UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE ENGENHARIA DE COMPUTAÇÃO

GUSTAVO VERGANI DAMBROS

Predicting Response Quality as a Proxy of Fatigue via Eye Tracking and EEG

Work presented in partial fulfillment
of the requirements for the degree of
Bachelor in Computer Engeneering

Advisor: Prof. Dr. Bruno Castro da Silva
Coadvisor: Dr. Martin Spüler

Porto Alegre
December 2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. Rui Vicente Oppermann
Vice-Reitora: Prof^{a}. JaneJane Fraga Tutikian
Pró-Reitor de Graduação: Prof. Vladimir Pinheiro do Nascimento
Diretora do Instituto de Informática: Prof^{a}. Carla Maria Dal Sasso Freitas
Coordenador do Curso de Engenharia de Computação: Prof. Renato Ventura Bayan Henriques
Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

"A ship is always safe at the shore,

but that is not what it's build for."

— Albert Einstein

ACKNOWLEDGEMENTS

First of all, I would like to thank my parents, Maria do Carmo Vergani Dambros and Sergio José Dambros, for the support they have always given me and for more than 23 years of teaching, without which I could not even imagine writing this thesis.

To my brother, Jonathan William Vergani Dambros, who induced me to choose Computer Engineering and whom I have as an example to follow.

To my grandparents, Ary Vergani (in memoriam), Inelia Terezinha Polidoro Vergani (in memoriam), Aurora Bonatto Dambros and Casemiro Dambros, and my godfather Fabiano Vergani (in memoriam) and godmother Adriana Vergani for the support they gave me whenever I needed it.

I also thank my friend Gabriel Bonezi for his fellowship, for the countless conversations we had and for correcting me whenever I was wrong.

To Susan Lusca da Silva for the affection, patience and essential support during the development of this work.

I only have to thank the countless teachers who, in my 18 years as a student, have somehow collaborated in my academic and personal training.

Finally, I thank my advisors Dr. Bruno Castro da Silva and Dr. Martin Spüler who supported me in the development of this thesis and to Judith Ungewiß for sharing her knowledge of both campimetric setup and campimetric experiment in behalf of this work.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos meus pais, Maria do Carmo Vergani Dambros e Sergio José Dambros, pelo apoio que ambos sempre me deram e por mais de 23 anos de ensinamentos sem os quais não poderia nem imaginar em escrever esta tese.

Ao meu irmão, Jônathan William Vergani Dambros, que me induziu a escolher o curso de Engenharia de Computação e a quem tenho como exemplo a ser seguido.

A meus avós, Ary Vergani (in memoriam), Inélia Terezinha Polidoro Vergani (in memoriam), Aurora Bonatto Dambros e Casemiro Dambros, e ao meu padrinho Fabiano Vergani (in memoriam) e madrinha Adriana Vergani pelo suporte que me deram sempre que precisei.

Agradeço também ao amigo Gabriel Bonezi pelo companheirismo, pelas inúmeras conversas que tivemos e por corrigir-me sempre que estive errado.

À Susan Lusca da Silva pelo carinho, pela paciência e pelo apoio essencial durante o desenvolvimento desse trabalho.

Só tenho a agradecer aos incontáveis professores que, nos meus 18 anos como estudante, de alguma forma colaboraram na minha formação acadêmica e pessoal.

Por fim, agradeço aos meus orientadores Dr. Bruno Castro da Silva e Dr. Martin Spüler que me apoiaram no desenvolvimento desta tese e a Judith Ungewiß por compartilhar seu conhecimento sobre campimetria em prol desse trabalho.

LIST OF ABBREVIATIONS AND ACRONYMS

CC	Correlation Coefficient
ECG	Electrocardiogram
EEG	Electroencephalography
EOG	Electrooculogram
ET	Eye Tracking
GP	Gaussian Process
HCI	Human-Computer Interaction
LASSO	Least Absolute Shrinkage and Selection Operator
OLS	Ordinary Least Squares
PERCLOS	Percent Eye Closure
PSD	Power Spectrum Density
RMSE	Root-Mean-Square Error
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression Machine

LIST OF FIGURES

LIST OF TABLES

CONTENTS

ABSTRACT

Many computer systems are capable of adapting their behavior depending on the degree of fatigue of their users, so it is of interest to estimate how tired users are. As there is no method to measure user fatigue directly, we investigate whether the use of data collected via eye tracking systems (ET) and electroencephalography (EEG) can predict user error rate during a simple visual experiment. Using the rate with which a user errs while solving a given task as a proxy of fatigue is possible because evidence exists that both values are correlated. We propose comparing different machine learning methods applied to the problem of predicting user error rate given ET and EEG measurements as inputs. The data used in this work was collected during a 40-minute campimetric task, where users were tasked with detecting visual stimuli (i.e., points) of different contrasts.

Keywords: Fatigue. electroencephalogram. eye tracking. prediction.

RESUMO

Vários sistemas computacionais apresentam a possibilidade de se adaptarem dependendo do grau de fatiga de seus usuários; por essa razão, é de interesse estimar o quão cansado os mesmos estão. Como não há um método para mensurar a fadiga diretamente, investigamos se o uso de dados coletados via rastreamento ocular e eletroencefalografia (EEG) podem predizer a taxa de erro durante um experimento visual simples. Usar a taxa de erros que um usuário comete ao resolver uma determinada tarefa como proxy da fadiga é possível porque existe evidência de que ambos os valores estão correlacionados. Compararemos diferentes métodos de aprendizado de máquina com o intuito de encontrar melhores predições, com base nos dados de ET e EEG. Os dados foram coletados durante uma tarefa campimétrica de 40 minutos, onde o usuário deve detectar estímulos visuais (ou seja, pontos) de diferentes contrastes.

Palavras-chave: Fatiga. eletroencefalograma. rastreamento ocular. predição.

1 INTRODUCTION

Vigilance is derived from the Latin word *vigilantia* and can be translated as wakefulness. As a technical term, vigilance describes primarily the degree of central nervous activation (HEAD, 1923) and for that, is used to describe a physiological state of alertness (CANISIUS; PENZEL, 2007). In contrast, reduced vigilance can be called *fatigue* (WEESS et al., 1998).

Knowledge about a user's fatigue is essential for properly conducting perceptual and behavioral studies. For example, a medical examination of the visual field can be exhausting for patients since it requires patients to repeatedly detect and respond to hardly-visible stimuli. The possible gains in performance and reliability resulting from increased test duration are antagonized by increasing fatigue, effectively limiting the achievable examination quality. The rate of response errors (e.g., missed detections that were above a given perceptual threshold) increases with the duration of the measurement, experiment, or examination to which a user is submitted. Although the examination time for visual field testing has been reduced over the last years (BENGTSSON et al., 1997), current examination techniques still lack a retest reliability (BENGTSSON; HEIJL, 2008). The difficulty of performing tasks reliably due to fatigue affects numerous situations: not only medical examinations but also the interaction of users with software systems, for example.

Many techniques have already been used to try to predict user's fatigue. For example, the most commonly-used task used to validate these methods involves making subjects drive for several hours until they become tired. A wide range of machine learning algorithms has been used in order to predict fatigue in this and other tasks. Eye tracking (ET), electroencephalogram (EEG) and electrocardiogram (ECG) are some examples of the physiological monitoring methods that are used to collect training features that serve as input to the machine learning algorithms that, based on them, predict the level of fatigue of a given user. These methods will be described in more details in a later section.

The main limitation that can be observed in the existing works for predicting fatigue with machine learning algorithms is that they use only one physiological monitoring methods. Users that use reading glasses, for instance, could be a serious problem to eye tracker systems, since glasses can cause glare and may be totally opaque to light, making it difficult for a camera to monitor eye movement. Besides that, in EEG, the very small signal-to-noise ratio (i.e. interference when capturing brain signals that cause loss of information) is an obstacle. This problem is common when the user moves their head abruptly, masticate or even due to cardiac signals interacting with EEG measurements. Noise can also be found in the ECG recording due to muscle contraction. The use of two different sources of physiological indicators can increase a machine learning-based predictor's robustness since poor predictions often occur when *individual* indicators or features cannot be acquired accurately or reliably. In this work, we evaluate the performance with which a machine learning algorithm can predict fatigue based on a *family* of physiological measurements, used in conjunction to try to mitigate limitations of any one individual measurement. We compare four different machine learning algorithms, including linear and non-linear models, and models that try to actively select which input features (physiological measurements) help the most in predicting fatigue based on the existing data.

In this work, we evaluate the proposed systems on *campimetry experiments*. Campimetry is the examination of the visual field on a flat surface (e.g., a computer screen), where lights are presented at different locations in the visual field and with different contrasts with respect to the background surface. If a stimulus is perceived, the subject is expected to provide feedback by pressing a button (TAFAJ et al., 2011). Through this procedure, the minimum perceivable stimulus-to-background contrast is determined at each location. In this task, fatigue affects the user's capability of rapidly indicating whether a given stimulus has been perceived, and often results in increased error rate — the user incorrectly indicates that a stimulus occurred when it did not, for instance.

In this work, we do not focus on predicting user fatigue for improving behavioral studies but in order to obtain accurate estimates of vigilance that can be used to improve adaptable computer systems. As will be described in more details later on, we choose to estimate fatigue by using eye tracking and electroencephalogram data collected while a user performs a campimetry task, as this type of task provides objective performance indicator via the response error rate. However, since reliable indicators (i.e. useful input physiological measurements that can be used as input to the machine learning algorithms) for this particular task are found, we expect it to be applicable to a broad range of similar tasks. We also evaluate if the two above-mentioned types of physiological features, when used jointly to train our fatigue-predicting systems, result in improved results with respect to a system that analyzes them independently.

We show that a subject's fatigue can be predicted by machine learning algorithms using eye tracking and EEG parameters data. As previously mentioned, the data used in this work was collected during a task in which the user had to respond to visual stimuli (i.e., points) of different contrasts and the rate with which a user errs is used as proxy of fatigue. Errors were defined as stimuli that were not responded to by the user or situations when a user responded to stimuli that did not occur. In this work we employed two different methodologies for training and evaluating different machine learning methods: 5-fold cross-validation and leave-one-out cross-validation. The precise meaning of these methodologies will be described in section 5.3. When using 5-fold cross-validation to train and evaluate different machine learning methods, we obtained a correlation up to 0.96 between the error rate and its prediction using Gaussian process with only EEG features. In the leave-one-out cross-validation, a correlation of 0.80 ± 0.07 using support vector regression (SVR) with both ET and EEG features was attained.

1.1 Outline

This thesis is structured as follows. In Section 2, we review the theoretical background needed to understand the technical sections that follow, providing an overview of physiological monitoring methods, vigilance indicators, and machine learning algorithms. In Section 3, we give a brief overview of the machine learning algorithms that we used in this thesis. In Section 4, we review related work on fatigue prediction via methods that use ET, EEG, or ECG features as inputs to various machine learning algorithms. In Section 5, we describe the experimental methodology used in this work (including a discussion on how data was collected and pre-processed and which features were constructed), as well as the different machine learning algorithms that we evaluate in the fatigue-prediction task and the metrics according to which they are compared. Section 6 presents the experimental results obtained by applying the selected algorithms when deployed in two ways: in order to construct a fatigue model for each user and to construct a more general model trained based on data from many users. Finally, in Section 7, we conclude this work by discussing the most important results obtained and providing a brief discussion over future work.

2 THEORETICAL BACKGROUND ON FATIGUE-RELATED CONCEPTS

In this chapter, we discuss important concepts related to fatigue—in particular, physiological monitoring methods and indicators of vigilance, such as eye tracking and Electroencephalography (EEG), which provide measurements about a user's state that are known to be correlated with fatigue. Information collected by these monitoring methods will be used in later chapters as training data for different machine learning algorithms.

2.1 Physiological Monitoring Methods

Physiologic monitoring methods are processes for monitoring vital physiological parameters of a person. In this section, we present two of these methods which are important when modeling user fatigue: Eye Tracking (ET) and Electroencephalography (EEG).

2.1.1 Eye Tracking

In simplest terms, eye tracking refers to a physiological monitoring process for measuring eye activity. By using an eye tracker system, it is possible to collect various types of information about a user's eyes; for instance, where a given subject is looking at (fixation), when they blink (blink rate and blink duration), how their eyes are moving from one gaze to the next (saccade), and what is the reaction of their pupils to different types of stimuli (pupil size and variability).

Eye tracking is not a new field of science. (JAVAL, 1878) defined the term *saccade* back in 1878 when he reported that humans eyes do not move continuously along a text line during a reading task, but instead, make short and rapid movements interleaved with brief stops. Over the years, many eye tracker systems and algorithms were developed, making it possible for more precise measurements. Some important concepts related to ET, which will be used throughout this work, are summarized below:

- *Perimetry* is the measurement of a person's field of vision;
- *Pupillometry* is the measurement of pupil size and reactivity to different types of stimuli;
- *Scotopic vision* refers to vision under low light conditions. It is produced exclusively through rod cells which are most sensitive to wavelengths of light around

498 nm (green-blue) and are not sensitive to wavelengths longer than 650 nm (red);

- *Photopic vision* refers to vision under well-lit conditions. It allows for color perception, mediated by cone cells, and a significantly higher visual acuteness and temporal resolution than available in scotopic vision;
- *Mesopic vision* is a combination of photopic vision and scotopic vision—vision in low but not quite dark lighting situations;
- *Heterophoria* is the failure of the visual axis (i.e. the line extending from the object seen to the center of the pupil) to remain parallel when not performing binocular fusion.

2.1.2 Electroencephalography (EEG)

Electroencephalography (EEG) refers to a physiological monitoring process for monitoring electrical activity of the brain (electrophysiological monitoring). It is typically noninvasive, performed via electrodes placed along the scalp with the support of a cap. EEG measures voltage fluctuations in the brain resulting from ionic current within the neurons. Researchers generally focus on the spectral content of EEG—i.e., analysis of neuron activity in the frequency domain. This allows for EEG signals that measure neural oscillations called brain waves.

Brain waves are observed throughout the central nervous system; in general, these oscillations can be characterized by their frequency, amplitude and phase. Such signal properties can be extracted from neural recordings using time-frequency analysis such as Burg's Method (BURG, 1975).

EEG monitoring is also not a new research field. German physiologist and psychiatrist Hans Berger was able to record a human EEG for the first time in 1924, describing the process in details five years later (BERGER, 1929). Nowadays, EEG amplifiers, a hardware that amplifies EEG signals, can reach very high frequencies and handle many channels, each channel receiving data from an electrode placed in the users scalp.

2.1.2.1 Electrooculography (EOG)

When measuring EEG signals, the occurrence of blinks and eye movements introduces artifacts (perturbations) to the recorded data. In this work we use EOG measurements to remove such artifacts from the measured EEG channels, as described in

(SCHLÖGL et al., 2007).

Electrooculography (EOG) is the record of the standing voltage between the front and back of the eye. It is known to be correlated with eyeball movement and can be obtained by electrodes suitably placed on the skin near the eye. In this work, it will be used to remove EOG artifacts in the EEG signals.

2.2 Indicators of Vigilance as Features for Machine Learning Algorithms

Based on the physiological monitoring methods described in the previous section, we now discuss the specific ones that will be used as *indicators of vigilance* and used as input features to fatigue-prediction machine learning algorithms. Here, we define indicators of vigilance as any physiological data about a user that may provide information about their level of fatigue while solving a given task.

The choice of input data, or indicators of vigilance, that will be given to a fatigue-prediction machine learning algorithm, depends on the particular task that the fatigued user will need to solve. In this work, we focus on **campimetry tasks**, or capimetry experiments. Campimetry is the examination of the visual field on a flat surface (e.g., computer screen), where light is presented at different locations of the visual field. If a stimulus is perceived, the subject provides feedback by pressing a button (TAFAJ et al., 2011; TAFAJ et al., 2010). Through this procedure, the minimum perceivable stimulus-to-background contrast is determined at each location. Increasing the number of stimuli presented allows either for a finer resolution of the location grid, a higher number of different contrast levels or more retests. Fatigue impacts the user's capability of doing the task so accurately; this is reflected in the user's error rate in that task—they incorrectly indicate that a stimulus occurred when it did not, for instance.

2.2.1 Eye tracking Data for Predicting Fatigue

Parameters of vigilance in eye tracking data are relatively well-studied. When analyzing user fatigue during campimetry and perimetry tasks, researchers have typically focused on fatigue waves; e.g., *pupillary oscillations* (HENSON; EMUH, 2010), although a wide variety of other vigilance parameters can theoretically be measured through the same device—a camera directed at the subject's eyes.

The term pupillary oscillations is often found in the context of vigilance and fatigue. The sleepier a person is (and the less they try to suppress sleep), the shorter the time of initial pupil constriction is, and the larger (and more frequent) pupillary oscillations are. These oscillation waves originate in the central nervous system (LOWENSTEIN; FEINBERG; LOEWENFELD, 1963). Such waves are known as pupillary **fatigue waves**. Henson and Emuh (HENSON; EMUH, 2010) have shown that oscillation waves can be used both to detect fatigue under scotopic, also under mesopic and even low photopic conditions, since they occur within visual field testing by campimetry.

Henson and Emuh showed that it is possible to monitor user vigilance during campimetry under photopic conditions using pupillometry methods (HENSON; EMUH, 2010). In darkness, pupil size oscillations come in two flavors: slow waves of dilatation and constriction of 4 to 40 seconds duration and amplitude of up to 0.5 mm; and superposed fast oscillations, of 0.5 to 1 second duration and with amplitudes below 0.1 mm—but being able to reach up to 0.3mm (LOWENSTEIN; FEINBERG; LOEWENFELD, 1963). Importantly, it is known that with decreasing vigilance the feedback loop that regulates the pupil diameter becomes unstable, resulting in much larger pupillary oscillations (WILHELM et al., 2001).

Another parameter is the vergence angle of both eyes. It is connected to vigilance in a not necessarily straight-forward way. Heterophoria has been found to grow with fatigue and when performing an unfamiliar task (KAUFMANN; STEFFEN, 2012). Some authors report that the vergence system is particularly fragile with fatigue (VERNET; KAPOULA, 2009).

Pupil size and variability are also commonly used as an indicator of fatigue. Both are determined by two neurophysiological reflexes, the pupillary light reflex regulating the amount of light entering the pupil, and accommodation, a change in the curvature of the lens.

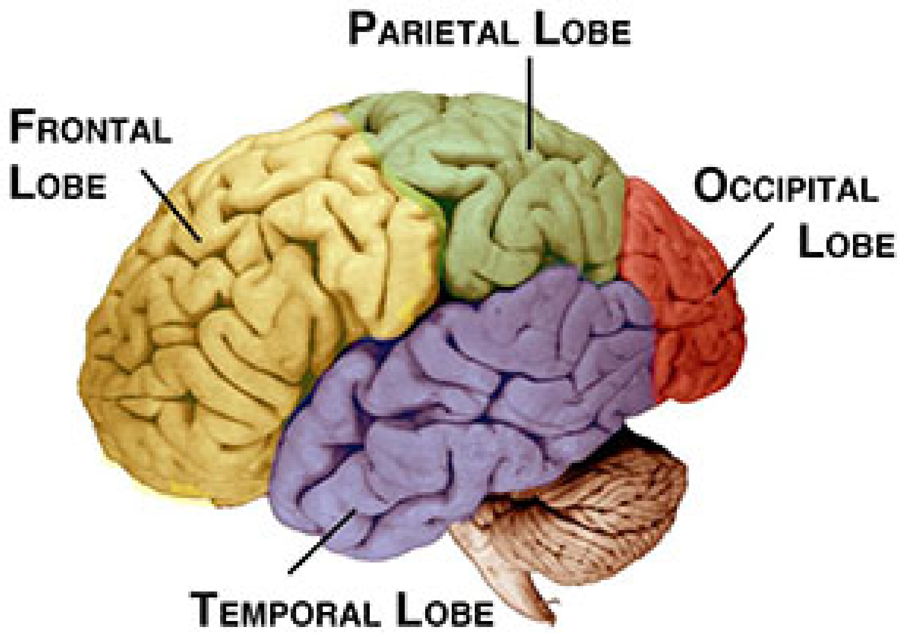
Blink rate, duration, and lid closure speed reflect the level of fatigue and sleepiness. Light fatigue is associated with an increase in blink frequency, sleepiness with an increase in blink duration (SCHLEICHER et al., 2008). Due to these, blink rate and blink duration can be used as parameters of vigilance.

2.2.2 EEG Data for Predicting Fatigue

Many researchers have been investigating methods to predict fatigue using EEG signals in the last years. One of the most used tools for doing so involves measurements of the power spectrum density (PSD) of brain waves. This technique transforms the EEG signal from time to frequency domain, making it possible to analyze how its strength is distributed as a function of frequency. Many methods for converting signals from the time domain to the frequency domain were developed over the years, such as the Fast-Fourier-transform, Welch's method, and Burg's Method. In neuroscience, the PSD curve is usually analyzed in four different frequency bands, ranging from 0.5 Hz to 4 Hz, 4 Hz to 8 Hz, 8 Hz to 13 Hz, and higher than 13 Hz (TEPLAN, 2002). These intervals are the same for delta, theta, alpha and beta waves when the signal is in the time domain, respectively.

(ÅKERSTEDT; KECKLUND; KNUTSSON, 1991) discovered a relation between alpha waves and sleepiness during the execution of night shift tasks. Furthermore, (TREJO et al., 2015) showed that the mental fatigue is associated with increased power in frontal theta and parietal alpha EEG rhythms during an exhausting arithmetic task (Figure 2.1). The results obtained by these authors also indicate that EEG spectra did not correlate with fatigue outside the theta and alpha intervals (GHARAGOZLOU et al., 2015) and (KONG et al., 2015) demonstrated that the same results were obtained when the subjects are evaluated during a tiresome driving task.

Figure 2.1: Brain regions (SARINANA, 2010)



3 OVERVIEW OF MACHINE LEARNING ALGORITHMS

In this section, we present short descriptions of the main machine learning (ML) algorithms that will be used in the remainder of this work as methods to predict user fatigue based on physiological data. We selected four ML to analyze fatigue-related training data: Suppor Vector Regression, LASSO, Gaussian Processes Regression, and Neural Networks. These algorithms were selected due to their wide applicability in many real-life machine learning tasks; below we present the main ideas underlying these algorithms.

3.1 Support Vector Regression (SVR)

Support Vector Machines are supervised learning models that analyze data in the context of classification problems. Given a set with n training examples in the form:

$$(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$$

where $\vec{x}_i$ is a vector of features describing the i-th training example, and where each y_i indicates whether that example belongs to one of two classes, an SVM training algorithm finds the maximum-margin hyperplane that divides the two classes. This hyperplane can subsequently be used to assign novel examples to one of these classes.

A variant of SVM was was proposed in 1995 by Harris Drucker et al. so that it could be used in regression problems (DRUCKER et al., 1997); i.e., when the value y being predicted by the algorithm is not a discrete class label, but a real value. This method is known as support vector regression (SVR). Training an SVR model requires solving the following minimization problem:

$$\text{minimize} \quad \frac{1}{2}\|\vec{w}\|^2$$

$$\text{subject to} \begin{cases} y_i - \langle \vec{w}, \vec{x}_i \rangle - b \leq \varepsilon \\ \langle \vec{w}, \vec{x}_i \rangle + b - y_i \leq \varepsilon \end{cases} \tag{3.1}$$

where $\vec{x}_i$ is a training sample, y_i is the respective real-valued target associated with that example, $\vec{w}_i$ is a weight vector, b is known as the model bias, ε is a free parameter that serves as a *threshold* (discussed in what follows), and where $\langle \vec{w}, \vec{x}_i \rangle$ is the inner product between $\vec{w}$ and $\vec{x}_i$.

When solving the above-mentioned optimization goal, all predictions have to be within an ε range of the true value. If the minimization problem is infeasible (no weights w achieve predictions within this accuracy), slack variables are usually added into the minimization problem (3.1) to allow for approximations. When presented with a novel input $\vec{x}_t$ whose corresponding output y_t needs to be predicted, the model computes this output by taking the inner product between the weight vector and the input, plus the bias variable, b:

$$y_t = \langle \vec{w}, \vec{x}_t \rangle + b \tag{3.2}$$

3.2 Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO is a machine learning regression algorithm that performs both *variable selection* and *regularization* to enhance prediction performance. Variable selection means that the algorithm will automatically determine which input features in the vector x are relevant to predicting outputs y; removing irrelevant features is known to make the prediction model more accurate. Regularization, on the other hand, refers to a technique by which a penalty is given for solutions to the optimization problem (i.e., weights w) that are large; this type of penalty implies that whenever possible, few weights will be non-zero, thus making the model mapping inputs x to outputs y use as few components of the vector x as possible, when computing a predicted output.

The LASSO algorithm was introduced in 1996 by Robert Tibshirani (TIBSHIRANI, 1996), and consists in minimizing the residual sum of squares of the difference between the real value and its prediction subject to the constraint that the sum of the absolute value of the weights w being used by the model is less than a parameter $\boldsymbol{t}$. This parameter is known as the regularization parameter, and its use causes some of the regressions coefficients w to shrink towards zero, thereby effectively selecting which features (components of the input x) will be used when making predictions. The algorithm assumes, as SVR does, that there are N training examples in the form:

$$(\vec{x}_1, y_1), \ldots, (\vec{x}_N, y_N)$$

where each $\vec{x}_i$ has M dimension. Then, the algorithm identifies the weights w that are a solution to the following optimization problem:

$$(b,\vec{w}) = \operatorname{argmin} \sum_{i=1}^{N} (y_i - b - \sum_{j=1}^{M} w_j x_{ij})^2$$

$$\text{subject to } \sum_{j=1}^{p} |w_j| \leq t$$

where $\vec{x}_i$ is a training sample, y_i is the respective target value, $\vec{w}_i$ is the weight vector, b is the bias and t is a free parameter that stipulates the amount of regularization. As with SVR, Equation (3.2) is used to calculate the predicted value of y_t for a new input $\vec{x}_t$.

3.3 Gaussian Processes Regression

A Gaussian Process (GP) is a statistical technique used to model observations x that occur within some continuous input domain. It uses a measure of the similarity (known as a *kernel*) between input points when predicting the output value of a novel input. The prediction made by a GP is not just an estimate for that input's output, but also the uncertainty of this prediction.

Given a set of n training examples in the form:

$$(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$$

a GP assumes that the set of y values can be modeled as a sample from a multivariate Gaussian distribution:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} \sim G(\vec{0}, \mathbf{K})$$

where $G(\vec{0}, \mathbf{K})$ is a multivariate Gaussian distribution with zero mean and $\mathbf{K}$ is a covariance matrix. Each entry K_{ij} is computed by a kernel function, which measures similarities between pairs of input points x_i and x_j. Below is an example of kernel function, known as the Squared Exponential kernel:

$$K_{ij} = e^{\frac{-(x_i - x_j)^2}{2}}$$

To calculate the prediction y_t for a new input $\vec{x}_t$, a GP model first needs to compute K_t:

$$K_t = \begin{bmatrix} K(t,1) \\ K(t,2) \\ \dots \\ K(t,n) \end{bmatrix}$$

According to (MURPHY, 2012), we can compute the expected value for y_t (the output predicted for the novel input x_t) as follows:

$$E[y_t] = K_t^T \mathbf{K}^{-1} \vec{y}$$

and the variance of this expected value as:

$$\sigma = -K_t^T \mathbf{K}^{-1} K_t + K_{tt}$$

3.4 Artificial Neural Networks

In machine learning, an artificial neural network (ANN or NN) is a learning algorithm inspired by the structure and functional aspects of biological neural networks. Computations performed by a highly interconnected group of artificial neurons, each one performing a relatively simple transformation of its inputs. ANNs are usually used to model complex relationships between inputs and outputs and to find patterns in training data. The input goes from the input to the output layer, passing through one or more intermediate layers. Each feature corresponds to one neuron in the input layer, while the output given by the output layer corresponds to the prediction. Each neuron of the hidden and output layers has an activation function that takes a weighted sum of the inputs given by the neurons of the previous layer to discover if they must be propagated to the next layer of neurons. This weighted sum is known as propagation function and is defined as follows:

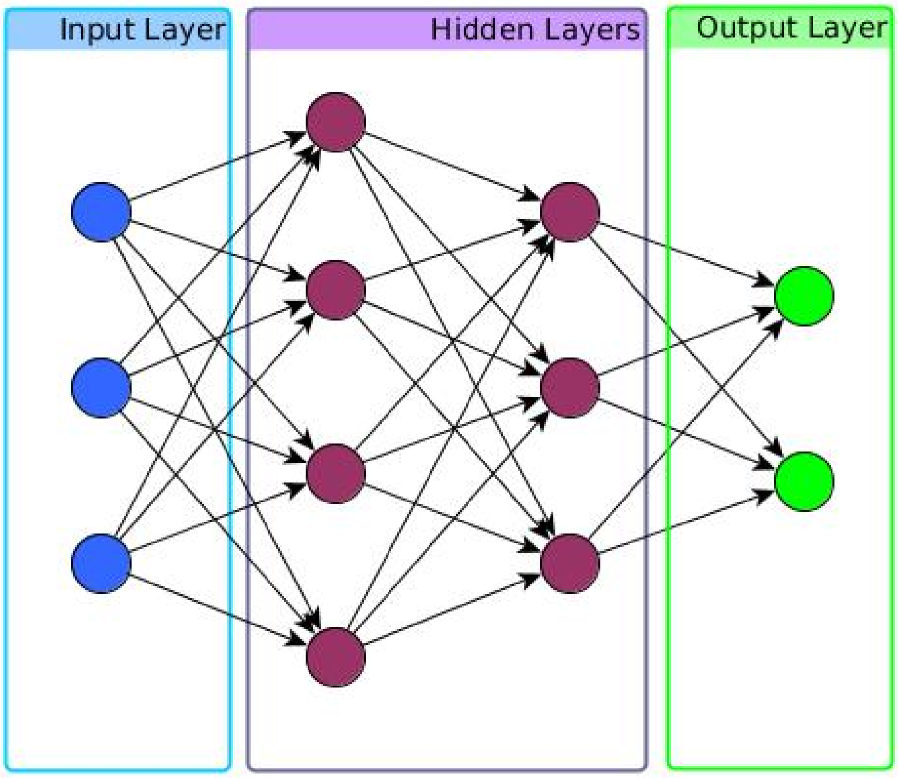
$$p_j(t) = \sum_i o_{ij}(t) * w_{ij}(t)$$

where p_j is the propagated value of the neuron j to the neurons of the next layer at a given time t, o_{ij} is the signal received by the neuron from the previous layer and w_{ij} is the connection weight.

There are many different types onction. The simplest is the step function (3.3). A step function results in 1 if the input is higher than a certain threshold θ, otherwise it's output will be 0.

$$f(x) = \begin{cases} 1, x > \theta \\ 0, x \leq \theta \end{cases} \tag{3.3}$$

Figure 3.1: Representation of an artificial neural network with 12 neurons. Circles represent neurons and arrows represent connections between neurons. Four layers are present in this example; the network has one input layer, two hidden layers, and one output layer.



Neural networks need to go through a learning process to find the optimal connection weights. For this, it is important to define cost function $C : F \rightarrow \mathbb{R}$, where F is a set of functions. The optimal solution is $f' \in F$ such that $C(f') \leq C(f)\ \forall\ f \in F$. The cost function is an important concept in neural networks, as it is a measure of how far away a particular solution is from an optimal solution to the problem to be solved.

The most famous learning algorithm is the backpropagation. The first steps of this algorithm is choosing the connection weights $w_{ij}(t_o)$, doing the prediction for an training sample and calculating the cost function for this setup. Then, the weight updates for the following training samples of backpropagation can be done via gradient descent as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \eta * \frac{\partial C}{\partial w_{ij}}$$

where η is the learning rate.

4 RELATED WORK

Predicting user fatigue using physiological data, such as ET and EEG measurements, is a current and active research topic. In this chapter we review related works that aim at predicting the error rate affecting users performing a given task. We discuss the physiological data/features used by these techniques and their advantages, disadvantages, and limitations.

4.1 Fatigue Prediction Using ET Features

In a previous work (DAMBROS et al., 2017), we used only ET features gathered during the same experiments to predict user's fatigue using an SVR with linear kernel. An average correlation of 0.72 ± 0.17 between the error rate and the predicted value was obtained when using leave-one-subject-out cross-validation. In this work, we improved these results by adding EEG features to complement the ET features and increased the number of machine learning algorithms, including non-linear models. Further, the 5-fold cross-validation approach was added.

Many techniques try to predict user fatigue via eye tracking features. Due to the number of accidents caused by driver's fatigue, for instance, detection of a driver's fatigue is the most common problem that many of these techniques aim at solving (BERGASA; NUEVO; SOTELO, 2006; JI; ZHU; YU-PIN, 2004) . A large range of input physiological features related to eye tracking has been used when constructing methods to predict driver fatigue. Among them, we highlight the use of blink rate, blink duration, pupil size, pupil width, percent eye closure (PERCLOS), and saccade velocity. Also, many machine learning algorithms were tested in order to learn mappings from these features to user error rates (used, again, as proxy measurements of user fatigue). (JO et al., 2011) used a support vector machine model to this end, while (BERGASA; NUEVO; SOTELO, 2006) used a fuzzy classifier to achieve a real-time system that can predict if the driver is either tired or awake.

Even though driver fatigue detection is the most famous application of ET-based fatigue detection, eye tracking can be used to predict fatigue or alertness in other tasks. Studies that relate eye tracking, alertness, and human–computer interaction (HCI) can be easily found in the literature. Estimation of a user's interest in a learning task is essential to reduce the learning time. (ASTERIADIS et al., 2009) uses eye tracking features to train a neuro-fuzzy system to improve the user's experience in an e-learning environment. The user does not need to interact with the proposed system, using the website as if he/she is not been monitored. (BRUNEAU; SASSE; MCCARTHY, 2002) uses eye movement, pupil size and blink information to detect fatigue using an eye-tracker to get an insight and enhance the user's usability in websites.

Although these related works also use eye tracking features to predict user's fatigue/interest during their respective task, all lack a second source of features that could improve the robustness since glasses and the lighting can oppose to the acquisition of indicators accurately. Another important difference is the amount of features used. (JO et al., 2011), despite using the same machine learning algorithm, uses only PERCLOS and PERLOOK (i.e. the percentage of time not looking forward) to predict fatigue while we use 6 ET-related indicators. Furthermore, PERLOOK cannot fit to our experiment since the user must look at a central fixation cross during all the experiment. We believe that the features that we used and the combination of two sources of inficators can also be used to predict driver's fatigue.

4.2 Fatigue Prediction Using EEG Features

Detection of a driver's fatigue is also a common problem when designing ML algorithms to predict fatigue based on EEG measurements/features. (YEO et al., 2009) proposed a method to predict fatigue using alpha and beta waves as features to train an SVM. (LIN et al., 2005) used EEG and EOG combined with independent component analysis, power-spectrum analysis, and a linear regression model to estimate a driver's drowsiness level. The authors tested their approach with ten EEG/EOG channels and also using only the two best drowsiness-related EEG channels. Both tests were able to obtain driver's fatigue with accuracy higher than 80%. (LIN et al., 2005) used error as a proxy of fatigue, being error defined as the deviation the center of the vehicle and the center of the cruising lane. Differently from this work, they used EOG as source of fatigue indicators (i.e. blink rate, blink duration, saccade velocity and saccade frequency). The eye tracker can obtain these and other indicators that EOG cannot, such as pupil size and variability, but glasses and lighting does not interfere with obtaining data via EOG.

Another works aim at estimating user fatigue via EEG signals, but in different tasks—for instance, one where the goal is to predict quality response based on EEG signals during an auditory vigilance task. In other words, the participants had to give a determined feedback to four different sound stimuli. The percentage of correct feedback was used to predict the level of user's fatigue. In this task, participants underwent a 25 hours sleep deprivation experiment. (SHEN et al., 2008) used EEG features to training both probabilistic-based multi-class SVM and a standard multi-class SVM, demonstrating that it is feasible to monitor mental fatigue, obtaining an accuracy higher than 85% with both machine learning algorithms. In comparison, the participants of our experiments were not required to undergo a 25 hours of sleep deprivation.

(TREJO et al., 2015) used a linear regression classifier to find that theta and alpha waves that are most highly correlated with a user's fatigue, achieving an accuracy higher than 90%. In this work, fatigue had to do with the capability of participants to perform a math task in which they had to repeatedly say whether different equations were true or false. Users underwent this task until they either quit from exhaustion or three hours had elapsed. The response time and accuracy were used to set the drowsiness level of each 15 minutes interval.

Although these works use error rate as proxy of fatigue, none of them combine eye tracking and EEG. ET features could make the model more robust since both EEG and EOG suffer from interference caused by muscle contraction. The above-mentioned works also use the data from the same participant in the training and testing set, while we have the leave-one-subject-out cross-validation approach that does not require that. Another point that is important to highlight is the lack of works that use EEG features to predict fatigue during a campimetry task.

4.3 Fatigue Prediction Using ECG Features

The ECG (electrocardiogram) features as inputs for machine learning algorithms predicting fatigue has also been studied. ECG signals are known to be correlated with the different stages of drowsiness/levels of user fatigue. (PATEL et al., 2011) used heart rate variability (HRV) to predict fatigue—in particular, they measured the variation in the time interval between heartbeats. This author states that when a person is changing from an awake to a drowsy state, the ratio of low heart frequency (0.04-0.15Hz) to high frequency (0.15-0.4Hz) decreases progressively. The data was gathered during a motor driver simulator task and a neural network was used to predict driver's fatigue. They were able to obtain an accuracy of 90% when using the same user's data in training and testing.

5 METHODOLOGY AND EXPERIMENTS

In this chapter we discuss how the different machine learning approaches proposed in this work (as ways of predicting user fatigue via ET and EEG measurements during a campimetric task) were set up and evaluated. First, we will explain our experimental setup—in particular, the hardware and software components used to collect and to process data. Then, we will discuss how the preprocessing of the acquired data was performed. At last, we will discuss the particular metrics that were chosen in order to compare the select machine learning algorithms, as well as the particular algorithms that will be applied to analyze the data and construct fatigue-predicting models.

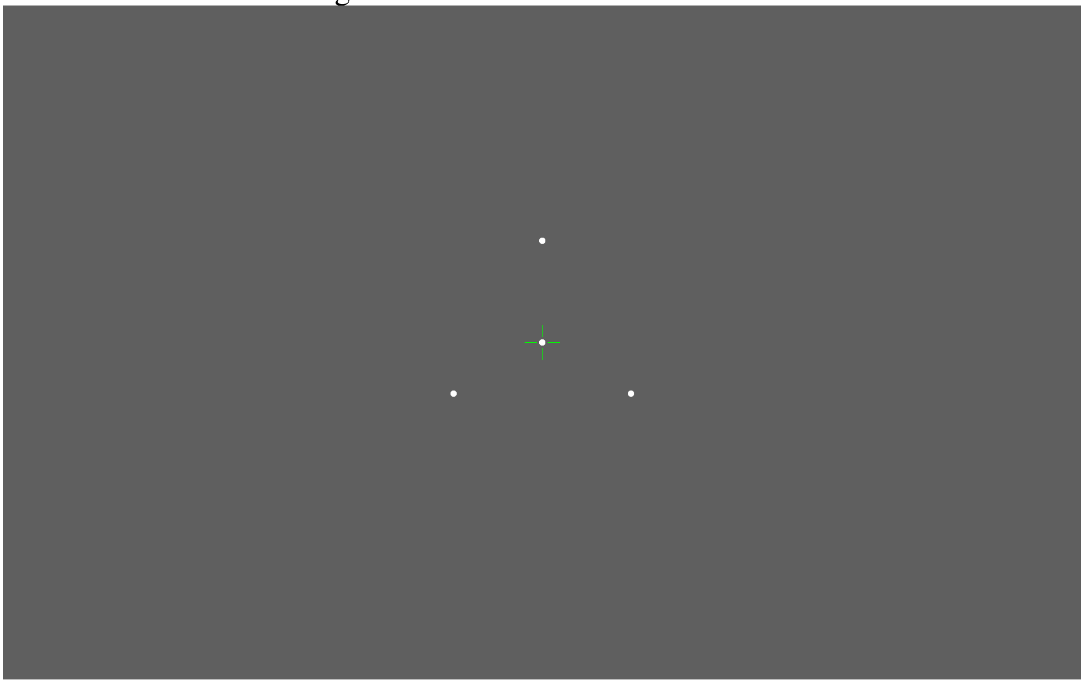
5.1 Experimental Setup

Nine subjects (4 females and 5 males, aged 20-32) participated in the campimetry experiment performed by me. During the campimetric task, the subjects were asked to press a button in response to a perceived light stimulus placed on a screen in front of them. A central fixation cross was shown in between stimuli in order to mark the gaze point that the user should track. Stimuli were presented at twelve different contrast levels on the screen and appeared at three different screen locations: (0°, 4.67°), (-4.07°, -2.34°) and (4.07°, -2.34°), and also at the center of the screen (i.e. (0°, 0°)); the latter occurred only in around 3 % of trials (UNGEWISS et al., 2016), in order to motivate looking back at the fixation cross. These locations are shown in Figure 5.1. Additionally, catch-trials were inserted: either trials where stimuli where shown with a contrast so high that an attentive subject was guaranteed to perceive it (positive catch-trials), or trials where stimuli were not visible at all (negative catch-trial). In the following, we refer to a *false negative* as a positive catch-trial that was not reacted to, and a *false positive* as a negative catch-trial (i.e., no stimulus presented) but for which the participant did press the response button.

The performance of the machine learning algorithms can be degraded when using only one source of fatigue indicators due to noise that can interfere with the quality of the data gathered. Our approach to mitigate this problem is to increase the amount of source of indicators. There may be noises in both at the same time, but the probability of this happening is small.

Eye movements and pupil diameter were recorded by means of a remote eye tracking device (*SMI RED250*) at a frequency of 250 Hz. A chin rest stabilized the position of

Figure 5.1: Possible stimuli locations.

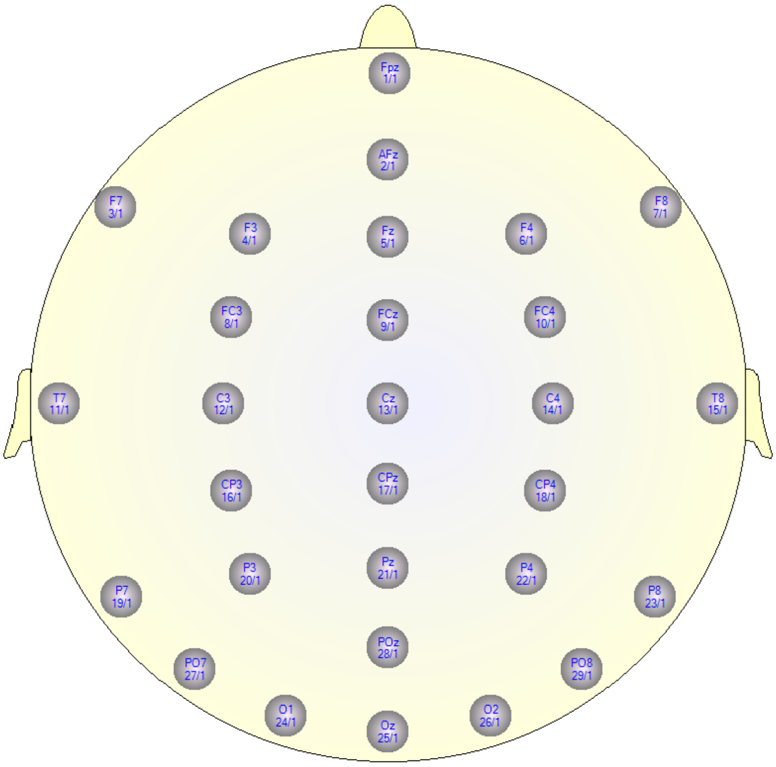


the head at 60 cm away from the screen. Twenty-nine EEG channels, (each channel corresponding to an electrode, placed as shown in Figure 5.2), were also recorded using two *g.USBamp* biosignal amplifiers/acquisition systems, at a frequency of 600 Hz. A standard system for electrode positioning, the *10/5 system*, was used to collect the EEG signals as shown in Figure 5.2. Three extra electrodes were used to collect electrooculogram (EOG) signals—i.e., the corneo-retinal resting electric potentials that exist between the front and the back of the eye. The time synchronization of the above-mentioned measurements was done using the timestamps offered by the eye tracker. The resolution of the monitor where stimuli were shown was of 1920x1200 pixels. In total 1,488 stimuli were presented to each participant, 25 % of which corresponded to positive catch-trials and 25 % to negative catch-trials. Each stimulus was shown for 200 ms and was followed by a 1300 ms response window during which it was expected that the user would respond. To increase the likelihood of fatigue, the experiments were conducted at two different times, 2:30 PM and 7:30 PM, and in a dark room where the experiment screen was the only source of light.

5.2 Data Pre-processing and Feature Construction

Due to the amount of data and its nature collected in our experiments, they need to be pre-processed prior to being used as input of machine learning algorithms. In this

Figure 5.2: EEG channels placement used in the experiments composed of 29 electrodes.



section we show how the features used to train our models were obtained from the raw data.

5.2.1 Error Rate

In this work we use error rate measurements as proxy of user fatigue. We recorded, for each presented stimulus, whether the user was correct or incorrect in their answer. A moving average over the stimuli with a window size of 200, corresponding to a total of 5 minutes of measurements, was used to create an error rate (ER) feature. This feature is defined as the sequence of smoothed-out errors ER_i, where each ER_i is as follows:

$$\mathrm{ER}_i = \frac{1}{200} \sum_{n=i}^{i+199} E_n \tag{5.1}$$

where E_n is 1 if there was either a false negative or a false positive in the stimulus n, and 0 otherwise. This causes the error rate to be a stepwise curve, making it necessary to smooth it.

5.2.2 Eye Tracking Features

The following ET indicators were collected in our experiments and which will be used as input training features:

- Blink duration;
- Blink rate;
- Average pupil Diameter;
- Pupil variability;
- Vergence;
- Fatigue wave (i.e., pupillary oscillations).

Blink rate, blink duration, average pupil diameter, and vergence were z-score normalized to the two first minutes of data as a baseline —i.e., we first calculated the mean (μ) and the standard deviation (σ) of the first two minutes of data and then the following formula was used to normalize each feature:

$$\mathrm{z}_i = \frac{x_i - \mu}{\sigma} \tag{5.2}$$

where $\mathrm{x_i}$ is one sample of the features mentioned above and $\mathrm{z_i}$ is the same sample after being z-scored.

Linear interpolation was used to fill in for moments where the user blinked, and in the case of non-blink related indicators, to deal with moments where a tracking loss occurred. Subsequently, vergence was filtered/smoothed via a moving-window filter selecting the 75%-percentile. Finally, all of the eye tracking related indicators were passed through a moving average filter (with window size of length 200 stimuli) in order to create a smoothed 1-1 correspondence between error rate measurements (described below) and each ET indicator. This smoothing process is as follows. Let I_i be the i-th eye tracking indicators, out of the ones mentioned above; the corresponding smoothed out, pre-processed indicator IP_i, is calculated as follows:

$$\mathrm{IP}_i = \frac{\sum_{n=t_j}^{t_{j+199}} I_n}{M} \tag{5.3}$$

where t_j is the time of j-th measured stimulus and M is the total number of samples collected for this indicator.

Sample pre-processed eye tracking features, as measured for one particular subject

(*NC*) are shown in Figures 5.3, 5.4 and 5.5.

Figure 5.3: Blink duration and blink rate of subject NC.

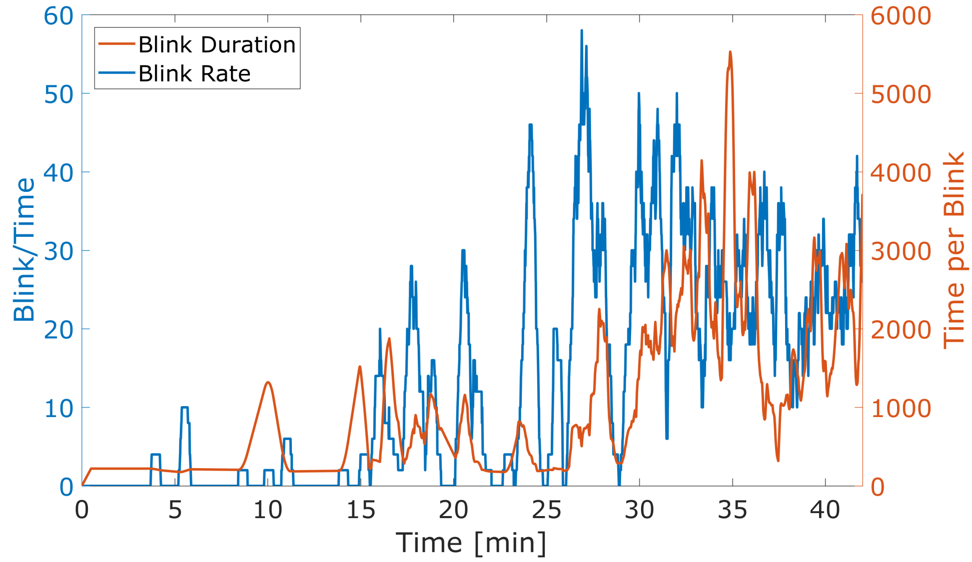
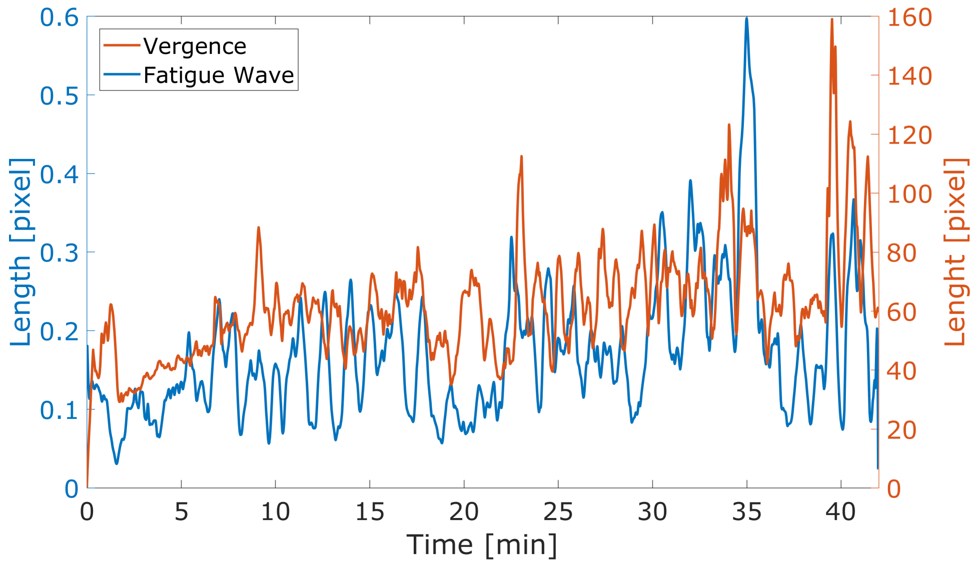


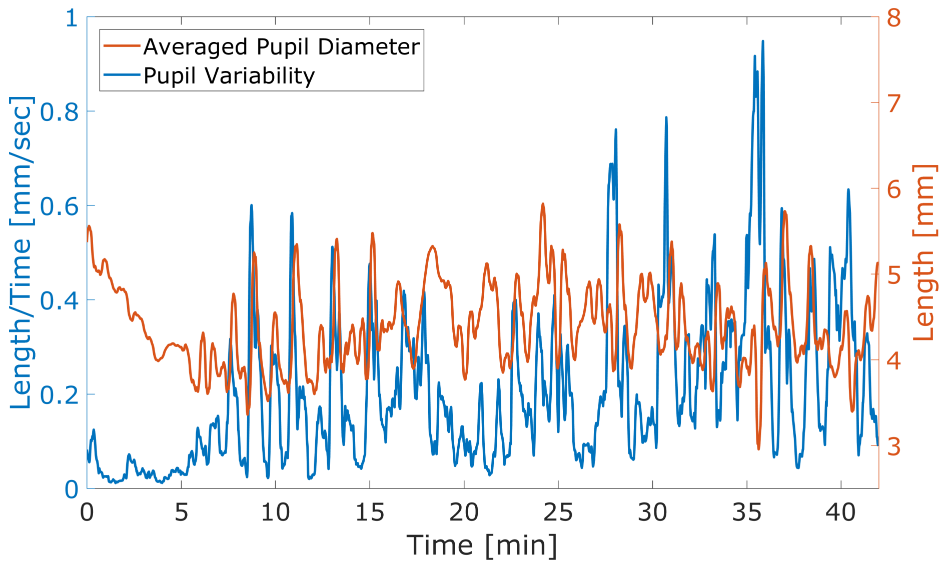
Figure 5.4: Vergence and fatigue waves of subject NC.



5.2.3 EEG Features

When measuring EEG signals, the occurrence of blinks and eye movements introduces artifacts (perturbations) to the recorded data. In this work we use EOG measurements to remove such artifacts from the measured EEG channels, as described in (SCHLÖGL et al., 2007). Artifacts in the EEG signal are stronger in the electrodes that are placed in the frontal lobe, and weaker in electrodes in the parietal, temporal and oc-

Figure 5.5: Average pupil diameter and pupil variability of subject NC.



cipital lobes. After removing artifacts, the electroencephalogram waves were also z-score normalized to the two first minutes of data as a baseline, as done with the eye tracking indicators (5.2). Then, the power spectrum density (PSD) in the data from one stimulus time to the next was calculated using Burg's method (BURG, 1975). Only the theta (4-8 Hz) and alpha (8-13 Hz) waves were used in this study since that these are known to be the best indicators of fatigue (TREJO et al., 2015; GHARAGOZLOU et al., 2015; KONG et al., 2015). This transformation generated four curves in frequency domain for each signal in time domain, making a total of 116 EEG features. Finally, these signals were filtered using the same process as was described for the eye tracking indicators (5.3).

5.2.4 List of Features

It is interesting to summarize the features used as input of the machine learning algorithms. In table 5.1, we can see, for each source, which data was used and how many different features each one corresponds to. After processing the data, we obtained 1,288 samples of each feature for each one of the subjects.

5.3 Machine Learning Algorithms for Modeling Fatigue

As described earlier, in this work we evaluated a range of different machine learning algorithms with different characteristics with the goal of modeling user fatigue. The

Table 5.1: List of features used

Source	Name	Number of Features
Eye tracking	Blink Duration	1
	Blink Rate	1
	Average Pupil Diameter	1
	Pupil Variability	1
	Vergence	1
	Fatigue Wave	1
EEG	Alpha Wave	58
	Theta Wave	58
		Total: 122

algorithms are listed below:

- Support Vector Regression;
- LASSO;
- Neural Network;
- Gaussian Processes.

For each one of these, a large number of combinations of parameters (e.g., number of neurons in a neural network) was tested in order to find the best configuration. To evaluate the quality of the prediction of each model we used Pearson's Correlation Coefficient (CC) and the root-mean-squared error (RMSE), as those are good performance indicators (SPÜLER et al., 2015). The CC is used as the first ranking criteria and RMSE as the tie breaker criteria.

The user error rate (feature ER) in response to catch-trials was predicted using two different approaches. The first approach was **5-fold cross-validation**, where the *data of each subject* was divided in 10 intervals. This was done because there is a temporal dependency in the data. In each run of this approach, from the 50 sets of data, 40 were used as training and the remaining as testing.

The second approach was predicting by a **leave-one-subject-out cross-validation**, in which *data of 4 subjects* was used to train the prediction model and it was tested on the remaining subject. This procedure was repeated for each subject.

For each one of these approaches, we ran each machine learning algorithm with three different sets of input features: first, using only the 6 ET-related features; only the 116 EEG related features; and all the 122 features.

6 EMPIRICAL RESULTS

In this chapter we will present the empirical performance results obtained by using the two different approaches for predicting user fatigue, comparing different machine learning algorithms in terms of performance. Some of the subjects that were part of the data-collecting experiments finished the experiment early due to abortion or technical difficulties; in this work we analyzed only the data from the five subjects with a complete recording. In a previous work (DAMBROS et al., 2017) all participants' data was analyzed.

6.1 5-Fold Cross-Validation

A comparison among the different machine learning methods (and different performance metrics—Pearson's Correlation Coefficient and RMSE) can be seen in Tables 6.1, 6.2, 6.3 for the case where only ET features, only EEG features, and both were used, respectively. The correlation and root-mean-square error between the error rate and the prediction is shown for each method used. All correlations are the mean over the five runs, one for each fold.

Table 6.1: Comparison between methods for the 5-fold cross-validation approach using only ET features.

Method	**Correlation**	**RMSE**
Gaussian Process	0.91	0.010
Neural Network	0.88	0.015
LASSO	0.88	0.023
Linear SVR	0.81	0.019

Table 6.2: Comparison between methods for 5-fold cross-validation approach using only EEG features.

Method	**Correlation**	**RMSE**
Gaussian Process	0.96	0.008
Linear SVR	0.94	0.011
LASSO	0.92	0.030
Neural Network	0.89	0.12

In the 5-fold cross-validation approach an average correlation of 0.94±0.08 was obtained using Gaussian Process regression to train the prediction model.

For the 5-fold cross-validation approach, the model that predicted the error rate more accurately was generated using Gaussian process EEG features. Different from what

Table 6.3: Comparison between methods for the 5-fold cross-validation approach using both ET and EEG features.

Method	Correlation	RMSE
Gaussian Process	0.94	0.008
LASSO	0.93	0.021
Linear SVR	0.90	0.033
Neural Network	0.84	0.019

we expected, the model generated by Gaussian process with both ET and EEG features had a slightly worse performance than with only EEG, but the difference is rather small to conclude that the former is better than the latter. This could have happened by chance or overfitting. Analyzing the Tables 6.1 and 6.2, we can conclude that, when the data of the subject is both in training and testing, the EEG indicators react more consistently to fatigue than ET.

6.2 Leave-one-subject-out cross-validation

In the leave-one-subject-out cross-validation approach an average correlation of 0.80 ± 0.07 was obtained by the best ML model that we evaluated; that is, a 0.08 increase in average performance (and also a 0.1 decrease in the standard deviation) compared with the previous work (DAMBROS et al., 2017).

In Tables 6.4, 6.5, 6.6, a comparison among the different evaluates machine learning methods is shown—both for the case where only ET features, only EEG features, and both were used, respectively. The correlation and root-mean-square error between the actual error rate and the predicted error rate are shown for each method. All correlations are means over the five runs, one for each subject.

Table 6.4: Comparison between methods for the leave-one-subject-out cross-validation approach using only ET features.

Method	Correlation	RMSE
LASSO	0.73	0.030
Gaussian Process	0.73	0.030
Linear SVR	0.72	0.031
Neural Network	0.56	0.030

Figures 6.1, 6.2, and 6.3 show a few representative examples of the fatigue-predicting models learned by different algorithms. They show, in particular the true error rate for a selected participant (AB) and its corresponding predicted error rate, when using the LASSO algorithm with either EEG features, or ET features, and also the predictions made

Table 6.5: Comparison between methods for the leave-one-subject-out cross-validation approach using only EEG features.

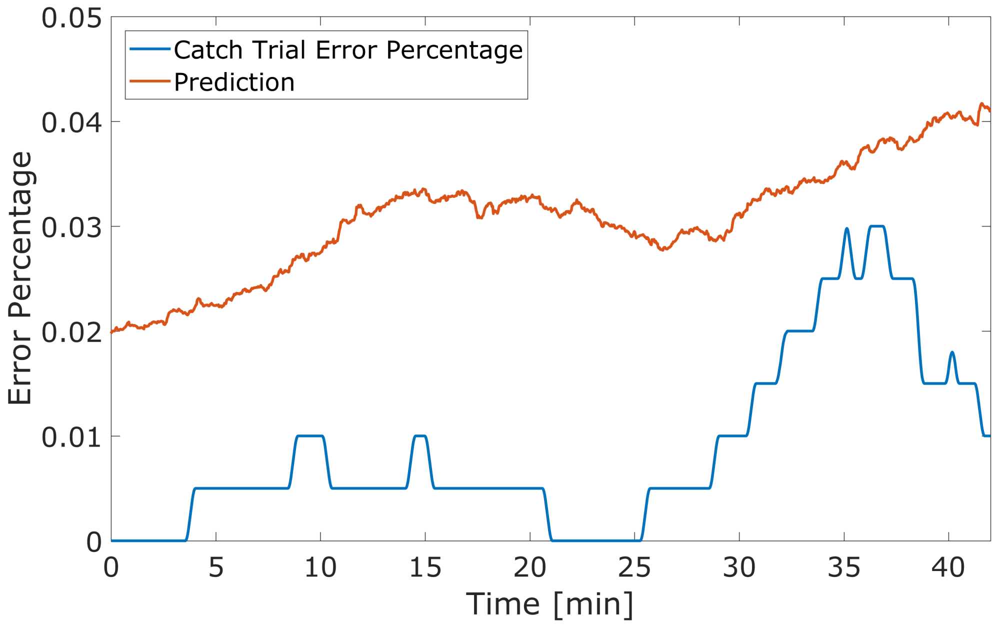
Method	Correlation	RMSE
LASSO	0.64	0.023
Linear SVR	0.59	0.099
Neural Network	0.35	0.040
Gaussian Process	0.22	0.192

Table 6.6: Comparison between methods for the leave-one-subject-out cross-validation approach using both ET and EEG features.

Method	Correlation	RMSE
Linear SVR	0.80	0.030
LASSO	0.76	0.029
Gaussian Process	0.75	0.032
Neural Network	0.57	0.019

by SVR with both types of features. Due to the large number of possible models that were evaluated in this work, resulting from combinations of types of training features and learning algorithms, we show in the main text only the above-mentioned examples. All the other plots can be found in the Appendix chapter.

Figure 6.1: Error rate of subject AB and prediction given by the LASSO algorithm trained with EEG features.



More detailed results for the leave-one-subject-out cross-validation approach can be seen in Table 6.7, where both the correlation and the root-mean-square error between the error rate and its corresponding prediction are shown for each participant. The model used to generate the data in this table was Support Vector Regression with a linear kernel,

Figure 6.2: Error rate of subject AB and prediction given by the LASSO algorithm trained with ET features.

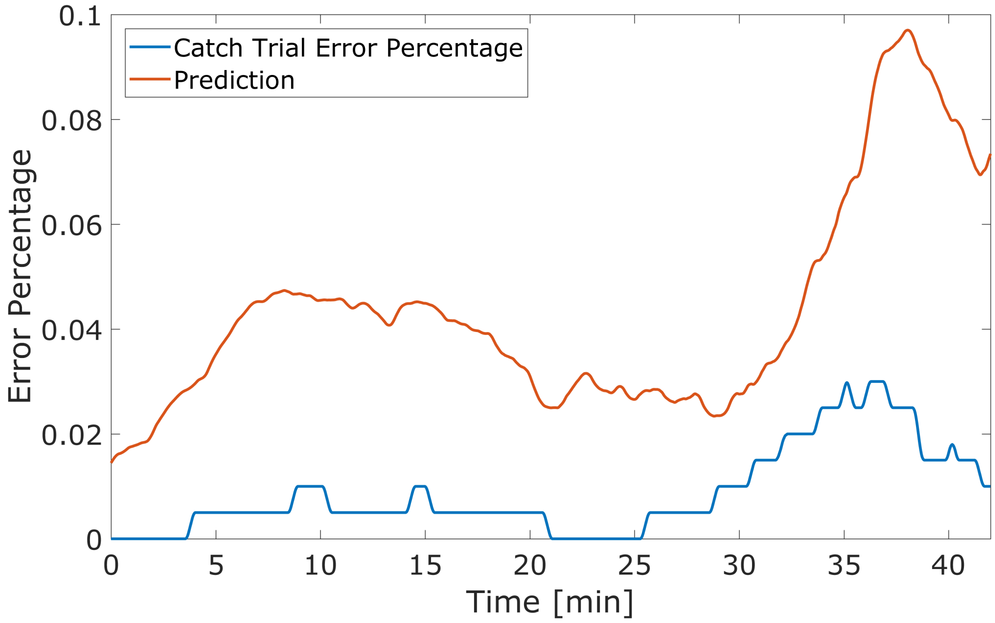
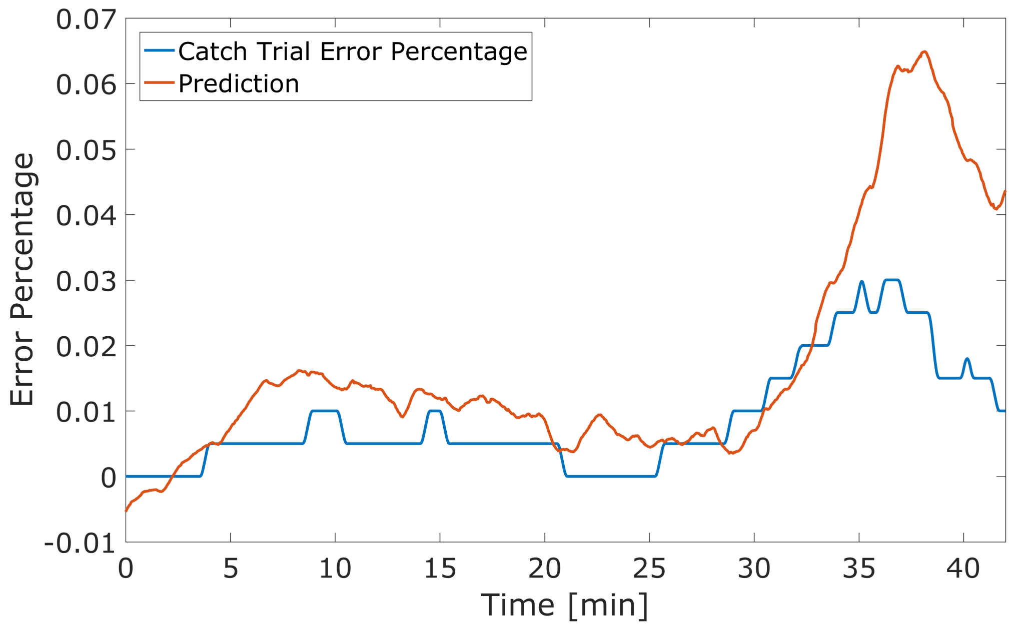


Figure 6.3: Error rate of subject AB and prediction given by the SVR algorithm trained with ET and EEG features.



operating over all of the 122 features available; as was seen in the summarized table (Table 6.6), this was the approach that performed best.

Table 6.8 shows the weights assigned to each of the ET indicators/features by the best performing linear support vector regression model, when trained on data from the 5 subjects. The reason why the vergence feature has one of the largest weights in

Table 6.7: Comparison table with correlation between catch error rate and prediction, and root-mean-square error of each subject for the three sets of training data.

	ET		**EEG**		**ET and EEG**	
Subject	**Correlation**	**RMSE**	**Correlation**	**RMSE**	**Correlation**	**RMSE**
NC	0.86	0.040	0.83	0.018	0.88	0.041
AB	0.77	0.038	0.66	0.023	0.81	0.019
PZ	0.63	0.016	0.60	0.015	0.71	0.012
LR	0.62	0.036	0.64	0.041	0.83	0.035
JD	0.77	0.016	0.49	0.018	0.76	0.015
Mean	0.73	0.029	0.64	0.023	0.80	0.024
Std	0.10	0.012	0.12	0.010	0.07	0.013

the prediction model is that the fixation cross in the experiments forced a small inward deviation of the subjects eyes; as the subject starts to get tired, the eye muscles become exhausted, forcing the eyes to approach a resting position. The positive and relatively large coefficient of blink duration indicates that the subjects reached a sleepiness state during the task.

Table 6.8: Coefficients for each ET feature obtained by linear regression SVR with linear kernel trained with the data from the 5 subjects. To make coefficients comparable, all features where z-score normalized.

Feature	**Coefficient**
Blink Duration	0.0971
Blink Rate	-0.0327
Average Pupil Diameter	-0.0155
Pupil Variability	-0.0615
Vergence	0.0758
Fatigue Wave	0.0654

Figure 6.9 shows the performance difference when using only the theta and alpha waves (116 features) compared when using all of the spectrum of the EEG signals (i.e. not only theta and alpha waves, but also delta and beta waves, a total of 1247 features). Both results were obtained using SVR. The results when using all of the spectrum are worst because the algorithms try to explain the error rate based on features that may not be correlated with fatigue, and in doing so end up overfitting to the data.

For this approach, the model that predicted the error rate more reliably was generated using linear SVR with ET and EEG features. The model learns common characteristics of people in the sense of how do theirs physiologic responses react to fatigue. This approach is useful for real applications because the model can be trained with data of many users and then any other person can make use of it without having to train the model again. Analyzing the Tables 6.4 and 6.5, we can also conclude that the eye-tracker indicators are more constant over subjects than EEG indicators, that is, the ET indicators of two subjects tend to respond to fatigue more similarly than the EEG indicators.

Comparing the Tables 6.1 and 6.2 with 6.4 and 6.5, we can observe that adding EEG data of a testing user to the training process is more effective due to the fact discussed above.

Table 6.9: Comparison table with correlation between catch error rate and prediction, and root-mean-square error for each subject, for all the spectrum and only theta and alpha waves.

	ET and EEG (0-100 Hz)		**ET and EEG (4-12 Hz)**	
Subject	**Correlation**	**RMSE**	**Correlation**	**RMSE**
NC	0.34	0.039	0.88	0.041
AB	0.31	0.073	0.81	0.019
PZ	-0.29	0.226	0.71	0.012
LR	0.81	0.053	0.83	0.035
JD	-0.60	0.081	0.76	0.015
Mean	0.11	0.112	0.80	0.024
Std	0.56	0.016	0.07	0.013

7 CONCLUSION AND FUTURE WORKS

This work evaluated different machine learning algorithms tasked with predicting user fatigue; we investigated the effectiveness of these algorithms when operating over different types of physiological data and when evaluated according to different performance metrics. The results indicate that using both features obtained by eye tracking devices and EEG can be used to construct a reliable prediction model of the error rate, and thereby can be used to estimate response quality in a task (campimetry examination).

Individual ET and EEG measurements are spiky and noisy, and depend highly on individual baselines. For example, blink behaviour is a highly predictive feeature for many subjects; however, the experimental procedure that we performed implied a certain optimal blink time that may mask this effect for other subjects. By training a regression model based on many different physiological measurements, we were to able to combine multiple indicators in order to obtain a more stable error rate prediction across all subjects.

The use of two different sources of indicators increased the predictor's robustness to noise, since (as the experiments showed) poor predictions may occur when individual indicators/features are used or cannot be acquired accurately or reliably. This problem is observed when users make use of glasses, for instance, which adds a serious problem to eye tracker devices. Besides that, in EEG measurements, the very small signal-to-noise ratio of the signal results in noisy features when the user moves their head abruptly; another source of imprecise feature measurements occurs due to cardiac signals interfering with the EEG measurements.

The performance results we obtained were better when using a 5-fold cross-validation approach. The reason why this occurs is that the data of all subjects is present during both the model training and testing stages. However, when transferring these results to an application to be deployed in real life, this procedure implies that training data for a new user would be needed *before* they would be allowed to use the system. The leave-one-subject-out cross-validation, by contrast, estimates the performance of the model under different conditions, and in a way that does not require such user-specific training data before the model can make predictions for novel participants. The second training approach is, therefore, more flexible and easy to deploy.

It can also be noticed in our experiments that the performance difference between algorithms when using the different combinations of features, is not significantly large when the methods are trained using the first approach (5-fold cross-validation). This happens because the algorithms were capable of memorizing the user's data due to the fact that the same user's data was both in the training and testing set. This problem does not occur in the second training approach (leave-one-subject-out cross-validation), thus making it clearer that the combination of the two types of features increases the performance of the model and that a new user would not need to go through a training phase before using the system.

As future work, we would like to apply our fatigue-prediction method to different tasks, besides campimetry, where no easily measurable performance indicators (such as error rate) are available. There are certain preconditions that need to be met when generalizing our approach to other experimental setups that might restrict the availability of individual indicators or physiological features/measurements. For example, training models based on vergence angle requires binocular tracking devices, and fatigue waves have only been found to be relevant in scotopic and mesopic conditions (HENSON; EMUH, 2010).

We believe that a reliable and robust fatigue prediction model constructed via eye tracking measurements and EEG signals will enable intelligent devices to adapt to the cognitive state and current abilities of their users, and could thereby allow such systems to interact with their users in a more efficient and productive way.

REFERENCES

ÅKERSTEDT, T.; KECKLUND, G.; KNUTSSON, A. Manifest sleepiness and the spectral content of the eeg during shift work. **Sleep**, v. 14, n. 3, p. 221–225, 1991.

ASTERIADIS, S. et al. Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. **Multimedia Tools and Applications**, v. 41, n. 3, p. 469–493, 2009.

BENGTSSON, B.; HEIJL, A. A visual field index for calculation of glaucoma rate of progression. **Am J Ophthalmol**, v. 145, p. 343–353, 2008.

BENGTSSON, B. et al. A new generation of algorithms for computerized threshold perimetry, sita. **Acta Ophthamol**, v. 75, n. 2, p. 368–375, 1997.

BERGASA, L. M.; NUEVO, J.; SOTELO, M. A. Real-time system for monitoring driver vigilance. **IEEE Intelligent Vehicles Symposium**, v. 7, p. 63–77, 2006.

BERGER, H. Über das elektrenkephalogramm des menschen. **Psychiatrie**, Springer, v. 87, 1929.

BRUNEAU, D.; SASSE, M. A.; MCCARTHY, J. The eyes never lie: The use of eye tracking data in hci research. **Proceedings of the CHI**, v. 2, p. 25–30, 2002.

BURG, J. P. **Maximum Entropy Spectral Analysis**. Thesis (PhD) — Stanford University, 1975.

CANISIUS, S.; PENZEL, T. Vigilance monitoring - review and practical aspects. **Biomedizinische Technik**, v. 52, p. 77–82, 2007.

DAMBROS, G. V. et al. Monitoring response quality during campimetry via eye-tracking. In: **22st International Conference on Intelligent User Interfaces, IUI 2017, ACM**. [S.l.]: Association for Computing Machinery, 2017.

DRUCKER, H. et al. Support vector regression machines. **Advances in neural information processing systems**, Morgan Kaufmann Publishers, v. 9, p. 155—-161, 1997.

GHARAGOZLOU, F. et al. Detecting driver mental fatigue based on eeg alpha power changes during simulated driving. **Sensors**, v. 15, p. 1693–1700, 2015.

HEAD, H. The conception of nervous and mental energy. ii. vigilance: A physiological state of the nervous system. **British Journal of Psychology**, v. 14, p. 125–147, 1923.

HENSON, D. B.; EMUH, T. Monitoring vigilance during perimetry by using pupillography. **Investigative Ophthalmology & Visual Science**, The Association for Research in Vision and Ophthalmology, v. 51, n. 7, p. 3540–3543, 2010.

JAVAL, L. E. Essai sur la physiologie de la lecture. **Annales d'Oculistique**, v. 79, p. 97–117, 1878.

JI, Q.; ZHU, Z.; YU-PIN, L. Real-time nonintrusive monitoring and prediction of driver fatigue. **IEEE Transactions on Vehicular Technology**, v. 53, n. 4, p. 1052 – 1068, 2004.

JO, J. et al. Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. **Optical Engineering**, v. 50, n. 12, 2011.

KAUFMANN, H.; STEFFEN, H. **Strabismus**. 4th. ed. Stuttgart: Georg Thieme Verlag, 2012.

KONG, W. et al. Investigating driver fatigue versus alertness using the granger causality network. **Iranian Journal of Public Health**, v. 44, n. 12, p. 19181—-19198, 2015.

LIN, C.-T. et al. Eeg-based drowsiness estimation for safety driving using independent component analysis. **IEEE Transactions on Circuits and Systems I**, v. 52, n. 12, p. 2726–2738, 2005.

LOWENSTEIN, O.; FEINBERG, R.; LOEWENFELD, I. E. Pupillary movements during acute and chronic fatigue a new test for the objective evaluation of tiredness. **Investigative Ophthalmology & Visual Science**, Association for Research in Vision and Ophthalmology, v. 2, n. 2, p. 138–157, 1963.

MURPHY, K. P. **Machine Learning - A Probabilistic Perspective**. Cambridge, Massachusetts: The MIT Press, 2012.

PATEL, M. et al. Applying neural network analysis on heart rate variability data to assess driver fatigue. **Expert Systems with Applications**, v. 38, n. 6, p. 7235–7242, 2011.

SARINANA, J. **Memory, Photography, and the Human Brain**. 2010. Available from Internet: <http://www.joshuasarinana.com/memory-photography-brain/>.

SCHLEICHER, R. et al. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? **Ergonomics**, Taylor & Francis, v. 51, n. 7, p. 982–1010, 2008.

SCHLÖGL, A. et al. A fully automated correction method of eog artifacts in eeg recordings. **Clinical Neurophysiology**, Elsevier, v. 118, n. 1, p. 98–104, 2007.

SHEN, K.-Q. et al. Eeg-based mental fatigue measurement using multi-class support vector machines with confidence estimate. **Clinical Neurophysiology**, v. 119, n. 7, p. 1524–1533, 2008.

SPÜLER, M. et al. Comparing metrics to evaluate performance of regression methods for decoding of neural signals. In: IEEE. **2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.], 2015. p. 1083–1086.

TAFAJ, E. et al. Vishnoo – an open-source software for vision research. In: IEEE. **24th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2011. p. 1–6.

TAFAJ, E. et al. Mobile and fast detection of visual field defects for elderly drivers as a necessary input into driver assistance systems for mobility maintenance. **GMM-Fachbericht-AmE 2010-Automotive meets Electronics**, VDE VERLAG GmbH, 2010.

TEPLAN, M. Fundamentals of eeg measurement. **Measurement Science Review — 2**, v. 2, p. 1–11, 2002.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society**, Wiley, v. 58, n. 1, p. 267–288, 1996.

TREJO, L. J. et al. Eeg-based estimation and classification of mental fatigue. **Psychology**, Scientific Research, v. 6, n. 5, p. 572–589, 2015.

UNGEWISS, J. et al. Monitoring quality and vigilance during automated static perimetry. **22th International Visual Field and Imaging Symposium**, 2016.

VERNET, M.; KAPOULA, Z. Binocular motor coordination during saccades and fixations while reading: a magnitude and time analysis. **Journal of Vision**, The Association for Research in Vision and Ophthalmology, v. 9, n. 7, p. 2–2, 2009.

WEESS, H. G. et al. Vigilanz, einschlafneigung, daueraufmerksamkeit, müdigkeit, schläfrigkeit. die messung müdigkeitsbezogener prozesse bei hypersomnien. theoretische grundlagen. **Somnologie**, v. 2, p. 32–41, 1998.

WILHELM, B. et al. Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. **Journal of Sleep Research**, Wiley Online Library, v. 10, n. 1, p. 1–7, 2001.

YEO, M. V. et al. Can svm be used for automatic eeg detection of drowsiness during car driving? **Safety Science**, v. 47, n. 1, p. 115–124, 2009.

APPENDIX

In this chapter, we will show the plots obtained for the leave-one-subject-out cross-validation approach under different learning algorithms and features. Recalling that, in this approach, data from four users were used for training and data from the fifth for testing.

Appendix A

In this appendix, we show each user's error rate and its corresponding prediction obtained using LASSO algorithm with 6 ET-related features.

Figure A.1: Error rate and prediction for subject NC

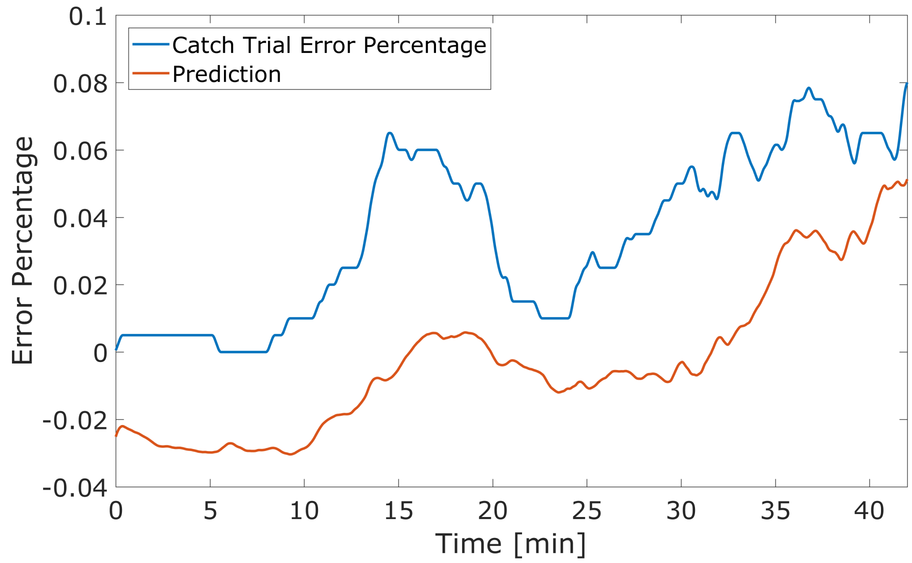


Figure A.2: Error rate and prediction for subject AB

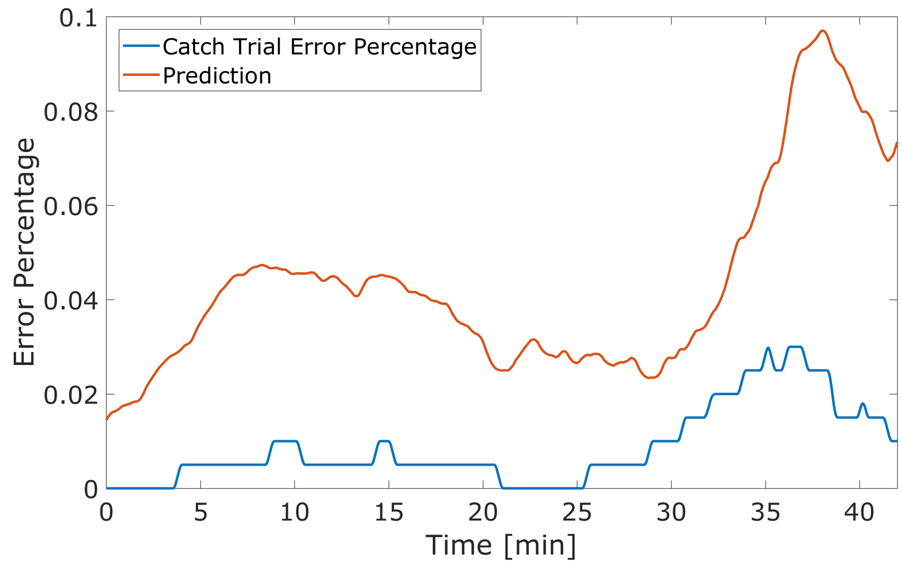


Figure A.3: Error rate and prediction for subject PZ

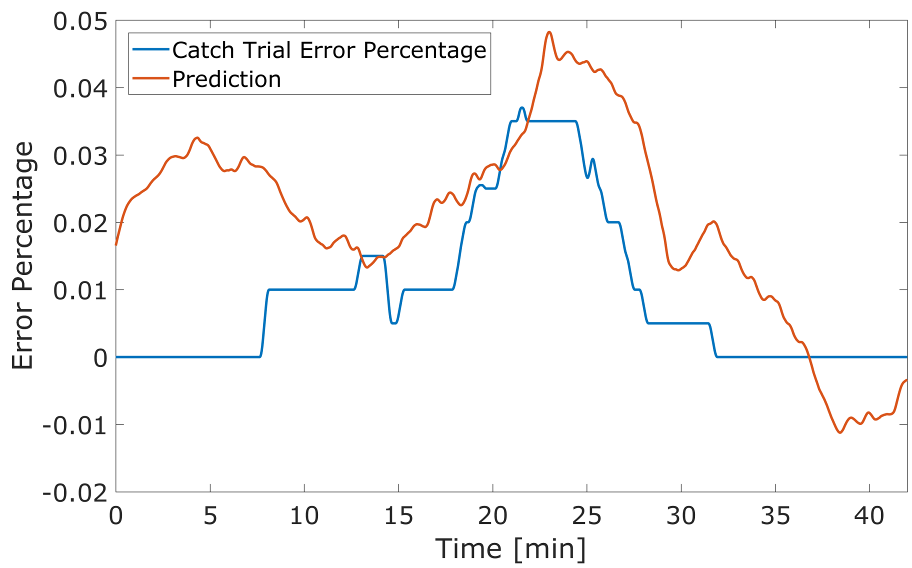


Figure A.4: Error rate and prediction for subject LR

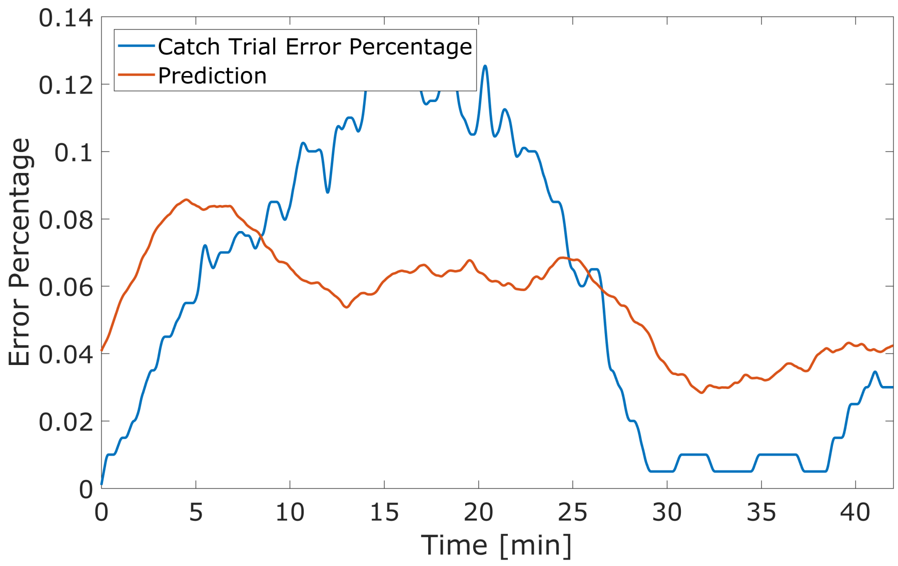
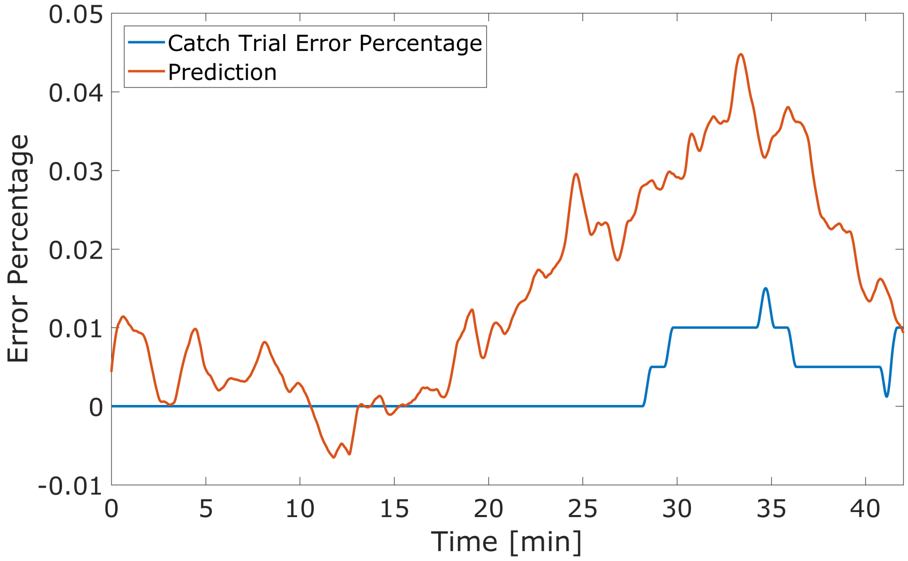


Figure A.5: Error rate and prediction for subject JD



Appendix B

In this appendix, we show each user's error rate and its corresponding prediction obtained using LASSO algorithm with 116 EEG-related features.

Figure B.1: Error rate and prediction for subject NC

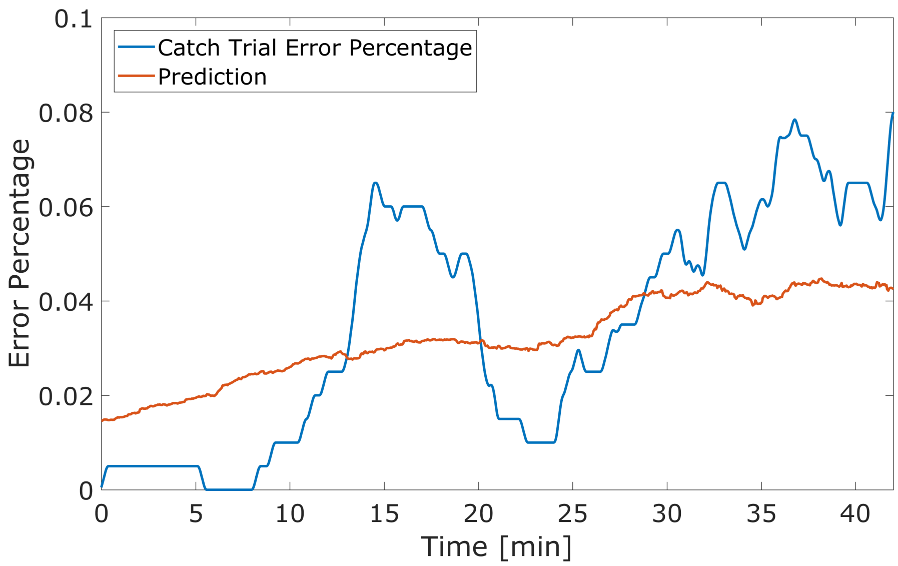


Figure B.2: Error rate and prediction for subject AB

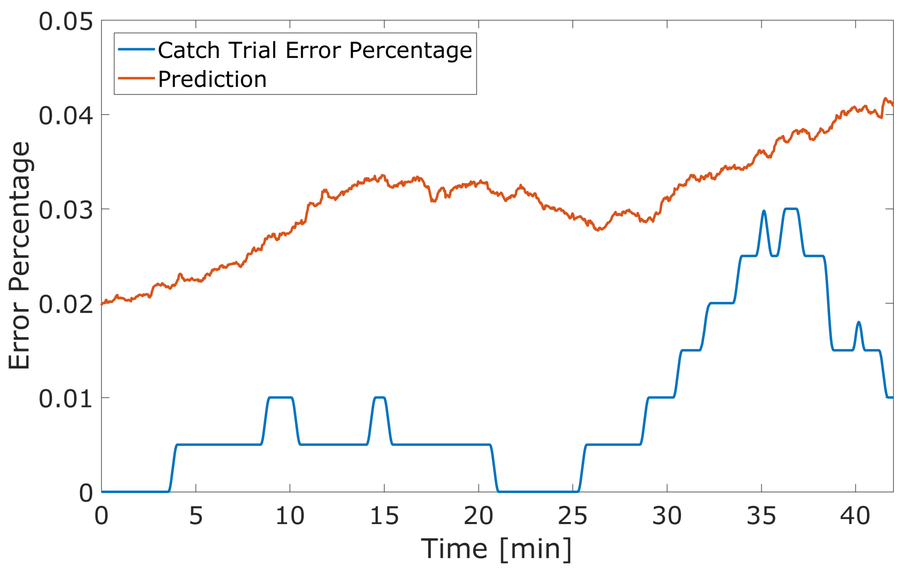


Figure B.3: Error rate and prediction for subject PZ

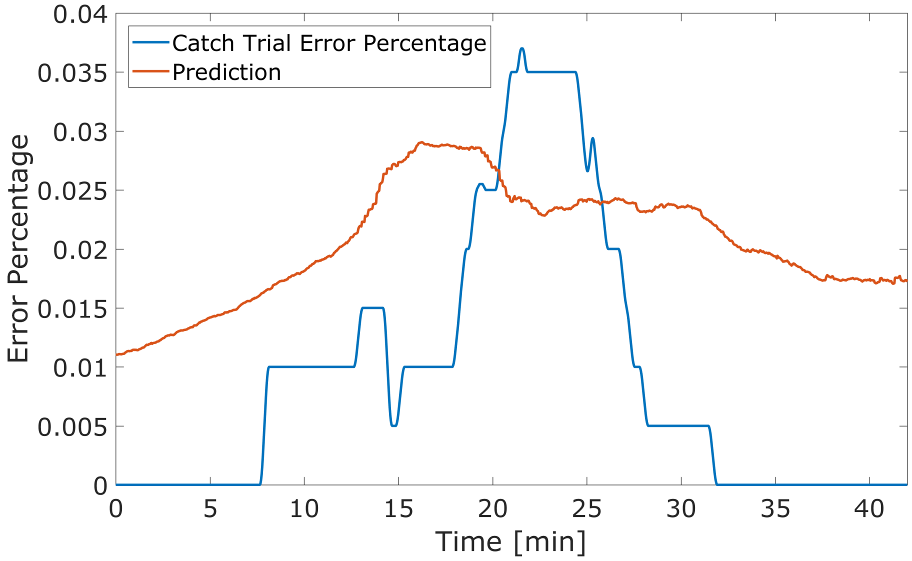


Figure B.4: Error rate and prediction for subject LR

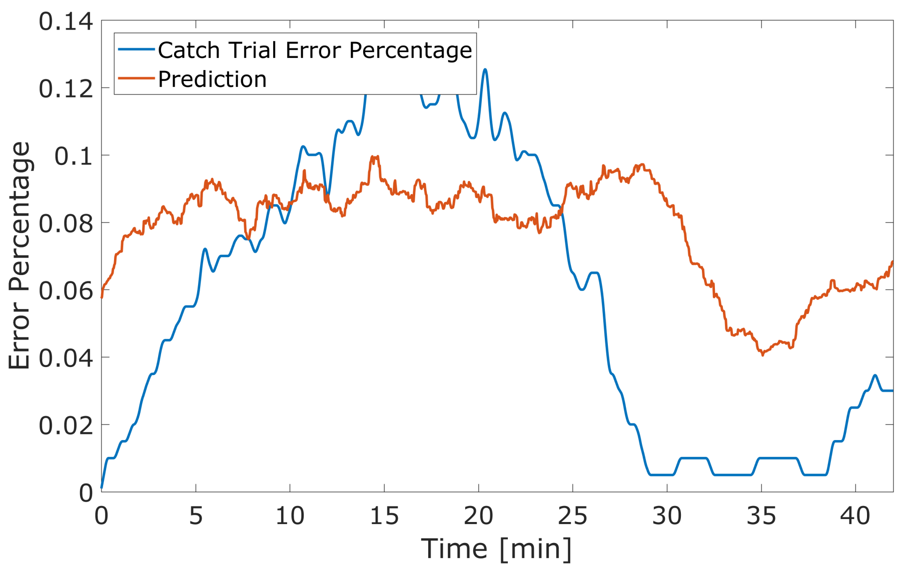
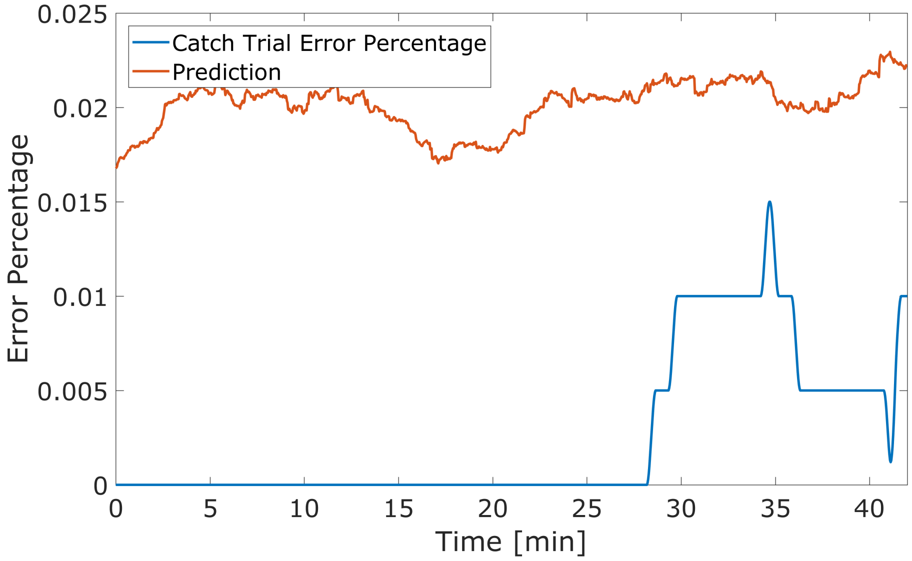


Figure B.5: Error rate and prediction for subject JD



Appendix C

In this appendix, we show each user's error rate and its corresponding prediction obtained using linear SVR algorithm with all the 122 ET and EEG features.

Figure C.1: Error rate and prediction for subject NC

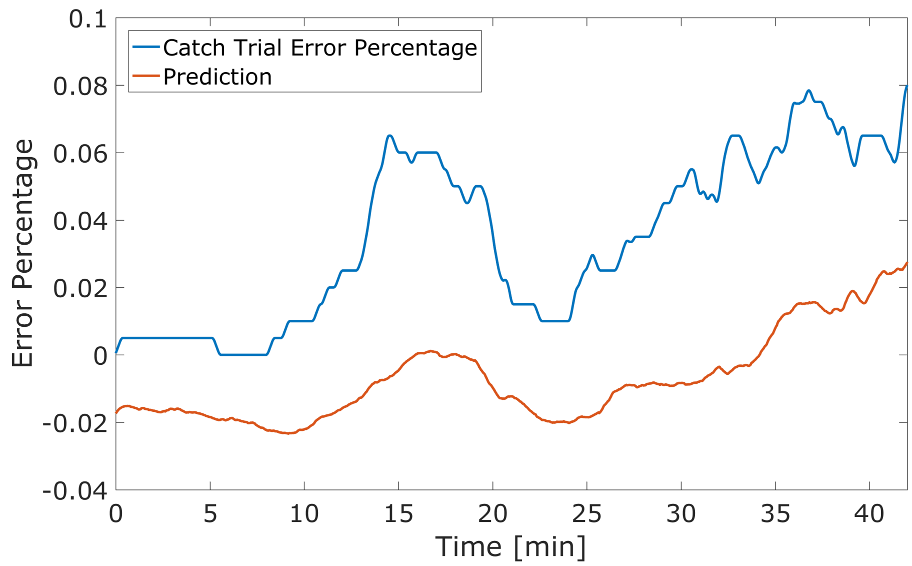


Figure C.2: Error rate and prediction for subject AB

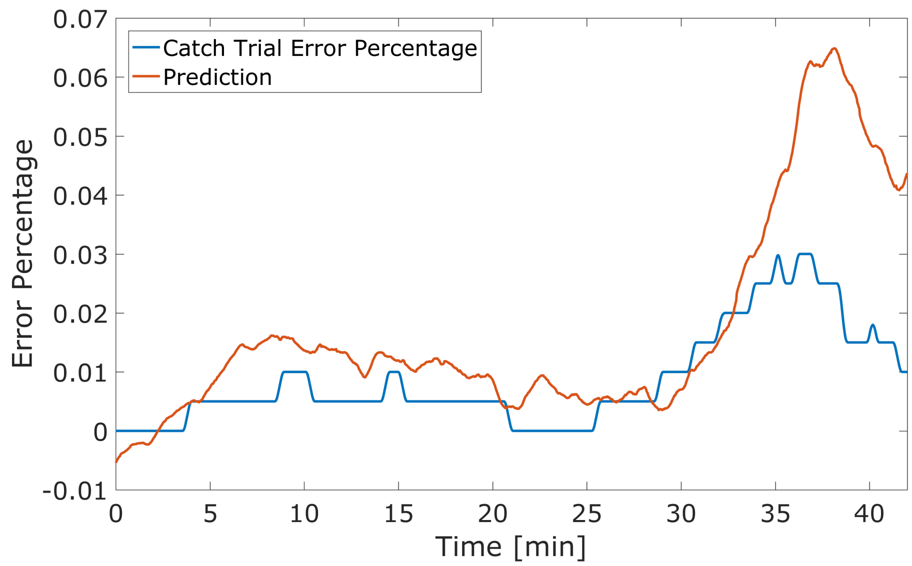


Figure C.3: Error rate and prediction for subject PZ

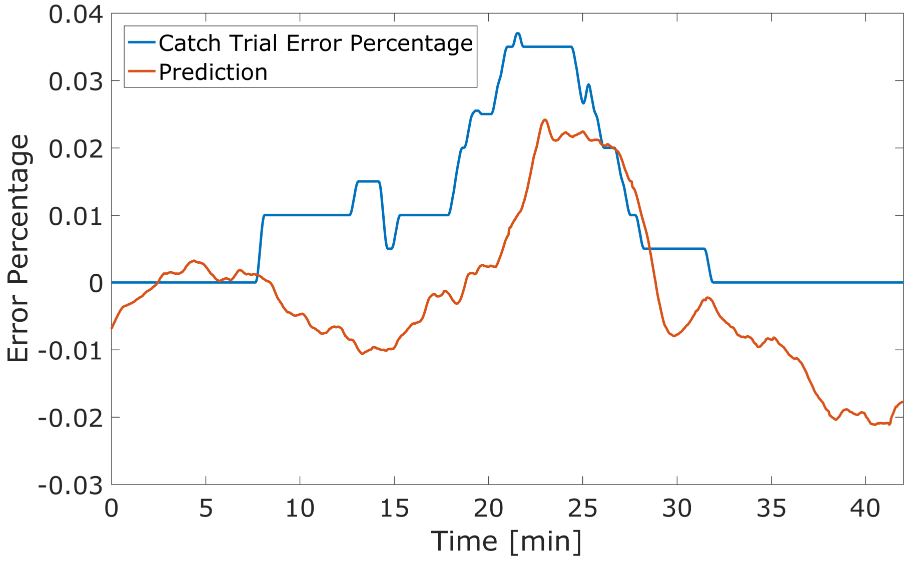


Figure C.4: Error rate and prediction for subject LR

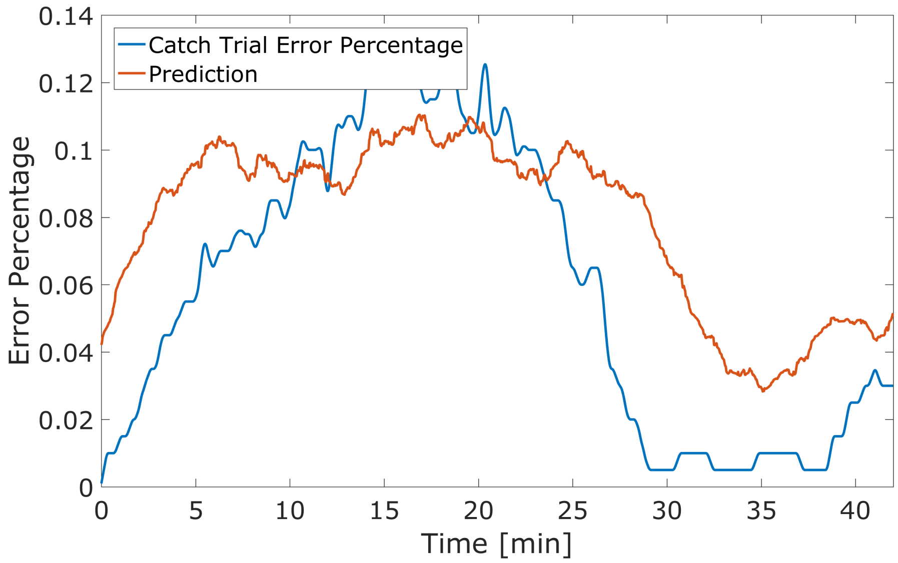
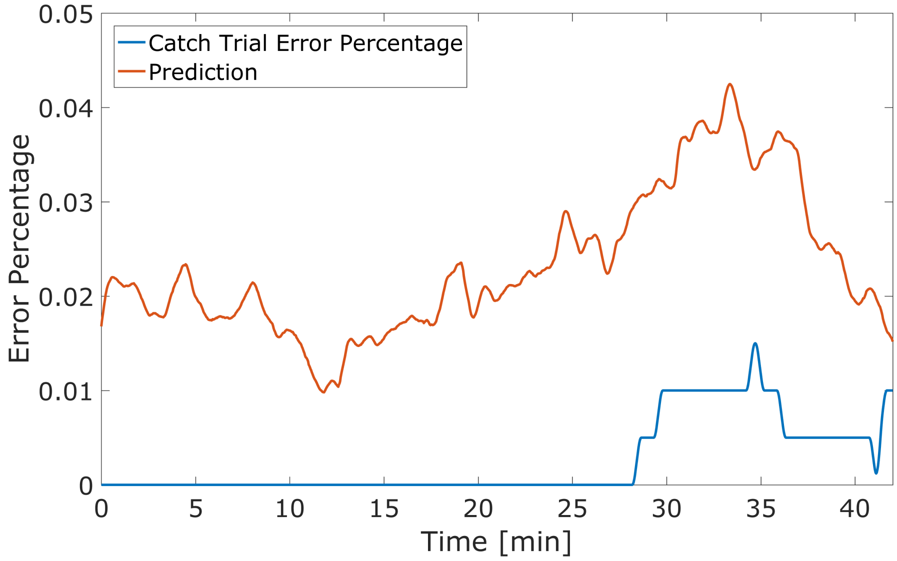


Figure C.5: Error rate and prediction for subject JD



Predicting Response Quality as a Proxy of Fatigue via Eye-Tracking and EEG

Gustavo Vergani Dambros[1], Bruno Castro da Silva[1], Martin Spüler[2]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]Computer Engineering Institute – Universität Tübingen
Tübingen – Germany

{gvdambros,bsilva}@inf.ufrgs.br, spueler@informatik.uni-tuebingen.de

Abstract. *Many computer systems present the possibility of adapting depending on the degree of fatigue of their users, so it is of interest to estimate how tired they are. As there is no method to measure fatigue directly, we investigated whether the use of data collected via eye-tracking (ET) and electroencephalography (EEG) can predict user error rate during a simple visual experiment. Using the rate with which an user errors while solving a given task as a proxy of fatigue is possible because there is evidence that both values are correlated. We will compare different machine learning methods in order to find the one that results in better predictions of user error rate, given ET and EEG features. The data were collected during a 40 minute campimetric task, where a given user must detect visual stimuli (i.e., points) of different contrasts.*

Resumo. *Vários sistemas computacionais apresentam a possibilidade de se adaptarem dependendo do grau de fatiga de seus usuários, sendo assim, é de interesse estimar o quão cansado os mesmos estão. Como não há um método para mensurar a fadiga diretamente, investigamos se o uso de dados coletados via rastreamento ocular e eletroencefalografia (EEG) podem predizer a taxa de erro durante um experimento visual simples. Usar a taxa com a qual os erros de um usuário ao resolver uma determinada tarefa como proxy da fadiga é possível porque existe evidência de que ambos os valores estão correlacionados. Compararemos diferentes métodos de aprendizado de máquina com o intuito de encontrar melhores predições, com base nos dados de ET e EEG. Os dados foram coletados durante uma tarefa campimétrica de 40 minutos, onde o usuário deve detectar estímulos visuais (ou seja, pontos) de diferentes contrastes*

1. Introduction

Vigilance is derived from the Latin word *vigilantia* and can be translated as wakefulness. As a technical term, vigilance describes primarily the degree of central nervous activation [Head 1923] and for that, is used to describe a physiological state of alertness [Canisius and Penzel 2007]. In contrast, reduced vigilance can be called fatigue [Weeß et al. 1998].

Knowledge about a user's fatigue is essential for properly conducting perceptual and behavioral studies. For example, a medical examination of the visual field is exhausting for patients as it involves repeatedly detecting and responding to hardly visible stimuli. The possible gain in accuracy and reliability with increased test duration is antagonized by increasing fatigue, effectively limiting the achievable examination quality. The rate of response errors (i.e., missed detections that were above a given perceptual threshold) increases with the duration of the measurement, experiment or examination. Although the examination time for visual field testing has been reduced over the last years [Bengtsson et al. 1997], current examination techniques still lack of a retest reliability [Bengtsson and Heijl 2008].

As we made use of campimetry in the experiments, it is important to define it. Campimetry is the examination of the visual field on a flat surface (e.g., computer screen), where lights are presented at different locations in the visual field and with different contrasts with the background surface. If a stimulus is perceived, the subject provides feedback by pressing a button [Tafaj et al. 2011]. Through this procedure, the minimum perceivable stimulus-to-background contrast is determined at each location.

In this work we focus not on user fatigue for improving behavioral studies, but for obtaining accurate estimates of vigilance that can be used to improve adaptable computer systems. As it will be described in more details later on, we chose to estimate fatigue by using eye-tracking and EEG data collected while a user performs a campimetry task, as it provides an objective performance indicator via the response error rate. However, once reliable indicators for this particular task are found, we expect it to be applicable to a broad range of similar tasks.

We show that error rate, as a proxy of subject's fatigue the error ate of a user can be used as a proxy of subject's fatigue and that it can be predicted by machine learning algorithms based on eye-tracking and EEG parameters.

The use of two different sources of indicators increases the predictor's robustness, since, poor predictions may occur when individual indicators or features cannot be acquired accurately or reliably. For example, users that use glasses could add a serious problem to eye trackers. Glasses can cause glare and may be totally opaque to light, making it difficult for a camera to monitor eye movement. Besides that, in EEG, the very small signal-to-noise ratio is an obstacle. Other common examples of such noise are cardiac signal and movement artifacts due to muscle contraction.

2. Theoretical Background

In this section, we will clarify the concepts related to fatigue, eye-tracking, EEG and machine learning that are not generally known and will be used in the sequence.

Chronotype is the tendency of an individual to sleep at a particular time during a 24-hour period.

Perimetry is the measurement of a person's field of vision.

Pupillometr is the measurement of pupil size and reactivity.

Scotopic vision is the vision of the eye under low light conditions.

Photopic vision is the vision of the eye under well-lit conditions.

Mesopic vision is a combination of photopic vision and scotopic vision in low but not quite dark lighting situations.

Heterophoria is the failure of the visual axes to remain parallel when not performing binocular fusion.

3. Related Work

Predicting fatigue is a currently and active research topic. An overview on eye-tracker-based prediction or EEG-based prediction are shows in this section.

3.1. Estimation of behavioral user state based on eye gaze and head pose — application in an e-learning environment

Estimation of an user's interest in a learning task is essential to reduce the learning time. [Asteriadis et al. 2009] uses eye-tracking features to train a neuro-fuzzy system to improve the user's experience in a e-learning environment. The user doesn't need to interact with the proposed system, using the web-site as if he/she is not been monitored. The features used are: head, eye and hand movements.

3.2. Real-time system for monitoring driver vigilance

The number of accidents caused by driver's fatigue is considerably large. Therefore, a real-time system that can detect oncoming fatigue and issue warnings could help preventing accidents. A fuzzy classifier was used by [Bergasa et al. 2006] to infer the level of a drive's fatigue. Six parameters were used: percent eye closure (PERCLOS), eye closure duration, blink frequency, nodding frequency, face position and fixed gaze.

3.3. EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate

Quality response during an auditory vigilance task via EEG also can be used as a proxy of fatigue measurement. The participants participate in a 25 hours sleep deprivation experiment. [Shen et al. 2008] used EEG features to training both probabilistic-based multi-class SVM and standard multi-class SVM, demonstrating that it is feasible to monitor mental fatigue.

3.4. EEG-Based Estimation and Classification o Mental Fatigue

[Trejo et al. 2015] uses kernel-based partial least squares to find that theta and alpha waves are highly correlated with the user's fatigue. The participants had to pass through a math task, saying that a given equation was true or false.

4. Indicators of vigilance

In this section we will discuss several possible features from which we can attempt to predict user's fatigue. The literature in this area is broad and the range of possibilities is vast. Some of the features largely used are head movement, facial expression and questionnaires on demographic data, strain, chronotype and sleep behaviour. In this work, we focus in two sources of data: eye-tracking and EEG. This choice was made because they are simpler to obtain than head movement and facial expression, and are not feedback depend as questionnaires.

4.1. Eye-tracking

Parameters of vigilance in eye-tracking data are relatively well studied. During campimetry and perimetry, researchers have focused on fatigue waves, i.e., pupillary oscillations [Henson and Emuh 2010], although a wide variety of other parameters is available and can theoretically be measured through the same device — a camera directed at the subject's eye.

The term pupillary oscillations is often found in the context of vigilance and fatigue. The sleepier a person is and the less he or she tries to suppress sleep, the shorter the time of initial pupil constriction and higher and more frequent pupillary oscillations occur. These oscillation waves show central nervous activation [Lowenstein et al. 1963]. They are known as pupillary **fatigue waves**. Henson and Emuh [Henson and Emuh 2010] have shown that oscillation waves can not only be used to detect fatigue under scotopic but also under mesopic and even low photopic conditions as they occur within visual field testing by campimetry.

Henson and Emuh showed that it is possible to monitor vigilance during campimetry under photopic conditions using pupillometry [Henson and Emuh 2010]. The equiluminant surrounding is in fact optimal to avoid pupil diameter changes due to varying illumination. In darkness, pupil size oscillations come in two flavors: slow waves of dilatation and constriction of 4 to 40 seconds duration and an amplitude of up to 0.5 mm, and superposed fast inextensive oscillations, of 0.5 to 1 second duration and with amplitudes of usually below 0.1 mm, but being able to reach up to 0.3mm [Lowenstein et al. 1963]. With decreasing vigilance the feedback loop that regulates the pupil diameter becomes unstable resulting in much larger pupillary oscillations [Wilhelm et al. 2001].

A **vergence** is the simultaneous movement of both eyes in opposite directions. Vergence angle of the eyes is connected to vigilance in a not necessarily straight-forward way. Heterophoria has been found to grow with fatigue and when performing an unfamiliar task [Kaufmann and Steffen 2012]. Some authors report that the vergence system is particularly fragile with fatigue [Vernet and Kapoula 2009].

Pupil size is determined by two neurophysiological reflexes, the pupillary light reflex regulating the amount of light entering the pupil, and accommodation, a change in the curvature of the lens.

Blink rates of healthy individuals range between 5 and 15$\times$ per minute [Barbato et al. 2000]. Blink rate and duration reflect the level of fatigue and sleepiness. Light fatigue is associated with an increase in blink frequency, sleepiness with an increase in blink duration [Schleicher et al. 2008].

4.2. EEG

Many researches have been investigating methods to predict fatigue using electroencephalography (EEG) signals in the last years. One of the most used tools is calculating the power spectrum density (PSD) of the brain waves. This transforms the signal from time to frequency domain, making it possible to analyse how its strength is distributed as a function of frequency. Many methods were developed during the years. Fast-fourier-transform, Welch's method and Burg's Method are some of them. In neuroscience, this curve is usually divided in 4 intervals, delta wave (0.5-4 Hz), theta wave (4-8 Hz), alpha wave (8-13 Hz) and beta wave (>13 Hz) [Teplan 2002]. Recent studies also define the gamma wave (>35 Hz). There is no precise agreement on the frequency ranges for each type.

Different frequency intervals provide different evidence on a person's fatigue or vigilance. [Åkerstedt et al. 1991], for instance, discovered in 1991 a relation between alpha waves and sleepness during a night shift. Furthermore [Trejo et al. 2015] showed that the mental fatigue is associated with increased power in frontal theta and parietal alpha EEG rhythms during a exhausting arithmetic task. These regions are shown in image 1. The results also indicate that EEG spectra did not correlate with fatigue outside the theta and alpha intervals. [Gharagozlou et al. 2015] and [Kong et al. 2015] demonstrate that the same effect occurs when the subjects are evaluated during a exhausting driving task.

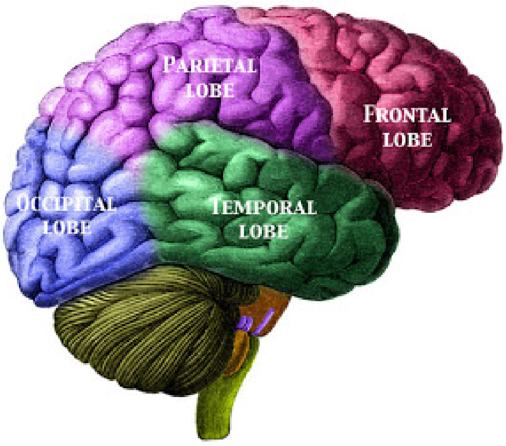


Figure 1. Brain regions.

5. Methodology

This work consists of three main parts: data gathering by means of experiments, data processing and different machine learning algorithm evaluation.

5.1. Data Gathering

Nine subjects (4 female and 5 male, aged 20-32) participated in the experiment. During the campimetric task they were asked to press a button in response to a perceived light stimulus. A central fixation cross was presented in between to mark the gaze point. Stimuli were presented at twelve different contrast levels and appeared at three screen locations (0°, 4.67°), (-4.07°, -2.34°) and (4.07°, -2.34°) and at the center (but only for around 3 % of trials to motivate looking at the fixation cross), as shown in image 2. Additionally, catch-trials were inserted, i.e., trials with a stimulus contrast so high that an attentive subject has to perceive it (positive catch-trial), or stimuli that were not visible at all (negative catch-trial). In the following, we refer to a *false negative* as a positive catch-trial that was not reacted to, and a *false positive* as a negative catch-trial (i.e., no stimulus presented), for which the participant pressed the response button.

Eye movements and pupil diameter were recorded by means of a remote eye tracking device SMI RED250 at a frequency of 250 Hz. A chin rest stabilized the position of

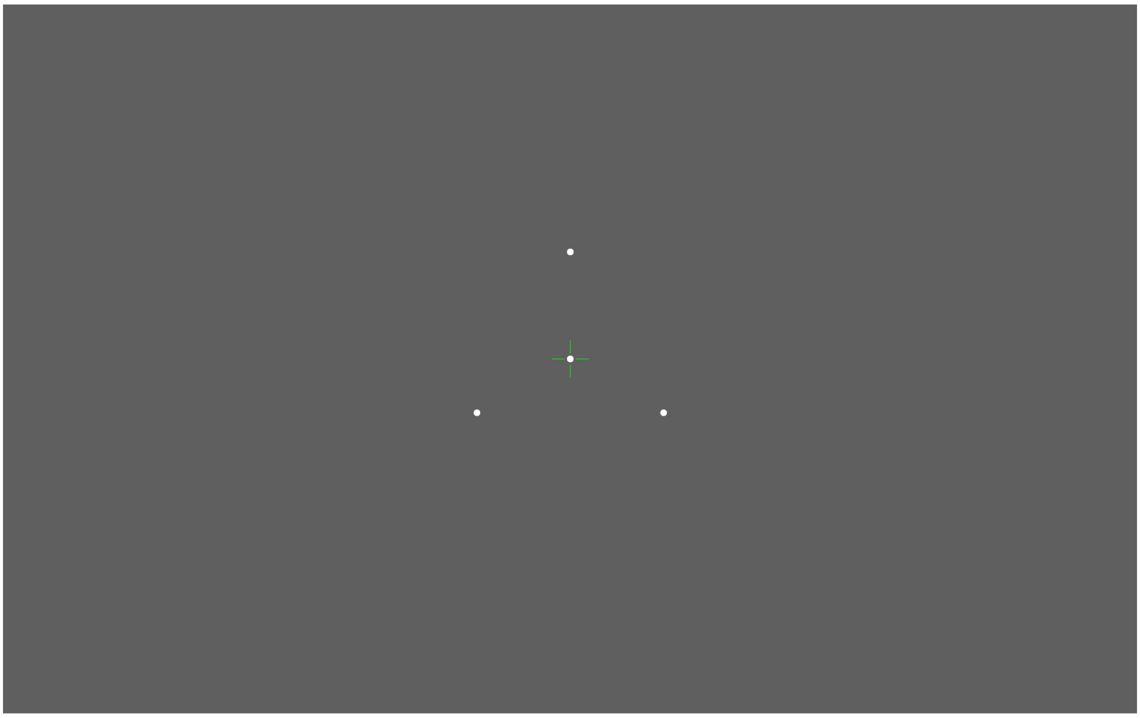


Figure 2. Stimuli placement.

the head 60 cm from the screen. Twenty-nine EEG channels were also recorded using two g.USBamp, a biosignal amplifier and acquisition system, at a frequency of 600 Hz. A standard system for electrode positioning, the 10/5 system, was used to collect the EEG signals as shown in image 3. Extra electrodes were used to collect the electrooculogram (EOG), the corneo-retinal resting eletric potential that exists between the front and the back of the eye. The synchronization was done using the timestamps offered by the eye-tracker. The resolution of the monitor was 1920x1200 pixels. In total 1,488 stimuli were presented, 25 % of which were positive and 25 % negative catch-trials. Each stimulus was shown for 200 ms followed by a 1300 ms response window. To increase the likelihood of fatigue, the experiments were conducted at two different times, 2:30 PM and 7:30 PM, and inside a dark room where the experiment screen was the only source of light.

The eye-tracking features gathered during the experiments can be seen preprocessed in the images 4, 5 and 6.

5.2. Data processing

Due to the amount of data and its nature, the input of the machine learning algorithms needed to be processed before training the model. In this section we show how the features were obtained from the raw data.

Error-rate: A moving average over the catch errors with window size of 200 was used to create the error-rate (ER). It can also be defined as the sequence of ER_i, where ER_i is as follows:

$$\mathrm{ER}_i = \frac{1}{200} \sum_{n=i}^{i+199} E_n \tag{1}$$

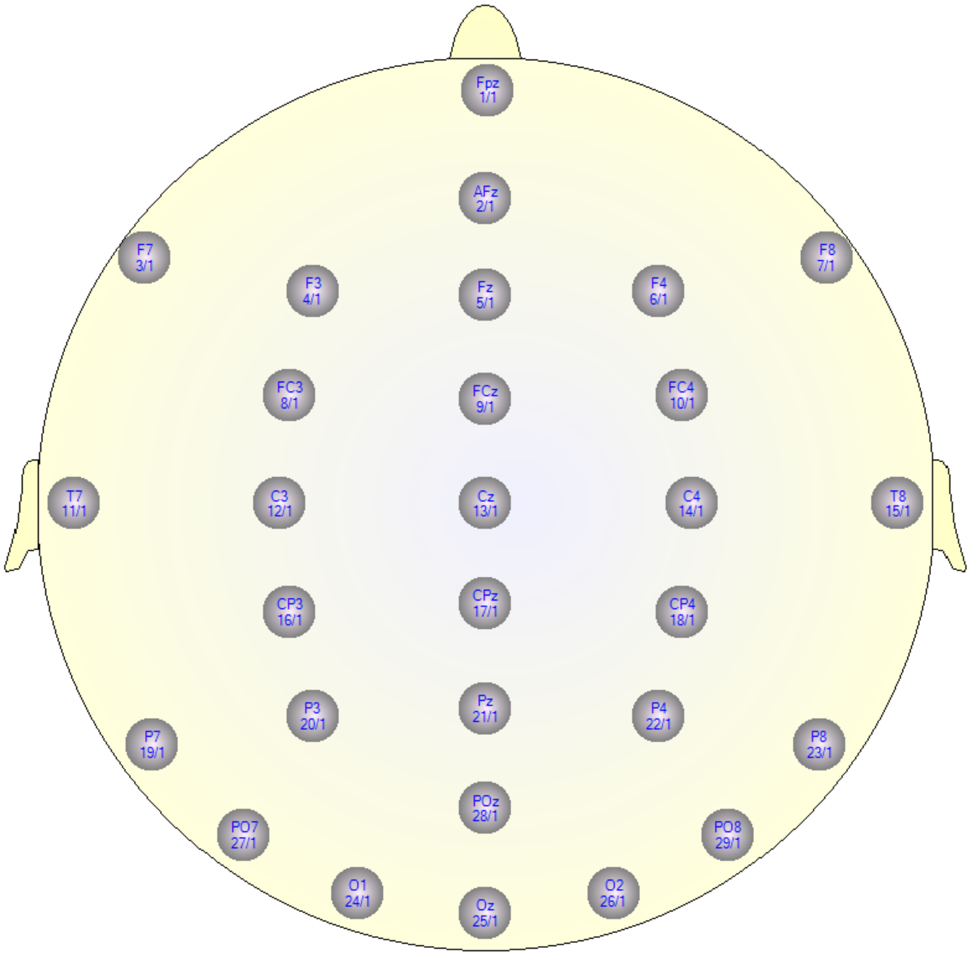


Figure 3. EGG channels placement.

where E_n is 1 if there was either a false negative or a false positive in the stimulus n, and 0 otherwise. This causes the error-rate to be a stepwise curve, making it necessary to smooth it. Other window sizes were also tested, but 200 was discovered to be the best choice.

Eye-tracking indicators: The ET indicators used are listed below:

- Blink duration
- Blink rate
- Average pupil Diameter
- Pupil variability
- Vergence
- Fatigue wave

Blink rate, blink duration, average pupil diameter and vergence were z-score normalized to the two first minutes of data as a baseline. Linear interpolation was used to fill blinks and tracking losses for all non-blink related indicators. Subsequently, vergence was filtered by a moving window filter selecting the 75%-percentile. Finally, all the eye related indicators passed through a moving average filter to create a 1-1 correspondence between error-rate and indicator. This process is described below.

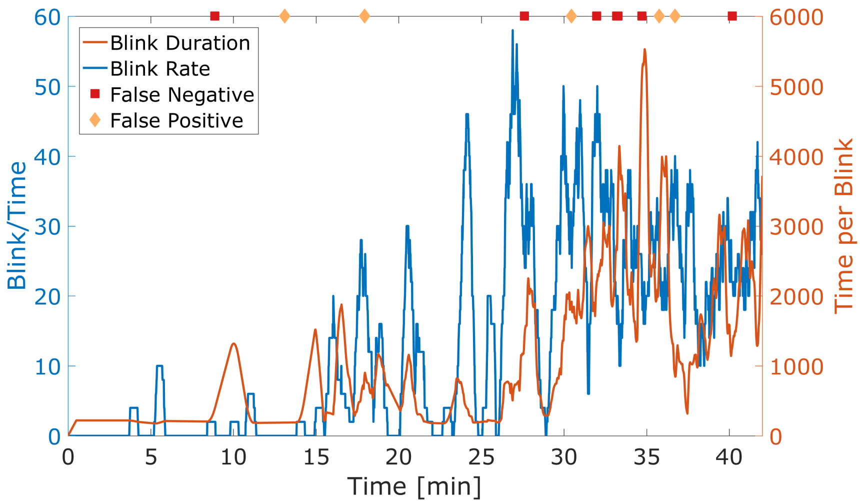


Figure 4. Blink duration and rate of subject NC.

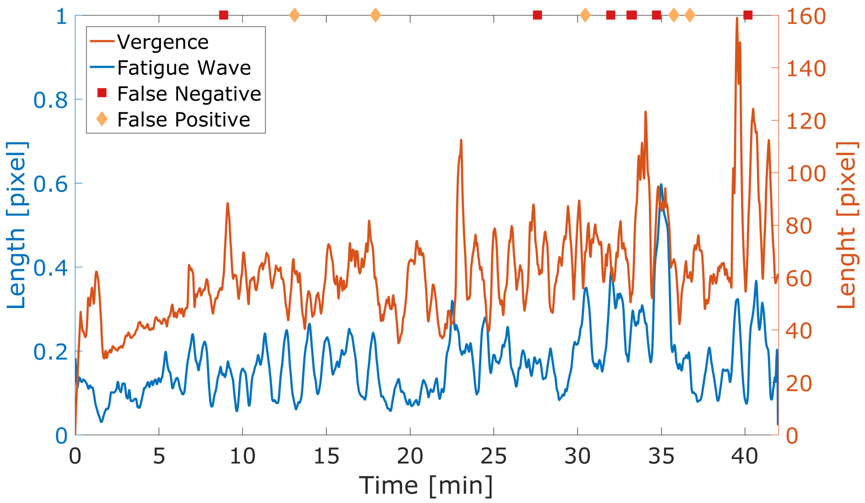


Figure 5. Vergence and fatigue waves of subject NC.

Let I be any one of the above-mentioned eye-tracking indicators. The post-processing indicator (IP) can be calculated by

$$\mathrm{IP}_i = \frac{\sum_{n=t_j}^{t_{j+199}} I_n}{M} \tag{2}$$

where t_j is the time of stimulus j, t_{j+199} is the time of stimulus i+199 and M is the number of samples in this period.

EEG indicators: The correction of EOG artifacts in the EEG channels was done as described in Schlölg et al. [Schlögl et al. 2007]. After that, the electroencephalogram waves were also z-score normalized to the two first minutes of data as a baseline. Then, from one stimulus time to the next, the power spectrum density (PSD) was calculated using the Burg's method [Burg 1975]. Only the theta (4-8 Hz) and alpha (8-

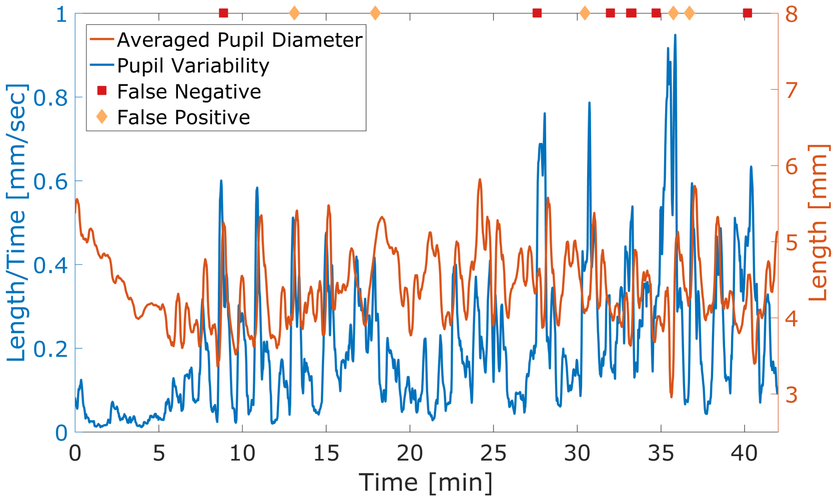


Figure 6. Average pupil diameter and pupil variability of subject NC.

13 Hz) waves were used in this study since that these are the best indicators for fatigue [Trejo et al. 2015, Gharagozlou et al. 2015, Kong et al. 2015]. Finally, they passed through the same process as the eye-tracking indicators (2).

5.3. Machine Learning Algorithms Evaluation

In this work we will try a range of different machine learning algorithms. We will evaluate these with two approaches: leave-one-subject-out cross-validation and simple cross-validation and they will be compared by means of correlation and root-mean-square error (RMSE). Some of the algorithms that will be used are listed below:

- Support Vector Regression (SVR)
- Linear Regression
- LASSO
- Neural Network
- Gaussian Processes

Until the actual date, only the leave-one-subject-out cross-validation using SVR with linear kernel was tested as a proof of concept.

6. Partial Results

This section presents the partial results obtained when training a support vector regression with linear kernel using eye-tracking and EEG features to predict the catch error rate.

Some of the subjects finished the experiment early due to abortion or technical difficulties. In this work we analyzed only the data from subjects with a complete recording. In the previous work [Dambros et al. 2017] the data of all participants was analyzed.

In the leave-one-subject-out cross-validation approach an average correlation of 0.80 ± 0.07 between the catch error rate and the prediction was obtained, that is, a 0.08 increase in the average and also a 0.1 decrease in the standard deviation compared with when only ET data was used, making the prediction of the error rate more consistent. The correlation C between two vectors with size n is calculated as follows:

$$C = \frac{\sum_n (A_n - \overline{A})(B_n - \overline{B})}{(\sum_n (A_n - \overline{A})^2)(\sum_n (B_n - \overline{B})^2)} \tag{3}$$

where $\overline{A}$ is the mean of A and $\overline{B}$ is the mean of B. The root-mean-square error RMSE between ground-truth and prediction can be calculated by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{4}$$

where $\hat{y}_i$ is the predicted value, y_i is the real value and n is the vectors size.

A more detailed results for the leave-one-subject-out cross-validation approach can be found in Table 1, where the correlation and root-mean-square error between the error rate and the prediction is shown for each participant, with and without the use of EEG features. The model used was support vector regression with a linear kernel.

	ET		**ET and EEG**	
Subject	**Correlation**	**RMSE**	**Correlation**	**RMSE**
NC	0.85	0.033	0.88	0.041
AB	0.78	0.027	0.81	0.019
PZ	0.79	0.016	0.71	0.012
LR	0.42	0.056	0.83	0.035
JD	0.78	0.018	0.76	0.015
Mean	0.72	0.030	0.80	0.024
Std	0.17	0.016	0.07	0.013

Table 1. Comparison table with correlation between catch error rate and prediction, and root-mean-square error for each subject, without and with EEG indicators.

Images 7 and 8 show the catch trial error percentage and prediction for the subjects AB and NC, respectively. One can see that they are highly correlated, but, mainly for the subject NC, it seems that the method is not able to fit completely. A drag up of the prediction would make it better, resulting in a much smaller RMSE. Another mistake that this algorithm makes is predicting negative values, it should give only values between and inclusive 0 and 1, given that the real curve is a percentage.

Table 2 shows the weight for the ET indicators of a linear support vector regression trained on data from the 5 subjects. The reason why the vergence is one of the biggest weights is that the fixation cross force a small inward deviation of the subjects eyes. As the subject starts to get tired, the muscles from the eye also become exhausted, forcing the eyes to approach the rest position. The positive and relatively substantial coefficient of blink duration indicates that the subjects reached a sleepiness state during the task.

7. Next Steps

The results presented until now show that features obtained by eye-tracking and EEG can be used for a reliable prediction of the error-rate and thereby can be used to measure response quality in a task (i.e., campimetric examination). The next step is to increase the

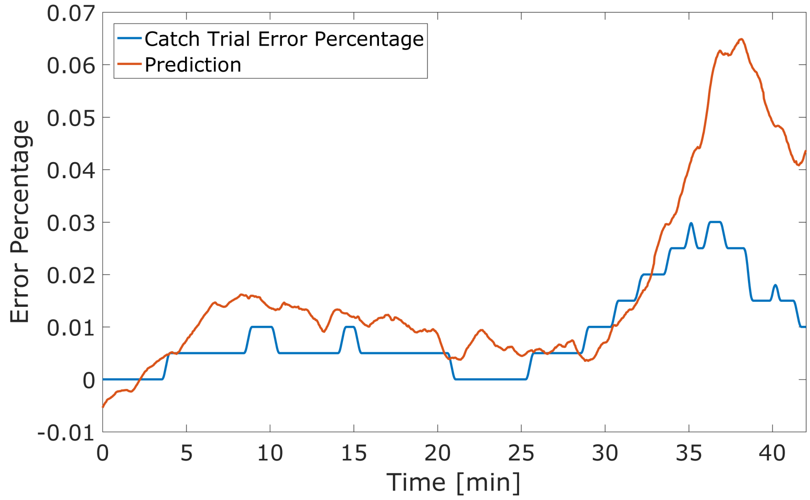


Figure 7. Catch trial error percentage and prediction for the subject AB

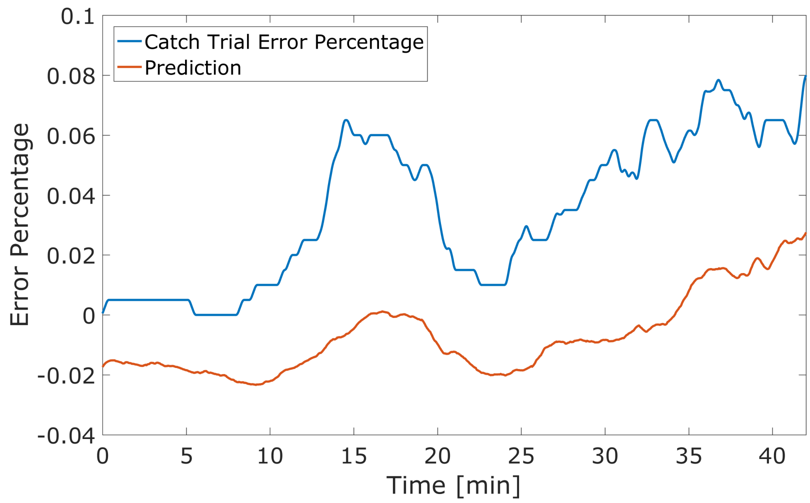


Figure 8. Catch trial error percentage and prediction for the subject NC

Feature	**Coefficient**
Blink Duration	0.0971
Blink Rate	-0.0327
Average Pupil Diameter	-0.0155
Pupil Variability	-0.0615
Vergence	0.0758
Fatigue Wave	0.0654

Table 2. Coefficients for each ET feature obtained by support vector regression with a linear kernel trained with the data from the 5 subjects. To make coefficients comparable, all features where z-score normalized.

number of machine learning algorithms used, including also non-linear algorithms. We believe that part of the RMSE obtained is due the inabilities of the linear SVR.

After that, feature selection and extraction will be used to remove useless or overlapping features. Another improvement to this work will be to add a new prediction approach. Instead of only using leave-one-subject-out cross-validation, we also want to try simple cross-validation, where the data of each subject was divided in 10 folds. From the 50 sets of data, 40 were used as training and the remaining as testing.

7.1. Schedule

In the image 9 it is shown the expected schedule. First it is given one week for each of the algorithms in order to learn about it , find an implementation and discover the best parameters that fit to this problem. Additionally to two weeks for further algorithms that may be useful.

One week should be enough to try various feature selection and feature extraction algorithms. Two weeks were chosen for adding the simple cross-validation approach. Implementation of this approach and running all the algorithms listed above are needed for comparison.

At last, seven weeks will be used for writing, correction and final changes, and two weeks to prepare the defense.

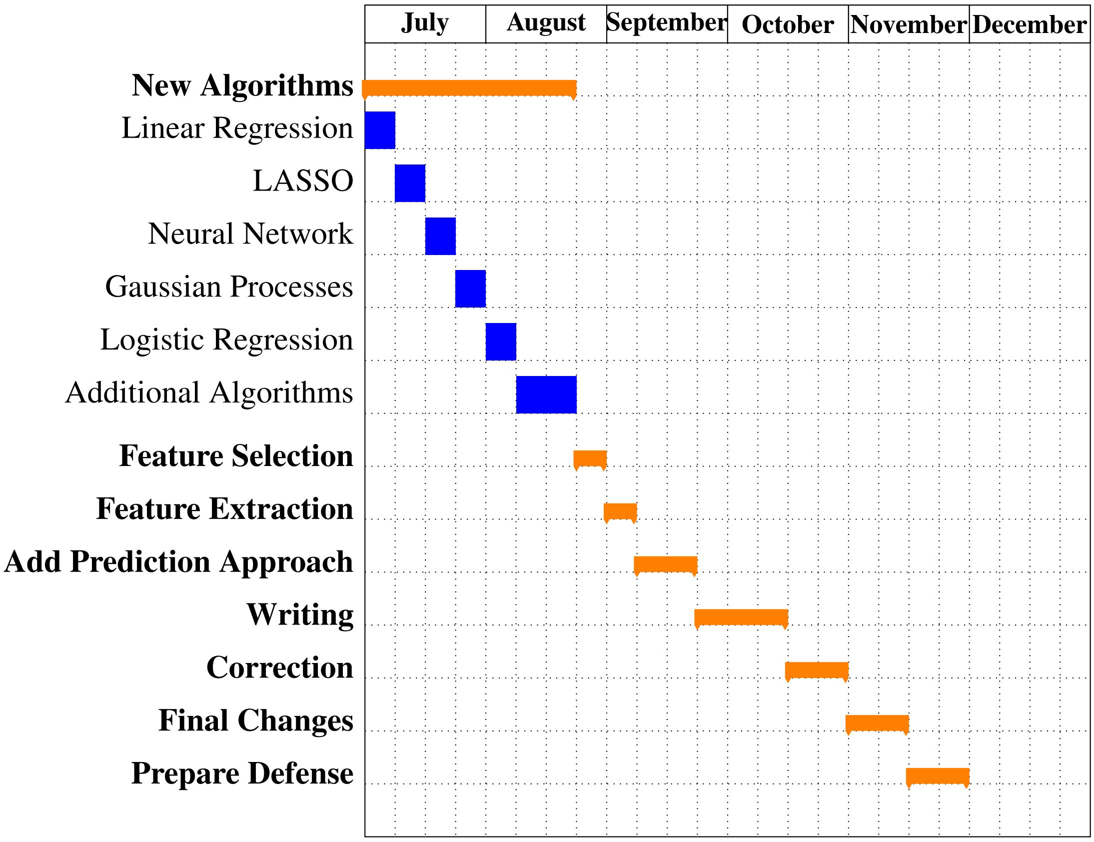


Figure 9. Expected schedule.

References

Åkerstedt, T., Kecklund, G., and Knutsson, A. (1991). Manifest sleepiness and the spectral content of the eeg during shift work. *Sleep*, 14(3):221–225.

Asteriadis, S., Tzouveli, P., Karpouzis, K., and Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. *Multimedia Tools and Applications*, 41(3):469–493.

Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., and Rinaldi, F. (2000). Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research*, 93(2):145–151.

Bengtsson, B. and Heijl, A. (2008). A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*, 145:343–353.

Bengtsson, B., Olsson, J., Heijl, A., and Rootzen, H. (1997). A new generation of algorithms for computerized threshold perimetry, sita. *Acta Ophthamol*, 75(2):368–375.

Bergasa, L. M., Nuevo, J., and Sotelo, M. A. (2006). Real-time system for monitoring driver vigilance. *IEEE Intelligent Vehicles Symposium*, 7:63–77.

Burg, J. P. (1975). *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford Universuty.

Canisius, S. and Penzel, T. (2007). Vigilance monitoring - review and practical aspects. *Biomedizinische Technik*, 52:77–82.

Dambros, G. V., Ungewiss, J., Kübler, T. C., Kasneci, E., and Spüler, M. (2017). Monitoring response quality during campimetry via eye-tracking. Association for Computing Machinery.

Gharagozlou, F., Saraji, G. N., Mazloumi, A., Nahvi, A., Nasrabadi, A. M., Foroushani, A. R., Kheradmand, A. A., Ashouri, M., and Samavati, M. (2015). Detecting driver mental fatigue based on eeg alpha power changes during simulated driving. *Sensors*, 15:1693–1700.

Head, H. (1923). The conception of nervous and mental energy. ii. vigilance: A physiological state of the nervous system. *British Journal of Psychology*, 14:125–147.

Henson, D. B. and Emuh, T. (2010). Monitoring vigilance during perimetry by using pupillography. *Investigative Ophthalmology & Visual Science*, 51(7):3540–3543.

Kaufmann, H. and Steffen, H. (2012). *Strabismus*. Georg Thieme Verlag, Stuttgart, 4th edition.

Kong, W., Lin, W., Babiloni, F., Hu, S., and Borghini, G. (2015). Investigating driver fatigue versus alertness using the granger causality network. *Iranian Journal of Public Health*, 44(12):19181—-19198.

Lowenstein, O., Feinberg, R., and Loewenfeld, I. E. (1963). Pupillary movements during acute and chronic fatigue a new test for the objective evaluation of tiredness. *Investigative Ophthalmology & Visual Science*, 2(2):138–157.

Schleicher, R., Galley, N., Briest, S., and Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982–1010.

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., and Pfurtscheller, G. (2007). A fully automated correction method of eog artifacts in eeg recordings. *Clinical Neurophysiology*, 118(1):98–104.

Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., and Wilder-Smith, E. P. (2008). Eeg-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, 119(7):1524–1533.

Tafaj, E., Kübler, T. C., Peter, J., Rosenstiel, W., and Bogdan, M. (2011). Vishnoo – an open-source software for vision research. In *24th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6. IEEE.

Teplan, M. (2002). Fundamentals of eeg measurement. *Measurement Science Review*, 2:1–11.

Trejo, L. J., Kubitz, K., Rosipal, R., Kochavi, R. L., and Montgomery, L. D. (2015). Eeg-based estimation and classification of mental fatigue. *Psychology*, 6(5):572–589.

Vernet, M. and Kapoula, Z. (2009). Binocular motor coordination during saccades and fixations while reading: a magnitude and time analysis. *Journal of Vision*, 9(7):2–2.

Weeß, H. G., Lund, R., Gresele, C., Boehning, W., Sauter, C., and Steinberg, R. (1998). Vigilanz, einschlafneigung, daueraufmerksamkeit, müdigkeit, schläfrigkeit. die messung müdigkeitsbezogener prozesse bei hypersomnien. theoretische grundlagen. *Somnologie*, 2:32–41.

Wilhelm, B., Giedke, H., Lüdtke, H., Bittner, E., Hofmann, A., and Wilhelm, H. (2001). Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. *Journal of Sleep Research*, 10(1):1–7.