivities through HRT and CVS.

NN heat maps show that more theta activity from pre-frontal to central regions is correlated with more consistency in CVS and HRT and, in the case of EO theta, with better CVSmean. However, none of these correlates appear in the best LOO-CV models of Figure 3.7. Interestingly, higher frontal and right parietal theta during EO are positively associated with slower RT. Similarly, in a study on teenage ADHD and control groups, higher theta power from the posterior and left frontal cortex had positive correlations with longer reaction time in the control group [169].

3.5.3 Roles of Frontal and Parietal Alpha

Increase in alpha ratios – more synchronization – in frontal and central regions is a correlate of lower CVS and higher variabilities in CVS and HRT. This is in line with the effect of smaller midline alpha during a visual conjunctive continuous performance task [170]. We also observe a close relationship between impaired visual attention and long-duration task-induced mental fatigue. Increase in alpha powers in both occipital and parietal regions –Brodmann's areas 18, 19, and 37– are linked with longer RT and lower CVSmean in this study and a simulated driving experiment [171]. Thus, participants who demonstrate clear patterns of *maintaining* their vigilance scores are able to regulate and desynchronize their parietal alpha powers to block attentional drifts and stay on the task despite its monotonous nature [35], [66].

3.5.4 Opposite Roles of Beta Sub-bands in Task Consistency

Temporal and parietal beta show differential hemispheric activities during emotional and cognitive processes [172]. Figure 3.6 shows beta-1 (12-16 Hz) is more similar to lower frequencies in being associated with slower responses while beta-2 and mid-beta (16-24 Hz) are correlates of improved and consistent performance and faster HRT. The latter is consistent with increased frontal beta being associated with lower CPT score variability of the ADHD group [173]. However, to the best of our knowledge, our finding on decreased RT variability from parietal low beta (12-16 Hz) is not reported in the literature.

3.5.5 Fronto-parietal Gamma Predicts Better Task Consistency

Increase in midline parieto-occipital gamma during EO and EC is a predictor of higher CVSmean and lower CVSvar and faster responses in the EC states. These findings fill the gap in resting-state correlates of upper beta and gamma bands for Go/NoGo stimuli selection and fatigue. Furthermore, our observations on higher pre-frontal gamma being correlated with fewer CEs and lower performance variability is in line with increased fronto-parietal gamma reported in experienced meditators [174]. Difference in gamma powers of two fronto-parietal networks was used to predict performance variations in an SMR-BCI task [99] and linked with the association of attentional shifts and gamma oscillations [175].

3.5.6 Temporal Gamma Predicts Lower Task Consistency

Increase in the left central and temporal gamma and upper beta during rest predicts slower RT, lower CVSmean, and more CVSvar. This disability in sustaining attention can be explained by differences between the high- and low-attention networks at temporal regions (Brodmann's areas 35 and 36), and the function of default mode network (DMN). DMN is active during wakeful resting states accompanied by daydreaming, and deactivated while attending to specific events and tasks [176]. Abnormal DMN activation is observed in individuals with depression, anxiety, schizophrenia, ADHD, Alzheimer's disease, and amyotrophic lateral sclerosis (ALS) [99], [176]. Controlling DMN activity through practicing meditation can improve attention in motor-imagery BCI [177] and be used in EEG-based assistive technologies [99].

3.5.7 Role of Short-Range and Long-range Connectivity

Studies on neural correlates of fatigue and drowsiness generally focus on the functional connectivity measures such as the directed (transfer) entropy or partial directed coherence from MRI and EEG recordings, and use these features as well as the direction of information flow and increase in the oxygenation levels to classify the alert or wakeful versus fatigued states [151], [156], [178], [179]. Few studies have looked at the correlations of pre-task resting states dynamics with performance in an attention task [124] and a learning task [180]. One study has also identified the resting networks responsible for information transfer during three different cognitive tasks [181]. They discovered that regions within the pre-frontal cortex and the left-lateralized default model network (DMN) were responsible for logic-rule task while successful sensory-rule mappings were related to information transfer between the visual and dorsal attention regions.

From our regression results, reductions in short-range and regional synchronization within the posterior cortex predicted higher scores and faster responses. But the variability of CVS and response time were generally predicted by frontal inter-hemispheric or fronto-parietal connections.

The topographic plots of Figures 3.8 and 3.9 show that reduction in synchrony of the left parietal channels with right and midline frontal cortex, lower synchronization within the parietal cortex, and smaller long-range frontal-parietal synchrony in gamma oscillations are correlates of fewer omission errors. The lower synchrony within the parietal cortex was a significant predictor of improved CVS mean, faster responses, and more consistent response time in the scatter plots of Figure 3.10. These observations regarding the improved short-range and high frequency synchronization inside the posterior region can be explained with the concept of promoted gamma oscillations inside the task-related areas during sustained-attention experiments [153]. This is while the right and left tempo-parietal channels during the EO state should be more synchronized with the opposite frontal and central channels as correlates of fewer commission errors. The EC patterns show that the smaller synchrony of left parietal channels with the pre-frontal cortex is a correlate of more consistency in CVS while the EC signals in PO8 needs to have stronger synchronies with the left frontal and central channels for more consistent response speeds and smaller number of commission and omission errors or better target/non-target detection. This observation is in agreement with the findings of Choi *et al.* during an easy auditory oddball task who explained the enhancement of long-range gamma connectivity was linked to “matching” of the working memory contents with the actual stimuli which is meditated by attentional resources [182]. Similarly, in an experiment with monkeys and during a pre-stimulus delay after a cue signal, long-range gamma (35-60 Hz) coherence was enhanced between the posterior parietal cortex, visual cortex, and frontal eye field while the 5 to 15 Hz coherence was suppressed between the poster pariteal and visual cortex [183]. Likewise, the EO connectivity plots of Figures 3.8 and 3.9 show that increase in short-range alpha synchrony within the posterior region is a correlate of more errors, higher CVS variability, and smaller CVSmean.

Finally, the regression results in Figure 3.10 indicate that increase in pre-task interhemispheric synchrony in the alpha and beta-bands predict more variability of CVS and response time. In an analysis of alertness and fatigue in a long-term cognitive task, it was observed that beta-band coherence and PLVs decreased in the central and parietal cortex had significantly decreased in the post-task, fatigued states with respect to the pre-task recordings [184], suggesting the role of fatigue in reduction of cooperative processing and functional coupling. In a more recent study on vigilance characterization during an air traffic controller task, network parameters for dynamic partial directed coherence (PDC) of the alpha band from the midline fronto-central, right fronto-central, and right parieto-occipital regions could significantly distinguish between the positive and negative vigilance levels [185]. Their results confirm the significant and undirected PSI weights of alpha band from the fronto-central and parieto-occipital regions in Figure 3.10. Finally, the beta band PLIs had the best performance in distinguishing the alert versus fatigue states from the beginning and end of a long driving tasks [157]. This result agrees with our findings where various beta band subsets were selected as predictors for all the four measures of sustained attention task.

3.6 Conclusion

In this work we have, for the first time, established the phase synchrony correlates of continuous-valued performance measures for average and variability of vigilance score and response time from the resting-state EEG, and identified the multivariate predictors of these measures from band-power and phase synchrony features. Our performance measures are completely objective which improves the applicability of our work in reality while several other studies look into classification of high versus low attention levels merely based on their experimental design [157]. Our results indicate that higher interhemispheric phase synchronies especially in the frontal and central regions are predictors of more variability or inconsistency in the vigilance score and reaction time. This work provides a clear picture on the role of resting-state dynamics in individual traits enabling the maintenance of consistent performance in a highly repetitive and uniform task. This study can be used to obtain markers before calibration sessions of BCI experiments to indicate the need for subsequent adaptation in the interface environment and classifier parameters.

The flexible selection of information is served by attention in association with current behavioral goals. In the visual modality, this neuro-cortical function is a critical component of cognition and a distinguishing feature of adaptive behavior. In psychological literature, 'sustained attention' has been characterized by individual readiness to detect sudden and/or unanticipated stimuli throughout a time period [153]. Visual attention facilitates processing of the visual inputs such as the stimuli that are physically more prominent (exogenous attention) and/or relevant to behavioral goals based on the specified motivation (endogenous attention), whereas the irrelevant stimuli are filtered out as an intrinsic response. In particular, endogenous attention is accomplished through top-down feedback from frontal and parietal areas [181]. As well, neuro-imaging research studies show that activation of frontal and parietal lobes (mostly in the right hemisphere) are relatively more associated with sustained attention performance [186]. Our results are highly compatible with those previous findings.

4 Deep Neural Networks for Vigilance Prediction from Pre-Trial Spatio-Spectral Features

In the previous chapter, we used the collective spatio-spectral features and neural dynamics of resting-state brain networks to predict the average and variability of vigilance scores and response time in a subsequently executed long sustained attention task. The resting-state networks are generally considered to reflect the intrinsic state of the brain and act as the baseline for subsequent cognitive and sensorimotor tasks, and understanding their dynamics in participants undertaking BCI calibration provides a clear picture of the need for assistive and adaptive technologies. However, the need for adaptation over time should be evaluated in a continuous manner during the execution of mental tasks. In systems used for monitoring of operators' performance, cognitive computing, and physical rehabilitation, detection of drowsy periods and intervals of low vigilance are of utmost interest for subsequent adaptation of experimental paradigms to the users' cognitive states. Focusing on the little-explored world of pairwise phase connectivity in classification and detection of drowsy states, in this chapter we provide an in-depth spatio-spectro-temporal analysis followed by the classification of vigilance states from the band-power (BP) and phase-locking value (PLV) features of EEG measurements during our long Sustained Attention to Response Task (SART).

In the first part of this chapter, Experiment 1, we report the power of utilizing averaged pre-trial phase synchrony indices (PSIs) and deep neural networks (DNNs) for estimating the block-wise vigilance and response time of all participants in a cross-validated, regression setting. The aforementioned experiment blocks last for approximately 8 minutes. Neural correlates of sustained performance and response time in these blocks represent the collective correlates of *tonic* vigilance variations from alpha and lower beta bands. This work has been presented in [36] and cited as one of the novel brain connectivity studies for vigilance prediction in [187].

In the second part of this chapter, Experiment 2, with a shift of focus to the performance of each individual separately, we analyze the *phasic* correlates of attention and classify the highest and lowest scoring vigilance trials from their immediate PLV features in the pre-trial and early onset intervals of the entire 105-minute experiments. A three-layer convolutional neural network (CNN) is used to extract spatial information from visual patterns of pairwise, symmetric PLV matrices. Classification results from BP features assembled in 14 regions of interest (ROI) act as the baseline for CNN outcomes. Finally, we present the most commonly distinguishing features of these phasic vigilance states across the entire participants. Unlike the work reported in [188], our labeling strategy does not require a visual scanning of hit and miss intervals to detect unresponsiveness, neither does it assume uniform alert and drowsiness states in the beginning and end of the task sessions as done in [110], [151], [157]. This work demonstrates the importance of considering individual psychophysiological differences for modeling and detecting the extremely alert and drowsy trials in long and monotonous experiments.

4.1 Motivation

Monitoring human alertness has been a subject of interest with several applications in driving simulations, operation of critical machinery, and tasks concerned with learning and memory. Increase in mental fatigue can demonstrate itself as slower reaction time while an occurrence of microsleep episodes, defined as the "complete and unintentional sleep-related losses of consciousness" up to 30 seconds, can be potentially hazardous during the operation of critical machinery [189], [190]. Detection of responsive versus microsleep events or long lapses of attention has attracted interest in recent years [191]. However, traditional studies on sleep and drowsiness detection had not utilized inter-channel or pairwise relationships of EEG features [192]. A few studies focused on using variations of common spatial pattern (CSP) filtering such as the regularized spatio-temporal filtering and classification (RSTFC) [193] to optimize temporal and spatial filters for maximizing the separability of microsleep and responsive classes. However, although the structural and functional connectivity differences between the alert and drowsy states are well documented in neuroimaging studies [156], [194], few studies have so far attempted to *predict* and *classify* these states using connectivity and pairwise measures.
The widespread and successful application of DNNs in many fields including computer vision and speech processing has brought a new generation of systems using EEG data for sleep stage classification, epileptic seizure detection, affective computing, and the like. A subset of these studies also focused on *visualizing* the nature of features that the CNNs and deep belief networks (DBNs) learn from multi-class EEG-based BCI datasets [26], [27]. In the context of DNNs, two-dimensional kernels of convolutional blocks could partially extract spatial information from multi-channel raw inputs, but their receptive field is limited to the nearby channels – sorted according to the EEG electrode numbering– due to the usually small dimension of utilized kernels. Notable solutions to tackle this limitation include using larger kernels in the shallow and deep ConvNet architectures [26], applying a sequence of temporal and depthwise convolutions in EEGNet for learning the frequency-specific spatial filters [27], designing spatial and temporal recurrent layers for horizontal and vertical scanning of channels [195], and using convolutional layers after meshes of scalp channels [196]. More recently, Zhang *et al.* used 20-frame sequences from 3D EEG cubes as the inputs to a deep CNN followed by a deep recurrent neural network (RNN) for classification of low versus high mental workload in a 70-minute sessions of spatial n-back and arithmetic sessions [197].

In this chapter, we address the interesting view points and shortcomings of both approaches, and predict the occurrence of alert and drowsy trials during a long Go/NoGo task using phase synchrony values [128], [129] extracted from pre-stimulus EEG signals. The novelties and contributions of our work can be listed as below.

1. We focus on extracting EEG-based features from pre-trial intervals since the alertness during the expectation state is shown to be associated with performance in visual oddball [64], motor imagery [198], and motor learning [123] tasks.

2. We identify the common pre-trial, band-power correlates of variations in vigilance level among all participants, and use them in Chapters 5 and 6 for prediction of continuous and discrete vigilance levels in an unsupervised manner when an objective and reliable ground truth does not exist.

3. We rely on the benefits of PLV in demonstrating different functional features between alert and drowsy states, and combine them with the learning power of DNNs and CNNs in extracting spatial information from visual patterns of pairwise, symmetric matrices.

   To the best of our knowledge, so far only one study has used deep neural networks for attention/no-attention classification of EEG signals from a non-driving and non-sleep scenario [199], and even they have used resting-state recordings for the no-attention class. Furthermore, although the phase synchrony index has been used for detection of microsleep versus response trials from a simulated driving task [191], the utilized features did not involve matrices that show larger spatial regions for synchronization or desynchronization. Our input matrices and proposed networks are considerably simpler compared to the extremely deep solutions recently proposed to tackle the limited receptive field of CNNs [26], [27], [195], [196] and the 3D convolutional-recurrent network architecture of [197] that is applied to the sequential frames of 3D EEG cubes.

4. Through our novel trial vigilance score (TVS), we appreciate and emphasize inter-individual differences in execution of long and monotonous Go/NoGo tasks [52]. The ground truths for vigilance scores are individually customized according to the fastest response time of initial trials, and then averaged over specific length windows. Furthermore, trials belonging to the high and low alert classes are extracted from the extreme tails of the score histograms and are *not* assumed to belong to the immediately initial or final experimental blocks.

The rest of this chapter is organized as follows. Section 4.2 presents a summary of the related work. Section 4.3 focuses on the calculation of adaptive vigilance scores as well as cross-subject regression results for 8-min blocks using the PSI features. Section 4.4 then presents details of dataset construction using BP-ROI and PLV features, introduces the proposed CNN architecture, and discusses common and individual correlates of high and low vigilance levels as well as the classification rates among different features. This chapter is concluded in Section 4.5 with a summary of important findings and implications.

4.2 Related Work

As one of the most established studies on the role of alpha oscillations in memory and cognitive capabilities, Klimesch *et al.* discussed that participants are at a state of expectancy between occurrence of a warning sign and appearance of a target or non-target visual stimulus [64]. When participants are expecting a target, a significant reduction in the lower alpha power – event-related desynchronization (ERD) – is observed that reflects their high alertness level. Furthermore, the power of lower alpha-2 oscillations decreases one second before the onset of any type of stimulus, thus reflecting a general expectancy state. Later on and if targets do appear, participants demonstrate larger reduction in lower alpha-2 power. Ozdenizci *et al.* [123] also reported large reduction in pre-trial fronto-parietal beta (15-30 Hz) power in a motor learning experiment. Furthermore, Bamdadian *et al.* reported that the ratio of theta to the total alpha and beta powers from the 2-second pre-trial interval was correlated with the classification accuracy of performing motor imagery versus mental counting trials [198].

In addition to studies on sleep level classification, detection of mental fatigue and assessment of drowsiness levels are usually carried out during simulated or real-world driving experiments [13], [156], [168], [192], [200]–[202]. The ground truth in these experiments is either obtained from averaging the number of errors or accidents, fixed thresholding on the reaction time, or using correlates of sleepiness from facial features such as variations in the eye blink rates and increase in the duration of eye closures. In the same context, a growing number of papers have been published on regression methods for vigilance or drowsiness estimation, usually from both electrooculography (EOG) and EEG recordings [111], [139], [203]–[206]. Less frequently, EEG signals from resting-state sessions recorded immediately before and after a series of cognitive tasks are used for drowsy versus alert state classification and regression [52], [110]. In another group of papers to which this chapter's studies belong, EEG activities recorded during the execution of mental tasks, Go/NoGo experiments, or vigilance sessions have been used for drowsiness detection with ground truths obtained in the form of hits and misses. Sun and Lu performed a monotonous visual task, and averaged the number of errors committed within 2-minute windows to label and classify fatigue levels with support vector machines (SVM) and parallel hidden Markov models (HMM) [112]. The same group used the first and last five minutes of a psychomotor vigilance test (PVT) and constructed partial directed coherence (PDC) for connectivity matrices before classification [151]. Comsa *et al.* assigned the first five minutes of an auditory semantic task to the responsive and another 5-minute long interval from subsequent windows to the unresponsive states and performed classification using the pre-stimulus EEG [188]. Fahimi *et al.* labeled all trials obtained from a Stroop color test [207] as attentive while the subsequent resting-state sessions were used for the non-attentive state [199]. Finally, although Chavarriaga *et al.* characterized the evoked potentials and spectral power during the anticipation period in a visual driving simulation [208], few studies have focused on the importance of *predicting* low-vigilance trials from intervals prior to the onset of visual stimuli to enable further intervention in the interface or adaptation of the underlying classifications algorithms.

Besides the different experimental paradigms and labeling strategies mentioned above, the nature of features extracted from EEG signals are highly diverse as well. Earlier studies on sleep stage classification used (a) statistical moments, (b) peak-to-peak voltage and area under the EEG and EOG signals, (c) time-frequency features (fast Fourier transform and power spectral density), (d) wavelet coefficients and entropy, (e) signals constructed from blind source separation such as the principal and independent component analyses, and (f) cross-correlation and autocorrelation of multi-channels signals [209]. However, structural and functional connectivity differences between the alert and drowsy states, although known in the literature, have been rarely used for classification of these mental states in the literature. It is, for example, widely known that the right fronto-parietal regions become activated in simple sustained attention tasks and with increased time-on-task, and become deactivated with increase in mental fatigue and decrease in task performance [149]–[151]. EEG-based phase synchronization and other connectivity measures have been also analyzed in Go/NoGo tasks and motor execution/inhibition with audio stimuli [152], [182], vigilance and working memory [210], [211], and transitions and implications of alert versus fatigued task execution [178], [179]. With a shift in classic spectral features and in their work on sleep microstate classification versus responsive trials, Buriro *et al.* extracted a combination of seven inter-channel features, namely, covariance, correlation, cross-spectral power, coherence, joint entropy, mutual information, and phase synchronization index, from band-passed EEG signals. They found that joint entropy, wavelet cross-spectral power, and covariance features outperformed the normalized features in an cross-subject classification [191]. Feature selection was applied in each fold of their leave-one-subject-out cross validation for a total of 8 participants, and the area under the receiver-operator characteristic (ROC) and precision-recall (PR) curves as well as the Matthew's correlation coefficient (phi) [212] were used due to the class imbalance in each cross-validation fold. Chen *et al.* used unweighted networks from phase lag index (PLI) of EEG signals to characterize and classify the correlates of alert –first three minutes– and drowsy –last three minutes– intervals in a one-hour simulated driving task [201]. In a different study on audio classification, the power spectral density ratio and weighted phase lag index (WLPI) were used for detection of responsive versus unresponsive state [188].

Almost simultaneously with [201] and [188], we presented and published our work on continuous estimation of vigilance levels using the collective, cross-validated pre-trial phase synchrony index of 8-min blocks of SART experiments at the EMBC 2019, and focused on using the same features, this time from individual trials of each participant separately, to classify their most extreme attention levels.

4.3 Experiment 1: Cross-Subject Regression of Vigilance Score and Response Time

In this section, we present the methods and results for prediction of continuous vigilance levels using averaged pre-trial phase synchrony index (PSI) from experimental blocks in a leave-one-subject-out cross-validation scheme.

4.3.1 Methods

Details of the experiment design, data acquisition, and labeling scheme are as described in Section 3.3. Monopolar EEG signals were recorded using 64 Ag/AgCl active electrodes mounted according to the 10-10 International Electrode Placement System and connected to a BioSemi ActiveTwo amplifier (Biosemi Inc., Amsterdam, The Netherlands). The common mode sense (CMS) active electrode and Driven Right Leg (DRL) passive electrodes served as the reference and ground of the system. The sampling rate for the EEG and 3-channel EOG signals was set to 2,048 Hz. Each 12-block SART session, equivalent to 2,700 trials, lasted for approximately 105 minutes and resulted in over 13.5×10^6 time samples. Two participants, S02 and S08, completed fewer than 12 blocks. The recordings in each block were bandpassed between 1 and 70 Hz and the logistic Infomax Independent Component Analysis (ICA) algorithm of Bell and Sejnowski [161] was applied to the pre-processed signals. Independent components with peak activities exceeding 9 standard deviations from their mean were rejected as indicators of non-cortical activities from muscular or ocular artifacts as well as abrupt and sharp movements. Finally, block-wise signals were back-propagated and reconstructed.

4.3.1.1 Phase-locking Value Computation

From each trial in the artifact-free EEG signals, time intervals from -200 ms up to 1,600 ms with respect to the digit onsets were extracted and downsampled to 512 Hz, resulting in intervals of 923 samples[1]. To demonstrate a measure of inter-trial variability [128], the phase locking value (PLV) is computed for all unique electrode pairs of band-passed signals in the α (8–12 Hz), lower β-1 (12–16 Hz), lower β-2 (16–20 Hz), mid-β (20–24 Hz), upper β (24–28 Hz), and wide-band γ (28–60 Hz) oscillations in all the time samples and trials. Each real-valued and band-passed signal $x(t)$ is convolved with $1/\pi t$ to obtain $x_{HT}(t)$, the Hilbert transform of $x(t)$ [163]. The inverse tangent of $x_{HT}(t)/x(t)$, $\Phi_i(t,n)$, represents the instantaneous phase of transformed signals for electrode i at time bin t and trial $n = 1,2,...,N$. The phase difference for each electrode pair (i, j) is then obtained

[1]A note on the commonly used terms in this paper: Time interval actually refers to the term epoch used in neurophysiological studies for windows of specific length extracted with respect to each digit (stimulus) in an experimental trial, and the term epoch refers to one full pass of training set in the deep learning terminology. Furthermore, a time sample refers to an instantaneous electric potential in the EEG recordings [213].

from the difference between their instantaneous phases at each time sample, i.e.,

$$\Phi_{ij}(t,n) = \Phi_i(t,n) - \Phi_j(t,n), \tag{4.1}$$

where $i = 1,2,...,63$, and $j = 2,3,...,64$, and $t = 1,2,...,923$ in our setup. In other words, this phase difference is extracted for all the time bins, trials, and electrode pairs.

As a general rule, if signals measured at electrodes i and j are at the maximum synchrony with each other or have a small phase difference at each time bin t and the majority of trials inside a block, their PLV is close to 1. Otherwise, their PLV is closer to 0 in case of large differences in instantaneous phase or random pairwise phase variations across different trials.

4.3.1.2 Phase-Synchrony Index Computation

For each block, PLV at time t is calculated by dividing the magnitude of sum of complex exponentials with the phase differences computed in Eq. 4.1 by the total number of trials in that block:

$$PLV_{bk}(t) = \frac{1}{N}\left|\sum_{n=1}^{N} e^{j\Phi_{ij}(t,n)}\right|, \tag{4.2}$$

where bk denotes the block number, N is equal to 225 for each block, and $t \in \{1,2,...,923\}$. Next, to analyze the entire pre-trial connectivity patterns for each block, PLVs from the downsampled intervals of 200 ms before the digit onsets were averaged for all trials regardless of their labels. This resulted in obtaining a 64 $\times$ 64 matrix for each block that represented the phase-synchrony index (PSI) for each electrode pair. In other words,

$$PSI_{bk} = \frac{1}{T}\sum_{t=1}^{T} PLV_{bk}(t), \tag{4.3}$$

where $T = 103$ for the downsampled time intervals. The final feature set for each frequency band, $X \in R^{2016\times 113}$, consisted of PSI values for 2,016 unique electrode pairs and all the 113 blocks of the whole 10 SART participants.

4.3.1.3 Adaptive Vigilance Labels

In this study, to design an objective measure for labeling sustained attention without interrupting the users to carry on their experimental tasks, we rely only on the response time and number of erroneous and correct events – unlike the percentage of eye closure (PERCLOS) obtained from EOG measures in simulated driving experiments [139]. Experiment trials were automatically labeled by the interface according to the type of the digit –target or non-target– and whether a click was executed or not. To come up with a multi-level measure purely dependent on the task performance, we calculate the 5-level Trial Vigilance Score (TVS) measurement system [52]. Next, the Cumulative Vigilance Score (CVS) at each trial is calculated with a moving average window from the TVS of

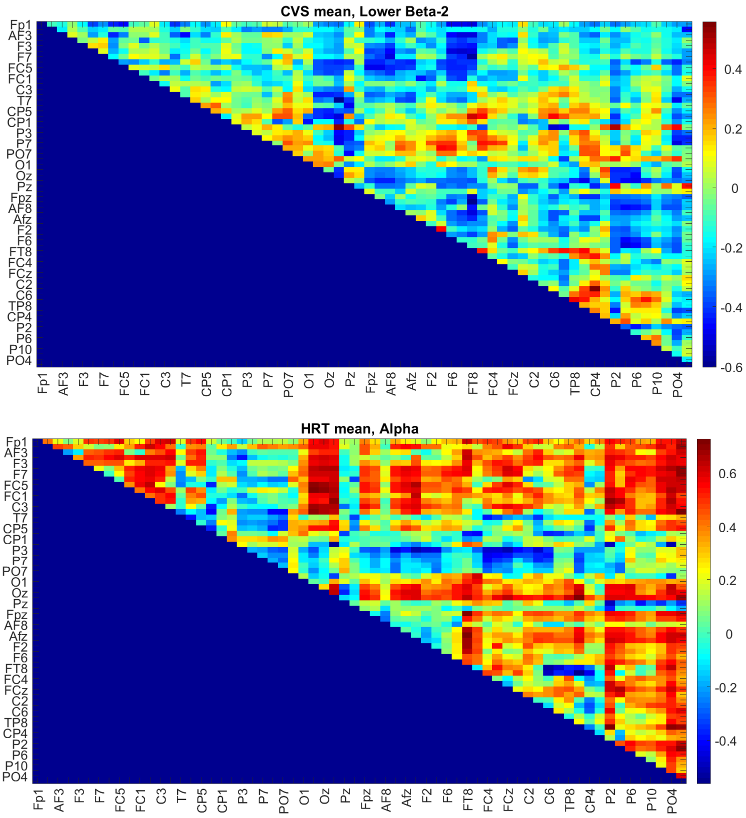


Figure 4.1: Heatmaps showing correlation patterns for the block-wise pre-trial PSIs with performance measures from 113 blocks of all participants. (Top) Mean CVS, lower beta-2, and (Bottom) Mean HRT (ms), alpha oscillations. Color bars denote the Pearson's coefficients, $p < 0.001$ when $|r| > 0.32$.

36 preceding trials that lasted for 4 sequences or 73 seconds, and subsequently divided by four to keep the range between 0 and 1. This measure allows us to characterize the long-term, tonic nature of attention and performance variations rather than the momentary, phasic correlates of missed clicks, and works similarly to the cumulative scores in sleep stage classifications [111].

4.3.1.4 Deep Neural Networks for Block-wise Regression

A total of 12 experiments were conducted to predict all the block-wise CVS mean and HRT mean values from five narrow-band and one wide-band pre-trial PSI feature sets. Feature vectors were fed to the designed deep networks consisting of an input layer, a dropout layer to prevent overfitting, a rectified linear unit (ReLU) activation layer followed by a fully-connected layer with 800 units, a drop-out layer and ReLU layer, a second fully-connected layer with 300 units, and a ReLU layer followed by the output layer with one unit. Two networks were trained with two different loss functions commonly used in regression problems, the Mean-Squared-Error (MSE) and Mean-Absolute-Error (MAE).

By observing the value of loss functions after each epoch, we found that, for this small dataset, 150 epochs were sufficient for the stochastic gradient descent (SGD) to reach

Table 4.1: The best RMSE, MAE, and Pearson's correlation coefficients for prediction of CVS Mean (min: 0.1768, max: 0.6110, median: 0.4818) using DNNs trained with MSE and MAE loss functions in their regression layers for 5 permutations of 4-fold cross-validation. Numbers in the parentheses denote the pair of mini-batch sizes and learning rates obtained through grid search for the best output of each network evaluated by the designated performance metric.

Perf. Metric	Root-Mean-Squared-Error		Pearson's Correlation	
Loss Function	MSE Loss	MAE Loss	MSE Loss	MAE Loss
Alpha	0.0465 (64, 0.044)	**0.0464 (128, 0.078)**	0.760 (16, 0.044)	**0.770 (128, 0.023)**
Lower Beta-1	**0.0457 (16, 0.062)**	0.0459 (16, 0.048)	**0.772 (16, 0.062)**	0.771 (16, 0.038)
Lower Beta-2	**0.0430 (16, 0.011)**	0.0440 (16, 0.030)	**0.806 (16, 0.014)**	0.802 (128, 0.100)
Mid-Beta	**0.0457 (16, 0.062)**	0.0466 (128, 0.048)	**0.768 (16, 0.078)**	0.762 (16, 0.018)
Upper Beta	**0.0443 (16, 0.011)**	0.0447 (16, 0.018)	0.786 (16, 0.048)	**0.790 (128, 0.078)**
Gamma	**0.0461 (16, 0.048)**	0.0461 (16, 0.030)	0.777 (16, 0.048)	**0.781 (16, 0.018)**

the minima. A grid search was then performed to optimize each loss function for 20 learning rates logarithmically increasing from 0.001 to 0.1 and four mini-batch sizes of 16, 32, 64, and 128. To assure fair evaluations were being conducted, we ran 5 dataset permutations, each consisting of 4-fold cross validations. For each combination of the learning rate and mini-batch size, the network performance was assessed using the root-mean-squared-error (RMSE) and the Pearson's correlation coefficient between the true and predicted labels.

4.3.2 Results

Figure 4.1 demonstrates the correlation patterns between the block-wise CVS mean with the pre-trial PSI features of the lower beta-2, and the block-wise HRT mean with the pre-trial PSIs of alpha oscillations which showed the most wide-spread and highest number of significant correlations. Two-way analysis of variance (ANOVA) revealed significant factors of frequency bands and electrode pairs on the obtained correlation coefficients for both performance measures, $p < 0.001$. Stronger desynchronization within the frontal network and from left centro-temporal channels with the midline parieto-occipital, and synchronies within the right centro-tempo-parietal cortex are correlates of improved CVS. The strongest correlates of delayed responses are observed from the alpha oscillations from synchronies within the left fronto-central and from there with the right parietal and occipital channels, and within the right parieto-occipital cortex.
Tables 4.1 and 4.2 show the lowest errors and highest correlations obtained for prediction of CVS mean and HRT mean after grid search on the learning rates and mini-batch sizes for all feature sets. Among the 6 tested feature sets, the lower beta-2 PSIs outperformed other bands in predicting the block-wise CVS mean with the lowest errors of 0.0430 and 0.0440 in networks with the MSE and MAE loss functions, respectively, both obtained with mini-batches of 16 samples. With a range of 0.1768 to 0.6110 and median of 0.4818, the error to median ratio is equal to 0.0892 for this regression problem. Linear

Table 4.2: The best RMSE, MAE, and Pearson's correlation coefficients for prediction of HRT Mean (min: 261.81 ms, max: 840.66 ms, median: 433.00 ms) using DNNs trained with MSE and MAE loss functions in their regression layers for 5 permutations of 4-fold cross-validation. Numbers in the parentheses denote the pair of mini-batch sizes and learning rates obtained through grid search for the best output of each network evaluated by the designated performance metric.

Perf. Metric	Root-Mean-Squared-Error		Pearson's Correlation	
Loss Function	MSE Loss	MAE Loss	MSE Loss	MAE Loss
Alpha	**51.91 (32, 0.030)**	54.75 (16, 0.009)	**0.903 (16, 0.018)**	0.896 (16, 0.023)
Lower Beta-1	**53.21 (16, 0.018)**	55.86 (16, 0.018)	**0.903 (16, 0.018)**	0.899 (16, 0.023)
Lower Beta-2	**52.22 (32, 0.038)**	55.28 (16, 0.023)	**0.899 (32, 0.038)**	0.893 (16, 0.023)
Mid-Beta	**53.28 (16, 0.030)**	55.39 (32, 0.078)	0.902 (16, 0.030)	**0.904 (16, 0.023)**
Upper Beta	55.33 (32, 0.062)	58.91 (32, 0.062)	0.892 (16, 0.048)	0.889 (16, 0.030)
Gamma	56.85 (16, 0.018)	60.32(16, 0.038)	0.890 (64, 0.078)	0.883 (16, 0.048)

correlations obtained from these networks were equal to 0.806 and 0.802.
The pre-trial PSIs of alpha band returned the lowest errors of 51.91 ms and 54.75 ms, and the correlations of 0.903 and 0.896 from the MSE and MAE loss functions for predicting the average HRT, respectively, closely followed by the lower beta-1 and mid-beta features. With a range of 261.81 to 840.66 ms and a median of 433.00 ms, this performance translates to an error/median of 0.12. Figure 4.2 demonstrates the performance curves versus learning rates of the best feature sets for predicting the block-wise CVS mean and HRT mean, obtained from four mini-batch sizes in networks with the MSE and MAE loss functions.

4.3.3 Discussion

In this study, we trained DNNs to assess the average and objective performance in a long SART experiment using phasic synchrony features of pre-trial EEG as a novel integration of functional connectivity into deep learning architectures. Contributions are multi-fold, starting with a labeling strategy independent of the participant's facial and ocular artifacts that does not require a built-in camera or a constant interruption of the natural variation of individuals' vigilance patterns. Second, the proposed system uses phasic PSIs of pre-trial intervals to assess the tonic performance of individuals in the order of minutes, thus reducing the need for fast updating any alarm system and providing a suitable approach for performance monitoring for users susceptible to hypo- or hyper-vigilance disorders and subsequent system tuning. Our findings are in line with the roles of alpha and beta coherence values in transition from awake to fatigue states [151], [156], the role of alpha desynchronization in attentional processes [152] and task learning [180], and role of beta synchronization over the sensorimotor cortex in motor learning. This work can be extended by using CNNs to assess the learned weights of hidden layers as shown in the next section.

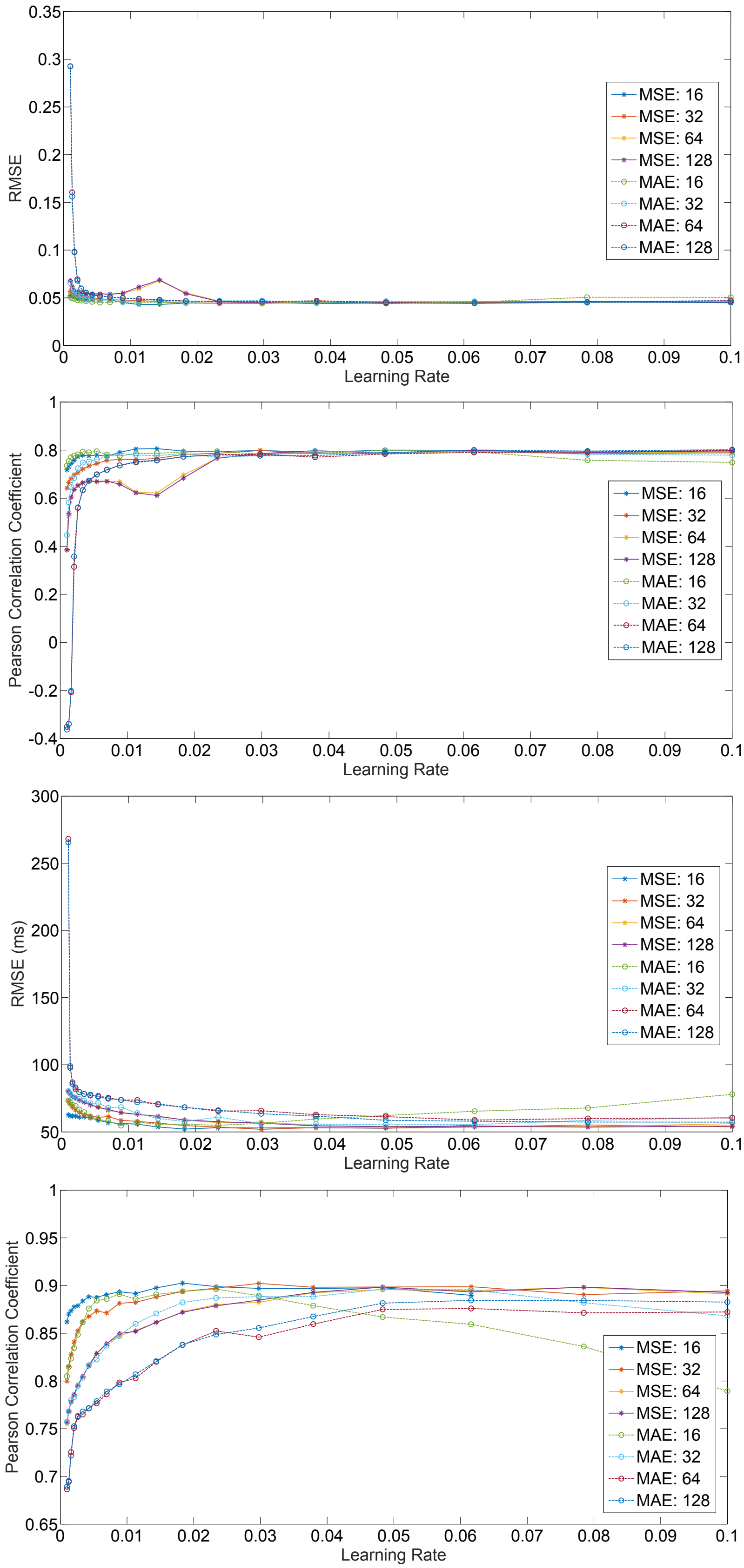


Figure 4.2: The RMSEs and correlation coefficients versus learning rates in predicting the (top) block-wise mean CVS from lower beta-2 pre-trial PSIs and (bottom) mean HRT from alpha-band pre-trial PSIs. Curves represent the validation metrics from experiments conducted on different mini-batch sizes by networks with the MSE and MAE loss functions.

4.4 Experiment 2: Classification of Drowsy and Alert Vigilance States

In this section, we present the methods and results for prediction of discrete drowsy and alert vigilance levels using pre-trial phase locking values of individual trials for each participant in the SART scheme.

4.4.1 Methods

In this part, the method of obtaining ground truth labels for the drowsy and alert classes are explained and two datasets generated from the band-power (BP) ratios and phase synchrony values (PLV) of each trial are introduced. Results of vigilance classification using the BP ratio features are treated as the baseline for the second group of features. We discuss the experiments conducted with classical machine learning schemes applied to the BP datasets and with a deep CNN architecture on the PLV datasets. For the first group, we aim to compare our findings with the literature on emotion recognition, BCI applications, and sleep or microsleep detection in which the classical learners are traditionally applied on Fourier transform coefficients and band-power features. For PLV features constructed as images in our datasets, we are interested in exploring the usability of CNNs in classification and visualization of learned features through inter- and intra-region connections.

4.4.1.1 Construction of Drowsy and Alert Classes

Table 4.3 includes the variability of CVS, defined as the ratio of its standard deviation to its average value, for each participant. Results of performance stability from CVS curves demonstrate that participants S02, S03, S05, S08, S09, and S11 were able to maintain relatively stable vigilance scores during the long SART experiment while S04, S06, S07, and S10 experienced extreme transitions between high alertness and full sleepiness and were labeled as having unstable CVS curves.
To construct the two discrete vigilance levels, Next, a CVS histogram was plotted for each participant and two thresholds were applied to obtain three vigilance states of drowsy, alert, and normal, so that either each extreme had less than 9% of the total samples or one of them had at least 200 samples. Table 4.3 shows the range of these custom thresholds for drowsy (low-CVS) and alert (high-CVS) classes and the number of trials in each class.

4.4.1.2 Analyzing Performance Stability from CVS Curves

A number of parametric curves including the linear, quadratic, and cubic polynomial functions, exponential functions, Fourier representations with one and two frequencies, and Gaussian functions with one and two terms were fitted to the CVS curve of each participant from the entire experiment. The best fit was determined based on the largest adjusted R^2 and smallest values of Akaike's information criterion (AIC) and Bayesian information criterion (BIC). These curves visualize the participants' abilities in sustaining their attention levels throughout this long experiment, and the sharp slopes in the

Table 4.3: CVS variability, threshold range (the difference between low-CVS and high-CVS thresholds), and the number of trials in the drowsy and alert classes of each participant. Numbers in the parentheses denote the ratio of drowsy trials in each dataset.

Participant	CVS Variability	Threshold Range	# Drowsy Samples	# Alert Samples
S02	0.057	0.064	284 (53.69%)	245
S03	0.047	0.073	241 (44.88%)	296
S04	0.335	0.300	256 (49.71%)	259
S05	0.062	0.085	198 (54.10%)	168
S06	0.115	0.120	244 (49.39%)	250
S07	0.234	0.260	244 (50.00%)	244
S08	0.056	0.090	192 (48.85%)	201
S09	0.112	0.170	206 (50.49%)	202
S10	0.228	0.300	226 (53.94%)	193
S11	0.096	0.120	245 (52.01%)	226

fitted curves are indicators of fast improvements or sudden declines in sustained attention levels.

4.4.1.3 Construction of BP-ROI Datasets

Fast Fourier Transform (FFT) was applied to compute the pre-trial band powers, and FFT coefficients inside the following non-overlapping frequency bands were obtained: 0.5 – 3.5 Hz (δ), 4 – 7.5 Hz (θ), 8 – 11.5 Hz (α), 12 – 15.5 (lower β- 1), 16 – 19.5 Hz (lower β-2), 20 – 23.5 Hz, ..., and 44 – 48 Hz (upper γ). These spectral powers, when extracted from various cortical regions during the pre-trial or in-trial interval in a number of sensory-motor and motor-imagery paradigms, were shown to be associated with attentional levels [30]. Since our earlier analysis had shown participants had different levels of base-band powers in the eyes-open and eyes-closed resting states and the actual task sessions, we computed the ratios of non-overlapping band powers with respect to the sum of coefficients from 0.5 Hz to 48 Hz for each trial from the magnitudes of the FFT coefficients. We also extracted the following mixed-band features from pre-stimuli intervals: $(\theta+\alpha)/\beta$, the reciprocal of an attention or engagement index used in neurofeedback experiments [214], $(\theta+\alpha)/(\alpha+\beta)$, α/β, a sign of mental inattentiveness and decreased arousal [214]–[216], and θ/β or TBR, an indicator of decreased attention in a number of ADHD subgroups [217]. For these mixed-band features, wide-band β was calculated from 12 to 28 Hz.

Next, for each participant, 10 BP features were extracted from the pre-trial intervals of all the low-CVS and high-CVS trials. These features included the narrow-band δ, θ, α, β-1, and β-2, wide-band γ, and the four mixed-band features mentioned above. To summarize the role of spatial variations in cortical activities, the 64 electrodes were grouped into the following 14 regions of interest (ROI): left, midline, and right pre-frontal (LPF, MPF, and RPF), frontal (LF, MF, and RF), central (LC, MC, and RC), parietal (LP, MP, and RP), and left and right temporal (LT and RT) regions. Thus, the BP-ROI dataset of each participant had a total of 140 spatial-spectral features.

4.4.1.4 Baseline Experiment: Classification of BP-ROI Datasets

For each participant, the BP-ROI datasets were permuted and split in a 0.8:0.2 ratio to obtain the training and test sets. Six classic learners including SVM with linear, quadratic, and Gaussian kernels, decision tree (DT), k-nearest neighbor (kNN), and Naive Bayes (NB), as well as four ensemble techniques (DT with bootstrap aggregation or bagging, adaptive boosting (AdaBoostM1), random undersampling boosting (RUSboost), and kNN with subspace learners) were utilized. A 5-fold cross-validation was applied on the training set to train each of these classification approaches and perform hyperparameter tuning on the remaining fold.

For each learner and validation fold, the best set of hyperparameters was obtained from the "best estimated feasible point" after 30 iterations in MATLAB® 2018b. This training-test split and subsequent cross-validation and testing operations were repeated five times for each frequency feature and the entire features in the BP-ROI dataset. The tuned hyperparameters consisted of the box constraint and kernel scale for SVM learners, number of neighbors and distance metric for k-nearest neighbor (kNN) learners and the number of learning cycles for kNN ensembles, the width for Naive Bayes, minimum leaf size for decision trees, and the number of learning cycles, learning rate, maximum number of splits, and minimum leaf size for ensemble decision trees with bootstrap aggregation (bagging), adaptive boosting (AdaBoostM1), and random undersampling boosting (RUSboost) algorithms.

4.4.1.5 Construction of PLV-abs Datasets

The usual method of averaging $\Phi_{ij}(t,n)$ computed in Eq. 4.1 involves dividing the average values of sum of complex exponentials with the aforementioned phase difference by the total number of trials from the same class or state, resulting in only one 64×64 matrix for each time bin of the entire trials. This method was utilized in Eq. 4.2 for Experiment 1 since we were interested in obtaining only one feature matrix from each block. Since this method would not allow us to populate an observation for each individual trial, we followed a different approach for within-subject classification of alert and drowsy states in Experiment 2. After obtaining $\Phi_{ij}(t,n)$ from Equation 4.1 for each trial of the drowsy and alert classes, we segmented the $[-200,1600]$ ms windows into shorter time intervals to analyze the effect of pre-digit, early onset, and late onset time intervals on the detection rates. As shown in Equation 4.4, the sum of complex exponentials with the calculated phase difference for each intervals was divided by the total number of time samples in that interval. In other words,

$$PLV_{[t_b:t_e]}(n) = \frac{1}{|\{t_b, t_b+1, ..., t_e\}|} \sum_{t=t_b}^{t_e} e^{j\Phi_{ij}(t,n)}, \tag{4.4}$$

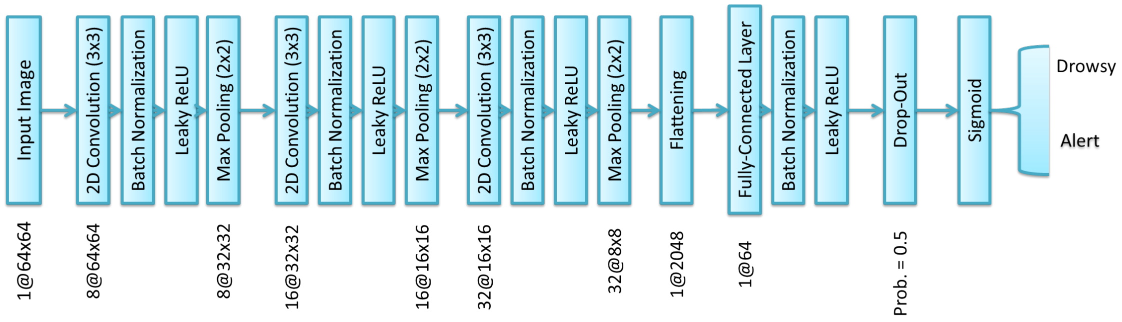


Figure 4.3: The CNN-based deep neural network architecture proposed for classification of drowsy versus alert states from 64-by-64, symmetric PLV matrices.

where t_b and t_e represent the beginning and end of the desired temporal interval with respect to the digit onset.

The magnitudes of these complex numbers were separately saved for each trial in symmetrical square matrices that contained 2,016 unique channel pairs, resulting in one 64×64 $PLV_{[t_b:t_e]}$ matrix for each trial among the drowsy and alert trials. We calculated PLV-abs matrices from the following intervals: $[-200, 0]$ ms (pre-trial), $[-200, 100]$ ms (pre-trial and early onset), $[-100, 100]$ ms (early onset), $[100, 300]$ ms (medium onset), and $[300, 500]$ ms (late onset). We thus obtained a total of 35 different datasets for each participant with $X_{low} \in R^{N_{low} \times 64 \times 64}$ and $X_{high} \in R^{N_{high} \times 64 \times 64}$ to represent the low-vigilance and high-vigilance trials from 7 different frequency bands of 5 different intervals.

4.4.1.6 The Proposed Convolutional Neural Network

For each time interval and frequency band, PLV-abs matrices were fed to a CNN with the architecture shown in Figure 4.3 that was implemented in TensorFlow with a Keras backend. The designed network consists of 3 blocks, each with one 2D convolutional layer with 3×3 kernels, a batch normalization layer, a Leaky Rectified Linear Unit (leakyReLU) activation function [218], and a max-pooling layer. The number of kernels in each convolutional layer is set to 8, 16, and 32, respectively. Outputs of the last max-pooling layer are flattened and fed to a fully-connected layer with 64 neurons followed by a batch normalization layer and another leakyReLU non-linearity. The output of leakyReLU is connected to the last fully connected layer with 2 neurons where the half of these connections are randomly dropped by the Dropout block during training for regularization. Weights in all the convolutional and dense layers are initiated with a random normal function, and a ℓ_2 kernel regularizer with a decay rate of 0.01 is used to avoid overfitting. The Adam algorithm, a method for adaptive moment estimation, is used for optimization with a learning rate of 1×10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ [165]. Binary cross-entropy is set as the loss function. Due to small levels of imbalance between the number of drowsy and alert samples in the datasets of Table 4.3, two objective functions are utilized: classification accuracy, defined as the number of correctly classified samples divided by the total number of samples in each batch, and a custom F1-mean, defined as the average of F1 scores for correct classification of the drowsy and alert samples.

For each frequency band, X_{high} and X_{low} trials were concatenated and randomly permuted, and split into a 0.8:0.1:0.1 ratio to construct the training, validation, and test sets. The network was trained for 300 epochs with a mini-batch size equal to 16. At the end of each epoch, network weights were saved if the F1-mean obtained from validation had increased with respect to the F1-mean of all the previous epochs. The "best model" with the highest F1-mean was retrieved at the end of training and used for predicting the test labels. The whole operation was repeated for five runs.

4.4.2 Results

In this section, experimental results for within-subject classification of drowsy (low-CVS) and alert (high-CVS) trials are presented. Special care is given to the success of each feature from the perspective of individual behavior styles and roles of these measurements in observing specific oscillations and synchronization patterns. PR-AUC is reported as the performance metric since it is more sensitive to the poor performance of a given classifier in an imbalanced dataset [219]. The effects of features, learners, and time intervals are assessed with the ANOVA, and the Wilcoxon signed rank test [220] is used to compare the medians of PR-AUC from conducted experiments. This section concludes by visualizing the activations of a trained CNN for a participant with the highest PR-AUC to explore the usability of CNNs in classification and visualization of learned features through inter- and intra-region connections.

4.4.2.1 BP-ROI Features and CVS Variability

Before obtaining the classification results using BP-ROI features, a correlation analysis was performed between the features and continuous-valued CVS vectors to identify the most informative features for the upcoming classification task. These features can potentially be used for assessment of vigilance levels in unsupervised studies from real-time recordings (ref. Chapters 5 and 6).

The blue curves in Figure 4.4 show the CVS obtained from 36-trial windows plotted versus the trial index for each participant. The red curves represent the parametric functions fitted to these CVS curves using the method described in Section 4.4.1.2. CVS variability, defined as the standard deviation of CVS divided by the its mean, was computed for all participants. We calculated Pearson's correlation coefficients between the CVS curves and cumulative versions of all features in the pre-trial BP-ROI dataset, and obtained heat maps with 10×14 cells for each participant as shown in Figure 4.4.

The dark red and dark blue cells, respectively, indicate the large positive and negative correlations. These correlation patterns, together with the specified CVS variability and range of low- and high-CVS thresholds, demonstrate the individual physiological and behavioral differences especially between the group who could maintain relatively stable vigilance scores with those who experienced extreme transitions between high alertness

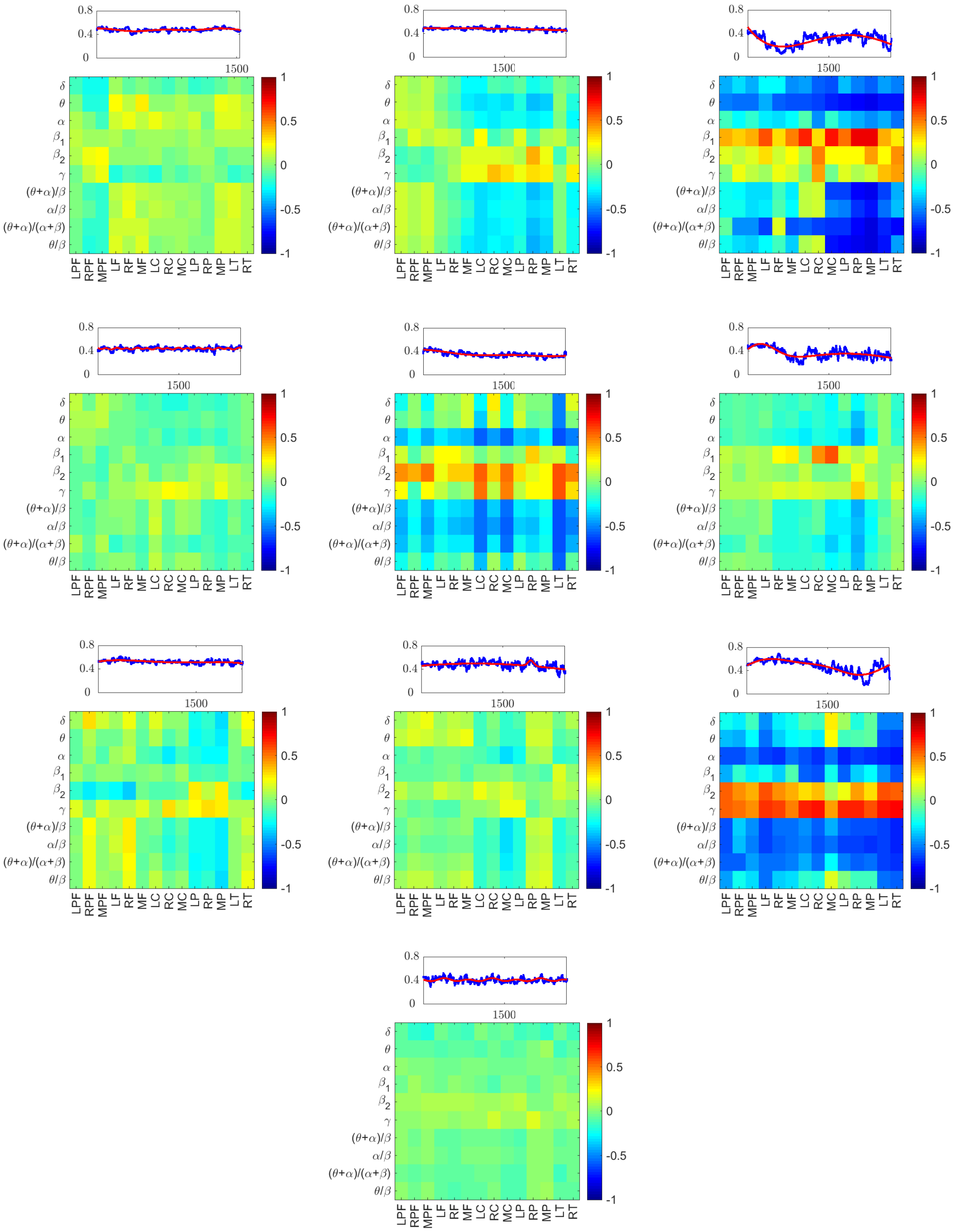


Figure 4.4: Behavioral differences in maintaining stable vigilance scores and electrophysiological differences in correlations between CVS and pre-trial BP ratios for 10 SART participants. The blue and red curves, respectively, represent the 36-trial averaged CVS curves and the parametric functions fitted to them. The heat map cells demonstrate the Pearson's correlation coefficients between the CVS scores and each of the 10 BP features obtained from 14 regions of interest from the entire SART experiment.

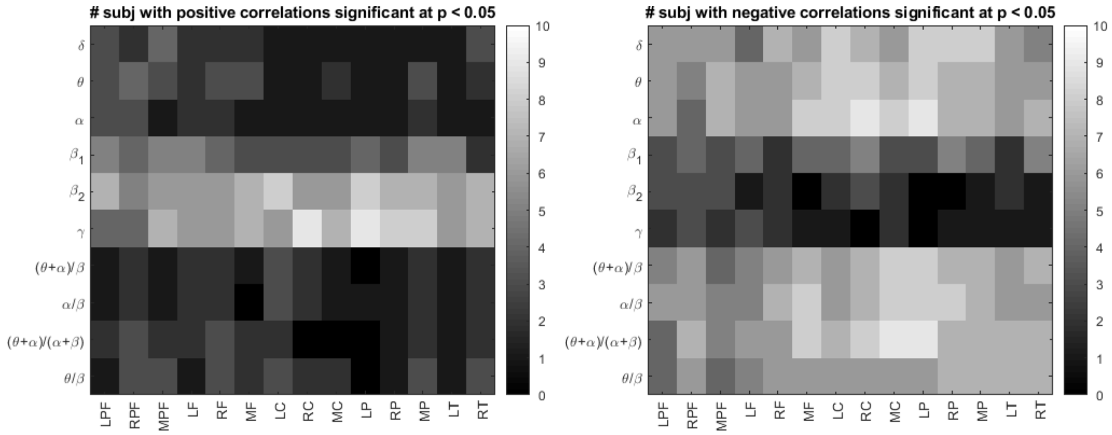


Figure 4.5: Total number of participants for whom the linear correlations between the cumulative vigilance scores and pre-trial BO-ROI features, averaged over 36-trial windows as shown in Figure 4.4, were significantly (a) positive and (b) negative at the 0.05 level.

and full sleepiness. S04, with the largest CVS variability of 0.335, slept during most of the second and third blocks and regained their alertness later. Increase in the ratio of lower β-1 especially from the frontal, parietal, and occipital regions, increase in β-2 from temporal and right central channels, overall decrease of θ ratios, and reduction in δ and the mixed-band ratios from parietal, temporal, and midline central regions were highly correlated with increase in the cumulative vigilance scores. S07, with a CVS variability of 0.234, slept during the fourth and fifth blocks and demonstrated similar correlations patterns but in smaller scales. S10, with a variability of 0.228, had no errors in the initial blocks and only experienced heavy sleepiness towards the end. An interesting case was S06 who had a smaller variability equal to 0.115 and showed a gradually deteriorating performance. Despite having no large peaks or falls in the CVS curve, S06's correlation patterns from the central, parietal, and temporal regions were similar to that of S10. Finally, the CVS patterns of S02 and S03 appeared to be very similar.

To summarize the common patterns, Figure 4.5 shows the total number of participants for whom these linear correlations were statistically significant at the 0.05 level. Increase in γ ratios from the left central and the whole parietal regions, and increase in β-2 from the left central and parietal channels was correlated with higher CVS in at least 80% of participants. Reductions in slower oscillations and their combined ratios from midline frontal, the whole central cortex, and the left parietal region generally demonstrated negative correlations with increased CVS during the experiment.

4.4.2.2 Classification Results from BP-ROI Features

Ten classical learners were used for drowsy versus alert state detection with 10 individual features as well as the whole BP-ROI dataset as described in Section 4.4.1.4. We noticed that a number of feature-learner pairs would result in unequal F1 scores for the drowsy and alert classes, and this difference could easily exceed 0.4 in the case of S02, S03, and S05 where Naive Bayes models resulted in the most imbalanced classifications. We thus only report the PR-AUC values in Figure 4.6, choosing the drowsy state as the positive class. For easier interpretation of individual results and between-subject differences, CVS variability and the difference between the alert and drowsy thresholds are specified.

For S10, the set of all BP features from all learners exceeded the 90% classification accuracy. Using all the BP features as well as θ for S04 and α and mixed-band ratio features for S06, these two participants could also achieve detection rates higher than 80%. The set of all features generally worked well for all participants except for S02, S09, and S11. Finally, the SVM classifiers and decision tree ensembles outperformed the rest of models for most participants.

The top bar plot of Figure 4.8 shows the PR-AUC from BP-ROI datasets averaged among all participants. Using SVM with Gaussian kernels on the set of all features results in the best averaged PR-AUC of 0.7251 while accuracy is maximized using the Bagged decision trees with a detection rate of 0.7011. One-way ANOVA demonstrated a significant effect of Learner type on the average of test accuracy from different features, $F(9,90) = 35.21$, $p < 0.001$. Signed rank test revealed that **SVM learners with Gaussian kernels outperformed SVMs with linear kernels and bagged decision trees,** $p < 0.05$**, while the Naive Bayes models returned the lowest PR-AUCs**. A similar ANOVA demonstrated an even stronger effect of Features among different learners, $F(10,90) = 74.07$, $p < 0.001$. Using all BP-ROI features resulted in the best detection rates while β-1 by itself performed the worst. When averaged across all participants, Wilcoxon signed rank test showed that using the set of all features resulted in the highest PR-AUC medians from different classifiers, followed by $(\theta+\alpha)/(\alpha+\beta)$ and α/β which significantly outperformed models trained by α, γ, and $(\theta+\alpha)/\beta$, $p < 0.1$.

4.4.2.3 PLV Features and CVS Variability

To determine the frequency bands and time intervals from which phase synchronization was potentially helpful for distinguishing the drowsy and alert states, two one-sided, two-sample t-tests were performed with the 0.001 significance level, and the number of electrode pairs for which the *magnitudes* of PLV values were significantly larger in each class were computed. Two-way tests of ANOVA applied to these results for separate frequency bands revealed a significant effect of Time Interval on the number of significantly larger PLVs in the alert samples with respect to the drowsy samples for mid-β-3 ($p < 0.1$), upper β ($p < 0.01$), wide-band β ($p < 0.1$), and wide-band γ ($p < 0.001$) PLVs, with the $[-200, 100]$ ms interval having more number of significantly different pairs in these bands. When focusing on the drowsy class, only for the upper β ($p < 0.1$) and γ ($p < 0.001$) bands a significant effect of Time Interval from the $[-200, 100]$ ms interval existed on significantly larger PLVs in the drowsy class. Thus, **the effect of temporal transitions on differences between the drowsy and alert states is most clearly observed in the upper β and wide-band γ phase synchrony features**. Similar tests of ANOVA showed a significant effect of Frequency Band for the $[-200, 0]$ ms ($p < 0.1$), $[-200, 100]$ ms ($p < 0.01$), $[-100, 100]$ ms ($p < 0.1$), $[100, 300]$ ms ($p < 0.1$),

Figure 4.6: AUC of precision-recall curves for within-subject drowsy-vs-alert state detection using 10 learners and 11 BP-ROI features.

and $[300, 500]$ ms ($p > 0.1$) intervals with γ band always having the most significantly different features between the two classes. Therefore, **the effect of spectro-spatial features on differences between the drowsy and alert classes is most clearly observed in the pre-trial and early onset phase synchrony features.**

4.4.2.4 Classification Results from PLV Features

For each frequency band, symmetric PLV matrices for five different time intervals were separately fed to the CNN architecture described in Section 4.4.1.6. The PR-AUC values of these experiments are displayed in Figure 4.7, choosing the drowsy state as the positive class. **Gamma-band PLVs resulted in detection rates of at least 0.9 in S04, S06, S07, and S10, and at least 0.8 in 60% of total participants**. Considering the results of participants separately for each time interval, the two-way ANOVA and Wilcoxon signed rank tests found that **γ-band PLVs outperformed all the lower frequencies,** $p < 0.05$. The bottom bar plot of Figure 4.8 demonstrate the averaged PR-AUC from PLV features. While all features from all time intervals achieved higher than chance level classification results, the pre-trial to early onset γ PLVs outperformed the rest and obtained an average PR-AUC of 0.79 with the deep CNN architecture. A two-way ANOVA revealed a strong effect of Frequency Band with $F(6, 24) = 89.33$, $p < 0.001$, on the PR-AUC averaged across all participants. S04 and S09 showed a difference of 0.2385 and 0.2381, respectively, between the best test accuracies of their respective γ and α PLVs, but S07 surpassed this difference at 0.2979. For this participant, strong γ synchrony of the left tempo-parietal channels with left and right frontal cortex, within the pre-frontal regions, and between the right parietal channels with the rest of the brain account to over 1,200 channel pairs with significantly larger synchronization. For the same subject, strong desynchronization between the left (right) central channels with the midline and left (right) pre-frontal cortex was observed.

Another two-way ANOVA also showed a significant factor of Time Interval with $F(4, 24) = 10.36$, $p < 0.001$, on the PR-AUC averaged across all participants. ANOVA and Wilconx tests revealed that, **although the effect of time interval was negligible for γ-band PLVs, for all the other frequency bands, the -200 ms to 100 ms features resulted in the highest PR-AUCs especially in comparison to the $[-100, 100]$ ms interval,** $p < 0.05$. Individually, this pattern was verified for all participants except for S02 and S11. For S02, most PLV pairs from different bands and intervals had AUCs just below 0.6 except for the upper β synchronies of the $[-100, 100]$ ms interval. The same applies for S11 with only pre-trial and early onset PLVs of γ band exceeding 0.6 in the AUC metrics. For S05, features from all the frequency bands and intervals resulted in higher-than-chance classification except for the mid-β oscillations from the late 300 ms to 500 ms interval.

4.4.2.5 Performance of BP-ROI and PLV Datasets

Gamma-band PLVs surpassed the best PR-AUCs obtained by BP-ROI features for 60% of our participants (mean = 0.122, SD = 0.077, min = 0.046 for S08, and max =

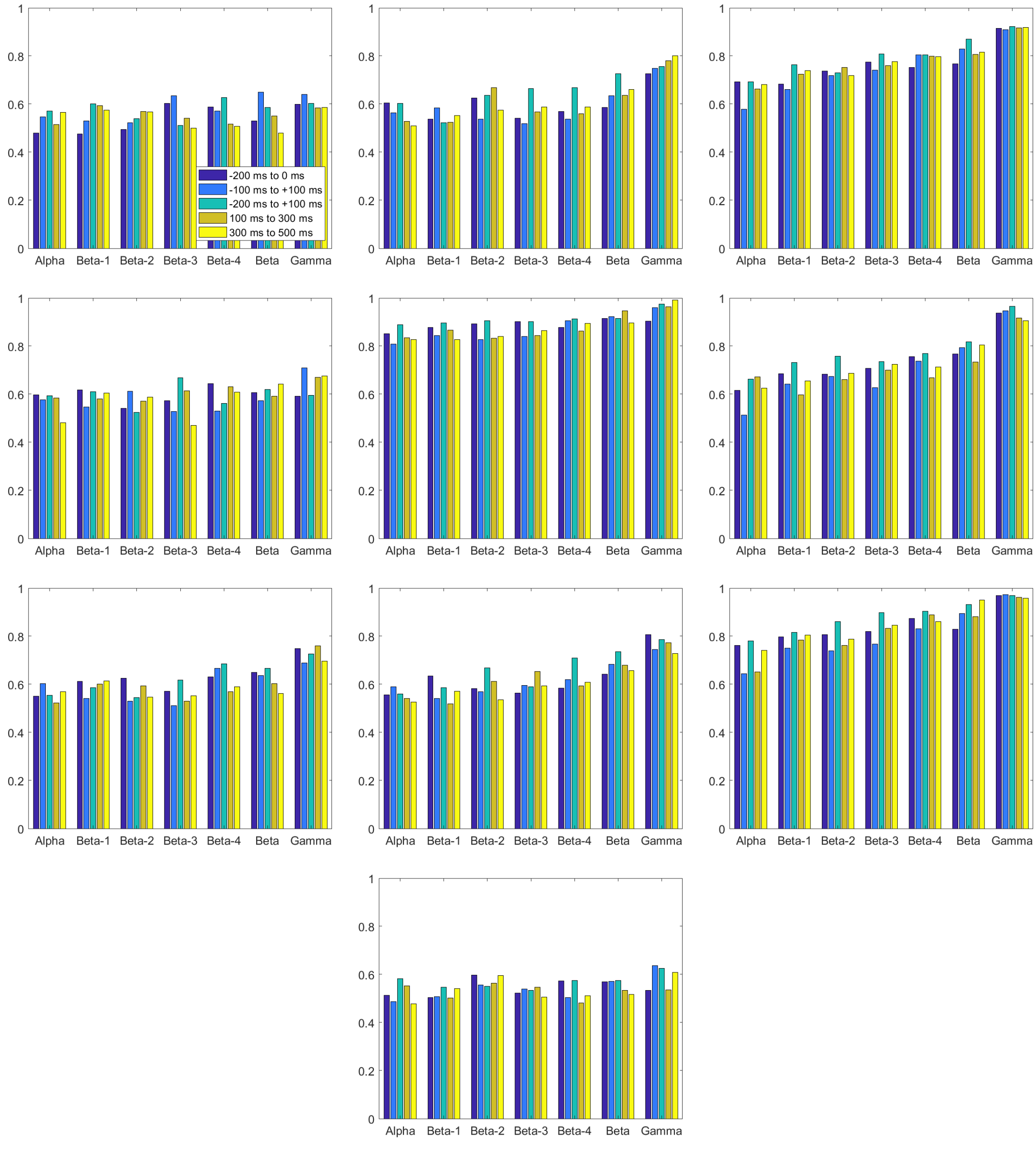


Figure 4.7: AUC of precision-recall curves for within-subject drowsy-vs-alert state detection using 5 time intervals and 7 PLV bands.

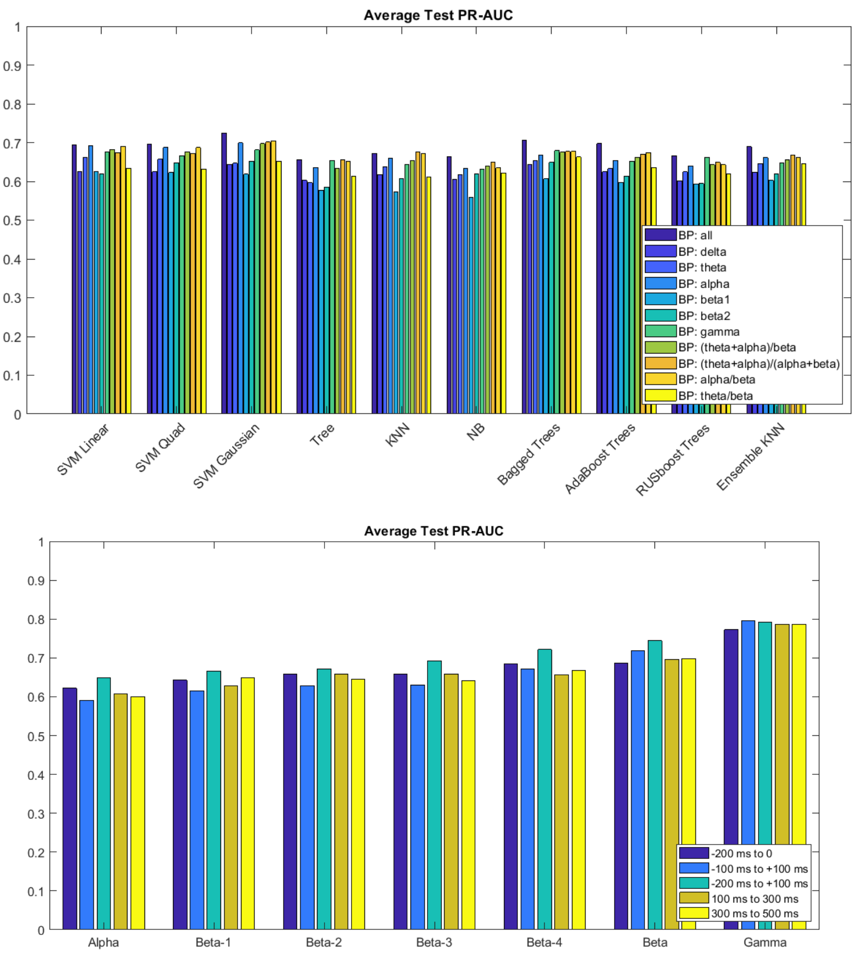


Figure 4.8: Grand-average of within-subject PR-AUC for drowsy-vs-alert state detection using (top) BP-ROI and (bottom) PLV features.

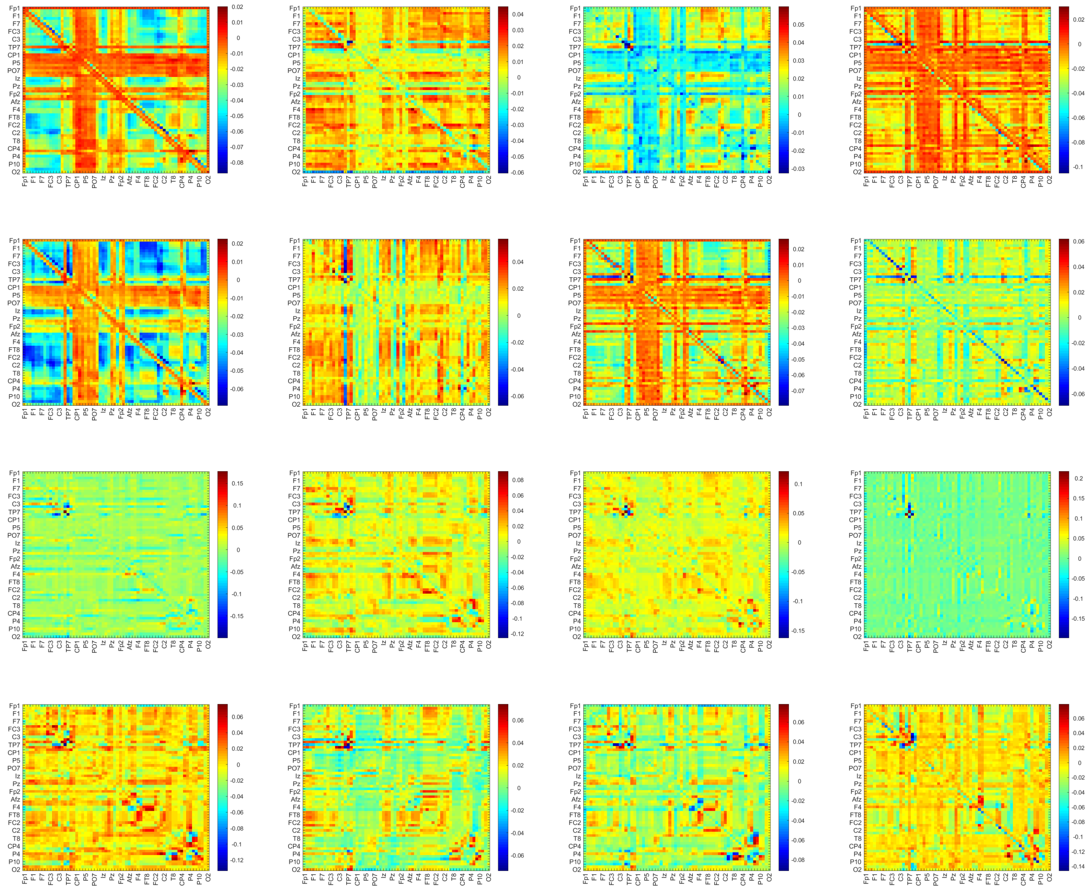


Figure 4.9: Output activations for eight kernels of the first Leaky ReLU layer in the proposed deep CNN measured between the average of alert and drowsy trials of S06. The depicted kernels belong to the α and γ PLVs from the $[-200, +100]$ ms time intervals. For improved readability, only one third of channel names have been included.

0.244 for S07). For others, the best PR-AUC from PLVs had slightly reduced in comparison to the best BP-ROI results with an average of 0.066 $\pm$0.051, min = 0.003 for S10 and max = 0.129 for S02.

After averaging the classification metrics across all participants, one-sided, two-way Student's t-tests were applied to the classification metrics of the two feature sets. PLV models of β-1, β-2, and γ oscillations outperformed the BP-ROI features from similar bands *and* and the set of all BP features in terms of F1 scores of each class, accuracy, and PR-AUC, $p < 0.001$. With an averaged PR-AUC in the range of [0.63-0.69], **pre-trial α ratios outperformed alpha synchrony features in all the classification metrics,** $p < 0.01$.

4.4.2.6 Visualization of Learned PLV Features from Convolutional Layers

Recent studies on EEG-based emotion recognition and detection of mental tasks in multi-class BCI datasets visualize and characterize activations of different layers to analyze the spatial relationships of features learned by CNNs and deep belief networks (DBF) [26], [27], [164]. In this work, as depicted in Figure 4.3, output matrices have the same dimen-

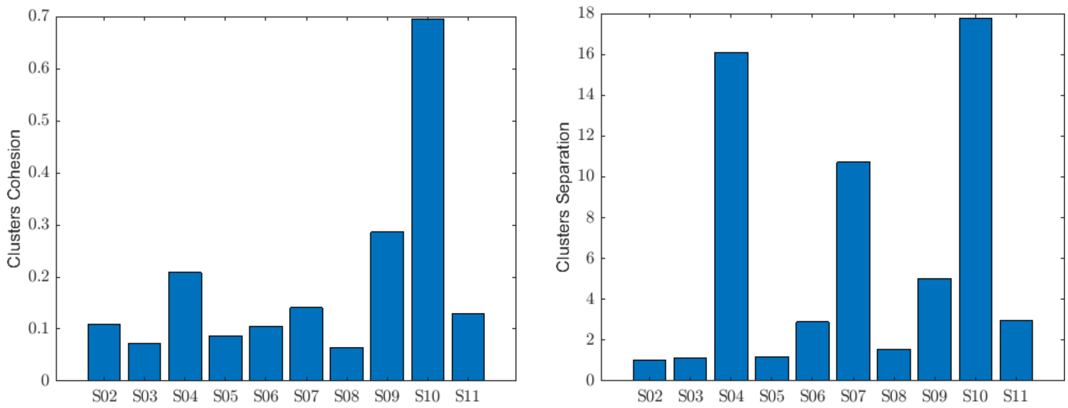


Figure 4.10: Cluster validity: The sum-of-squares based (a) cohesion and (b) separation for the drowsy and alert clusters constructed from the tails of the CVS histograms.

sions as the original input matrix up to the first max-pooling layer and their activations can be mapped back to the original channel pairs. We chose one participant with the highest detection rates to visualize the differences between the activations of hidden unit outputs of their drowsy and alert trials: S06 with a classification accuracy of 88.98% and 97.55% from the α and γ PLVs in the $[-200, 100]$ ms interval. We obtained the outputs of the first Leaky ReLU layer for all the drowsy and alert trials in the training set of one network run, and calculated the average of these activations for each class separately divided by the square root of number of samples in each class. This was done to accommodate for the imbalanced number of trials in the training set. We then obtained the differences between these averaged activations for each kernel for the alert class with respect to the drowsy class.

Figure 4.9 demonstrates the 8 resulting 64×64 activation matrices for α and γ bands. These heat maps demonstrate the network mostly responds to differences in phase synchronization between the left centro-parieto-occipital channels with the rest of the brain, and to a smaller degree, to the synchorny of inter-hemispheric frontal cortex and that of right fronto-temporal with left pre-frontal channels as shown in kernels 1, 3, and 5. For γ PLVs, the trained network is activated in response to differences in synchrony within the left centro-parietal and within the right parietal regions. Here, the absolute value of γ activations in this participant are less than 2 times stronger than those from α PLVs.

4.4.3 Discussion

In this study, we have, for the first time, reported the application of PLV features extracted from pre-trial and early onset time intervals in detecting low versus high vigilance trials using a deep convolutional neural network. In the rest of this section, we present evidence for the wellness of separation between the constructed low-CVS (drowsy) and high-CVS (alert) clusters, demonstrate not all participants are necessarily extremely alert in the very beginning and drowsy towards the session end, and discuss the implications of our findings on PLV correlates of alert and vigilant task performance.

4.4.3.1 Validation Analysis for Constructed Clusters

One common pitfall of trial labeling in Go/NoGo experiments is that individual trial vigilance scores solely represent one's momentary performance affected by phasic lapses of attention, demonstrated as delayed clicks or missed responses. To overcome this issue, we introduce the TVS as an objective, non-intrusive, and multi-level measure of vigilance that is adapted to changes in the response time of each individual with respect to their own initial reaction time [52]. In this study, the CVS, a cumulative measure to characterize the tonic variations during 36-trial windows is utilized as the ground truth for vigilance labeling over approximately 72-sec intervals. Dataset construction, feature analysis, and classification presented in this work are based on the assumption that all participants demonstrated two distinct alert and drowsy states. To test this assumption, we calculated the inter- and intra-cluster distance metrics for each individual. The Euclidean and squared Euclidean values for the complete linkage, average linkage, and centroid linkage distances were larger than the complete diameter, average diameter, and centroid diameter distances for all participants, verifying the validity of constructed low-CVS and high-CVS clusters. Figure 4.10 demonstrates the computed cohesion and separation values to visualize the inter-individual differences in the mental states that participants experienced during this long task. Cohesion here is equal to the total within-clusters' sum-of-squared (WSS) distance from the mean, and separation is computed from the between sum-of-squared (BSS) distance of the clusters' centroids from the overall mean. As can be seen from this Figure, S10, S04, S07, and S09 outperform other participants in terms of wellness of their clusters' separations and S10 shows the most centered drowsy and alert samples. These rankings are in line with the values of low-high thresholds reported in Table 4.3.

4.4.3.2 Temporal Distribution of Drowsy and Alert Samples

A number of studies on vigilance assessment in the context of BCI and simulated driving assume all participants experience the fixed order of alert, medium, and drowsy states linearly and uniformly [179]. To demonstrate inter-subject differences in mental state transition during our long SART sessions, we obtained the number of trials labeled as drowsy or alert in each of the 12 blocks, and calculated the cumulative sum of these events divided by the total number of drowsy and alert trials for each participant. As Figure 4.11 shows, although S06, S07, S08, S10, and S11 surpass 50% of their high-CVS trials within the first four experimental blocks, participants S02, S04, and S05 exceed 50% of their low-CVS trials during the same time period which implies they went through a relatively more drowsy state in the beginning of the task and had a different temporal pattern in their alertness.

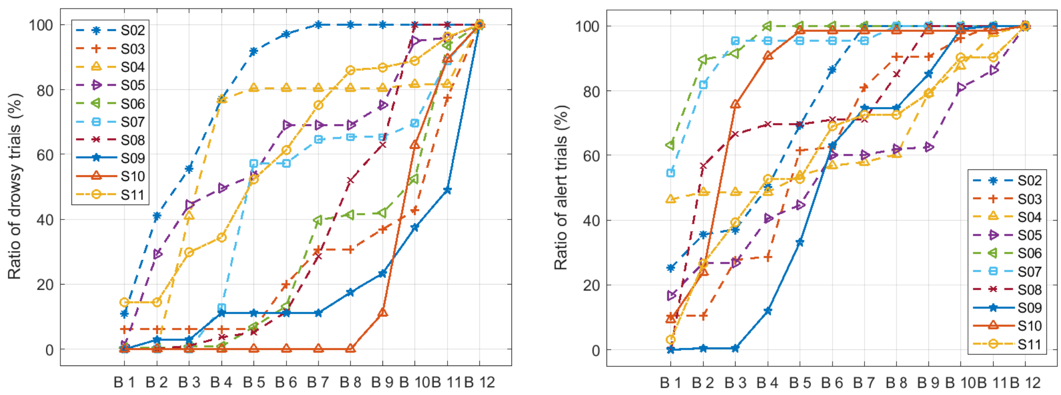


Figure 4.11: Individual differences in mental state transition during the long SART session: The cumulative ratio of trials labeled as drowsy (top) and alert (bottom) to the total number of similar trials plotted versus the SART blocks.

4.4.3.3 Role of Connectivity and Phase Synchronization Patterns

A number of studies have looked at the role of brain connectivity networks in the context of attention and drowsiness characterization. In a 20-minute continuous target selection task, α-band connections within the frontal regions had become stronger in the last block - labeled as the fatigued state - while connections and communications with other brain networks were disrupted [179]. However, introducing a break in the middle of the task resulted in an increase in the last block's connectivities among the temporal, parietal, and frontal regions, an indicator of an improved mechanism for information transfer for task execution. Disruptions in large-scale synchronization are known to be associated with cognitive fatigue [178]. More importantly, the long-range coherence in α band has been suggested to attenuate the noise caused by task-unrelated thoughts and neural activity [66]. The heat maps of Figure 4.9 from activations of the first convolution blocks for participant S06 with superior detection rates show a similar pattern: long-range synchronies of alpha oscillations from pre-trial to early onset phase are improved between the left posterior with the pre-frontal cortex, and inter-hemispheric, short-range γ connections are also enhanced in the frontal cortex. The high number of channel pairs with PLVs significantly larger in the drowsy state with respect to the drowsy state in this participant could reflect the same associations.

The pairwise γ synchronization also outperformed the slower oscillations in detecting drowsy versus alert states. It is know that γ synchronizations are enhanced inside the task-related regions with selective excitation [153], and such synchronies over the visual, auditory, or tactile regions can result in improved attention to the sensory stimuli. However, γ oscillations also require coupling with low-frequency and long-range connections for superior sustained attention. For example, in a subject-dependent classification of verbal versus quantitative tasks, δ and γ PLVs from within the pre-frontal and occipital regions and between the pre-frontal and occipital channels were selected as the most distinctive features [221]. In an auditory oddball task, higher γ band phase synchrony (GBPS) between frontal and posterior regions with increased task difficulty was interpreted as showing active interactions [182]. The same study also analyzed the factor of temporal evolution on the number of pairwise PLVs significantly different from the pre-stimulus intervals: the largest increases in γ and θ, both happening more strongly in the easy condition, belonged to the 300 to 400 ms interval in the midline frontal-parietal channels and 250-350 ms in the midline frontal-parietal, frontal-temporal, and temporal-parietal regions with respect to the stimulus onset. Hong *et al.* calculated matrices of phase syncrhony indices in a modified visual Go/NoGo experiment and averaged them for all the intervals in the baseline interval, from -500 to 0 ms, as well as the $[200, 700]$ ms intervals of the Go and NoGo conditions [222]. They observed that frontal-central θ synchronization was enhanced during response inhibition while the wide-band β synchrony had improved over the central-parietal regions during response execution. In a semantic auditory task, increased connectivity of frontal theta and disintegration of posterior alpha were observed at the sleep onset and were associated with suppressing the responsiveness [188].

Beta-band synchronization patterns are discussed less than the other frequency bands in the literature. In a study of fatigue caused by a 2-hour driving task, for 13–30 Hz oscillations, the interhemispheric central and parietal PLVs as well as the frontal-parietal, central-parietal, and middle and left frontal-central PLVs all significantly decreased after the task [201]. For S06, for example, the wide-band beta coherence was larger in the alert states between the posterior sites with the rest of the brain, and smaller between the left central/temporal with the right parieto-occipital region. These results are in line with those of an inter-subject regression using pre-trial phase synchrony index (PSI) in which lower $\beta-2$ features outperformed other features in predicting improved block-wise CVS [36]. Stronger desynchronization of lower $\beta-2$ in the right frontal network and between left centro-temporal channels with midline parieto-occipital channels, and increased synchrony within the right centro-parietal cortex were correlates of higher CVS.

4.4.3.4 Comparison with Similar Classification Studies

Our work demonstrated that a 3-layer CNN architecture applied to 64×64 images of gamma-band phase-locking values could achieve an average PR-AUC of 0.79 in 10 participants while exceeding 90% for 40% of participants. We reported the PR-AUC as a fair performance metric in imbalanced data sets and obtained our extreme alert and drowsy classes based on the objective CVS score and not the experimental conditions, but not all other studies have chosen to do so.

Figure 4.12 presents a number of state-of-the-art, EEG-based classification systems for extreme drowsiness or low performance detection with the types of features and classifiers, classification metrics, and major reported results. For a within-subject detection of extremely alert versus drowsy driving conditions using preprocessed EEG signals, EEG-Conv and EEG-Conv-R, a modified version of convolutional layers combined with residual learning [223], obtained an average accuracy of 91.79% and 92.68% and outperformed SVM and simple LSTM networks with accuracies of 88.07% and 85.13% [224].

The authors did not report their PR-AUC results. In a work on good and poor performance detection based on fixed-threshold response time in simulated driving, DNNs and CNNs obtained an average area under the curve (ROC-AUC) of 80% while CNNs applied on channel-wise segments achieved a detection rate of 86% in within-subject classification [168]. In another study on within-subject classification of drowsy versus states purely based on adaptive RT thresholds in 10 participants, ICA sources achieved an ROC-AUC of 0.745 while power-based features only gained an equivalent value of 0.671 [13]. Finally, for a cross-task mental workload assessment and using spatio-spectro-temporal features, kNN had the lowest accuracy of just around 0.7 and the linear SVM and LDA had performed better while still falling below the proposed model with a concatenated RNN and 3D CNN architecture [197]. Their proposed R3DCNN model had an average accuracy of 0.889 that was higher than the DEEP CNN model of [26].

4.5 Conclusion

We have demonstrated associations between response styles and distinctions among the obtained spatio-spectro-temporal features. Participants with lower abilities to maintain their attention levels also demonstrated improved detection rate using PLV features from the pre-trial and early post-digit EEG markers. The implemented preprocessing and feature extraction steps are completely automated and do not rely on human expertise to identify the nature of calculated independent components for artifact rejection. Furthermore, our multi-level scoring scheme only relies on each participant's initial response speed to determine the upper RT threshold. To sum up, combining these blocks with our feature extraction techniques and the proposed deep CNN architecture for PLV classifications results in a successful end-to-end vigilance detection system.

Paper	Task	Labeling	Participants	Duration	EEG Channels	Segments or Epochs	Input Features	Classifiers	Classification Metric	Classification Results	
Buriro 2018	Microsleep vs. responsive states	Manual annotation	8	Two 1-hour session per participant	16	EEG sub-bands, decimated to 5-s epochs, 250 ms steps	7 inter-channel features	LDA and linear SVM	PR-AUC	LOO-CV LDA: 0.42 (Cov), 0.38 (WCSP), 0.44 (JE), 0.41 (WSP)	LOO-CV SVM: 0.45 (Cov), 0.36 (WCSP), 0.32 (JE)
Shoorangiz 2019	"	"	"	"	"	"	Regularized spatio-temporal filtering and classification (RSTFC)	"	"	0.52 (LDA), 0.42 to 0.50 (SBL), and 0.48 (VBLR)	
Zeng 2018	Extremely alert vs. drowsy driving states	Enforced by block-wise design (engaging vs. boring)	10	2 hours (~25 min used for classification) per participant	15	0.5-s 15 x 128 matrices	Raw data	EEG-Conv (4 convolutional layers), EEG-Conv-R (with residual learning)	Accuracy	Within-subject: 0.918 (EEG-Conv), 0.927 (EEG-Conv-R), 0.880 (SVM), 0.851 (LSTM)	Cross-subject: 0.829 (EEG-Conv), 0.844 (EEG-Conv-R), 0.818 (SVM), 0.755 (LSTM)
Zhang 2019	Cross-task (spatial vs. arithmetic) mental workload assessment	Enforced by experiment design	20	70 min per participant	16	20 epochs, 1.8 s x 40	EEG cubes of spatio-spectro-temporal features	R3DCNN (8 convolutional layers and 2 bidirectional LSTM layers)	Accuracy	Average of cross-task, cross-subject: 0.889 (R3DCNN), 0.7 (KNN), 0.842 (deep CNN)	
Hajinoroozi 2016	Good vs. poor driving performance	Fixed RT thresholding	37	80 sessions in total	30	1-s: 30 x 150 matrices	Raw data	Channel-wise CNN (CCNN), CCNN with Restricted Boltzmann Machine (CCNN-R)	ROC-AUC	Within-subject: 0.861 (CCNN), 0.807 (CCNN-R)	Cross-subject: 0.829 (EEG-Conv), 0.844 (EEG-Conv-R), 0.818 (SVM), 0.755 (LSTM)
Hsu 2017	Alert vs. drowsy driving performance	Adaptive RT thresholding	10	90 min per participant	30	90-s blocks with 30-s overlap	ICA sources and distance of BP from training data	Model deviation index (MDI)	ROC-AUC	0.745 (ICA), 0.671 (power)	
Proposed BP-ROI Dataset	Extremely alert vs. drowsy Go/NoGo states	Errors and adaptive RT thresholding	10	105 min per participant	64	14 x 10 matrices	Pre-trial BP features from 14 ROIs	SVM (linear, quadratic, Gaussian), DT, kNN, NB, DT with Bagging, AdaBoostM1, RUSboost, kNN with subspace learners	PR-AUC	Within-subject: 0.725 (Gaussian SVM, all features)	
Proposed PLV-abs Dataset	"	Errors and adaptive RT thresholding	"	"	"	64 x 64 matrices	Pre-trial and early onset PLV from band-passed EEG	Proposed CNN Model	PR-AUC	Within-subject: 0.79 (pre-trial to early onset Gamma)	

Figure 4.12: State-of-the-art, EEG-based classification systems for extreme drowsiness or low performance detection.

5 Bayesian Models for Changepoint Detection in Vigilance Time-Series

During the classification and regression approaches of the last two chapters, we paid a special attention to inter-individual differences of maintaining consistent performance that demonstrate themselves as highly different patterns of the CVS curves as shown in Figure 4.4. An important question raised at this point is whether variations in vigilance curves and their EEG-based predictors could be tracked and modeled across time, and whether the exact *moments* at which a vigilance transition occurs can be detected or predicted. A potential solution would help with building adaptive BCIs that detect the onset of vigilance changes and, consequently, modify the classification parameters in covert adaptation or the interface parameters in the overt adaptation schemes. For these reasons, in this chapter we focus on the dynamic modeling of vigilance curves and their EEG predictors in general and on changepoint detection (CPD) from those time-series in particular.

Changepoints are defined as the onsets of abrupt changes in the underlying properties of time-series data [225] and can be detected using supervised methods or a variety of statistical, probabilistic, kernel-based, subspace-based, or clustering algorithms [226]. CPD from non-stationary data has been moderately explored in finance and stock markets [227], speech recognition for detecting temporal borders between speech, noise, and silence [228], continuous patient monitoring through tracking of physiological signals such as EEG and brain imaging data for seizure onset detection [229], [230], ECG for heart rate monitoring [231] and workload detection [232], glucose data for miscarriage risk detection [233], and human activity recognition from a variety of wearable sensors [234]. However, to the best of our knowledge, CPD has been rarely used for drowsiness detection in dynamic models of cognitive functions and even less frequently for continuous vigilance modeling from the behavioral and neural/EEG signals. Similar to automatic seizure onset prediction that prevents accidents and saves lives by alarming the users about an upcoming intracranial seizure, changepoint detection from vigilance time-series can notify the operators about the onset of drowsiness and enable autonomous or semi-autonomous systems such as vehicles and radars as well as psychologists and clinicians, BCI experimenters, educators, and parents to present manual or automatic adaptations in response to attention variations in their users, patients, and learners. However, unlike EEG-based epilepsy datasets in which onsets of various seizures as well as their time intervals are annotated by clinicians for easier labeling of pre-ictal periods [49], [235],

exact moments of vigilance decline are not readily clear in our calculated CVS curves and their accompanying EEG datasets. This brings an additional challenge regarding the ground truth annotation in physiological and cognitive datasets (Ref. Section 1.2).
In this chapter, we utilize two offline and online Bayesian changepoint detection (BCPD) algorithms to locate transitions of vigilance curves and their EEG predictors. We validate the performance of best features and models on SEED-VIG, a dataset labeled with continuous-valued eye-closure events from driving simulations [35]. This novel fusion of band-power vigilance predictors with the purely statistical CP detection is an indicator of success of unsupervised inference models that will be subsequently used in Chapter 6 for designing an adaptive alertness-aware classification system. It should be noted that we analyzed a variety of changepoint detection algorithms such as the Autoregressive Moving-Average (ARMA), Metropolis-Hastings (MH) [236], [237], and Gibbs sampling [238] based on pre-trial BP features and their distances with respect to the beginning of experimental sessions; however, the Bayesian changepoint detection models reported here outperform all of them in terms of the number of true changepoints (TCP) detected from vigilance time-series and the insensitivity to the choice of initial parameters.

5.1 Motivation

We propose that a BCI that is to be adapted to transitions in sustained attention level of its users has to be *aware* of the underlying vigilance level in real time or near real time, and *react* to vigilance variations through updating the classification or environment parameters. The work presented in this chapter is motivated by the following methodological questions and observations as a result of analyzing CVS curves and their correlations with pre-trial EEG features in Figure 4.4.

1. Is it possible to track changes in the statistical distributions of behavioral vigilance curves, vector y, and their EEG predictors, matrix X, in a dynamic setting to detect when a person's long-term, tonic attention starts to decline?

   Here we distinguish between the short-term and phasic or event-related variations caused by momentarily lapses of attention and tonic changes that are caused by fatigue and drowsiness [64].

2. In an actual BCI experiment, vigilance scores or CVS curves may not be known since the user does not provide a physically detectable response to the observed stimuli. Therefore, relying mainly on the EEG features is preferable. Will the CPD algorithms detect identical or close changepoints from the vector y and neural time-series X? In other words, can changepoints obtained from EEG correlates analyzed in the previous chapters point to correct transitions in vigilance performance curves, and subsequently, be used to infer vigilance transitions even in unlabeled cognitive and BCI datasets?

3. The spatio-spectral properties of EEG features in matrix X impose strong correlations among neighboring channels and frequency bands. Therefore, in the CPD algorithm of choice, should we use single or multi-variate time series for learning these change points? In [239], Xuan and Murphy propose two Gaussian-based solutions, the Independent Features Model (IFM) and Full Covariance Model (FCM). In both cases, the joint densities are obtained according to the method of conjugate priors. How can one deal with the multidimensionality of EEG features X for finding the common changepoints of these time series?

4. Because of the final goal of "adapting" a BCI classifier to the underlying cognitive state of the user, we prefer to detect moments of decline in alertness in a "real-time manner" or with a short delay. Can we implement an ε-real time algorithm that needs a minimum of ε samples in each new batch for CPD [226], or an online changepoint detection system as in, for example, the Bayesian Online Changepoint Detection (BOCD) of Adams and MacKay [240] or online inference algorithm of Fearnhead and Liu [241]?

In response to these questions, the following methodological and technical contributions for vigilance changepoint detection from behavioral and neural time-series data are presented in this chapter:

1. Inferring vigilance transitions for unsupervised changepoint detection: By acknowledging large inter-individual variations in vigilance traits and without enforcing any assumption on the maximum number of possible changepoints or their distances in the time-series, Bayesian offline and online CPD algorithms are applied on vigilance-related performance curves – when available in the SEED-VIG and our SART dataset – to obtain onsets of vigilance level transitions. This approach, in our opinion, is the least constrained and most applicable choice for modeling human behavior and cognitive activities. To be more precise, after applying these algorithms on the objective, continuous-valued vigilance curves to obtain changepoint locations, hereafter known as true changepoints or TCPs, we assess the performance of EEG-based features in reporting similar change onsets, hereafter referred to as the EEG-CPs, in the vicinity of detected TCPs.

2. Detecting changepoints from EEG-based time-series: Realizing that EEG time-series may report several changepoints due to momentary transitions, both the behavioral curve and and EEG time series are first accumulated over a window of size w. Although PLV and PSI features outperformed the pre-trial BP-ROI datasets during the regression and classification tasks, the high dimensionality of the former modality – 2,016 unique channel pairs in the 64-channel EEG setup – makes it extremely demanding for use in single or multi-dimensional changepoint detection problems. BP-ROI features also had more consistent correlation patterns with vigilance curves (Ref. Figure 4.4 for the SART and Figure 6.9 for SEED-VIG datasets).

BP-ROI features act as a proof-of-concept method with an acceptable performance, and we envision to upgrade the system with functional connectivity or PLV feature in the future.

3. Lower complexity and improved localization: The ARMA, Metropolis-Hastings, and Gibbs sampler methods for changepoint detection require a large number of parameters for initial and proposed symmetric or independent kernels that quickly add up in modeling non-stationary EEG features [242]. Furthermore, our preliminary analysis demonstrated the immense sensitivity of detected changepoints to the selected time segments and choice of initial parameters. This is while, in our experiments, the Bayesian online and offline models generalize better for the behavioral curves from which the ground truth are to be obtained and EEG feature time-series, have a time complexity of $O(n)$ if solved approximately, and are our algorithms of choice in this work. We plan to return to this problem in the future and develop an autoregressive Bayesian changepoint model that would detect changepoints with similar or better accuracy.

4. No prior information on the number of segments or states: As a powerful inferential technique, HMMs need to have prior information on the number of states to estimate the transition and emission probabilities. Our experiments on CVS curves demonstrated that a number of participants never go through more than one or two alertness levels during their training data while falling into extreme drowsiness long after. Therefore, it is not possible to set the exact number of states a priori. An infinite HMM would be a better algorithm of choice assuming old alertness states can be revisited during the experiments [239]. One recent study used the LOO-CV scheme to estimate the number of brain states [243]. One could also use the product partition model (PPM) with independent or dependent parameters across segments based on assumptions on the occurrence of old states [244].

5. Performance evaluation: The online, ε-real time, or offline (∞-real time) algorithms [226] return different changepoints when applied to a time-series. We used those TCPs as the separate ground truth for vigilance curves y, and introduced an evaluation window of length L_{eval} to detect the occurrence or localization of EEG-based changepoints in a supervised manner by reporting the precision and recall values [245] as is the common practice in expert-annotated datasets. An unsupervised learning view can also be implemented to report a distance metric such as the mean absolute error (MAE) or the Jaccard index and its derivations [226], [245].

The rest of this chapter is organized as follows. Section 5.2 provides a summary of important literature on changepoint detection and their applications for cognitive and medical applications. Sections 5.3 introduces the problem formulation for changepoint detection using a product partition model, and presents details of a scheme for online and offline BCPDs from the SEED-VIG dataset. Section 5.4 includes results and comparisons between the three utilized algorithms using the supervised performance metrics. This chapter is concluded in Section 5.5 with a summary of important findings.

5.2 Related Work

The literature on changepoint detection covers a variety of parametric and non-parametric algorithms, namely the probabilistic models, subspace and kernel based methods, probability density ratio estimation such as cumulative sum (CUSUM) and autoregressive (AR) models [215], [225], and classification techniques including hidden Markov models (HMMs) and Gaussian mixture models (GMMs). A number of studies also merged the fields of Bayesian and AR models and developed Bayesian autoregressive changepoint detectors [242], [246].

In this chapter, we focus on probability-based, Bayesian inference techniques due to their low computational cost ($O(n)$ if solved approximately), small number of parameters, and ease of calculating the closed-form formula. In this context, Fearnhead and Xuan have several pieces of work on changepoint detection from single and multivariate time-series using the product partition model (PPM) [239], [247]. Adams and MacKay have a similar PPM algorithm with independent segment parameters for online CPD based on run length distributions [240].

In the context of brain imaging data and EEG signals, several studies have classified experimental recordings into multiple states for sleep stage classification using deep architectures and infinite HMMs [248], [249] or performed alert versus microsleep and alert versus drowsy interval classification using source separation and pairwise inter-channel features [191], [250]. Very few studies have, however, attempted to use dynamic inference for real-time detection of changepoints. For example, Gao *et al.* presented a linear autoregressive model that detected changepoints in an expert-annotated EEG dataset using a sum of entropy values from time-domain features [251]. Zheng *et al.* used Pruned Exact Linear Time (PELT) [252] to detect epileptic seizures in calcium imaging video data from zebrafish [229]. Chen *et al.* applied a sliding window approach and calculated the similarity metrics between subsequent power features to detect focal and non-focal EEG segments from labeled datasets [230]. Guo *et al.* used the least squares method to detect anxiety during driving from eye movement data [253]. The ground truth for extracted sessions were obtained from subjective declarations of participants and a number of psychological questionnaires. A hidden semi-Markov model (HSMM) was applied on fMRI data to estimate time-varying brain networks during anxiety-induced sessions [243]. A LOO-CV scheme was used to obtain the optimal number of states across all participants. In a different work, graph-theoretical methods were applied on functional connectivity matrices for temporal changepoint detection [254]. The most comprehensive review of Bayesian connectivity changepoint detection on EEG time-series is however presented in [255] for application for clustering subjects into different groups rather than detecting objective transition moments from their individual time-series.

5.3 Methods and Experiments

We start this section by presenting a formal problem formulation for detecting changepoints from an ordered time sequence or time series, and continue by presenting details of a proposed CPD scheme using online and offline algorithms for detection and evaluation of vigilance transition moments using behavioral and EEG time-series.

5.3.1 Problem Formulation

Let $y_{1:T} = (y_1, y_2, ...y_T)$ denote a set of observations from T trials in an experiment. If the statistical properties of the segment $(y_1, y_2, ..., y_\tau)$ is different from those of segment $(y_{\tau+1}, ..., y_T)$, then a changepoint has occurred at time τ. Extending it to a multiple changepoint framework, the sequence of $n+1$ ordered changepoints $\tau_{0:n}$ divides the time-series into n segments where $\tau_0 = 1$ and $\tau_n = T$ is the last point. Segments are usually assumed to have been generated from similar distributions with different parameters, with segment j located between τ_{j-1} and τ_j and generated using parameters θ_j, $j = 1, ..., n$.
A common approach for single changepoint detection is using a likelihood framework based on a null hypothesis that all observations in the sequence y are extracted from a single distribution with similar statistical properties versus the alternative hypothesis that $(y_{1:\tau}) \sim \theta_1$ and $(y_{\tau+1:T}) \sim \theta_2$. The changepoint location τ is then equal to the maximum likelihood estimate (MLE) of the point of transition between the two statistical models under the alternative hypothesis. But here we follow a product partition model (PPM) in which the number of non-overlapping partitions or segments n is unknown while the data themselves are assumed to be independent across these partitions or segments, i.e.,

$$p(y_{1:T}|\pi) = \prod_{j=1}^{n} p(y_{\pi_j}), \tag{5.1}$$

where π_j denotes the prior density for parameters θ_j of segment j [239], [247]. Note that Fearnhead has a different perspective in which he assumes dependence among the partition parameters [244]. He also assumes the length of each segment follows a geometric distribution. This prior on segment length affects the partition prior $p(\theta)$ as well.
To find the optimal number of changepoints and their priors, the posterior probability function of $p(\pi_{1:n}|y_{1:T})$ has to be maximized, resulting in the following maximum a posteriori (MAP) estimation:

$$(n^*, \pi^*) = \underset{n, \pi_{1:n}}{\arg\max}\, p(y|\pi_{1:n})\, p(\pi_{1:n}). \tag{5.2}$$

[247] defines another prior function $g(t)$ for the time between two successive points in the range of 1 to T which should be strictly positive and independent across all segments. A uniform or negative binomial distribution would satisfy this constraint. Following a series of recursions, the following probabilities need to be computed:

1. The observation log-likelihood $\log P(t,s) = \log P(y_{t:s}|t,s$ in the same segment$)$ or

no changepoint occurring within their segment, $s \geq t$, is calculated from

$$P(t,s) = \int_{\theta} \prod_{i=t}^{s} f(y_i|\theta)\pi(\theta)d\theta. \tag{5.3}$$

2. $\log Q(t) = \log P(y_{t:T}|\tau_{t-1})$ or the log-likelihood of $y_{t:T}$ conditioned on a changepoint happening at time $t-1$. Here, $t = 2,3,...,T$.

3. $\log Pcp(j,t) = \log P(y_{t:T}|\tau_j = t)$ or the log-likelihood that the j-th changepoint happens at time t, $t = 2,3,...,T$.

Note that when the number of changepoints n is unknown, a prior $\pi(n)$ is needed from which the following posterior distribution can be computed:

$$P(n|y_{1:T}) = \pi(n)P(y_{1:T}|n) \tag{5.4}$$

In what follows, we describe different models for observation likelihood functions.

5.3.1.1 Piecewise Gaussian Observation Model

In the piecewise Gaussian Observation Model (GOM), we assume that each single-dimensional observation y_i located in segment j has a Gaussian distribution with an unknown mean and known variance, i.e., $y_i \sim N(\mu, z)$ where $\mu \sim N(\nu, \kappa)$ and $z \sim IG(\alpha, \beta)$. IG represents an inverse Gamma distribution. Using Equation 5.3,

$$P(t,s) = \int_z \int_{\mu} \prod_{i=t}^{s} \phi(y_i;\mu,z)\, p(\mu)\, p(z)\, d\mu\, dz \tag{5.5}$$

where $\phi(y_i;\mu,z) = \frac{1}{\sqrt{2\pi z}} exp(\frac{-(y_i-\mu)^2}{2s})$. The log-likelihood function of $P(t,s)$ will then be calculated and used in the changepoint detection algorithm to obtain the changepoint likelihood function, Pcp, for each time point.

5.3.1.2 Independent Features Model

If $y \in R^{d \times T}$, we can start by assuming independence among its d dimensions so that,

$$p(y_{t:s}) = \prod_{k=1}^{d} p(y_{t:s},k), \tag{5.6}$$

hence the name Independent Features Model (IFM). Since $p(y_{t:s}) = \int p(y_{t:s}|\theta_k).p(\theta_k)d\theta_k$, we obtain the following expression for the probability of this segment's observation having been generated by a model with the parameter θ_k

$$p(y_{t:s}) = \prod_{k=1}^{d} p(y_{t:s},k) = \prod_{k=1}^{d} \left(\int \left(\prod_{i=t}^{s} p(y_{i,k}|\theta_k) \right) p(\theta_k)d\theta_k \right). \tag{5.7}$$

Now assume that observation in dimension k from time i is extracted from a normal distribution with a zero mean and an unknown variance that follows an inverse Gamma distribution [239]. In other words, $y_{i,k} \sim N(0,\sigma_k^2)$ where $x = \sigma_k^2 \sim IG(N_0/2, V_{0k}/2)$. The joint density in front of the integral can then be written as follows:

$$p(y_{t:s}|\theta_k).p(\theta_k) = \frac{(V_{0k}/2)^{N_0/2}}{\Gamma(N_0/2)}\, e^{-\dfrac{V_{0k}}{2x}}\, x^{-N_0/2-1} \prod_{i=t}^{s} \frac{1}{\sqrt{2\pi x}}\, e^{-\dfrac{(y_i-0)^2}{2x}} \tag{5.8}$$

which can be rewritten as

$$p(y_{t:s}|\theta_k).p(\theta_k) = \frac{(V_{0k}/2)^{N_0/2}}{\Gamma(N_0/2)}\, (2\pi)^{-m/2}\, exp(-\frac{V_{0k}+\sum_{i=t}^{s} y_{i,k}^2}{2x})\, x^{-m/2-N_0/2-1}. \tag{5.9}$$

Here, m is the length of segment. Integrating with respect to θ_k and using the formula of Γ function, we obtain

$$p(y_{t:s},k) = \pi^{-m/2}\, \frac{V_{0k}^{N_0/2}}{V_{mk}^{(N_0+m)/2}}\, \frac{\Gamma(N_0/2)^{-1}}{\Gamma((N_0+m)/2)^{-1}}, \tag{5.10}$$

where $V_{mk} = V_{0k} + \sum_{i=t}^{s} y_{ik}^2$. By setting N_0 to the dimension d and empirically calculating V_0 from the variance of flattened version of $y_{t:s}$ from its entire features, one can calculate the log likelihood of this segment having been generated by the prior density j, $\log p(y_{t:s}|\pi_j)$.

5.3.1.3 Full Covariance Model for Multidimensional Time-Series

The Full Covariance Model (FCM) is developed to consider the inter-feature correlations. Let us assume the observation at each time point follows a multivariate normal distribution, i.e., $y_i \sim N(0,\Sigma)$, and $x = \Sigma \sim IG(N_0, V_0)$. The marginal likelihood then simplifies to

$$p(y_{t:s}) = \pi^{-md/2}\, \frac{|V_0|^{N_0/2}}{|V_m|^{(N_0+m)/2}}\, \frac{\Gamma_d(N_0/2)^{-1}}{\Gamma_d((N_0+m)/2)^{-1}}, \tag{5.11}$$

where $V_m = V_0 + S$, $S = \sum_{i=t}^{s} y_i y_i^T$, and $\Gamma_d(N_0/2) = \pi^{d(d-1)/4} \prod_{i=0}^{m-1} \Gamma(m-i/2)$.

The parameter V_0 in this setup is set to a diagonal matrix with previously calculated pooled variance. This model is said to work well for up to five dimensions since there might not be enough data to estimate the covariance matrix Σ of relatively shorter segments [239]. Therefore, we try to limit the number of dimensions and use subsets of features X to use for this model rather than using the whole pre-trial BP-ROI features.

5.3.1.4 Online Detection using Bayesian Inference

Suppose that data samples arrive sequentially in a real-time experiment, and that the probability of each new observation belonging to either the previous hyperparameters or to a new model needs to be calculated. Following a product partition model for non-overlapping segments as before, the model of Adams and McKay [240] assumes the run length r_t increases linearly as a function of time until a changepoint occurs, at which time it becomes equal to zero. Using the posterior distribution $P(r_t|y_{1:t})$ and the predictive distribution conditioned on a given run length, $P(y_{t+1}|r_t, y_t^r)$, the marginal predictive distribution $P(y_{t+1}|y_{1:t})$ is computed as follows:

$$P(y_{t+1}|y_{1:t}) = \sum_{r_t} P(y_{t+1}|r_t, y_t^r)\, P(r_t|y_{1:t}). \tag{5.12}$$

This algorithm then recursively obtains the joint distribution over the run length and observed data, $P(r_t, y_{1:t}) = \sum_{r_{t-1}} P(r_t, r_{t-1}, y_{1:t})$.
The outputs of this algorithm include the matrix R that contains $= P(r_t|y_{1:t})$, the posterior probability for the current run length given a changepoint had happened r_t steps earlier, and the vector v_{map} that contains the maximum a posteriori (MAP) estimate or $\hat{r}_{tMAP}$ from $P(r_t|y_{1:t})$, denoting the maximum run length at time t.

5.3.2 Proposed Online and Offline Vigilance Changepoint Detection Schemes

The SEED-VIG dataset [35] is used for the first validation attempt for offline and online changepoint detection algorithms. This dataset is recorded from 21 participants who performed a virtual driving task for 118 minutes. Each recording includes 17 EEG channels located on the centro-parietal, parietal, occipital, and temporal cortices and 4 forehead EEG and EOG channels which recorded signals at the sampling rate of 200 Hz. This dataset includes PERCLOS labels that represent the PERcentage of eye CLOSure duration during 8-second non-overlapping windows through direct computation with SMI eye-tracking glasses. Thus, each participant has a vigilance time-series y with 885 observations or labels between zero and one. This measure goes beyond the usual and autonomous eye blink events and thus represents increases in drowsiness, and in more extreme cases, the absolute sleep status of each individual. Figures 5.1 to 5.4 demonstrate PERCLOS curves for four different SEED-VIG participants.
In our proposed scheme, EEG-based BP-ROI features of Chapter 4 are calculated from one-second time intervals and averaged across similar non-overlapping 8-second windows. Due to the presence of EEG electrodes across the temporal and parietal regions in the SEED-VIG dataset, only 6 non-empty ROIs are formed which result in a 60-dimensional feature set. A correlation analysis demonstrates that 90% of participants had significantly positive correlations between their PERCLOS labels and the α/β features from all the six ROIs. In other words, an increase in these features is associated with increase in drowsiness and reduction in vigilance. Thus, the six α/β features are selected as the common neural markers and EEG predictors in these CPD experiments

and organized in matrix X.

The following CPD scheme is then applied on the behavioral labels and EEG-based correlates of vigilance variation:

1. A moving average window of length 10 is applied to vector y, the PERCLOS time-series of each participant to obtain summarized tonic vigilance variations represented in an ordered sequence $y_{1:T}$, $T = 876$.

2. Offline TCP detection from the vector y: Using IFM and GOM observation likelihoods from y, the T-by-T matrices P_{IFM} and P_{GOM}, and (T-1)-by-(T-1) matrices Pcp_{IFM} and Pcp_{GOM} of vector y are computed using this toolbox as a foundation. Pcp matrices are marginalized to obtain the marginal likelihoods of having a changepoint at each time t between from 1 to T. A lower threshold of λ = 0.01 is applied to smooth the density function, and all the peaks above that threshold are calculated to represent true changepoints τ_{IFM} and τ_{GOM} for each PERCLOS time-series y.

3. Offline EEG-CP detection from the matrix X: Similar to [239], applying the full covariance model (FCM) on the 6 α/β features resulted in an unrealistic overestimation of changepoints. Thus, we apply IFM and GOM on each dimension of matrix X separately. After using a similar moving average window of length 10, EEG changepoints $EEG\tau_{IFM}$ and $EEG\tau_{GOM}$ are calculated similar to step 2 for each feature in $X \in R^{6\times T}$.

4. Online TCP detection from the vector y: Using a uniform distribution between 1 to 250 for the Hazard function and a student's t-distributions for the vector of predictive probabilities $\pi_t^{(r)}$ [239] with $(\alpha, \beta, \kappa, \mu)$ = $(0.1, 0.01, 1, 0)$, the maximum run length estimation $\hat{r}_{MAP}$ is computed that demonstrates a sawtooth-pattern. Locations of τ_{online} are obtained from $\hat{r}_{MAP}$ local peaks.

5. Online EEG-CP detection from the matrix X: The method in step 4 is used to compute $EEG\tau_{online}$ for each of the 6 individual features in matrix X.

6. Using τ_{IFM}, τ_{GOM}, and τ_{online} as the ground truth for locations of transitions in the PERCLOS vectors, the number of detected EEG-CP for each approach is obtained inside an evaluation window with L_{eval} = 10 trials. Subsequently, supervised performance metrics are calculated to determine which observation likelihood could lead to the highest precision and recall rates for detecting changepoints from EEG features with a 10-sample lag. Precision and recall are defined as

$$Precision = \frac{TP}{TP+FP}, \qquad Recall = \frac{TP}{TP+FN}, \tag{5.13}$$

where $TP+FP$ indicates the number of unique EEG-CPs detected by a model from 6 EEG features, and TP indicates the number of true positives, i.e., the number of unique EEG-CPs detected using that model and located within 10 samples from a TCP. Finally, $TP + FN$ consists of the total number of TCPs located from the PERCLOS using a model. It should be mentioned that due to the different ground truths used in online and offline algorithms, the reported precision and recall values are compared separately.

5.4 Results

Three models used for the observation likelihood function consisting of the online Bayesian model of [240], the offline individual features model (IFM) [239], and the piecewise Gaussian model (GOM) [247] were run separately on the PERCLOS time-series y and six-dimensional α/β matrices X. Figures 5.1 to 5.4 demonstrate the detected change-points for participants S01, S16, S18, and S21 from the SEED-VIG dataset. The PERCLOS curves smoothed with a 10-point window are shown on the top plots with red points denoting locations of the detected τ_{online} calculated from the peaks of the run length MAP. The second plot demonstrates the locations of all the detected $EEG\tau_{online}$ which occasionally overestimate the number of τ_{online}. Plots 3 and 4 demonstrate the marginal likelihoods for CPD detection from PERCLOS using the offline IFM and GOM algorithms as well as their detected peaks τ_{IFM} and τ_{GOM}. Finally, the marginal likelihoods from all the six α/β features as well as the $EEG\tau_{IFM}$ and $EEG\tau_{GOM}$ locations detected offline are presented at the bottom of each figure.

The online, IFM, and GOM obtain an average number of 31.24, 5.19, and 9.95 true changepoints from the PERCLOS curves. An analysis of variance (ANOVA) demonstrates a significant factor of algorithm, $F(2,60) = 109.38$, $p < 0.001$, among the number of detected ground truths from PERCLOS curves from all participants. It can also be observed that the online algorithms is more sensitive to phasic, short-term variations in PERCLOS curves and some of its detected changepoints may not lead to true onsets of vigilance variations. Table 5.1 demonstrates the average precision and recall for all participants using the online and offline changepoint detection algorithms. As can be seen, EEG features are able to detect the locations of τ_{online} with an average precision and recall of 0.82 ± 0.26 and 0.56 ± 0.14 which may point to the sensitivity of the online algorithm to small variations in the curves that result in CPD not necessarily corresponding to transitions in vigilance time-series. The performance of CPD from EEG time-series using offline algorithms are more comparable in terms of their precision values although their τ_{IFM} and τ_{GOM} are essentially different from each other.

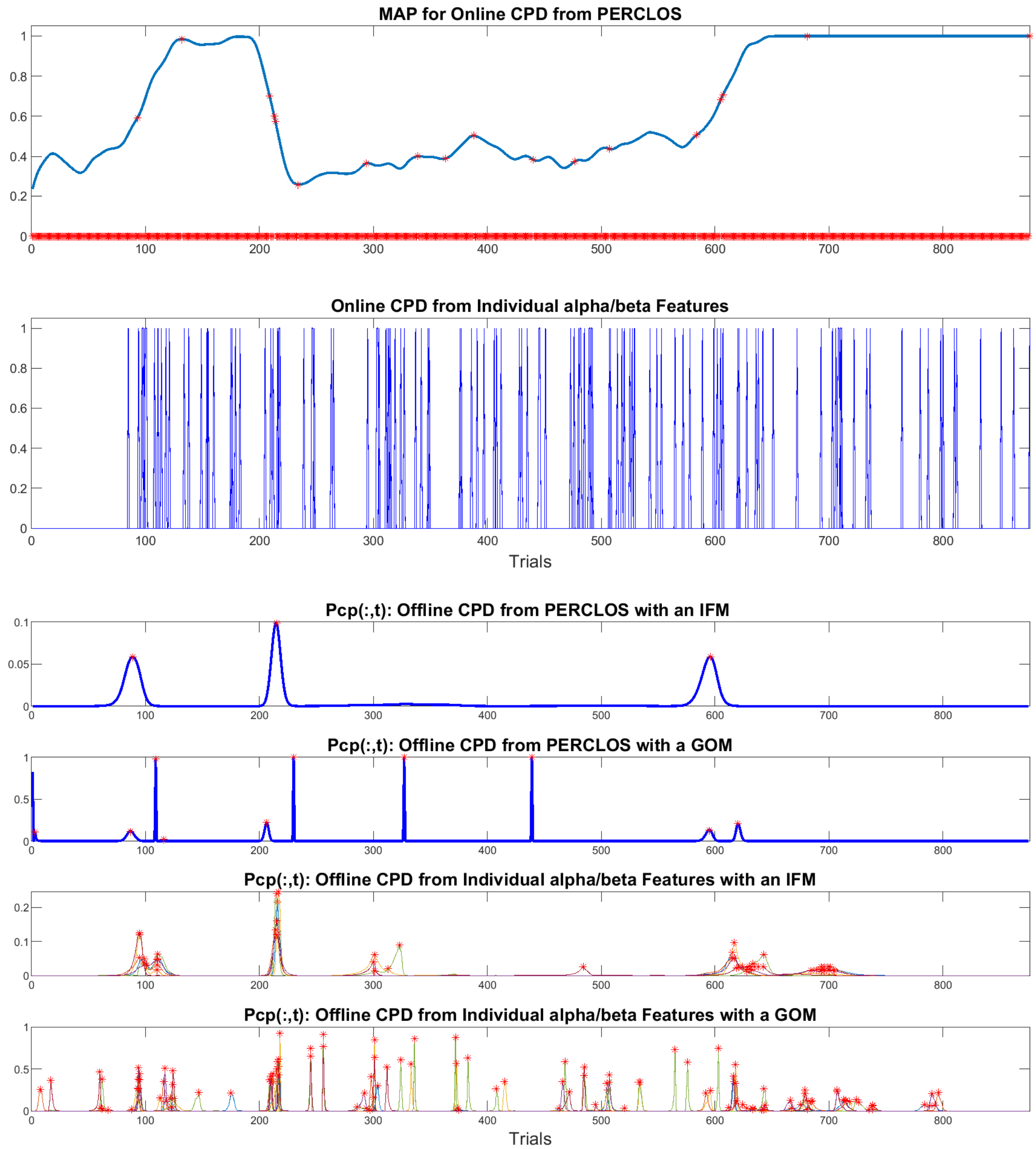


Figure 5.1: Changepoint detection from PERCLOS and EEG α/β time-series for participant S01. In the top plot, the blue curve represents the original PERCLOS, and red points denote locations of TCPs detected using the online algorithm. EEG-CPs from the same online algorithm are shown in the second plot. In the middle plots, blue curves indicate the independent features model (IFM) and piecewise Gaussian model (GOM) Pcp curves from PERCLOS, and red points indicate their peaks or TCPs from offline algorithms. Offline Pcp curves from all individual α/β features and their EEG-CPs are demonstrated in the bottom plots.

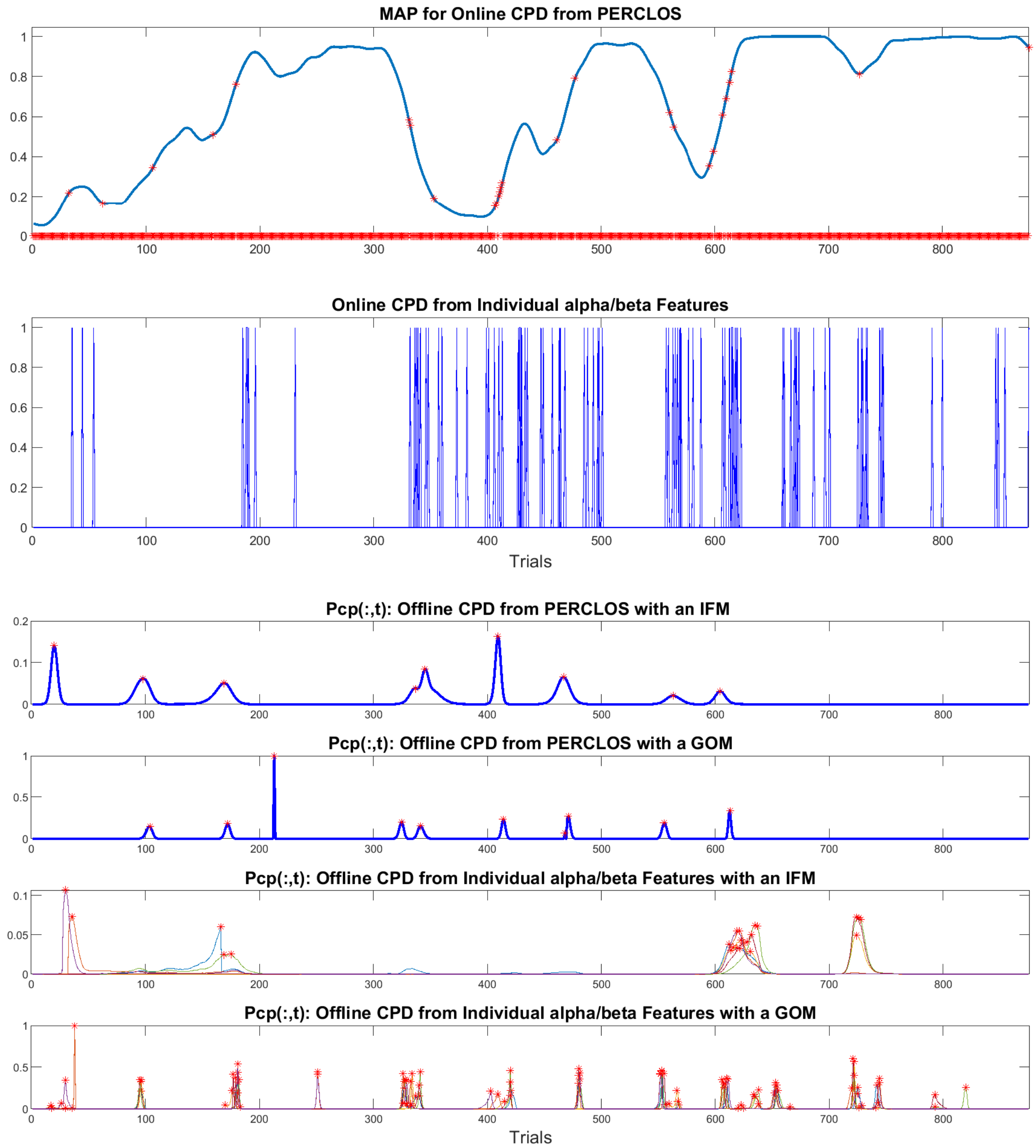


Figure 5.2: Changepoint detection from PERCLOS and EEG α/β time-series for participant S16. In the top plot, the blue curve represents the original PERCLOS, and red points denote locations of TCPs detected using the online algorithm. EEG-CPs from the same online algorithm are shown in the second plot. In the middle plots, blue curves indicate the independent features model (IFM) and piecewise Gaussian model (GOM) Pcp curves from PERCLOS, and red points indicate their peaks or TCPs from offline algorithms. Offline Pcp curves from all individual α/β features and their EEG-CPs are demonstrated in the bottom plots.

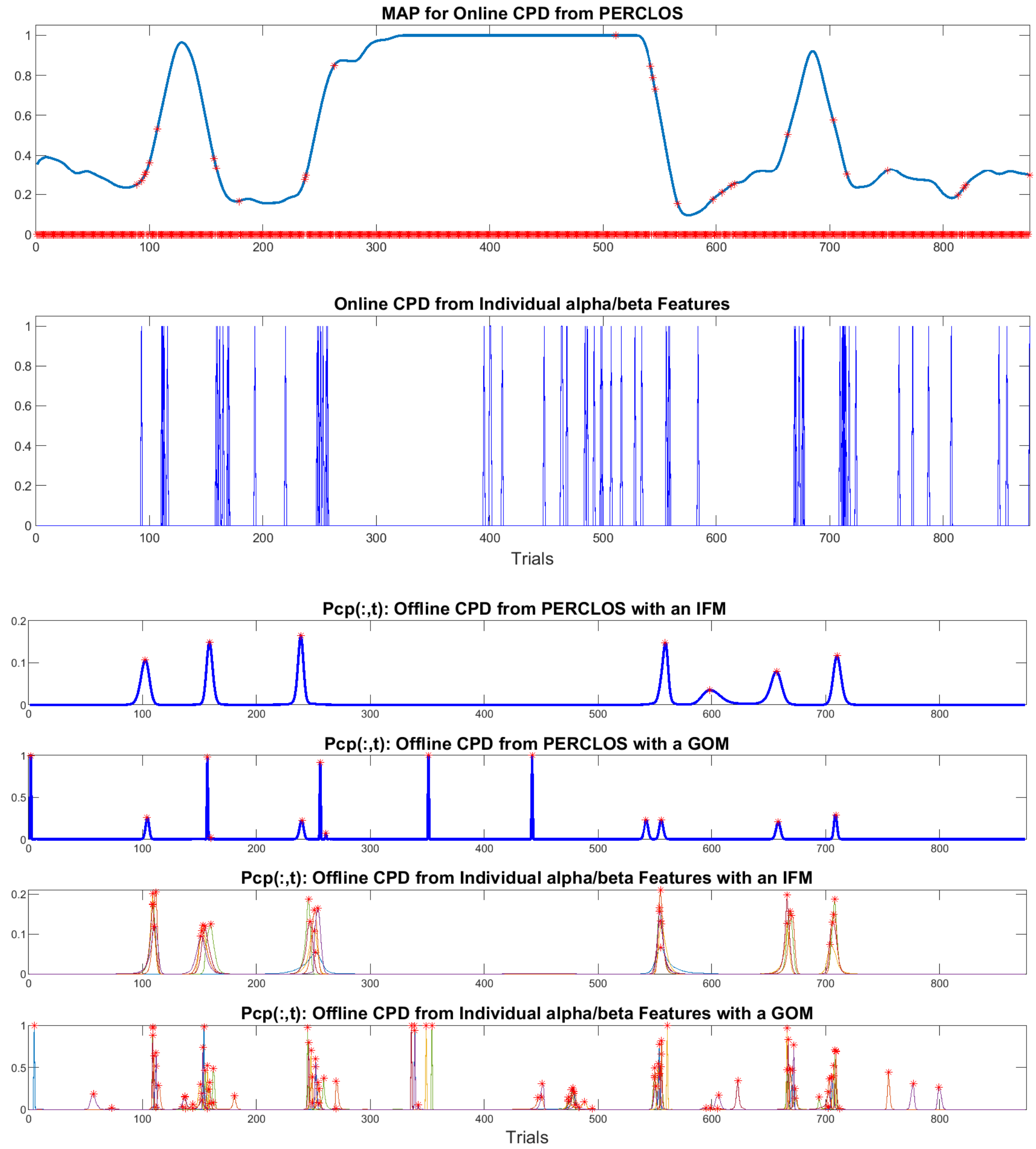


Figure 5.3: Changepoint detection from PERCLOS and EEG α/β time-series for participant S18. In the top plot, the blue curve represents the original PERCLOS, and red points denote locations of TCPs detected using the online algorithm. EEG-CPs from the same online algorithm are shown in the second plot. In the middle plots, blue curves indicate the independent features model (IFM) and piecewise Gaussian model (GOM) Pcp curves from PERCLOS, and red points indicate their peaks or TCPs from offline algorithms. Offline Pcp curves from all individual α/β features and their EEG-CPs are demonstrated in the bottom plots.

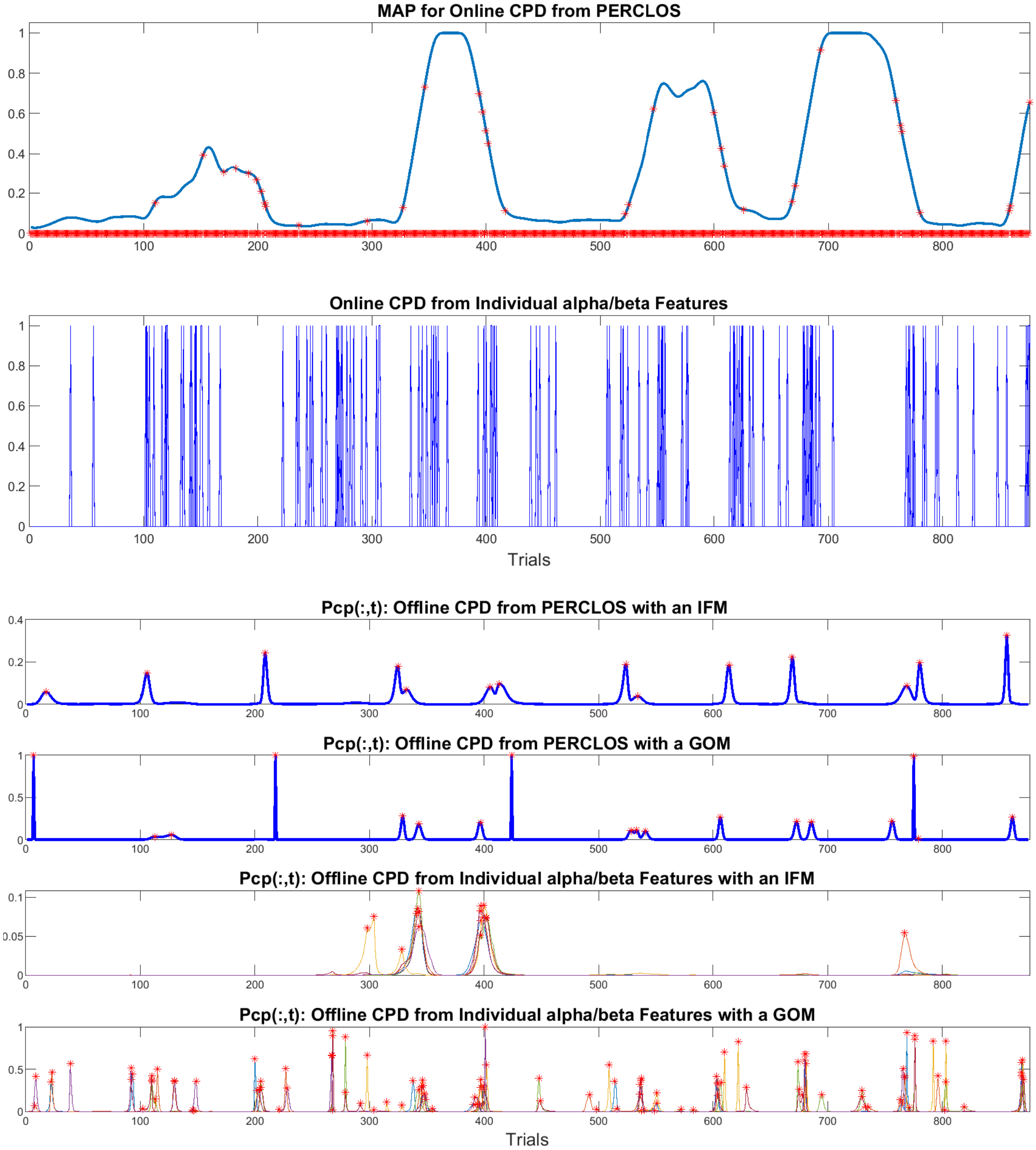


Figure 5.4: Changepoint detection from PERCLOS and EEG α/β time-series for participant S21. In the top plot, the blue curve represents the original PERCLOS, and red points denote locations of TCPs detected using the online algorithm. EEG-CPs from the same online algorithm are shown in the second plot. In the middle plots, blue curves indicate the independent features model (IFM) and piecewise Gaussian model (GOM) Pcp curves from PERCLOS, and red points indicate their peaks or TCPs from offline algorithms. Offline Pcp curves from all individual α/β features and their EEG-CPs are demonstrated in the bottom plots.

Table 5.1: Mean and standard deviation of precision and recall for online and offline Bayesian CPD algorithms using α/β features from 21 participants in the SEED-VIG dataset. Performance metrics of the online, IFM, and GOM algorithms are obtained based on their corresponding and individually detected changepoints τ_{online}, τ_{IFM}, or τ_{GOM}, respectively, and are not to be compared with each other.

	Online CPD	Offline IFM CPD	Offline GOM CPD
Precision	0.82±0.26	0.38±0.27	0.36±0.20
Recall	0.56±0.14	0.19±0.16	0.46±0.15

5.5 Conclusion

In this chapter, we demonstrated the applicability of online and offline algorithms for detecting changepoints from vigilance time-series and their predictor EEG feature sequences without having any constraint on the number and location of changepoints. The performance of each algorithm in detecting changepoints from EEG time series were evaluated based on the ground truths obtained from the drowsiness curves using the same algorithm. Our results shows that the online Bayesian CPD algorithm detects higher number of vigilance changepoints from PERCLOS curves compared to the offline individual feature model and piecewise Gaussian models. Using the online algorithm, EEG predictors also obtain higher precision and recalls in average. A visual inspection points to the probable oversensitivity of the online algorithm in detecting points which may not correspond to true variations in vigilance levels. This shortcoming can be improved by smoothing the curves for a longer time window length, w, which causes in a delay or ε-real time algorithm. This novel application of dynamic and sequential inference techniques has immense potentials for implementation in unsupervised vigilance monitoring systems for the final goal of alarming users about onsets of drowsiness. Similar models will be applied on PLV matrices in the future to evaluate the performance of these pairwise, spatio-spectral vigilance correlates in terms of changepoint localization. In the next chapter, we demonstrate the success of vigilance level clustering as another unsupervised method for assessment of vigilance time-series for the final goal of BCI adaptation.

6 Adaptive Alertness-Aware Classification for Motor Imagery-based Brain-Computer Interfaces

Brain-computer interfaces (BCIs) provide a direct path to control external devices through mental commands, and motor imagery (MI) has been one of the most commonly used paradigms in lab-based and commercial EEG-based BCI systems [256]. The interest in utilizing MI-based BCIs arises from brain-mapping studies that demonstrate the imagination of motor movements activates regions related to motor execution [62] and, if implemented within neurorehabilitation sessions, speeds up the limb and gait recovery and reduces the risk of falling down after brain strokes and knee arthroplasty [2]. Furthermore, improving the efficacy and usability of MI-based BCIs has been partly motivated by observing the feeling of autonomy in patients who are able to independently navigate while mentally controlling a wheelchair or a car without using peripheral nerves [42], or those who can control a robotic arm after having been in a locked-in state for a long time. Due to its non-invasive nature and high temporal resolution, EEG is the most favored signal acquisition method in BCI systems [257]. However, factors such as cognitive state variations, alertness levels, the ability to stay focused during the operation of a BCI-controlled device, and lack of familiarity with or fear in working with technology affect the extracted EEG-based features and the users' experience with BCI systems [31], [33]. This is also due to the fact that long experiments with monotonic and steady audiovisual stimuli increase the boredom in participants, create idle phases in the cortical networks, and reduce alertness [79].

As shown in previous chapters through the correlations between neural and behavioral correlates of long Go/NoGo execution during SART, any prolonged mental task that fails to maintain arousal of brain networks and block irrelevant stimuli causes inevitable decrements in sustained attention and delays response time (Ref. Section 4.4.2.1). We thus hypothesize that inferring the users' vigilance levels during the execution of mental commands can improve the understanding of their current level of task engagement and reaction time, and in turn increase the reliability of BCI systems. With this intention, this chapter focuses on the effects of alertness on the performance of MI BCI as a common mental control paradigm. In the first part of this chapter, Experiment 1, a new protocol is proposed to predict MI performance decline through alertness-related pre-trial spatio-spectral EEG features. This work has been presented at the 42nd International IEEE Conference of Engineering in Medicine and Biology Society (EMBC2020) [53].

In the second part of this chapter, Experiment 2, we propose an adaptive MI BCI classification approach based on the continuous assessment of alertness/vigilance information. The proposed approach considers adaptation in (1) extraction of the best time intervals for MI performance from each trial based on the inferred vigilance level, and (2) in using classifiers trained and subsequently tested separately for each vigilance level. To be more precise, without falling in the critical trap of pausing the experiment to collect subjective information about fatigue and drowsiness, vigilance information is estimated during the training session of MI recordings based on the clustering of related pre-trial EEG features discussed in Chapters 4 and 5. After vigilance clustering of training sessions trials are complete, the best time intervals for MI performance are selected separately for each vigilance level. These time intervals are used for MI feature extraction from the test session trials, and the previously trained classifiers are evaluated separately for each vigilance level. In short, we demonstrate that using the first experimental session as a training/calibration dataset for both MI and vigilance clustering enables the system to improve its prediction of vigilance-dependent performance of EEG-based BCI systems in the upcoming test/evaluation session. Results of this alertness-aware classification scheme are presented for our SPIS MI-BCI dataset and a two-class version of BCI Competition IV dataset 2a [71] in Experiment 2.

6.1 Contributions

In previous chapters, spatio-spectral features were used for building models that predicted attention levels during the execution of long vigilance tasks. The ground truths for inattentive periods and drowsy states in Chapters 3 and 4 were acquired from our objective cumulative vigilance score (CVS) that had summarized the occurrence of errors and delayed response time while, in Chapter 5, the PERCLOS curves reflected periods of increased eye closures. However, during the continuous execution of motor imagination when the user is completely focused on the mental task of imagining limb movements, no objective performance measure such as the CVS can be naturally obtained due to the lack of a recorded sequence of errors and response time. Therefore, a proposed scheme is to focus on EEG features already highly associated with existing ground truth in vigilance-labeled datasets, and utilize them to infer variations of attention levels in an unsupervised manner from recordings that lack objective levels on attention and drowsiness.

With this view point, we present two pieces of work that evaluate the effects of users' vigilance levels on the tonic performance of MI-based BCIs. Concerned with detecting signs of attention decline, the first work proposes and demonstrates the effectiveness of a variety of pre-trial spatio-spectral alertness features in predicting MI classification performance. Information for inference of vigilance labels is directly obtained from a cumulative classification score based on the outputs of the common spatial pattern (CSP) filtering and LDA classifier. In the second work, focusing on the band-power, pre-trial EEG correlates of attention variations, an unsupervised, clustering-based scheme is proposed to infer vigilance level of each trial in the training and test sessions before adapting

the classifier. This new information is used in the context of an alertness-aware adaptive classification method to obtain improvements in the inference of two-class motor imagery.
The novelty and contributions of these pieces of work are as follow:

1. Since the alertness level of a user has a critical effect on their reaction time and ability to correctly execute motor imagination, we hypothesize that alertness and vigilance levels subsequently affect the performance of an MI BCI system. Thus, we propose two contributions in Experiment 1 to directly evaluate these effects: 1. We evaluate vigilance based on an objective MI performance score, and 2. We demonstrate the effectiveness of several pre-trial spatio-spectral alertness features in predicting MI classification performance. The pre-trial band-power (BP) ratio features from Section 4.4.1.4 as well as their distances with respect to the beginning of each session and the pre-task resting-states are used as features or predictors of this binary problem for classification of high-vs-low attention levels. We report Cohen's kappa as the classification evaluation metric. When the proposed EEG-based features indicate the loss of attention, one can initiate the process of either adapting the classifier to the cognitive state of the users or taking steps to restore their attention. Hence, our work offers an objective methodology that can help prevent BCI performance decline due to attention variations.

   To the best of our knowledge, so far only one research group has objectively monitored and labeled the fatigue level of users in the course of MI BCI execution, and proposed an adaptive scheme for common spatial pattern (CSP) filtering based on the underlying fatigue levels that are obtained subjectively at the end of each run [45], [258]. However, this group has not provided any classification result and has only reported the "improved separability" of MI EEG features extracted by their proposed adaptive CSP scheme.

2. In Experiment 2, in order to validate an unsupervised learning scheme for continuous detection of vigilance levels, we propose and evaluate different static and dynamic clustering schemes on a dataset labeled with eye-closure events, SEED-VIG [35], and report significant correlations between the continuous vigilance labels and cluster indices. These schemes are used to provide information for vigilance estimation for our SPIS MI-BCI EEG recordings. We assume that the best EEG time interval to be used for classification of MI-based BCI, hereafter referred to as the winning time interval, should be the same across similar vigilance levels of training and test sessions for each participant. Therefore, after clustering the vigilance features during training session, MI CSP+LDA classifiers are evaluated in a cross-validation scheme on trials of each vigilance level in the training set. Once the best (winning) time interval for MI execution is obtained for each level, the same MI time interval is extracted from test trials predicted to have that cluster index, and their MI accuracy is obtained from the corresponding trained CSP+LDA classifier.

Three different versions of this adaptive classification are introduced in this work. Therefore, in this novel approach, we propose and evaluate an alertness-aware adaptive classification for motor imagery paradigm.

As will be clarified in Section 6.4.1.5, the term "adaptive" in this work corresponds to (1) extraction of best EEG time interval for MI execution according to the inferred vigilance level, and (2) applying the CSP+LDA parameters according to the inferred vigilance cluster. We report improvements in the overall test accuracy of adaptive versions with respect to the original, non-adaptive baseline for the dataset collected in our laboratory, referred to as the SPIS MI-BCI dataset, and the BCI Competition IV - Dataset 2a for comparison [71]. For both datasets, the dynamic and trial-based clustering schemes perform the best which points to the importance of phasic alertness in correct decoding of trial-wise sensorimotor rhythms.

3. In both experiments and especially in Experiment 2, special attention is paid to maintaining the temporal sequence of experiment trials during the clustering schemes to preserve the original pattern of drowsiness and regaining of alertness as experienced during SART sessions (see CVS curves in Section 4.4.2.1). Similarly, the evaluation of MI performance in cross-validation folds is performed without random permutation of trials. Both considerations are in line with the recommendations for block design experiments for neural data [38] and time-series treatment of data in Chapter 5. For these reasons, we did not perform two-class MI classification of left hand versus right hand on the 4-class datasets of BCI Competition mentioned above since extracting half of experimental trials would completely disrupt the temporal sequence of cluster indices.

The rest of this chapter is organized as follows. Section 6.2 presents a summary of the related work and state-of-the-art methods. Section 6.3 focuses on the first piece of work on prediction of discrete-valued MI performance scores. Section 6.4 then presents the second piece of work on evaluating the consistency of best MI time intervals in similar vigilance levels across different sessions as well as the alertness-aware adaptive BCI scheme. This chapter is concluded in Section 6.4.3 with a summary of important findings and implications.

6.2 Related Work

Adaptation methods mentioned in Section 1.1 all belong to the class of covert adaptation techniques since the system's classifier is updated in the background according to the varying statistical distributions of the incoming signals by changing the decision-making criteria. Overt adaptation techniques, on the other hand, attempt to update the experiment interface and experiment flow to decrease the participant's boredom and enhance the interaction outcome. In either case, any classification scheme that aims to incorporate the underlying attention and alertness level of the user should be continuously updated with the new level of attention.

As one of the earlier attempts on the role of alertness in BCI performance, it was shown that when training sessions use wakeful alertness levels, the P300 classifiers' accuracy significantly increases in experiments focusing on attention, concentration, and control [115]. A more recent study reported that increased cognitive load, induced by presenting visual distractors during the execution of MI BCI, could significantly predict reductions in BCI performance of users whose undisturbed accuracy was below 75% [34]. The notion of visual distractors is closely related to our previously conducted Go/NoGO experiment with SART and leads to our understanding of sustained attention, i.e., the ability to maintain the cognitive focus in a long period of time while avoiding mind wandering [143]. Gaume et al. designed a cognitive BCI for a continuous task to discriminate high and low sustained attention states. More recently, another group of researchers have analyzed the separability of two-class motor imagery features once information on the trial-based alertness levels has been taken into account [45], [258]. In their motor imagery paradigm, they performed an unsupervised CSP adaptation, and the fatigue state at the end of each run was rated using a subjective fatigue scale. The kernel partial least square (KPLS) from spatio-soectral features was used to calculated the fatigue score. The adaptation paradigm was applied in two modes: offline, where the CSP matrix was updated and applied for each high fatigue run, and near real-time or online, where the CSP matrix was updated on any high fatigue trial and subsequently used for the upcoming ones. However, to the best of our knowledge, these researches or others have not directly reported the effects of alertness levels and improvements in classification accuracy of MI-based BCI datasets as a result of incorporating vigilance or fatigue information in a fully adaptive classification paradigm.

6.3 Experiment 1: Prediction of Motor Imagery Performance from Pre-Trial Alertness Features

In this experiment, having collected EEG signals from a two-class MI session, we introduce a cumulative MI performance score (CMPS) that represents the tonic performance of MI classification rather than the momentary or trial-wise mistakes. By keeping the experimental flow as consistent as possible, the alertness level is hypothesized as the main factor affecting the MI performance. As will be explained, trials in the first half of each session are used for optimization of the best time interval for MI classification while the second half is used for the implementation and evaluation of the proposed alertness protocol using pre-trial band-power (BP) ratio and distance features.

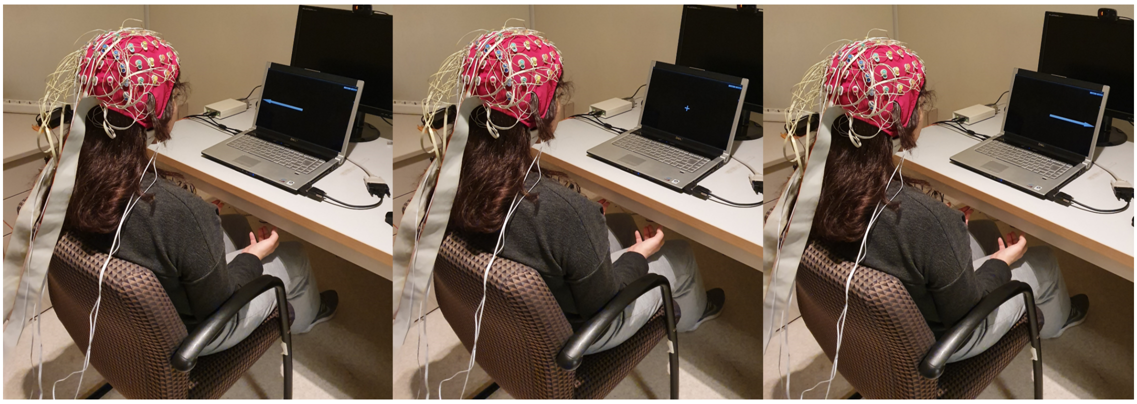


Figure 6.1: A user attending a session of the two-class motor imagery experiment that generated the SPIS MI-BCI dataset.

6.3.1 Methods

6.3.1.1 Participants

Ten healthy individuals (5 males and 5 females), aged 24.5 to 38 years (mean: 29.8), attended single data collection sessions starting at 10:30 AM when they were relatively alert. Participants had normal or corrected-to-normal vision, and all but one were right handed. Participants were not under any drowsiness-inducing medications within three days prior to the experiment date and reported no history of neurological disorders. All individuals were notified of experimental goals prior to attendance. The recruitment and experimental procedures were approved by the Sabanci University Research Ethics Council and all participants signed informed consents.

6.3.1.2 EEG Data Acquisition

Data collection was performed in an EEG room inside a Faraday cage. EEG activity was collected via 64 Ag/AgCl active electrodes placed according to the 10-10 International Electrode Placement System and connected to a BioSemi ActiveTwo set (Biosemi Inc., Amsterdam, the Netherlands). Signals were sampled at 2,048 Hz and bandpass filtered between 1 and 70 Hz using a Butterworth filter.

Once participants were seated in a comfortable armchair 60 cm away from a 17-inch LCD monitor, one resting-state session with eyes open (EO) lasting for 2.5 minutes was recorded to be used as the baseline. Next, they performed an MI practice session under the supervision of experimenters. The actual 200-trial MI session was then conducted for 20 minutes. The experimental paradigm consisted of a two-class cue-based MI to imagine movement of right and left hands as shown in Figure 6.1. Each trial composed of a 2-s fixation period prompted by a plus sign, hereinafter referred to as the pre-trial interval, followed by the appearance of an arrow pointing to the right or left for 4 s during which the motor imagination is to be performed. Facial videos of participants were recorded during EEG acquisition for verification of their alertness levels. The visual interface was written in Visual C♯. Figure 6.2 demonstrates the experimental flow.

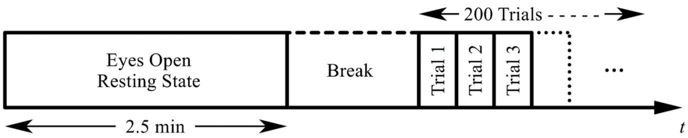


Figure 6.2: The experimental flow for the 200-trial cue-based two class motor imagery session.

Table 6.1: Spatio-spectral features for sustained attention analysis extracted from pre-trial intervals of MI trials.

Feature Names	Spectral Features	ROIs	Ratio Vectors	Distance Vectors
Full	δ, θ, α, $\beta1$, $\beta2$, γ, $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\alpha+\beta)$, θ/β	14 scalp-wise	v_{Full}	$d^{Init.}_{Full}$, d^{Rest}_{Full}
Fronto-parietal	θ, α, $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\alpha+\beta)$, θ/β	12 frontal & parietal	v_{FP}	$d^{Init.}_{FP}$, d^{Rest}_{FP}
Frontal	θ, α, $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\alpha+\beta)$, θ/β	6 prefrontal & frontal	v_F	$d^{Init.}_{F}$, d^{Rest}_{F}
Parietal	θ, α, $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\alpha+\beta)$, θ/β	6 central & parietal	v_P	$d^{Init.}_{P}$, d^{Rest}_{P}

6.3.1.3 Motor Imagery Signal Processing

From the 64 scalp electrodes, 14 channels placed over the sensorimotor cortex (C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, P1, Pz, PO, and POz) are used for MI signal processing and classification. EEG signals are downsampled to 256 Hz. A fifth-order Butterworth bandpass filter is applied between 8 and 30 Hz to contain μ and β bands as major MI related frequency range [259]. The well-known common spatial patterns (CSP) method is then applied for feature extraction. The first and last three rows of the projected EEG signals Z are used to calculate the log-variance matrix $S = \log(\mathrm{Var}(Z))$ for obtaining the MI class discriminating features. The LDA scheme is then applied to classify the MI features.

Based on monitoring the recorded facial videos, we established that participants are at a sufficiently high alertness level during the first 10 minutes of the MI experiment. For each participant, the best time interval to be used by the classifier, also known as the settle time, is determined by a cross-validation (CV) process on the first 100 trials, hereafter known as the training data. Four time intervals of 2-s duration are extracted starting from 0.5 s post-arrow onset with 0.5-s overlaps. Ten-fold CV is applied, and the macro-averaged F1-score, defined as the average of per-class F1-scores, is utilized as the selection criterion. For each participant, the MI time interval resulting in the highest score is chosen.

6.3.1.4 Sustained Attention Feature Extraction

In our previous work involving a long and monotonic sustained attention to response task (SART), we identified a set of EEG features that provide information about sustained attention levels [52], also ref. Section 4.4. In this experiment, we analyze the relationships between similar types of features and MI performance. To obtain potential predictors of variations in the MI performance based on cognitive processes during the stimulus expectation interval, several spatio-spectral features are extracted from pre-trial EEG signals for evaluation of attention levels. Two types of features are considered, the band-power (BP) ratio features and their distances with respect to (a) several initial trials of the same session in which participants are assumed to be at their most alert states, and (b) the EO resting-state in which the task-related networks are idle while the visual cortex is active [117]. These distance features, i.e., variations in sample covariance matrices of the obtained ratio features during an MI session, are proposed in this work to objectively reflect the user's current alertness with respect to their initial cognitive state.

1) BP ratio features. One-second pre-trial EEG signals are downsampled to 512 Hz. A 128-sample Hanning window with 64-sample overlaps computes the short-time Fourier transform (STFT) coefficients over 2-48 Hz with a 0.1 Hz resolution. The logarithmic BP is calculated for 2-4 Hz and 4-Hz non-overlapping bands for the remaining spectra. Due to different power levels in the resting-state and MI task sessions across participants, BP ratios are normalized with respect to the total power. Ten BP and combined BP ratio features are computed: δ (2-4 Hz), θ (4-8 Hz), α (8-12 Hz), $\beta 1$ (12-16 Hz), $\beta 2$ (16-20 Hz), γ (28-48 Hz), $(\theta+\alpha)/\beta$, $(\theta+\alpha)/(\alpha+\beta)$, α/β, and θ/β, where β is from 12 to 28 Hz. To focus on the specific cortical regions and avoid inter-subject differences in channel placements, all 64 electrodes are grouped into 14 regions of interest (ROI): left, midline, and right prefrontal, frontal, central, and parietal regions, and left and right temporal ROIs. BP ratios are averaged among electrodes of each ROI to form v_{Full}, a 140-d feature vector.

Sustained attention is generally associated with the prefrontal region, and attention decline can turn into drowsiness in more extreme cases. Furthermore, the connection between frontal and parietal regions is critical in intrinsic alertness and phasic alertness control [150], [260]. Therefore, as shown in Table 1, aforementioned indicators of attention and drowsiness are extracted from the prefrontal and frontal regions, $v_F \in R^{1\times 36}$, central and parietal cortices, $v_P \in R^{1\times 36}$, and their concatenation, $v_{FP} \in R^{1\times 72}$.

2) Distance features. BP ratio features represent the phasic patterns of individuals awaiting the new stimuli. Their corresponding *initial-distance* and *rest-distance* features are extracted to reflect tonic changes in sample feature covariance matrices from initial trials of each session and the EO resting-state signals, respectively.

Assume $X \in R^{N_t \times n}$ is a BP ratio feature set where N_t and n represent the number of trials and number of features (140, 72, or 36), respectively. The initial and EO covariance matrices cov^{Init} and cov^{Rest} are, respectively, calculated from the pairwise covariance elements of the first m samples of matrix X and the 150 one-second epochs of EO signals. Subsequently, $cov^k \in R^{n\times n}$ is the covariance matrix of samples $1+k$ to $m+k$ from matrix X.

Next, the pairwise city block distance between cov^k and cov^{Init} or cov^{Rest} is proposed for extracting distance features. The intuition behind this proposal is to measure changes in collective variance of spectral features in a specific region with respect to the state when users are assumed to be at higher alertness level. Although a variety of distance metrics could be utilized, we focused on the city block distance which represents the absolute value of changes between two instances. The pairwise city block distance between row s of matrix W and row t of matrix Y is obtained from the following formula

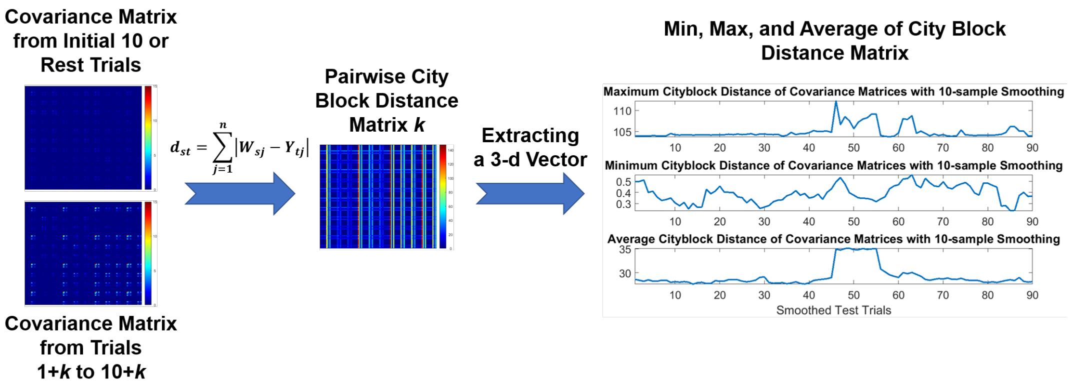


Figure 6.3: Initial and subsequent covariance matrices and 3 distance vectors for one participant.

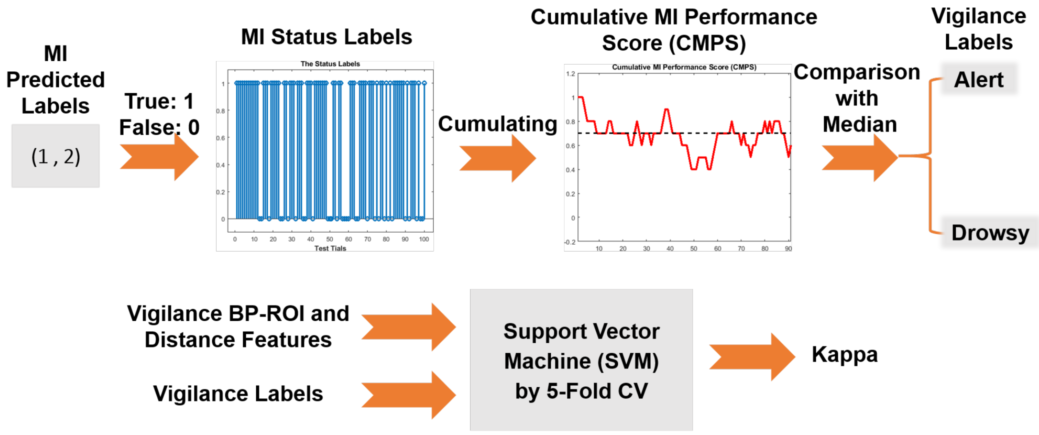


Figure 6.4: Pipeline for predicting the MI BCI performance using EEG-based sustained attention features.

$$d_{st} = \sum_{j=1}^{n} |W_{sj} - Y_{tj}|. \tag{6.1}$$

This results in an n-by-n distance matrix D composed of the obtained elements. To summarize this information, 3-dimensional vectors of maximum, minimum, and average values of the resulting matrices are considered as the distance feature vectors. Distance metrics are calculated for all the spatio-spectral feature groups as elucidated in Table 6.1. Figure 6.3 presents the initial and subsequent covariance matrices and corresponding distance vectors for the participant S06.

6.3.1.5 BCI Performance Prediction based on Vigilance Level

Figure 6.4 demonstrate the pipeline for predicting the MI BCI performance using EEG-based sustained attention features explained in the previous section. The classification results achieved by applying the trained MI system on the test trials using the features of the selected time interval is fed to the alertness assessment protocol. To assess the effects of users' underlying alertness levels on their MI BCI performance, a cumulative MI performance score (CMPS) is defined that would focus on the tonic performance of

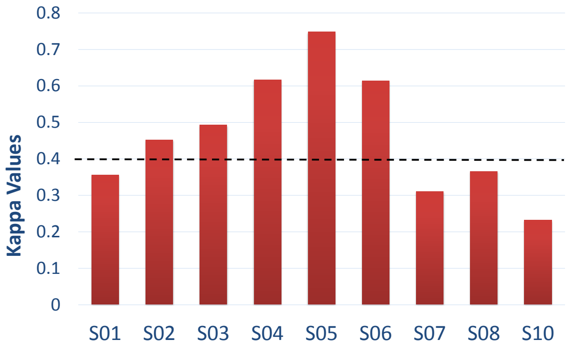


Figure 6.5: Alertness Kappa from the v_{Full} feature set.

Table 6.2: Kappa values for predicting MI-BCI performance using pre-trial spatio-spectral and distance features.

		S01	S02	S03	S04	S05	S06	S07	S08	S10
MI Evaluation	CV F1	0.70	0.63	0.59	0.79	0.61	0.74	0.64	0.70	0.60
	Test F1	0.65	0.53	0.55	0.71	0.56	0.62	0.43	0.51	0.58
Alertness Evaluation	v_{Full}	0.356	0.452	0.493	**0.513**	**0.748**	0.614	0.311	0.366	0.232
	$v_{Full}, d_{Full}^{Init.}$	0.351	0.452	0.493	**0.513**	0.694	0.764	0.283	0.376	0.195
	$v_{Full}, d_{Full}^{Rest.}$	0.321	**0.481**	0.493	**0.513**	0.445	0.673	0.303	0.376	0.232
	v_{FP}	0.189	0.248	0.462	0.210	0.675	0.650	0.459	0.292	0.024
	$v_{FP}, d_{FP}^{Init.}$	0.213	0.286	0.397	0.343	0.537	0.652	0.426	0.221	0.093
	$v_{FP}, d_{FP}^{Rest.}$	0.213	0.140	**0.525**	0.237	0.407	0.566	0.547	**0.509**	0.032
	v_F	0.104	0.354	-0.046	0.315	-0.068	**0.941**	**0.607**	0.208	0.263
	$v_F, d_F^{Init.}$	**0.382**	0.247	-0.205	0.286	0.354	**0.941**	0.372	0.208	0.287
	$v_F, d_F^{Rest.}$	0.326	0.414	0.052	0.343	0.429	0.793	**0.607**	0.418	**0.318**
	v_P	0.195	0.129	0.462	0.479	0.302	-0.051	0.275	0.198	0.232
	$v_P, d_P^{Init.}$	0.218	0.065	**0.525**	0.319	0.205	0.059	0.167	0.030	0.263
	$v_P, d_P^{Rest.}$	0.114	0.026	0.462	0.296	0.308	0.039	0.139	0.055	0.296

MI classification rather than the momentary or trial-wise mistakes: First, a binary-valued vector called *classification status*, s_{MI}, is defined for the test set classification output such that s_{MI} for each trial is equal to 1 if its MI label is correctly estimated by the MI classifier. Next, s_{MI} is smoothed using a moving window of length m to obtain the CMPS vector, s_{MI}^m. By keeping the experimental flow as consistent as possible, the alertness level is hypothesized as the main factor affecting the MI performance. Monitoring facial videos of the users can validate this assumption that changes in alertness levels correspond to variations in the MI performance. Therefore, a s_{MI}^m close to 1 is assumed to correspond to high tonic alertness while a value close to 0 indicates decreased attention or drowsiness. The s_{MI}^m is hence quantized based on its median value to form two *alertness* level labels. The 20 different feature sets composed of features listed in Table I and the concatenations of BP and their corresponding distance features are applied to predict the quantized s_{MI}^m labels. Support vector machine (SVM) with a linear kernel is used in a 5-fold CV scheme respecting the sequential order of trials. Due to the probable inequality in the distributions of two defined alertness classes in each fold, Cohen's kappa is reported as the classification evaluation metric.

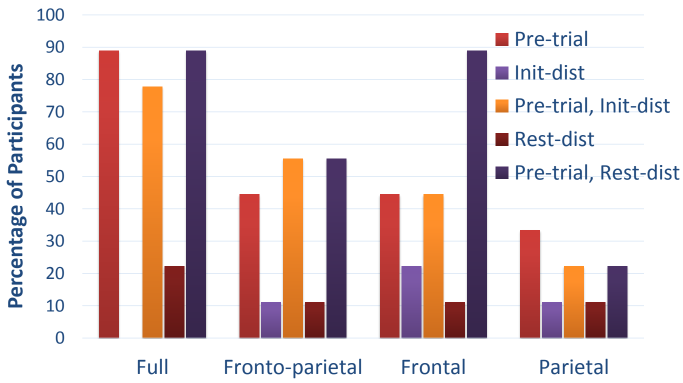


Figure 6.6: Percentage of participants achieving Cohen's kappa over 0.3 from different spatio-spectral feature sets.

6.3.2 Results

The obtained results are reported for all participants but S09 whose CV MI classification did not exceed the chance level – a possible indicator of insufficient MI performance. The alertness feature parameter window length, m, is set to 10 that represents the underlying cognitive activity of the recent ($10\times6=$) 60 s.

1) MI performance prediction based on alertness. To evaluate the hypothesis of predictability of MI performance based on pre-trial alertness features, the full feature set is first applied and its kappa results are shown in Figure 6.5. For 5 out of 9 participants, a kappa over 0.4 is achieved that corresponds to an accuracy of 0.70 in a balanced, binary classification task. S04 and S05 obtain noteworthy kappas of 0.617 and 0.748, respectively, and S01 and S08 have kappa scores close to 0.4. These results indicate the efficiency and applicability of our approach in predicting the alertness levels during MI execution and demonstrate that a good relationship exists between the proposed pre-trial alertness features and the MI performance. Alertness is absolutely not the only factor in BCI performance decline; however, the obtained results suggest that it can be successfully used as a predictor together with other effective factors.

2) Alertness feature assessment. The evaluation is then performed using all the proposed 20 feature sets and results are presented in Table 6.2. Results based on distance features without combining them with BP ratio features are eliminated due to their low achieved kappa in most cases. The feature set resulting in the high kappa for each participant is bolded. The CV and test MI F1-scores are also summarized in the table. Different features yield the best result for different participants, which suggests subject-based feature selection may yield more promising results. Overall, our results indicate the predictive power of frontal features in a subset of participants, which is in line with the localization of alerting, orienting, and executive attention networks over frontal anatomic structures [40]. Particularly, S06 achieved a highly promising kappa of 0.94 based on frontal features.

Additionally, to introduce the best predictors over all users, the percentage of participants for which a kappa value over 0.3 is achieved for each feature set is reported in Figure 6.6.

Utilizing a regional perspective, v_{Full}, v_{Full} concatenated with d_{Full}^{Rest}, and v_F concatenated with d_F^{Rest} outperform other feature sets for 90% of participants. In addition, although distance features alone do not perform as well as BP ratio features, adding d_F^{Rest} features to the v_F set increases the aforementioned percentage of participants. This positive effect is also observed in Table II where adding d_{FP}^{Rest} to the fronto-parietal features improves kappa scores in 6 out of 9 participants.
It is worth noting that S04 and S07 experienced the lowest and highest MI performance reduction, respectively, across the training and test sessions. This matches the visual inspection of recorded videos that demonstrate excessive sleepiness accompanied with continuous eye closures and head tilts for S07 as opposed to the stable and alert position of S04, and provides another means for validating the role of alertness decline in this study.

6.3.3 Conclusion

We have demonstrated that MI BCI performance can be predicted by alertness related pre-trial EEG features. Our work can be used to develop BCIs that react to inferred loss of alertness during their use either by performing a classifier adaptation to restore BCI performance or by initiating a process to restore the user's alertness. Comparing effects of different BCI decoders on generating labels for alertness levels will be considered in a future extension of this work.

6.4 Experiment 2: Adaptive Alertness-Aware Classification for Motor Imagery-based BCI

In the previous section, we focused on predicting the cumulative MI performance score (CMPS) based on vigilance-related pre-trial EEG features, with the underlying assumption that any decline in MI performance score is a result of reduction in attention levels and increased mind wandering. To improve MI classification accuracy due to attention decline, in this extension of Experiment 1, we present an unsupervised clustering-based technique to estimate trial vigilance levels and propose an adaptive and alertness-aware classification approach whose performance will be evaluated with respect to the regular, non-adaptive classification approach.

6.4.1 Methods

The datasets used in this experiment are composed of the SEED-VIG dataset [35], BCI Competition IV - Dataset 2a [71], and an extended version of our previously reported MI BCI dataset, now dubbed SPIS MI-BCI. In contrast with our previous work in which the first and last 100 trials of one session were used, respectively, for MI classifier optimization and alertness classification, in Experiment 2, participants have completed two separate, training and test, MI sessions. here, the goal was to demonstrate that using the first session as the training/calibration dataset for both MI and vigilance clustering enables the system to predict vigilance-dependent performance of that MI task in the upcoming test/evaluation session.

6.4.1.1 SPIS MI-BCI Dataset

Eight healthy individuals (5 males and 3 females), aged 24.6 to 37.7 years (mean: 28.59, SD: 4.68), attended two subsequent data collection sessions starting at 10:30 AM when they were relatively alert. Participants had normal or corrected-to-normal vision, and all but one were right handed. Participants were not under any drowsiness-inducing medications within three days prior to the experiment date and reported no history of neurological disorders. All individuals were notified of experimental goals prior to attendance. The recruitment and experimental procedures were approved by the Sabanci University Research Ethics Council and all participants signed informed consents. The dataset composed of these participants is hereinafter referred to as the SPIS MI-BCI dataset.

Experimental flow and EEG data acquisition were similar to those explained in Section 6.3.1.2. The only exception was that the experiment consisted of two sessions in one setting: training (calibration) and test (evaluation). For participants 1 to 3, the training session consisted of 200 MI trials and lasted for 20 minutes while for participants 4 to 8, it had 300 MI trials and was completed in 30 minutes. All participants completed a 300-trial test session after a 10-minute break in which they talked with the experimenters about their ability to perform motor imagination, answered questions regarding their sleepiness levels, and rested. As before, facial videos were recorded during EEG acquisition for verification of their alertness levels. Channel P10 was used as the reference due to its proximity to the right ear. MI features were extracted from 2-second time intervals that started 0.5 seconds after the cue onset.

6.4.1.2 BCI Competition IV, Dataset 2a

Searching for long MI BCI recordings that would demonstrate clear correlates of declined alertness and sustained attention was a challenging task. Our group came across the 11-participant recordings of Talukdar et al. [45], [258] composed of a maximum of 8 runs of 4-class motor imagery for a total of 96 minutes, and was the best fit to validate our proposed algorithm due to its self-annotated vigilance levels. However, the dataset was not available for download at the time of writing of this thesis. Instead, we focused on the publicly available BCI Competition IV, dataset 2a [71]. This dataset is collected with a cue-based paradigm recorded from 9 healthy participants, each having completed one 288-trial session of four-class motor imagery on two different days. The flow of its 8-second trials are shown in Figure 6.7. Each trial starts by demonstrating a cross on the screen for 2 seconds, followed by a cue in the shape of one of the four classes lasting for 1.25 seconds. The four classes of motor imagery consisted of instructions to imagine movement of the left hand, right hand, feet, or tongue that had to be executed for 3 seconds. There was a 2-second break at the end of each trial in which participants should

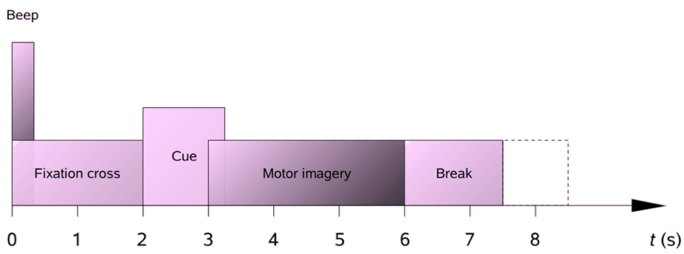


Figure 6.7: The timing flow of 8-second trials in BCI Competition IV - Dataset 2a [71].

refrain from imagining any movement. Each session lasts for 38.4 minutes, and recordings consisted of 22-channel EEG and 3-channel EOG signals sampled at 250 Hz using the left mastoid as the reference and the right mastoid as ground. The initial preprocessing steps including a bandpass filtering from 0.5 Hz to 100 Hz and notch filtering at 50 Hz were already applied on the signals.

6.4.1.3 Clustering of Vigilance Levels

Consider the data set $X \in R^{N_t \times d}$ composed of pre-trial, band-power EEG features where N_t represents the total number of trials and d is the number of features. While maintaining the order of trials, the dataset is split into the training or calibration set X_{Train} and the test or evaluation set X_{Test}. The following steps are performed on the training set to cluster its features:

1. K-means clustering algorithm is performed on all N_{Train} samples of X_{Train} using the square Euclidean distance metric for a pre-specified number of clusters, k.

2. The optimum number of training set clusters, k^*_{Train}, is selected using the Silhouette clustering evaluation criterion. The Silhouette value measures the degree of similarity of each point in a cluster to all other points in the same cluster versus others, and is calculated from the following formula for each point j:

$$S^j = \frac{b^j - a^j}{max(a^j, b^j)}, \tag{6.2}$$

   where a^j is the average distance between point j and all other points in its cluster, and b^j represents the minimum average distance between point j and other points in different clusters. A S^j closer to 1 represents higher similarity of point j within its own cluster. The Silhouette clustering evaluation criterion in this algorithm reports the Silhouette value averaged across all points and clusters.

3. Upon detection of k^*_{Train}, centroids of the optimal scheme are sorted so that centroids with smallest coordinates are assigned to cluster 1. This step maintains a consistent order for features' levels, corresponding to a consistent vigilance order.

Before predicting vigilance cluster indices of test session trials, we notice that, due to the nonstationarity of brain dynamics in general and EEG signals in particular, dynamic models are of superior performance in capturing time-varying features especially for testing and prediction purposes [261]–[263]. Therefore, we follow a gradual framework for predicting vigilance clusters of test trials, starting from a statistic prediction system as a baseline in the C_1 scheme, moving towards a real-time and dynamic prediction scheme in C_2 and C_2', and smoothing the predictions in a real-time manner to decrease the effect of momentary outliers in C_3. Thus, the following clustering schemes are conducted on the test set:

- Scheme C_1 - the baseline: Static prediction of test clusters. All N_{Test} samples in X_{Test} are clustered based on their distances from k^*_{Train} centroids to obtain the test cluster indices, $CI_{Test}^{1:N_{Test}}$. Here it is assumed that distances between centroids of training vigilance clusters will stay the same during the subsequent test sessions, meaning that the relative distances of vigilance features are constant across these sessions.

- Scheme C_2: Dynamic updating of test clusters after arrival of new test samples. Upon arrival of each new m samples $j-(m+1)$ to j from X_{Test}, all samples in X_{Train} and $\bigcup_{i=1}^{j} x_{Test}^{i}$ are concatenated for computation of new centroids. In other words, the optimum number of clusters and their corresponding centroid coordinates are reevaluated upon arrival of new samples which are then assigned to one cluster.

- Scheme C_2': Predict the test clusters first before dynamically updating their centroids. Upon arrival of each new m samples $j-(m+1)$ to j from X_{Test}, their cluster index is first predicted using the already existing centroids. Next, X_{Train} and $\bigcup_{i=1}^{j} x_{Test}^{i}$ are concatenated for computation of new centroids. This method is more applicable for implementation of real-time experiments.

- Scheme C_3: Smoothing the training and test cluster indices. High-frequency ripples in train or test clustering indices and variations in clustering decisions due to existence of movements and outliers might not be ideal or representative of the brain's alertness levels. Thus, once k^*_{Train} is determined, cluster indices are smoothed using a 3-point moving average and quantized to have a maximum of three levels. Next, for every new arrival of sample j from X_{Test}, its true cluster index or CI_{Test}^{j} is predicted, and a moving average window of length 3 is applied on the last three predicted labels $CI_{Test}^{j-2:j}$ to obtain CI_{smooth}. Since the smoothed curve contains decimal numbers, it is quantized to have a maximum of three levels. In another scheme, a sliding average window of length 3 is applied on the entire predicted test clusters up to sample j, $CI_{Test}^{1:j}$, to obtain $CI_{movmean}$ before quantizing it.

It should be mentioned that the schemes C_2, C_2', and C_3 are all based on causal algorithms. C_2 has the drawback of having to wait for new samples before recalculating the centers of vigilance clusters while C_2' speeds up the real-time implementation of this dynamic clustering scheme. Furthermore, obtaining cluster indices for C_1, C_2, and C_2' rely on prestimulus amplitudes of EEG features which demonstrate a short-time and phasic perspective in capturing the alertness levels. Scheme C_3, however, considers the assigned indices of preceding trials as well for smoothing out its predictions, and thus has a more tonic perspective similar to our treatment of cumulative vigilance scores. In the subsequent experiments of section 6.4.2, we explore which alertness inference approach improves the prediction of trial-wise motor imagery labels.

6.4.1.4 Detection of Consistent MI Features during Similar Vigilance Levels

In this experiment, the assumption is that the best time interval resulting in the highest classification has to be consistent for similar vigilance levels of any executed BCI session. Assume an MI BCI paradigm is executed for separate training and test sessions, S_1 and S_2, and that clustering the training and test α/β features demonstrate that each session consists of two different vigilance levels VL_1 and VL_2. Then we hypothesize that the vigilance level is a key factor affecting the reaction time and concentration of participants which subsequently affects the MI classification accuracy in each session. In addition, the best time interval that wins the classification accuracy for similar vigilance levels should be identical for two sessions.

1. Unsupervised labeling of vigilance levels: Pre-trial BP-ROI features are extracted from one second before the cue onset, and α/β features from all regions available in the dataset are used to compute cluster indices (CIs) in sessions S_1 and S_2. Having specified a maximum of k clusters, trials with vigilance levels VL_1 to VL_k are identified in both sessions. Note that there are 14 ROIs available in our SPIS MI-BCI dataset due to having 64 electrodes covering the whole scalp while the BCI Competition provides only 7 ROIs due to having electrodes only over the sensorimotor cortex. This is a challenge for studies involving the latter dataset since no feature from the pre-frontal or frontal regions can be obtained.

2. Classification of MI trials for discrete vigilance levels and sessions: MI EEG features are extracted from four time intervals of 2-s duration starting from 0.5 s post-cue (left or right arrow) onset with 0.5-s overlaps, i.e., [0.5 , 2.5], [1 , 3], [1.5 , 3.5], and [2, 4] seconds. For each vigilance level VL_j, $k = 1$ to k, motor imagery-labeled trials are classified: A multi-fold cross-validation (CV) is applied on trials belonging to VL_j of each session without any permutation to maintain the original temporal sequence of trials while assuring that almost equal number of samples from each MI class is included in each fold. If, for a specific vigilance level, less than two samples from a class are included in a fold, MI is not classified for that vigilance level. Finally, the average CV accuracy is reported and the time interval $TI_{j,i}$ resulting in the highest MI accuracy in VL_j of Session i, S_i, reported as the winning time interval.

3. Our adaptation approach is based on the assumption that each participant has similar time intervals being successful in classifying their motor imagery tasks for similar

vigilance levels across different sessions. For example, for a 2-level vigilance clustering scheme performed in two sessions, we hypothesize that the $TI_{1,1}$ and $TI_{1,2}$ should be the same for a participant as should be the $TI_{2,1}$ and $TI_{2,2}$. For each dataset, the vigilance clustering scheme that results in the highest number of participants having consistent MI performance is reported as the winning vigilance detection scheme.

6.4.1.5 Proposed Adaptive Alertness-Aware Classification Approach

The proposed adaptive BCI approach considers adaptation in (1) extraction of the best time intervals for MI execution in each trial according to its inferred vigilance level, and (2) in setting its classifier parameters according to the same inference. As the highlight of this adaptive alertness-aware classification method, we present the results of the original, non-adaptive MI classification that does not utilize any vigilance level information, and the proposed adaptive methods that incorporate information on the vigilance level of each clustered trial. As an important contribution to the literature, the proposed method includes both an adaptive feature extraction time interval and an adaptive setting for the CSP + LDA classification.

1. Unsupervised labeling of vigilance levels: Pre-trial BP-ROI features are extracted, and α/β features from all regions available in the dataset are used to compute cluster indices (CIs) in sessions S_1 and S_2. Having specified a maximum of k clusters, trials with vigilance levels VL_1 to VL_k are identified in both sessions.

2. Classification of MI trials for discrete vigilance levels and sessions: Four time intervals of 2-s duration starting from 0.5 s post-cue onset with 0.5-s overlaps are taken into account. In the "Original" or non-adaptive paradigm, the best interval is selected using K-fold cross-validation (CV) over all the trials of the train data. The classic CSP and LDA are applied for MI feature extraction and classification respectively. By selecting the best interval, the BCI system (CSP+LDA) is trained by train data and evaluated over test data based on the chosen best interval.

   (a) Adaptive Classification, Version 1: Here, the best interval is selected separately for each vigilance level by using K-fold cross validation over each level of train data. If two clusters are detected in the training trials, two systems of CSP_1+LDA_1 and CSP_2+LDA_2 are built by K-fold CV on their corresponding training trials and evaluated on test data separately for vigilance levels VL_1 and VL_2 by considering the corresponding chosen time interval TI_1 and TI_2. Finally, the overall accuracy achieved from all the test trials is reported as the result of "Adaptive" vigilance based BCI system.

   Data recorded from participants at their alert state is assumed to have been a better representative of their MI execution abilities. Additionally, the small number of drowsy trials for most participants in our existing datasets reduces

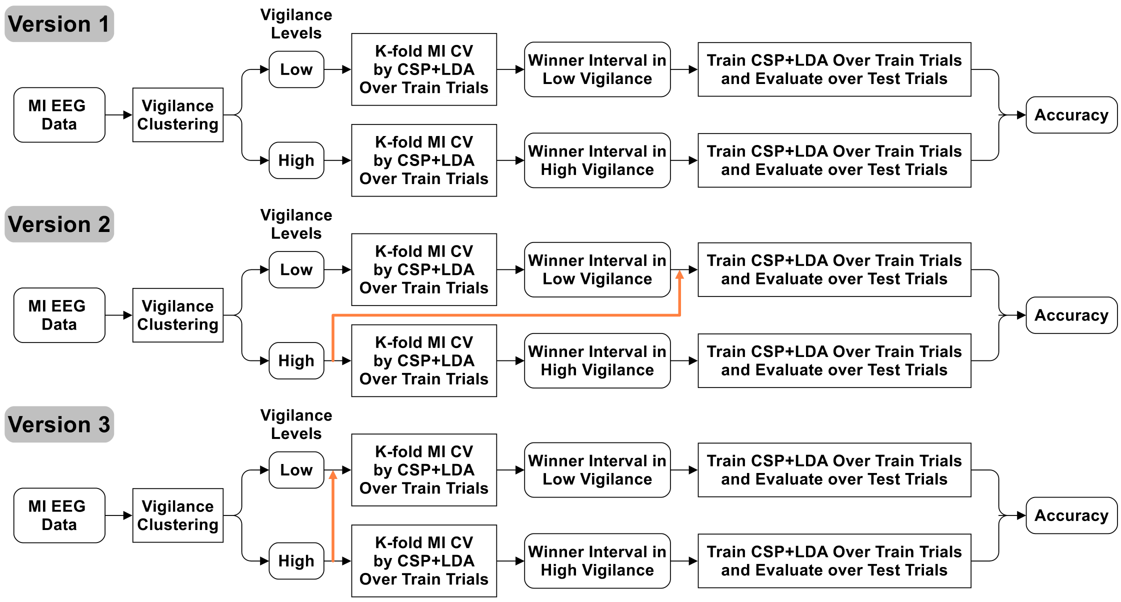


Figure 6.8: Three versions of the proposed adaptation approach for alertness-aware MI-BCIs. High and low vigilance clusters correspond to the alert and drowsy or VL_1 and VL_2 clusters, respectively.

the amount of valuable information for detecting the best MI interval and training the BCI system in their drowsy state due to the possible overfitting or not building a reliable model. Therefore, two other versions of the proposed adaptive BCI are also conducted, as follows:

(b) Adaptive Classification, Version 2: The selection of the best time interval for VL_2 is performed based on its corresponding trials from the training data as before. The training of the CSP_2+LDA_2 system is, however, conducted based on all trials of the training data by considering the chosen interval TI_2.

(c) Adaptive Classification, Version 3: In this version, both selecting the best interval TI_2 and training the CSP_2+LDA_2 are performed based on all trials of training data for evaluation on VL_2 test trials.

Figure 6.8 demonstrates the three versions of this proposed adaptation approach for alertness-aware MI-BCIs. Light red arrows indicate using trials of the alert cluster, VL_1, in training the drowsy classifiers in version 2, or using those trials for obtaining both the best time interval TI_2 and training the drowsy classifier in version 3.

6.4.2 Results

6.4.2.1 Alertness-Aware Clustering of SEED-VIG Dataset

The SEED-VIG dataset [35], developed by the Center for Brain-like Computing and Machine Intelligence (BCMI) and introduced in Section 5.3.2, is used for evaluation of alertness-aware clustering techniques. The dataset was recorded from 21 participants who performed a virtual driving task for 118 minutes. Vigilance levels were labeled with the PERCLOS indicator every 8 seconds using the SMI eye-tracking glasses, creating 885 labels for each session. Each recording includes 17 EEG channels located on the centro-parietal, parietal, occipital, and temporal cortices and 4 forehead EEG and EOG channels, all sampled at 200 Hz. Vector $y \in R^{885\times 1}$ contains continuous vigilance values, ranging from 0 to 1, obtained from 8-second non-overlapping windows.

In calculating pre-trial BP-ROI features from our SART and MI-BCI datasets, we were aware of the exact onset of visual stimuli presentation to the users. In the SEED-VIG setup, however, participants were constantly following the visual stream of cars and roads on the screen and no cue was presented to help with extracting the "pre-trial" intervals. Thus, noticing that PERCLOS labels are computed from 8-second long windows, BP-ROI features averaged across similar non-overlapping 8-second windows. It should be noted that due to the presence of EEG signals only across the temporal and parietal regions, only 6 non-empty ROIs are formed which results in a 60-dimension feature set. In other words, the SEED-VIG dataset of BP-ROI features is represented as $X \in R^{885\times 60}$. We form various subsets of X formed by band-specific features: X_θ, X_α, $X_{\theta+\alpha)/\beta}$, $X_{\theta+\alpha)/(\alpha+\beta)}$, $X_{\alpha/\beta}$, and $X_{\theta/\beta}$, as features potentially be related to the drowsiness levels.

As demonstrated in heatmaps of Figure 6.9, over 90% of participants consistently demonstrated significantly positive correlations between their α/β features from all 6 ROIs and PERCLOS labels. For θ features from midline central, left and midline parietal, and right temporal regions, and α features from midline central and left/right/midline parietal channels, this accounted for 80% of participants. **We thus select α/β features as the common neural markers of increased eye closure, in turn a behavioral marker of increased sleepiness and decreased vigilance.**

These results on BP-ROI correlates of increased drowsiness from SEED-VIG are in agreement with the heatmaps of Figure 4.5 in which negative correlations of BP-ROI features with CVS curves was an indicator of increased drowsiness. Positive correlation of all centro-parieto-temporal α/β features with increased PERCLOS was observed in more than 90% of SEED participants, while such strong associations with decreased CVS were observed only from the right and midline central and left parietal regions for more than 80% of SART participants. Positive correlations of left parietal α with drowsiness is also obtained for over 90% of individuals from both datasets. Patterns of θ associations are also common in central features. Finally, in SART heatmaps, an increase in γ ratios from the right central and left parietal regions was an indicator of improved performance. This strong association has shifted towards the left and right temporal regions as seen in SEED-VIG heatmaps. The smaller number of channels used inside each ROI in the latter dataset could be a reason for these small anatomical shifts.

Considering $N_t = 885$ as the total number of samples obtained from 8-second windows, the 6-dimension training set of $X_{\alpha/\beta}$ is extracted to include the first $\lfloor Nt/4 \rfloor$ samples of α/β features, accounting for 222 samples or the first 29.5 minutes of the driving session. Subsequently, the test set of $X_{\alpha/\beta}$ contains the remaining 663 samples, corresponding to

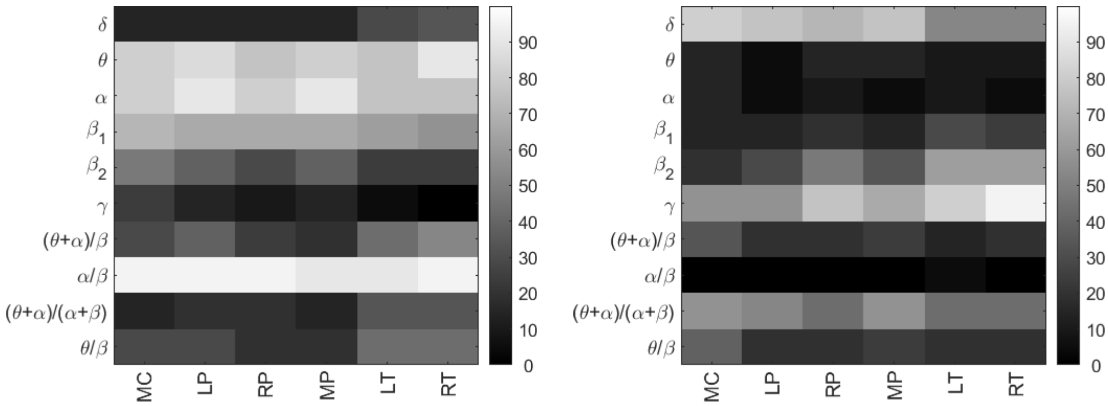


Figure 6.9: Percentage of SEED-VIG participants with statistically significant positive (left) and negative (right) correlations between various BP-ROI features and PERCLOS labels, $p < 0.1$. α/β features from 6 ROIs were positively correlated with increase in sleepiness in 90% of participants, followed by α and θ features. Right temporal γ is correlated with decrease in sleepiness and eye closure in 95% of participants.

the last 88.9 minutes of the driving session. The three clustering schemes explained in Section 6.4.1.3 are applied on the training and test sets of each participant. The dynamic schemes are especially evaluated since samples in the test set could be at extremely higher or lower levels of the utilized α/β features, and, subsequently, at a previously unobserved vigilance level.

Starting from $k = 3$ number of clusters for the training set, Table 6.3 presents k^*_{Train} in schemes C_1 and C_2 with $m = 1$, k^*_{Test} for scheme C_2, and Pearson correlation coefficients between the labels y (PERCLOS) and cluster indices (CI) for the training set, test set, and their concatenations. Table 6.4 then demonstrates k^*_{Train} and correlation coefficients between the PERCLOS labels and cluster indices for the C_3 scheme. The standard deviation (SD) of eye-closure labels is also reported since highly constant labels or predicted clusters would result in undefined Pearson coefficients. Bold numbers demonstrate the higher correlation between the two quantized versions of CI_{smooth} versus $CI_{movmean}$ for each participant. These higher correlation coefficients are above 0.5 for 11 out of 21 participants. The paired sample Student's t-test demonstrates no statistically significant difference exists between the means of the two groups, $p > 0.2$. Therefore, both quantized versions were computed for subsequent experiments on BCI Competition IV Dataset 2b and our own SPIS MI-BCI dataset.

Considering all three schemes, in 11 out of 21 participants, the quantized CI_{smooth} based on α/β features obtains the highest correlation coefficients, and in two cases, it generates a fixed cluster index for all samples in the test set, hence a NaN result for correlation. A one-way analysis of variance (ANOVA) demonstrates a non-significant effect of clustering scheme on the obtained correlations of overall participants, $F(3,78) = 0.66$, $p >0.5$. To visually present the effectiveness of unsupervised labeling of trial vigilance using the quantized smooth scheme, α/β features obtained from C_3 and their corresponding PERCLOS labels from the entire driving sessions of four participants are shown in Figure 6.10. **Since increase in the α/β level is positively correlated with more duration**

Table 6.3: Optimum number of clusters and Pearson's linear correlation coefficients between PERCLOS labels and Cluster Indices (CI) in schemes C_1 and C_2 with $m = 1$ when starting with $k = 3$ clusters.

	Scheme C_1				Scheme C_2				
	k^*_{Train}	ρ_{Train}	ρ_{Test}	$\rho_{Train \bigcup Test}$	k^*_{Train}	k^*_{Test}	ρ_{Train}	ρ_{Test}	$\rho_{Train \bigcup Test}$
S01	2	0.833	0.850	0.846	2	2	0.833	0.867	0.858
S02	2	0.070	0.353	0.332	2	2, 3	0.070	0.358	0.343
S03	3	0.243	0.319	0.353	3	2, 3	0.242	0.334	0.299
S04	2	0.169	0.225	0.446	2	2	0.169	0.343	0.354
S05	2	0.648	0.690	0.691	2	2	0.648	0.697	0.699
S06	2	0.016	-0.109	-0.078	2	2	0.016	-0.079	-0.047
S07	3	-0.136	0.316	0.061	3	2, 3	-0.136	0.182	-0.048
S08	2	0.010	0.670	0.644	3	2, 3	0.118	0.497	0.333
S09	2	0.015	0.089	0.016	2	2	0.015	0.071	0.011
S10	2	0.313	0.093	0.099	2	2	0.313	0.084	0.091
S11	3	0.191	0.432	0.355	3	2, 3	0.262	0.372	0.273
S12	3	0.506	0.635	0.606	3	2, 3	0.482	0.575	0.543
S13	2	0.662	0.628	0.654	2	2	0.662	0.598	0.629
S14	2	0.320	0.478	0.462	2	2, 3	0.345	0.470	0.438
S15	2	-0.023	-0.038	0.013	2	2	-0.023	-0.026	0.025
S16	2	0.610	0.763	0.762	2	2, 3	0.610	0.520	0.565
S17	2	0.793	0.687	0.757	2	2, 3	0.793	0.685	0.738
S18	3	0.721	0.813	0.807	3	2, 3	0.800	0.847	0.853
S19	2	0.424	0.007	0.229	2	2	0.502	0.100	-0.012
S20	2	0.359	0.064	0.157	2	2	0.359	0.045	0.138
S21	2	0.250	0.689	0.630	2	2, 3	0.250	0.580	0.529

Table 6.4: Optimum number of clusters, SD of PERCLOS labels, and Pearson's linear correlation coefficients between PERCLOS labels and Cluster Indices (CI) in scheme C_3 when starting with $k = 3$ clusters.

	k^*_{Train}	SD of y_{Train}	SD of y_{Test}	ρ_{Train}	ρ_{Test}	$\rho_{quantizedsmooth}$	$\rho_{quantizedmovmean}$	$\rho_{Train \bigcup Test}$
S01	2	0.274	0.292	0.833	0.850	**0.929**	0.893	0.846
S02	2	0.041	0.184	0.092	0.360	**0.484**	0.418	0.339
S03	3	0.053	0.164	0.306	0.307	**0.581**	0.372	0.364
S04	2	0.036	0.172	0.169	0.225	0.019	**0.220**	0.446
S05	2	0.199	0.339	0.648	0.690	**0.819**	0.726	0.691
S06	2	0.071	0.080	0.016	-0.109	NaN	-0.101	-0.078
S07	3	0.054	0.107	-0.136	0.316	0.205	**0.342**	0.061
S08	3	0.016	0.254	-0.010	0.612	**0.752**	0.666	0.599
S09	2	0.089	0.130	0.015	0.089	NaN	0.082	0.016
S10	2	0.099	0.144	0.307	0.097	**0.162**	0.061	0.105
S11	3	0.089	0.115	0.191	0.432	**0.587**	0.485	0.355
S12	3	0.318	0.391	0.510	0.645	**0.789**	0.685	0.615
S13	2	0.186	0.198	0.662	0.628	**0.818**	0.696	0.654
S14	2	0.103	0.243	0.345	0.484	0.515	**0.553**	0.472
S15	2	0.055	0.167	-0.023	-0.038	**0.004**	-0.048	0.013
S16	2	0.269	0.291	0.610	0.763	0.826	**0.844**	0.762
S17	2	0.275	0.124	0.793	0.687	0.666	**0.740**	0.757
S18	2	0.236	0.349	0.785	0.810	0.851	**0.860**	0.813
S19	2	0.065	0.091	0.424	0.007	0.017	**0.020**	0.229
S20	2	0.378	0.326	0.359	0.064	0.018	**0.059**	0.157
S21	2	0.120	0.358	0.250	0.689	**0.833**	0.763	0.630

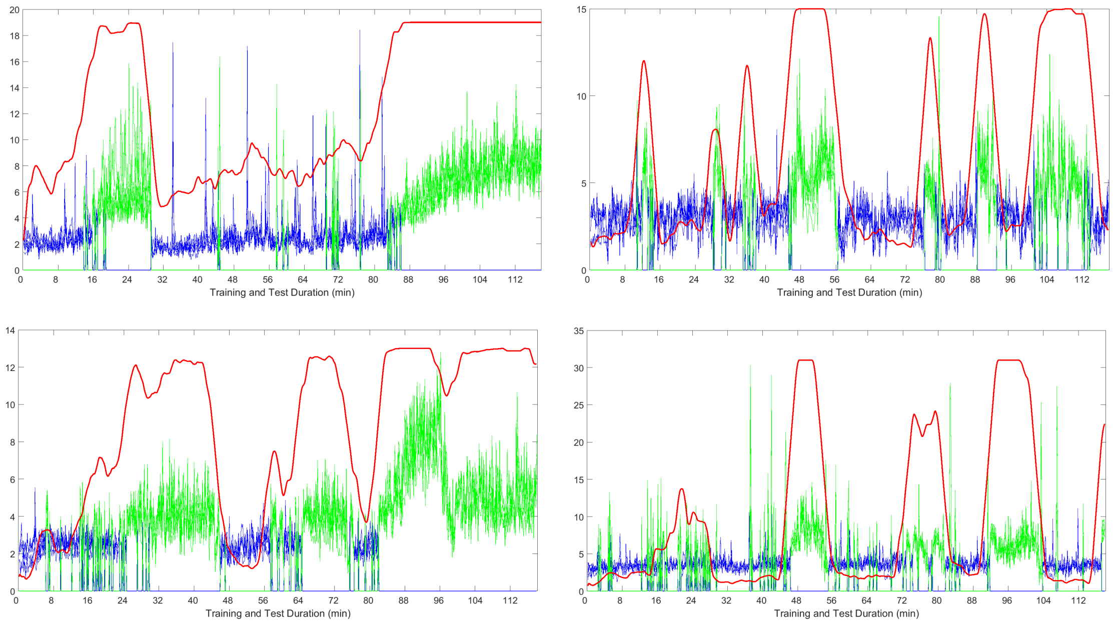


Figure 6.10: Clustering results of scheme C_3 using the quantized smooth method for four SEED participants, S01, S05, S16, and S21. Blue and green curves indicate the α/β features of clusters 1 and 2, respectively, while the red curves correspond to the scaled PERCLOS labels for the entire session.

of eye closure and sleepiness, sorting the cluster centroids assures that vigilance level 1 always corresponds to the highest vigilance level and subsequent clusters represent lower vigilance and higher sleepiness. It should be noted that we did run clustering schemes using other BP-ROI features, but their correlations coefficients were low and completely outperformed by the α/β features.

6.4.2.2 Consistent MI Features during Similar Vigilance Levels

Once the clustering of α/β features using the quantized CI_{smooth} and $CI_{movmean}$ was verified for the majority of SEED-VIG participants, the same procedure was applied for our SPIS MI-BCI dataset and the BCI Competition IV Dataset 2a. To detect consistent performance of MI time intervals during similar vigilance levels across different sessions, the methodology of Section 6.4.1.4 is applied on the two-class SPIS MI-BCI dataset as follows:

1. Vigilance-related features explained in Section 4.4.1.4 are extracted from 1-second intervals before the cue onset, and α/β cluster indices are calculated according to the schemes C_1 to C_3. In each experiment, each trial is labeled as either belonging to vigilance level VL_1 or VL_2.

2. The training session consisted of 200 trials for the first three and 300 trials for the last five participants while the test session had 300 trials for everyone. Here we attempted to detect imagination of class 1 (left hand) versus class 2 (right hand) using a combination of CSP and LDA on trials labeled with similar cluster indices.

Table 6.5: Winning $TI_{j,i}$ from two clustering schemes with the highest number of consistent time intervals for similar vigilance levels in the binary MI classification of 8-participant SPIS MI-BCI. Number of sessions for whom 3-fold CV was skipped due to the low number of samples is also reported.

Clustering Scheme	C_2', $m=5$, $k=3$				C_2, $m=5$, $k=2$			
	VL_1		VL_2		VL_1		VL_2	
Participant	S1	S2	S1	S2	S1	S2	S1	S2
S1	1	1	2	3	1	1	2	3
S2	1	1	4	1	1	1	4	1
S3	2	2	4	2	2	2	4	1
S4	1	1, 2	1	1	1	1	1	1
S5	1	1	3	0	1	1	3	0
S6	1	1	1	1	1	1	1	4
S7	3, 4	4	1	1	3, 4	4	1	1
S8	4	3	4	4	4	3	4	4
# Consistent Levels	7		4		7		3	

Table 6.5 presents the winning time intervals $TI_{j,i}$ from two clustering schemes that result in the highest number of consistent time intervals across all participants. As before, VL_1 and VL_2 correspond to relatively higher and lower vigilance levels (or, lower and higher sleepiness levels). Each clustering scheme has one zero session, i.e., a session for which the 3-fold CV was skipped due to the low number of samples and not because of inconsistent time intervals. It can be seen that from the total 32 $VL\times$ Session pairs observed in this table, these two clustering schemes have successfully obtained similar time intervals for 30 pairs. Interestingly, the former scheme has more agreeable sessions for VL_2 vigilance level, denoting the success of prediction using old centroids before using the whole data for updating new cluster coordinates.

Focusing on the 9-participant BCI Competition dataset and to detect consistent performance of MI time intervals during similar vigilance levels across different sessions, the methodology of Section 6.4.1.4 was applied as follows:

1. Features explained in Section 4.4.1.4 are extracted from 1-second intervals before the cue onset, and α/β cluster indices are calculated according to the three clustering schemes C_1 to C_3. In each scheme, each trial is labeled as either belonging to vigilance level VL_1 or VL_2.

2. Due to the 4-class nature of this dataset, having only 288 trials in a session, and an occasional low number of trials in VL_2 in most sessions, 2-class motor imagery classification is performed by grouping two of four classes together at a time to form positive and negative samples. In the first MI experiment, “1&2 vs. 3&4”, classes 1 and 2 (left and right hands) are classified versus classes 3 and 4 (both feet and tongue). In the second MI experiment, “1&4 vs. 2&3”, classes 1 and 4 (tongue and left hand) are classified against classes 2 and 3 (right hand and both feet). These experiments are performed due to the higher separability of features across all participants of this dataset [259]. To be more precise, since we were to

Table 6.6: Winning $TI_{j,i}$ from the clustering scheme with the highest number of consistent time intervals for similar vigilance levels in the binary MI classification of "1&2 vs. 3&4", BCI Competition IV - Dataset 2a. Number of trials from each vigilance level and each session is denoted in parentheses.

Clustering Scheme	C_3: quantized CI_{smooth}, k =2			
	VL_1		VL_2	
Participant	S1	S2	S1	S2
A1	2 (122)	1, 2 (214)	2 (166)	1 (74)
A2	2 (131)	2 (264)	2 (157)	0 (24)
A3	1 (123)	4 (52)	3 (165)	1, 3 (236)
A4	1 (180)	1 (157)	1, 3 (108)	1 (131)
A5	1 (214)	2 (282)	1 (74)	0 (6)
A6	1 (189)	1 (148)	1 (99)	3 (140)
A7	1 (164)	3 (203)	1, 2 (124)	1 (85)
A8	4 (139)	3 (222)	2 (149)	3, 4 (66)
A9	1 (207)	1 (96)	2 (81)	1 (192)
# Consistent Levels	5		3	

focus on the effect of vigilance on the MI BCI performance, we tried to utilize the already tested scenarios in terms of high separability in MI EEG features.

Thus, 10-fold CV is performed on separate vigilance levels of each session and the winning time intervals are reported. Tables 6.6 and 6.7 present the clustering schemes that resulted in the highest cases of consistent time intervals from similar vigilance levels across different sessions in each experiment. If two different time intervals resulted in exactly equal classification accuracies, both intervals are reported. As before, VL_1 and VL_2 correspond to relatively higher and lower vigilance levels (or, lower and higher sleepiness levels).

In MI classification of '1&2 vs. 3&4", the quantized version of smoothed CIs from both training and test sessions resulted in the highest number of agreeable intervals as shown in Table 6.6. As seen in this table, participants experienced similar $TI_{1,j}$. This number reduced to 3 participants with similar winning time intervals in the lower vigilance level of VL_2 across different sessions. However, more participants demonstrate consistent time intervals across two sessions when MI classes 1 and 4 were classified against classes 2 and 3, as shown in Table 6.7. When clustering train and test trials after 5 new samples with a maximum of two clusters, C_2', 6 and 4 participants had consistent time intervals in VL_1 and VL_2, respectively. When both training and test CIs were smoothed, this was in favor of the lower vigilance level which saw 5 out of 9 participants experiencing consistent time intervals.

A note on the winning adaptive alertness-aware method for lower vigilance levels: For the same BCI Competition dataset, we tested the scenario in which MI EEG features were extracted from trials of different lengths all starting at 0.5 seconds post-cue onset. In other words, six new time intervals were formed covering the [0.5, 1], [0.5, 1.5], [0.5,

Table 6.7: Winning $TI_{j,i}$ from clustering schemes with the highest number of consistent time intervals for similar vigilance levels in the binary MI classification of “1&4 vs. 2&3”, BCI Competition IV - Dataset 2a. Number of trials from each vigilance level and each session is denoted in parentheses.

Clustering Scheme	C_2', $m = 5$, k =2				C_3' smooth CI_{Train} & CI_{Test}, k =2			
	VL_1		VL_2		VL_1		VL_2	
Participant	S1	S2	S1	S2	S1	S2	S1	S2
A1	1, 4 (162)	1 (220)	1, 4 (126)	1 (68)	1 (122)	1 (214)	1 (166)	1 (74)
A2	4 (168)	4 (245)	4 (120)	0 (43)	1 (131)	4 (264)	4 (157)	0 (24)
A3	1 (163)	1 (120)	2 (125)	2 (168)	1 (123)	2 (52)	4 (165)	2 (236)
A4	4 (195)	1 (210)	3 (93)	4 (78)	4 (180)	2 (157)	4 (108)	1 (131)
A5	2 (218)	3 (259)	4 (70)	0 (29	2 (214)	2 (282)	4 (74)	0 (6)
A6	3 (220)	1 (180)	2 (81)	2 (108)	2 (189)	2 (148)	2 (99)	2 (140)
A7	3 (186)	3 (199)	3 (102)	1 (89)	3 (164)	3 (203)	2 (124)	2 (85)
A8	1 (173)	1 (220)	1 (115)	1 (68)	1 (139)	1 (222)	1 (149)	1 (66)
A9	1 (211)	1 (164)	1 (77)	3 (124)	1 (207)	3 (96)	1 (81)	1 (192)
# Consistent Levels	6		4		5		5	

2], [0.5, 3], [0.5, 3.5], and [0.5, 4] second intervals after onset of arrows. In this way, we were exploring whether increasing the amount of data during the periods of sleepiness would assist with increased MI accuracy and lead to using time intervals of equal length in both sessions. As shown on Table 6.8, clustering scheme C_3 with 2 clusters where the training CIs were smoothed and used to predict the test CIs results in the largest number of participants having consistent interval lengths across different sessions of VL_2.

6.4.2.3 Adaptive Alertness-Aware Classification for MI BCI

Following the discussion in Section 6.4.1.5 for adaptive feature extraction and classification of MI BCI, the vigilance information based on clustering of α/β features was utilized to perform adaptive classification for 2-class motor imagery in the SPIS MI-BCI dataset. Tables 6.9 and 6.10 demonstrate these results when using Version 1 of adaptation, ref. Section 6.4.1.5. Clustering and indexing vigilance levels were performed using a maximum of 3 clusters with the scheme C_2 (updating the centroids before prediction, $m = 5$), and C_2' (predicting cluster indices first before updating centroids after 5 new trials), respectively. The columns under “Original MI, No Adaptation” include results of a 3-fold CV on training trials only using the left and right MI labels, the time intervals with highest CV accuracy, and test accuracy using those MI features of those time intervals. Columns under “Vigilance Level 1” contain similar results when only trials labeled as VL_1 (the base vigilance level) by the clustering scheme and their best time intervals were used for classification. Similarly, results under “Vigilance Level 2” were obtained when only VL_2 trials from the lower vigilance level (higher drowsiness) and their winning time intervals were used for MI classifications. Finally, the total number of correctly classified MI trials from the two adapted classifications divided by the entire 300 trials in the test

Table 6.8: Winning time intervals from the clustering scheme C_3 with a maximum of k =2 that resulted in the highest number of consistent interval lengths for similar vigilance levels in the binary MI classification of “1&4 vs. 2&3”, BCI Competition IV - Dataset 2a. Numbers correspond to 6 intervals starting at 0.5 s post-cue onset.

	VL_1		VL_2	
Participant	S1	S2	S1	S2
A1	6	4	5	2, 5
A2	5	6	2, 6	0
A3	3	5	4	4
A4	6	1	1, 2, 4	1
A5	4	5	4, 5	0
A6	6	6	3	3
A7	1	6	2, 5	2
A8	4, 5, 6	4	6	3, 6
A9	2	3	3	4
# Consistent Levels	2		6	

session are reported under the “Vigilance” column.

As can be seen from Tables 6.9 and 6.10, the overall Acc_{Test} has improved for 5 out of 8 participants and remained the same for one participant in both clustering schemes. And the overall Acc_{Test} in the full adaptation case shows an average rate of change of 1.99±5.80 under C_2 and 0.99±5.21 under C_2'. Student’s t-test reported no significant statistical difference between the results of two clustering schemes, $p > 0.2$.

Two notable results are worth a special discussion: Participant S5 who had the highest Acc_{Test} of 70% without any adaptation saw a perfect MI detection rate for trials labeled as VL_2 under the C_2' clustering while C_2 had pointed to wrong MI time intervals. Analyzing the test CIs demonstrated that S5 only had one trial labeled as VL_2 in his test session. On the contrary, the overall Acc_{Test} of participant S6, 49.33%, had not changed under the train and update scheme C_2', but increased to 54.67% under the train and test scheme of C_2. For this participant, the Acc_{Test} for separate VL_1 and VL_2 classifications were also higher under the C_2 clustering. A review of facial videos showed that S6 had excessive sleepiness periods during their MI sessions. This brought us to the conclusion that they had performed motor imagination only for a short period during the training session.

For the same participants and using identical clustering schemes C_2 and C_2', adaptation paradigms Version 2 and Version 3 which modify the method of training CSP+LDA system for drowsy trials in VL_2 were also performed. Table 6.11 presents the average and standard deviation for improvements in overall Acc_{Test} in three Adaptation versions with respect to the Original, non-adaptive classification results. Using Version 2, the classification accuracy changed between -8.66% and +6.21% after adaptation while, using Version 3, improvements between -1.96% and 6.05% were obtained. Results indicate that Adaptation Version 3 outperforms the other two versions and obtains the highest improvement. Furthermore, the overall adaptive test accuracy had significantly increased with respect to the non-adaptive values for C_2 scheme in Version 2 and both clustering schemes in

Table 6.9: Classification accuracy for the SPIS MI-BCI dataset without adaptation (Original MI), and adaptation Version 1. Clustering is performed using the C_2 scheme with $m = 5$ and a maximum of 3 clusters. Highlighted cells demonstrate improved test accuracy after adaptation while bold cells indicate no change in the test accuracy. TI: Time Interval; Acc: Accuracy; SD: Standard Deviation.

	Original MI, No Adaptation			Vigilance Level 1			Vigilance Level 2			Vigilance
Participant	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	Acc_{Test} (%)
S1	50.00	1	49.67	54.62	1	55.69	51.43	2	49.62	53.00
S2	55.00	3	57.67	54.61	1	52.05	61.02	4	55.56	53.00
S3	66.00	1	71.67	56.91	2	76.28	64.94	2	74.12	75.67
S4	52.33	1	53.67	59.02	1	54.44	42.86	2	55.77	54.67
S5	64.00	1	**70.67**	66.36	2	70.90	57.83	4	0.00	**70.67**
S6	52.67	4	49.33	52.49	4	54.89	49.37	2	54.49	54.67
S7	50.67	1	58.52	55.91	4	60.00	47.50	1	58.57	59.11
S8	40.33	2	51.00	49.30	4	51.10	44.83	4	39.29	50.00
Average	53.88		57.78	56.15		59.42	52.47		48.43	58.85
SD	8.14		8.94	5.04		9.25	7.96		21.83	9.29

Table 6.10: Classification accuracy for the SPIS MI-BCI dataset without adaptation (Original MI), and adaptation Version 1. Clustering is performed using the C_2' scheme with $m = 5$ and a maximum of 3 clusters. Highlighted cells demonstrate improved test accuracy after adaptation while bold cells indicate no change in the test accuracy. TI: Time Interval; Acc: Accuracy; SD: Standard Deviation.

	Original MI, No Adaptation			Vigilance Level 1			Vigilance Level 2			Vigilance
Participant	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	Acc_{Test} (%)
S1	50.00	1	49.67	54.62	1	55.69	51.43	2	49.62	53.00
S2	55.00	3	57.67	54.61	1	51.39	61.02	4	55.95	52.67
S3	66.00	1	71.67	57.14	2	75.93	64.86	4	66.67	73.33
S4	52.33	1	53.67	53.97	1	58.96	45.83	1	51.02	57.67
S5	64.00	1	70.67	65.42	1	71.91	53.49	3	100.00	72.00
S6	52.67	4	**49.33**	50.45	1	47.79	48.75	1	50.61	**49.33**
S7	50.67	1	58.52	55.91	4	59.77	47.50	1	59.15	59.33
S8	40.33	2	51	48.62	4	50.18	42.90	4	40.74	49.33
Average	53.88		57.78	55.09		58.95	51.97		59.22	58.34
SD	8.14		8.94	5.02		10.19	7.57		18.14	9.53

Table 6.11: Average improvements, in percent, in overall Acc_{Test} of three Adaptation versions with respect to the Original, non-adaptive MI classification results for the SPIS MI-BCI dataset, $N = 8$. Results of one-sided, paired Student's t-test between the adaptive and non-adaptive test accuracy are indicated inside the parentheses.

Clustering Scheme	Version 1	Version 2	Version 3
C_2 , $m = 5$	1.99 ±5.80	2.32±2.66 ($p < 0.05$)	3.06±5.95 ($p < 0.1$)
C_2' , $m = 5$	0.99±5.21	1.50±5.08	2.85±5.22 ($p < 0.1$)

Table 6.12: Classification accuracy for the SPIS MI-BCI dataset without adaptation (Original MI), and Adaptation Version 3. Clustering is performed using the C_2' scheme with $m = 5$ and a maximum of 3 clusters. Highlighted cells demonstrate improved test accuracy after adaptation while bold cells indicate no change in the test accuracy. TI: Time Interval; Acc: Accuracy; SD: Standard Deviation.

	Original MI, No Adaptation			Vigilance Level 1			Vigilance Level 2			Vigilance
Participant	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	Acc_{Test} (%)
S1	50.00	1	49.67	54.62	1	55.69	51.43	2	53.38	54.67
S2	55.00	3	57.67	54.61	1	52.05	61.02	4	60.49	54.33
S3	66.00	1	71.67	56.91	2	76.28	64.94	2	71.67	75.00
S4	52.33	1	53.67	59.02	1	54.44	45.86	2	51.92	54.00
S5	64.00	1	70.67	66.36	2	70.90	57.83	4	0	70.67
S6	52.67	4	49.33	52.49	4	54.89	49.37	2	55.69	55.33
S7	50.67	1	58.52	55.91	4	60.00	47.50	1	61.43	60.89
S8	40.33	2	51.00	49.30	4	51.10	44.83	4	42.86	50.33
Average	53.88		57.78	56.15		59.42	52.47		49.69	59.40
SD	8.14		8.94	5.05		9.25	7.96		21.75	8.85

Version 3. Finally, Table 6.12 demonstrates the non-adaptive and adaptive classification results using the clustering scheme C_2 and adaptation Version 3. It should be clarified that only one test trial belonged to the second cluster for S5.

Finally, the vigilance information based on clustering of α/β features was utilized to perform adaptive "1&4 vs. 2&3" classification for BCI Competition Dataset 2a. Observing that the Adaptation version 3 had outperformed the other two in the SPIS MI-BCI dataset, here we only choose to present the classification result from the same version. Table 6.13 demonstrates these results when clustering and labeling of vigilance levels were performed using a maximum of 3 clusters with the scheme C_2' (predicting cluster indices first before updating centroids after 5 new trials). As can be seen in this table, Acc_{Test} has increased for 6 out of 9 participants and remained constant for one when using the whole vigilance information. Overall, an average improvement of 1.72±6.80 (%) is observed in the Acc_{Test} results of Adaptive versus Original classification.
To investigate the reason for the large reduction in the test accuracy of participant A4, the Euclidean distances between cluster centroids throughout the test session updates are computed for each participant. Figure 6.11 demonstrates these distances when the clustering scheme C_2', i.e., prediction of cluster indices and subsequent updates are performed

Table 6.13: Results of the binary MI classification of “1&2 vs. 3&4”, BCI Competition IV - Dataset 2a, without adaptation (Original MI) and adaptation Version 3. Clustering is performed using the C_2' scheme with $m = 5$ and a maximum of 3 clusters. Highlighted cells demonstrate improved test accuracy after adaptation while bold cells indicate no change in the test accuracy. TI: Time Interval; Acc: Accuracy; SD: Standard Deviation.

	Original MI, No Adaptation			Vigilance Level 1			Vigilance Level 2			Vigilance
Participant	10-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	10-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	3-fold Acc_{Train}	Winner Train TI	Acc_{Test} (%)	Acc_{Test} (%)
A1	90.97	2	92.36	93.21	2	93.64	89.68	2	89.71	92.71
A2	73.61	2	68.06	73.24	3	69.80	75.34	2	65.12	69.10
A3	81.60	1	85.07	79.04	1	83.76	75.21	4	85.38	84.72
A4	54.86	1	71.18	55.62	1	60.20	60.40	4	70.65	63.54
A5	61.81	3	57.64	53.21	1	66.02	52.86	1	65.52	65.97
A6	67.36	4	51.04	61.54	1	58.89	58.75	3	49.07	55.21
A7	86.46	2	84.38	85.47	1	86.60	80.73	2	84.04	85.76
A8	73.96	4	72.57	71.10	1	70.91	73.04	2	79.41	72.92
A9	60.76	1	**76.39**	61.14	1	74.39	63.64	1	79.03	**76.39**
Average	72.38		73.19	70.40		73.80	69.96		74.21	74.04
SD	12.31		13.26	13.73		11.98	11.84		12.82	12.05

with a maximum of 3 clusters. For A4, during the first 100 trials and later between trials 205 and 230, the distance between centroids is highly fluctuating. This participant had 2 samples from VL_3 in trials 39 and 47 of their test set which were concatenated with VL_2 for practical classification. It seems the algorithm was struggling with the computation of a new cluster throughout the test session, but the procedure of clustering was not stable for this participant.

Excluding the results of participant A4, Table 6.14 demonstrates the average and standard deviation for improvements in the overall Acc_{Test} in three Adaptation versions with respect to the Original, non-adaptive classification results of BCI Competition IV dataset 2a. Results indicate that here as well, the Adaptation Version 3 outperforms the other two versions and obtains the highest improvement. Furthermore, the overall adaptive test accuracy had significantly increased with respect to the non-adaptive values for C_2' scheme in Version 3.

6.4.3 Conclusion

In this chapter, we demonstrated that performance of MI BCI can be predicted by pre-trial EEG features from which vigilance information can be inferred. Since different spatio-spectral feature subsets yield the best result for different participants, we suggested that subject-based feature selection may yield more promising results. The predictive power of frontal features in a subset of participants was also in line with the localization of alerting and executive attention networks over frontal anatomic structures.

Furthermore, we presented an adaptive alertness-aware MI classification system built upon the assumption that vigilance level of users affect their reaction time and the ability to focus during mental imagination tasks, and hence affect the accuracy of decoding algo-

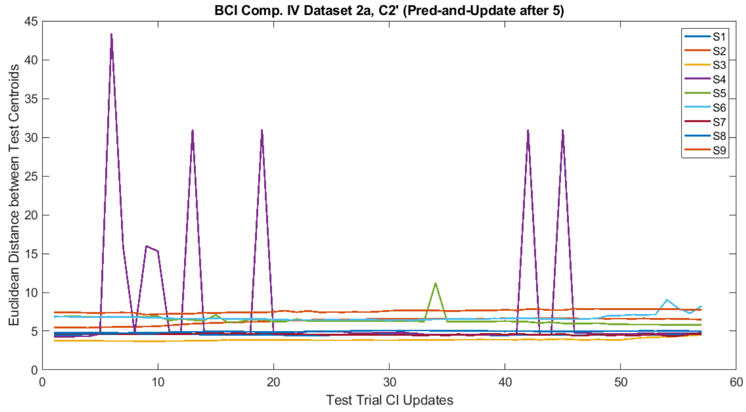


Figure 6.11: Euclidean distance between cluster centroids in the test session of BCI Competition IV - Dataset 2a, under scheme C_2' with a maximum of 3 clusters. Predictions and updates were performed after arrival of each 5 test samples.

Table 6.14: Average improvements, in percent, in overall Acc_{Test} of three Adaptation versions with respect to the Original, non-adaptive MI classification results for BCI Competition IV Dataset 2a, $N = 8$, excluding A4. Results of one-sided, paired Student's t-test between the adaptive and non-adaptive test accuracy are indicated inside the parentheses.

Clustering Scheme	Version 1	Version 2	Version 3
C_2' , $m = 5$	1.58±8.56	2.36±6.91	3.28±5.28 ($p < 0.1$)

rithms. We showed that successful MI time intervals are the same across similar vigilance levels of different experimental sessions, and demonstrated that, using the first session as a training/calibration dataset for both MI and vigilance clustering enables the system to predict vigilance-dependent performance of EEG-based BCI systems in the upcoming test/evaluation sessions. We verified the statistically significant improved classification accuracy on our SPIS MI-BCI and BCI Competition IV datasets. Unfortunately, the EEG recordings of Talukdar *et al.* [45] were not readily available for evaluation of the proposed paradigm.

When comparing Tables 6.11 and 6.14, it is clear that the utilized BCI competition dataset reports a higher improvement in the overall classification accuracy as a result of adaptation version 3. Analyzing results of respective classification tables point to the higher training accuracy for this dataset, 73.19 ± 13.26 (%), with respect to that of the SPIS BCI dataset, 57.78 ± 8.94 (%). This lower accuracy could have been caused by the lack of sufficient motor imagery skills for our participants which points to the importance of prior spatial and MI training sessions, preferably augmented through neurofeedback [29], before conducting long MI experiments. Since retraining the same group of participants is not possible at this point, future attempts could employ novel classifiers based on the DNN and CNN architectures, for example, to increase the training accuracy for the SPIS BCI dataset before or while incorporating the adaptive alertness-aware classification approach.

The reported results for both datasets also demonstrate the success of dynamic clustering schemes C_2 and C_2' compared to the static baseline of C_1 and smoothed indices of C_3 in obtaining the highest MI classification accuracy. This observation can be explained by the nonstationary nature of EEG signals while pointing to the association of short-term, phasic alertness in correct prediction of trial-wise MI performance.

7 Contributions and Future Work

7.1 Summary of Contributions

Existing experimental BCI systems are usually trained in a supervised fashion and then evaluated during test sessions. With increasing demands for daily and long-term use of BCIs such as in semi-autonomous cars, these systems have been tested on longer sessions during which researchers have observed considerably lower performance of trained systems partly due to variations in the underlying attention levels of operators. Vigilance level variation and failing to maintain consistent attention and performance levels is one of the major obstacles in applicability of the BCI systems in real-life environments. The literature and our own results in Figures 4.4 have shown that individuals demonstrate extremely different behavioral and neurophysiological patterns of maintaining their attention levels – sustained attention or vigilance – or falling into microsleep or even longer drowsiness periods. For the first part of this thesis, Chapters 3 and 4, we ran an experimental session of sustained attention to response task (SART) that lasted for 105 minutes and drove brain networks to intense boredom due to its repetitive nature and the low ratio of target-to-non-target stimuli. We observed significant inter-individual differences in attention maintenance. After proposing an objective and multi-level adaptive performance measure and its continuous-valued version, the Cumulative Vigilance Score (CVS) that summarized the tonic patterns in error and response time, we proposed to determine the EEG-based correlates of large variations in personal vigilance traits. We first asked the question of whether brain dynamics during the resting states recorded immediately before the cognitive task could explain common correlates of performance variation. We then presented two multi-variate regression models based on feature relevance analysis and a deep neural network (DNN) architecture that showed the role of various spatio-spectral EEG features, including opposite roles of lower and higher beta components and the effect of high gamma from left temporal – the default mode network (DMN) – in predicting higher mean and lower variability of objective attention performance measures. The DNN solution was specially useful in visualizing hidden weights of trained networks that are presented for the first time in the literature.

Our analysis was not limited to band-power features and we considered pairwise phase synchrony across the whole cortical regions to build descriptive EEG models of function connectivity. We report that the right and left tempo-parietal channels during the eyes-open and closed resting states should be more synchronized with the opposite frontal and central channels as correlates of fewer commission errors and more consistent CVS and response time. Resting-state interhemispheric synchrony in the alpha and beta-bands also predicted more variability of CVS and response time. This is the first time that EEG-based PSI predictors of both commission and omission errors as well as continuous-valued vigilance scores and response time and their variability are presented in the literature. In another regression scheme using deep neural networks, symmetric matrices of phase synchrony features from EEG signals in a one-second time windows before the occurrence of visual stimuli were computed for cross-subject prediction. Phase synchrony index (PSI) from lower beta-2 (12-16 Hz) and alpha band (8-12 Hz), respectively, returned the largest correlation coefficient and smallest error for prediction of average CVS and response time in long-term, 8-second experimental blocks across all participants. We thus demonstrated the consistency of our EEG-based phase synchrony features in two different settings while adding to the literature on technical capability of vigilance modeling with deep neural networks.

In the second part of Chapter 4, we focused on between-subject, two-class analysis of extreme vigilance performance for each participant. Although several studies listed in Table 4.12 limit their definition of drowsy and alert cognitive states to the beginning and end of experimental sessions or choose the easiest and hardest blocks for the dataset generation, we considered the highest and lowest CVS scores of each participant in an adaptive and objective manner. Our novel extraction of phase locking values from various time intervals before and after the stimulus onset paired with a deep convolutional neural network (CNN) outperforms the classification results of [191] and [193] who also explored role of pairwise spatio-spectro-temporal features and were essentially different from the driving conditions of other studies. Gamma-band PLVs especially outperform the rest of phase-locking values features for this classification and pointed to the role of high-frequency associations across different cortical regions for improved information transfer and improved activation in task-related networks [153]. We believe our classification results can be improved by incorporating both recurrent neural networks (RNNs) or LSTMs and CNNs. Attention mechanisms can also be incorporated to learn temporal dependencies across neighboring trials.

In the second part of this thesis, Chapters 5 and 6, we focus on the fact that objective labeling of vigilance/alertness/sustained attention is a challenge during purely cognitive BCI tasks as the participants do not provide any physical input in the form of clicks or response time. Since vigilance curves and their EEG predictors are continuous-valued time-series, we propose two unsupervised models for identifying the exact *moments* at which a vigilance transition occurs. To study feasibility of inferring vigilance level information from an unlabeled time-series, in Chapter 5 we utilize two offline and online Bayesian changepoint detection (BCPD) algorithms to locate transitions of vigilance curves and their EEG-based band-power features on SEED-VIG, a dataset labeled with continuous-valued eye-closure events from driving simulations [35]. We demonstrate that online BCPD from these EEG features results in an average precision of 0.82 and average recall of 0.56, and significantly outperforms the offline algorithms for detecting vigilance transitions themselves detected objectively by similar algorithms. This novel application of dynamic and sequential inference techniques for chanegepoint detection from band-power features can be implemented in unsupervised vigilance monitoring systems to alarm users about onsets of drowsiness.

Another application of such unsupervised vigilance detection methods is to build adaptive BCIs that detect the onset of vigilance changes and, consequently, modify the classification parameters in covert adaptation or the interface parameters in overt adaptation schemes. Therefore, in Chapter 6, an adaptive MI BCI classification approach based on the continuous assessment of alertness/vigilance information is proposed that considers adaptation in both the best time intervals for MI feature extraction as well as in setting the classifier parameters. Without pausing the experiment to collect subjective information about fatigue and drowsiness, information for vigilance estimation during MI recordings are obtained from unsupervised learning of EEG correlates for increased drowsiness based on our earlier studies [52], [53]. Once vigilance clustering of training trials is complete, the best time interval for MI inference is selected and classifiers are trained and tested separately for each vigilance level. Three different versions of this adaptive classification framework are introduced. We report improvements in the overall test accuracy of adaptive versions with respect to the original, non-adaptive baseline from the dataset collected in our laboratory, referred to as the SPIS MI-BCI dataset, and the BCI Competition IV Dataset 2a.

To the best of our knowledge, this is the first time that a fully covert adaptive classification paradigm is proposed that utilizes adaptation based on alertness information in both its feature extraction and classification. Although extensive work has been done in the past 5 years for characterizing neural correlates of drowsiness, fatigue, and vigilance in the context of simulated driving and, less frequently, sustained task execution using EEG and EOG signals, the most recent work on incorporating this information with MI execution has resulted in only reporting improved separability of MI-based CSP features, measured by a number of distance indices, and not improved classification accuracy under the proposed adaptations [258].

The improvements in classification accuracy reported in Tables 6.9 and 6.12 are meaningful considering the personal traits of attention maintenance. The success of dynamic clustering schemes C_2 and C_2' compared to the static baseline or smoothed cluster indices can be explained by the nonstationarity of EEG signals and the link between phasic alertness and correct prediction of trial-wise MI performance. Shorter periods of attention drifts are more common as observed by the recorded facial videos and cause variations in the learned features and covariate shifts not only across sessions but also between trials of the same session [5]. Our proposed adaptive alertness-aware MI classification approach achieves improved MI accuracy even for one particular participant, S6, who had long periods of drowsiness in the 30-minute training session (from 49.33% to 55.33% using version 3, thus a maximum increase of 12.16%), and provides a promising perspective for training of users with highly unstable vigilance traits. In another participant, S4, who was highly drowsy as well and had a baseline accuracy of 53.67%, version 1 of the adaptive BCI approach obtained a test accuracy of 57.67%, showing a 7.45% increase. Meanwhile, for participant S5 who was already highly alert and had a baseline accuracy of 70.67%, version 1 of this adaptive approach provided an improvement of 1.88%, thus showing no serious need for adaptation in this case.

The second set of results belongs to the BCI Competition IV Dataset 2a obtained from 9 participants of a 4-class MI paradigm in which the adaptive approach shows an average improvement of 3.28%. We envision that the observed increase in classification accuracy for users with highly unstable attention patterns provide a very promising path for improving the usability of BCI systems in and outside laboratory settings.

7.2 Future Research Directions

The proposed extensions for this work fall in two main lines of work: (1) dynamic modeling and changepoint detection from BP and PLV EEG time-series for alarming users of transitions in their vigilance levels, and (2) improving the performance of proposed alertness-aware adaptive classifier.

For the first line of work, we have already tested implementations with the HMM and RNN/LSTM architectures. A problem we faced for predicting SART time-series with HMMs is that often it is not possible to include the number of states in the model design or even calculate the transition and emission probabilities during training since the model does not observe enough variations in a participant with high performance in the first half of experiment. To this end, an infinite HMM (IHMM) with a Dirichlet process approach, as used for determining number of brain states in simulated sleep datasets [249], would be an interesting approach as it would not need to be aware of the number of states. However, the proposed solution with the Gibbs sampler requires a large number of iterations that prohibits an efficient and real-time implementation. Furthermore, we have already used LSTM in the context of our SART dataset for prediction of average response time and CVS, but the high number of dimensions and channels made the system susceptible to overfitting.

We instead turn our attention on Bayesian regression structures that consider spatio-temporal dependencies among their input features, including the Continuous Conditional Neural Fields (CCNF) [113] and especially its chain-CCNF version that is capable of solving regression for time-series with varying lengths. CCNF is an undirected graphical model that learns the condition probabilities of outputs based on inputs. Continuous Conditional Random Fields (CCRF) [264] is another suitable model, but it needs initial predictions from unstructured regressions such as support vector regression (SVR). An application for predicting eye movements using horizontal and vertical components of EOG signals was presented for the SEED-VIG dataset using these architectures [35], and the best correlation results were reported for CCNF and temporal signals using differential entropy (DE) in a 5-fold cross-validation approach. We propose to extend that work by predicting the PERCLOS label itself using our BP and PLV features and extending them to use in SART dataset. Our current results indicate chain-CCNF with α/β features followed by CCRF with α/β features outperform the rest of BP-ROI datasets in terms of mean correlation metric. It would be interesting to feed PLV matrices to grid-CCNF and analyze performance of these regression architectures in extracting spatial information across different sequence lengths. Similarly, changepoint detection from temporal sequences of two-dimensional PLV matrices would be another interesting line of work to explore.

The alertness-aware adaptive BCI paradigm introduced in Chapter 6 was to be evaluated online on new participants in our laboratory; however, at the time of writing this dissertation, this plan had to be discarded due to the ongoing lockdown. We still believe that new BCI users should be trained by receiving informative visual feedback to increase their mastery of motor imagination [29] before evaluating variations in their underlying sustained attention levels in a long MI BCI session. However, the proposed attention/alertness-aware paradigm uses pre-stimulus features that are associated with general drowsiness, and thus can be used as the neural correlates of attention variations in tasks other than the motor imagery-based BCI. Multi-class spellers based on P300 event-related potentials or SSVEPs could benefit from attention inference and subsequent adaptations as well. It would also be interesting to explore the fields of auditory stimuli to see if their classification could be improved using such alertness-aware paradigms. Since attention variation is a key challenge for learners affected by attention deficit disorders or people working in unstimulating environments, the proposed unsupervised changepoint detection and the clustering schemes could have immense potentials as two augmenting and assistive methodologies for improving the performance of children and adults in cognitive tasks, mathematical and analytical paradigms, and continuous learning and reading applications.

In another important extension, we will focus on improving the classification accuracy of the proposed algorithm in Section 6.4.1.5. Noticing the low MI training accuracy of the SPIS MI BCI dataset and the fact that new training is not possible, one could improve the generalization of drowsy classifiers in Figure 6.8 by using extra drowsy trials from the test sessions of the same participants to validate the applicability of the proposed adaptive classifiers. This task can even be extended to use a leave-one-subject-out approach and use all the trials labeled as low vigilance from other participants to train the CSP+LDA classifiers that otherwise suffer from insufficient drowsy trials.

This work was the first attempt for proposing a solution for BCI systems using online inference of attention variations towards the challenging goal of developing neuroadaptive BCIs [265]. Such adaptations can be done in two ways, one by overt adaptation – updating properties of the visual interface to increase alertness and concentration of the otherwise drowsy user – or covert adaptation by updating the classifier parameters. For the first method, once the BCI decision software is notified of a transition in underlying vigilance level either through clustering or changepoint detection, the color, frequency, and size of stimulus on the screen could be modified to attract more attention and engage the user back in the task of mental imagination. In the second method, in addition to extracting fixed-length time intervals for MI execution as we proposed in this work, the raw EEG signals could be fed to a deep learning architecture ideally composed of both temporal and spatio-spectral layers for feature learning. Employing novel classifiers based on the DNN and CNN architectures that use the attention mechanism and explore a variety of temporal resolutions from EEG signals can increase the accuracy of adaptive alertness-aware classification without requiring the need for pre-detecting the winning MI time intervals. This feature will also help to generalize the classification approach to other BCI paradigms. Information on estimated vigilance levels will then be used for adaptive feature extraction and for applying a unique combination of hidden weights to obtain improved prediction of brain dynamics under a variety of task paradigms.

Bibliography

[1] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, N. Ramsey, P. Salomon, and G. R. Müller-Putz, "BNCI Horizon 2020: towards a roadmap for the BCI community," *Brain-Computer Interfaces*, vol. 2, no. 1, pp. 1–10, 2015, ISSN: 2326-263X. DOI: 10.1080/2326263X.2015.1008956.

[2] L. Zapparoli, L. M. Sacheli, S. Seghezzi, M. Preti, E. Stucovitz, F. Negrini, C. Pelosi, N. Ursino, G. Banfi, and E. Paulesu, "Motor imagery training speeds up gait recovery and decreases the risk of falls in patients submitted to total knee arthroplasty," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, ISSN: 20452322. DOI: 10.1038/s41598-020-65820-5. [Online]. Available: http://dx.doi.org/10.1038/s41598-020-65820-5.

[3] C. Jeunet, D. Hauw, and J. d. R. Millán, "Sport Psychology: Technologies Ahead," *Frontiers in Sports and Active Living*, vol. 2, no. February, pp. 1–3, 2020. DOI: 10.3389/fspor.2020.00010.

[4] M. Cocks, C. A. Moulton, S. Luu, and T. Cil, "What surgeons can learn from athletes: Mental practice in sports and surgery," *Journal of Surgical Education*, vol. 71, no. 2, pp. 262–269, 2014, ISSN: 18787452. DOI: 10.1016/j.jsurg.2013.07.002. [Online]. Available: http://dx.doi.org/10.1016/j.jsurg.2013.07.002.

[5] H. Raza, D. Rathee, S. M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface," *Neurocomputing*, vol. 343, pp. 154–166, 2019, ISSN: 18728286. DOI: 10.1016/j.neucom.2018.04.087. arXiv: 1805.01044. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.04.087.

[6] T. O. Zander and S. Jatzev, "Context-aware brain-computer interfaces: Exploring the information space of user, technical system and environment," *Journal of Neural Engineering*, vol. 9, no. 1, 2012, ISSN: 17412560. DOI: 10.1088/1741-2560/9/1/016003.

[7] I. Yilmaz, S. Demir, T. Tasdizen, and M. Cetin, "Semi-supervised Adaptation of Motor Imagery Based BCI Systems," in *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, IEEE, 2015, pp. 1841–1844, ISBN: 9781467373869.

[8] Y. Choi, J. Park, and D. Shin, "A semi-supervised inattention detection method using biological signal," *Annals of Operations Research*, vol. 258, no. 1, pp. 59–78, 2017, ISSN: 0254-5330. DOI: 10.1007/s10479-017-2406-6.

[9] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, and Y. Zhang, "A Survey on Deep Learning based Brain Computer Interface: Recent Advances and New Frontiers," *arXiv preprint arXiv:1905.04149*, 2019, ISSN: 01635964. arXiv: 1905.04149. [Online]. Available: http://arxiv.org/abs/1905.04149.

[10] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. D. R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006, ISSN: 15344320. DOI: 10.1109/TNSRE.2006.875642.

[11] M. Palus, "Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos.," *Biological cybernetics*, vol. 75, no. 5, pp. 389–396, 1996, ISSN: 03401200. DOI: 10.1007/s004220050304.

[12] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface," *Neural computation*, vol. 28, no. 8, pp. 2146–2171, 2013. DOI: 10.1162/NECO.

[13] S. H. Hsu and T. P. Jung, "Monitoring alert and drowsy states by modeling EEG source nonstationarity," *Journal of Neural Engineering*, vol. 14, no. 5, 2017, ISSN: 17412552. DOI: 10.1088/1741-2552/aa7a25.

[14] C. Vidaurre, M. Kawanabe, P. Von Bünau, B. Blankertz, and K. R. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3 PART 1, pp. 587–597, 2011, ISSN: 00189294. DOI: 10.1109/TBME.2010.2093133.

[15] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A Fully On-Line Adaptive BCI," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1214–129, 2006. DOI: 10.1109/TBME.2006.873542.

[16] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K. R. Müller, "Towards adaptive classification for BCI," *Journal of Neural Engineering*, vol. 3, no. 1, 2006, ISSN: 17412560. DOI: 10.1088/1741-2560/3/1/R02.

[17] D. Wu, "Online and Offline Domain Adaptation for Reducing BCI Calibration Effort," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 550–563, 2017, ISSN: 21682291. DOI: 10.1109/THMS.2016.2608931. arXiv: 1702.02897.

[18] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain Topography*, vol. 23, no. 2, pp. 194–198, 2010, ISSN: 08960267. DOI: 10.1007/s10548-009-0121-6.

[19] M. J. Bryan, S. A. Martin, W. Cheung, and R. P. N. Rao, "Probabilistic co-adaptive brain–computer interfacing," *Journal of Neural Engineering*, vol. 10, no. 6, p. 066 008, 2013, ISSN: 1741-2560. DOI: 10.1088/1741-2560/10/6/066008.

[20] C. Vidaurre, C. Sannelli, K. R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, 2011, ISSN: 17412560. DOI: 10.1088/1741-2560/8/2/025009.

[21] X. Yong, M. Fatourechi, R. K. Ward, and G. E. Birch, "Adaptive classification in a self-paced hybrid brain-computer interface system," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3274–3279, 2012, ISSN: 1557170X. DOI: 10.1109/EMBC.2012.6346664.

[22] A. Llera Arenas, "Adaptive classification on Brain Computer Interfaces using reinforcement signals," *Neural Computation*, vol. 24, no. 11, pp. 2900–2923, 2012.

[23] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, 2008, ISSN: 1098-7576. DOI: 10.1109/IJCNN.2008.4634130.

[24] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K. R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2010, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.03.022.

[25] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, 2017, ISSN: 15344320. DOI: 10.1109/TNSRE.2016.2627016.

[26] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017, ISSN: 10970193. DOI: 10.1002/hbm.23730. arXiv: 1703.05051.

[27] 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," *Journal of Neural Engineering*, vol. 15, no. 5, aace8c, 2018, ISSN: 17412552. DOI: 10.1088/1741-2552/aace8c.

[28] A. Jafarifarmand and M. A. Badamchizadeh, "Real-time multiclass motor imagery brain-computer interface by modified common spatial patterns and adaptive neuro-fuzzy classifier," *Biomedical Signal Processing and Control*, vol. 57, p. 101 749, 2020, ISSN: 17468108. DOI: 10.1016/j.bspc.2019.101749. [Online]. Available: https://doi.org/10.1016/j.bspc.2019.101749.

[29] L. Pillette, "Redefining and Adapting Feedback for Mental-Imagery based Brain-Computer Interface User Training to the Learners' Traits and States," Ph.D. dissertation, Universite de Bordeaux, 2019.

[30] C. Jeunet, B. Nkaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns," *PLoS ONE*, vol. 10, no. 12, pp. 1–21, 2015, ISSN: 19326203. DOI: 10.1371/journal.pone.0143962. arXiv: journal.pone.0143962 [10.1371].

[31] C. Jeunet, B. N. Kaoua, and F. Lotte, "Advances in User-Training for Mental-Imagery Based BCI Control: Psychological and Cognitive Factors and their Neural Correlates," *Progress in Brain Research*, vol. 228, pp. 3–35, 2016. DOI: 10.1016/bs.pbr.2016.04.002.

[32] J. S. Muller, C. Vidaurre, M. Schreuder, F. C. Meinecke, P. V. Bunau, and K.-R. Müller, "A mathematical model for the two-learners problem," *Journal of Neural Engineering*, vol. 14, no. 3, p. 036 005, 2017.

[33] A. Myrden and T. Chau, "Effects of user mental state on EEG-BCI performance," *Frontiers in Human Neuroscience*, vol. 9, no. June, pp. 1–11, 2015, ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00308.

[34] Z. Emami and T. Chau, "The effects of visual distractors on cognitive load in a motor imagery brain-computer interface," *Behavioural Brain Research*, vol. 378, no. September 2019, p. 112 240, 2020, ISSN: 18727549. DOI: 10.1016/j.bbr.2019.112240. [Online]. Available: https://doi.org/10.1016/j.bbr.2019.112240.

[35] W. L. Zheng and B. L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of Neural Engineering*, vol. 14, no. 2, 2017, ISSN: 17412552. DOI: 10.1088/1741-2552/aa5a98. arXiv: 1611.08492.

[36] M. Torkamani-Azar, S. D. Kanik, S. A. Ali Ahmed, S. Aydin, and M. Cetin, "Prediction of Response Time and Vigilance Score in a Sustained Attention Task from Pre-trial Phase Synchrony using Deep Neural Networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 676–679, ISBN: 9781538613115. DOI: 10.1109/embc.2019.8856291.

[37] M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers in Neuroscience*, vol. 6, no. JULY, pp. 1–31, 2012, ISSN: 16624548. DOI: 10.3389/fnins.2012.00055.

[38] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.11.004. arXiv: 0904.3664v1.

[39] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, 2007, ISSN: 17412560. DOI: 10.1088/1741-2560/4/2/R01.

[40] M. I. Posner and M. K. Rothbart, "Research on Attention Networks as a Model for the Integration of Psychological Science," *Annual Review of Psychology*, vol. 58, no. 1, pp. 1–23, 2007, ISSN: 0066-4308. DOI: 10.1146/annurev.psych.58.110405.085516.

[41] T. G. Monteiro, C. Skourup, and H. Zhang, "Using EEG for Mental Fatigue Assessment: A Comprehensive Look into the Current State of the Art," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 6, pp. 599–610, 2019, ISSN: 21682305. DOI: 10.1109/THMS.2019.2938156.

[42] A. Jafarifarmand and M. A. Badamchizadeh, "EEG Artifacts Handling in a Real Practical Brain-Computer Interface Controlled Vehicle," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1200–1208, 2019, ISSN: 15580210. DOI: 10.1109/TNSRE.2019.2915801.

[43] J. H. Song and K. Nakayama, *Hidden cognitive states revealed in choice reaching tasks*, 2009. DOI: 10.1016/j.tics.2009.04.009.

[44] A.-M. Brouwer, T. O. Zander, and J. B. F. van Erp, *Using neurophysiological signals that reflect cognitive or affective state*, APR. Frontiers Media SA, 2015, vol. 9, p. 193, ISBN: 9782889196135.

[45] U. Talukdar, S. M. Hazarika, and J. Q. Gan, "Adaptation of Common Spatial Patterns based on mental fatigue for motor-imagery BCI," *Biomedical Signal Processing and Control*, vol. 58, p. 101 829, 2020, ISSN: 17468108. DOI: 10.1016/j.bspc.2019.101829. [Online]. Available: https://doi.org/10.1016/j.bspc.2019.101829.

[46] V. Bekhtereva, C. Sander, N. Forschack, S. Olbrich, U. Hegerl, and M. M. Müller, "Effects of EEG-vigilance regulation patterns on early perceptual processes in human visual cortex," *Clinical Neurophysiology*, vol. 125, no. 1, pp. 98–107, 2014, ISSN: 13882457. DOI: 10.1016/j.clinph.2013.06.019.

[47] G. Hossain, "Rethinking self-reported measure in subjective evaluation of assistive technology," *Human-centric Computing and Information Sciences*, vol. 7, no. 1, p. 23, 2017. DOI: 10.1186/s13673-017-0104-7.

[48] C. Aydin, "The differential contributions of visual imagery constructs on autobiographical thinking," *Memory*, vol. 26, no. 2, pp. 189–200, 2018, ISSN: 14640686. DOI: 10.1080/09658211.2017.1340483.

[49] o. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, no. May, pp. 24–37, 2018, ISSN: 18790534. DOI: 10.1016/j.compbiomed.2018.05.019.

[50] S. I. Dimitriadis, C. Salis, and D. Linden, "A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates," *Clinical Neurophysiology*, vol. 129, no. 4, pp. 815–828, 2018, ISSN: 18728952. DOI: 10.1016/j.clinph.2017.12.039.

[51] S. Olbrich, C. Sander, H. Matschinger, R. Mergl, M. Trenner, P. Schönknecht, and U. Hegerl, "Brain and Body: Associations Between EEG-Vigilance and the Autonomous Nervous System Activity During Rest," *Journal of Psychophysiology*, vol. 25, no. 4, pp. 190–200, 2011, ISSN: 0269-8803. DOI: 10.1027/0269-8803/a000061.

[52] M. Torkamani-Azar, S. D. Kanik, S. Aydin, and M. Cetin, "Prediction of Reaction Time and Vigilance Variability from Spatio-Spectral Features of Resting-State EEG in a Long Sustained Attention Task," *IEEE Journal of Biomedical and Health Informatics*, vol. 2194, no. c, pp. 1–9, 2020. DOI: 10.1109/JBHI.2020.2980056.

[53] M. Torkamani-Azar, A. Jafarifarmand, and M. Cetin, "Prediction of Motor Imagery Performance based on Pre-Trial Spatio-Spectral Alertness Features," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020, pp. 3062–3065.

[54] U. Hoffmann, G. Garcia, J. M. Vesin, K. Diserenst, and T. Ebrahimi, "A boosting approach to P300 detection with application to brain-computer interfaces," *2nd International IEEE EMBS Conference on Neural Engineering*, pp. 97–100, 2005, ISSN: 1062922X. DOI: 10.1109/CNE.2005.1419562.

[55] J. L. Sirvent Blasco, E. Iáñez, A. Úbeda, and J. M. Azorín, "Visual evoked potential-based brain-machine interface applications to assist disabled people," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7908–7918, 2012, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.01.110.

[56] C. Ulas and M. Cetin, "Incorporation of a Language Model into a Brain Computer Interface Based Speller through HMMs," *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1138–1142, 2013.

[57] T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner, "Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces," in *Brain-computer interfaces*, Springer, 2010, pp. 181–199, ISBN: 978-3-642-02090-2. DOI: 10.1007/978-3-642-02091-9. arXiv: arXiv:1011.1669v3.

[58] F Lotte, "A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces," *Guide to Brain-Computer Music Interfacing*, 2014.

[59] A.-M. Brouwer, T. O. Zander, J. B. Van Erp, and A. W. Korteling, Johannes E Bronkhorst, "Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls," *Frontiers in Neuroscience*, vol. 9, p. 136, 2015, ISSN: 1662453X. DOI: 10.3389/fnins.2015.00136.

[60] C. Roda, *Human Attention in Digital Environments*. Cambridge University Press, 2011, ISBN: 9780521765657.

[61] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. 1995, Oxford University Press, USA.

[62] M. Lotze, P. Montoya, M. Erb, E. Hülsmann, H. Flor, U. Klose, N. Birbaumer, and W. Grodd, "Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study," *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 491–501, 1999, ISSN: 0898929X. DOI: 10.1162/089892999563553.

[63] S. J. Luck, *An Introduction to the Event-related Potential Technique*. The MIT Press, 2005.

[64] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews*, vol. 29, no. 2-3, pp. 169–195, 1999, ISSN: 01650173. DOI: 10.1016/S0165-0173(98)00056-3.

[65] W. Klimesch, P. Sauseng, and C. Gerloff, "Enhancing cognitive performance with repetitive transcranial magnetic stimulation at human individual alpha frequency," *European Journal of Neuroscience*, vol. 17, no. 5, pp. 1129–1133, 2003, ISSN: 0953816X. DOI: 10.1046/j.1460-9568.2003.02517.x.

[66] P. M. Dockree, S. P. Kelly, J. J. Foxe, R. B. Reilly, and I. H. Robertson, "Optimal sustained attention is linked to the spectral content of background EEG activity: greater ongoing tonic alpha (10 Hz) power supports successful phasic goal activation," *European Journal of Neuroscience*, vol. 25, no. 3, pp. 900–907, 2007, ISSN: 0953816X. DOI: 10.1111/j.1460-9568.2007.05324.x.

[67] T. H. Grandy, M. Werkle-Bergner, C. Chicherio, F. Schmiedek, M. L{\"o}vd{\'e}n, and U. Lindenberger, "Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults," *Psychophysiology*, vol. 50, no. 6, pp. 570–582, 2013, ISSN: 14698986. DOI: 10.1111/psyp.12043.

[68] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels," *Biomedical Signal Processing and Control*, vol. 38, pp. 302–311, 2017, ISSN: 17468108. DOI: 10.1016/j.bspc.2017.06.016. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2017.06.016.

[69] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP\AM-BA-SVM Approach for Motor Imagery BCI System," *IEEE Access*, vol. 6, pp. 49 192–49 208, 2018, ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2868178.

[70] H. Abbaspour, N. Mehrshad, and S. M. Razavi, "An Effective Brain-Computer Interface System Based on the Optimal Timeframe Selection of Brain Signals," *International Clinical Neuroscience Journal*, vol. 5, no. 1, pp. 35–42, 2018, ISSN: 2383-1871. DOI: 10.15171/icnj.2018.07. [Online]. Available: https://doi.org/10.15171/icnj.2018.07.

[71] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and P. G., "BCI Competition 2008 – Graz data set A," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[72] S. Weintraub, "Neuropsychological Assessment of Mental State," in *Principles of behavioral and cognitive neurology*, Oxford University Press, 2000, pp. 121–173.

[73] J. A. Deutsch and D. Deutsch, "Attention: Some theoretical considerations," *Psychological Review*, vol. 70, no. 1, pp. 80–90, 1963. DOI: 10.1037/h0039515.

[74] M. I. Posner, C. R. R. Snyder, and R. L. Solso, "Attention and cognitive control," *Cognitive psychology: Key readings*, pp. 205–223, 2004.

[75] P. Pennekamp, R. Bosel, A. Mecklinger, and H. Ott, "Differences in EEG-theta for responded and omitted targets in a sustained attention task," *Journal of Psychophysiology*, vol. 8, no. 2, pp. 131–141, 1994, ISSN: 02698803.

[76] W. C. McCallum, *Attention*, 2015. [Online]. Available: https://global.britannica.com/topic/attention.

[77] National Institutes of Health, "NIH Toolbox Training Manual – English version, September 2012," [Online]. Available: http://www.nihtoolbox.org/HowDoI/HowToAdministerTheToolbox/Pages/TrainingManual.aspx.

[78] M. Sarter, B. Givens, and J. P. Bruno, "The cognitive neuroscience of sustained attention: where top-down meets bottom-up," *Brain Research Reviews*, vol. 35, no. 2, pp. 146–160, 2001, ISSN: 01650173. DOI: 10.1016/S0165-0173(01)00044-3.

[79] B. Oken, M. Salinsky, and S. Elsas, "Vigilance, alertness, or sustained attention: physiological basis and measurement," *Clinical Neurophysiology*, vol. 117, no. 9, pp. 1885–1901, 2006, ISSN: 13882457. DOI: 10.1016/j.clinph.2006.01.017. arXiv: NIHMS150003.

[80] A. S. Smit, P. A. Eling, and A. M. Coenen, "Mental effort causes vigilance decrease due to resource depletion," *Acta Psychologica*, vol. 115, no. 1, pp. 35–42, 2004, ISSN: 00016918. DOI: 10.1016/j.actpsy.2003.11.001.

[81] M. M. Lorist, E. Bezdan, M. ten Caat, M. M. Span, J. B. Roerdink, and N. M. Maurits, "The influence of mental fatigue and motivation on neural network dynamics; an EEG coherence study," *Brain Research*, vol. 1270, pp. 95–106, 2009, ISSN: 00068993. DOI: 10.1016/j.brainres.2009.03.015.

[82] E. K. Papies, T. M. Pronk, M. Keesman, and L. W. Barsalou, "The benefits of simply observing: Mindful attention modulates the link between motivation and behavior.," *Journal of Personality and Social Psychology*, vol. 108, no. 1, pp. 148–170, 2015, ISSN: 1939-1315. DOI: 10.1037/a0038032.

[83] S. M. Doran, H. P. Van Dongen, and D. F. Dinges, *Sustained attention performance during sleep deprivation: evidence of state instability*, 2001.

[84] S Drummond, "The Effects of Total Sleep Deprivation on Cerebral Responses to Cognitive Performance," *Neuropsychopharmacology*, vol. 25, no. 5, S68–S73, 2001, ISSN: 0893133X. DOI: 10.1016/S0893-133X(01)00325-6.

[85] W. S. Helton and P. N. Russell, "Rest is best: The role of rest and task interruptions on vigilance," *Cognition*, vol. 134, pp. 165–173, 2015, ISSN: 00100277. DOI: 10.1016/j.cognition.2014.10.001.

[86] P. M. Dockree, S. P. Kelly, R. A. Roche, M. J. Hogan, R. B. Reilly, and I. H. Robertson, "Behavioural and physiological impairments of sustained attention after traumatic brain injury," *Cognitive Brain Research*, vol. 20, no. 3, pp. 403–414, 2004, ISSN: 09266410. DOI: 10.1016/j.cogbrainres.2004.03.019.

[87] J. C. Ballard, "Computerised assessment of sustained attention: A review of factors affecting vigilance performance," *Journal of Clinical and Experimental Neuropsychology*, vol. 18, no. 6, pp. 843–863, 1996.

[88] C. K. Conners and G. Sitarenios, "Conners' Continuous Performance Test (CPT)," in *Encyclopedia of Clinical Neuropsychology*, J. S. Kreutzer, J. DeLuca, and B. Caplan, Eds., Springer New York, 2011, pp. 681–683. DOI: 10.1007/978-0-387-79948-3_1535.

[89] A. Sadeh, O. Dan, and Y. Bar-Haim, "Online assessment of sustained attention following sleep restriction," *Sleep Medicine*, vol. 12, no. 3, pp. 257–261, 2011, ISSN: 13899457. DOI: 10.1016/j.sleep.2010.02.001.

[90] N. H. Mackworth, "The Breakdown of Vigilance during Prolonged Visual Search," *Quarterly Journal of Experimental Psychology*, vol. 1, no. 1, pp. 6–21, 1948, ISSN: 0033-555X. DOI: 10.1080/17470214808416738.

[91] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, "'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects," *Neuropsychologia*, vol. 35, no. 6, pp. 747–758, 1997, ISSN: 00283932. DOI: 10.1016/S0028-3932(97)00015-8. arXiv: NIHMS150003.

[92] C. Fassbender, K. Murphy, J. Foxe, G. Wylie, D. Javitt, I. Robertson, and H. Garavan, "A topography of executive functions and their interactions revealed by functional magnetic resonance imaging," *Cognitive Brain Research*, vol. 20, no. 2, pp. 132–143, 2004, ISSN: 09266410. DOI: 10.1016/j.cogbrainres.2004.02.007.

[93] P. M. Dockree, S. P. Kelly, I. H. Robertson, R. B. Reilly, and J. J. Foxe, "Neurophysiological markers of alert responding during goal-directed behavior: A high-density electrical mapping study," *NeuroImage*, vol. 27, no. 3, pp. 587–601, 2005, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2005.05.044.

[94] A. Gale, "Some EEG correlates of of sustained attention," in *Vigilance*, Boston, MA.: Springer, 1977, pp. 263–283.

[95] H. Laufs, K. Krakow, P. Sterzer, E. Eger, A. Beyerle, A. Salek-Haddadi, and A. Kleinschmidt, "Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest," *Proceedings of the National Academy of Sciences*, vol. 100, no. 19, pp. 11 053–11 058, 2003, ISSN: 0027-8424. DOI: 10.1073/pnas.1831638100.

[96] G. Matthews, J. S. Warm, L. E. Reinerman-Jones, L. K. Langheim, D. A. Washburn, and L. Tripp, "Task Engagement, Cerebral Blood Flow Velocity, and Diagnostic Monitoring for Sustained Attention," *Journal of Experimental Psychology: Applied*, vol. 16, no. 2, pp. 187–203, 2010, ISSN: 1076898X. DOI: 10.1037/a0019572.

[97] T. O. Zander, C. Kothe, S. Welke, and M. Roetting, "Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5638 LNAI, 2009, pp. 759–771, ISBN: 364202811X. DOI: 10.1007/978-3-642-02812-0_86. [Online]. Available: http://link.springer.com/10.1007/978-3-642-02812-0{_}86.

[98] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general," *Journal of Neural Engineering*, vol. 8, no. 2, 2011, ISSN: 17412560. DOI: 10.1088/1741-2560/8/2/025005. arXiv: 1741-2560/8/2/025005 [10.1088].

[99] M. Grosse-Wentrup and B. Schlkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces," *Journal of Neural Engineering*, vol. 9, no. 4, 2012, ISSN: 17412560. DOI: 10.1088/1741-2560/9/4/046001.

[100] J. Schumacher, C. Jeunet, F. Lotte, J. Schumacher, C. Jeunet, F. Lotte, T. Explanatory, and J. Schumacher, "Towards Explanatory Feedback for User Training in Brain – Computer Interfaces," pp. 3169–3174, 2015. DOI: 10.1109/SMC.2015.550.

[101] A. O. Argunsah and M. Cetin, "AR-PCA-HMM approach for sensorimotor task classification in EEG-based brain-computer interfaces," *Proceedings - International Conference on Pattern Recognition*, pp. 113–116, 2010, ISSN: 10514651. DOI: 10.1109/ICPR.2010.36.

[102] J. F. Delgado Saa and M. Cetin, "Discriminative Methods for Classification of Asynchronous Imaginary Motor Tasks From EEG Data," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 5, pp. 716–724, 2013, ISSN: 1534-4320. DOI: 10.1109/TNSRE.2013.2268194.

[103] J. Saa and M Cetin, "Modeling Differences in the Time-Frequency presentation of EEG Signals Through HMMs for classification of Imaginary Motor Tasks," *Sabanci University*, 2011.

[104] J. F. D. Saa and M. Cetin, "Hidden conditional random fields for classification of imaginary motor tasks from EEG data," in *Signal Processing Conference, 2011 19th European*, 2011, pp. 171–175.

[105] J. F. D. Saa and M. Çetin, "A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026 020, 2012, ISSN: 1741-2560. DOI: 10.1088/1741-2560/9/2/026020.

[106] J. F. Saa, A. D. Pesters, D. McFarland, and M. Çetin, "Word-level language modeling for P300 spellers based on discriminative graphical models," *Journal of Neural Engineering*, vol. 12, no. 2, p. 26 007, 2015, ISSN: 17412552. DOI: 10.1088/1741-2560/12/2/026007.

[107] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," *Information Sciences*, vol. 180, no. 10, pp. 1942–1954, 2010, ISSN: 00200255. DOI: 10.1016/j.ins.2010.01.011.

[108] C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, "A Multi-Modal Driver Fatigue and Distraction Assessment System," *International Journal of Intelligent Transportation Systems Research*, vol. 14, no. 3, pp. 173–194, 2016, ISSN: 1348-8503. DOI: 10.1007/s13177-015-0112-9.

[109] P. Arico, "Mental States Monitoring through Passive Brain-Computer Interface systems," Ph.D. dissertation, University of Bologna, 2014.

[110] J. Liu, C. Zhang, and C. Zheng, "EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters," *Biomedical Signal Processing and Control*, vol. 5, no. 2, pp. 124–130, 2010, ISSN: 17468094. DOI: 10.1016/j.bspc.2010.01.001.

[111] J.-N. Gu, H.-T. Lu, and B.-L. Lu, "An integrated Gaussian mixture model to estimate vigilance level based on EEG recordings," *Neurocomputing*, vol. 129, pp. 107–113, 2014, ISSN: 09252312. DOI: 10.1016/j.neucom.2012.10.042.

[112] H. Sun and B. L. Lu, "EEG-based fatigue classification by using parallel hidden Markov model and pattern classifier combination," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7666 LNCS, no. PART 4, pp. 484–491, 2012, ISSN: 03029743. DOI: 10.1007/978-3-642-34478-7_59.

[113] T. Baltrušaitis, P. Robinson, and L. P. Morency, "Continuous conditional neural fields for structured regression," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, pp. 593–608, 2014, ISSN: 16113349. DOI: 10.1007/978-3-319-10593-2_39.

[114] M. Andujar and J. Gilbert, "Let's learn!: enhancing user's engagement levels through passive brain-computer interfaces," *CHI'13 Extended Abstracts on Human Factors in*, pp. 703–708, 2013, ISSN: 9781450319522. DOI: 10.1145/2468356.2468480.

[115] C. E. Lakey, D. R. Berry, and E. W. Sellers, "Manipulating Attention via Mindfulness Induction Improves P300-based Brain-Computer Interface Performance," *Journal of Neural Engineering*, vol. 8, no. 2, pp. 139–148, 2011, ISSN: 1527-5418. DOI: 10.1038/nmeth.2839.A. arXiv: 15334406.

[116] O. W. Klaproth, M. Halbrügge, L. R. Krol, C. Vernaleken, T. O. Zander, and N. Russwinkel, "A Neuroadaptive Cognitive Model for Dealing With Uncertainty in Tracing Pilots' Cognitive State," *Topics in Cognitive Science*, vol. 12, pp. 1012–1029, 2020, ISSN: 17568765. DOI: 10.1111/tops.12515.

[117] R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee, and J. A. Rushby, "EEG differences between eyes-closed and eyes-open resting conditions," *Clinical Neurophysiology*, vol. 118, pp. 2765–2773, 2007. DOI: 10.1016/j.clinph.2007.07.028.

[118] M. W. Cole, D. S. Bassett, J. D. Power, T. S. Braver, and S. E. Petersen, "Intrinsic and task-evoked network architectures of the human brain," *Neuron*, vol. 83, no. 1, pp. 238–251, 2014, ISSN: 15378276. DOI: 10.1158/0008-5472.CAN-10-4002.BONE. arXiv: NIHMS150003.

[119] M. W. Cole, T. Ito, D. S. Bassett, and D. H. Schultz, "Activity flow over resting-state networks shapes cognitive task activations," *Nature Neuroscience*, vol. 19, no. 12, pp. 1718–1726, 2016, ISSN: 15461726. DOI: 10.1038/nn.4406.

[120] M. Lamoš, R. Mareček, T. Slavíček, M. Mikl, I. Rektor, and J. Jan, "Spatial-temporal-spectral EEG patterns of BOLD functional network connectivity dynamics," *Journal of Neural Engineering*, vol. 15, no. 3, 2018, ISSN: 17412552. DOI: 10.1088/1741-2552/aab66b.

[121] P. Núñez, J. Poza, C. Gomez, V. Barroso-García, A. Maturana-Candelas, M. Á. Tola-Arribas, M. Cano, and R. Hornero, "Characterization of the dynamic behavior of neural activity in Alzheimer's disease: exploring the non-stationarity and recurrence structure of EEG resting-state activity," *Journal of Neural Engineering*, 2020, ISSN: 1741-2560. DOI: 10.1088/1741-2552/ab71e9.

[122] J. L. Reichert, S. E. Kober, C. Neuper, and G. Wood, "Resting-state sensorimotor rhythm (SMR) power predicts the ability to up-regulate SMR in an EEG-instrumental conditioning paradigm," *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2068–2077, 2015, ISSN: 18728952. DOI: 10.1016/j.clinph.2014.09.032.

[123] O. Ozdenizci, M. Yalcin, A. Erdogan, P. Volkan, M. Grosse-Wentrup, and M. Cetin, "Electroencephalographic identifiers of motor adaptation learning," *Journal of Neural Engineering*, vol. 14, no. 4, p. 046 027, 2017. DOI: 10.1088/1741-2552/aa6abd.

[124] T. M. Madhyastha, M. K. Askren, P. Boord, and T. J. Grabowski, "Dynamic Connectivity at Rest Predicts Attention Task Performance," *Brain Connectivity*, vol. 5, no. 1, pp. 45–59, 2015, ISSN: 2158-0014. DOI: 10.1089/brain.2014.0248.

[125] M. Irrmischer, S.-S. Poil, H. D. Mansvelder, F. S. Intra, and K. Linkenkaer-Hansen, "Strong long-range temporal correlations of beta/gamma oscillations are associated with poor sustained visual attention performance," *European Journal of Neuroscience*, vol. 48, no. 8, pp. 2674–2683, 2018, ISSN: 0953816X. DOI: 10.1111/ejn.13672.

[126] D. Karamacoska, R. J. Barry, G. Z. Steiner, E. P. Coleman, and E. J. Wilson, "Intrinsic EEG and task-related changes in EEG affect Go/NoGo task performance," *International Journal of Psychophysiology*, vol. 125, no. December 2017, pp. 17–28, 2018, ISSN: 18727697. DOI: 10.1016/j.ijpsycho.2018.01.015.

[127] Y. Si, F. Li, K. Duan, Q. Tao, C. Li, Z. Cao, Y. Zhang, B. Biswal, P. Li, D. Yao, and P. Xu, "Predicting individual decision-making responses based on single-trial EEG," *NeuroImage*, vol. 206, 2020, ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.116333.

[128] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring Phase Synchrony in Brain Signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999. DOI: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C.

[129] R. Bruña, F. Maestú, and E. Pereda, "Phase locking value revisited: Teaching new tricks to an old dog," *Journal of Neural Engineering*, vol. 15, no. 5, 2018, ISSN: 17412552. DOI: 10.1088/1741-2552/aacfe4. arXiv: 1710.08037.

[130] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007, ISSN: 10659471. DOI: 10.1002/hbm.20346.

[131] S. Aydore, D. Pantazis, and R. M. Leahy, "A note on the phase locking value and its properties," *NeuroImage*, vol. 74, pp. 231–244, 2013, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.02.008. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.02.008.

[132] S. Aydın, S. Demirtaş, M. A. Tunga, and K. Ateş, "Comparison of hemispheric asymmetry measurements for emotional recordings from controls," *Neural Computing and Applications*, vol. 30, no. 4, pp. 1341–1351, 2018, ISSN: 09410643. DOI: 10.1007/s00521-017-3006-8. [Online]. Available: http://doi.org/10.1007/s00521-017-3006-8.

[133] C. Beauchene, S. Roy, R. Moran, A. Leonessa, and N. Abaid, "Comparing brain connectivity metrics: A didactic tutorial with a toy model and experimental data," *Journal of Neural Engineering*, vol. 15, no. 5, 2018, ISSN: 17412552. DOI: 10.1088/1741-2552/aad96e.

[134] J. P. Van Haneghan and L. Baker, "Congitive Monitoring in Mathemtics," in *Cognitive Strategy Research*, Springer, 1989, pp. 215–238.

[135] P. Arico, G. Borghini, I. Graziani, F. Taya, Y. Sun, A. Bezerianos, N. V. Thakor, F. Cincotti, and F. Babiloni, "Towards a multimodal bioelectrical framework for the online mental workload evaluation," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, no. Figure 1, pp. 3001–3004, 2014, ISSN: 1557170X. DOI: 10.1109/EMBC.2014.6944254.

[136] R. b. Roy, A. Breust, S. Bonnet, J. Porcherot, S. Charbonnier, C. Godin, and A. Campagne, "Influence of workload on auditory evoked potentials in a single-stimulus paradigm," *PhyCS 2015 - 2nd International Conference on Physiological Computing Systems, Proceedings*, pp. 104–111, 2015.

[137] S. Aydin, "Deep Learning Classification of Neuro-Emotional Phase Domain Complexity Levels Induced by Affective Video Film Clips," *IEEE Journal of Biomedical and Health Informatics*, 2019, ISSN: 2168-2194. DOI: 10.1109/jbhi.2019.2959843.

[138] C. Cannard, T. Brandmeyer, H. Wahbeh, and A. Delorme, "Self-health monitoring and wearable neurotechnologies," in *Handbook of Clinical Neurology*, vol. 168, Elsevier, 2020, pp. 207–232, ISBN: 9780444639349. DOI: 10.1016/B978-0-444-63934-9.00016-0.

[139] N. Zhang, W.-L. Zheng, W. Liu, and B.-L. Lu, "Continuous vigilance estimation using LSTM neural networks," in *International Conference on Neural Information Processing*, Springer, 2016, pp. 530–537, ISBN: 978-3-319-46672-9; 978-3-319-46671-2. DOI: 10.1007/978-3-319-46672-9. arXiv: 1602.08225.

[140] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, H. Wang, K. Sgarbas, N. Thakor, A. Bezerianos, and Y. Sun, "Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations between Driving and Vigilance Tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 740–749, 2018, ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2791936.

[141] T. Kustermann, B. Rockstroh, J. Kienle, G. A. Miller, and T. Popov, "Deficient attention modulation of lateralized alpha power in schizophrenia," *Psychophysiology*, vol. 53, no. 6, pp. 776–785, 2016, ISSN: 14698986. DOI: 10.1111/psyp.12626.

[142] K. Jimura and R. A. Poldrack, "Analyses of regional-average activation and multivoxel pattern information tell complementary stories," *Neuropsychologia*, vol. 50, no. 4, pp. 544–552, 2012, ISSN: 00283932. DOI: 10.1016/j.neuropsychologia.2011.11.007.

[143] A. Gaume, G. Dreyfus, and F. B. Vialatte, "A cognitive brain-computer interface monitoring sustained attentional variations during a continuous task," *Cognitive Neurodynamics*, vol. 4, pp. 257–269, 2019. DOI: 10.1007/s11571-019-09521-4.

[144] M. K. Van Schie, R. D. Thijs, R. Fronczek, H. A. Middelkoop, G. J. Lammers, and J. G. Van Dijk, "Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness," *Journal of Sleep Research*, vol. 21, no. 4, pp. 390–395, 2012, ISSN: 09621105. DOI: 10.1111/j.1365-2869.2011.00979.x.

[145] S. Coelli, R. Barbieri, G. Reni, C. Zucca, and A. M. Bianchi, "EEG indices correlate with sustained attention performance in patients affected by diffuse axonal injury," *Medical and Biological Engineering and Computing*, vol. 56, no. 6, pp. 991–1001, 2018, ISSN: 17410444. DOI: 10.1007/s11517-017-1744-5.

[146] F. C. Wilson and T. Manly, "Sustained attention training and errorless learning facilitates self care functioning in chronic ipsilesional neglect following severe traumatic brain injury," *Neuropsychological Rehabilitation*, vol. 13, no. 5, pp. 537–548, 2003, ISSN: 0960-2011. DOI: 10.1080/09602010343000200.

[147] J. Grange and G. Houghton, *Task switching and cognitive control*. Oxford University Press, USA, 2014, ISBN: 9780199921959.

[148] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13 040–13 045, 2009, ISSN: 0027-8424. DOI: 10.1073/pnas.0905267106.

[149] J. Coull, R. Frackowiak, and C. Frith, "Monitoring for target objects: activation of right frontal and parietal cortices with increasing time on task," *Neuropsychologia*, vol. 36, no. 12, pp. 1325–1334, 1998, ISSN: 1044-8632. DOI: 10.1080/10448632.2014.930608.

[150] W. Sturm, A. De Simone, B. J. Krause, K. Specht, V. Hesselmann, I. Radermacher, H. Herzog, L. Tellmann, H. W. Müller-Gärtner, and K. Willmes, "Functional anatomy of intrinsic alertness: Evidence for a fronto-parietal-thalamic-brainstem network in the right hemisphere," *Neuropsychologia*, vol. 37, no. 7, pp. 797–805, 1999, ISSN: 00283932. DOI: 10.1016/S0028-3932(98)00141-9.

[151] Y. Sun, J. Lim, J. Meng, K. Kwok, N. Thakor, and A. Bezerianos, "Discriminative Analysis of Brain Functional Connectivity Patterns for Mental Fatigue Classification," *Annals of Biomedical Engineering*, vol. 42, no. 10, pp. 2084–2094, 2014, ISSN: 15739686. DOI: 10.1007/s10439-014-1059-8.

[152] J. Tallet, J. Barral, and C.-A. Hauert, "Electro-cortical correlates of motor inhibition: A comparison between selective and non-selective stop tasks," *Brain Research*, pp. 68–76, 2009. DOI: 10.1016/j.brainres.2009.05.058.

[153] M. S. Clayton, N. Yeung, and R. Cohen Kadosh, "The roles of cortical oscillations in sustained attention," *Trends in Cognitive Sciences*, vol. 19, no. 4, pp. 188–195, 2015, ISSN: 1879307X. DOI: 10.1016/j.tics.2015.02.004.

[154] R. G. O'Connell, P. M. Dockree, M. A. Bellgrove, A. Turin, S. Ward, J. J. Foxe, and I. H. Robertson, "Two Types of Action Error: Electrophysiological Evidence for Separable Inhibitory and Sustained Attention Neural Mechanisms Producing Error on Go/No-go Tasks," *Journal of Cognitive Neuroscience*, vol. 21, no. 1, pp. 93–104, 2009, ISSN: 0898-929X. DOI: 10.1162/jocn.2009.21008.

[155] F. C. Fortenbaugh, J. Degutis, and M. Esterman, "Recent theoretical, neural, and clinical advances in sustained attention research," *Ann N Y Acad Sci.*, vol. 1396, no. 1, pp. 70–91, 2017. DOI: 10.1111/nyas.13318.Recent.

[156] 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, no. September 2016, pp. 474–482, 2017, ISSN: 18728286. DOI: 10.1016/j.neucom.2016.09.057.

[157] H. Wang, X. Liu, J. Li, T. Xu, A. Bezerianos, Y. Sun, and F. Wan, "Driving Fatigue Recognition with Functional Connectivity Based on Phase Synchronization," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2020, ISSN: 2379-8920. DOI: 10.1109/tcds.2020.2985539.

[158] W.-L. Zheng, K. Gao, G. Li, W. Liu, C. Liu, J.-Q. Liu, G. Wang, and B.-L. Lu, "Vigilance Estimation Using a Wearable EOG Device in Real Driving Environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, pp. 1–15, 2019, ISSN: 1524-9050. DOI: 10.1109/TITS.2018.2889962.

[159] C. Braboszcz and A. Delorme, "Lost in thoughts: Neural markers of low alertness during mind wandering," *NeuroImage*, vol. 54, no. 4, pp. 3040–3047, 2011, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.10.008.

[160] V. R. Ribas, R. G. Ribas, J. d. A. N. Nobrega, M. V. Da, J. A. d. A. Especie, M. T. Calafange, C. d. O. M. Calafange, and H. A. d. L. Martins, "Pattern of anxiety, insecurity, fear, panic and/or phobia observed by quantitative electroencephalography (QEEG)," *Dementia & Neuropsychologia*, vol. 12, no. 3, pp. 264–271, 2018.

[161] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computationomputation*, vol. 7, no. 6, pp. 1129–1159, 1995, ISSN: 0899-7667. DOI: doi:10.1162/neco.1995.7.6.1129. arXiv: arXiv:1511.06440v1.

[162] A. Delorme and S. Makeig, "EEGLAB: an open sorce toolbox for analysis of single-trail EEG dynamics including independent component anlaysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2003.10.009. arXiv: j.jneumeth.2003.10.009. [10.1016].

[163] J. Bhattacharya and H. Petsche, "Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise," *Signal Processing*, vol. 85, no. 11, pp. 2161–2177, 2005, ISSN: 01651684. DOI: 10.1016/j.sigpro.2005.07.007.

[164] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015, ISSN: 19430604. DOI: 10.1109/TAMD.2015.2431497.

[165] D. P. Kingma and J. L. Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," *arXiv preprint arXiv:1412.6980*, vol. 9, 2015. arXiv: 1412.6980.

[166] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the royal statistical society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[167] G. S. Larue, A. Rakotonirainy, and A. N. Pettitt, "Real-time performance modelling of a sustained attention to response task," *Ergonomics*, vol. 53, no. 10, pp. 1205–1216, 2010, ISSN: 00140139. DOI: 10.1080/00140139.2010.512984.

[168] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Processing: Image Communication*, vol. 47, pp. 549–555, 2016, ISSN: 09235965. DOI: 10.1016/j.image.2016.05.018.

[169] D. F. Hermens, E. X. C. Soei, S. D. Clarke, M. R. Kohn, E. Gordon, and L. M. Williams, "Resting EEG theta activity predicts cognitive performance in attention-deficit hyperactivity disorder," *Pediatric Neurology*, vol. 32, no. 4, pp. 248–456, 2005, ISSN: 08878994. DOI: 10.1016/j.pediatrneurol.2004.11.009.

[170] A. Behzadnia, M. Ghoshuni, and S. A. Chermahini, "EEG Activities and the Sustained Attention Performance," *Neurophysiology*, vol. 49, no. 3, pp. 226–233, 2017, ISSN: 15739007. DOI: 10.1007/s11062-017-9675-1.

[171] R. S. Huang, T. P. Jung, and S. Makeig, "Tonic changes in EEG power spectra during simulated driving," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5638 LNAI, no. July, pp. 394–403, 2009, ISSN: 03029743. DOI: 10.1007/978-3-642-02812-0_47.

[172] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, no. 4700, pp. 750–752, 1985, ISSN: 00368075. DOI: 10.1126/science.3992243.

[173] S. K. Loo, T. S. Hale, J. Macion, G. Hanada, J. J. McGough, J. T. McCracken, and S. L. Smalley, "Cortical activity patterns in ADHD during arousal, activation and sustained attention," *Neuropsychologia*, vol. 47, no. 10, pp. 2114–2119, 2009, ISSN: 00283932. DOI: 10.1016/j.neuropsychologia.2009.04.013.

[174] A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, and R. J. Davidson, "Long-term meditators self-induce high-amplitude gamma synchrony during mental practice," *Proceedings of the national Academy of Sciences*, vol. 101, no. 46, pp. 16 369–16 373, 2004. DOI: 10.1073/pnas.0407401101.

[175] O. Jensen, J. Kaiser, and J. P. Lachaux, "Human gamma-frequency oscillations associated with attention and memory," *Trends in Neurosciences*, vol. 30, no. 7, pp. 317–324, 2007, ISSN: 01662236. DOI: 10.1016/j.tins.2007.05.001.

[176] P. J. Fitzgerald and B. O. Watson, "Gamma oscillations as a biomarker for major depression: an emerging topic," *Translational Psychiatry*, vol. 8, no. 1, 2018, ISSN: 21583188. DOI: 10.1038/s41398-018-0239-y.

[177] P. Eskandari and A. Erfanian, "Improving the performance of brain-computer interface through meditation practicing," *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 662–665, 2008, ISSN: 1557-170X. DOI: 10.1109/IEMBS.2008.4649239.

[178] J. P. Liu, C. Zhang, and C. X. Zheng, "Estimation of the cortical functional connectivity by directed transfer function during mental fatigue," *Applied Ergonomics*, vol. 42, no. 1, pp. 114–121, 2010, ISSN: 00036870. DOI: 10.1016/j.apergo.2010.05.008.

[179] J. Li, J. Lim, Y. Chen, K. Wong, N. Thakor, A. Bezerianos, and Y. Sun, "Mid-Task Break Improves Global Integration of Functional Connectivity in Lower Alpha Band," *Frontiers in Human Neuroscience*, vol. 10, no. June, pp. 1–12, 2016, ISSN: 1662-5161. DOI: 10.3389/fnhum.2016.00304.

[180] A. L. Manuel, A. G. Guggisberg, R. Thézé, F. Turri, and A. Schnider, "Resting-state connectivity predicts visuo-motor skill learning," *NeuroImage*, vol. 176, no. May, pp. 446–453, 2018, ISSN: 10959572. DOI: 10.1016/j.neuroimage.2018.05.003.

[181] T. Ito, K. R. Kulkarni, D. H. Schultz, R. D. Mill, R. H. Chen, L. I. Solomyak, and M. W. Cole, "Cognitive task information is transferred between brain regions via resting-state network topology," *Nature Communications*, vol. 8, no. 1, 2017, ISSN: 20411723. DOI: 10.1038/s41467-017-01000-w.

[182] J. W. Choi, K. Y. Jung, C. H. Kim, and K. H. Kim, "Changes in gamma- and theta-band phase synchronization patterns due to the difficulty of auditory oddball task," *Neuroscience Letters*, vol. 468, no. 2, pp. 156–160, 2010, ISSN: 03043940. DOI: 10.1016/j.neulet.2009.10.088.

[183] M. Siegel, T. H. Donner, and A. K. Engel, "Spectral fingerprints of large-scale neuronal interactions," *Nature Reviews Neuroscience*, vol. 13, no. 2, pp. 121–134, 2012, ISSN: 1471003X. DOI: 10.1038/nrn3137.

[184] C. Zhang, X. Yu, Y. Yang, and L. Xu, "Phase synchronization and spectral coherence analysis of EEG activity during mental fatigue," *Clinical EEG and Neuroscience*, vol. 45, no. 4, pp. 249–256, 2014, ISSN: 21695202. DOI: 10.1177/1550059413503961.

[185] S. Xie and Y. Li, "EEG effective connectivity networks for an attentive task requiring vigilance based on dynamic partial directed coherence," *Journal of Integrative Neuroscience*, vol. 19, no. 1, pp. 111–118, 2020, ISSN: 1757448X. DOI: 10.31083/J.JIN.2020.01.1234.

[186] H. Hart, L. Lim, M. A. Mehta, A. Chatzieffraimidou, C. Curtis, X. Xu, G. Breen, A. Simmons, K. Mirza, and K. Rubia, "Reduced functional connectivity of fronto-parietal sustained attention networks in severe childhood abuse," *PLoS ONE*, vol. 12, no. 11, pp. 1–17, 2017, ISSN: 19326203. DOI: 10.1371/journal.pone.0188744.

[187] J. Li, A. Bezerianos, and N. Thakor, "Cognitive State Analysis, Understanding, and Decoding from the Perspective of Brain Connectivity," pp. 1–37, 2020. arXiv: 2005.12191. [Online]. Available: http://arxiv.org/abs/2005.12191.

[188] I. M. Comsa, T. A. Bekinschtein, and S. Chennu, "Transient Topographical Dynamics of the Electroencephalogram Predict Brain Connectivity and Behavioural Responsiveness During Drowsiness," *Brain Topography*, vol. 32, no. 2, pp. 315–331, 2019, ISSN: 15736792. DOI: 10.1007/s10548-018-0689-9.

[189] L. J. Trejo, K. Kubitz, R. Rosipal, R. L. Kochavi, and L. D. Montgomery, "EEG-Based Estimation and Classification of Mental Fatigue," *Psychology*, vol. 06, no. 05, pp. 572–589, 2015, ISSN: 2152-7180. DOI: 10.4236/psych.2015.65055.

[190] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, M. Matsuo, and H. Kadotani, "Heart Rate Variability-Based Driver Drowsiness Detection and Its Validation With EEG," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1769–1778, 2019, ISSN: 15582531. DOI: 10.1109/TBME.2018.2879346.

[191] A. B. Buriro, R. Shoorangiz, S. J. Weddell, and R. D. Jones, "Predicting Microsleep States Using EEG Inter-Channel Relationships," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 12, pp. 2260–2269, 2018, ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2878587.

[192] R.-S. Huang, T.-P. Jung, and S. Makeig, "Multi-Scale EEG Brain Dynamics During Sustained Attention Tasks," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 4, IEEE, 2007, pp. IV–1173–IV–1176, ISBN: 1-4244-0727-3. DOI: 10.1109/ICASSP.2007.367284.

[193] R. Shoorangiz, A. B. Buriro, S. J. Weddell, and R. D. Jones, "Detection and Prediction of Microsleeps from EEG using Spatio-Temporal Patterns," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2019, pp. 522–525, ISBN: 9781538613115. DOI: 10.1109/EMBC.2019.8857962.

[194] M. Awais, N. Badruddin, and M. Drieberg, "EEG Brain Connectivity Analysis to Detect Driver Drowsiness Using Coherence," *Proceedings - 2017 International Conference on Frontiers of Information Technology, FIT 2017*, vol. 2017-Janua, pp. 110–114, 2018. DOI: 10.1109/FIT.2017.00027.

[195] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-Temporal Recurrent Neural Network for Emotion Recognition," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 839–847, 2017, ISSN: 21682267. DOI: 10.1109/TCYB.2017.2788081. arXiv: 1705.04515.

[196] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-Basaed Person Identification Using the Deep Learning Approach," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–10, 2019. arXiv: 1807.03147.

[197] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning Spatial-Spectral-Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, 2019, ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2884641.

[198] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance," *Journal of Neuroscience Methods*, vol. 235, pp. 138–144, 2014, ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2014.06.011.

[199] F. Fahimi, Z. Zhang, W. B. Goh, T. S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *Journal of Neural Engineering*, vol. 16, no. 2, 2019, ISSN: 17412552. DOI: 10.1088/1741-2552/aaf3f6.

[200] R. Fu, H. Wang, and W. Zhao, "Dynamic driver fatigue detection using hidden Markov model in real driving condition," *Expert Systems with Applications*, vol. 63, pp. 397–411, 2016, ISSN: 09574174. DOI: 10.1016/j.eswa.2016.06.042.

[201] J. Chen, H. Wang, Q. Wang, and C. Hua, "Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males," *Neuropsychologia*, vol. 129, no. April, pp. 200–211, 2019, ISSN: 00283932. DOI: 10.1016/j.neuropsychologia.2019.04.004.

[202] Y. Deng, Y. Jiao, and B.-L. Lu, "Driver Sleepiness Detection Using LSTM Neural Network," in *International Conference on Neural Information Processing*, Springer International Publishing, 2018, pp. 622–633, ISBN: 9783030042127. DOI: 10.1299/jsmemag.90.823_758.

[203] L.-c. Shi, Y.-y. Jiao, and B.-L. Lu, "Differential Entropy Feature for EEG-based Vigilance Estimation," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, IEEE, 2013, pp. 6627–6630, ISBN: 9781467319690.

[204] Y. Cui and D. Wu, "EEG-Based Driver Drowsiness Estimation Using Convolutional Neural Networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10635 LNCS, pp. 822–832, 2017, ISSN: 16113349. DOI: 10.1007/978-3-319-70096-0_84. arXiv: arXiv:1809.00929v1.

[205] H. Li, W.-l. Zheng, and B.-l. Lu, "Multimodal Vigilance Estimation with Adversarial Domain Adaptation Networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2018, pp. 1–6.

[206] W. Wu, Q. M. J. Wu, W. Sun, Y. Yang, X. Yuan, W.-L. Zheng, and B.-L. Lu, "A regression method with subnetwork neurons for vigilance estimation using EOG and EEG," *IEEE Transactions on Cognitive and Developmental Systems*, vol. PP, no. X, pp. 1–1, 2018, ISSN: 2379-8920. DOI: 10.1109/TCDS.2018.2889223.

[207] S. M. Golden, Charles J and Freshwater, "Stroop Color and Word Test," 1978.

[208] R. Chavarriaga, M. Ušćumlić, H. Zhang, Z. Khaliliardali, R. Aydarkhanov, S. Saeedi, L. Gheorghe, and J. D. R. Millán, "Decoding Neural Correlates of Cognitive States to Enhance Driving Experience," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 4, pp. 288–297, 2018, ISSN: 2471285X. DOI: 10.1109/TETCI.2018.2848289.

[209] R. P. Balandong, R. F. Ahmad, M. N. Mohamad Saad, and A. S. Malik, "A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver," *IEEE Access*, vol. 6, pp. 22 908–22 919, 2018, ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2811723.

[210] J. Wei, Y. Ke, C. Sun, X. An, H. Qi, D. Ming, and P. Zhou, "The timing of theta phase synchronization accords with vigilant attention," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2442–2445, 2017, ISSN: 1557170X. DOI: 10.1109/EMBC.2017.8037350.

[211] A. L. Proskovec, A. I. Wiesman, E. Heinrichs-Graham, and T. W. Wilson, "Beta Oscillatory Dynamics in the Prefrontal and Superior Temporal Cortices Predict Spatial Working Memory Performance," *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018, ISSN: 20452322. DOI: 10.1038/s41598-018-26863-x.

[212] H. Cramer, *Mathematical Methods of Statistics*. Princeton U. Press, Princeton, 1946. DOI: 10.1214/aoms/1177730503.

[213] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051 001, 2019, ISSN: 17412552. DOI: 10.1088/1741-2552/ab260c.

[214] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological Psychology*, vol. 40, no. 1-2, pp. 187–195, 1995, ISSN: 03010511. DOI: 10.1016/0301-0511(95)05116-3.

[215] N. H. Liu, C. Y. Chiang, and H. C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors," *Sensors (Switzerland)*, vol. 13, no. 8, pp. 10 273–10 286, 2013, ISSN: 14248220. DOI: 10.3390/s130810273.

[216] S. Coelli, R. Sclocco, R. Barbieri, G. Reni, C. Zucca, and A. M. Bianchi, "EEG-based index for engagement level monitoring during sustained attention," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 1512–1515, ISBN: 9781424492701.

[217] M. Arns, C. K. Conners, and H. C. Kraemer, "A Decade of EEG Theta/Beta Ratio Research in ADHD: A Meta-Analysis," *Journal of Attention Disorders*, vol. 17, no. 5, pp. 374–383, 2013, ISSN: 10870547. DOI: 10.1177/1087054712460087.

[218] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *ICML '13*, vol. 28, p. 6, 2013.

[219] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, pp. 1–21, 2015, ISSN: 19326203. DOI: 10.1371/journal.pone.0118432.

[220] J. D. Gibbons and S. Chakraborti, *Nonparametric statistical inference*. Springer, 2011.

[221] H. Atasoy and E. Yildirim, "Classification of Verbal and Quantitative Mental Tasks Using Phase Locking Values between EEG Signals," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 7, pp. 383–390, 2016.

[222] X. Hong, Y. Liu, J. Sun, and S. Tong, "Age-related differences in the modulation of small-world brain networks during a Go/NoGo task," *Frontiers in Aging Neuroscience*, vol. 8, no. MAY, pp. 1–15, 2016, ISSN: 16634365. DOI: 10.3389/fnagi.2016.00100.

[223] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 770–778, 2016, ISSN: 10636919. DOI: 10.1109/CVPR.2016.90. arXiv: arXiv:1512.03385v1.

[224] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cognitive Neurodynamics*, vol. 12, no. 6, pp. 597–606, 2018, ISSN: 18714099. DOI: 10.1007/s11571-018-9496-y.

[225] Y. Kawahara and M. Sugiyama, "Sequential change-point detection based on direct density-ratio estimation," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 5, pp. 114–127, 2012. DOI: https://doi.org/10.1002/sam.10124.

[226] S. Aminikhanghahi and D. D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, 2017, ISSN: 02193116. DOI: 10.1007/s10115-016-0987-z.

[227] S. Thies and P. Molnár, "Bayesian change point analysis of Bitcoin returns," *Finance Research Letters*, vol. 27, no. February, pp. 223–227, 2018, ISSN: 15446123. DOI: 10.1016/j.frl.2018.03.018. [Online]. Available: https://doi.org/10.1016/j.frl.2018.03.018.

[228] M. F. Chowdhury, S. A. Selouani, and D. O'Shaughnessy, "Bayesian on-line spectral change point detection: A soft computing approach for on-line ASR," *International Journal of Speech Technology*, vol. 15, no. 1, pp. 5–23, 2012, ISSN: 13812416. DOI: 10.1007/s10772-011-9116-2.

[229] J. Zheng, H. Fushing, and L. Ge, "A Data-driven Approach to Predict and Classify Epileptic Seizures from Brain-wide Calcium Imaging Video Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5963, no. c, pp. 1–1, 2019, ISSN: 1545-5963. DOI: 10.1109/tcbb.2019.2895077.

[230] G. Chen, G. Lu, W. Shang, and Z. Xie, "Automated Change-Point Detection of EEG Signals Based on Structural Time-Series Analysis," *IEEE Access*, vol. 7, pp. 180 168–180 180, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2956768.

[231] M. Sinn, A. Ghodsi, and K. Keller, "Detecting change-points in time series by maximum mean discrepancy of ordinal pattern distributions," *Uncertainty in Artificial Intelligence - Proceedings of the 28th Conference, UAI 2012*, no. October 2012, pp. 786–794, 2012.

[232] A. Hoover, A. Singh, S. Fishel-Brown, and E. Muth, "Real-time detection of workload changes using heart rate variability," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 333–341, 2012, ISSN: 17468094. DOI: 10.1016/j.bspc.2011.07.004. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2011.07.004.

[233] M. Shauly-Aharonov and O. Barenholz-Goultschin, "Real-Time Change-Point Detection Algorithm with an Application to Glycemic Control for Diabetic Pregnant Women," *Methodology and Computing in Applied Probability*, vol. 21, no. 3, pp. 931–944, 2019, ISSN: 15737713. DOI: 10.1007/s11009-019-09716-6.

[234] S. Aminikhanghahi and D. J. Cook, "Enhancing activity recognition using CPD-based activity segmentation," *Pervasive and Mobile Computing*, vol. 53, pp. 75–89, 2019, ISSN: 15741192. DOI: 10.1016/j.pmcj.2019.01.004. [Online]. Available: https://doi.org/10.1016/j.pmcj.2019.01.004.

[235] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction — ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10, pp. 618–630, 2018, ISSN: 17594766. DOI: 10.1038/s41582-018-0055-2. [Online]. Available: http://dx.doi.org/10.1038/s41582-018-0055-2.

[236] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953, ISSN: 00219606. DOI: 10.1063/1.1699114.

[237] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, p. 97, 1970, ISSN: 00063444. DOI: 10.2307/2334940.

[238] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984, ISSN: 01628828. DOI: 10.1109/TPAMI.1984.4767596.

[239] X. Xuan and K. Murphy, "Modeling changing dependency structure in multivariate time series," *ACM International Conference Proceeding Series*, vol. 227, pp. 1055–1062, 2007. DOI: 10.1145/1273496.1273629.

[240] R. P. Adams and D. J. C. MacKay, "Bayesian Online Changepoint Detection," *arXiv preprint arXiv:0710.3742*, 2007. arXiv: 0710.3742.

[241] P. Fearnhead and Z. Liu, "On-line inference for multiple changepoint problems," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 69, no. 4, pp. 589–605, 2007, ISSN: 13697412. DOI: 10.1111/j.1467-9868.2007.00601.x.

[242] F. Sadia, S. Boyd, and J. M. Keith, "Bayesian change-point modeling with segmented ARMA model," *PLoS ONE*, vol. 13, no. 12, pp. 1–23, 2018, ISSN: 19326203. DOI: 10.1371/journal.pone.0208927.

[243] H. Shappell, B. S. Caffo, J. J. Pekar, and M. A. Lindquist, "Improved state change estimation in dynamic functional connectivity using hidden semi-Markov models," *NeuroImage*, vol. 191, no. October 2018, pp. 243–257, 2019, ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.02.013. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2019.02.013.

[244] P. Fearnhead and Z. Liu, "Efficient Bayesian analysis of multiple changepoint models with dependence across segments," *Statistics and Computing*, vol. 21, no. 2, pp. 217–229, 2011, ISSN: 09603174. DOI: 10.1007/s11222-009-9163-6. arXiv: arXiv:0910.3099v1.

[245] G. J. J. van den Burg and C. K. I. Williams, "An Evaluation of Change Point Detection Algorithms," pp. 1–33, 2020. arXiv: 2003.06222. [Online]. Available: http://arxiv.org/abs/2003.06222.

[246] R. Cmejla and P. Sovka, "Recursive Bayesian autoregressive changepoint detector for sequential signal segmentation," in *2004 12th European Signal Processing Conference*, IEEE, 2004, pp. 245–248, ISBN: 9783200001657.

[247] P. Fearnhead, "Exact and efficient Bayesian inference for multiple changepoint problems," *Statistics and Computing*, vol. 16, no. 2, pp. 203–213, 2006, ISSN: 09603174. DOI: 10.1007/s11222-006-8450-8.

[248] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J. F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018, ISSN: 17468108. DOI: 10.1016/j.bspc.2017.12.001.

[249] A. H. Song, L. Chlon, H. Soulat, J. Tauber, S. Subramanian, D. Ba, and M. J. Prerau, "Multitaper Infinite Hidden Markov Model for EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2019, pp. 5803–5807, ISBN: 9781538613115. DOI: 10.1109/EMBC.2019.8856817.

[250] C. C. Hsu and C. W. Lin, "CNN-Based joint clustering and representation learning with feature drift compensation for large-scale image data," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 421–429, 2018, ISSN: 15209210. DOI: 10.1109/TMM.2017.2745702.

[251] Z. Gao, G. Lu, P. Yan, C. Lyu, X. Li, W. Shang, Z. Xie, and W. Zhang, "Automatic change detection for real-time monitoring of EEG signals," *Frontiers in Physiology*, vol. 9, no. APR, 2018, ISSN: 1664042X. DOI: 10.3389/fphys.2018.00325.

[252] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012, ISSN: 01621459. DOI: 10.1080/01621459.2012.737745. arXiv: arXiv:1101.1438v3.

[253] Y. Guo, X. Wang, Q. Xu, F. Liu, Y. Liu, and Y. Xia, "Change-point analysis of eye movement characteristics for female drivers in anxiety," *International Journal of Environmental Research and Public Health*, vol. 16, no. 7, 2019, ISSN: 16604601. DOI: 10.3390/ijerph16071236.

[254] M. Dai, Z. Zhang, and A. Srivastava, "Discovering common change-point patterns in functional connectivity across subjects," *Medical Image Analysis*, vol. 58, pp. 1–38, 2019, ISSN: 13618423. DOI: 10.1016/j.media.2019.101532. arXiv: 1904.12023.

[255] B. Liu, "Bayesian Methods in Brain Connectivity Change Point Detection with EEG Data and Genetic Algorithm," *Mathematics Dissertations*, 2017. [Online]. Available: https://scholarworks.gsu.edu/math{_}diss/47.

[256] G. Hirsch, M. Dirodi, R. Xu, P. Reitner, and C. Guger, "Online Classification of Motor Imagery Using EEG and fNIRS: A Hybrid Approach with Real Time Human-Computer Interaction," in *International Conference on Human-Computer Interaction*, Springer, 2020, pp. 231–238. DOI: https://doi.org/10.1007/978-3-030-50726-8_30.

[257] H. Yuan and B. He, "Brain – Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1425–1435, 2014. DOI: 10.1109/TBME.2014.2312397.

[258] U. Talukdar, S. M. Hazarika, and J. Q. Gan, "Adaptive feature extraction in EEG-based motor imagery BCI: Tracking mental fatigue," *Journal of Neural Engineering*, vol. 17, no. 1, 2020, ISSN: 17412552. DOI: 10.1088/1741-2552/ab53f1.

[259] A. Jafarifarmand, M. A. Badamchizadeh, S. Khanmohammadi, M. A. Nazari, and B. M. Tazehkand, "A New Self-Regulated Neuro-Fuzzy Framework for Classification of EEG Signals in Motor Imagery BCI," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1485–1497, 2018, ISSN: 10636706. DOI: 10.1109/TFUZZ.2017.2728521.

[260] S. Sadaghiani, M. D'Esposito, R. Scheeringa, K. Lehongre, B. Morillon, A.-L. Giraud, and A. Kleinschmidt, "Alpha-band phase synchrony is related to activity in the fronto-parietal adaptive control network," *Journal of Neuroscience*, vol. 32, no. 41, pp. 14 305–14 310, 2012, ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1358-12.2012. arXiv: NIHMS150003.

[261] W. D. Penny and S. J. Roberts, "Dynamic models for nonstationary signal segmentation," *Computers and Biomedical Research*, vol. 32, no. 6, pp. 483–502, 1999, ISSN: 00104809. DOI: 10.1006/cbmr.1999.1511.

[262] J. R. Stroud, P. Müller, and B. Sansö, "Dynamic models for spatiotemporal data," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 63, no. 4, pp. 673–689, 2001, ISSN: 13697412. DOI: 10.1111/1467-9868.00305.

[263] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature Neuroscience*, vol. 20, no. 3, pp. 340–352, 2017, ISSN: 15461726. DOI: 10.1038/nn.4497.

[264] V. Imbrasaite, T. Baltrušaitis, and P. Robinson, "CCNF for continuous emotion tracking in music: Comparison with CCRF and relative feature representation," *2014 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2014*, 2014. DOI: 10.1109/ICMEW.2014.6890697.

[265] L. R. Krol, P. Haselager, and T. O. Zander, "Cognitive and affective probing: A tutorial and review of active learning for neuroadaptive technology," *Journal of Neural Engineering*, vol. 17, no. 1, 2020, ISSN: 17412552. DOI: 10.1088/1741-2552/ab5bb5.