TOWARD A PASSIVE BRAIN COMPUTER INTERFACE FOR SIMULTANEOUS DETECTION OF MENTAL WORKLOAD AND STRESS

By

Mahsa Bagheri

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Faculty of Engineering & Applied Science

Memorial University of Newfoundland

Abstract

A passive brain-computer interface (pBCI) is a system that continuously adapts a human-computer interaction to the user's mental state. An example would be a system that aims to prevent traffic accidents by sending alerts to a truck driver when a state of drowsiness is detected. Key to the efficacy of such a system is the reliable estimation of the user's state via neural signals, acquired through non-invasive methods like electroencephalography (EEG). Typically, in pBCI studies, the state being explored (e.g., fatigue, frustration, boredom, attention) is considered in isolation, and no other aspect of the user's state is taken into account. In real-life scenarios, however, different aspects of the user's state are likely to be changing simultaneously - for example, their cognitive (e.g., level of mental workload) and affective (e.g., level of stress/anxiety) states. This inevitable confounding of different states needs to be accounted for in the development of state detection algorithms in order for them to remain effective when taken outside the lab.

In this work, simultaneous classification of two mental states via EEG is investigated for the first time. Specifically, mental workload and stress are explored since detection of both of these states would be useful in a variety of applications, including for improving safety in high risk work environments. Individually, both mental workload and stress have been studied extensively in the passive BCI literature, however in real-life scenarios they often vary concurrently within an individual. First, the effect of varying each state on classification of the other state was investigated to indicate if/how mental workload and stress confound one another. Then, different classification algorithms were proposed and evaluated to mitigate the confounding effects of variation in mental workload on the detection of stress and vice versa. Finally, a processing pipeline suitable for realizing an online BCI for simultaneous detection of mental workload and affective state was

investigated. This work represents a step toward the ultimate goal of realizing a functional, reliable, and robust passive BCI capable of detecting both mental workload and stress.

General Summary

Brain-computer interface (BCI) technologies aim to allow the control of external devices based on the brain activity of the user. Initially, the motivation of BCI research was the development of a movement-free means of communication and environmental control for people with severe physical disabilities. However, due to the considerable advancements in BCI algorithms achieved through decades of research, the potential of using BCIs in new applications, even for healthy users, has become highly interesting. One emerging area of BCI research, which has a wide range of potential applications for all users, are passive BCIs. Passive BCIs do not aim in the active control of devices, rather, the BCI is programmed to recognize different mental states of interest in the user (e.g., cognitive or emotional) from brain activity that arises naturally during the task at hand. An example would be a BCI that detects states of drowsiness and provides alerts to a truck driver, helping to increase safety by avoiding potential accidents.

In this research, two different mental states that are of particular importance in high risk work environments, mental workload level and stress, were investigated. Specifically, the ability to detect these two states in a user at the same time, using a non-invasive functional imaging technology called electroencephalophrahy (EEG), was explored. These states have been investigated before in other research studies, but always individually. The long term goal of this research is the realization of a functional, reliable, and robust passive BCI capable of detecting both the mental workload level and stress level of the user simultaneously, since having this more detailed information would provide a better picture of the user's mental state, and their risk of error. Such a system could have significant impact by helping to prevent industrial accidents and their associated human, economic, and environmental costs.

Acknowledgements

I would like to express my sincere gratitude and appreciation to my supervisor, Dr. Sarah Power, for her continued support and guidance. Without your support, continuous engagement, and prompt and thoughtful feedback, this dissertation would not have been completed. You are a true definition of ingenuity, dedication and kindness and I am proud and grateful for my time working with you.

I am grateful for the constructive feedback and kind encouragement offered by my supervisory committee, Dr. Cecilia Moloney and Dr. Heather Carnahan as well as by my examination committee, Dr. Tiago Falk, Dr. Brian Veitch, and Dr. Jiang Xianta.

I am extremely thankful for my family and friends' unconditional, unequivocal, and loving support.

I would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

Contents

Lists of Tables

Lists of Figures

Lists of Abbreviations

aCAMS Automation-enhanced cabin air management system

ANN Artificial neural network

ANOVA Analysis of variance

ALS Amyotrophic lateral sclerosis

ATM Air traffic management

BCI Brain computer interface

CFS Correlation-based feature selection

CNN Convolutional neural network

CSP Common spatial pattern

ECG Electrocardiogram

ECOG Electrocorticography

EEG Electroencephalography

EMG Electromyography

EOG Electrooculography

ERD Event-related desynchronization

ERP Event related potential

fMRI Functional magnetic resonance imaging

fNIRS Function near-infrared spectroscopy

GSR Galvanic skin response

IAF Individual alpha frequency

ICA Independent component analysis

ICEHR Interdisciplinary committee on ethics in human research

KNN K-nearest neighbors

LDA Linear discriminant analysis

MATB	Multi-attribute task battery
MEG	Magnetoencephalography
MIST	Montreal imaging stress task
MRI	Magnetic resonance imaging
mRmR	Minimum redundancy maximum relevance
NB	Naive bayes
NSERC	Natural sciences and engineering research council of Canada
pBCI	Passive brain computer interfaces
PET	Positron electron tomography
RG	Relative Gamma
RSME	Rating scale of mental effort
SCWT	Stroop color-word test
STAI	State-trait anxiety inventory
SVM	Support vector machine
TSST	Trier social stress task
WML	Working memory load

Chapter 1 : Introduction

1.1 Brain Computer Interfaces (BCIs)

Brain-computer interface technologies aim to provide a method of communication between a human and a device that is based solely on neural signals measured from the brain [1, 2]. Initially, the motivation of BCI research was the development of a movement-free means of communication and environmental control for people with severe physical disabilities, and specifically those with late-state amyotrophic lateral sclerosis (ALS) who may retain no residual volitional motor control, in any part of their body, with which they could operate more conventional interfaces (e.g., keyboard, mouse, pushbutton, etc.) [1, 3]. In such a BCI, a user would initiate different commands by intentionally generating specific patterns of activity in their brain. Typically, this would be done either by selectively focusing on different external stimuli (in a so-called "reactive" BCI), or by performing different mental tasks (in a so-called "active" BCI), that would each be associated with a different command. The BCI system would detect which stimuli the user was observing, or which task they were performing, based on neural signals measured via one of a number of possible imaging modalities, and output the appropriate command to the connected device. For example, imagining movement of the right hand might translate to a "move cursor right" command, and imagining movement of the left hand might translate to a "move cursor left" command. Reactive/active BCI research has been ongoing since the late 1980s/early 1990s [4, 5, 6], and the interested reader is referred to [7-9] for a recent review of this literature.

Recently, the field of BCI research has started to branch in new directions. Due to the considerable advancements in BCI algorithms achieved through decades of research, the potential of using BCIs in new applications, even for healthy users, has become highly interesting [1]. In particular, research in the area of "passive BCIs" has grown significantly since the concept was introduced approximately a decade ago [10]. Rather than deriving its outputs from brain activity which is directly and consciously controlled by the user for the purpose of controlling a device, a passive BCI is one that is designed to recognize different mental states of interest (e.g., cognitive or emotional) that arise naturally during the task at hand, and use this implicit information to adapt the environment in a useful way. An example application would be a BCI that provides alerts to a truck driver when a state of drowsiness is detected, helping to increase safety by avoiding potential traffic accidents, or a BCI that adjusts the difficulty level of a video game based on neurally-derived measures of boredom and frustration in the player. The original research described in this thesis falls within the category of passive BCI research.

1.2 BCI design

Generally, a BCI, whether active or passive, consists of a variety of intermediary components that perform specific functions in first detecting the user's mental state and then issuing an appropriate command to the connected device or application. Figure 1.1 illustrates a block diagram of the main functional elements of a BCI and their principal interactions. Typically, a BCI model consists of the six main components of signal acquisition, signal pre-processing, feature extraction, feature selection, classification, and feedback, each of which will be explained in the following sections [11]. In Figure 1.1, the whole process is comprised of two consecutive phases: the training/calibration phase and the testing/use phase. In the training/calibration phase, labeled samples of neural data from a set of pre-defined mental states are successively processed by the different modules to train a classifier capable of discriminating amongst the mental states. Then, in the testing/use phase, unlabeled samples of neural data are identified by the classifier and translated into a command to control an external electronic device (e.g., to operate a wheelchair in the case of an active BCI, or to change/adapt, in some useful way, the behavior of the interface that the user is interacting with in the case of a passive BCI). Note that the feature selection step is only used in the calibration phase to identify the most discriminatory signal features to use in the classifier; these selected features are then used directly in the testing/use phase.

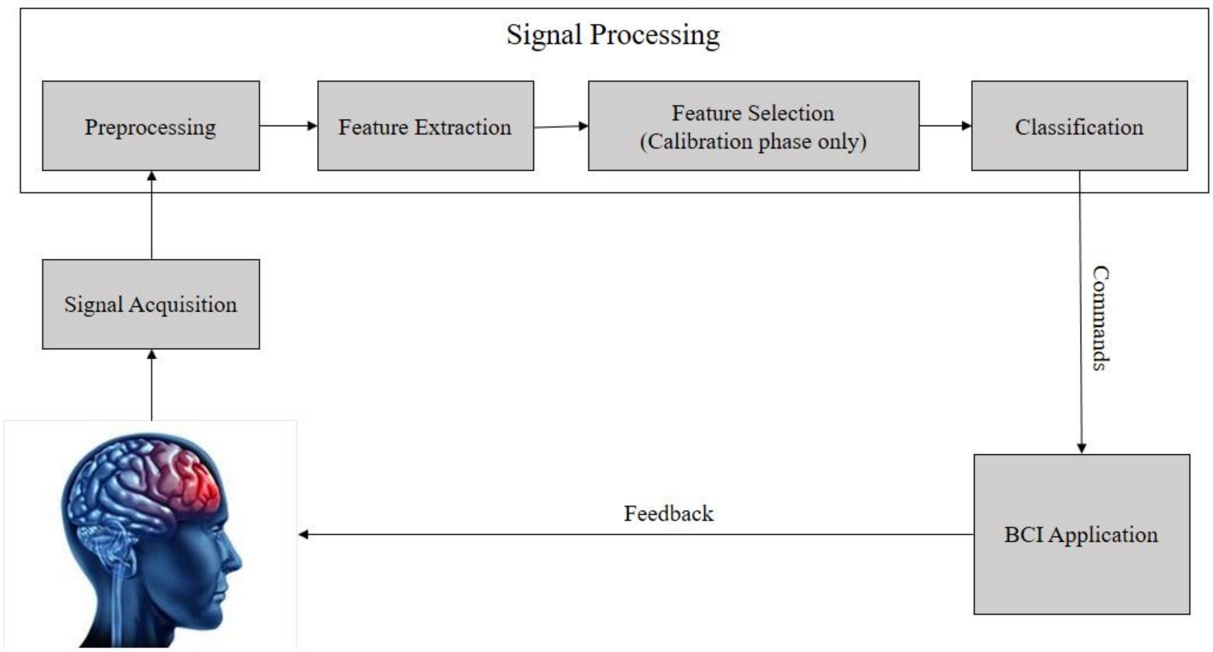


Figure 1.1: The block diagram of a typical BCI system

1.2.1 Physiological phenomenon and signal acquisition

Electroencephalography (EEG), which measures the electrical activity of the brain via electrodes placed on the scalp, is the most prevalent method of signal acquisition used in passive BCI research due to its non-invasiveness, portability, high temporal resolution, relative low cost, and suitability for use in real-life scenarios [11]. Table 1.1 provides a comparison of some other functional imaging modalities and their relative advantages and disadvantages for passive BCI applications [9]. Hereafter, this thesis will focus on the discussion of EEG-based BCI technologies.

The electrical activity measured by EEG originates primarily in the neurons of the cerebral cortex (the outer surface of the brain), and thus reflects various higher level mental processes such as consciousness, thought, emotion, reasoning, language, sensory processing, motor control, and memory [12]. The cerebral cortex is divided into four lobes (frontal, temporal, parietal, and occipital), each of which is associated with different functions. EEG signals do not reflect the activity of individual neurons, but rather they measure the electric fields resulting from the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientation. Pyramidal neurons of the cerebral cortex are thought to contribute the most to the measured EEG signal because they are well-aligned and fire together [13]. EEG is thus a direct measure of neural activity and has excellent time resolution in the range of milliseconds suitable to examine real-time and ultra-fast neurodynamics and information processing, however its spatial resolution is relatively poor. Because the EEG electrodes are separated from the neural signal sources within the cortex by the meninges, cerebrospinal fluid, skull, and scalp, this results in the smearing of electrical potentials. More specifically, the EEG signal obtained at any given electrode will not reflect activity exclusively from the cortical area directly around that electrode, but will be a mix of signals originating from neurons at different spatial locations from different sources [14]. EEG has seen widespread use in numerous fields since its discovery by Hans Berger in the 1920's [15], therefore the techniques and technology of signal acquisition through this method are well-understood, and have been standardized [16].

Table 1.1: Neural/brain activity measures

Methods	Advantage	Disadvantage
EEG	▪ Excellent temporal resolution ▪ No real safety restrictions ▪ Cost efficient	▪ Low spatial resolution
fMRI	▪ Excellent spatial resolution	▪ Indirect measure of neural activity ▪ High latency ▪ Not portable ▪ Not practical for real-life scenarios ▪ Expensive
fNIRS	▪ Good spatial resolution ▪ Portable ▪ Suitable for real-life scenarios	▪ Indirect measure of neural activity ▪ High temporal latency of measured response ▪ Relative high cost
MEG	▪ Both excellent temporal and spatial resolution	▪ Not portable ▪ Not practical for real-life scenarios ▪ Expensive
PET	▪ High spatial resolution	▪ Indirect measure of neural activity ▪ Low temporal resolution ▪ Expensive ▪ Significant safety restrictions
ECoG	▪ Excellent signal-to-noise (SNR)	▪ Invasive, so unsuitable for passive BCI applications

In conventional EEG, the recording is obtained by placing one or more electrodes on the head, and using a conductive gel or paste to reduce impedance with the scalp and maximize signal-to-noise ratio. For ease of electrode placement, multi-channel/multi-electrode EEG systems typically use caps or nets into which electrodes are embedded and secured to the head. Electrode placement typically follows the “International 10/20 System” [16] as shown in Figure 1.2a. The "10" and "20" refer to the fact that the distances between adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull. For higher density systems, it is also common to use a variant of the 10/20 system that fills in the spaces between the electrodes with additional electrodes (e.g., the “10/10 System”, Figure 1.2b) [17].

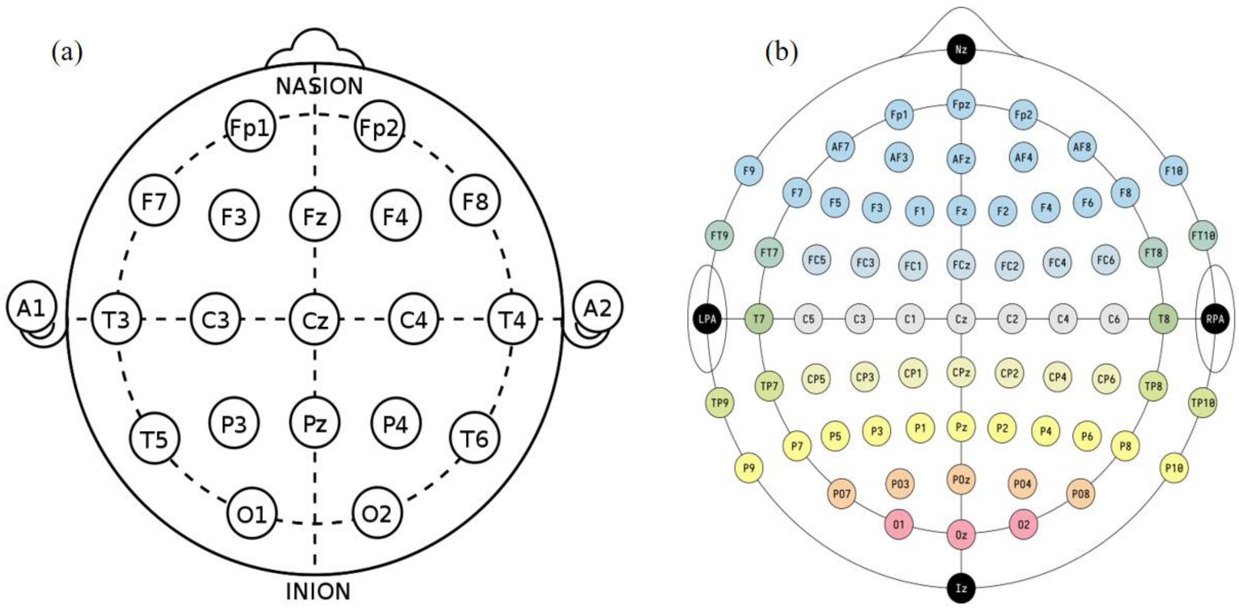


Figure 1.2: a) International 10/20 system. b) International 10/10 system.

1.2.2 Signal pre-processing

Once the EEG data is acquired, it then usually undergoes pre-processing. This typically involves the removal of artifacts from the data so that the resulting signals more accurately reflect the true underlying neural activity.

EEG artifacts are electrical signals that are recorded at the electrodes but that were not generated by neural activity. There are two main types of artifacts: electrophysiological artifacts such as electrooculography (EOG; movement of the eyes), electrocardiography (ECG; cardiac activity), and electromyography (EMG; muscle activity) originate from the body of the user, while non-physiological artifacts originate from the environment and may include things like motion artifacts caused by the movement of the electrodes, power line noise (50 or 60 Hz), or high frequency noise caused by poor connection of the electrode with the scalp. Artifact rejection can be done either manually by simply looking at the data (the most serious artifacts tend to have recognizable patterns) and removing the contaminated sections of the signals, through automated artifact-removal algorithms, or a combination of both. When EEG data is to be analyzed "offline" (i.e., using data that has been previously recorded and stored) as is common in preliminary BCI studies, then either manual or automated approaches are feasible. However, for "online" analysis (i.e., where the data is analyzed in real-time as soon as it is recorded) automated algorithms are required. Practical BCI systems, which analyze the user's brain activity and predict their mental state in real-time, require automated pre-processing techniques, and a balance must be struck between the effectiveness of the chosen techniques for removing the unwanted artifacts, and the speed/complexity of the algorithms. There is no universally adopted EEG pre-processing pipeline in BCI research, and researchers typically exercise some freedom in choosing the most appropriate techniques to use for their given dataset and application.

1.2.3 Feature extraction

Feature extraction aims to reduce the amount of resources needed to describe a large dataset accurately while at the same time minimizing the loss of important information embedded in the data [18]. Therefore, after the pre-processing phase, different features of the EEG signals that are thought or known to capture underlying cognitive processes and therefore might be useful in identifying the target mental states, are calculated. In EEG-based BCI design, there are three main types of information that are typically targeted when extracting features [19].

- Spatial information: This describes where the relevant signal comes from. In practice, this means selecting only specific EEG electrodes that are more relevant for a particular mental state discrimination task, or focusing more on some than on others. Some BCI systems use spatial filtering algorithms to combine several sensors, generally through linear combination, in order to form a new (virtual) sensor from the extracted features. This is useful not only because it reduces the number of original sensors to a small number of spatially filtered signals, but also because it has a neurophysiological meaning. Indeed, as discussed in section 1.2.1, EEG signal obtained at any given electrode will not reflect activity exclusively from the cortical area directly around that electrode, but will be a noisy mix of signals originating from neurons at different spatial locations from different sources. Therefore, spatial filtering makes it possible to help recover the original signal by gathering relevant information that was scattered over different sensors [19]. One of the most successful and well-known methods widely used for feature extraction in EEG-based BCI is the common spatial pattern (CSP). The CSP algorithm is used to enhance the spatial resolution of EEG and maximize the discriminability of two classes [19].

- Spectral/frequency information: This describes how the power of the EEG signal varies in some specific frequency bands. In practice, this is equivalent to using signal band power as features.
- Temporal information: This describes how EEG signals vary over time. In practice, this means using the values of EEG signals for different specific time intervals or different time windows. Temporal features are typically used in "reactive BCIs" where the relevant neural signal is time-locked to a known stimulus, but are less commonly used in active and passive BCIs.

Among different features which have been investigated in EEG-based BCI studies, signal power over frequency bands of interest, and at different spatial locations, are most commonly used in BCI studies [20].

1.2.3.1 EEG frequency bands

EEG signals are typically divided into several frequency bands as shown in Figure 1.3. A majority of BCI research focuses on the alpha band (8-13 Hz) and the beta band (14-30 Hz) [10]. The beta band is sometimes considered to have an extended range of up to 60 Hz with the gamma band indicating all signals greater than 30 Hz.

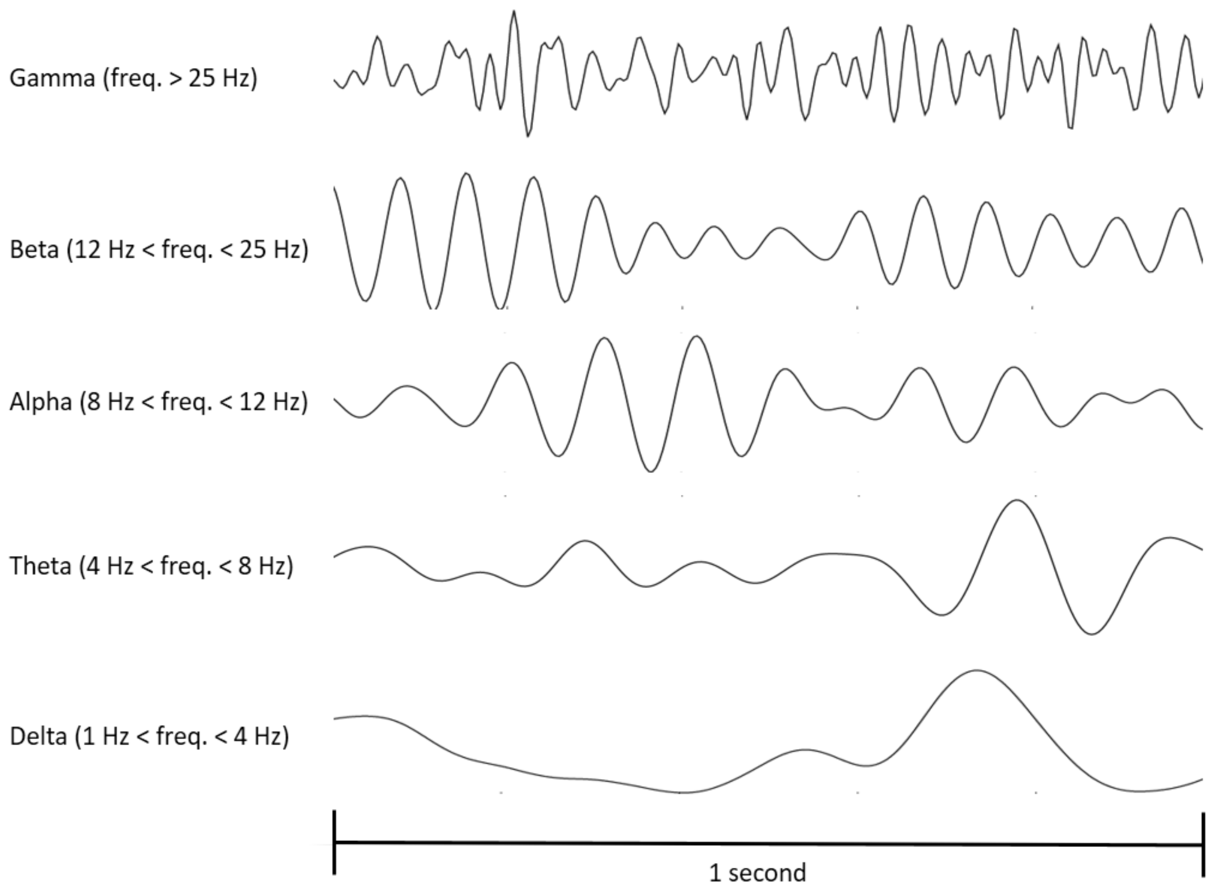


Figure 1.3: EEG frequency bands

1.2.4 Feature selection

The feature selection process plays an important role in the performance of the BCI algorithm by eliminating redundant and irrelevant features from the overall feature pool that was calculated, and selecting a small set of the most discriminative features (i.e., the features that most clearly distinguish between the different mental states of interest). When considering a large number of features, reducing the feature set is important because of the phenomenon called the “curse of dimensionality” which states that the number of training samples needed to properly describe the different classes increases exponentially with the dimension of the feature vector (i.e. the number of features used) [21-23]. It has been recommended to use from 5 to 10 times as many training examples per class as the size of the feature vector [24].

Feature selection techniques are categorized in two main groups: filter methods and wrapper methods [25]. Filter methods involve evaluating the effectiveness of individual features, or combinations of features, for discriminating the states to be classified based on some pre-defined measure that is independent from the classification algorithm that is to be applied (e.g., correlation-based feature selection (CFS), ReliefF, minimum redundancy-maximum relevance (mRmR)) [26]. Wrapper methods, on the other hand, use the resulting performance of the classification algorithm (e.g., accuracy) in discriminating the target states to evaluate the effectiveness of individual features, or combinations of features [27].

Feature selection is only done in the calibration/training phase of the BCI algorithm. In the testing/use phase the features selected in the calibration phase are used directly in the classifier.

1.2.5 Classification

Machine learning techniques are employed to predict the user’s mental state based on their neural signals. The selected features are passed to a classifier which use them to predict which of a pre-defined set of target mental states the user is experiencing. There are numerous types of linear and non-linear classification algorithms, such as linear discriminant analysis (LDA), logistic regression, naïve Bayes, support vector machines (SVM), k-nearest neighbor (KNN), random forest, and artificial neural networks [28]. As is the case in most machine learning applications, there is no "ideal" classification algorithm for use in BCI systems. As always, the aim when selecting a classification algorithm is to maximize performance and practicability for the chosen application.

Three classification algorithms commonly used in BCI research - LDA, SVM and KNN - are briefly described below.

1.2.5.1 Linear discriminant analysis (LDA)

Linear discriminant analysis seeks to separate two or more classes by using hyperplanes, which is obtained by searching for the projection that maximizes the distance between the class means and minimizes the interclass variance as shown in Figure 1.4. This technique has a very low computational requirement and is relatively simple to use [29]. LDA has been used with success in various types of passive BCI systems such as cognitive workload recognition [e.g., 30-36], mental stress detection [e.g., 37-39], and fatigue/drowsiness detection [e.g., 40-46].

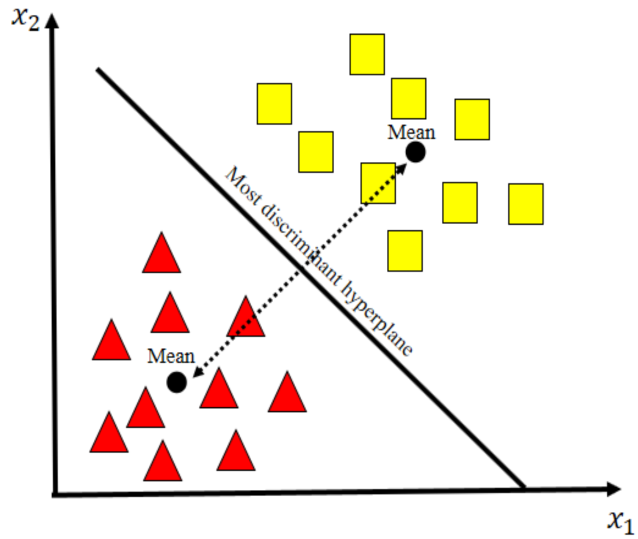


Figure 1.4: The basic illustration of the LDA classifier used for a binary classification problem. Samples from Class 1 are represented by red triangles and samples from Class 2 are represented by yellow squares. $x_1$and x_2 are features of the EEG signal being used in the classifier.

1.2.5.2 Support vector machines (SVM)

The support vector machine algorithm is similar to LDA in that it exploits a discriminant hyperplane to predict classes. However, in the case of SVM, the selected hyperplane is the one that maximizes the distance between the nearest training points of the different classes. This optimal hyperplane is described by the vectors which lie on the margin, which are called support vectors (see Figure 1.5). SVM can be easily extended to complex instances that are not linearly separable. This is done by mapping the original input space to a higher-dimensional space where they become linearly separable by the use of kernel functions on training sets. SVM is known to have good generalization properties and to be insensitive to over-training and to the curse-of-dimensionality [29].

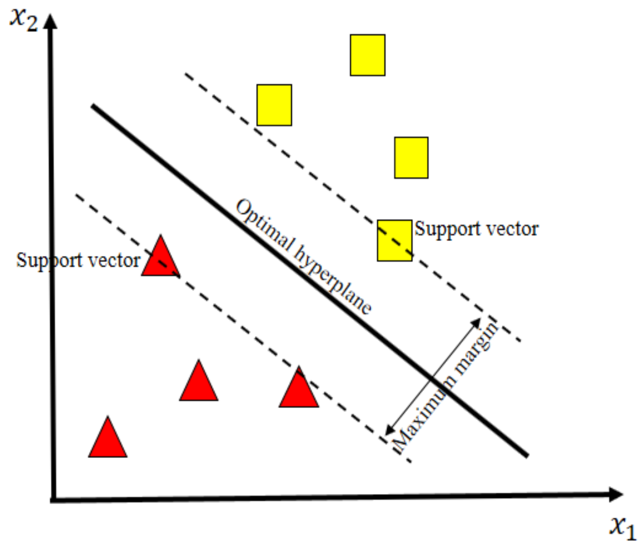


Figure 1.5: The basic illustration of the SVM classifier used for a binary classification problem. Samples from Class 1 are represented by red triangles and samples from Class 2 are represented by yellow squares. $x_1$and x_2 are features of the EEG signal being used in the classifier.

1.2.5.3 K-nearest neighbors (KNN)

In the K-nearest neighbor algorithm, a test sample is assigned to the majority class among its K nearest neighbors (as determined by a distance metric) within the training set. Figure 1.6 shows the basic illustration of KNN classification for a binary classification problem. In the figure, if K=3, the test sample is labeled as Class 1 (indicated by red triangles) and if K=5, the test sample is labeled as Class 2 (indicated by yellow squares). KNN is a non-linear classifier. The KNN algorithm's main advantage is its simplicity [29]. However, sensitivity to the curse-of-dimensionality is considered the main drawback that affects its performance in BCI systems [29].

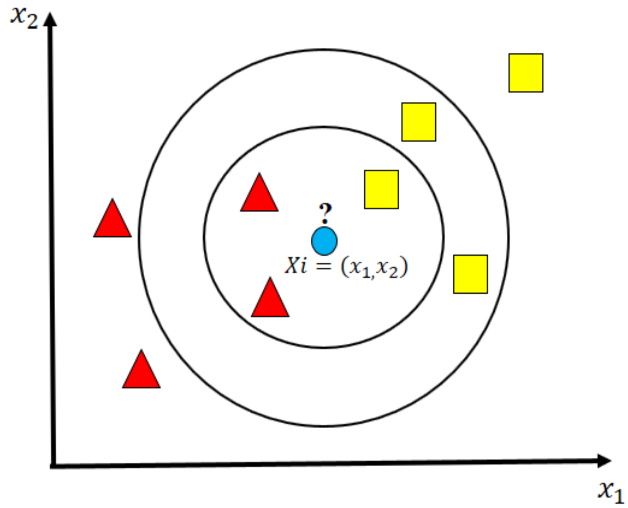


Figure 1.6: The basic illustration of KNN classification used for a binary classification problem. Samples from Class 1 are represented by red triangles and samples from Class 2 are represented by yellow squares. x_1 and x_2 are features of the EEG signal being used in the classifier. If K=3, the test sample is labeled as Class 1 and if K=5, the test sample is labeled as Class 2.

1.2.6 Feedback

Based on the predicted mental state, commands are sent to change/adapt, in some useful way, the behavior of the interface that the user is interacting with. For example, in the case of drowsiness detection this might be the activation of an alarm to alert and arouse the driver. This feedback in turn affects the user's mental state, thus closing the control loop.

1.2.7 BCI performance evaluation

1.2.7.1 Performance metrics

Many metrics have been proposed for quantifying the performance of BCI systems [47]. However, the most widely reported metric is accuracy, calculated as the percentage of the total test samples that are correctly classified (Equation 1.1).

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad , \tag{1.1}$$

where:

- Tp is the number of true positives classified by the model, (i.e., the number of samples from the positive class that are correctly predicted as being positive).
- Tn is the number of true negatives classified by the model, (i.e., the number of samples from the negative class that are correctly predicted as being negative).
- Fp is the number of false positives classified by the model, (i.e., the number of samples from the negative class that are incorrectly predicted as being positive).
- Fn is the number of false negatives classified by the model (i.e., the number of samples from the positive class that are incorrectly predicted as being negative).

Accuracy is intuitive and easily interpretable, and is informative as a BCI performance metric as long as two conditions are met: first, the data must be balanced amongst all classes, and second, the decisions of the BCI must be unbiased (i.e. approximately equal performance should be achieved for each class). When these conditions are not met then the accuracy can be biased, and the F_1-score is a good alternative BCI performance metric.

The F_1-score (Equation 1.2) is defined as the harmonic mean of the model's precision and recall, where precision (Equation 1.3) is the fraction of true positive samples among the samples that the model classified as positive, and recall (Equation 1.4) is the fraction of examples classified as positive among the total number of positive examples.

$$F_1\text{-}score = \frac{2}{\frac{1}{Precision} \times \frac{1}{Recall}} = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{Tp}{Tp + \frac{1}{2}(Fp + Fn)} \quad , \tag{1.2}$$

where:

$$Precision = \frac{Tp}{Tp + Fp} \quad , \tag{1.3}$$

$$Recall = \frac{Tp}{Tp + Fn} \quad . \tag{1.4}$$

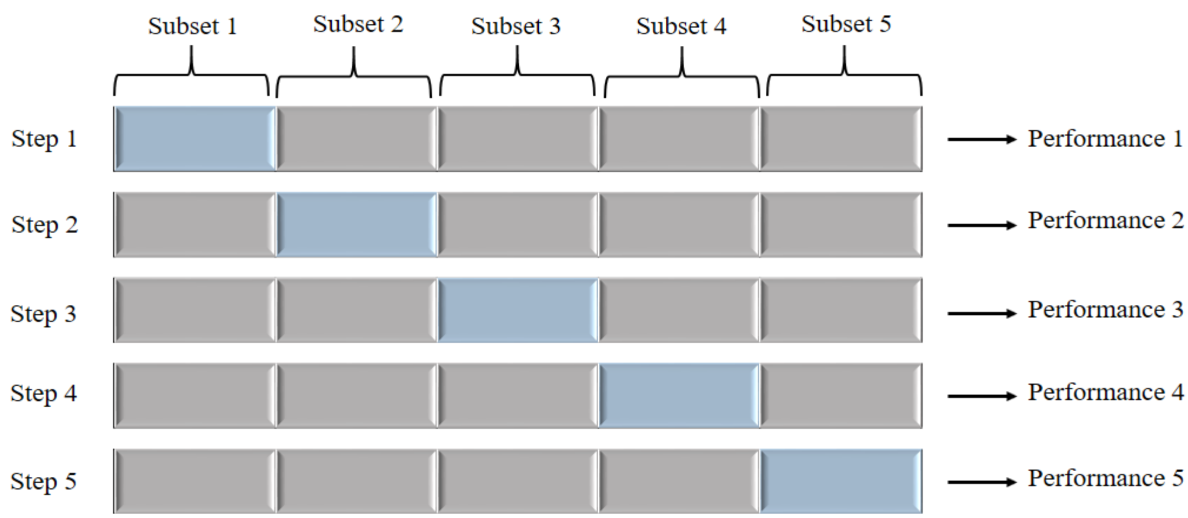
The highest possible value of an F_1-score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0, if either the precision or the recall is zero.

1.2.7.2 Training and testing schemes

An important question when discussing BCI performance evaluation is the way in which the data are handled. With a pre-recorded dataset (i.e., when doing offline analysis), the whole dataset is divided into two groups: the training set and the test set. The training set is used to select the features and train the classifier, and the test set is used to evaluate the performance of the classifier. One very common strategy of dividing the data set into the training and test sets for offline analysis is called "k-fold cross-validation". In this approach, the data set is randomly divided into k subsets of approximately equal size, ideally with classes balanced within each subset. In each step, i, of the procedure, the i^{th} subset is held out as the test set, and the training set is comprised of the remaining $k-1$ subsets. The performance of the classifier in the i^{th} step is estimated using an appropriate metric. This is repeated for k steps, until each of the k subsets has been used as the test set exactly once. Finally, the overall classifier performance is estimated as the average of the

k performance estimates from each step. Figure 1.7 clarifies the k-fold cross-validation procedure. k-fold cross-validation is a popular approach in applied machine learning as it generally results in a less biased and more generalizable estimate of the classifier performance than other methods, such as a simple train/test split. The process can be repeated, and the performance averaged, for multiple "runs" of the k-fold cross-validation in order to further reduce variability. A common issue that often appears with cross-validation is called data leakage. Data leakage refers to a problem where information about the holdout dataset, such as a test or validation set, is made available to the model in the training set. This results in unrealistically high levels of performance of the model on the test set, because that model is being developed on data that it had already seen in some capacity in the training set. When such a model is then used on truly unseen data, the performance of that model will be much lower than expected. There are different examples of the data leakage in machine learning applications; however, a common example with EEG-based BCI systems is the data leakage due to the EEG epoching. More clearly, when EEG data is epoched using sliding windows, neighboring samples/epochs may have some data leakage in them which results in the unreliable prediction outcome after model development.

When doing online BCI classification, a sufficient amount of data from each of the mental states to be classified must be collected in a calibration session, and used to train the classifier. Then, the classifier is used to predict the mental state of the user based on their EEG signals collected and processed in real-time (in the testing/use phase).



Overall estimate of classifier performance $= \frac{1}{5} \sum_{i=1}^{5} Performance\ (i)$

Figure 1.7: k-fold cross-validation. The division of data into training and test sets for k-fold cross-validation, shown here for k=5. In each step, the training set is composed of the grey subsets, and the blue subset is the test set.

1.3 Mental state classification for passive BCI

As is clear from section 1.2, the first and most crucial step in realizing a BCI is developing methods (e.g., identifying useful EEG features, classification algorithms, etc.) for the accurate classification of the mental states of interest based on neural signals. This is true of both active and passive BCIs, but the task is arguably more challenging in passive BCI systems. Rather than having a specific, discrete set of target states that are consciously controlled by the user, as is the case in active BCI, in passive BCI the states being detected are spontaneously occurring during naturalistic scenarios, and therefore by nature have much more variability and are less well-defined.

To date, many studies have investigated the possibility of classifying different cognitive or emotional states for the ultimate purpose of realizing passive BCI systems. States that have been investigated include fatigue/drowsiness [e.g., 40-46], sustained attention [e.g., 48, 49], task engagement [e.g., 50], expertise/skill acquisition [e.g., 34, 51-53], various emotions [e.g., 54-56], cooperation in a team [e.g., 57, 58], and perception of advertisements [e.g., 41, 59-62]. Common application domains include driving [e.g., 43, 49, 54, 63-68], aviation including piloting and air-traffic control [e.g., 33, 40, 69-72], surgery [e.g., 73], education [e.g., 48], and neuromarketing [e.g., 60-62, 74].

Two other states that are, because of their relevance to a wide variety of real-life applications, among the most commonly studied for passive BCI applications are mental workload, and stress. These states are the focus of this thesis research.

1.3.1 Mental workload detection

The detection of mental workload has received particular attention in the pBCI literature in a variety of application domains, but particularly for improving safety in high risk occupations like airplane pilots and air-traffic controllers. The overarching goal of such a pBCI is to derive a continuous, objective estimation of the cognitive load on an individual based on neural signal variables, so that appropriate adaptation strategies can be engaged to reduce the potential for error during periods of extreme demand or overload. Mental workload recognition has been applied to various domains, such as education (e.g., students learning online [75] and web browsing [76]), public transport (e.g., driving vehicle [77], airplane [35], air traffic management [33]), health care [78], and other safety critical occupations (e.g., engineers of nuclear power plants [79]).

1.3.1.1 Typical experimental paradigms

Mental workload is defined as the perceived relationship between an individual's total mental processing capability and the amount required by the task at hand; the closer the requirements are to the actual capabilities, the higher is the perceived workload [80]. Therefore, a common strategy for workload manipulation in pBCI research is the variation of task difficulty. Often, the paradigms used to induce mental workload experimentally are just simple cognitive tasks that are often borrowed from cognitive research (e.g., n-back working memory task [81], Sternberg working memory task [31], mental arithmetic tasks [82, 83], IQ test [84], silent reading [85], and visual degradation task [86]), where the level of mental workload can be fairly reliably modulated. Many other studies employ more realistic, multi-faceted tasks specific to the target application domain, often performed in simulated environments, to induce different levels of workload. Common examples include the Air Traffic Management (ATM) task [87], driving simulation tasks [84], flight simulation tasks [35, 88-90], and complex operation multi-tasks such as the NASA Multi-Attribute Task Battery (MATB) [91, 92] and the automation-enhanced Cabin Air Management System (aCAMS) [93].

The number of mental workload levels investigated in a single study has ranged from two to seven, though the majority of studies attempt to induce/classify between two and four workload levels [31, 32, 35, 82, 83, 88, 89, 94-109].

1.3.1.2 EEG markers of mental workload

Several studies have revealed associations between varying workload levels and power alterations in EEG frequency bands. It has been shown that the most prominent event was the increase of the EEG power spectrum in the theta frequency band over the prefrontal cortex, and the decrease of the EEG power spectrum in the alpha frequency band [20, 42, 89, 110-119]. Studies reported the increase of the theta power at parietal areas in response to an increased task demand [120], at the frontal cortex in relation to an increase of focused attention during the task [121], at the frontal and central brain areas during a time pressure task [122], and at the prefrontal cortex in flight and air-traffic control simulations tasks [110, 112, 113]. Besides alpha and theta bands, powers in the delta, beta, and gamma bands have been reported to associate with varying workloads [94, 123, 124]. In a related study, a decrease of the EEG power spectrum in the delta frequency band was found in response to increasing workload in a silent reading task [123].

1.3.1.3 Classification Methods

A classification or a regression model can be designed to detect mental workload. The output of a classifier is a discrete value, whereas the output of a regressor is a continuous variable. Since the majority of pBCI studies aim to induce/detect discrete workload levels (e.g., high vs. low), this section focuses on classification problems.

In terms of analytical methods, the cognitive workload recognition studies can be divided into those using classical machine learning models and those using deep learning models. The classical machine learning models use steps including data preprocessing, feature extraction, feature selection, classification, and performance evaluation. Different studies employed different algorithms/methods for the mentioned steps. However, frequency domain features along with SVM [86, 105, 125-131] and LDA [30-34, 83] classifiers were mostly used in the literature. Some other classifiers used in cognitive workload recognition, include KNN [97] and the Bayes-based model [101, 102, 132].

In contrast to the classical machine learning models, which typically extract features from the temporal and spectral views separately, deep learning can learn to acquire complex information of multiple domains simultaneously [133, 134]. Therefore, mental workload researchers have begun using deep learning (e.g., Convolution Neural Network [77, 133, 135], Recurrent Neural Network [136, 137], Denoising Auto-Encoder [93, 138-140], and Deep Belief Network [141, 142]) to learn robust EEG representations. It is worth nothing that while the deep learning models might improve the evaluation performance of multi-class classification through powerful nonlinear feature representation, they require large amounts of training data and more training time to tune the structure and parameter [134].

In terms of performance evaluation strategies, a majority of studies have used offline analysis, with k-fold cross-validation being mostly employed when doing subject-specific mental workload detection (i.e., the BCI classifier is trained specifically for use by one individual, based on their own training data) [30-32, 35, 36, 58, 81, 86, 102, 125, 126, 128-130, 132, 137, 140, 143-147], and leave-one-subject-out cross-validation being most commonly used when doing subject-independent mental workload detection (i.e., the BCI classifier is trained for use by any user, based on training data collected from a number of other individuals) [82, 94, 95, 104, 133, 135, 136, 142, 148-150]. In online studies, of which there have been significantly fewer [32, 151-153], the experiment of real-time workload detection generally consists of two consecutive or separated sessions: first a calibration session to collect training data with which to train the BCI classifier, and a second "online" session where the classifier is then used to detect the mental workload level of the user in real-time. In the online session, the classifier has been used to detect mental workload either in the same tasks that were used in the calibration session [32, 151, 152], or in different tasks [153].

Generally, there is a wide range of classification results reported, and it is difficult to compare one study to the next due to differences in the experimental (e.g., type of task, number of workload levels, difference between workload level conditions, number of electrodes, offline or online) and analytical (e.g., pre-preprocessing techniques, EEG features, feature selection and classification algorithms) methods employed [154]. That said, typically a majority of mental workload detection studies in recent years have produced classification results at or above 80% [31, 32, 35, 75, 77, 80, 82, 93-95, 98-102, 123, 125-127, 129, 133-137, 139-141, 143, 145-151, 155], which is promising.

Further research on the application of pBCI in ecologically-valid scenarios will be needed to determine whether this performance is transferable to real-life applications, and then whether it is sufficient so as to be effective in meeting the objectives of the BCI in a given application (e.g., to reduce the number of errors due to cognitive overload committed by an operator).

1.3.2 Stress detection

Another application that has received a lot of interest in the pBCI research area is the monitoring of the user's affective state, particularly stress/anxiety. Stress is known to negatively impact cognitive efficiency and performance [156] as well as decision-making (especially when performing unfamiliar tasks) [157], thus a pBCI designed to detect the user's stress level could initiate appropriate task or environmental adaptation strategies to mitigate these potential negative effects.

1.3.2.1 Typical experimental paradigms

There are many techniques that have been used to induce levels of stress in lab settings. A detailed review of the experiments to elicit stress is given in [158]. However, the most widely used paradigms to induce mental stress are the Stroop Color-Word Test (SCWT) [159-161], the Trier Social Stress Test (TSST) [162, 163], and the Montreal Imaging Stress Task (MIST) [37, 164-170], as well as various mental arithmetic tasks [171-175], music videos [176-178], and exam stress [179].

In different studies, the TSST, MIST and SCWT paradigms have been modified to meet the specific needs of the research; however, a brief review of their original versions is as follows.

The TSST generally consists of three steps [162]. Stress induction begins with the participant being taken into a room where a panel of judges are waiting, along with a video-camera and audio recorder. The first 10-minute step is the anticipatory stress phase, during which the judges ask the participant to prepare a 10-minute presentation. In most studies this presentation is framed as part of a job interview or a general public speaking task. During the 10-minute presentation step, the judges observe the participant without comment or other feedback, to increase anxiety. The presentation is then followed by a verbal mental arithmetic task.

The Montreal Imaging Stress Task (MIST) consists of a series of mental arithmetic tasks, along with social evaluative threat components that are built into the program or presented by the investigator [180]. The MIST has three test conditions of rest, control and experiment. In the rest condition, subjects look at a static computer screen on which no tasks are shown. In the control condition, a series of mental arithmetic tasks are displayed on the computer screen, and subjects submit their answers by means of a response interface. In the experimental condition, the difficulty and time limit of the tasks are manipulated to be just beyond the individual's mental capacity. Along with the time limitation in the task, social evaluative threat components such as negative feedback when answering incorrectly or failing to answer each question within the time limit, or performance evaluation strategies, are added to further increase the stress experienced by the participants.

In the most common version of the Stroop Color-Word Test (SCWT), subjects are presented with color words (i.e., the words "red", "green", "blue", etc.) displayed in matching and non-matching font colors [181]. Subjects are required to read three different tables as fast as possible. Two of them represent the "congruous condition" in which participants are required to 1) read names of colors printed in black ink, and 2) name the color of different colored patches. In the third table, named the "color-word condition", color words are printed in an inconsistent colored ink (e.g., the word "red" printed in green ink). In this "incongruent condition", participants are required to name the color of the ink instead of reading the word. This condition is quite difficult, and induces mental stress in the participant.

It is worth noting that the SCWT, MIST and mental arithmetic tasks are mental stress-inducing tasks that also induce mental workload, whereas psychological tasks such as viewing music videos, or the anticipation phase of an exam or public speaking task, induce the stress state while not affecting mental workload.

There are a wide range of studies that aimed to simply detect the state of stress as compared to a relaxed/rest/control condition [37, 161, 166-168, 170, 172-177, 182-187], whereas others aimed to induce/recognize different levels of stress (two to four levels) [39, 160, 169, 188-192].

1.3.2.2 EEG markers of stress

Many studies have explored power spectrum, or relative power, indices to assess the behavior of the human brain in emotional stressed states, with various patterns reported. Many of the reviewed studies have reported a decrease of the EEG power spectrum in the alpha frequency band during the stressful conditions compared to the relaxed conditions [170, 193, 194, 195]. Regarding the beta band, studies have reported high beta activity as a result of stress [168, 193, 196]. The prefrontal relative gamma (RG) power has also been suggested as a marker for stress assessment, since the RG was shown to be more discriminative between stress levels than alpha asymmetry, theta, alpha, beta, and gamma power in the prefrontal cortex, with a positive correlation with stress level [197]. In [198], an increase in the ratio of the power of the beta waves over the alpha waves as a results of stress were found. In another study, the change in the ratio of the slow waves over fast waves was detected for both delta/beta and theta/beta configurations among states of relaxed and stressed [199].

EEG has been shown to be sensitive to localized brain activity in areas responsible for the stress response [200]. According to [169], the most dominant cortical structure that is involved in stress detection is the right prefrontal cortex. Furthermore, studies also reported higher right prefrontal activity in stressful contexts, compared to left prefrontal activity [169, 179, 201-203].

1.3.2.3 Classification results

Like the mental workload detection studies, mental stress detection studies have varied significantly in terms of the experimental and analytical methods used, including the number of subjects, number of EEG channels, type of stressors, amount of data collected, pre-processing methods, feature extraction mechanisms, and types of classifier.

While various types of classifiers have been used to assess mental stress, the most common and significant classifiers are SVM [38, 39, 160, 163, 165-170, 174, 183-185, 187, 188, 190, 192, 204, 205], KNN [38, 39, 160, 176, 177, 183, 184, 190], LDA [37-39], naive Bayes (NB) [167, 192, 206] and convolutional neural network (CNN) [161, 191]. Most of the reviewed studies conducted offline experiments and only a few studies focused on real-time mental stress detection algorithms [164, 207]. It is again difficult to compare the results of different mental stress detection studies due to the diversity of methods used, however regardless of the task environments and system setups, the mean accuracies obtained are at or above 80% in most studies [37, 38, 161, 163, 166-170, 173, 175, 183, 184, 185-188, 189, 191, 208-210].

1.4 Practical challenges with passive BCI research

Passive BCI research has been growing rapidly in recent years. However, this promising field is still in its infancy and there is still a long way to go due to some significant challenges that limit transferability of research from lab to real life applications. Two such challenges are:

- Amount of calibration data needed: Most preliminary studies that investigate the ability to classify different mental states are done offline, with the classification accuracy estimated using cross-validation. As mentioned, cross-validation reduced the bias of the performance estimate as compared to a single train/test split, and provides a more accurate estimate of how the classifier will perform in general when used to predict the class of new/unseen data. It usually takes several hours to collect the whole dataset in a single experimental session, and typically at least 80% of the data is used in the training set in each step of the cross-validation (i.e., k-fold cross-validation where $k \geq 5$). Moreover, most studies use "subject-specific" classifiers; that is, the training and testing data all come from the same person, usually collected in a single session. While this approach is reasonable and informative for initial offline analysis, it is not directly transferable for implementation in real, online BCI applications since long calibration sessions would be required to collect comparable amounts of training data from the user to calibrate the system each time they wish to use it. Being able to use subject-independent classifiers where the system would be pre-trained using data collected from a separate group of individuals, or even subject-specific classifiers where the calibration data is pre-collected on a different day, would mitigate this issue. However, due to significant inter-subject differences in EEG signals and in the neural responses to different tasks and stimuli, and even inter-session differences within the same individual, it is much more challenging to achieve sufficient BCI performance this way. In fact, the underlying EEG sensor signals are always non-stationary if they are sampled from different experimental sessions or subjects [211, 212]; which results in the deterioration of the classification performance. Therefore, it is often not possible to reliably classify EEG patterns across subjects with conventional classifier approaches used in offline BCI studies. To mitigate the adverse effects of the non-stationarity of EEG sensor signals, the use of transfer learning approaches has been proposed [213]. In general, transfer learning algorithms allow knowledge learned in one domain (e.g., data collected from one group of subjects) to be transferred into a different but related domain (e.g., data collected from another subject) [213]. There are two main advantages of transfer learning in EEG signal analysis. 1) Match individual difference: When collected from different subjects or from the same subject on different days, the difference between the training and testing data is huge [214], which increases the difficulty in analysis as explained before. Transfer learning can make adjustments so that the model flexibly matches the data collected from different individuals and/or through different tasks. There have been several recent BCI studies proposing such algorithms that are adaptive to different subjects and individuals [215, 216]. 2) Reduce data requirement: In EEG signal analysis, problems of data scarcity and insufficient labeling hinder the learning of the target task [217, 218]. Transfer learning methods learn the target task according to a priori knowledge learnt in a similar domain enhanced by a small amount of data in the target domain to adjust the classifier, which reduces the requirements for available data. Different transfer learning models are proposed to solve the problem of small training datasets in various applications [219, 220]. There are four transfer learning methods commonly used in EEG signal analysis called: 1) domain adaptation, 2) improved common spatial patterns algorithms, 3) deep neural networks, and 4) subspace learning. For a throughout review of each method see [213].

- Moving from controlled to uncontrolled environment: Another very significant challenge in transferring passive BCI research from the lab to real-life scenarios is related to the introduction of potentially confounding factors that are typically absent in the experimental condition but cannot be controlled in real environments. For example, most passive BCI experiments are done in controlled environments and involve very specific and prescribed tasks or situations, whereas in real-life scenarios the range of activities in which the BCI user may be engaged is much more variable. Another example is that in most experiments, participants are shielded (at least to a certain extent) from environmental noise and other distractions and are very limited in their task-unrelated talking and moving; this would not be the case in a natural setting. The addition of these different factors that were not accounted for in the experiments will almost certainly affect the individual's neural signals, and therefore negatively impact the transferability of the experimental results.

1.5 Motivation for the proposed work

As mentioned, a significant challenge exists in transferring passive BCI research from the lab, where conditions are typically very controlled, to practical applications where they are much more variable. Another important example of this is that typically in pBCI studies, the mental states of interest are considered in isolation, which is very different from real-life situations where there could be the simultaneous coexistence of many mental and emotional states. In real life, different aspects of the user's state are likely to be changing simultaneously - for example, their cognitive state (e.g., level of mental workload) and affective state (e.g., level of stress/anxiety). This inevitable confounding of different states needs to be accounted for in the development of state detection algorithms in order for them to remain effective when taken outside the lab. This would be of particular concern for mental workload and stress, since these two states are both highly relevant for many situations for which passive BCIs would be useful, such as high risk work environments where individuals often perform duties in stressful situations that carry a high cost of error, and also because they are known to be highly related to one another. It has been discussed in sections 1.3.1.2 and 1.3.2.2 that EEG is sensitive to both mental workload and stress, thus the simultaneous experience of these states by the user could affect the distribution of the physiological variables on which the individual mental workload and stress classification algorithms are based. If the mental state detection algorithms are developed considering each state alone, the performance could potentially suffer when taken outside the lab where other mental states cannot be controlled.

Indeed, the only study specifically investigating the effect of stress state on automatic mental workload detection found that resulting classifiers failed to generalize well across different stress conditions. Muhl et al. investigated the effect of stress on mental workload detection and found that classification performance suffered when training and testing data came from different stress conditions [117]. Also, more recently, Grissmann et al. assessed the impact of affective valence (i.e., positive or negative emotion) on the classification of working memory load (WML) [221]. Their results indicated that even though WML could be automatically detected with good classification accuracies over 70% in presence of contextual changes, the affective context could significantly affect classification accuracy. These findings suggest a need to develop mental workload detection algorithms that are robust to affective state. The impact of varying mental workload level on automatic stress detection has not yet been investigated.

While detection of both task difficulty and stress level [117, 221] from neural signals has been attempted before on an individual basis, estimating both states simultaneously, with each confounding the other, is a challenge that has not yet been addressed. I aim to address this research gap in my thesis. The long term goal of this work is the realization of a functional, reliable, and robust passive BCI capable of detecting both the mental workload level and stress level of the user simultaneously, since having this more detailed information would provide a better picture of the user's mental state, and their risk of error. Such a system could have significant impact by helping to prevent industrial accidents and their associated human, economic, and environmental costs.

1.6 Research objectives

The motivation of this research is to work toward the development of an online EEG-based passive BCI that is able to accurately detect both the level of difficulty of the task the individual is performing and their stress/anxiety level simultaneously. To this end, the main objectives of this thesis are to:

1) Investigate if/how mental workload (as modulated by task difficulty) and stress confound one another, and the implications for the development of reliable automatic detection algorithms for each state using EEG. Specifically, a) the effect of varying stress level on the ability to classify mental workload, and b) the effect of varying mental workload on the ability to classify stress, are explored.
2) Investigate classification techniques to mitigate any confounding effects of variation in mental workload on the detection of stress, and vice versa, thus making classification of each state more robust.
3) Develop a classification pipeline for the simultaneous classification of mental workload and affective state that is suitable for an online BCI.

1.7 Thesis organization

The current chapter has presented relevant background information on brain-computer interface development. Chapters 2 through 4 of this thesis each represent a version of an independent manuscript that has been published in a peer-reviewed journal. Each details a distinct study that addresses one of the three thesis objectives and associated research questions, as noted in section 1.6.

Chapter 2 has been published in the Journal of Neural Engineering. This study investigated whether the ability to detect one mental state via EEG signals is affected by variation in another state. Specifically, the effect of varying affective state on the automatic classification of mental workload, and the effect of varying mental workload on the automatic classification of affective state, was explored. This chapter addresses research objective 1 from section 1.6.

Chapter 3 has been published in the journal Brain Computer Interfaces. This study investigated different classification approaches to improve the performance of 1) EEG-based mental workload detection in the presence of variation in affective state and 2) EEG-based affective-state detection in the presence of variation in mental workload level by explicitly considering this variation in the development of the classification algorithms. This chapter addresses research objective 2 from section 1.6.

Chapter 4 has been published in the journal Sensors. This study investigated the ability to classify both mental workload level and affective state simultaneously using methods appropriate for implementation in an online BCI. This chapter addresses research objective 3 from section 1.6.

The final chapter, Chapter 5, summarizes the major original contributions of this thesis and discusses potential avenues for future work.

Because this thesis takes the manuscript-style of presentation the introduction and/or methods sections of some chapters may contain repeated information that the reader may wish to skip as they see fit. Permission to reproduce articles in this thesis was obtained as necessary.

Chapter 2 : EEG-based detection of mental workload level and stress: the effect of variation in each state on classification of the other

Co-authorship statement. A version of this chapter has appeared in the Journal of Neural Engineering as the article titled "EEG-*based detection of mental workload level and stress: the effect of variation in each state on classification of the other"* in October 2020. The author, Mahsa Bagheri, carried out the study design, data collection and data analysis. Mahsa Bagheri also prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-author's feedback as well as the comments received from the peer review process. The co-author, Dr. Sarah Power provided guidance on study design, data collection and data analysis. Dr. Power also helped in reviewing, editing and revising the manuscript. All authors read and approved the final draft.

2.1 Abstract

Objective. A passive brain-computer interface (pBCI) is a system that continuously adapts human-computer interaction to the user's state. Key to the efficacy of such a system is the reliable estimation of the user's state via neural signals, acquired through non-invasive methods like electroencephalography (EEG) or function near-infrared spectroscopy (fNIRS). Many studies to date have explored the detection of mental workload in particular, usually for the purpose of improving safety in high risk work environments. In these studies, mental workload is generally modulated through the manipulation of task difficulty, and no other aspect of the user's state is taken into account. In real-life scenarios, however, different aspects of the user's state are likely to be changing simultaneously - for example, their cognitive state (e.g., level of mental workload) and affective state (e.g., level of stress/anxiety). This inevitable confounding of different states needs to be accounted for in the development of state detection algorithms in order for them to remain effective when taken outside the lab. ***Approach.*** In this study we focused on two different states that are of particular importance in high risk work environments, specifically mental workload and stress, and explored the effect of each on the ability to detect the other using EEG signals. We developed an experimental protocol in which participants performed a cognitive task under two different levels of workload (low workload and high workload) and at two levels of stress (Relaxed and Stressed) and then used a linear discriminant classifier to perform classification of workload level and stress level independently. ***Main results.*** We found that the detection of both mental workload level (e.g., low workload vs. high workload) and stress level (e.g., Stressed vs. Relaxed) were significantly diminished if the training and test data came from different as opposed to the same level of the other state (e.g., for mental workload classification, training on data from a Relaxed condition and testing on data from a Stressed condition, rather than both training and testing on the Relaxed condition). The reduction in classification accuracy observed was as much as 15%. ***Significance.*** The results of this study indicate the importance of considering multiple aspects of a user's state when developing detection algorithms for pBCI technologies.

2.2 Introduction

2.2.1 Problem statement

A passive brain-computer interface (pBCI) is a system that enriches human-machine interaction by providing implicit information on a user's mental (e.g., cognitive, affective) state and adapting the environment accordingly [1]. One potential application that has received particular attention in the pBCI research community is the monitoring of mental workload [20, 40, 88, 94, 95, 114, 144, 155, 222-224], particularly for safety-critical occupations like pilots, air-traffic controllers, and other industrial operators. The overarching goal of a pBCI, then, is to derive a continuous, objective estimation of the cognitive load on an individual from neural signal variables, so that appropriate adaptation strategies can be engaged to reduce the potential for error during periods of extreme demand or overload. Such a technology could have significant industrial and economic impact by preventing accidents related to operator error, and their associated human, economic, and environmental losses. Mental workload detection also has great potential value in other domains, including gaming [225], adaptive training [99, 226], and user interface design [227], to enhance and personalize user experience.

The first step in realizing such a technology is developing methods for the accurate detection of mental workload derived from neural signals. While functional magnetic resonance imaging (fMRI) is the gold standard in functional imaging, the high cost, along with the practical limitations of the imaging process, make it unsuitable for use in pBCI technologies, which demand that the sensing technology be relatively inexpensive, highly portable and non-intrusive. Therefore, because of their practical advantages, both electroencephalography (EEG) and function near-infrared spectroscopy (fNIRS) have emerged as the leading neural imaging modalities used in pBCI research, and indeed in BCI research generally [228-232].

Mental workload is defined as the perceived relationship between an individual's total mental processing capability and the amount required by the task at hand [233]. The closer the requirements are to the actual capabilities, the higher is the perceived workload. Therefore, a common strategy for workload manipulation is the variation of task difficulty. There have been many studies to date investigating the ability to detect mental workload levels due to the performance of tasks of varying difficulty using EEG [34, 69, 81, 83, 89, 90, 94, 105, 106, 116, 234] and NIRS [235]. Though results vary across studies depending on the specifics of the tasks investigated (e.g., n-back task [81, 224], mental arithmetic [83, 235], flight simulator task [89, 90], air traffic control task [34, 69, 116]), and the classification methods used (e.g., type and number of features, duration of signal used, type of classifier, size of data set, etc.) generally they are, in aggregate, quite promising, with two or more levels of workload consistently being classified with accuracies significantly exceeding chance.

Studies on mental workload detection generally consider workload exclusively in terms of task demands as manipulated through the variation of task difficulty. However, precise estimation of mental workload is not possible when exclusively considering the properties of the task because individual factors will affect the mental effort needed to perform the task [236]. We argue that the individual's affective state, and particularly stress/anxiety, is of particular relevance in mental workload detection, particularly those aimed at safety-critical occupations where individuals often perform duties in stressful situations that carry a high cost of error. According to the Processing Efficiency Theory [237], stress reduces the storage and processing capacity of the working memory system, and necessitates an increase in on-task effort to maintain performance. Thus, a task performed under stress may require considerably more effort to achieve a certain performance than it would under relaxed conditions (i.e., performance effectiveness may be maintained, but at the expense of reduced cognitive efficiency).

Physiologically, EEG and NIRS have both been shown to be sensitive to emotion [238, 239]; thus, anxiety could impact not only the level of mental workload experienced by the user during task performance, but also the distribution of the physiological variables on which the mental workload classification algorithms are based, suggesting that performance of detection algorithms developed considering task difficulty alone will suffer outside the lab where affective context cannot be controlled. Indeed, the only study specifically investigating the effect of affective state on automatic mental workload detection found that classification performance suffered when training and testing data came from different stressed conditions [117]. This suggests a need to develop mental workload detection algorithms that are robust to affective state. Beyond addressing the technical implications for workload detection, however, we argue that having an objective, real-time measure of the affective state of the individual would in itself be very valuable. Stress is known to negatively impact cognitive efficiency and performance [156] as well as decision-making (especially when performing unfamiliar tasks) [157], thus having information about the individual's stress level would be valuable when determining the adaptation strategies to be initiated by the pBCI. While detection of both task difficulty and stress level [167, 194] from neural signals has been attempted before on an individual basis, estimating both states simultaneously, with each confounding the other, is a challenge that has not yet been addressed.

This study represents the first step toward the long term objective of developing a pBCI that performs simultaneous detection of mental workload and stress level. In this work, we investigate if/how each of the two states confounds the other, and the implications for the development of reliable automatic state detection algorithms. Specifically, the objective here is to investigate: 1) the effect of varying stress level on the ability to classify mental workload 2) the effect of varying mental workload level on the ability to classify stress. For this purpose, EEG signals were used due to the excellent temporal resolution and relatively low cost as compared to other methods such as fNIRS, which has high temporal latency of the measured response and higher cost.

2.3 Background

2.3.1 EEG correlates of mental workload

Mental workload is defined as the interaction between task demands and the capacity of the operator [80]. The Multiple Resources Theory [240, 241] posits that performing different tasks necessitates a subject to tap into a set of separate resources, which are both limited in capacity and allocatable amongst different tasks. These resources are defined along four dimensions: the information processing stage (perception or cognition vs. response), the processing code (verbal vs. spatial), the input modality (visual vs. auditory), and the visual channel (focal vs. ambient) [241, 242].

Mental workload arises from the interaction of a number of factors including the properties of a task, the task environment, and the individual characteristics and abilities of the individual [243]. More specifically, it involves the interaction between 1) the requirements of a task such as task difficulty, 2) the circumstances under which it is performed such as environmental noises and distractions, and 3) the skills, behaviors and perceptions of the operator such as emotional instability, fatigue and motivation [243]. Therefore, mental workload can be modulated by adjusting a subgroup of these factors while controlling for the rest [224].

Most studies of mental workload based on brain function have used electroencephalography (EEG) as the imagine modality [244]. The sensitivity of the human EEG to changes in mental effort has been known since Hans Berger (1929) reported a decrease in the amplitude of the alpha rhythm of the EEG during mental arithmetic. Due to its very high time resolution, subtle changes in mental states like alertness, attention and workload can be accurately captured via EEG. Many studies conducted in laboratory, simulation and operational environments have demonstrated significant correlations between EEG indices of cognitive state changes and performance [107, 245-250].

EEG spectral power at different channel locations has been investigated in many studies of mental workload [108, 236, 251-257]. For example, theta band power changes at the frontal midline channels have been found to be linked to the development of mental workload [255, 256]. Alpha band power changes over the centro-parietal and parietal areas have been shown to be sensitive to mental workload, and mental effort in attentive stimulus processing and expectancy [49, 51, 119, 257-259]. Various indices based on beta band power and/or the ratio of beta band power to either alpha or theta band power have also been investigated [260]. Among these studies, the majority used a classical cognitive paradigm with repetitive stimuli [258, 49, 260-262] which while useful for facilitating subsequent data analysis (e.g., event related potential (ERP) analysis [108, 263]), limit their applicability to the investigation of intrinsic dynamics of mental factors in real-world tasks.

Alternatively, real-world task situations (e.g., driving, aircraft landing and takeoff, etc.) varying in levels of difficulty or load have been employed in some studies [46, 68, 69, 236, 250-252, 263, 264].

2.3.2 EEG correlates of stress

Hans Selye popularized stress as a medical scientific idea in 1926, but the way in which the term stress has been used in the literature has not been consistent [265]. Cohen defined it as "a process in which environmental demands tax or exceed the adaptive capacity of an organism, resulting in psychological and biological changes" [266]. In practice, it is a general term referring to a wide range of negative emotional states including unhappiness, anxiety, agitation, frustration, irritability, anger, and overstimulation. While some studies suggest there may be some positive effects of stress [267] that may vary based on individual characteristics [268], in general, it can be said that stress is a state of negative valence and positive arousal. From a biological perspective, stress is a physiological reaction that occurs in response to a stressor, i.e., a perceived harmful event, attack, or threat to survival [269]. Bodily changes such as increased heart rate, blood pressure, breathing rate, and perspiration are symptoms of what is called the "stress response", which helps the body to adapt to the stressor [270]. In general, although many parts of the body are affected by stress, the hippocampus [271], the amygdala, and the prefrontal cortex [272] are the main brain regions with a critical involvement in the stress response.

EEG is a reliable tool reflecting upper cognitive functions and mental or psychological states. EEG has been shown to be sensitive to localized brain activity in areas responsible for the stress response, or activity associated with increased arousal or specific psycho-emotional states [200]. The left anterior region of the brain appears to be involved in the expression and experience of approach-related emotions and the right anterior region appears to be involved in the expression and experience of avoidance-related emotions [273]. Lewis et al. reported a shift from greater left frontal activity during low stress to greater right frontal activity during high stress [179]. Moreover, studies show that positive moods or reactions are related to relatively greater left prefrontal activity, whereas, negative moods or reactions are related to relatively greater right prefrontal activity [203]. Seo et al. reported higher right prefrontal activity in stressful contexts, compared to left prefrontal activity [202].

Many studies have explored power spectrum or relative power indices to assess the behavior of the human brain in emotional stressed states [56, 184, 204-206]. Higher spectral activity has been correlated with arousal, cognitive processing or emotional activity [178]. Many studies investigating the effect of stress on EEG power reported a decrease in the alpha power and an increase in the beta power in the presence of stressors [45, 168, 169, 193, 195-197]. The frontal/prefrontal cortex appears to play a significant role in emotional processing, presenting increased prefrontal theta activity in reduced anxiety states [174]. Minguillon et al. [197] suggest that relative gamma power is more discriminative between stress levels than alpha asymmetry, theta, alpha, beta, or gamma power in the prefrontal cortex, with a positive correlation with stress level. More recent studies suggest that EEG correlates of stress may display even greater frequency-specificity. Alonso et al. [193] utilized two different stressors, the Stroop test and sleep deprivation to assess stress using EEG. They found a decrease in high alpha (11 to 12 Hz) power and an increase in high beta (23 to 36 Hz) power for both stressors, whereas the low alpha (7.5 to 10.5 Hz) indicated an increase of attention and alertness for the Stroop test and a decrease for the sleep deprivation. The theta activity clearly increased for the sleep deprivation case but no significant changes were obtained for the Stroop test [193].

2.3.3 Relation to previous work

The influence of mental state changes during training and testing on active BCI (i.e., BCI in which mental states are intentionally generated by the user for the purpose of controlling an external device) performance has been examined in some studies. Surprisingly, Reuderink et al. reported a significant performance increase during frustrating vs. relaxed periods when using a BCI based on event-related desynchronization (ERD) [275]. A study investigating the effects of fatigue on EEG signal characteristics and workload classification performance reported a decline in classification performance with increasing fatigue [44]. To date only a few studies have investigated the influence of mental state on BCI performance, however, it has been found that BCIs are susceptible to changes in task-unrelated mental states (e.g., attention, fatigue or mood) during classification.

Also, it is generally accepted that one of the most challenging issues in brain computer interface development is the non-stationarity of brain signals due to internal sources of variability such as task-unrelated mental states [117, 211].

One previous study has specifically investigated the effect of stress on the automatic detection of mental workload via EEG. In their paper, Muhl et al. [117] designed a study in which 24 participants performed trials of a visual n-back task at two different levels of difficulty (specifically, 0-back and 2-back) in both a relaxed and a stressed condition. They found that the accuracy with which they could classify the workload levels was affected by whether the training and testing data came from the same or different stressed conditions. When the classifier was trained on data from the relaxed (stressed) condition and tested on data from the relaxed (stressed) condition – i.e., the "within-state" case - mean accuracy was approximately 72%. However, when the classifier was trained on data from the relaxed (stressed) condition and tested on data from the stressed (relaxed) condition – i.e., the "across-state" case - the mean accuracy dropped to approximately 69%. It was found that this effect could be mitigated to some extent by including data from both stressed conditions in the training set. The results summarized here were for frequency domain EEG signal features, but they observed similar trends for time-domain features (i.e., ERPs) as well.

The study by Muhl et al. provided a thorough and rigorous investigation of the effect of stressed condition on the classification of mental workload using EEG signal features, but did not include a similar analysis of the effect of mental workload condition on the classification of stress. Like mental workload, the automatic detection of stress based on neural and physiological (e.g., ECG) signals is a very active research area (for a thorough review, see [276]). Some studies use emotionally salient music [208], videos [183], or pictures [209, 210] to induce stress under a no-task condition. Such studies do not consider that when the proposed stress detection system is taken outside the lab for use in real-life scenarios, the user would likely be performing different tasks and activities, and experiencing the associated changes in cognitive load, which could reduce the reliability of the stress detection algorithm developed under a single workload (or no load) context. We aim to address this research gap in the current study.

Another limitation noted in [117] was that the stress induction paradigm manipulated affective context only once; that is, one block of workload trials was completed under one affective context, followed by a second block of trials completed under the other affective context. Therefore, the stress manipulation was synonymous with a change in time, and the possibility that the observed results were due simply to time-related signal changes cannot be excluded. The non-stationarity of EEG is indeed a well-documented phenomenon [211, 212]. To ensure the validity of the results, the experiment needs to be repeated using interleaved stressed conditions. For the present study, we devised a stress induction protocol that was similar to that used in [117], but that incorporated this counterbalancing of the stressed conditions, allowing us to rule out the possibility that the results were merely due to the effect of time.

2.4 Material and methods

2.4.1 Participants

Eighteen right-handed subjects (mean age: 26 ± 8 years; 7 females) participated in this study. Participants were excluded if they 1) were not between the ages of 18 and 65 years; 2) had a history of neurological disease, disorder or injury, or cognitive impairment; 3) did not have normal or corrected-to-normal vision and hearing. Subjects were asked to refrain from exercise, smoking, or consuming caffeine or alcohol for at least 4 hours prior to the session. Informed consent was obtained from the participants and documented in writing prior to the experiment. The study was approved by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) at Memorial University of Newfoundland (approval #20190461-EN, Appendix 1).

2.4.2 Physiological signal acquisition

EEG data was recorded via a 64-channel electrode system (ActiCHamp, Brain Products, GmBH). The position of the electrodes on the scalp was based on the International 10-20 system for EEG electrode placement. The reference electrode was at FCz. The impedance of recording electrodes was monitored for each subject prior to data collection and was kept below 10 kΩ. Electrocardiogram (ECG) data was collected using three electrodes connected to an auxiliary channel of the EEG amplifier. Both the EEG and ECG signals were recorded at a sampling rate of 500 Hz.

2.4.3 Experimental procedure

A protocol was designed in which subjects performed a cognitive task at two different levels of mental effort (Easy and Difficult), each under two different affective contexts (Relaxed and Stressed). The workload task was designed in the MatLab Cogent2000 toolbox. The structure of the experimental session is depicted in Figure 2.1.

The experiment consisted of four main blocks, which alternated between a “Relaxed” and a “Stressed” condition. For half of the participants the order of the blocks was R1-S1-R2-S2 while for the remainder the order was S1-R1-S2-R2. At the beginning of each block, an affective state induction protocol was performed to induce the desired state, and the participant’s subjective rating of their stress level was then captured via form Y-1 of the State-Trait Anxiety Inventory (STAI) questionnaire [277]. Each block then consisted of 10 trials: four eyes-open “Baseline” trials, three

“Easy” arithmetic trials, and three “Difficult” arithmetic trials. Trials within a block were in pseudo-random order (blocks always began with a Baseline trial, and no two consecutive trials were of the same type), and the order was different for each block. After each trial, subjects were asked to rate the level of mental effort required to perform the trial via a modified version of the Rating Scale of Mental Effort (RSME) [278] (see Figure 2.1). All trials were approximately 67 seconds in duration. Other than the stress induction measures, there was no difference between the four blocks in terms of the workload trials. To capture any change in the perceived affective state over the course of the workload trials, the participant again completed form Y-1 of the STAI questionnaire at the end of each block.

At the beginning of the experiment (i.e., before Block #1), an eyes-closed baseline trial of 67 seconds duration was collected.

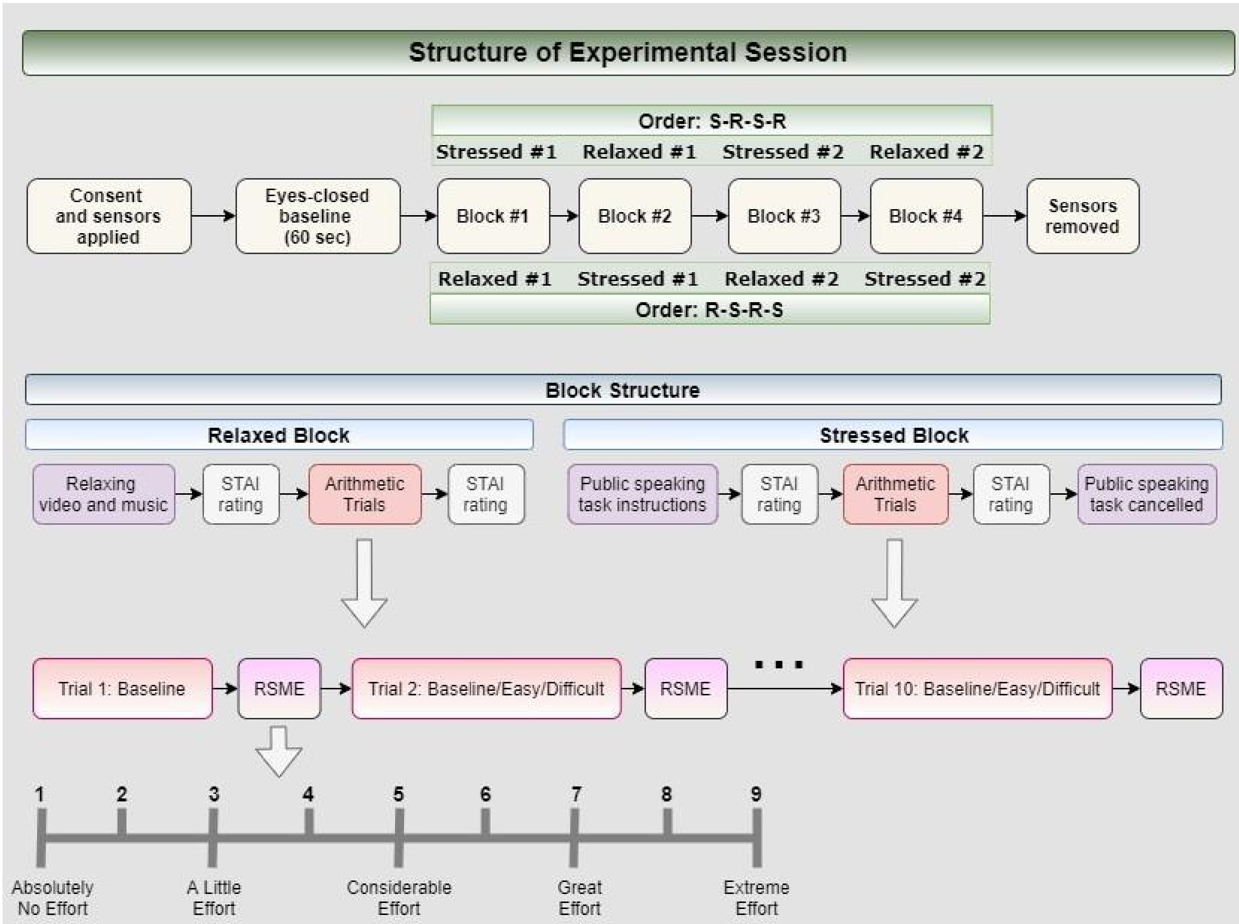


Figure 2.1: Structure of the experimental session.

2.4.3.1 Workload trials

For the Baseline trials, participants were asked to focus their eyes on a cross that appeared in the centre of the screen and to sit quietly for the duration of the trial. No arithmetic task was performed.

Both the Easy and Difficult arithmetic trials involved answering math equations in the form "num1 modulus num2 = num3". When each trial began, a single equation in this form appeared on the screen and participants were asked to indicate if the equation was correct or incorrect by pressing one of two keyboard buttons (the buttons were the same for every trial). The equation stayed on the screen until the participant responded with a key press, then a different equation would appear. No time restriction was put on the participant's response, because time pressure is known to elicit stress [169, 279] and we wanted to decouple stress induction and workload induction as much as possible. The subject's performance (% correct = # correct/# completed) was displayed in the right corner of the screen and updated after each response. The number of equations completed in a given trial varied based on how quickly the participant responded to the questions. No equation was presented more than once during the experimental session (i.e., all equations from all trials were unique).

For the Easy arithmetic trials, num2 in the equation above was restricted to either 2, 5 or 10 (e.g., "20 mod 5 = 0" would be correct, "32 mod 2 = 1" would be incorrect), since division by 2, 5, and 10 is relatively simple and intuitive for most people. For the Difficult arithmetic trials, num2 was either 3, 4, 6 or 9 since division by these numbers is generally quite difficult in comparison to the "Easy" condition. The "Easy" and "Difficult" trials represent the "low workload" and "high workload" conditions, respectively.

2.4.3.2 Induction of affective states

The stress induction protocol used in [117] was based on the Trier Social Stress Task (TSST) [162], and they induced the relaxed state by having participants either relax in silence or while listening to calming music. The TSST is a very effective and reliable method of inducing stress and anxiety [280, 281], however it includes only one instance of the stressed condition. As previously mentioned, it is critical for the research questions we are attempting to answer in this work that the stressed conditions be interleaved. We thus developed a stress induction protocol, also based on the TSST, which allowed us to introduce the Stressed conditions twice, while allowing participants to return to the Relaxed state in between.

At the beginning of the session, the participant was not given specific information about the tasks they would have to perform during the experiment but rather were just informed that they would have to perform a series of different tasks, some of which were meant to be relaxing and others stressful. As will be explained, the protocol involved some deception and this vagueness was intentional, so that the participant would not be able to predict what would happen later in the session.

To induce a Relaxed state prior to the Relaxed Blocks, participants sat comfortably and watched a 2-minute video with relaxing imagery and music. They were then told that they would have to perform a series of arithmetic trials of varying difficulty. Before beginning the workload trials, the participant was told that while they should try to perform the arithmetic task as well as possible, their performance would not be saved or compared to other participants. The same protocol was followed for both Relaxed Blocks #1 and #2.

To induce a Stressed state for Stressed Block #1, the subjects were told that the first task would be to perform a series of arithmetic trials at varying levels of difficulty, and that their performance during the arithmetic trials was extremely important to the study results and would be recorded and compared to other participants. They were then told that following the arithmetic task, they would have to do a public speaking task. They were told that this task involved giving a 10-minute presentation on a topic of the experimenter's choosing to a small panel of evaluators (the study's principal investigator and some of her colleagues), one of whom was a body language expert. They were told that they would be given five minutes to prepare, and that they could use the internet to research the topic and could make notes. They were also told that their presentation might be video recorded for future analysis. To increase anxiety, the experimenter made a "fake" phone call to the principal investigator letting her know that the arithmetic task was about to begin and that the panel would be needed in about 15 minutes. After completing the STAI questionnaire, the participant completed the arithmetic trials for Stressed Block #1. Thus in this protocol, it was the anticipation of having to perform the public speaking task that induced stress during the arithmetic trials.

So that the experimental session was not impractically long (and since, again, it was the anticipation of the public speaking task that induced the Stressed state) we did not require the participant to actually do the public speaking task. After completing the workload trials of Stressed Block #1 and completing the STAI questionnaire, participants were told that "based on their physiological data and their performance during that block", we did not require them to do the public speaking task after all. This allowed them to return to a Relaxed state for the next block (which was either Relaxed Block #1 or #2, depending on the participant's block order).

At the beginning of Stressed Block #2, the participant was again told that they would have to do the more stressful arithmetic task (where their performance was highly important to the study outcome, and would be monitored and compared with others) followed by the public speaking task, but this time they were told that we needed the data from this task regardless of their performance/data in the arithmetic task. Because of the deception used in Stressed Block #1, there was a risk that the stress induction would be less effective for Stressed Block #2 due to the participant doubting our truthfulness. To increase believability, this time the experimenter made a "real" phone call to the principal investigator, and spoke to her on speaker phone so that the participant could hear the arrangements regarding the panel being made. After completing the arithmetic trials for Stressed Block #2 and completing the STAI questionnaire, the participant was again told that they would not actually have to do the public speaking task, and it was explained why the deception used was necessary for the objectives of the study.

Note that for both Stressed Blocks #1 and #2, the STAI questionnaire was completed immediately after the participant was told about the public speaking task, and then again before they were told they did not actually have to do it (see Figure 2.1).

2.4.4 Data analysis

2.4.4.1 Validation of affective state induction protocol

In order to validate the affective state induction protocol, we analyzed both subjective data (via the STAI questionnaire), and objective data (via the participant's heart rate calculated from the collected ECG signals). Our hypothesis was that if our protocol was effective, both the heart rate and STAI questionnaire scores would be significantly higher for the Stressed Blocks than the Relaxed Blocks.

Heart rate was calculated for each trial from the ECG signals. We performed a two-way repeated measures ANOVA with within-subject main factors of affective state (two levels: Relaxed and Stressed) and workload (two levels: Easy and Difficult) for the averaged-over-condition heart rates. We also considered the interaction of the main effects.

Unlike the other measures, the STAI scores were only available for each block, rather than for each trial, and so we could only test the effect of the affective state (not workload) on this variable. To make sure that the protocol was effective for inducing stress in both Stressed Block #1 and Stressed Block #2, we performed a two-way repeated measures ANOVA with within-subject main factors of affective state (two levels: Relaxed and Stressed) and block (levels: Block #1 and Block #2).

2.4.4.2 Validation of workload level induction

Similarly, in order to validate that our arithmetic task induced different levels of workload, we analyzed both subjective data (via the RSME ratings), and objective data (via the participants' performance in the arithmetic trials, specifically their response times and response accuracies). Our hypothesis was that if our protocol was effective, the RSME ratings and reaction times would be significantly higher, and the response accuracies significantly lower, for the Difficult arithmetic trials than the Easy arithmetic trials.

We performed two-way repeated measures ANOVAs with within-subject main factors of affective state (two levels: Relaxed and Stressed) and Workload (two levels: Easy and Difficult) for the averaged-over-condition RSME ratings, response times, and response accuracies. We also considered the interaction of the main effects.

2.4.4.3 EEG-based classification of mental workload and affective state

We aim to classify the level of mental workload as well as the affective state, and determine the effect of each on the other. First, we investigated the effect of variation in affective state on the ability to classify levels of mental workload (i.e., Easy vs. Difficult). We considered three different classifier training paradigms: 1) within-affective-state classification (classifiers were trained and tested on data from the same affective state), 2) across-affective-state classification (classifiers were trained on data from one state and tested on data from another state), and 3) combined-affective-state classification (classifiers were trained on data from both states combined but tested on data from a single state). All possible combinations of training and test data for each training paradigm were investigated.

Next, classification of affective state (i.e., Relaxed vs. Stressed) was examined for three different classifier training paradigms: 1) within-workload-level classification (classifiers were trained and tested on data from the same workload level, 2) across-workload-level classification (classifiers were trained on data from one workload level and tested on data from the other workload level) and 3) combined-workload-level classification (classifiers were trained on data from both workload levels combined, but tested on data from only one level). All possible combinations of training and test data for each training paradigm were investigated. Please note that in all cases, within-subject classification was performed.

For each classification problem of interest, the following steps had to be taken prior to classification: preprocessing, feature extraction, feature selection, and classification. Then statistical analysis was performed on the results.

2.4.4.3.1 EEG pre-processing

Since EEG signals contain noise and artifacts from several sources such as the subject's body movements, eye blinks, cardiac signals or muscle contractions, pre-processing methods need to be applied to remove these artifacts and produce clean data. First, we applied a band-pass filter with a low cut-off frequency of 1 Hz to remove the DC components of the signals and also the drifts, and a high cut-off frequency of 50 Hz. In the next step, EMG and motion artifact contaminated segments of the signals were manually rejected. Finally, independent component analysis (ICA) [282] was applied and artifact components such as eye blinks and saccades were identified and removed. Next, the data was down sampled from 500 Hz to 256 Hz. Pre-processing steps including ICA were applied using the MatLab Toolbox EEGLAB [283, 284]. "Runica" was selected for the ICA type and it was applied to all of a subject's data, which is approximately 45 minutes in length.

2.4.4.3.2 EEG signal feature calculation

Feature extraction aims to minimize the loss of important information embedded in the data and simplify the amount of resources needed to describe a large dataset accurately [18]. Here, frequency domain features of the EEG signals were calculated, specifically the signal power in seven common EEG frequency bands. Frequency bands were calculated with respect to each participant's individual alpha frequency (IAF) [285] as follows: delta from (IAF – 8) Hz to (IAF - 6) Hz; theta from (IAF – 6) Hz to (IAF – 4) Hz; alpha1 from (IAF – 4) Hz to (IAF - 2) Hz; alpha2 from (IAF – 2) Hz to (IAF) Hz; alpha3 from (IAF) Hz to (IAF + 2) Hz; beta from (IAF + 2) Hz to (IAF + 20) Hz; gamma from (IAF + 20) Hz to (IAF + 30) Hz. The IAF for each participant was determined using the eyes-closed baseline trial. Specifically, the frequency with the maximum signal power for the eyes-closed baseline was taken as the IAF. Power signals for each frequency band were obtained via the filter-Hilbert method [286].

To make sure that the workload-level classification was based on differences in neural activity related to the workload levels and not to the incidental differences in the motor requirements of the Easy and Difficult conditions (since response frequency was greater for the Easy condition), we excluded the electrodes over the motor and sensorimotor brain regions for the Easy vs. Difficult classification problems. Specifically, the seven central and seven centro-parietal electrodes (Cz, C1-C6, CPz, CP1-CP6) were excluded. This was not a concern for the Stressed vs. Relaxed classification problems since there was no difference in the motor requirement in the Relaxed and Stressed conditions, and so the central and centro-parietal electrodes were not excluded. This resulted in a total of 63 electrodes x 7 frequency bands = 441 features for affective-state classification problems and a total of 49 electrodes x 7 frequency bands = 343 features for workload-level classification problems.

Each of the four blocks consisted of three 67-second trials of the Easy, and Difficult workload conditions, for a total of 402 seconds per workload level/stress state combination (i.e., 2 blocks per affective state x 3 trials per workload level per block x 67 seconds per trial = 402 seconds). Epochs of 4 seconds in length were calculated for each electrode and frequency band via a sliding window with 50% overlap. This resulted in a total of (approximately, depending on data loss from artifacts) n=200 epochs (or samples) per workload level/stress state combination.

2.4.4.3.3 Feature selection

To have a successful classification algorithm, it is necessary to eliminate redundant and irrelevant features and select a small set of informative features. A greedy forward selection search algorithm was used to select an optimized ten-dimensional feature set. The Fisher score [28, 287] served as the fitness criterion.

2.4.4.3.4 Classification

The ten selected power features were used to train a linear discriminant classifier. The performance of the classifier was assessed using five runs of six-fold cross-validation, performed separately for each participant. In each "fold" of the cross-validation, no test data was used in either feature selection or classifier training.

For the "within-affective-state" and "within-workload-level" classification cases, the conventional cross-validation method was applied. For the "across" and "combined" cases, however, it is not possible to do normal cross-validation, since the training and test data come from completely different datasets. For these cases a "pseudo-cross-validation" technique was applied, as described in Figure 2.2. Note that in the "combined" case, the size of the dataset used for analysis was made equal to the "within" and "across" cases via random sampling of the full dataset.

For all classification paradigms, the classifier performance was estimated as the average classification accuracy across all runs of the cross-validation.

2.4.4.3.5 Statistical analysis

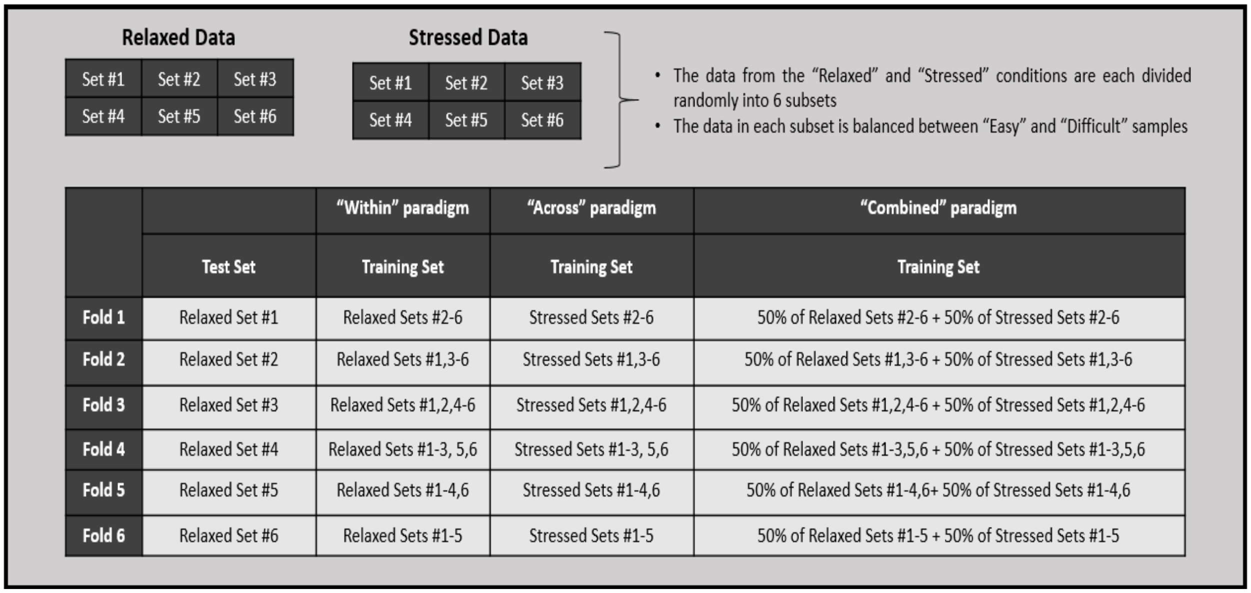
In terms of workload detection, a two-way repeated measures ANOVA was performed considering within-subject main factors of affective state of the test data (two levels: Relaxed and Stressed) and classifier training paradigm (three levels: within-affective-state, across-affective-state, and combined-affective-states). Post-hoc Tukey-Cramer tests were performed to compare each pair of conditions.

In terms of stress detection, for the Relaxed vs. Stressed classification problem, a two-way repeated measures ANOVA was performed with within-subject main factors of workload level of test data (two levels: Easy and Difficult) and classifier training paradigm (three levels: within-workload-level, across-workload-level, and combined-workload-levels). Post-hoc Tukey-Cramer tests were performed to compare each pair of conditions.

2.5 Results

2.5.1 Validation of affective state induction protocol

The ANOVA results indicated a significant main effect of affective context on heart rate ($F_{(1,17)} = 4.43$; $p = 0.05$), with the heart rate being higher in the Stressed condition (mean of 84.6 bpm) than the Relaxed condition (mean of 82.2 bpm). There was no significant difference in heart rate between the Easy and Difficult conditions ($F_{(1,17)} = 0.3$; $p = 0.58$), nor was there a significant interaction effect ($F_{(1,17)} = 0.25$; $p = 0.62$).



		"Within" paradigm	"Across" paradigm	"Combined" paradigm
	Test Set	**Training Set**	**Training Set**	**Training Set**
Fold 1	Relaxed Set #1	Relaxed Sets #2-6	Stressed Sets #2-6	50% of Relaxed Sets #2-6 + 50% of Stressed Sets #2-6
Fold 2	Relaxed Set #2	Relaxed Sets #1,3-6	Stressed Sets #1,3-6	50% of Relaxed Sets #1,3-6 + 50% of Stressed Sets #1,3-6
Fold 3	Relaxed Set #3	Relaxed Sets #1,2,4-6	Stressed Sets #1,2,4-6	50% of Relaxed Sets #1,2,4-6 + 50% of Stressed Sets #1,2,4-6
Fold 4	Relaxed Set #4	Relaxed Sets #1-3, 5,6	Stressed Sets #1-3, 5,6	50% of Relaxed Sets #1-3,5,6 + 50% of Stressed Sets #1-3,5,6
Fold 5	Relaxed Set #5	Relaxed Sets #1-4,6	Stressed Sets #1-4,6	50% of Relaxed Sets #1-4,6+ 50% of Stressed Sets #1-4,6
Fold 6	Relaxed Set #6	Relaxed Sets #1-5	Stressed Sets #1-5	50% of Relaxed Sets #1-5 + 50% of Stressed Sets #1-5

Figure 2.2: Illustration of the six-fold cross-validation techniques used in the "within", "across" and "combined" classifier training paradigms. The example shown is for the "Easy vs. Difficult" classification problem, for test data from the "Relaxed" condition.

The ANOVA results for the STAI scores also showed a significant difference between the participants' perception of the affective states between the Relaxed and Stressed states ($F_{(1,17)} = 14.66$; $p = 0.001$), with the STAI score being higher in the Stressed condition (mean of 36.8) than the Relaxed condition (mean of 28.4). There was no significant effect of Block ($F_{(1,17)} = 3.63$; $p = 0.07$) on STAI score, nor was there a significant interaction effect ($F_{(1,17)} = 0.03$; $p = 0.86$).

2.5.2 Validation of workload level induction

The ANOVA results indicated a significant main effect of workload level on RSME score ($F_{(1,17)} = 82.14$; $p < 0.001$), with the RSME score being significantly higher in the Difficult condition (mean of 2.7) than the Easy condition (mean of 1.2). There was no significant main effect of affective state on the RSME scores ($F_{(1,17)} = 0$; $p = 0.95$), nor was there a significant interaction effect ($F_{(1,17)} = 0.12$; $p = 0.73$).

The ANOVA results also indicated a significant main effect of workload level on response time ($F_{(1,17)} = 21.33$; $p < 0.001$), with the response time being significantly higher in the Difficult condition (mean of 7.6 seconds) than the Easy condition (mean of 4.0 seconds). There was no significant main effect of affective state on the response time ($F_{(1,17)} = 2.54$; $p = 0.1292$). There was a significant interaction effect ($F_{(1,17)} = 5.67$; $p = 0.0292$), indicating that the difference in response time between the Easy and Difficult conditions was more pronounced in the Stressed condition than the Relaxed condition.

The ANOVA results also indicated a significant main effect of workload level on response accuracy ($F_{(1,17)} = 16.07$; $p =<0.001$), with the response accuracy being significantly lower in the Difficult condition (mean of 92.5%) than the Easy condition (mean of 97%). There was no significant main effect of affective state on the response accuracy ($F_{(1,17)} = 0.03$; $p = 0.87$). There was a significant interaction effect ($F_{(1,17)} = 5.28$; $p = 0.03$), indicating that the difference in response accuracy between the Easy and Difficult conditions was more pronounced in the Stressed condition than the Relaxed condition.

2.5.3 EEG-based classification

Table 2.1 shows the detailed workload level classification results (averaged across participants). The results are given for each classifier training paradigm, and for test data from each affective state, separately.

Table 2.2 shows the detailed affective state classification results (averaged across participants). The results are given for each classifier training paradigm, for test data from each workload level separately.

The confidence limits of "chance" depends on the number of trials per class, with the upper confidence limit for a significance level of α=5% and n=200 samples per class being approximately 54.3% by the binomial test. The classification results in Tables 2.1 and 2.2 thus show above chance accuracies for all cases.

In terms of workload classification, the ANOVA results indicated that there was a significant effect of "classifier training paradigm" on classification accuracy ($F_{(2,34)} = 151.67$; $p < 0.001$). Post hoc Tukey-Kramer tests indicated that there were significant differences between all pairwise combinations of the three classifier training paradigms ($t_{(34)} > 7.63$; $p < 0.001$). There was no significant effect of "affective state of the test set" on the classification accuracy ($F_{(1,17)} = 0.62$; $p = 0.44$). There was also a slightly significant interaction effect ($F_{(2,34)} = 3.49$; $p = 0.04$), with the difference in classification accuracy among the three classifier training paradigms being slightly larger when test data came from the Stressed condition as compared to the Relaxed condition.

In terms of affective state classification, the ANOVA results revealed a significant main effect of "classifier training paradigm" on the classification accuracy for Relaxed vs. Stressed ($F_{(2,34)} = 39.36$; $p < 0.001$). Post hoc Tukey-Kramer tests indicated that there were significant differences between all pairwise combinations of the three classifier training paradigms ($t_{(34)} > 4.84$; $p < 0.001$). There was no significant main effect of "workload level of the test set" on the classification accuracy ($F_{(1,17)} = 0.16$; $p = 0.70$), nor was there a significant interaction effect ($F_{(2,34)} = 2.9$; $p = 0.07$).

Table 2.1: Workload level classification results (averaged over all participants) for different classifier training paradigms. For the training set composition, "same affective state" and "other affective state" are relative to the test set composition.

		Mental workload classification results: Easy vs. Difficult		
		Training set composition		
		Same affective state only ("within" paradigm)	Other affective state only ("across" paradigm)	Both affective states ("combined" paradigm)
Test set Compos	Relaxed	71.4 ± 5.2	56.4 ± 6	67.1 ± 5.4
	Stressed	73.8 ± 5.5	56.3 ± 5.5	66.8 ±5.7
Mean:		72.6	56.3	66.9

Table 2.2: Affective state classification results (averaged over all participants) for different classifier training paradigms. For the training set composition, "same affective state" and "other affective state" are relative to the test set composition.

		Affective state classification results: Relaxed vs. Stressed		
		Training set composition		
		Same workload level only ("within" paradigm)	Other workload level only ("across" paradigm)	Both workload levels states ("combined" paradigm)
Test set Compos	Easy	85.8 ± 6.9	75.0 ± 9	82.3 ± 7.3
	Difficult	84.8 ± 6.6	72.6 ± 8.4	82.6 ± 7
Mean:		85.3	73.8	82.4

2.5.4 Feature analysis

In order to get insight on the source of the class differentiability, we investigated the features selected from each frequency band and each brain region via the automatic feature selection algorithm for the "within" classification training paradigm. Figures 2.3 and 2.4 show the percentage of selected features coming from each considered a) brain region, and b) frequency band, for the mental workload and affective state classification problems, respectively. Given that we performed 5 runs of 6-fold cross-validation, and in each fold ten features were selected, the percentages given for each classification problem are out of a total of 300 selected features. Recall that the central and centro-parietal electrodes were excluded for the mental workload classification problems.

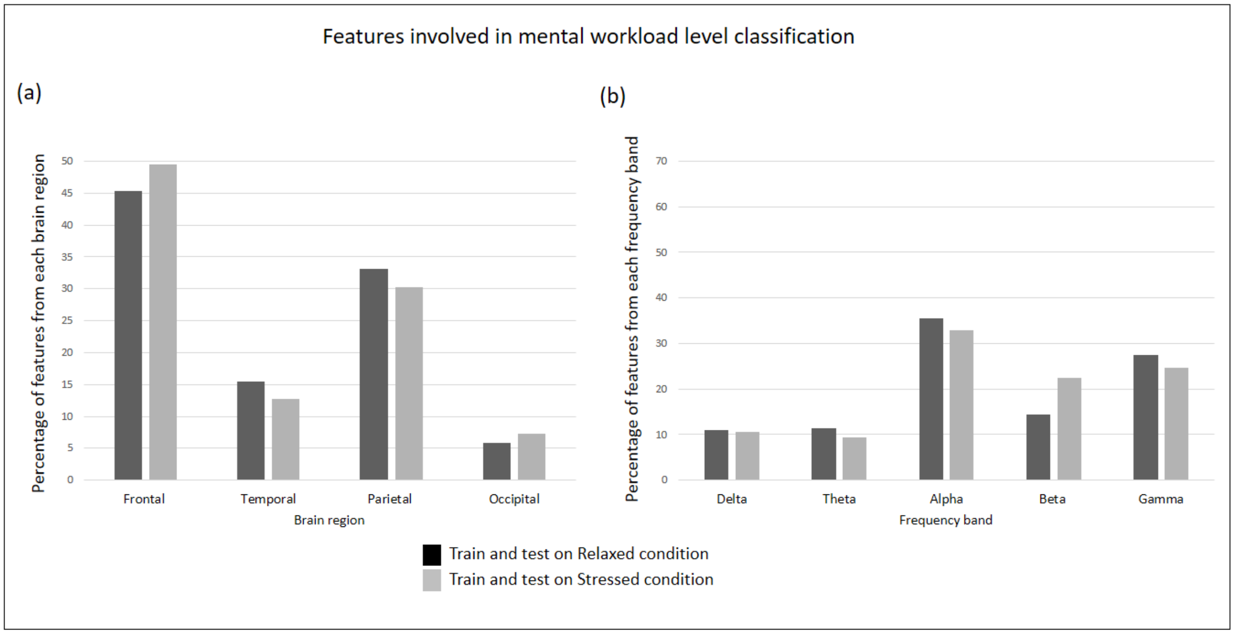


Figure 2.3: Feature analysis for the mental workload classification problem (i.e., Easy vs. Difficult) for the "within-affective-state" classifier training paradigm. Percentage of all features selected (over 5 runs of 6-fold cross-validation) that came from a) each brain region considered, and b) each frequency band considered.

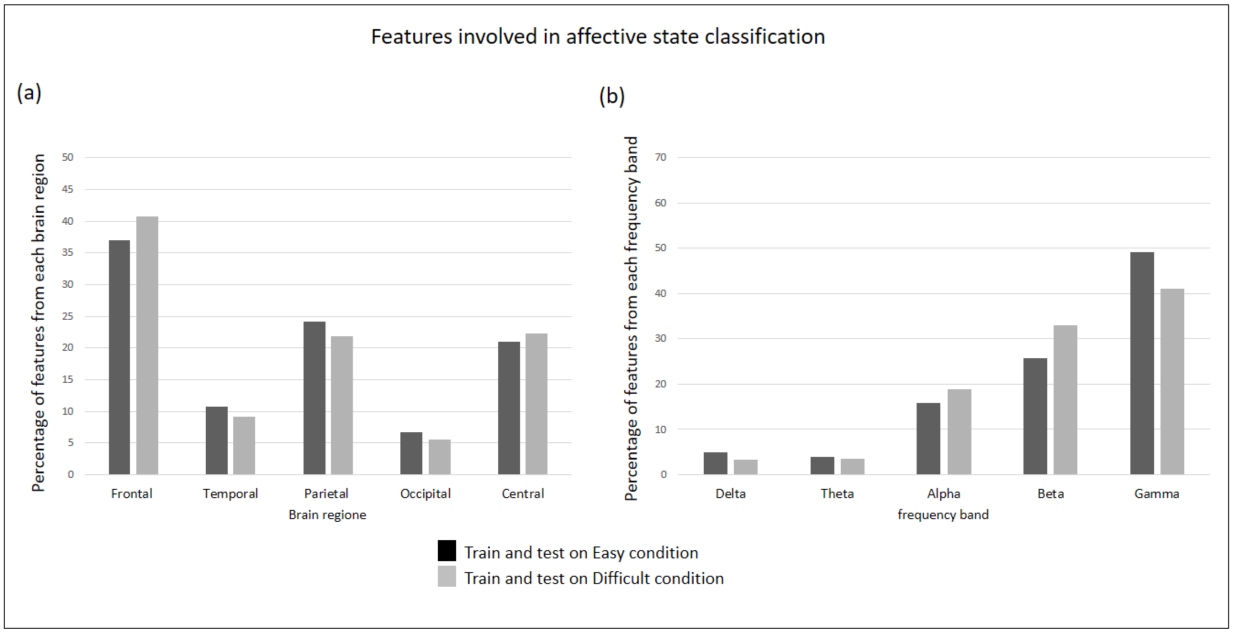


Figure 2.4: Feature analysis for the affective state classification problem (i.e., Relaxed vs. Stressed) for the "within-workload-level" classifier training paradigm. Percentage of all features selected (over 5 runs of 6-fold cross-validation) that came from a) each brain region considered, and b) each frequency band considered.

2.6 Discussion

The main objective of this study was to investigate the effect of varying affective state on the automatic classification of mental workload, and the effect of varying mental workload on the automatic classification of affective state (specifically stress or no stress), via EEG signals. For this, we designed an experimental protocol that required participants to perform a task at two levels of difficulty (Easy and Difficult), each under two different affective states (Relaxed and Stressed).

Subjective and objective measures were used to validate our workload and affective state induction protocols. The results of statistical tests comparing both the heart rate and the STAI scores between the Relaxed and Stressed Blocks indicate that our protocol was effective in inducing the desired affective states. Given the novelty of our protocol in terms of interleaving the Relaxed and Stressed Blocks, we were concerned that the second instance of the stress induction protocol (Stressed

Block #2) might be ineffective (or less effective) in inducing stress, so to test this we performed a two-way ANOVA with “block” (two levels: Block #1 and Block #2) included as a main effect. The test revealed no statistically significant main effect of block on the STAI scores, and no interaction effect of block and affective state, indicating that in general stress induction was equally effective in both the first and second Stressed Blocks.

Similarly, statistical comparison of the participants’ response times, response accuracies and RSME ratings among the different task difficulties indicate that distinct levels of mental workload/effort were induced in the Easy (“low workload”) and Difficult (“high workload”) trials. It is interesting that the reaction times in particular were on average almost twice as long for the Difficult trials as for the Easy trials. Significant interaction effects indicated that the difference in both response accuracy and response time between the Easy and Difficult conditions was more pronounced in the Stressed condition than the Relaxed condition. The fact that cognitive performance is impaired under anxiety is well-reported [288-290]. These results provide further support that our stress induction protocol was effective.

It is worth noting that while the difference in the RSME ratings between the Easy and Difficult conditions was statistically significant, it was not very large (1.2 vs. 2.7 on a scale from 1 to 9). This was not unexpected, and was due to the fact that when designing our task workload levels, we really prioritized the need to not induce stress in the high workload condition. It was imperative for the objectives of this work that we induce mental workload and stress as independently of one another as possible. Therefore, we endeavored to make the high workload task as different from the low workload task as possible in terms of difficulty while keeping it manageable enough that the participants would not feel any additional stress due to the task. This is the reason (as mentioned in the section 2.4.3.1) that we did not put any time limit on the task, for example. Given the unfamiliar scenario of participating in a research study, we were concerned that pushing the difficult condition too high would result in our participants feeling task-related stress. Because of this, we were relatively conservative in the design of our high workload condition. The fact that the EEG classification results are well above chance and in line with other similar studies in the literature [117], combined with the difference in the task performance measures, is indicative that distinct workload levels were indeed induced in the Easy and Difficult conditions, and we felt that this was sufficient to meet the objectives of this research, i.e., to investigate the effect of variation in stress on the ability to classify mental workload (and vice versa).

To test how a workload classifier would be effected by variability due to changes in affective state, we considered three different classifier training paradigms: 1) "within-affective-state" classification (classifiers were trained and tested on data from the same affective state), 2) "across-affective-state" classification (classifiers were trained on data from one state and tested on data from the other state), and 3) "combined-affective-state" classification (classifiers were trained on data from both states combined but tested on data from a single state). The results showed that in the "within-affective-state" classification paradigm, the Easy and Difficult conditions could be distinguished from one another with accuracies significantly exceeding chance, about 71% in the Relaxed condition, and 74% in the Stressed condition. We found that regardless of whether the test data were from the Relaxed or Stressed condition, the "across-affective-state" classifier training paradigm resulted in quite significantly reduced classification accuracies – around 15-17% lower than the "within-affective-state" case. These results support those reported in [117], where they also observed a reduction in classification accuracy for the across-affective-state as compared to the within-affective-state training paradigm, however our observed effect is much larger (they observed an approximately 3% reduction in accuracy for low vs. high workload classification). This discrepancy could be due to the different experimental protocols used. In [117] the participants were told about a public speaking task, and then actually had to go through with the task prior to performing the workload trials. Likely, the participants' stress level was at a peak during the performance of the public speaking task, and then reduced drastically once it was finished, prior to the workload trials. In our paradigm, we told the participants about the upcoming public speaking task, and then they completed the workload trials under the anticipation of this imminent, very stressful situation. It is possible that our paradigm was more effective in inducing stress during the workload trials, and this is why we observed a larger effect of the "across-affective-state" training paradigm. The observed difference could also be due to the fact that we used a different workload task than was used in 117].

For the Easy vs. Difficult problem, including data from both the Relaxed and Stressed conditions in the training set – as in the "combined-affective-state" training paradigm – significantly improved the classification as compared to the "across-affective-state" paradigm, but does not increase it to the level of the optimal, "within-affective-state" paradigm.

Similar results were observed when investigating the effect of varying workload on the classification of affective state. When performing classification of Relaxed vs. Stressed, the "within-workload-level" training paradigm yielded an accuracy of 85%, which was significantly reduced to 74% on average for the "across-workload-level" training paradigm. When data from both workload levels were included in the training set (the "combined-workload-level" training paradigm), the accuracy was significantly improved to approximately 82%, but again did not reach the level of the optimal ("within-workload-level") paradigm.

Our classification results for the Easy vs. Difficult condition (for the “within-affective-state” training paradigm) are comparable to other results reported in the literature, including those in [117] which are approximately equal, on average, to ours.

For the mental workload classification, Figure 2.3a indicates that a majority of the selected features came from the frontal and parietal lobes. This is in line with other studies that have shown these regions to be associated with mental workload [116, 291-300]. In terms of frequency bands, Figure 2.3b indicates that approximately 56% of features selected were from lower frequency bands (i.e. delta, theta and alpha) and approximately 44% from high frequency bands (i.e. beta and gamma). Some studies have suggested that high frequency bands can be contaminated with EMG from the frontalis and/or temporalis muscles [301]. Since many people tend to contract the frontalis muscle when concentrating, it seemed possible that there could potentially have been more such EMG activity during the Difficult than the Easy task, and that this could have contributed to the differentiability of these conditions. Indeed, [302] found EMG of the frontalis muscle to be sensitive to variation in cognitive load. To make sure that the classification was indeed based primarily on neural activity rather than muscle activity, we repeated the “within-affective-state” classification with the high frequency bands of beta and gamma excluded from the feature pool. With the high frequency bands excluded, the classification accuracy for Easy vs. Difficult decreased by approximately 4% on average, from about 72.6% to 68.3%. These results are still well in excess of chance, indicating that the workload classifier does rely primarily on neural information from low frequency bands. If frontalis muscle activity did indeed contribute minimally to the classifier, however, we argue that this would not necessarily be a negative. The application of mental workload detection is for able-bodied users so there are no concerns that the results would not transfer to the target population as is often a concern in active BCI. Furthermore, many studies investigating mental workload detection incorporate other physiological signals like ECG, GSR, and pupil dilation [303] to improve results. We argue that it would be in a similar vein to incorporate information from EMG activity generated due to workload-induced contraction of the frontalis muscle.

For affective state classification, Figure 2.4 indicates that while features from all brain regions and frequency bands were selected with some frequency, a majority of selected features came from the frontal, parietal and central brain regions and from the higher frequency (gamma and beta) bands. These findings are in line with results reported in the literature [197]. The gamma band has been reported to be more discriminative between stress levels than lower bands, and it has been suggested that both low and high frequency bands should be considered in order to assess stress level [197].

The results of this study confirm those reported in [117] that affective state can have a very significant effect on the classification of workload levels via EEG. In the present study, there is no possibility that the observed effect is simply due to time-related changes in the signals since the Stressed and Relaxed conditions in our experimental protocol were interleaved. Our results further suggest that workload level can have a similar effect on the classification of affective state. In order for stress and workload detection algorithms to be useful outside the lab where users are likely to experience changes in both their cognitive load and stress level as they perform tasks in real life scenarios, these confounding effects must be considered and accounted for. These results suggest that including samples from both affective states in the training data for mental workload detection, and including samples from different workload levels in stress detection, will make the algorithms somewhat more robust, however it is not an optimal solution.

Now that we have identified this issue, the next steps of this work will be to try to improve the concurrent classification of mental workload and stress by incorporating techniques such as hierarchical classification strategies and transfer learning methods. Future work will also investigate mental workload and stress at the highest end of workload, where the two are often (though not necessarily) highly correlated. While in many scenarios the induction of mental workload and stress may be inextricably linked, we hope to determine if detection of the states can still be done independently.

2.7 Conclusion

In this study, we investigated whether the ability to detect one mental state via EEG signals is affected by variation in another state. Specifically, we looked at the confounding effects of mental workload and stress. We found that detection of both states is significantly diminished in the presence of variation of the other state. In order to be effective in real-life scenarios where simultaneous variation of both mental workload and stress is inevitable, this effect will have to be accounted for in the development of the EEG-based state detection algorithms.

2.8 Acknowledgements

The authors acknowledge with gratitude the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 3 : Investigating hierarchical and ensemble classification approaches to mitigate the negative effect of varying stress state on EEG-based detection of mental workload level - and vice versa

Co-authorship statement. A version of this chapter has appeared in the journal Brain Computer Interfaces as the article titled "Investigating *hierarchical and ensemble classification approaches to mitigate the negative effect of varying stress state on EEG-based detection of mental workload level - and vice versa"* in August 2021. The author, Mahsa Bagheri, carried out the study design, data collection and data analysis. Mahsa Bagheri also prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-author's feedback as well as the comments received from the peer review process. The co-author, Dr. Sarah Power provided guidance on study design, data collection and data analysis. Dr. Power also helped in reviewing, editing and revising the manuscript. All authors read and approved the final draft.

The same dataset is used in this study as was described in Chapter 2.

3.1 Abstract

Research studies on EEG-based mental workload detection generally consider workload exclusively in terms of task demands, as manipulated through the variation of task difficulty. However, mental workload cannot be estimated precisely with the properties of the task alone because different aspects of the user's state could affect the mental effort needed to perform the task. We argue that affective state, specifically stress/anxiety, is of particular relevance. Thus the

overall objective of this work is to develop a passive brain-computer interface that detects mental workload and affective state simultaneously. In the study described in Chapter 2, we found that variation in affective state negatively effects the EEG-based classification of mental workload, and vice versa. Thus, such variation should be explicitly considered when developing detection algorithms that will remain effective outside the lab. In the work described in this chapter, we investigated five classification approaches for detecting mental workload and affective state which explicitly consider variation in the other state. Significant improvements in classification accuracy were achieved for both states.

3.2 Introduction

Passive brain-computer interfaces (BCI) are systems that aim to monitor the mental state of the user and use the information to enhance an ongoing human-machine interaction [1, 10]. Information on the mental state is derived from neural signals obtained through an appropriate functional imaging technology. Due to its portability, relatively low cost, and excellent temporal resolution, electroencephalography (EEG) has emerged as perhaps the most promising modality. One highly impactful potential application of passive BCI is the monitoring of mental workload [20, 40, 88, 94, 95, 114, 144, 155, 222-224] in high risk and safety critical occupations like air traffic controllers, pilots, and other industrial operators, where the consequences of human error can be severe. The system would detect potentially dangerous states of high cognitive demand or overload and take measures to help mitigate the risk of error.

Mental workload is defined as the perceived relationship between an individual's total mental processing capability and the amount required by the task at hand [80]. The closer the task requirements are to the individual's capability, the higher is the perceived workload. Mental workload arises from the interaction of several factors including the properties of the task being performed (e.g., difficulty), the characteristics of the individual performing the task (e.g., cognitive capacity, level of training in the task, mood), and the task environment (e.g., noisy, distracting) [243].

Mental workload detection via EEG (sometimes in combination with other physiological variables such as electrocardiogram or pupil dilation) is a very active and growing field, with a large number of published studies [e.g., 34, 36, 69, 81-83, 89, 90, 105, 106, 109, 116, 117, 134, 147, 234, 304, 305]. Such studies generally consider workload exclusively in terms of task demands, as manipulated through the variation of task difficulty. However, while this is certainly a major factor, precise estimation of mental workload is not possible when exclusively considering the properties of the task because, as mentioned above, other factors will affect the mental effort needed to perform the task [236]. We argue that the potential effect of the user's affective state, and specifically their stress/anxiety level, is of particular relevance, especially in passive BCI applications aimed at safety-critical occupations where individuals often perform duties in stressful situations that carry a high cost of error. According to the Processing Efficiency Theory, stress reduces the storage and processing capacity of the working memory system, and necessitates an increase in on-task effort to maintain the same level of performance [237, 306]. So the full picture regarding the cognitive state of the individual that is relevant to their performance of the task at hand and their potential for error - which is the information the passive BCI is intended to provide – should include a combination of the difficulty of the task they are performing, and their level of anxiety/stress. It would be useful to know, for example, if the individual is performing a difficult task very calmly or with significant anxiety. The latter may be a much riskier situation, and thus the environmental adaptation strategies that should be implemented by the passive BCI should be much different in these two scenarios. Thus, the overall aim of this research is to develop a passive

BCI that is able to detect both the difficulty of the task the individual is performing and their stress/anxiety level simultaneously, in a sort of "two-dimensional" measure of cognitive load.

While detection of both task difficulty and stress/anxiety from EEG signals have been attempted before on an individual basis, estimating both simultaneously, with each state confounding the other, is a technical challenge that has only recently been addressed. In our recent study [304], we found that while both states (i.e., stress, and mental workload due to task difficulty) could be simultaneously classified at levels significantly exceeding chance, variation in each state had deleterious effects on the classification of the other by EEG. We found that the accuracy with which workload levels (as defined exclusively in terms of task difficulty) could be classified was affected by whether the classifier training and testing data were collected under the same or different affective states. When the classifier was trained on data from a relaxed (stressed) condition and tested on data from a relaxed (stressed) condition – i.e., the "within-affective-context" classification paradigm - mean classification accuracy across 18 participants for the "Easy" vs. "Difficult" task conditions was approximately 73%. However, when the classifier was trained on data from the relaxed (stressed) condition and tested on data from the stressed (relaxed) condition – i.e., the "across-affective-context" classification paradigm - the mean accuracy dropped to approximately 56%. This suggests that in a real passive BCI system, if the mental workload detection algorithm were to be trained using data collected during a calibration session where the individual experienced a single - likely relaxed - affective state, performance could deteriorate significantly when applied to real-life scenarios which are likely to induce varying levels of stress in the user. It was found that including data from both the stressed and relaxed states in the training set (i.e., the "combined-affective-context" classification paradigm) improved the classifier performance to approximately 67% - significantly higher than in the "across-affective-context" case, but still significantly below the ideal "within-affective-context" case. These results were in line with the results reported in [117]. Furthermore, we found in [304] that variation in the difficulty level of the task that the individual was performing had similarly negative effects on the performance of algorithms classifying affective state. The mean classification accuracy for the "Relaxed" vs. "Stressed" conditions across 18 participants was approximately 85% and 74% for the "within-workload-level" and "across-workload-level" classification paradigms, respectively. Including data from both workload conditions in the training set (i.e., the "combined-workload-level" classification paradigm) resulted in an accuracy of 82%, which was again significantly better than the "across-affective context" paradigm, but still significantly below the ideal "within-workload-level" classification paradigm.

In this paper, we expand on our earlier work and investigate ways to improve the classification of each state (i.e., mental workload and stress) in the presence of variation in the other state. Specifically, we consider four classification approaches that, based on our previous results, we hypothesized would improve accuracy of classification as compared to the "combined-affective-context" and "combined-workload-level" paradigms reported in [304]. The total of five approaches we investigated are described below for the case of mental workload classification (the same approaches were taken for affective state classification).

1) Combined approach – Each test sample is classified using a mental workload classifier trained on data from both affective states combined.
2) Hierarchical (without thresholding) approach – For each test sample, the individual's affective state is predicted first, then mental workload is predicted based on a "stress-state-specific" workload classifier trained only on data from the predicted stress state.

3) Hierarchical (with thresholding) approach – This is the same as the first approach, except the use of the “stress-state-specific” workload classifier is predicated on the confidence of the affective state prediction.
4) Posterior probability-based approach – Each test sample is classified using two different mental workload classifiers, independently: one trained on data from the relaxed state only, and one trained on data from the stressed state only. The final classification result is that of the classifier with the highest posterior probability associated with the prediction.
5) Majority vote-based approach – Each test sample is classified using three different mental workload classifiers, independently: one trained on data from both stress states, one trained on data from the relaxed state only, and one trained on data from the stressed state only. The final classification result is based on majority vote of the three classifiers.

In the following sections we describe our experimental methods, provide more details about the classification approaches, and report and discuss our findings.

3.3 Material and methods

3.3.1 Experimental methods

The experimental methods, including participants, physiological data acquisition, experimental procedure and validation of stress and workload induction, are the same in this study as were described in Chapter 2. Please refer to section 2.4.1 for information on participants, section 2.4.2 for physiological data acquisition, section 2.4.3 for experimental procedure, and sections 2.4.4.1, 2.4.4.2, 2.5.1 and 2.5.2 for information on validation of stress and workload induction.

3.3.2 Data analysis

3.3.2.1 EEG-based classification

In this paper, we investigate five classification schemes to improve mental workload classification (i.e., Easy vs. Difficult) in the presence of varying affective-state, and affective-state classification (i.e., Relaxed vs. Stressed) in the presence of varying workload level. The classification analysis was performed separately for each subject.

For each classification problem of interest, the following steps had to be taken prior to classification: preprocessing, feature extraction, and feature selection.

3.3.2.1.1 EEG pre-processing

EEG signals contain artifacts from sources such as movement of the electrodes, eye blinks and eye movements (electrooculography, EOG), cardiac signals (electrocardiography, ECG), and muscle activity (electromyography, EMG), therefore pre-processing techniques must be applied to remove them. We first applied a 1-50 Hz band-pass filter to remove signal components outside of our frequency range of interest. Next, signal segments contaminated by movement artifacts and EMG were manually rejected. Then, independent component analysis (ICA) [282] was applied and signal components related to eye blinks and saccades were identified and removed. Finally, the data was down sampled from 500 Hz to 256 Hz. Pre-processing steps including ICA were applied using the MatLab Toolbox EEGLAB [283, 284]. "Runica" was selected for the ICA type and it was applied to all of a subject's data, which is approximately 45 minutes in length.

3.3.2.1.2 EEG signal feature calculation

To make sure that the workload level classification was based on differences in neural activity related to the workload levels and not to any incidental differences in motor activity (since the response frequency was higher for the Easy as compared to the Difficult condition), we did not include the electrodes over the motor and sensorimotor brain regions for the mental workload classification problem. Specifically, the seven central and seven centro-parietal electrodes (Cz, C1-C6, CPz, CP1-CP6) were excluded. These electrodes were not excluded for the affective state classification problem since there was no difference in the motor requirement in the Relaxed and Stressed conditions. A total of 49 electrodes were considered for the mental workload classification, and 63 for the affective state classification.

Frequency domain features of the EEG signals were calculated. Specifically, the signal power in seven common EEG frequency bands were calculated with respect to each participant's individual alpha frequency (IAF) [285] as follows: delta from (IAF – 8) Hz to (IAF - 6) Hz; theta from (IAF – 6) Hz to (IAF – 4) Hz; alpha1 from (IAF – 4) Hz to (IAF - 2) Hz; alpha2 from (IAF – 2) Hz to (IAF) Hz; alpha3 from (IAF) Hz to (IAF + 2) Hz; beta from (IAF + 2) Hz to (IAF + 20) Hz; gamma from (IAF + 20) Hz to (IAF + 30) Hz. The IAF for each participant was determined using the eyes-closed baseline trial collected at the beginning of the session. Specifically, the frequency with the maximum signal power for the eyes-closed baseline was taken as the IAF. Power signals for each frequency band were obtained via the Filter-Hilbert method [286]. Then, average power was calculated over non-overlapping 2-second epochs. The power features were calculated for each electrode individually.

For the affective-state (Stressed vs. Relaxed) classification problem, measures of relative gamma (RG) activity were also computed [307] as follows: the power spectral distribution was computed on the z-transformed EEG in each of the 4-second windows. This distribution was averaged through all electrodes. The RG was computed as the power ratio between gamma and other rhythms (i.e, gamma/delta, gamma/theta, gamma/alpha1, gamma/alpha2, gamma/alpha3 and gamma/beta). The inverse of each frequency band (i.e., 1/delta, 1/theta, 1/alpha1, 1/alpha2, 1/alpha3, 1/beta and 1/gamma) were also computed [197].

This resulted in a total of 49 electrodes x 7 frequency bands = 343 features for the mental workload classification problem and a total of 63 electrodes x 7 frequency bands + 13 RG features = 454 features for the affective state classification problem. There were approximately 400 seconds of data collected for each combination of workload level and affective state (2 blocks per stress condition x 3 trials per workload level per block x 67 seconds per trial), which with 2-second epochs yielded approximately n=200 samples (i.e., epochs) per workload level/affective state combination.

Finally, each feature was normalized to a range between 0 and 1.

3.3.2.1.3 Classification algorithm

In each of the classification approaches investigated, any individual binary classifications were done using a regularized linear discriminant analysis (LDA) classifier [308] from the MatLab Toolbox BCILAB [284, 309].

3.3.2.1.4 Classification approaches

For both the workload level classification and the affective state classification, five classification approaches were investigated, as described below. In the following explanations, we will refer to the state to be classified (either mental workload or affective state) as "State A", and the other state as "State B". So the goal of the classification analyses is to predict the level of State A in the presence of variation in the level of State B. See Figure 3.1 for flowchart representations of each approach.

a) <u>"State-B-combined" approach:</u> The level of State A is predicted using a classifier trained using data from both levels of State B. This is the same as the "combined-workload-level" and "combined-affective-state" classification paradigms reported in [304]. The performance of the remaining approaches will be evaluated against this approach.

b) <u>Hierarchical (without thresholding) approach:</u> First, the level of State B is predicted, then based on that result, the appropriate State-B-specific classifier (i.e., one trained on data coming from only the predicted level of State B) is used to predict the level of State A. The State B classifier used in the "first stage" of the classifier is trained using data from both levels of State A (i.e., it is a "State-A-combined" State B classifier). The rationale for investigating this approach was our previous finding that for both mental workload and affective state, classification accuracy was higher when training data came exclusively from the same level of the "other" state as the sample to be classified.

c) <u>Hierarchical (with thresholding) approach:</u> First, the level of State B is predicted, then if there is sufficient confidence in that prediction (specifically, if the posterior probability of the predicted class is greater than or equal to 0.8), an appropriate State-B-specific classifier is used to predict the level of State A. If the confidence in the State B prediction is not sufficient, the "State-B-combined" classifier is used to predict the level of State A.

d) Posterior probability-based approach: Two separate "State-B-specific" State A classifiers were trained, one for each level of State B. The final prediction for the level of State A was taken to be the output of the classifier with the highest posterior probability.

e) Majority vote-based approach: Three separate State A classifiers were trained – a "State-B-combined" State A classifier and two "State-B-specific" State A classifiers, one for each level of State B. The final prediction for the level of State A was determined through majority vote of these three classifiers.

The performance of the all classification approaches was assessed using the average accuracy across five runs of six-fold cross-validation. In each "fold" of the cross-validation, no test data was used in either feature selection or classifier training.

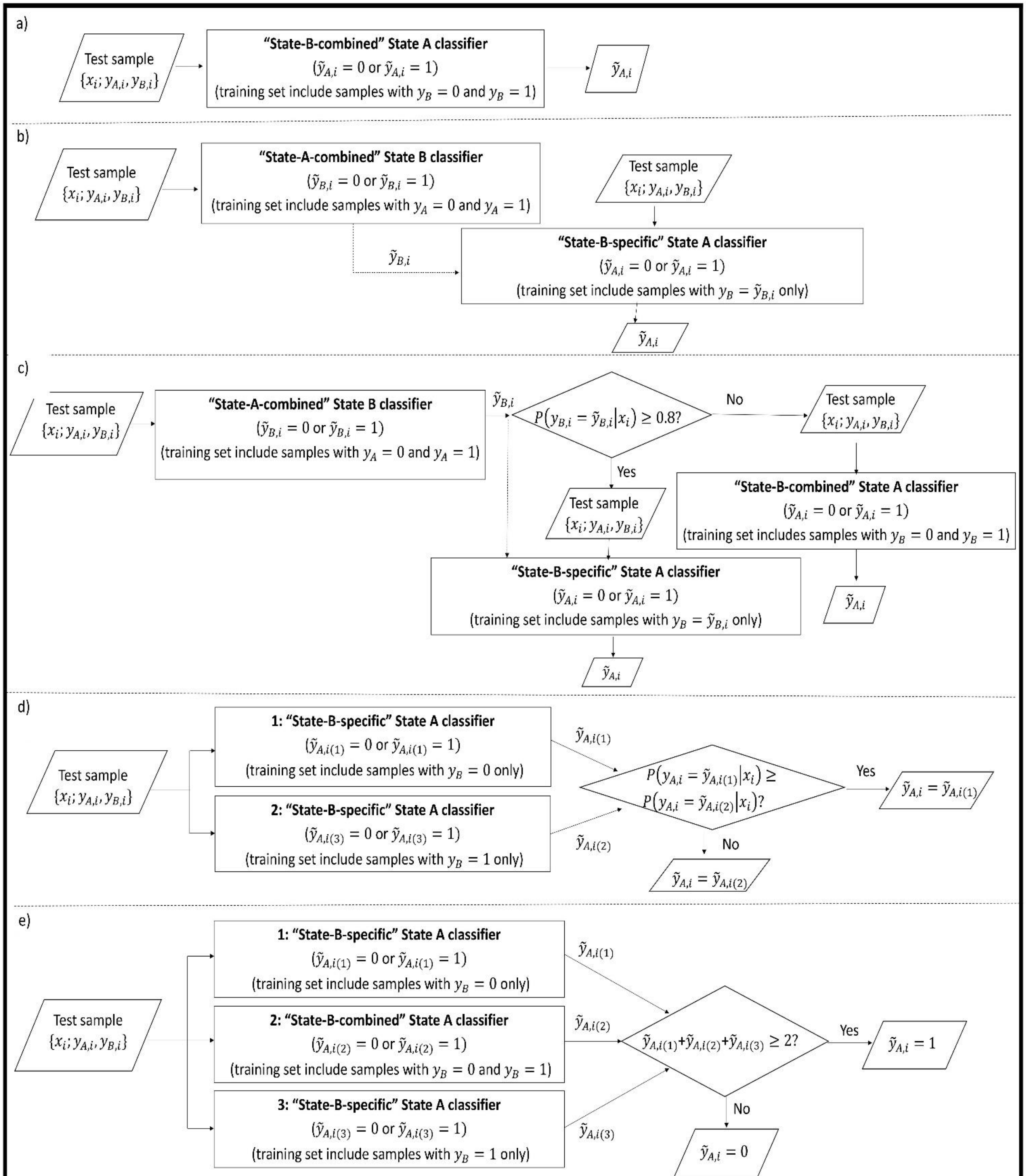


Figure 3.1: Classification approaches. For the i^{th} test sample: x_i represents the EEG feature vector; y_{Ai} and y_{Bi} represent the true class labels for State A and State B, respectively; $\tilde{y}_{Ai}$ and $\tilde{y}_{Bi}$ represent the predicted class labels for State A and State B, respectively.

3.3.2.1.5 Statistical analysis

One-way repeated measures ANOVA with post-hoc Tukey-Kramer tests was performed to compare the performance of the five different classification approaches. Cohen's d values were calculated to indicate the effect sizes of each classification approach compared to the "State-B-combined" approach.

3.4 Results

3.4.1 EEG-based classification

Table 3.1 shows the results of workload level classification (Easy vs. Difficult) for all participants for the five classification approaches. The hierarchical classification, with and without the thresholding, and the majority-vote-based classification approaches gave the best results with increases of 2.7%, 1.8%, and 2.5%, respectively, over the combined-affective-state approach. Post-hoc Tukey-Kramer tests indicated that these increases were all significant ($|t_{(68)}| > 5.49; p < .001$). The result for hierarchical classification with thresholding was significantly higher than that without thresholding ($t_{(68)} = 2.88; p = .04$), but there were no other significant differences among these three best approaches ($|t_{(68)}| < 2.19; p > .19$). The posterior-probability-based classification approach had a significantly lower accuracy than the affective-state-combined approach ($t_{(68)} = 5.42; p < .001$).

Table 3.1: Workload-level classification results for all participants for the five different classification approaches. Mean accuracies that are significantly (α=0.05) higher than that of the "affective-state-combined" approach are indicated by *. Cohen's d values are given to indicate effect sizes in these cases. A Cohen's d value of 0.2 indicates a small effect, while a value of 0.5 indicates a medium effect [310].

Mental Workload Classification Results (Easy vs. Difficult)					
Subjects	Affective-state-combined	Hierarchical (with thresholding)	Hierarchical (without thresholding)	Posterior-probability-based	Majority-vote-based
1	68.5	71.8	71.2	66.6	70.9
2	59	60.3	59.2	56.8	60.2
3	66	70.2	69.5	64.3	63.2
4	71.6	72.6	72.1	68.2	72.5
5	61.9	64.6	62.9	62.3	64.3
6	69.3	71.4	69.9	67.3	72.2
7	63.7	67.9	67.6	64.7	66.4
8	61.3	63	63.5	59.9	63
9	67.5	71	71	67.1	70.4
10	71.4	75.1	74.4	71	74.9
11	69.8	72.1	71.9	68.6	71.7
12	63.9	68.7	66.8	64.7	69.5
13	77	79.5	77.5	75.8	79.2
14	79	80.9	78.3	72.6	81
15	75.2	77.6	77	72.8	77.7
16	67.3	70.8	70.5	64.7	69
17	73	73.4	72.7	68.7	75.6
18	66.6	68.5	67	65.2	67.8
Mean:	68.4 ± 5.4	71.1 ±5.3*	70.2 ±5.1*	66.7 ±4.6	70.9 ±5.5*
Cohen's d	---	0.49	0.32	---	0.44

Table 3.2 shows the results of affective-state classification (Relaxed vs. Stressed) for all participants for the five classification approaches. The best results came from the majority-vote-based classification approach with average accuracy of 86.3%. This result exceeded the accuracy for the combined approach by 2.1%, however a post-hoc Tukey-Kramer test revealed that this difference just reached significance ($\left|t_{(68)}\right| = 2.8; p = .0499$). All of the other classification approaches yielded results very similar to, or lower than, that of the combined approach.

Table 3.2: Affective-state classification results for all participants and different classification approaches. Mean accuracies that are significantly (α=0.05) higher than that of the “workload-level-combined” approach are indicated by *. Cohen’s d values are given to indicate effect sizes in these cases. A Cohen’s d value of 0.2 indicates a small effect, while a value of 0.5 indicates a medium effect [310].

Affective State Classification Results (Relaxed vs. Stressed)					
Subjects	Workload-level-combined	Hierarchical (with thresholding)	Hierarchical (without thresholding)	Posterior-probability-based	Majority-vote-based
1	84.2	84.5	79.9	85.2	87.1
2	76.7	75.7	75.2	79	79
3	82.9	82.7	81	82	86
4	88.7	88.1	86	89.7	90.2
5	69.2	70.3	69.6	71.2	71.1
6	86.3	86.6	85.1	86.1	87.6
7	86.5	85.6	84.7	88.4	88.5
8	83.7	83.1	81.6	83.8	85.5
9	76.6	75.6	73.8	76	77.8
10	97.7	98.2	96.5	98.3	98.4
11	83.7	84.7	83.7	86.5	86.1
12	80	79.3	76.2	80.3	82.2
13	85.5	88.8	88.8	88	90.5
14	81	80.9	78.9	82.6	82.8
15	98.1	98.2	98.4	80.3	98.9
16	87.1	85.7	84.8	87.2	90.2
17	85.8	87.7	84.8	88	89.2
18	82.3	80.6	77.3	81.5	82.1
Mean:	84.2±6.8	84.2±7	82.6±7.3	84.1±5.9	86.3±6.7*
Cohen’s d	---	---	---	---	0.30

Note that for both mental workload and affective state classification, the “State-B-combined” State A classification results were a bit higher than those reported in [304], and referenced in the introduction section of this paper. This is due in part to the fact that in the previous paper the amount of data used for the classification analysis was balanced across the different classifier training paradigms in order to allow fair comparison, which resulted in half the amount of data being used for classification in the “combined” case. Here, the full dataset is used in each of the classification approaches investigated and, unsurprisingly, the significant increase in the amount of data resulted in an increase in the classification accuracy. Also, slightly different classification techniques were used here (specifically, 2 s non-overlapping epochs instead of 4 s overlapping epochs, and regularized LDA instead of LDA without regularization).

3.5 Discussion

The overall aim of this research is to develop a passive BCI that is able to detect both the difficulty of the task the individual is performing and their stress/anxiety level simultaneously, in a sort of "two-dimensional" measure of cognitive load. In our recent study [304], we found that while both stress and mental workload (due to task difficulty) could be simultaneously classified at levels significantly exceeding chance, variation in each state had quite significant negative effects on the classification of the other by EEG when this variation was not considered in the development of the classification algorithms. It was found that for both mental workload classification and affective state classification, including data from both levels of the "other state" helped, but accuracies were still significantly below the "ideal case" of no variation in the state.

The main objective of this study was to investigate different classification approaches to improve the performance of 1) EEG-based mental workload detection in the presence of variation in affective state and 2) EEG-based affective-state detection in the presence of variation in mental workload level (as modulated by task difficulty), by explicitly considering this variation in the development of the classification algorithms. We considered four different classification approaches that were based on hierarchical and ensemble methods and compared them to the approach of simply combining both levels of the "other state" in the training set. For the mental workload classification problem, the hierarchical approach with thresholding and the majority vote-based approaches significantly improved the classification accuracy as compared to the "combined" approach, with increases of 2.7% and 2.5%, respectively on average across the 18 participants. In an individual participant basis, these methods either made no difference or improved the classification accuracy as compared to the "combined" approach for all participants. For the affective state classification, only the majority vote-based approach resulted in a statistically significant improvement (2.1 %). It is unsurprising that the hierarchical approach was effective for the mental workload classification but not the affective state classification, since the first stage of this approach is based on classification of the other state using the “combined approach”, and the performance of the affective-state-combined mental workload classifier is much higher than the workload-level-combined affective state classifier (68.4% vs. 84.2%). This suggests that the hierarchical approach will be more effective in cases where the first stage of the classification (i.e., the “State B classifier”) is sufficiently accurate. Even in the case of mental workload, though, the hierarchical classification approach performed as least as well as the “combined” approach. Given that no method outperformed the majority vote approach in either the mental workload or the affective state classification, the results suggest that this may be the most reliable approach.

Though the increases were modest, we did identify classification approaches that significantly improved classification of mental workload in the presence of variation on affective state, and classification of affective state in the presence of variation in mental workload. It is worth noting that there was a limit on how much the accuracies could be improved, at least with this dataset, since the accuracy of the “ideal” (and not practically feasible) case of using a “State-B-specific” classifier that precisely matches the true state of the test sample 100% of the time is only 72.6% for mental workload classification and 85.3% for affective state classification. Despite the modest increases seen, we have shown the potential efficacy of these approaches, and it is possible that the improvements could be more substantial in other datasets. Furthermore, the approaches investigated here for use in the simultaneous detection of mental workload and affective state could be useful when applied to other mental states.

For the affective-state (Stressed vs. Relaxed) classification problem, measures of relative gamma (RG) activity were computed and added to the feature pool as they have been previously shown to be effective for stress detection [197, 307]. In order to investigate the contribution of these features to the classification performance, we repeated the classification of affective state, using the combined classification approach, with the RG features excluded from the feature pool and found a statistically significant decrease of approximately 1% in the average accuracy over 18 subjects ($F = 11.28, p = .004$).

To the best of our knowledge at the time of writing, the studies reported in this and our earlier paper [304] represent the first attempts to perform simultaneous classification of mental workload and stress, though there are many studies that have looked at one or the other condition individually. Generally, there is a wide range of classification results reported, and it is difficult to compare one study to the next due to differences in the experimental (e.g., type of task, difference between high/low workload conditions, number/length of trials) and analytical (e.g., pre-preprocessing techniques, EEG features used, feature selection and classification algorithms) methods employed. For example, studies aimed at stress detection use a wide variety of stress induction techniques including cognitive tasks like mental arithmetic (the Montreal Imaging Stress Task, MIST) [167, 168], exposure to emotionally salient pictures/videos [177] or music [208], or the Stroop task [160, 188]. In the studies cited, classification accuracies ranged from 73.3 to 94.6 %. In [163], the Trier Social Stress Test was used to induce stress, and a maximum accuracy of 91.2% was achieved; this is higher but comparable to our result, despite our data including variation due to workload, and our use of a significantly modified version of the TSST. Our classification results for the Easy vs. Difficult condition are also comparable to other results reported in the literature, including those in [117] which are approximately equal, on average, to those reported herein.

Future work will involve realizing the mental workload and stress detection algorithms online and evaluating them in more realistic, ecologically-valid scenarios in which users experience different levels of workload and affective-state. Incorporating multi-label classification models will also be considered.

3.6 Conclusion

The overall aim of this research is to develop a passive BCI that is able to detect both the difficulty of the task the individual is performing and their stress/anxiety level simultaneously, as both of these factors will influence the user's performance and potential for error. Previous studies, including by the authors, have shown that the performance of EEG-based mental workload classifiers deteriorate in the presence of varying affective state, and vice versa. Since the aim of passive BCIs is to monitor an operator's workload in real-world scenarios which are likely to vary from relaxing to stressful or from a low level of workload to a high level, this is a significant issue. To ensure the reliability and generalizability of the classification algorithms detecting each state, variation in the "other" state must be considered. In this paper, we investigated four classification approaches with potential to improve 1) mental workload-level detection in the presence of varying affective-state and 2) affective-state detection in the presence of varying workload-level as compared to the case of the "combined" classification approach that was investigated/introduced in Chapter 1. Statistically significant improvements in classification accuracy were observed for both mental workload and affective state classification using some of the proposed methods.

3.7 Acknowledgments

The authors acknowledge with gratitude the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 4 : Simultaneous classification of both mental workload and stress level suitable for an online passive brain-computer interface

Co-authorship statement. A version of this chapter has appeared in the journal Sensors as the article titled "*Simultaneous classification of both mental workload and stress level suitable for an online passive brain-computer interface"* in January 2022. The author, Mahsa Bagheri, carried out the study design, data collection and data analysis. Mahsa Bagheri also prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-author's feedback as well as the comments received from the peer review process. The co-author, Dr. Sarah Power provided guidance on study design, data collection and data analysis. Dr. Power also helped in reviewing, editing and revising the manuscript. All authors read and approved the final draft.

The same dataset is used in this study as was described in Chapter 2.

4.1 Abstract

Research studies on EEG-based mental workload detection for a passive BCI generally focus on classifying cognitive states associated with the performance of tasks at different levels of difficulty, with no other aspects of the user's mental state considered. However, in real-life situations, different aspects of the user's state such as their cognitive (e.g., level of mental workload) and affective (e.g., level of stress/anxiety) states will often change simultaneously, and performance of a BCI system designed considering just one state may be unreliable. Moreover, multiple mental states may be relevant to the purposes of the BCI—for example both mental workload and stress level might be related to an aircraft pilot's risk of error—and the simultaneous

prediction of states may be critical in maximizing the practical effectiveness of real-life online BCI systems. In this study we investigated the feasibility of performing simultaneous classification of mental workload and stress level in an online passive BCI. We investigated both subject-specific and cross-subject classification approaches, the latter with and without the application of a transfer learning technique to align the distributions of data from the training and test subjects. Using cross-subject classification with transfer learning in a simulated online analysis, we obtained accuracies of 77.5 ± 6.9% and 84.1 ± 5.9%, across 18 participants for mental workload and stress level detection, respectively.

4.2 Introduction

Passive brain-computer interfaces (BCI) are systems that aim to monitor the mental state (either cognitive or affective) of a user and exploit this information to adapt an ongoing human-machine interaction in some useful way [1, 10]. For example, a passive BCI could improve road safety by first detecting states of extreme fatigue or drowsiness in the driver of a car or transport truck and then using this information to initiate alarms or other safety measures to help avoid an accident. In passive BCI systems, information on the user's state is extracted from neural signals collected using an appropriate functional imaging modality. Electroencephalography (EEG), which measures the electrical activity of the brain, is considered perhaps the most promising modality given its portability, high temporal resolution, non-invasiveness, and relatively low cost [311].

One highly impactful potential application of passive BCI is the monitoring of mental workload in high risk and safety critical occupations like air traffic controllers, pilots, and other industrial operators, where incidents of human error can have severe consequences. Such a system would detect potentially dangerous states of high cognitive demand or overload and initiate measures to help mitigate the risk of error (e.g., have the system temporarily automate some tasks). Mental workload detection also has potential value in other domains, including gaming [225], adaptive training/learning [99, 226, 312], and user interface design applications [227], to enhance and personalize user experience. Because of its potential usefulness in a range of applications, mental workload detection via EEG is a very active and expanding field, with a large number of published studies [e.g., 34, 36, 69, 81-83, 89, 90, 105, 106, 109, 116, 117, 134, 147, 234, 304, 305, 313].

Mental workload is defined as the perceived relationship between an individual's total mental processing capability and the amount required by the task at hand; perceived workload is higher when the task requirements are closer to the individual's capability [80]. Mental workload is influenced by a number of factors including the properties of the task being performed (e.g., difficulty), the task environment (e.g., noisy, distracting), and the characteristics of the individual performing the task (e.g., cognitive capacity, level of training in the task, mood) [243]. Despite this, studies of EEG-based mental workload detection generally consider workload exclusively in terms of the task demands as manipulated through the variation of task difficulty, and do not consider any other factors. We argue that consideration of the user's affective state, and specifically their stress/anxiety level, is of particular importance. Stress is known to negatively impact cognitive efficiency and performance [156] as well as decision-making, especially when performing unfamiliar tasks [157]. Thus, a more complete picture of the user's cognitive state as it relates to their potential task performance and risk for error would include a combination of both the difficulty of the task they are performing, and their level of anxiety/stress. An individual performing a difficult task very calmly versus while experiencing a significant amount of anxiety are two very different scenarios with very different associated risk, and the environmental adaptation strategies implemented by the passive BCI should be much different in these cases. Thus, we argue that it is very important to develop a passive BCI that is able to detect both the difficulty of the task the individual is performing and their stress/anxiety level simultaneously, in a sort of “two-dimensional” measure of cognitive load. Detecting two mental states simultaneously via EEG, with each state confounding the other, is a challenge that had not previously been addressed in the literature.

In our previous work, we found that while both states (i.e., stress, and mental workload due to task difficulty) could indeed be simultaneously classified at levels significantly exceeding chance, variation in each state negatively affected the classification of the other by EEG [304]. Expanding on this work, we aimed to improve the classification of each of the two states in the presence of variation in the other state; we proposed a majority vote-based approach that modestly but significantly increased the classification accuracy of each state [313]. While the results of these studies were promising, the preliminary analyses were done offline and not in a manner suitable for real-time classification. It is therefore not clear how the results will translate to a practical, online passive BCI system for the simultaneous monitoring of mental workload level and stress.

In this paper, we advance our previous work by implementing simultaneous classification of mental workload level and stress using analysis techniques entirely compatible with online classification. We investigate both entirely subject-specific and cross-subject (i.e., using training data from subjects other than the test subject) classification approaches. In the latter case, a transfer learning technique recently proposed by [314] is employed to improve classification accuracy. Because of restrictions on research involving human participants arising due to the Covid-19 global pandemic, we were unable to conduct a separate experiment for this work, and instead used the data previously collected and reported in [304] and [313]. While the analysis for the current paper was by necessity performed offline, the methods used simulate exactly an online classification scenario and are completely suitable for direct online implementation.

In the following sections we describe our experimental methods, provide more details about the classification approaches investigated, and report and discuss our findings.

4.3 Materials and methods

4.3.1 Experimental Methods

The experimental methods including participants, physiological data acquisition, experimental procedure and validation of stress and workload induction are the same as Chapter 2. Please refer to section 2.4.1 for information on participants, section 2.4.2 for physiological data acquisition, section 2.4.3 for experimental procedure, and sections 2.4.4.1, 2.4.4.2, 2.5.1 and 2.5.2 for information on validation of stress and workload induction.

4.4 Data analysis

4.4.1 EEG-based classification of mental workload level and affective state

In this paper, we performed simultaneous EEG-based classification of mental workload level (i.e., "Easy" vs "Difficult") and affective state (i.e., "Relaxed" vs. "Stressed") using methods directly transferable to an online BCI. The classification was performed separately on each participant's data.

An entirely subject-specific as well as two cross-subject methods were investigated. In the entirely subject-specific case, the training data consisted of the first two blocks of data (one Relaxed and one Stressed) from the test participant, and the test data consisted of their final two blocks of data (one Relaxed and one Stressed). In the cross-subject cases, the training data consisted of the first two blocks of data from the test participant as well as all data from the 17 other subjects, while the test data again consisted of the final two blocks of data from the test participant. In the second cross-subject case, a transfer learning (TL) method recently proposed by [314] was used to match the distributions of the training and test data. In all cases, the “baseline” trials were omitted from analysis.

Prior to classification, the following preprocessing and feature extraction steps were applied to the training and test data.

4.4.2 EEG pre-processing

First, a band-pass filter was employed with low and high cut-off frequencies of 1 Hz and 50 Hz. The data were then downsampled from 500 Hz to 256 Hz, and Artifact Subspace Reconstruction (ASR) was used to reject signal segments containing EMG and motion artifacts.

To simulate an online application of the system, these preprocessing steps were applied to individual windowed segments (4 s length, 50% overlap) of the test data, in the order in which they were collected (i.e., chronological order). Specifically, each raw segment passed through all steps of the preprocessing, feature calculation, and classification, and predictions for both mental workload and stress level were made for that segment, before the next segment was processed. This is different from offline processing, where, typically, each step of the analysis (preprocessing, feature calculation, and classification) are performed on all test samples before proceeding to the next step. The analysis was performed on the same PC on which the pre-recorded EEG data were stored.

4.4.3 EEG signal feature calculation

Frequency domain features were computed to represent the characteristics of the EEG signals. Power signals in seven common frequency bands were extracted via the filter-Hilbert method [286], and average power was calculated over 4-s epochs with a sliding window with 50% overlap.

The individual alpha frequency (IAF) [285] was calculated using the one-minute eyes-closed baseline trial collected at the beginning of the experimental session and used to define the seven EEG frequency bands for each participant as follows: Delta (IAF-8 to IAF-6), Theta (IAF-6 to IAF-4), Alpha1 (IAF-4 to IAF-2), Alpha2 (IAF-2 to IAF), Alpha3 (IAF to IAF+2), Beta (IAF+2 to IAF+20), and Gamma (IAF+20 to IAF+30).

For the affective state classification problem (Stressed vs. Relaxed) all EEG channels were included, which resulted in a total of 63 electrodes × 7 frequency bands = 441 features. For the workload level classification problem (Easy vs. Difficult), since response frequency was greater for the Easy condition, we excluded the electrodes over the motor and sensorimotor brain regions (Cz, C1–C6, CPz, CP1–CP6) to ensure that incidental differences in the motor requirements of the Easy and Difficult conditions did not contribute to the classification. This resulted in a total of 49 electrodes × 7 frequency bands = 343 features for the workload level classification.

4.4.4 Classification

For both the mental workload level classification and the affective state classification, three classification paradigms were investigated and compared, as described below.

1. Subject-specific paradigm: For each target subject, the classifier was trained on the subject's first two blocks of data and tested on the subject's final two second blocks. A regularized Linear Discriminant Analysis (LDA) algorithm was used for classification [308].
2. Cross-subject without TL paradigm: For each target subject, the classifier was trained on the subject's first two blocks of data combined with all data from the other 17 subjects, and tested on the subject's final two blocks of data. No transfer

learning algorithm was applied. A regularized LDA algorithm was used for classification [308].

3. Cross-subject with TL paradigm: For each target subject, the classifier is trained on the subject's first two blocks of data combined with all data from the other 17 subjects, and tested on the subject's final two second blocks of data. The InstanceEasyTL transfer learning method was applied to reject the differences between data coming from different subjects; this method was originally proposed in [314] and is described in detail in the section below.

For each of the above classification problems, a majority vote-based approach proposed in our previous work to improve the classification of each state in the presence of the other was employed [313]. Specifically, in each case three classifiers were trained and the final predicted class was the result of a majority vote among the three. For the case of mental workload level classification, the first classifier was trained on data from the Relaxed state only, the second classifier was trained on data from the Stressed state only, and the third classifier was trained on data from both states. Similarly, for the affective state classification, the first classifier was trained on data from the Easy condition only, the second classifier was trained on data from the Difficult condition only, and the third classifier was trained on data from both conditions.

Note that the final two blocks of data consist of 12 trials, each of duration 67 seconds, for a total of 804 seconds of data. With 4 second epochs with 50% overlap, this yielded a classification every 2 seconds, for a total of 402 test samples per subject.

4.4.4.1 InstanceEasyTL algorithm

EEG-based mental state classification typically requires a large amount of labeled data for training. Given the non-stationarity of EEG data within a subject, as well as significant inter-subject variation, using data previously collected from other subjects, or even from the same subject on a previous day, often leads to poor classification accuracy [211, 212]. This is a challenge for developing real-world online BCI systems since impractically long calibration sessions are needed in order to gather sufficient data from the user to train the classifiers immediately before each use. Recently, transfer learning (TL) methods have been applied to mitigate this issue [213]. TL models are efficient methodologies that aim at transferring the previously extracted features from a labeled domain to a similar but different domain with limited or no labels, to perform some specific decision task on the different domain [315-318]. With such methods, for a given mental state classification task, EEG signals previously recorded from a set of subjects can be used to train a BCI system that will be suitable for any user, whether their data is in the training set or not. This is called cross-subject classification.

In this paper, we apply the InstanceEasyTL method, which was recently proposed in [314] to improve cross-subject EEG-based fatigue detection, to our objective of simultaneous classification of both mental workload level and stress. InstanceEasyTL is based on the EasyTL method [317] which was developed for image classification applications, but is adapted to work on EEG data, where there are often much larger differences in the target (test subject) and source (training subjects) domain distributions. Indeed, in [314], this method led to an increase of more than 15% compared to other existing transfer-learning methods such as transfer component analysis (TCA) [319], geodesic flow kernel (GFK) [320], and domain-adversarial neural networks (DANN) [321]. InstanceEasyTL is based on a "strategy of alignment with certain weights to align EEG samples collected from both source and target domains" [314]. To do this, InstanceEasyTL takes some EEG samples from the target domain Ω_t (i.e., the test subject's data) and combines it with all data from the original source domain Ω_s (i.e., the training subjects' data) to form a new source domain for training. This increases the amount of data available for training without increasing the amount of time needed for calibration prior to BCI use. As shown in Figure 4.1, the test subject's data (the initial target domain, Ω_t) is divided into two parts: S and T_{td}; T_{td} is added to the training subjects' data (the initial source domain, Ω_s, also called T_{sd} here) to form the new source domain Ω'_s, and S is the data from the test participant that ultimately undergoes classification (i.e., the new target domain, Ω'_t).

For our data set, we take the first 50% of the test subject's data (coming from the first two blocks) as T_{td} and combine it with the other 17 subjects' data (Ω_s) to create the new source domain Ω'_s, which is used for training the classifier. We then take the final 50% of the test subject's data (coming from the final two blocks) as S, the test data.

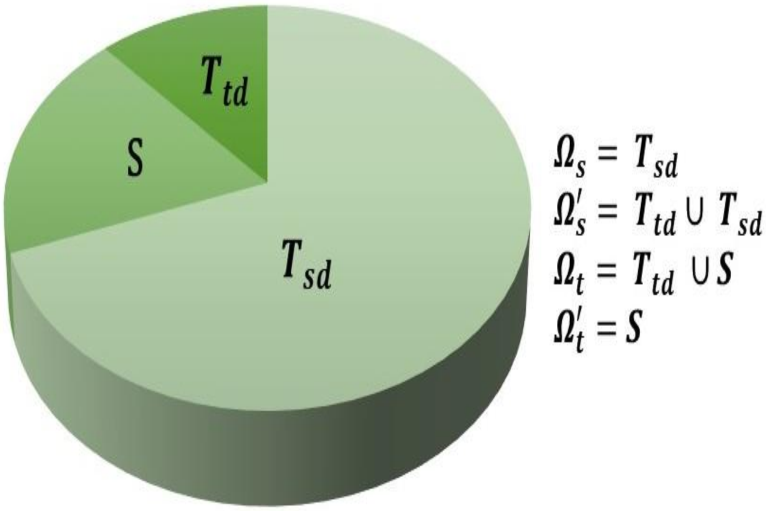


Figure 4.1: The original and new source and target domains. Ω_s and Ω_t are the original source and target domains, while Ω_s' and Ω_t' represent the new source and target domains used in the EasyInstanceTL algorithm. Figure adapted from [314].

A complete mathematical description of InstanceEasyTL algorithm was first proposed in [314]; for the convenience of the reader the main steps are summarized below.

First based on the method proposed in [317], intra-programming builds a classifier model by proposing a Probability Annotation Matrix W as in Equation 4.1, the rows of W denote the class label $c \in \{1, 2, \ldots, C\}$, and the column x_j^t represents the target samples. The element W_{cj} indicates the annotation probability of x_j^t belonging to class c. Based on the matrix W, the class of target samples are predicted. Note that the class labels of the target sample x_j^t that we choose are the corresponding ones with the maximum of $\{W_{cj}\}, j \in \{1, 2, \ldots, n_t\}$. For instance, if we have a Probability Annotation Matrix as in Equation 4.2, the class label of x_1^t will belong to class C_1 since it has the maximum probability of 0.7 among all elements $\{W_{11}, W_{21}\}$ with the probabilities of $\{0.7, 0.3\}$, respectively.

$$W = \begin{bmatrix} P(x_1^t \ |c = C_1) & P(x_2^t \ |c = C_1) & P(x_3^t \ |c = C_1) & \cdots & P(x_{n_t}^t \ |c = C_1) \\ P(x_1^t \ |c = C_2) & P(x_2^t \ |c = C_2) & P(x_3^t \ |c = C_2) & & P(x_{n_t}^t \ |c = C_2) \end{bmatrix} \tag{4.1}$$

$$W = \begin{bmatrix} 0.7 & 0.4 & 0.8 & \cdots & 0.3 \\ 0.3 & 0.6 & 0.2 & & 0.7 \end{bmatrix} \tag{4.2}$$

For the first iteration, t, of the InstanceEasyTL algorithm, initial weights are assigned to the data from both the training source domain, T_{sd}, and training target domain, T_{td}, (both from Ω_s') via Equation 4.3 [314]

$$\begin{gathered} W_{sd}^1 = \bigcup_{i=1}^{n_s} w_{sd}^i \\ W_{td}^1 = \bigcup_{i=n_s+1}^{n_s+m} w_{td}^i \\ W^1 = W_{sd}^1 \cup W_{td}^1 \end{gathered} \tag{4.3}$$

here, n_s and m are the number of samples in T_{sd} and T_{td}, respectively and w_{sd}^i and w_{td}^i are the weights for the i-th sample from T_{sd} and T_{td}, respectively. W_{sd}^1, W_{td}^1 and W^1 are the sets of weights after one iteration.

Next, the assigned weights for both T_{sd} and T_{td} are divided by the summation of all weights and stored as p_t, as shown in Equation 4.4 [314]

$$\begin{gathered} sum^t = \sum_{w \in W^t} w \\ p^t = \{\frac{w}{sum^t} \ ; w \in W^t\} \end{gathered} \tag{4.4}$$

After $t \in \{1, 2, \ldots, N\}$ iterations (N is the maximum number of iterations), w, W^t and sum^t represent the weight of one sample, the set of weights of all samples, and the sum of w in Ω_s', respectively, and p^t means the set of the weight w of each sample in Ω_s' in proportion of sum^t.

The training sample set $T = T_{sd} \cup T_{td}$ in Ω_s', p^t, and the test set S in Ω_t' are taken as input to the

InstanceEasyTL algorithm (though note that S is not used for the updating of weights in each iteration of the algorithm, but rather just for determining the predicted class labels in the final iteration) and the expected class label of h_t is calculated based on the intra-domain programming method (also called EasyTL) first detailed in [317].

Next, the error, ϵ_t, between the predicted class labels, h_t, and the real class labels, $y(x)$, is calculated via Equation 4.5 [314]

$$\epsilon_t = \frac{1}{W_t} * \sum_{\substack{x \in T_{td} \\ w_x \in W_{td}^t}} w_x |h_t(x) - y(x)| \tag{4.5}$$

The weights of T_{sd} and T_{td} are updated by the β_t-based function through Equation 4.6, detailed in [318]

$$\begin{gathered} \beta_t = \epsilon_t / (1 - \epsilon_t) \\ \beta = 1/(1 + (2 \ln n_s / N)^{\frac{1}{2}}) \end{gathered} \tag{4.6}$$

The weights are then updated via Equation 4.7 [314]

$$W^{t+1} = \bigcup_{\substack{x \in T_{sd} \\ w_x \in W_{sd}^t}} w_x \beta^{|h_t(x) - y(x)|} \cup \bigcup_{\substack{x \in T_{td} \\ w_x \in W_{td}^t}} w_x \beta^{|h_t(x) - y(x)|} \tag{4.7}$$

These steps (Equations 4.4 through 4.8) are repeated for N iterations. When $t = N$, the predicted class labels, $h_f(x)$, for the data in the test set, S, are calculated by Equation 4.8 [314]

$$h_f(x) = \begin{cases} 1, & \prod_{t=\lceil N/2 \rceil}^{N} \beta_t^{-h_t[x]} \geq \prod_{t=\lceil N/2 \rceil}^{N} \beta_t^{-1/2} \\ & 0, otherwise \end{cases} \tag{4.8}$$

4.5 Results

Table 4.1 shows the results of the mental workload level classification (Easy vs. Difficult) for all participants for the three classification paradigms. The cross-subject with transfer learning approach produced the best results with an average accuracy of 72.2% ± 5.3. This was 12.3% and 15.11% higher than the subject-specific and cross-subject without TL approach, respectively. A one-way ANOVA revealed a significant effect of classification approach ($F_{2,34} = 278.86; p < .001$), and post-hoc Tukey-Kramer tests indicated these increases were both significant ($|t_{(68)}| > 18.06; p < .001$). The result for the subject-specific approach was significantly higher than for the cross-subject without TL approach ($t_{(68)} = 4.14; p < .001$).

Table 4.1: Mental workload level classification results

Mental Workload Level Classification Results (Easy vs. Difficult)						
	Subject-Specific		Cross-Subject without TL		Cross-Subject with TL	
Subjects	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
1	66.7	0.66	58.2	0.58	73.5	0.73
2	49.3	0.51	47.2	0.48	64.2	0.65
3	64	0.63	59.7	0.59	74.8	0.74
4	65.8	0.65	59	0.6	75.0	0.75
5	53.0	0.54	51.3	0.52	65.9	0.66
6	52.5	0.53	53.8	0.53	68.8	0.67
7	56.4	0.56	50.5	0.5	66.5	0.65
8	58.1	0.57	50.9	0.5	63.4	0.63
9	57.7	0.58	58.6	0.58	73.5	0.74
10	66.5	0.66	59.7	0.59	77.2	0.76
11	59.2	0.6	56.8	0.55	71.6	0.72
12	57.6	0.57	55.9	0.56	69.5	0.68
13	66.6	0.66	65.5	0.66	79.8	0.79
14	67.7	0.67	67.7	0.67	81.4	0.81
15	60.3	0.61	61.3	0.62	80.5	0.8
16	60.8	0.6	56.0	0.56	72.3	0.71
17	58.8	0.58	58.5	0.58	72.7	0.72
18	57.7	0.57	57.3	0.57	69.6	0.7
Mean:	59.9 ± 5.3	0.59 ± 0.04	57.1 ± 5.1	0.56 ± 0.05	72.2 ± 5.3	0.71 ± 0.05

Table 4.2 shows the results of affective-state classification (Relaxed vs. Stressed) for all participants for the three classification paradigms. The best results again came from the cross-subject with transfer learning approach with an average accuracy of 74.2% ± 5.1, which exceeded

the accuracy for the subject-specific approach by 9.78% and for the cross-subject without TL paradigm by 15.91%. A one-way ANOVA revealed a significant effect of classification approach ($F_{2,34} = 154.95; p < .001$) and post-hoc Tukey-Kramer tests revealed that these differences were both significant ($|t_{(68)}| > 10.72; p < .001$). The result for subject-specific was significantly higher than cross-subject without TL approach ($t_{(68)} = 6.73; p < .001$).

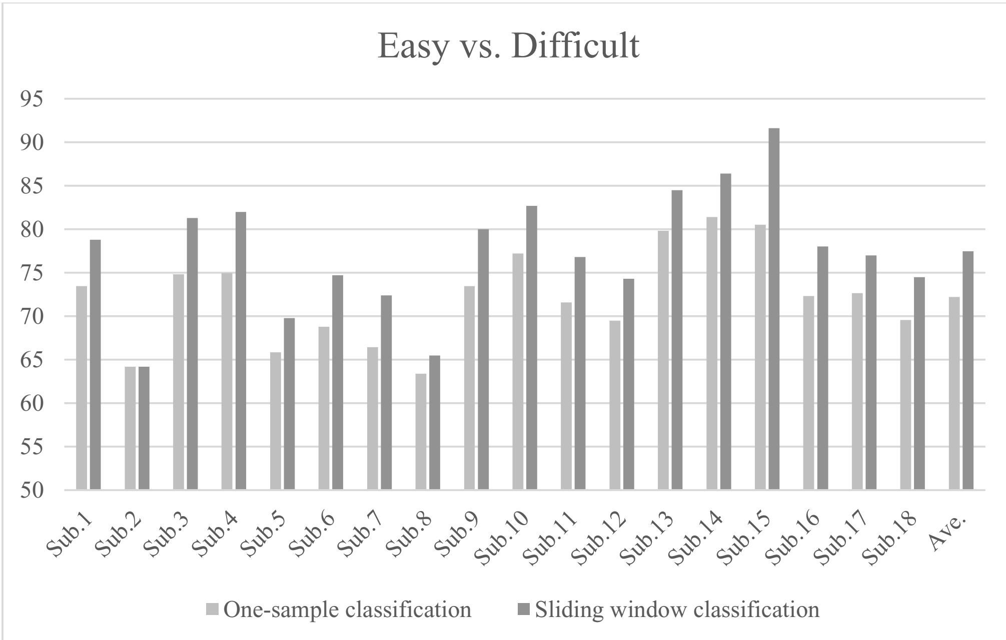
Since the classes are balanced for both the mental workload level classification and the affective state classification problems, the accuracy should be a valid measure of the classifier performance; however, the F1 score is given in Tables 4.1 and 4.2 as well (please refer to Appendix 1 for more detail on F1 score).

Table 4.2: Affective state classification results

Affective State Classification Results (Relaxed vs. Stressed)						
	Subject-Specific		Cross-Subject without TL		Cross-Subject with TL	
Subjects	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
1	68.8	0.69	65.1	0.64	78.6	0.78
2	60.0	0.59	47.7	0.47	67.4	0.66
3	61.6	0.61	62.4	0.61	79.0	0.78
4	64.4	0.64	64.2	0.63	80.0	0.79
5	60.8	0.61	51.4	0.52	71.5	0.73
6	65.5	0.65	50.9	0.51	68.4	0.68
7	60.5	0.6	54.8	0.54	67.8	0.67
8	60.6	0.61	56.5	0.56	66.9	0.67
9	64.4	0.64	56.1	0.56	74.2	0.75
10	64.0	0.63	64.9	0.63	78.5	0.77
11	59.0	0.59	57.6	0.56	71.3	0.71
12	65.3	0.64	56.0	0.56	69.9	0.7
13	72.8	0.71	65.0	0.64	81.9	0.81
14	73.6	0.74	66.1	0.67	79.0	0.78
15	67.2	0.67	58.7	0.58	80.2	0.81
16	62.3	0.62	59.2	0.58	77.0	0.78
17	66.8	0.66	57.2	0.57	74.3	0.74
18	62.7	0.63	56.1	0.56	70.3	0.71
Mean:	64.5 ± 4.1	0.64 ± 0.04	58.3 ± 5.4	0.57 ± 0.05	74.2 ± 5.1	0.74 ± 0.05

By the binomial test [322], the lower limit for statistical significance in a binary classification problem with $n = 402, p = 0.5$ and α=0.05 is 54.2%. For the cross-subject with TL approach, accuracies for all subjects, for both classification problems, exceed this value by at least 9.2%.

To further improve online classification accuracy by reducing incorrect predictions resulting from sudden changes in the EEG signals, we applied a sliding window classification in which the final predicted class for each sample was determined by majority vote of the output of the InstanceEasyTL classifier (i.e., the cross-subject with TL approach) for that sample and the two previous samples. This method results in a prediction every 2 seconds, same as before. Using this sliding window classification method for the cross-subject with TL paradigm, the average accuracy for the Easy vs. Difficult classification significantly increased by 5.3% to 77.5% ± 6.9 ($t_{17} = -10.12; p < .001$) and the average accuracy for Relaxed vs. Stressed classification significantly increased by 9.9% to 84.1% ± 5.9 ($t_{17} = -16.98; p < .001$) over all subjects. Figure 4.2 shows the classification accuracies using one-sample prediction as compared to the sliding window classification for all subjects.



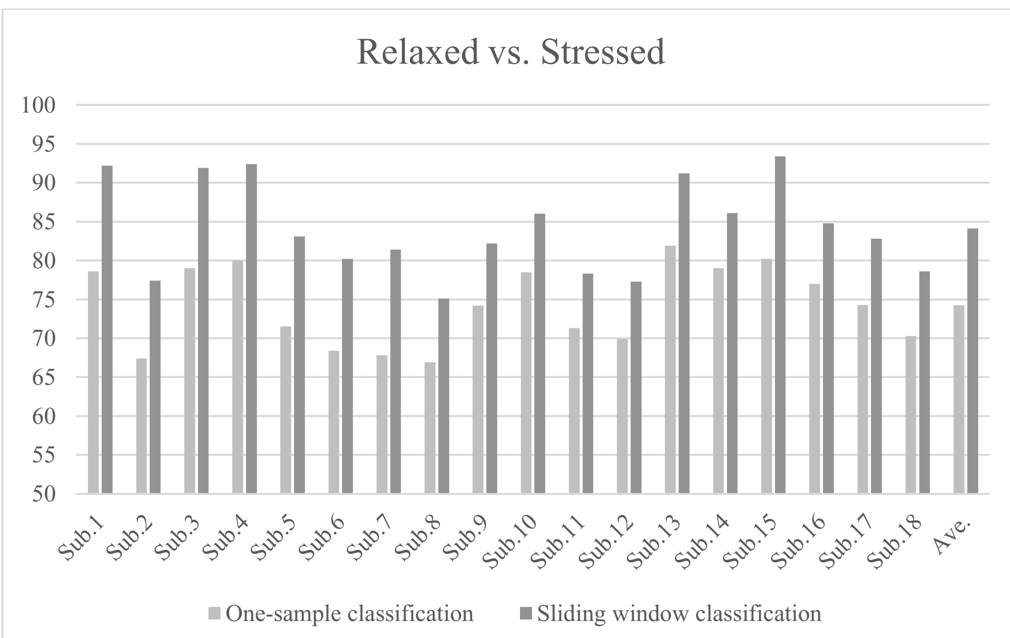


Figure 4.2: Classification accuracy of the cross-subject with TL approach with and without the application of sliding window classification.

Figure 4.3 shows a simulation of online classification for one participant (Subject #15) for both the workload level and affective state classification problems. The figure shows the predicted class output using the sliding window classification every 2 seconds (recall classification was done over 4 second epochs with a sliding window with 50% overlap) for the final two blocks of the session. The classifier for each individual sample was trained using the cross-subject with TL approach. As can be seen in the figure, the predicted classes follow the actual class labels with high accuracy - 91.6% and 93.4% for workload level and affective state classification problems, respectively.

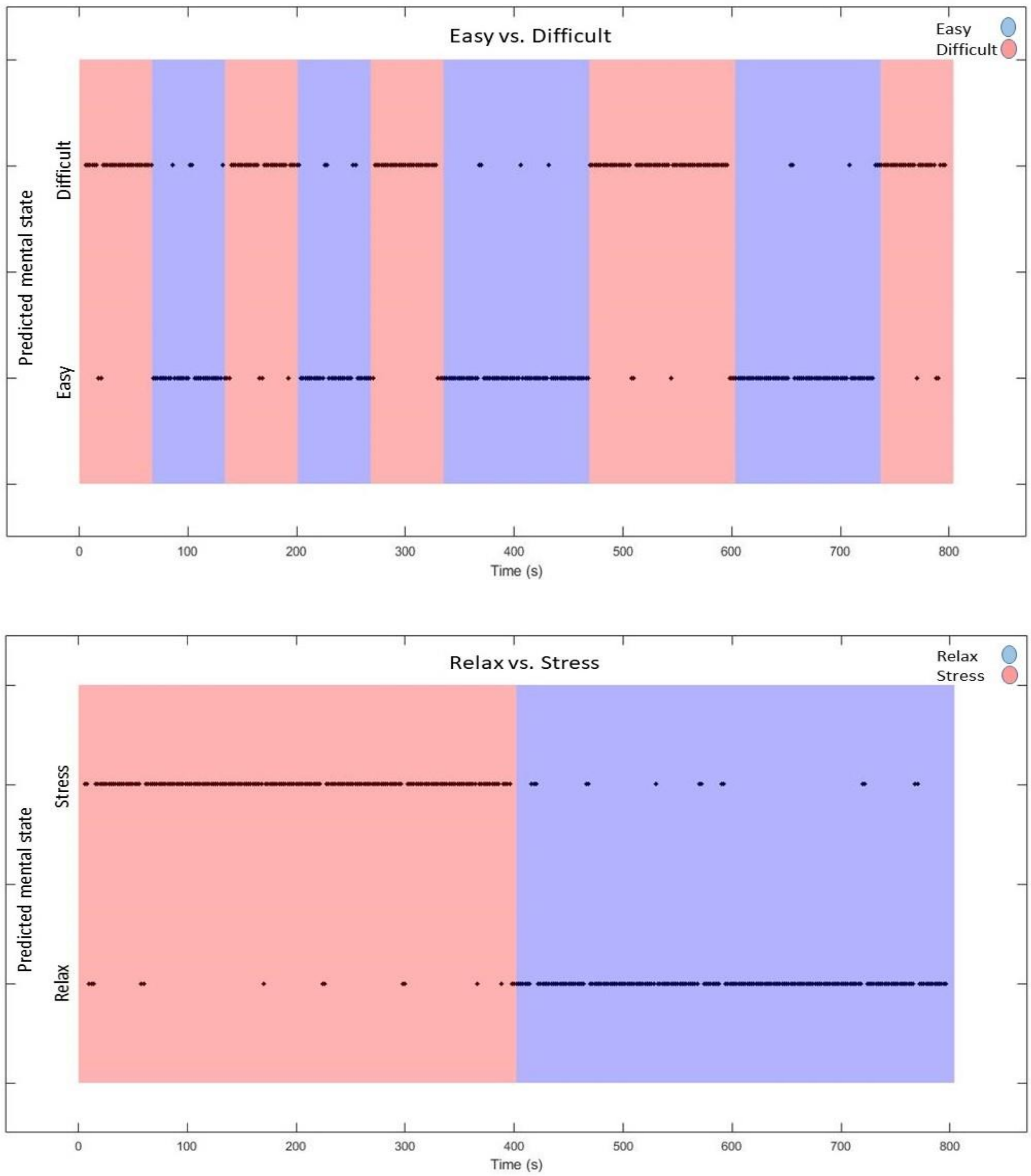


Figure 4.3: Simulated online output of the system for Subject 15 for both the Easy vs. Difficult and Relaxed vs. Stressed classification problems. After training the classifier on the first two blocks of data (combined with the data from all other subjects), consecutive samples from the final two blocks were classified (epochs were of length 4 s, with 2 s overlap). Classification was done using the cross subject with TL approach, with a sliding window classification over three samples. The shaded intervals indicate the actual mental state, while the black dots indicate the predicted state.

4.6 Discussion

The main objective of this study was to investigate the feasibility of performing simultaneous classification of mental workload and stress level in an online passive BCI. We investigated both subject-specific and cross-subject classification approaches, the latter with and without the application of a transfer learning technique called InstanceEasyTL [314] to align the distributions of data from the training and test subjects. In the subject-specific case, the first 50% of the individual's data was used to train a regularized LDA classifier, and the final 50% of data was used as test data. For the cross-subject cases, the data from the 17 other participants was also added to the training set. Though done offline, all steps of the analysis - including pre-processing, feature selection and classification - were done in a manner completely compatible with online implementation.

Our results showed that mental workload level (Easy vs. Difficult) and affective state (Relaxed vs. Stressed) could be classified in manner suitable for online implementation with accuracies of 77.5% ± 6.9 and 84.1% ± 5.9 respectively, across 18 participants. Accuracies significantly exceeded chance (54.2% in this case) for all participants, for both classification problems. These results were achieved using cross-subject classification with transfer learning, which gave significantly better results than the other methods (i.e., entirely subject-specific or cross subject without TL applied, with the same amount of training data taken from the test subject).

To the best of our knowledge at the time of writing, this study represents the first attempt to perform simultaneous classification of mental workload level and stress in an online passive BCI, though there are many studies that have looked at one or the other condition individually through offline applications. Actually, online BCI studies are generally rather scarce. Recently, [323] presented an EEG-based classification of four mental states (fatigue, workload, distraction, and the normal state) for seven pilots in both offline and pseudo-online analyses. They proposed a multiple feature block-based convolutional neural network (MFB-CNN) with spatio-temporal EEG filters to recognize the pilot's current mental states. In the pseudo-online analysis, they conducted an evaluation between one of the mental states (fatigue, workload and distraction) and rest states and obtained the detection accuracy of 72%, 72%, and 61% for fatigue, workload, and distraction, respectively. In [324], a novel strategy named adaptive subspace feature matching (ASFM) for cross-domain EEG-based emotion recognition was proposed. ASFM integrates both the marginal and conditional distributions. Both offline and online evaluations were performed and, the average classification accuracy of 75.1% ± 7.6 was achieved in the subject-to-subject evaluation for 15 subjects for the online analysis. It is worth mentioning that we employed the ASFM transfer learning algorithm to our data as well but did not achieve promising results. Even though it is difficult to directly compare our study to [323] and [324] due to differences in the experimental (e.g., type of task, number/length of trials) and analytical (e.g., pre-preprocessing techniques, EEG features used and classification algorithms) methods employed, the accuracies obtained are similar. And the fact that our study involved a potentially much more complex scenario – that is, simultaneous classification of two states, where both states were confounding one another – makes our results even more encouraging.

Note that the InstanceEasyTL algorithm allowed us to achieve satisfactory classification accuracies despite a relatively small amount of training data taken from the test subject immediately prior to "online" classification. To simulate an online application of the system, the classifier was trained on the target subject's first two blocks of data and tested on the subject's final two blocks, to keep the data continuity in time. Therefore, about 14.5 minutes (12 trials x 67 seconds/trial + 67-second eye-closed baseline trial for extracting IAF) of training data was used from the test subject; that is a fairly reasonable length of time for system calibration prior to use. In the simulated online testing, about 402 epochs were tested one by one as time progressed. The time taken for data preprocessing, feature extraction and classification of each 4-second epoch was 0.22, 0.59 and 0.03 seconds, respectively, for a total of less than 1 second (0.84 second) per epoch. All the algorithms were done using MATLAB R2019b with an Intel core i7-8th Gen processor.

Due to restrictions on research involving human participants arising due to the Covid-19 global pandemic, we were unable to conduct a separate experiment for online implementation and instead used the data previously collected and reported in [304] and [313] to simulate exactly an online classification scenario completely suitable for direct online implementation. Future work will involve realizing the cross-subject mental workload level and stress detection algorithms in an actual online application, and evaluating them in more realistic, ecologically-valid task scenarios in which users experience different levels of workload and affective-state.

4.7 Conclusions

In this study, we investigated the ability to classify both mental workload level and affective state simultaneously using methods appropriate for implementation in an online BCI. Using the InstanceEasyTL transfer learning algorithm proposed in [314], we achieved accuracies of 77.5% and 84.1% for mental workload level and affective state classification, respectively, using a database of “previous” subjects and just 13.5 minutes of training data from the test subject. Classification was performed every two seconds. These results are very promising, and support the feasibility of developing a practical, online passive BCI for use in realistic scenarios where both the cognitive and affective state of the user will be changing over time.

4.8 Acknowledgments

The authors acknowledge with gratitude the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 5 : Conclusion

5.1 Summary of contributions

This thesis makes several original contributions to the fields of biomedical engineering, and specifically to the domain of passive brain-computer interfacing.

Specifically, in this thesis I have:

1. Developed an experimental paradigm for the simultaneous and independent induction of mental workload and stress.
2. Found that while two mental states (i.e., stress, and mental workload due to task difficulty) could be simultaneously classified at levels significantly exceeding chance, variation in each state had deleterious effects on the classification of the other by EEG; and detection of both states is significantly diminished in the presence of variation of the other state.
3. Found that the accuracy with which workload levels (as defined exclusively in terms of task difficulty) could be classified was affected by whether the classifier training and testing data were collected under the same or different affective states. When the classifier was trained on data from a relaxed (stressed) condition and tested on data from a relaxed (stressed) condition – i.e., the “within-affective-context” classification paradigm – the mean classification accuracy across 18 participants for the “Easy” vs. “Difficult” task conditions was approximately 73%. However, when the classifier was trained on data from the relaxed (stressed) condition and tested on data from the stressed (relaxed) condition – i.e., the “across-affective-context” classification paradigm - the mean accuracy dropped to approximately 56%. This suggests that in a real passive BCI system, if the mental workload detection algorithm were to be trained using data collected during a calibration session

where the individual experienced a single - likely relaxed - affective state, performance could deteriorate significantly when applied to real-life scenarios which are likely to induce varying levels of stress in the user. It was found that including data from both the stressed and relaxed states in the training set (i.e., the "combined-affective-context" classification paradigm) improved the classifier performance to approximately 67% - significantly higher than in the "across-affective-context" case, but still significantly below the ideal "within-affective-context" case (i.e., where there is no variation in affective state).

4. Found that variation in the difficulty level of the task that the individual was performing had similarly negative effects on the performance of algorithms classifying affective state. The mean classification accuracy for the "Relaxed" vs. "Stressed" conditions across 18 participants was approximately 85% and 74% for the "within-workload-level" and "across-workload-level" classification paradigms, respectively. Including data from both workload conditions in the training set (i.e., the "combined-workload-level" classification paradigm) resulted in an accuracy of 82%, which was again significantly better than the "across-affective context" paradigm, but still significantly below the ideal "within-workload-level" classification paradigm (i.e., where these is no variation in mental workload).
5. Proposed and compared five classification approaches to improve 1) mental workload-level detection in the presence of varying affective-state and 2) affective-state detection in the presence of varying workload-level. Statistically significant improvements in classification accuracy were observed for both mental workload and affective state classification using some of the proposed methods.
6. Showed that mental workload level (Easy vs. Difficult) and affective state (Relaxed vs. Stressed) could be simultaneously classified in a manner suitable for online implementation

with accuracies of 77.5% ± 6.9 and 84.1% ± 5.9 respectively, across 18 participants. Accuracies significantly exceeded chance (54.2% in this case) for all participants, for both classification problems. These results were achieved using cross-subject classification with transfer learning, which gave significantly better results than the other investigated methods (i.e., entirely subject-specific or cross subject without TL applied, with the same amount of training data taken from the test subject).

Collectively, this research study represents the first attempt to perform simultaneous classification of mental workload level and stress in the pBCI literature. Such a technology could have significant industrial and economic impact by preventing accidents related to operator error, and their associated human, economic, and environmental losses during unsafe mental states, such as high mental workload and stress. The results from this study will help not only in simultaneous detection of mental workload level and affective state but perhaps other mental states as well. The ultimate goal of this research is to make passive BCIs closer to being integrated into daily life applications.

5.2 Study limitations and future work

The idea that stress and workload are linked was the main and first motivation to pursue this work. In the current study, it was found that the two states indeed confound each other and variation in one state negatively affects the classification of the other state, and that is why the subject's affective state should be considered when classifying workload level (and vice versa, the subject's workload level should be considered when classifying affective state) for a BCI to be effective in real-life scenarios. However, it was imperative for the purpose of this research to induce mental workload and stress as independently of one another as possible so that I would be able to investigate the effect of variation in each state on the classification of the other state. In other words, I wanted to make sure the workload induction protocol, by itself, didn't inadvertently induce stress, and that the stress induction protocol, by itself, didn't inadvertently induce workload. Therefore, I tried to make the high workload task as different from the low task as possible in terms of difficulty while keeping it manageable enough that the participants would not feel any additional stress due to the task itself in this condition. Similarly, I opted for a stress induction protocol that would not directly add workload to the task (for example, adding any time limit on the task to induce stress may require more mental effort). However, future work will also investigate mental workload and stress at the highest end of workload, where the two are often (though not necessarily) highly correlated. While in many scenarios the induction of mental workload and stress may be inextricably linked, there is a need to determine if detection of the states can still be done independently but simultaneously.

In addition, in this work, two levels of mental workload (i.e. Easy/Low and Difficult/High) and two levels of stress (i.e. Relaxed and Stressed) were considered, future work will involve increasing the number of workload and affective state levels, and investigating the effect of variation in each state on classification of the other states. In a real application, when the user is experiencing different levels of workload while his/her affective state is changing simultaneously and continuously, a BCI must be able to decide the user's affective state and operate based on that trying to increase the performance of the system.

Due to restrictions on research involving human participants arising due to the Covid-19 global pandemic, I was unable to conduct a separate experiment for online implementation and instead used the data previously collected to simulate exactly an online classification scenario completely suitable for direct online implementation (reported in Chapter 5). Future work will involve realizing the cross-subject mental workload level and stress detection algorithms in an actual online application, and evaluating them in more realistic, ecologically-valid task scenarios. In fact, BCIs should be evaluated and tested in circumstances like those of real-life applications. This testing can reveal how well the BCI adapts to spontaneous variation in the signal features when it does not have the advantage of knowing what the output is supposed to be. In this study, precautions were taken in designing affective state induction protocol so that stress can be induced by a real life scenario (i.e. public speaking task). However, future work will investigate different affective state induction protocols and workload tasks applicable in real life scenarios to determine if the proposed detection algorithms remain effective.

Furthermore, users are likely to differ greatly over time in the prominence and stability of specific signal features related to the mental states of interest. Those few research programs that have acquired long-term data have found that marked variations in performance typically occur over minutes, hours, days, weeks, and months. In the current study, the Stressed and Relaxed conditions in our experimental protocol were interleaved twice allowing to rule out the possibility that the results were merely due to the effect of time. However, it is suggested to gather and analyze data from each user many times, over substantial periods. The result obtained over a short session experiment, while encouraging, is not sufficient to be confident that the classification algorithm will remain stable across a longer session/across sessions.

Further work should also involve the investigation, or incorporation, of additional mental states.

5.3 Resulting publications

1. M. Bagheri, and S. Power, “EEG-based detection of mental workload level and stress: the effect of variation in each state on classification of the other,” Journal of Neural Engineering, vol. 17, 2020.

2. M. Bagheri, and S. Power, “Investigating hierarchical and ensemble classification approaches to mitigate the negative effect of varying stress state on EEG-based detection of mental workload level - and vice versa,” Brain Computer Interfaces, vol. 8, no. 1-2, 2021.
3. M. Bagheri, and S. Power, “Simultaneous classification of both mental workload and stress level suitable for an online passive brain-computer interface,” Sensors, vol. 22, no. 2, Jan 2022.

Bibliography

[1] C. Brunner et al., “BNCI Horizon 2020: towards a roadmap for the BCI community,” *Brain Computer Interfaces*, vol. 2, no. 1, pp. 1-10, 2015.

[2] J. R. Wolpaw, “Brain–computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–79, 2002.

[3] J. R. Wolpaw, and D. J. McFarland, "Multichannel EEG-based brain-computer communication," *Electroencephalography and clinical Neurophysiology*, vol. 90, no. 6, pp. 444-449, 1994.

[4] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, “Brain-computer interface—A new communication device for handicapped persons,” *J. Microcomp. Appl*., vol. 16, pp. 293–299, 1993.

[5] J. R. Wolpaw et al., “An EEG-based brain-computer interface for cursor control,” *Electroencephalogr Clin Neurophysiol*. vol. 78, no. 3, pp. 252-9, 1991. doi: 10.1016/0013-4694(91)90040-b

[6] L. A. Farwell, and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalogr Clin Neurophysiol.* vol. 70, no. 6, pp. 510-23, 1988. doi: 10.1016/0013-4694(88)90149-6.

[7] R. Mamunur et al., “Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review”, *Front. Neurorobot*., vol. 14, 2020. https://doi.org/10.3389/fnbot.2020.00025.

[8] S. Simanto et al., “Progress in Brain Computer Interface: Challenges and Opportunities,” *Front. Syst. Neurosci.,* vol. 15, 2021. https://doi.org/10.3389/fnsys.2021.578875

[9] L. F. Nicolas-Alonso, and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211-79, 2012. doi: 10.3390/s120201211.

[10] 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. doi: 10.1088/1741-2560/8/2/025005

[11] A. Vallabhaneni, T. Wang, and B. He, "Brain—Computer Interface," In: *He B. (eds) Neural Engineering. Bioelectric Engineering,* Springer, Boston, MA, 2005. https://doi.org/10.1007/0-306-48610-5_3.

[12] E. Kandel et al., *Principles of Neural Science*, 5th Edition, New York, N.Y.: McGraw-Hill Education LLC., 2013.

[13] P. L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, New York, 1981.

[14] P. L. Nunez et al., "A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging," *Electroencephalography and Clinical Neurophysiology*, vol. 90, pp. 40-57, 1994.

[15] M. Tudor, L. Tudor, K. I. Tudor, "Hans Berger (1873-1941)-the history of electroencephalography," *Acta Med Croatica,* vol. 59, no. 4, pp. 307-13, 2005.

[16] J. Malmivuo, and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, Oxford University press, 2012. DOI: 10.1093/acprof:oso/9780195058239.001.0001

[17] J. Puikkonen, and J. A. Malmivuo, *Theoretical investigation of the sensitivity distribution of point EEG-electrodes on the three concentric spheres model of a human head: An application of the reciprocity theorem*, Tampere University of Technology, 1987.

[18] A. S. Al-Fahoum, and A. A. Al-Fraihat, “Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domain,” *Hindawi Publishing Corporation ISRN Neuroscience,* vol. 2014.

[19] F. Lotte, and M. Congedo, “EEG feature extraction,” in *Brain-Computer Interfaces 1: Methods and Perspectives*, Wiley-ISTE, 2016.

[20] A. M. Brouwer et al., “Estimating workload using EEG spectral power and ERPs in the n-back task,” *Journal of Neural Engineering*, vol. 9, no. 4, 2012.

[21] J. H. Friedman, “On bias, variance, 0/1-loss, and the curse-of-dimensionality,” *Data Mining and Knowledge Discovery,* vol. 1, no. 1, pp. 55–77, 1997.

[22] A. Jain et al., “Statistical pattern recognition: A review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp.4-37, 2000.

[23] F. Lotte, “A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces,” in *Guide to Brain-Computer Music Interfacing,* E. R. Miranda and J. Castet, Springer, 2014.

[24] S. J. Raudys, and A. K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 13, no. 3, pp. 252-264, 1991.

[25] F. Lotte et al., “A review of classification algorithms for EEG-based brain–computer interfaces: a 10-year update,” *Journal of Neural Engineering*, vol. 4, pp. 28, 2018.

[26] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, “Filter Methods for Feature Selection – A Comparative Study,” In: *Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds) Intelligent Data Engineering and Automated Learning,* Springer, Berlin, Heidelberg, 2007. https://doi.org/10.1007/978-3-540-77226-2_19

[27] R. Kohavi, and G. John, “Wrappers for feature subset selection,” *Artificial Intelligence Journal*, Special issue on relevance, vol. 97, no. 1-2, pp. 273–324, 1997.

[28] C. C. Aggarwal, *Data Classification: Algorithms and Applications*, 1st. edition, Chapman & Hall/CRC, 2014.

[29] F. Lotte et al., “A review of classification algorithms for EEG-based brain–computer interface,” *Journal of Neural Engineering*, vol. 4, pp. 24, 2007.

[30] C. Tremmel et al., “Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG,” *Frontiers in Human Neuroscience*, vol. 13, no. 401, 2019.

[31] R. N. Roy et al., “Efficient mental workload estimation using task-independent EEG features,” *Journal of Neural Engineering*, vol. 13, no. 2, pp. 026019, 2016.

[32] D. Guido et al., "Improving Human Performance in a Real Operating Environment through Real-Time Mental Workload Detection," in *Toward Brain-Computer Interfacing*, MIT Press, 2007, pp. 409-422.

[33] P. Aricò et al., “A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks,” *Progress in Brain Research*, vol. 228, pp. 295-328, 2016.

[34] G. Borghini et al., “EEG-Based Cognitive Control Behavior Assessment: an Ecological study with Professional Air Traffic Controllers,” *Sci Rep*, vol. 7, no. 1, pp. 547, 2017.

[35] F. Dehais et al., “Monitoring Pilot’s Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions,” *Sensors,* vol. 19, no. 6, 2019.

[36] I. Kakkos et al., “Mental Workload Drives Different Reorganizations of Functional Cortical Connectivity Between 2D and 3D Simulated Flight Experiments,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 27, no. 9, pp. 1704-1713, 2019.

[37] Y. Zhang et al., “Investigating different stress-relief methods using Electroencephalogram (EEG),” *In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* Montreal, QC, Canada, 2020; pp. 2999–3002.

[38] O. Attallah, “An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes,” *Diagnostics,* vol. 10. pp. 292, 2020.

[39] V. Vanitha, and P. Krishnan, “Real time stress detection system based on EEG signals,” *Biomed Res. Int.*, vol. 27, pp. 271–275, 2016.

[40] D. Dasari et al., “ICA-derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task,” *Front. Neurosci*. vol. 11, pp. 297, 2017.

[41] G. Borghini et al., “Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness,” *Neurosci. Biobehav. Rev.,* vol. 44, pp. 58–75, 2014.

[42] R. N. Roy et al., “Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI” *Conf. Proc. IEEE Eng. Med. Biol. Soc*. pp 6607–10, 2013.

[43] G. Yang et al., “A Driver Fatigue Recognition Model based on Information Fusion and Dynamic Bayesian Network,” *Inf. Sci.*, vol. 10, pp. 1942–1954, 2010.

[44] R. Roy et al., “Time-on-task effect on workload level discriminability through electroencephalography,” *in Proceedings of BIOMAG 2012 (Paris),* 2012.

[45] Y. Tran et al., "Detecting neural changes during stress and fatigue effectively: a comparison of spectral analysis and sample entropy," in 3rd International IEEE Conference on Neural Engineering, 2007, pp. 350-353.

[46] A. Craig et al., “Regional brain wave activity changes associated with fatigue,” *Psychophysiology,* vol. 49, pp. 574–582, 2012. doi: 10.1111/j.1469-8986.2011. 01329.x

[47] G. Dornhege et al., "Evaluation Criteria for BCI Research," in *Toward Brain-Computer Interfacing*, MIT Press, 2007, pp.327-342.

[48] L. W. Ko et al., “Sustained attention in real classroom settings: an EEG study,” *Front. Hum. Neurosci*. vol. 11, p. 388, 2017.

[49] C. T. Lin et al., “Tonic and phasic EEG and behavioral changes induced by arousing feedback,” *Neuroimage*, vol. 52, pp.633–642, 2010. doi: 10.1016/j.neuroimage.2010.04.250

[50] K. J. Verdière, R. N. Roy, and F. Dehais, “Detecting Pilot’s engagement using fNIRS connectivity features in an automated versus manual landing scenario*,” Front. Hum. Neurosci.* vol. 12, 2018.

[51] G. Borghini et al., “Quantitative assessment of the training improvement in a motor-cognitive task by using EEG, ECG and EOG signals,” *Brain Topogr*. vol. 29, pp. 149–61, 2016.

[52] B. Kenny et al., “Assessment of changes in neural activity during acquisition of spatial knowledge using EEG signal classification,” *Journal Neural Engineering*, vol. 16, no. 3, 2019.

[53] R. Biswas et al., “Investigation of an EEG-based Indicator of Skill Acquisition as Novice Participants Practice a Lifeboat Maneuvering Task in a Simulator*,” International Journal of Human-Computer Interaction*, vol. 36, no. 8, pp. 1-11., 2019. doi: 10.1080/10447318.2019.1687235

[54] H. Cai, and Y. Lin, “Modelling of Operators’ Emotion and Task Performance in a Virtual Driving Environment,” *Int. J. Hum. Comput. Stud*. Vol. 69, no. 9, pp. 571–586, 2011.

[55] A. M. Brouwer et al., 2017 “Physiological effects of adaptive cruise control behaviour

in real driving," *Proc. 2017 ACM Workshop on An Application-Oriented Approach to BCI Out of the Laboratory*, 2017. pp. 15-19. https://doi.org/10.1145/3038439.3038441

[56] K. Giannakaki et al., "Emotional state recognition using advanced machine learning techniques on EEG data," in *30th IEEE International Symposium on Computer-Based Medical Systems*, 2017.

[57] J. Toppi et al., 2016 "Investigating cooperative behavior in ecological settings: an EEG hyperscanning study," *PLoS One,* vol. 11, no. 4, 2016.

[58] N Sciaraffa et al., "Brain interaction during cooperation: evaluating local properties of multiple-brain network," *Brain Sci*., vol. 7, no. 90, 2017.

[59] Y. H. Yu et al., "New flexible silicone-based EEG dry sensor material compositions exhibiting improvements in lifespan, conductivity, and reliability," *Sensors*, vol. 16, 2016.

[60] J. Guixeres et al., "Consumer neuroscience-based metrics predict recall, liking and viewing rates in online advertising," *Front. Psychol*,. vol. 8, 2017.

[61] P. Cherubino et al., "Neuroelectrical Indexes for the Study of the Efficacy of TV Advertising Stimuli," In: *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska (eds), Springer Proceedings in Business and Economics. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-28419-4_22

[62] G. Cartocci et al., "Electroencephalographic, heart rate, and galvanic skin response assessment for an advertising perception study: application to antismoking public service announcements," *J. Vis. Exp*., vol. 28, 2017.

[63] S. Wang et al., "Online Prediction of Driver Distraction Based on Brain Activity Patterns," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 136-150, Feb. 2015, doi: 10.1109/TITS.2014.2330979.

[64] H. S. Almahasneh et al., "EEG based driver cognitive distraction assessment," *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS),* 2014, pp. 1-4, doi: 10.1109/ICIAS.2014.6869460

[65] Y. Liang et al., "Prediction of drowsiness events in night shift workers during morning driving'" *Accid. Anal. Prev.,* vol. 126, pp. 105-114, 2017. https://doi.org/10.1016/j.aap.2017.11.004

[66] Z. Khaliliardali et al., "Single trial classification of neural correlates of anticipatory behavior during real car driving," *Proc. 6th Int. Brain–Computer Interface Meeting* (https://infoscience.epfl.ch/ record/218935), 2016.

[67] J. Perrier et al., "Driving performance and EEG fluctuations during on the-road driving following sleep deprivation," *Biol. Psychol*., vol. 121, no. 1, 2016.

[68] J. Kohlmorgen et al., "Improving human performance in a real operating environment through real-time mental workload detection," in *Toward Brain Computer Interfacing,* eds G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Cambridge, MA: MIT Press, 2007, pp. 409–422

[69] P. Aricò et al., "Adaptive automation triggered by EEG-based mental workload index: a passive brain–computer interface application in realistic air traffic control environment," *Front Hum. Neurosci*., vol. 26, 2016.

[70] F. Dehais et al., "EEG-engagement index and auditory alarm misperception: an inattentional deafness study in actual flight condition," *Advances in Neuroergonomics and Cognitive Engineering (Advances in Intelligent Systems and Computing)*, vol. 586, 2018.

[71] L. L. Di Stasi et al., "Task complexity modulates pilot electroencephalographic activity during real flights," *Psychophysiology*, vol. 52, pp. 951–6, 2015.

[72] C. A. Scholl et al., “Classification of pilot-induced oscillations during inflight piloting exercises using dry EEG sensor recordings,” *38th Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc. (EMBC),* 2016, pp 4467–70.

[73] T. O. Zander et al., “Automated task load detection with electroencephalography: towards passive brain–computer interfacing in robotic surgery,” *J. Med. Robot. Res.*, vol. 2, no. 1, 2017. http://dx.doi.org/10.1142/S2424905X17500039

[74] H. Gauba et al., “Prediction of advertisement preference by fusing EEG response and sentiment analysis,” *Neural Netw.* vol. 92, pp. 77–88, 2017.

[75] M. Mazher et al., “An EEG-Based Cognitive Load Assessment in Multimedia Learning Using Feature Extraction and Partial Directed Coherence,” in *IEEE Access*, vol. 5, pp. 14819-14829, 2017.

[76] J. M. Angel, R. Cristian, and L. Hernan, “Using Psychophysiological Sensors to Assess Mental Workload During Web Browsing,” *Sensors*, vol. 18, no. 2, pp. 458, 2018.

[77] M. A. Almogbel, A. H. Dang, and W. Kameyama, "Cognitive Workload Detection from Raw EEG-Signals of Vehicle Driver using Deep Learning." *International Conference on Advanced Communication Technology (ICACT),* pp. 1-6, 2018.

[78] A. Abrantes et al., “Classification of EEG Features for Prediction of Working Memory Load,” In: *Advances in The Human Side of Service Engineering*, *Advances in Intelligent Systems and Computing*, T. Ahram and W. Karwowski (eds), Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-41947-3_12

[79] G. Liang et al., “Evaluation and prediction of on-line maintenance workload in nuclear power plants,” *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 19, no. 1, pp. 64-77, 2009.

[80] R. Boff, L. Kaufman, and J. P. Thomas, *Handbook of Perception and Human Performance*, vol. 2, Cognitive Processes and Performance. New York, NY: John Wiley & Sons.

[81] S. Wang, J. Gwizdka, and W. A. Chaovalitwongse, "Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 424-435, 2016.

[82] P. Zarjam, J. Epps, and N. H. Lovell, "Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, pp. 301-310, 2015.

[83] Z. Y. Chin et al., "EEG-based discrimination of different cognitive workload levels from mental arithmetic," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2018.

[84] N. Friedman et al., "EEG-Based Prediction of Cognitive Load in Intelligence Tests," *Frontiers in Human Neuroscience*, vol. 13, no. 191, 2019.

[85] A. Knoll et al., "Measuring cognitive workload with low-cost electroencephalograph." *International Conference on Human Computer Interaction*, pp. 568-571, 2011.

[86] K. Yu et al., "Cognitive workload modulation through degraded visual stimuli: a single-trial EEG study," *Journal of Neural Engineering*, vol. 12, no. 4, pp. 046020, 2015.

[87] P. Arico et al., "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2015, pp. 7242-7245.

[88] M. K. Johnson et al., "Probe-independent EEG assessment of mental workload in pilots," *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 581-584, doi: 10.1109/NER.2015.7146689.

[89] M. E. Smith et al., “Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 3, pp. 366-80, 2001.

[90] P. K. Jao et al., “Analysis of EEG Correlates of Perceived Difficulty in Dynamically Changing Flying Tasks,” *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* 2018.

[91] L. R. Fournier, G. F. Wilson, and C. R. Swain, “Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training,” *International Journal of Psychophysiology,* vol. 31, no. 2, pp. 129-145, 1999.

[92] L. J. Prinzel et al., “Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component,” *Human Factors*, vol. 45, no. 4, pp. 601-614, 2003.

[93] Z. Yin et al., “Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework,” *Neurocomputing,* vol. 347, pp. 212-229, 2019.

[94] P. Zarjam et al., “Estimating cognitive workload using wavelet entropy-based features during an arithmetic task,” *Computers in Biology and Medicine*, vol. 43, no. 12, pp. 2186-2195, 2013.

[95] G. N. Dimitrakopoulos et al., “Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1940-1949, 2017.

[96] G. F. Wilson, and C. A. Russell, “Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding,” *Human Factors*, vol. 49, no. 6, pp. 1005-1018, 2007.

[97] S. I. Dimitriadis et al., “102,” *Annals of Biomedical Engineering*, vol. 43, no. 4, pp. 977-989, 2015.

[98] G. F. Wilson, and C. A. Russell, “Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task,” *Human Factors*, vol. 45, no. 3, pp. p.381-389, 2003.

[99] C. L. Baldwin, and B. N. Penaranda, “Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification,” *Neuroimage,* vol. 59, no. 1, pp. 48-56, 2012.

[100] B. N. Penaranda, and C. L. Baldwin, “Temporal Factors of EEG and Artificial Neural Network Classifiers of Mental Workload,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* vol. 56, no. 1, pp. 188-192, 2012.

[101] J. Tao et al., “Individual-Specific Classification of Mental Workload Levels Via an Ensemble Heterogeneous Extreme Learning Machine for EEG Modeling,” *Symmetry*, vol. 11, no. 7, pp. 944, 2019.

[102] M. Arvaneh, A. Umilta, and I. H. Robertson, "Filter bank common spatial patterns in mental workload estimation," *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 4749-4752.

[103] A. Appriou, A. Cichocki, and F. Lotte, "Towards Robust Neuroadaptive HCI: Exploring Modern Machine Learning Methods to Estimate Mental Workload from EEG Signals," *Human Factors in Computing Systems,* 2018.

[104] Z. Dai et al., “Mental workload classification in n-back tasks based on single-trial EEG,” *Chinese Journal of Scientific Instrument*, vol. 38, pp. 1335-1344, 2017.

[105] W. Yi et al., “Estimation of mental workload induced by different presentation rates in rapid serial visual presentation tasks, *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2019.

[106] A. Duraisingam, R. Palaniappan, and S. Andrews, “Cognitive task difficulty analysis using EEG and data mining,” *2017 Conference on Emerging Devices and Smart Systems (ICEDSS),* 2017.

[107] C. Berka et al., “Real-time analysis of EEG indices of alertness, cognition, and memory with a wireless EEG headset,” *Int J Hum Comput Interact*, vol. 17, pp. 151– 70, 2004.

[108] E. Galy, M. Cariou, and C. Mélan, “What is the relationship between mental workload factors and cognitive load types?” *Int. J. Psychophysiol*., vol. 83, pp. 269–275, 2012. doi: 10.1016/j.ijpsycho.2011.09.023

[109] A. Zammouri et al., “Brain waves-based index for workload estimation and mental effort engagement recognition,” *Journal of Physics: Conference Series,* vol. 904, 2017, doi: 10.1088/1742-6596/904/1/012008.

[110] G. Borghini et al., “Brainshield HREEG study of perceived pilot mental workload,” *Italian Journal of Aerospace Medicine,* vol. 5, pp. 34–47, 2011.

[111] F. Yamada, “Frontal midline theta rhythm and eye blinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics,” *Ergonomics*, vol. 41, no. 5, pp. 678–688, 1998.

[112] A. Smith et al., "Monitoring task loading with multivariate EEG measures during forms of human-computer interaction," *Human Factors*, vol. 43, no. 3, pp. 366–380, 2001.

[113] C. Dussault et al., "EEG and ECG changes during simulator operation reflect mental workload and vigilance," *Aviation, Space, and Environmental Medicine*, vol. 76, pp. 344–351, 2005.

[114] C. Berka et al., "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, Space, and Environmental Medicine,* vol. 78 no. 5, pp. B231–B244, 2007.

[115] M. E. Smith, L. K. McEvoy, and A. Gevins, "The impact of moderate sleep loss on neurophysiologic signals during working-memory task performance," *Sleep*, vol. 25, no. 7, pp. 784–794, 2002.

[116] Brookings JB, Wilson GF, Swain CR. "Psychophysiological responses to changes in workload during simulated air traffic control," *Biol Psychol.,* vol. 42, no. 3, pp. 361-77, 1996. doi: 10.1016/0301-0511(95)05167-8.

[117] C. Mühl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," *Frontiers Neurosci.,* vol. 8, p. 114, 2014.

[118] G. Borghini et al., "Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC),* Aug. 2012, pp. 6442–6445.

[119] I. E. de Vries, J. Van Driel, and C. N. Olivers, "Posterior alpha EEG dynamics dissociate current from future goals in working memory guided visual search," *J. Neurosci*., vol. 37, pp. 1591–1603, 2017. doi: 10.1523/JNEUROSCI.2945-16.2016

[120] S. H. Fairclough, L. Venables, and A. Tattersall, “The influence of task demand and learning on the psychophysiological response,” *International Journal of Psychophysiology,* vol. 56, no. 2, pp. 171–184, 2005.

[121] M. Doppelmayr, T. Finkenzeller, and P. Sauseng, “Frontal midline theta in the pre-shot phase of rifle shooting: differences between experts and novices,” *Neuropsychologia,* vol. 46, no. 5, pp. 1463–1467, 2008.

[122] S. M. Slobounov et al., “Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance,” *Cognitive Brain Research,* vol. 9, no. 3, pp. 287–298, 2000.

[123] P. Zarjam, J. Epps, and F. Chen, “Characterizing working memory load using EEG delta activity,” in *Proc. 19th Eur. Signal Process. Conf.,* 2011, pp. 1554–1558.

[124] C. Dijksterhuis et al., “Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns,” *Frontiers Neurosci*., vol. 7, p. 149, 2013.

[125] P. Bashivan, M. Yeasin, and G. M. Bidelman, "Single trial prediction of normal and excessive cognitive load through EEG feature fusion," *IEEE Signal Processing in Medicine and Biology Symposium,* pp. 1-5, 2015

[126] C. Walter et al., “Using Cross-Task Classification for Classifying Workload Levels in Complex Learning Tasks,” *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 876-881.

[127] P. Zarjam, J. Epps, and F. Chen, "Spectral EEG features for evaluating cognitive load," *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 2011, pp. 3841-3844, doi: 10.1109/IEMBS.2011.6090954.

[128] W. K. Y. So et al., “An evaluation of mental workload with frontal EEG,” *Plos One*, vol. 12, no. 4, pp. e0174949, 2017.

[129] J. Zhang, Z. Yin, and R. Wang, “Recognition of Mental Workload Levels Under Complex Human–Machine Collaboration by Using Physiological Features and Adaptive Support Vector Machines,” *IEEE Transactions on Human Machine Systems*, vol. 45, no. 2, pp. 200-214, 2015.

[130] J. Zhang, Z. Yin, and R. Wang, “Pattern Classification of Instantaneous Cognitive Task-load Through GMM Clustering, Laplacian Eigenmap, and Ensemble SVMs,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, pp. 947-965, 2017.

[131] Z. Yin, and J. Zhang, “Operator functional state classification using least-square support vector machine based recursive feature elimination technique,” *Computer Methods & Programs in Biomedicine*, vol. 113, no. 1, pp. 101-115, 2014.

[132] Z. Wang et al., “Cross-subject workload classification with a hierarchical Bayes model,” *Neuroimage,* vol. 59, no. 1, pp. 64-69, 2012.

[133] Z. Jiao et al., “Deep Convolutional Neural Networks for mental load classification based on EEG data,” *Pattern Recognition*, vol. 76, pp. 582-595, 2018.

[134] P. Zhang et al., “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.

[135] Y. Zhang, and Y. Shen, “Parallel Mechanism of Spectral Feature-Enhanced Maps in EEG-Based Cognitive Workload Classification,” *Sensors*, vol. 19, no. 4, pp. 808, 2019

[136] S. Kuanar et al., "Cognitive Analysis of Working Memory Load from EEG by a Deep Recurrent Neural Network," *International Conference on Acoustics, Speech, and Signal Processing*, 2018, pp. 2576-2580.

[137] R. G. Hefron et al., "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognition Letters*, vol. 94, pp. 96-104, 2017.

[138] Y. Zhong, and J. Zhang, "Recognition of Cognitive Task Load Levels Using Single Channel EEG and Stacked Denoising Autoencoder," *Proceedings of the 35th Chinese Control Conference IEEE*, 2016, pp. 3907-3912.

[139] Z. Yin, and J. Zhang, "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model," *Biomedical Signal Processing and Control*, vol. 33, pp. 30-47, 2017.

[140] S. Yang et al., "Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders," *Computers in Biology and Medicine*, pp. 159-170, 2019.

[141] J. Zhang, and S. Li, "A deep learning scheme for mental workload classification based on restricted Boltzmann machines," *Cognition, Technology & Work*, vol. 19, no. 4, pp. 607-631, 2017.

[142] Z. Yin, and J. Zhang, "Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights," *Neurocomputing*, vol. 260, no. 18, pp. 349-366, 2017.

[143] M. Plechawska-Wójcik et al., "A Three-Class Classification of Cognitive Workload Based on EEG Spectral Data," *Applied sciences*, vol. 9, no. 24, pp. 5340, 2019.

[144] Y. Ke et al., "An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task," *Frontiers in Human Neuroscience*, vol. 8, pp. 703, 2014.

[145] J. A. Blanco et al., "Quantifying Cognitive Workload in Simulated Flight Using Passive, Dry EEG Measurements," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 373-383, 2018.

[146] W. Qiao, and X. Bi, "Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload," *Biomedical Signal Processing and Control*, vol. 57, pp. 101745, 2020.

[147] P. Zhang et al., "Spectral and Temporal Feature Learning with Two-Stream Neural Networks for Mental Workload Assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1149-1159, 2019.

[148] P. Bashivan et al., "Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks," *arXiv,* 2015, https://doi.org/10.48550/arXiv.1511.06448

[149] Y. Kwak et al., "Multilevel Feature Fusion with 3D Convolutional Neural Network for EEG Based Workload Estimation," *IEEE Access*, no. 99, pp. 16009-16021, 2020

[150] R. G. Hefron et al., "Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks," *Sensors*, vol. 18, no. 5, pp. 1339, 2018

[151] G. F. Wilson, and C. A. Russell, "Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks," *Human Factors*, vol. 45, no. 4, pp. 635-644, 2003.

[152] R. Matthews et al., "Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes." *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, 2008, pp. 5871-5875.

[153] G. Zhao, Y. Liu, and Y. Shi, “Real-Time Assessment of the Cross-Task Mental Workload Using Physiological Measures During Anomaly Detection,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 149-160, 2018.

[154] C. Jeunet et al., “Mind the Traps! Design Guidelines for Rigorous BCI Experiments” in *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances,* S. N. Chang, A. Nijholt and F. Lotte., CRC Press pp.1-33, 2018.

[155] S. Samima, and M. Sarma, “EEG-Based Mental Workload Estimation,” *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2019.

[156] M. Parent et al., “The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS?” *Int J Psychophysiol.*, vol. 146, pp. 139-147, 2019, doi: 10.1016/j.ijpsycho.2019.09.005.

[157] R. Flin et al., *Decision-Making Under Stress: Emerging Themes and Applications*, Routledge; 1 edition London, UK, 1998

[158] N. Skoluda et al., “Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity,” *Psychoneuroendocrinology,* vol. 51, pp. 227-236, 2015.

[159] J. H. M. Tulen et al., “Characterization of stress reactions to the Stroop Color Word Test,” *Pharmacology Biochemistry and Behavior*, vol. 32, pp. 9-15, 1989.

[160] X. Hou et al., "EEG based stress monitoring," *In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics*, Hong Kong, China, 9–12 October 2015; pp. 3110–3115.

[161] K. Masood, K., and M. A. Alghamdi, "Modeling Mental Stress Using a Deep Learning Framework," *IEEE Access*, vol. 7, pp. 68446–68454, 2019.

[162] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test' - A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting," *Neuropsychobiology*, vol. 28, pp. 76-81, 1993.

[163] S. Lotfan et al., "Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures," *Biocybern. Biomed. Eng.*, vol. 39, pp. 199–213, 2019.

[164] J. Minguillon et al., "Portable system for real-time detection of stress level," *Sensors*, vol. 18, pp. 2504, 2018.

[165] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomed. Signal Process. Control*, vol. 46, pp. 18–32, 2018.

[166] F. Al-Shargie, T. B. Tang, and M. Kiguchi, "Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: An fNIRS-EEG study," *Biomed. Opt. Express*, vol. 8, pp. 2583–2598, 2017.

[167] A. R. Subhani et al., "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access,* vol. 5, pp. 13545–13556, 2017.

[168] F. Al-Shargie et al., "Mental stress assessment using simultaneous measurement of EEG and fNIRS," *Biomed. Opt. Express*, vol. 7, pp. 3882–3898, 2016.

[169] F. Al-Shargie et al., "Mental stress quantification using EEG signals," In *Proceedings of the International Conference for Innovation in Biomedical Engineering and Life Sciences*, Putrajaya, Malaysia, 6–8 December 2015; pp. 15–19.

[170] F. Al-shargie, "Simultaneous measurement of EEG-fNIRS in classifying and localizing brain activation to mental stress," *In Proceedings of the 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA),* Kuala Lumpur, Malaysia, 19–21 October 2015; pp. 282–286.

[171] C. H. Poh et al., "The effect of antireflux treatment on patients with gastroesophageal reflux disease undergoing a mental arithmetic stressor," *Neurogastroenterology & Motility*, vol. 23, pp. e489-e496, 2011.

[172] F. Al-Shargie, "Prefrontal cortex functional connectivity based on simultaneous record of electrical and hemodynamic responses associated with mental stress," *arXiv 2021*, https://doi.org/10.48550/arXiv.2103.04636

[173] M. Zanetti et al., "Assessment of mental stress through the analysis of physiological signals acquired from wearable devices," In *Proceedings of the Italian Forum of Ambient Assisted Living*, Lecce, Italy, 2–4 July 2018; pp. 243–256.

[174] J. W. Ahn, Y. Ku, and H. C. Kim, "A novel wearable EEG and ECG recording system for stress assessment," *Sensors*, vol. 19, no. 9, p. 1991, 2019.

[175] A. Saidatul et al., "Analysis of EEG signals during relaxation and mental stress condition using AR modeling techniques," In *Proceedings of the 2011 IEEE International Conference on Control System, Computing and Engineering*, Penang, Malaysia, 25–27 November 2011; pp. 477–481.

[176] D. Shon et al., “Emotional stress state detection using genetic algorithm-based feature selection on EEG signals,” *Int. J. Environ. Res. Public Health,* vol. 15, pp. 1–11, 2018.

[177] M. J. Hasan, and J. M. Kim, “A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals,” *Brain Sci*, vol. 9, p. 376, 2019.

[178] G. Giannakakis, D. Grigoriadis, and M. Tsiknakis, “Detection of stress/anxiety state from EEG features during video watching,” *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2013.

[179] R. S. Lewis, N. Y. Weekes, and T. H. Wang, “The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health,” *Biol. Psychol.,* vol. 75, pp. 224–239. 2007, doi: 10.1016/j.biopsycho.2007.03.004

[180] D. Katarina et al. “The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain.” *Journal of psychiatry & neuroscience: JPN,* vol. 30, no.5, pp. 319-25, 2005.

[181] J. R. Stroop, “Studies of interference in serial verbal reactions,” *Journal of Experimental Psychology,* vol. 18, no. 6, pp. 643–662, 1935. https://doi.org/10.1037/h0054651.

[182] S. A. Hosseini and M. A. Khalilzadeh, "Emotional Stress Recognition System Using EEG and Psychophysiological Signals: Using New Labelling Process of EEG Signals in Emotional Stress State," *2010 International Conference on Biomedical Engineering and Computer Science*, 2010, pp. 1-6, doi: 10.1109/ICBECS.2010.5462520.

[183] R. Khosrowabadi, “Stress and Perception of Emotional Stimuli: Long-term Stress Rewiring the Brain,” *Basic Clin. Neurosci*, vol. 9, no. 2, pp. 107-120, 2018.

[184] R. Khosrowabadi et al., “A Brain-Computer Interface for classifying EEG correlates of chronic mental stress,” In *Proceedings of the 2011 International Joint Conference on Neural Networks*, San Jose, CA, USA, 31 July–5 August 2011; pp. 757–762

[185] S. Betti et al., “Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers,” *IEEE Trans. Biomed. Eng*., vol. 65, pp. 1748–1758, 2017.

[186] A. Darzi, H. Azami, and R. Khosrowabadi, “Brain functional connectivity changes in long-term mental stress,” *J. Neurodev. Cogn.*, vol. 1, pp. 16–41, 2019.

[187] Z. Halim, and M. Rehan, “On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning,” *Inf. Fusion,* vol. 53, pp. 66–79, 2020.

[188] G. Jun, and K. G. Smitha, "EEG based stress level identification," *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003270-003274, doi: 10.1109/SMC.2016.7844738.

[189] M. Zanetti et al., “Multilevel assessment of mental stress via network physiology paradigm using consumer wearable devices,” *J. Ambient Intell. Humaniz. Comput*., vol. 12, pp. 4409–4418, 2019.

[190] H. Jebelli, S. Hwang, and S. Lee, “EEG-based workers’ stress recognition at construction sites,” *Autom. Constr.* vol. 93, pp. 315–324, 2018.

[191] D. Devi et al., “Brain wave based cognitive state prediction for monitoring health care conditions,” *Mater. Today Proc.* pp. 1–7, 2020. https://doi.org/10.1016/j.matpr.2020.09.616.

[192] A. Secerbegovic et al., “workload vs. stress differentiation using single-channel EEG,” In *CMBEBIH 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 511–515.

[193] S. Alonso et al., "Stress assessment based on EEG univariate features and functional connectivity measures," *Physiological measurement*, vol. 36, p. 1351, 2015.

[194] F. Al-Shargie et al., “Towards multilevel mental stress assessment using SVM with ECOC: An EEG approach,” *Med Biol. Eng. Comput*., vol. 56, pp. 125–136, 2018.

[195] A. Demerdzieva, "EEG characteristics of generalized anxiety disorder in childhood," *Acta Informatica Medica,* vol. 19, p. 9, 2011.

[196] S. Seo, and J. Lee, "Stress and EEG", in *Convergence and Hybrid Information Technologies*. London, United Kingdom: IntechOpen, 2010 [Online]. Available: https://www.intechopen.com/chapters/10986 doi: 10.5772/9651.

[197] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, "Stress Assessment by Prefrontal Relative Gamma," *Frontiers in Computational Neuroscience*, vol. 10, 2016.

[198] T. Waili et al., “Stress recognition using Electroencephalogram (EEG) signal”, *Journal of Physics: Conference Series*; vol. 1502, no. 1, 2020. DOI:10.1088/1742-6596/1502/1/012052

[199] S. Cheemalapati et al., "A real-time classification algorithm for emotion detection using portable EEG," *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI),* 2013, pp. 720-723, doi: 10.1109/IRI.2013.6642541.

[200] U. R. Acharya et al., "Application of nonlinear and wavelet based features for the automated identification of epileptic EEG signals," *Int J Neural Syst,* vol. 22, 2012.

[201] J. Hewig et al., “Associations of the cortisol awakening response (CAR) with cortical activation asymmetry during the course of an exam stress period,” *Psychoneuroendocrinology*, vol. 33, pp. 83–89, 2008. doi: 10.1016/j.psyneuen.2007.10.004

[202] S. Seo, Y. Gil, and J. Lee, “The relation between affective style of stressor on EEG asymmetry and stress scale during multimodal task,” In *the proceedings of Convergence and Hybrid Information Technology*, 2008, pp. 461-466, 2008.

[203] J. A. Coan, and J. J. B. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion,” *Biol. Psychol.,* vol. 67, pp. 7-49, 2004.

[204] S. A. Hosseini et al., "Higher order spectra analysis of EEG signals in emotional stress states," in *Conference on Information Technology and Computer Science*, 2010.

[205] N. Sharma and T. Gedeon, "Modeling a stress signal," *Applied Soft Computing*, vol. 14, pp. 53-61, 2014.

[206] K. E. Ko, H. C. Yang, and K. B. Sim, "Emotion recognition using EEG signals with relative power values and Bayesian network," *International Journal of Control, Automation and Systems,* vol. 7, p. 865, 2009.

[207] P. D. Purnamasari and A. Fernandya, "Real Time EEG-based Stress Detection and Meditation Application with K-Nearest Neighbor," *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC) (47129)*, 2019, pp. 49-54, doi: 10.1109/R10-HTC47129.2019.9042488

[208] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Computers in Biology and Medicine*, 2019. https://doi.org/10.1016/j.compbiomed.2019.02.015

[209] S. A. Hosseini, and M. B. Naghibi-Sistani, "Classification of Emotional Stress Using Brain Activity", in *Applied Biomedical Engineering*. London, United Kingdom: IntechOpen, 2011 [Online]. Available: https://www.intechopen.com/chapters/18027 doi: 10.5772/18294

[210] M. A. Khalilzadeh et al., "Qualitative and quantitative evaluation of brain activity in emotional stress," *Iranian Journal of Neurology*, vol. 8, pp. 605-618, 2010.

[211] D. J. Krusienski et al., (2011). “Critical issues in state-of-the-art brain–computer interface signal processing,” *J. Neural Eng.*, vol. 8, 2011. doi: 10.1088/1741-2560/8/2/025002

[212] G. E. Hine, E. Maiorana, and P. Campisi, “Resting-State EEG: A Study on Its Non-Stationarity for Biometric Applications,” *2017 International Conference of the Biometrics Special Interest Group (BIOSIG),* 2017.

[213] Z. Wan et al., “A review on transfer learning in EEG signal analysis,” *Neurocomputing,* vol. 421, pp. 1-14, 2021.

[214] R. A. Ramadan, and A. V. Vasilakos, “Brain computer interface: control signals review,” *Neurocomputing,* vol. 223, pp. 26–44, 2017.

[215] H. Dose et al., “An end-to-end deep learning approach to MI-EEG signal classification for BCIs,” *Expert Syst. Appl.,* vol. 114, pp. 532–542, 2018.

[216] Y. P. Lin, and T. P. Jung, “Improving EEG-based emotion classification using conditional transfer learning,” *Front. Human Neurosci.,* vol. 11, p. 334, 2017.

[217] D. P. Subha et al., “EEG signal analysis: a survey,” *J. Med. Syst.,* vol. 34, pp. 195–212, 2010.

[218] J. C. Shaw, “Correlation and coherence analysis of the EEG: a selective tutorial review,” *Int. J. Psychophysiol.,* vol. 1, pp, 255–266, 1984.

[219] A. M. Azab et al., “Dynamic time warpingbased transfer learning for improving common spatial patterns in brain computer interface,” *J. Neural Eng.,* vol. 17, p. 16061, 2020.

[220] H. Phan et al., “Deep transfer learning for single-channel automatic sleep staging with channel mismatch,” *2019 27th European Signal Processing Conference (EUSIPCO),* 2019, pp. 1–5.

[221] S. Grissmann et al., "Context Sensitivity of EEG-Based Workload Classification Under Different Affective Valence," *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 327-334, 2020, doi: 10.1109/TAFFC.2017.2775616.

[222] Ayaz H, et al. “Optical brain monitoring for operator training and mental workload assessment”, *Neuroimage,* vol. 59, no. 1, pp. 36-47, 2012.

[223] T. Gateau et al., “Real-Time State Estimation in a Flight Simulator Using fNIRS”, *PloS One,* vol. 10, no. 3, 2015.

[224] H. Aghajani, M. Garbey, and A. Omurtag, “Measuring Mental Workload with EEG+fNIRS,” *Frontiers in Human Neuroscience*, vol. 11, 2017.

[225] B. Z. Allison, and J. Polich, “Workload assessment of computer gaming using a single-stimulus event-related potential paradigm,” *Biological psychology*, vol. 77, no. 3, pp. 277-283, 2008.

[226] M. Chaouachi, I. Jraidi, and C. Frasson, “Adapting to Learners’ Mental States Using a Physiological Computing Approach”, In *Florida Artificial Intelligence Research Society Conference*, 2015.

[227] L. M. Hirshfield et al., “This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy”, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2011, pp. 373-382.

[228] T. Nagasawa et al., “Improving fNIRS-BCI accuracy using GAN-based data augmentation,” *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER),* 2019.

[229] N. K. Qureshi et al., “Comparison of classification performance for fNIRS-BCI system,” *2016 2nd International Conference on Robotics and Artificial Intelligence (ICRAI),* 2016.

[230] G. Zhu et al., “A simplified hybrid EEG-fNIRS Brain-Computer Interface for motor task classification,” *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER),* 2017.

[231] S. B. Borgheai et al., “Towards a Single Trial fNIRS-based Brain-Computer Interface for Communication,” *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER),* 2019.

[232] T. Yang et al., “Image-based Motor Imagery EEG Classification using Convolutional Neural Network,” *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI),* 2019.

[233] S. G. Hart, and L. Staveland, “Development of NASA-TLX (Task Load Index): results of empirical and theoretical research,” in *Human Mental Workload,* eds P. A. Hancock and N. Meshkati, Amsterdam: Elsevier, 1988, pp. 139–183.

[234] E. Koyas et al., “Detection of intention level in response to task difficulty from EEG signals,” *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP),* 2013.

[235] A. Rahimpour et al., “Classification of fNIRS based brain hemodynamic response to mental arithmetic tasks,” *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA),* 2017.

[236] G. Durantin et al., “Using near infrared spectroscopy and heart rate variability to detect mental overload,” *Behavioural Brain Research,* vol. 259, pp. 6-23, 2014. https://doi.org/10.1016/j.bbr.2013.10.042.

[237] M. W. Eysenck, and M. G. Calvo, “Anxiety and performance: The processing efficiency theory,” *Cognition & Emotion*, vol. 6, no. 6, pp. 409-434, 1992.

[238] S. A. Hosseini, and M. B. Naghibi-Sistani, “Emotion recognition method using entropy analysis of EEG signals,” *International Journal of Image, Graphics and Signal Processing*, vol. 3, no. 5, pp. 30, 2011.

[239] K. Tai, and T. Chau, “Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface,” *Journal of NeuroEngineering and Rehabilitation*, vol. 6, no. 39, 2009.

[240] D. Navon, and D. Gopher, “On the economy of the human-processing system,” *Psychol Rev*. vol. 86, pp. 214–255, 1979. doi: 10.1037/0033-295X.86.3.214

[241] C. D. Wickens, “Multiple resources and performance prediction,” *Theor. Issues Ergon. Sci*., vol. 3, pp. 159–177, 2002. doi: 10.1080/14639220210123806

[242] W. J. Horrey, and C. D. Wickens, “Multiple Resource Modeling of Task Interference in Vehicle Control, Hazard Awareness and In-vehicle Task Performance”, *Driving Assesment Conference*, 2003, pp.7-12. doi: https://doi.org/10.17077/drivingassessment.1082

[243] L. Longo, “Mental workload in medicine: foundations, applications, open problems, challenges and future perspectives,” *in IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS) (Dublin),* 2016, pp. 106–111.

[244] J. Wolpaw, and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*, 1st Edn. Oxford; New York, NY: Oxford University Press, 2012.

[245] A. Gevins et al., “Monitoring working memory load during computer-based tasks with EEG pattern recognition methods,” *Hum Factors,* vol. 40, pp. 79–91, 1998.

[246] S. Makeig, and T. P. Jung, “Changes in alertness are a principal component of variance in the EEG spectrum,” *Neuroreport*, vol. 7, pp. 213–6, 1995.

[247] C. W. Pleydell-Pearce, S. E. Whitecross, and B. T. Dickson, “Multivariate analysis of EEG: predicting cognition on the basis of frequency decomposition, inter-electrode correlation, coherence, cross phase, and cross power,” *Proceedings of the 36th Annual Hawaii International Conference on System Sciences,* 2003, p. 131.

[248] M. B. Sterman, C. A. Mann, and D. A. Kaiser, “Quantitative EEG patterns of differential in-flight workload,” *Space Operations, Applications, and Research Proceedings*. 1992. Sepulveda VA Medical Center: NASA Conference Publication; 1992; pp. 466–73.

[249] G. F. Wilson, and F. T. Eggemeier, “Physiological measures of workload in multi-task environments,” In: *Multiple-task performance*. D. L. Damos ed., London: Taylor & Francis, Ltd. 1991; pp. 329–360.

[250] G. F. Wilson, “An analysis of mental workload in pilots during flight using multiple psychophysiological measures,” *Int J Aviat Psychol*, vol. 12, pp. 3–18, 2004.

[251] M. Sterman et al., "Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1993, Seattle, WA: SAGE Publications, pp. 118–121.

[252] M. Sterman, and C. A. Mann, "Concepts and applications of EEG analysis in aviation performance evaluation," *Biol. Psychol.,* vol. 40, pp. 115–130, 1995. doi: 10.1016/0301-0511(95)05101-5.

[253] C. Berka, D. J. Levendowski, and G. Davis, "EEG indices distinguish spatial and verbal working memory processing: implications for real-time monitoring in a closed-loop tactical Tomahawk weapons simulation," *in Proceedings of the 1st International Conference on Augmented Cognition,* 2005.

[254] W. S. Helton, and P. N. Russell, "Working memory load and the vigilance decrement," *Exp. Brain Res.*, vol. 212, pp. 429–437, 2011. doi: 10.1007/s00221-011-2749-1

[255] S. Yamamoto, and S. Matsuoka, "Topographic EEG study of visual display terminal (VDT) performance with special reference to frontal midline theta waves," *Brain Topogr.* vol. 2, pp. 257–267, 1990. doi: 10.1007/BF01129654

[256] A. Gundel, and G. F. Wilson "Topographical changes in the ongoing EEG related to the difficulty of mental tasks," *Brain Topogr.* vol. 5, pp. 17–25, 1992. doi: 10.1007/BF01129966

[257] Ray, W. J., and Cole, H. W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. Science 228, 750–752. doi: 10.1126/science.3992243

[258] A. Keil et al., “Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials,” *Clin. Neurophysiol.,* vol. 112, pp. 2057–2068, 2001. doi: 10.1016/S1388-2457(01)00654-X.

[259] A. Brookings et al., “Towards measurement of brain function in operational environments,” *Biol. Psychol.,* vol. 40, pp. 169–186, 1995. doi: 10.1016/0301-0511(95)05105-8

[260] P. Ullsperger, A. M. Metz, and H. G. Gille, “The P300 component of the event-related brain potential and mental effort,” *Ergonomics*, vol. 31, pp.1127–1137, 1988. doi: 10.1080/00140138808966752

[261] M. A. Boksem, T. F. Meijman, and M. M. Lorist, “Effects of mental fatigue on attention: an ERP study,” *Cogn. Brain Res.,* vol. 25, pp. 107–116, 2005. doi: 10.1016/j.cogbrainres.2005.04.011

[262] J. Onton, A. Delorme, and S. Makeig, “Frontal midline EEG dynamics during working memory,” *Neuroimage,* vol. 27, pp. 341–356, 2005. doi: 10.1016/j.neuroimage.2005.04.014

[263] L. J. Trejo, A. F. Kramer, and J. A. Arnold, “Event-related potentials as indices of display-monitoring performance,” *Biol. Psychol.,* vol. 40, pp. 33–71, 1995. doi: 10.1016/0301-0511(95)05103-1

[264] B. Blankertz et al., “The Berlin brain-computer interface: progress beyond communication and control,” *Front. Neurosci.,* vol. 10, 2016. doi: 10.3389/fnins.2016.00530

[265] Hans Selye, *The Stress of Life*, McGraw-Hill Education, 1984.

[266] S. Cohen, R. C. Kessler, and L. U. Gordon, *Measuring stress: A guide for health and social scientists*, Oxford University Press on Demand, 1997.

[267] F. S. Dhabhar, "Immune Function, Stress-Induced Enhancement," in *Encyclopedia of Stress,* Second Edition, ed New York: Academic Press, 2007, pp. 455-461.

[268] A. J. Crum, P. Salovey, and S. Achor, "Rethinking stress: The role of mindsets in determining the stress response," *Journal of personality and social psychology*, vol. 104, p. 716, 2013.

[269] W. B. Cannon, *The wisdom of the body*, New York, W.W. Norton & Company, 1932.

[270] C. Setz et al., "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, pp. 410-417, 2010.

[271] L. Jacobson and R. Sapolsky, "The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis," *Endocrine reviews*, vol. 12, pp. 118-134, 1991.

[272] B. S. McEwen et al., "Mechanisms of stress in the brain," *Nature neuroscience*, vol. 18, p. 1353, 2015.

[273] J. Larsen, and T. Cacioppo, "The psychophysiology of emotion," *Handbook of emotions*, 2nd edition, pp. 173-191, 2000.

[274] D. J. Schutter et al., “Effects of slow rTMS at the right dorsolateral prefrontal cortex on EEG asymmetry and mood,” *Neuroreport,* vol. 12, no. 3, pp. 445-447, 2001.

[275] B. Reuderink, M. Poel, and A. Nijholt, “The impact of loss of control on movement BCIs,” *IEEE Trans. Neural Syst. Rehab. Eng.,* vol. 19, pp. 628–637, 2011. doi:10.1109/ TNSRE.2011.2166562

[276] G. Giannakakis et al., "Review on Psychological Stress Detection Using Biosignals," in *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440-460, 2022, doi: 10.1109/TAFFC.2019.2927337.

[277] C. D. Spielberger, R. L. Gorsuch, and R. E. Lushene, *Manual for the State-Trait Anxiety Inventory*, Palo Alto, CA: Consulting Psychologists Press.

[278] F. R. H. Zijlstra, "Efficiency in Work Behaviour: A Design Approach for Modern Tool," Ph.D. thesis, Delft University of Technology.

[279] A. L. George, "Adaptation to stress in political decision making: the individual, small group and organizational contexts'. In: *Coping and adaptation*, G.V. Coelho, P.A. Hamburg and J.E. Adams (eds.), New York: Basic Books. 1974, pp. 176-245.

[280] J. Hellhammer, and M. Schubert, "The physiological response to trier social stress test relates to subjective measures of stress during but not before or after the test," *Psychoneuroendocrinology*, vol. 37, pp. 119–124, 2012. doi: 10.1016/j.psyneuen.2011. 05.012.

[281] S. S. Dickerson, and M. E. Kemeny, "Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research," *Psychol. Bull.,* vol.130, pp. 355–391, 2004. doi: 10.1037/0033-2909.130.3.355.

[282] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 2000.

[283] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[284] A. Delorme et. al., "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing", *Computational Intelligence and Neuroscience,* vol. 2011, 2011. https://doi.org/10.1155/2011/130714.

[285] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res Brain Res Rev*., vol. 23, no.2-3, pp. 169-95, 1999.

[286] M. X. Cohen, "Bandpass Filtering and the Hilbert Transform," in *Analyzing neural time series data: theory and practice*. Cambridge, Massachusetts: The MIT Press, 2014.

[287] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification,* Wiley-Interscience Publication, 2001.

[288] N. Derakshan, and M. W. Eysenck, "Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory," *European Psychologist*., vol. 14, no. 2, pp. 168–76, 2009.

[289] E. J. Hermans et al., "Dynamic adaptation of large-scale brain networks in response to acute stressors," *Trends in neurosciences*., vol. 37, no. 6, pp. 304–14, 2014.

[290] Kohn N, Hermans EJ and Fernández G. "Cognitive benefit and cost of acute stress is differentially modulated by individual brain state," *Social cognitive and affective neuroscience*., vol. 12, no. 7, pp. 1179–87, 2017.

[291] G. Borghini et al., *Industrial Neuroscience in Aviation: Evaluation of Mental States in Aviation Personnel,* Cham: Springer International Publishing, 2017. https://psycnet.apa.org/doi/10.1007/978-3-319-58598-7

[292] G. Borghini et al., "Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks," In *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2013, pp 6619–6622.

[293] M. E. Smith, and A. Gevins, "Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator," *Proc. SPIE 5797, Biomonitoring for Physiological and Cognitive Performance during Military Operations*, 2005. https://doi.org/10.1117/12.602181.

[294] A. Gevins, and M. E. Smith, "Neurophysiological measures of cognitive workload during human-computer interaction," *Theor Issues Ergon, Sci,* vol. 4, 2003.

[295] A. Gevins et al., "High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice," *Cereb Cortex*, vol. 7, no. 4, pp. 374–385, 1991.

[296] A. Holm A (2010). "Developing neurophysiological metrics for the assessment of mental workload and the functional state of the brain," Ph.D. dissertation, Aalto University School of Science and Technology, Espoo, Finland.

[297] A. Holm et al., "Estimating brain load from the EEG," *The Scientific World Journal,* vol. 9, pp. 639–651, 2009. doi:10.1100/tsw.2009.83.

[298] L. Venables, and S. H. Fairclough, "The influence of performance feedback on goal-setting and mental effort regulation," *Motiv Emot*, vol. 33, pp. 63–74, 2009. doi:10.1007/s11031-008-9116-y.

[299] J. D. Cohen, T. S. Braver, and R. C. O'Reilly, "A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges," *Philos Trans R Soc Lond B Biol Sci,* vol. 351, pp. 1515–1527, 1996. doi:10.1098/rstb.1996.0138.

[300] G. Shou, L. Ding, and D. Dasari, "Probing neural activations from continuous EEG in a real-world task: time-frequency independent component analysis," *J Neurosci Methods,* vol. 209, pp. 22–34, 2012. doi:10. 1016/j.jneumeth.2012.05.022.

[301] I. Goncharova et al., "EMG contamination of EEG: spectral and topographical characteristics," *Clin. Neurophysiol.*, vol. 114, pp. 1580–1593, 2003. doi: 10.1016/S1388-2457(03)00093-2.

[302] J. Mead, M. Middendorf, and C. Gruenwald, "Investigating facial electromyography as an indicator of cognitive workload," Oak Ridge Institute for Science and Education Belcamp United States, 2017.

[303] E. Debie et al., "Multimodal Fusion for Objective Assessment of Cognitive Workload: A Review," in *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1542-1555, 2021, doi: 10.1109/TCYB.2019.2939399.

[304] M. Bagheri and S. Power, "EEG-based detection of mental workload level and stress: the effect of variation in each state on classification of the other," *Journal of Neural Engineering*, vol. 17, 2020.

[305] M. Bagheri and S. D. Power, "EEG-based classification of visual and auditory monitoring tasks," *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* Toronto, ON, 2020, pp. 4032-4037, doi: 10.1109/SMC42975.2020.9283336.

[306] K Mandrick et al. "Neural and psychophysiological correlates of human performance under stress and high mental workload." *Biological psychology,* vol. 121, pp. 62-73, 2016.

[307] A. Lutz et al., “Long-term meditators self-induce high-amplitude gamma synchrony during mental practice,” *PNAS*, vol. 101, no. 46, 2004.

[308] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of multivariate analysis,* vol. 88, no. 2, pp. 365-411, 2004.

[309] A. Delorme et al., MATLAB-Based Tools for BCI Research, in *Brain-Computer Interface. Human Computer Interaction Series.* D. Tan, A. Nijholt (eds) Springer, London, 2010. Doi: 10.1007/978-1-84996-272-8_14.

[310] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences,* 2nd edition, Psychology Press, 1988.

[311] A. Roman-Gonzalez, “EEG Signal Processing for BCI Applications,” In *Human-Computer Systems Interaction: Backgrounds and Applications*; Springer: Berlin/Heidelberg, Germany, 2012; vol. 98, pp. 571–591.

[312] L. M. Andreessen et al., “Toward neuroadaptive support technologies for improving digital reading: A passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading,” *User Model. User-Adapt. Interact.*, vol. 31, pp. 75–104, 2020.

[313] M. Bagheri, and S. D. Power, “Investigating hierarchical and ensemble classification approaches to mitigate the negative effect of varying stress state on EEG-based detection of mental workload level- and vice versa,” *Brain-Computer. Interfaces*, vol. 8, pp. 26–37, 2021.

[314] H. Zeng et al., “InstanceEasyTL: An Improved Transfer-Learning Method for EEG-Based Cross-Subject Fatigue Detection,” *Sensors*, vol. 20, no. 24, pp. 7251, 2020.

[315] H. Zeng et al., “A LightGBM-Based EEG Analysis Method for Driver Mental States Classification,” *Comput. Intell. Neurosci.,* vol. 2019, 2019.

[316] Z. Deng et al., “Transductive Joint-Knowledge-Transfer TSK FS for Recognition of Epileptic EEG Signals,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, pp. 1481–1494, 2018.

[317] J. Wang et al., “Easy Transfer Learning by Exploiting Intra-Domain Structures,” In *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME),* Shanghai, China, 8–12 July 2019; pp. 8–12.

[318] W. Dai et al., “Boosting for transfer learning,” In *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, USA, 20–24 June 2007; ACM: New York, NY, USA, 2007; pp. 193–200.

[319] S. J. Pan et al., “Domain Adaptation via Transfer Component Analysis,” *IEEE Trans. Neural Netw.*, vol. 22, pp. 199–210, 2010.

[320] B. Gong et al., “Geodesic flow kernel for unsupervised domain adaptation,” In *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, Providence, RI, USA, 16–21 June 2012.

[321] Y. Ganin et al., “Domain-Adversarial Training of Neural Networks,” *J. Mach. Learn. Res.,* vol. 17, pp. 1–35, 2016.

[322] B. Frey, *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc., 2018. doi:10.4135/9781506326139

[323] D. H. Lee et al., “Continuous EEG Decoding of Pilots’ Mental States Using Multiple Feature Block-Based Convolutional Neural Network,” *IEEE Access*, vol. 8, pp. 121929–121941, 2020.

[324] X. Chai et al., “A Fast, Efficient Domain Adaptation Technique for Cross-Domain Electroencephalography(EEG)-Based Emotion Recognition,” *Sensors*, vol. 17, no. 5, p. 1014, 2017.

Appendix 1: Ethics approval

From:	dgulliver@mun.ca
To:	Bagheri Mahsa(Principal Investigator)
Cc:	Power, Sarah Dianne; Office of Research Services; Gulliver, Deborah
Subject:	ICEHR Clearance # 20190461-EN – EXTENDED
Date:	September 7, 2021 9:00:56 AM

Interdisciplinary Committee on
Ethics in Human Research (ICEHR)

ICEHR Approval #:	20190461-EN
Researcher Portal File #:	20190461
Project Title:	*Toward the development of an emotion-aware passive brain-computer interface for mental workload monitoring- a pilot study*
Associated Funding:	20161820
Supervisor:	Dr. Sarah Power
Clearance expiry date:	**August 31, 2022**

Dear Miss Mahsa Bagheri:

Thank you for your response to our request for an annual update advising that your project will continue without any changes that would affect ethical relations with human participants.

On behalf of the Chair of ICEHR, I wish to advise that the ethics clearance for this project has been extended to **August 31, 2022**. The *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (TCPS2) requires that you submit another annual update to ICEHR on your project prior to this date.

We wish you well with the continuation of your research.

Sincerely,

DEBBY GULLIVER
Interdisciplinary Committee on Ethics in Human Research (ICEHR)
Memorial University of Newfoundland
St. John's, NL | A1C 5S7
Bruneau Centre for Research and Innovation | Room IIC 2010C
T: (709) 864-2561 |