EEG Based Mental Workload Assessment using Machine Learning

By

Umar Shahid

00000276613

Supervisor

Dr. Shahzad Rasool

Department of Computational Engineering

Research Center for Modelling and Simulation (RCMS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

September 2020

EEG Based Mental Workload Assessment using Machine Learning

By

Umar Shahid

00000276613

Supervisor

Dr. Shahzad Rasool

A thesis submitted in conformity with the requirements for the degree of *Master of Science* in Systems Engineering

Department of Computational Engineering

Research Center for Modelling and Simulation (RCMS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

September 2020

Declaration

I, *Umar Shahid* declare that this thesis titled "EEG Based Mental Workload Assessment using Machine Learning" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
5. I have acknowledged all main sources of help
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

Umar Shahid

00000276613

Copyright Notice

This thesis is dedicated to my parents specially to Umm-e-Umar. Without their endless support and encouragement I would never been able to complete my studies. You have very special place in my heart and I appreciate everything that you have done for me.

Acknowledgments

I want to thank my mother for supporting me spiritually and emotionally and for keeping me motivated. I will never be able to return back what you have done for me.

I wish to thank my advisor Dr. Shahzad Rasool for his guidance and support through out my studies and thesis. His comments and question were very helpful and beneficial form me. I want to thank Dr. Ammar Mushtaq and Dr. Adnan Maqsood for their valuable advice in my research and serving me as GEC. Also, I would like to Thanks principal RCMS Dr. Rizwan Riaz for his guidance sessions and motivational advises.

I want to thank my friends for being available for me every time. I always had your support, and encouragement through out my life even in time of crisis. I am the fortunate person for having you in my life. I cannot list all the names but you are always in my mind. Thank you for being in my life.

Thank you, Allah for always being there for me.

Abstract

Excessive mental workload effects mental and physical health along with the performance of individuals. There is a need to monitor the mental workload of operators performing critical tasks. Electrical signals produced by neural structures in the brain can be captured through EEG and information about mental state of an operator can be inferred. The power distribution in various frequency bands of these signals has been utilized to assess the mental workload. These noisy signals require significant filtering and preprocessing because of their low signal to noise ratio. A number of factors such as window size, filter cut-off, etc. influence the accuracy of EEG-based workload assessment. In this thesis, we analyze the performance of workload assessment pipeline with respect to these factors on an open-source workload dataset. Moreover, performance of workload assessment is analyzed using signals acquired from individual lobes of the brain instead of the entire brain and using different frequency bands instead of the entire frequency spectrum. Lastly, we also compare the performance of a number of classifiers for three level workload classification. For the preprocessing stage, a sliding window of 256 samples with an overlap of a quarter and an Artifact Subspace Reconstruction (ASR) threshold of 5 provide maximum assessment accuracy of 71.12%. Frontal and occipital lobes of the brain seem to contain the highest workload related information as they provide an average assessment accuracy of 65.83%. For frequency bands analysis, our findings validate that θ and α bands are most relevant for workload assessment as they provide 72.23% assessment accuracy which is highest among all bands. Finally, Support Vector Machines (SVM) is able to classify mental workload with an average accuracy of 66.22% which is the highest among the classifiers compared.

Contents

1 Introduction **1**

1.1 Background 1

1.2 Motivation 3

1.3 Problem Statement 3

1.4 Objectives 4

1.5 Thesis Organization 4

2 Literature Review **5**

2.1 Mental workload 5

2.2 Types of measures 5

2.2.1 Subjective measures 6

2.2.2 Objective measures 6

2.2.3 Correlation between subjective and objective measures 6

2.3 EEG 7

2.3.1 EEG signals 8

2.3.2 10-20 System International 8

2.3.3 EEG Frequencies 9

2.4 Correlation between EEG and mental workload 11

2.5 EEG based workload assessment 12

2.5.1 Recording brain activities 12
2.5.2 Data Cleaning 13
2.5.3 Frequency band extraction and labeling 14
2.5.4 Workload assessment 14
2.6 Research Gap 15

3 Methodology 16

3.1 Dataset 16
3.1.1 Data collection methods and materials 16
3.2 Workload measurement 18
3.3 Problem description 21
3.4 Methodology 22
3.4.1 Baseline Pipeline Validation 24
3.4.2 Experiment 1: Factors influencing workload assessment accuracy 25
3.4.3 Experiment 2: Region-based analysis 27
3.4.4 Experiment 3: Frequency-band analysis 29
3.4.5 Experiment 4: Classifier analysis 31
3.5 Chapter Summary 31

4 Results and Discussion 32

4.1 Validation experiment 32
4.2 Experiment 1: Analysis of parameter influence on workload assessment accuracy 33
4.2.1 Evaluation of combined effect of ASR, window size and overlap size on accuracy 34

4.2.2 Evaluation of combined effect of window size and overlap size on time 36
4.3 Evaluation Experiment 2: Region-based Analysis 38
4.4 Evaluation of Experiment 3: Frequency band analysis 40
4.5 Evaluation of different classifiers and their effect on accuracy 41

5 Conclusion 45
5.1 Conclusions 45
5.2 Limitations 46
5.3 Future directions 47

References 48

List of Figures

2.1 Emotiv EPOC+ 14-channel EEG device. 7

2.2 14-channels electrodes placement location according to 10-20 system [1] 9

2.3 Frequency bands (Delta, Theta, Alpha, Beta, Gamma) 10

2.4 Basic steps of EEG based workload assessment 13

3.1 SIMKAP test GUI. There are two segments: a comparison window and and an auditory response panel. A timer is displayed on the upper right corner [2] 17

3.2 EEG data recording pipeline. 17

3.3 Workload measurement pipeline 18

3.4 Visual description of window size and step size. 26

3.5 Individual region data extraction workflow 28

3.6 Individual band extraction flowchart 30

4.1 Heatmap of combined effect of ASR, Window and Overlap size on accuracy. 36

4.2 Effect of Window and Overlap size on time. x-axis represents the computation time and y-axis represents 9 different overlaps. 37

4.3 Effect of workload in different brain regions 39

4.4 Red circles shows the electrodes that are more responsive for mental workload. 39

4.5 Effect of Classifiers on accuracy 42

List of Tables

2.1 Frequency bands and ranges 10

3.1 Description of experiments performed 23

3.2 Different variable parameters and the values we have used 24

3.3 List of variable parameters effecting workload assessment accuracy . . . 25

4.1 Parameter values used for pipeline validation 33

4.2 Effect of different ASR values on accuracy 33

4.3 Effect of different window and overlap sizes on accuracy for ASR cut-off = 3 34

4.4 Effect of different window and overlap size on accuracy. 35

4.5 Effect of region based analysis on accuracy with 8 different windows and overlaps. 38

4.6 Analysis of individual channel 40

4.7 Workload assessment results in individual δ, θ, α, β and γ and combined θ and α frequency bands 41

4.8 Effect of region based frequency bands on accuracy for 8 different windows and overlaps. 41

4.9 Effect of different classifiers on accuracy 42

4.10 Result of experiments performed 44

Chapter 1

Introduction

This chapter is about background of this research, motivation of this research, objective of this research and organization of thesis.

1.1 Background

The use of wearable devices for monitoring our daily physiological activities along with the widespread availability of these devices has increased rapidly in the present decade. Devices such as smartwatches/gears, health and fitness bands are very popular these days. These devices have increasingly more features and embedded sensors for count steps, heart/pulse rate, etc. One aspect of the human body that can be tracked using a wearable device is the human brain activity. Electroencephalography (EEG) is one of the techniques used to observe/record it. These devices measure the electrical potential generated by activation of various neurons of the human brain. The person who first tried to record human brain activities using EEG was the German physiologist named Hans Berger in 1924 [3]. Later, EEG was adopted in clinical environment to analyze and diagnose human brain activities and the diseases such as seizure, epilepsy, stress and several types of brain disorder. In last decade, EEG was limited to clinical environment and the EEG devices were very costly and not portable while trained and expert people were needed for acquisition of EEG. However, in the present decade with the improvement of technology, the use of EEG is common in medical departments as the devices are portable, often wireless and easy to use with no professional aid required for EEG acquisition. These devices capture electrical signals from brain using electrodes. The signals captured from brain contains some information about brain activities. The meaningful information in the signals according to[3–5] is in δ, θ, α, β and γ. bands and the frequency ranges of these bands are (0 - 4 Hz), (4 - 8 Hz), (8 – 13 Hz), (13 – 30 Hz) and (> 30 Hz) respectively. Fatigue, workload, seizure, epilepsy and stress could be analyzed through the band powers ratios of these 4 bands. However, extracting meaningful information is challenging as the obtained signal is noisy therefore, a lot of preprocessing is required. EEG is not the only technique to capture brain activities, there are some other alternatives available for brain imaging such as, functional Magnetic Resonance Imaging (fMRI), functional Near Infrared Spectroscopy (fNIRS) and Magnetoencephalography (MEG) [4]. All of these are different with respect to their characteristic, their spatial and temporal resolution, their portability and their cost. fMRI devices are very expensive, are not portable, are operated in very controlled environments and are most commonly used in medical environments. fNRIS devices are portable, often wireless and easy to operate without professional aid. Its spatial resolution is very high, but its temporal resolution is very low. EEG devices are also portable, easy to operate and wireless. Its temporal resolution is very high, but its spatial resolution is low. EEG is being used for clinical as well as personal entertainment purpose. EEG device manufacturing companies often provide EEG based applications for gaming controls, mental problem-solving capability assessment and many other brain related applications. In various studies EEG is used for monitoring mental stress [5, 6], mental fatigue [7], stress level at construction sites [8] and workload [9]. In this thesis, we study the measurement of mental workload using EEG. We investigate which regions of the brain and which EEG bands contribute the most for assessment of mental workload. We analyze the performance of existing workload metrics when number of channels and EEG bands are reduced.

1.2 Motivation

EEG based workload assessment methods often use the complete spatial dimension of the brain. For example, if a 14-channel device is used then workload assessment is dependent on the acquired data of all 14 channels to retain its classification accuracy. However, the brain is subdivided into 4 lobes and each lobe is responsible for different activity. For example, the motor cortex is responsible for muscular movements, occipital lobe is responsible for eye focus, frontal and temporal lobes are responsive in stressed condition and frontal and parietal lobes are responsible for workload. So, all mental activities are related with different regions of brain. Each mental activity generates different brain signals and band powers of these signals also vary with respect to the mental activity that occurs. In this thesis, we explore the role of different regions of the brain in mental workload measurement. Such region based analysis of brain with EEG is important especially in case of workload as often workload needs to be measured in the natural environment instead of a controlled setup. Since frontal and parietal lobes are more responsive in case of mental workload, the analysis of these two regions of brain with EEG is performed in this research work. This will help in understanding the minimum number of required electrodes to reliably measure mental workload. So, if we fix only a few electrodes in some wearable device (like headphones), we would know which individual electrode must be included to give most reliable results.

1.3 Problem Statement

In this research, we evaluate the performance of EEG-based mental workload assessment techniques using the benchmark STEW dataset. Our evaluation has a number of dimensions. First, we study the effect of various parameters (window length, overlap, Artifact Subspace Reconstruction (ASR), etc.) on accuracy of workload assessment. Second, we study the role of different brain regions on workload assessment. Third, we study the role of different frequency bands on workload assessment. Lastly, we compare a number of classifiers and see which is most appropriate for the said problem.

1.4 Objectives

- To investigate the effect of each parameter in the workload estimation pipeline on the accuracy of assessment.
- To examine if we can reduce the number of electrodes used in workload estimation without compromising accuracy. If so how much reduction in number of electrodes is possible. How much accuracy do we compromise if we use only one electrode and which electrode gives maximum accuracy?
- To study if the conventional method of using frontal alpha and parietal theta for workload estimation is reasonable. What bands provide maximum information about workload?
- To compare the performance of a number of classifiers for measuring workload.

1.5 Thesis Organization

The organization of thesis is as follows. Chapter 2 reviews the literature of EEG based mental workload assessment. We first discuss mental workload, subjective and objective measure of mental workload and the correlation between both. Then, we describe Electroencephalogram (EEG), its characteristics, why it is important and the correlation between EEG and mental workload. At the end of chapter 2 we discuss EEG based workload assessment, how brain activities are recorded and how the novel aspects of our work. Chapter 3 describes the proposed methodology and implementation details. We first overview the problem and then describe the pipeline to address the underlying problem. In chapter 4, we discuss the parameters influencing workload assessment accuracy and how these parameters can effect computation time. We then compare the workload assessment accuracy from different brain regions and frequency bands. At the end, we compare the performance of different classifiers. In chapter 5, we conclude our thesis and present avenues of follow up work.

Chapter 2

Literature Review

This literature review focuses on subjective and objective measures of mental workload and the correlation between them. It covers the research that has already been carried out on EEG and brain signals and its correlation with cognitive states such as stress, workload and fatigue.

2.1 Mental workload

Every task needs some cognitive or mental resources to be utilized to complete it. The amount of cognitive or mental resources required to complete a certain task is called mental workload. Some tasks need more cognitive resources and some tasks need less cognitive resources. There is a direct relationship between required cognitive or mental resources and degree of mental workload. If degree of mental workload increases, the risk of error could also increase. So, mental workload is a very important parameter in the field of people's performance to reduce the risk of error associated with high mental workload [9].

2.2 Types of measures

There are two ways to measure mental or cognitive workload.

1. Subjective measures

2. Objective measures

2.2.1 Subjective measures

Subjective measure is used to get feedback of subjects' experience based on their feelings during the time of measurement. This includes surveys, open and close ended questionnaire, ranking and rating based on subjects' feelings. For example, NASA-TLX (NASA-Task Load Index), SURJ-TLX (Surgical-Task Load Index), RSME (Rating scale of mental effort), Bedford scale for workload assessment and SWAT (Subjective workload Assessment Technique), etc. [10]. Subjective measures are very important, but it could be difficult to comprehend without understanding the context of experience.

2.2.2 Objective measures

Objective measures are biological measures that are not dependant on examiner's judgement. These are recorded with measuring devices that work within allowable error range. These biological measures indicate the subjects' performance. For example, blood pressure measurement, ocular measurements (blink rate, pupil dilation), electrocardiac activity measurements (heart beat rate), respiration measurements (airflow and gas analysis), skin based measures (tissues blood volume, skin conductance level and response) and electrical brain activity measures, etc. [10].

2.2.3 Correlation between subjective and objective measures

Both subjective and objective measures are correlated to each other. Objective measures are simply the evidences for the subjective measures. Objective measures describe that how biological things are functioning whereas subjective measures help us to know about a subject's psychology and his feelings. For example, if a person is suffering from high workload, his EEG record will indicate how his brain is functioning whereas his subjective test will indicate the symptoms of workload.

2.3 EEG

EEG is one of the techniques used to get objective measurement of brain activity. It is the neuro-imaging technique used to record electrical activities occurring inside the brain. Different amounts of electrical activities occur in different lobes of brain like frontal lobe, parietal lobe, motor cortex, temporal lobe and occipital lobe during different tasks. To record these activities some electrodes needs to be placed on these lobes. These electrodes act as channels connected for recording with these devices. Channels receive very low amplitude data usually in micro volts (μV).

There are three types of EEG. First, invasive brain computer interface (BCI), in this technique electrodes need to be placed inside the brain lobes beneath the skull. Second, partial invasive brain computer interface in which electrodes need to be placed inside the scalp but only on the surface of the brain lobes and not inside. Third, non-invasive brain computer interface, in which electrodes are placed on the scalp. In this thesis, We focus on non-invasive brain computer interface technique as they are most common and safe for use in application other than medicine. See 14-channel non-invasive device in figure 2.1.

Figure 2.1: Emotiv EPOC+ 14-channel EEG device.

EEG is a famous neuro imaging technique that is easy to record because recent devices are portable, and it has more tolerance to subjects' movement during a recording session. Its temporal resolution is very high. It captures the dynamics of cognitive activities in the time period in which cognitive activity occurs. It has ability to record within hundred to thousands of milliseconds. It measures brain activities (Change in voltage) directly and the data recorded with EEG devices is multidimensional. Conceptually, it looks like two dimensional data. First is temporal which comes under the umbrella of change in voltage per unit time called sampling rate of EEG device and second is spatial that comes under the umbrella of space covered by the placed electrodes of the device used. But actually, it is four-dimensional data which are time, space, frequency and power [11].

EEG is not suitable for many research questions for example, where the importance of spatial resolution is high and where slow cognitive processes need to be measured. fMRI and fNIRS are more suitable for low temporal and high spatial resolution [11].

2.3.1 EEG signals

Electrical pulses are generated in the nervous system when neurons fire and the variations in potential occur at the scalp. These variations in potential called EEG signal which can be measured with an EEG device. These measured signals correspond to different lobes of the brain with respect to the event trigger. EEG signals are measured in micro volts (μV). These micro volts can be recorded using electrodes placed on the head. These micro volts are the change in electrical activity between the electrode and the reference electrode [11]. To record the electrical potentials various commercially available devices are used such as, NeuroSky MindWave Mobile 2 headset (single channel), Muse 2 Headset (4-Channel), Emotiv EPOC+ (14-Channel), etc. All these devices have some electrode channels that need to be placed on scalp according to a predefined electrode placement location system called 10-20 system international [1].

2.3.2 10-20 System International

The international 10-20 system [1, 12] is the first internationally recognized standard that is being followed for the electrodes names and placements locations. According to this system electrodes should be placed on the scalp in the way that there should be either 10% or 20% surface distance from front to back and left to right between two adjacent electrodes. The name 10-20 indicates the distance between the two adjacent electrodes. Electrode locations are associated with alphanumeric names (e.g. F4, T8, P7 etc.). The letters F, P, T, O and C represent the brain lobes frontal, parietal, temporal, occipital and central. And the (even) numbers 2, 4, 6, 8 represent the position on right sides of scalp and (odd) 1, 3, 5, 7 represent the position on left side of scalp and 'z' stands for zero that indicate the position on the mid-line of scalp. Some new alphabet prefixes are introduced in extended 10-20 system that are AF, FT, TP and FC that represent Anterior Frontal, Frontal Temporal, Temporal Parietal and Frontal Central locations respectively [13]. Figure 2.2 shows locations where the 14 electrodes of EMOTIV EPOC+ device are placed.

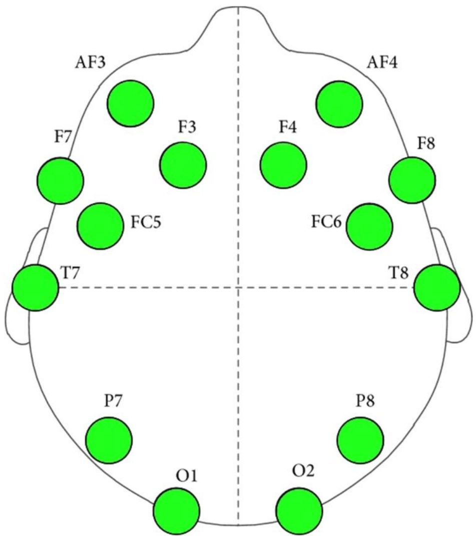


Figure 2.2: 14-channels electrodes placement location according to 10-20 system [1]

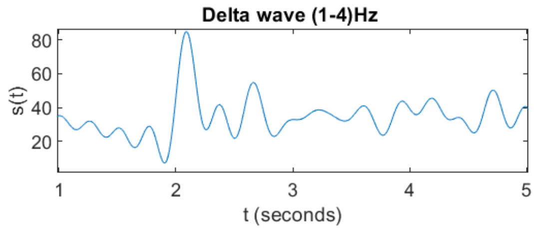
2.3.3 EEG Frequencies

EEG signals recorded from the brain are comprised of oscillations of different frequencies which are classified into frequency bands Delta, Theta, Alpha, Beta and Gamma, based on frequency ranges [14].

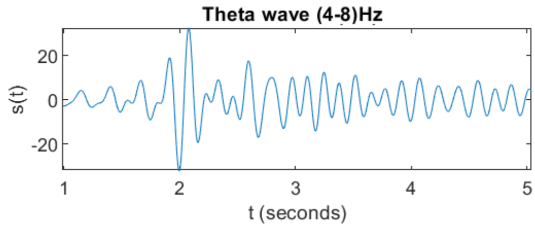
Delta wave is very slow wave and completes 1 to 4 cycles per second, see Figure 2.3a. This band influenced by external noise such as eye blinks and muscles movement and

BANDS	DELTA	THETA	ALPHA	BETA	GAMMA
Frequency Ranges	1 -4 Hz	4 - 8 Hz	8 - 13 Hz	13 - 30 Hz	> 30 Hz

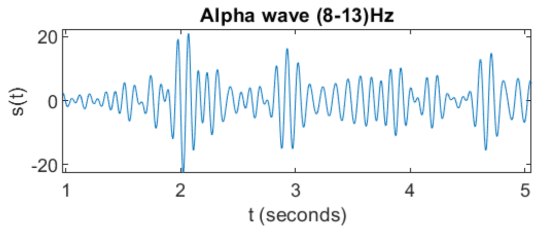
Table 2.1: Frequency bands and ranges



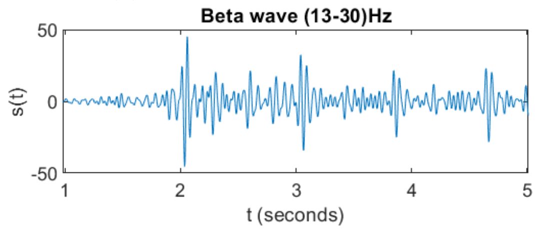
(a) Delta frequency band



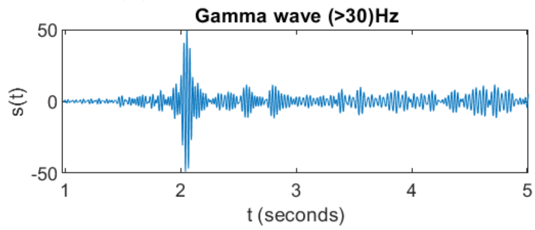
(b) Theta frequency band



(c) Alpha frequency band



(d) Beta frequency band



(e) Gamma frequency band

Figure 2.3: Frequency bands (Delta, Theta, Alpha, Beta, Gamma)

neck movement [15, 16]. Theta wave completes 4 to 8 cycles per second and it is faster than delta wave as shown Figure 2.3b. This band is related to the memory load and memory performance and and get effected during memory recall [17, 18]. Alpha wave, shown in Figure 2.3c, completes 8 to 13 cycles per second and it is faster than theta wave and correlated with relaxed state of mind [18–20]. Beta wave completes 13 to 30 cycles per second and is faster than alpha wave, see Figure 2.3d. It is associated with attention and concentration and causes stress if someone receive too much beta wave [21]. Gamma wave is a very fast wave and completes more than 30 cycles per seconds as shown in Figure 2.3e. This band is associated with high cognitive demand or high level cognitive processing and cause mental disorder or mental disease such as seizure and schizophrenia [22].

2.4 Correlation between EEG and mental workload

A number of studies have reported that mental workload can be measured/assessed with EEG [9, 10, 23]. A number of studies concluded that increase in theta and alpha activities cause the mental workload. In [14], Jap et al. reported if band power of alpha in frontal lobe and beta in parietal lobe is greater then all other bands that means the subject is fatigued. Since fatigue is a brain state induced due to high workload, so frontal alpha and parietal beta is correlated with mental workload. Hou et al. [24] reported that theta and alpha bands are correlated with mental workload because high band power in theta band indicates high memory load and high mental workload. Mu et al. [7] also reported that, when drivers' fatigue increases the power of theta and alpha bands of their brain activity increases. In [9], Lim et al. created a dataset of simultaneous task load and reported that high work loaded subjects have high theta and alpha band power in their brain activity recordings. In almost all studies people conducted time-frequency analysis of EEG signals to extract the band powers of the frequency bands. For time-frequency analysis signals needed to be transform into Fourier domain. Fourier transform gives us power of each frequency component. From this time-frequency transform power of each band can be calculated. In [25] it is claimed that increase in ratio of slow wave such as θ and α to fast wave such as β indicates increase in fatigue. Four algorithms of ratio of slow to fast wave are mentioned in [14] $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\alpha+\beta)$ and θ/β. If these increase, then mental fatigue would also increase. Moreover, increase in frontal delta (δ), frontal, temporal and occipital theta (θ), occipital alpha (α) and temporal and frontal beta (β) indicates the development of fatigue [25]. In [26] it is stated that increase in band power of frontal theta (θ) and parietal alpha (α) leads to development of mental fatigue. So different studies have different viewpoints. These four variables frontal, temporal, parietal and occipital are the lobes of brain and to acquire the brain waves from these lobes the electrodes of headset should be fixed on these specific lobes. For this purpose in various studies different headsets are used such as 32-channel headset is used in[25], 14-channel Emotiv EPOC headset is used in [9], 4-channel Emotiv Insight device is used in [27, 28] and a 26-Channel headset is used in [24, 29–31].

2.5 EEG based workload assessment

EEG based workload assessment has four basic steps [9, 30, 31] as shown in Figure 2.4. First step is recording brain activities with a measuring device as well as with subjective test so that these subjective test results can be used as ground truth. Second step is data cleaning with high pass, low pass and/or notch filters to remove the unnecessary/irrelevant data and noise. Third step is δ, θ, α, β and γ frequency band extraction with band pass filter. Last step is workload assessment after feeding the labeled data to the machine learning classifier.

2.5.1 Recording brain activities

As it is described above, to record brain activities of a subject an EEG device is required. That device has some electrodes and all those electrodes need to be placed on the scalp according the famous 10-20 system international. For better recording EEG, the environment should be calm and noiseless, and subject's movement should

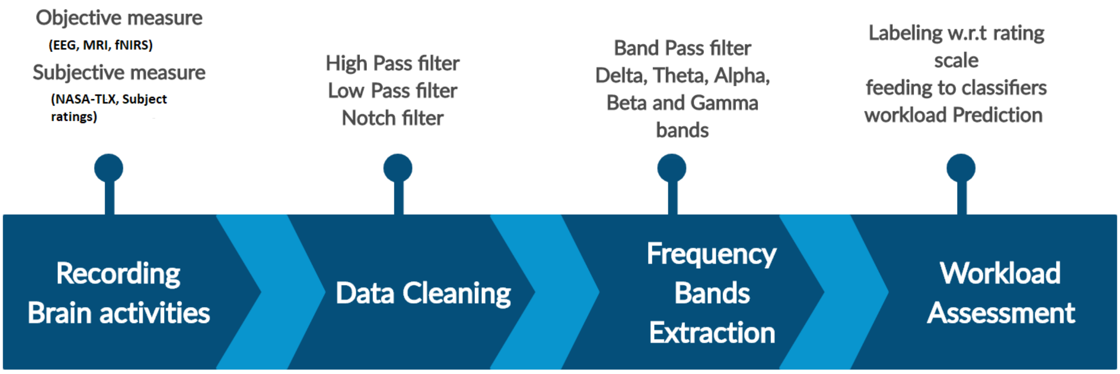


Figure 2.4: Basic steps of EEG based workload assessment

be minimal. In [25], a fatigue dataset is acquired from 52 drivers (36 males and 16 females) of age 20 to 70 years using a 32-channel headset. In [26], a fatigue dataset of 33 subjects is gathered. In [31], a stress, workload and emotion dataset of 7 subjects is acquired with a 14-channels EEG device. All these datasets are not freely available. A freely available dataset of 48 subjects for 2.5 minutes from 14-channels is collected in [9]. In [32], 6 electrodes C3, C4, P3, P4, O1, O2 and 2 references A1 and A2 are used for recording brainwaves. So, the accuracy of these systems varies with respect to the devices used, specific electrodes used, there number of participants, recording period and recording environment, etc.

2.5.2 Data Cleaning

EEG signals are noise prone and need to be cleaned first. EEG signals may have different type of noise such as, high amplitude artifacts, electrical line noise, white noise and the frequencies other than the required bands. All these effect the meaningful information in signals. So, the acquired signal needs to be cleaned with different kind of filters [9] such as a high pass filter, a low pass filter, a notch/band reject filter, etc. High pass filter is used to filter out all frequencies that are higher than a predefined threshold, low pass filter is used to filter all frequencies that are lower than the threshold and notch filter stops a specific frequency band and let all other frequencies pass through it.

2.5.3 Frequency band extraction and labeling

After cleaning EEG signals to extract frequency bands, signals need to be transformed in frequency domain and power spectrum of all bands needs to be calculated. For this purpose, power spectral density (PSD) needs to be measured. These band powers can be used as unique features of the signals [30, 31]. For feature extraction there are different methods used in different studies. In [24] Short Fourier transform is used for feature extraction. In [33] PSD Power Spectral Density, in [7] four entropies (spectrum, approximate, sample and fuzzy entropies) method is used for feature extraction purpose.

Two other factors of utmost importance are the ground truth used for labeling the data and workload induction methods. In [31] the Checklist Individual Strength (CIS) questioner is filled by subjects during the EEG recording and obtained marks from this questioner are used as ground truth of the respective recording and Psycho-Motor Vigilance Test (PVT), a cognitive test, is used to induce fatigue. In [32] Raven's Advance Progressive Metric Test (RAPM) is used to induce fatigue. In [9] a multitasking workload test called SIMKAP multitasking test is used to induce fatigue and for ground truth 1 to 9 rating of subject about his stress is used.

2.5.4 Workload assessment

As the step for workload assessment, the calculated PSD features should be fed to the classifiers to predict the workload. For prediction different types of machine learning classifiers are used in various studies and each classifier has its pros and cons. In [26], the classifiers used for features selection or training the system are kernel partial least squares linear discrimination analysis (KPLS-LDR) and kernel partial least squares support vector classifier (KPLS-SVC). In [32], Support Vector Machine (SVM), Multi layer Perceptron (MLP) Naive Bayes (NB) and K-Nearest Neighbor (K-NN) classifiers are used. In [31], Linear Support Vector Machine (LSVM), Naive Bayes (NB) and K-NN are used for classification. In [9], Near Component Analysis (NCA) and Support Vector Machine (SVM) classifiers are used for classification while [33] used Bayesian

Neural Network (BNN). For validation most of the studies used k-fold cross validation [9, 26, 32].

In two class mental fatigue classification, accuracy of class1 (low fatigue) 97.37% and class2 (high fatigue) 95.57% with 11 Hz low pass filter and class1 (low fatigue) 98.78% and class2 (high fatigue) 96.97% with 18 Hz low pass filter is reported in [26]. In [32] 93.33% , in [7] 98.75% classification accuracy with combined entropy features using SVM, in [33] 76% accuracy with open eyes and 75.3% with closed eyes with Bayesian Neural Network is achieved. In [31] 93.45% accuracy with subject dependent basis and 39.80% with subject independent basis is achieved. A dataset called STEW, is used in [9] and 69% classification accuracy is achieved after 5-fold cross validation.

2.6 Research Gap

1. In various studies different workload assessment methods and accuracy has been reported, however the analysis of various parameters effecting the accuracy and computation time has not been reported in literature. This is very important for effective use of parameters influencing the workload assessment accuracy and computation time.

2. In most of existing studies EEG data of all brain regions is used for brain state monitoring. Since, each brain state is a response of specific brain regions and specific frequency bands, workload assessment from specific regions and specific frequency bands needs to be investigated. These two aspects of workload assessment will help to isolate workload from rest of brain activities.

3. Lastly, different machine learning classifiers have been used on different datasets, but performance comparison of these classifiers for workload assessment is also missing in literature. Such a comparison will help select the best classifier for workload assessment.

Chapter 3

Methodology

In this chapter, we describe the benchmark dataset used and discuss our proposed methodology in detail.

3.1 Dataset

In this work, We have used a mental workload dataset called STEW: Simultaneous Task EEG Workload dataset [34]. This is a freely available dataset in which a benchmark test is used for workload induction. The dataset consists of pre and post workload recordings of 50 subjects. The number of subjects participated is sufficiently large. This dataset is recorded with a 14-channel EEG device called EMOTIV EPOC+. These 14 channels represents most of the brain region. Six channels cover frontal region, two channels cover temporal region, two cover parietal region, two cover occipital region and two cover the motor cortex.

3.1.1 Data collection methods and materials

For data collection Lim et al. [34] hired 50 graduate students from university and all the subjects have no neurological and brain related disease history. For workload induction they conducted a test called SIMKAP (Simultaneous Capacity) test [2]. This is a GUI based programmed psychological assessment test used to assess multitasking and stress tolerance in people. The GUI of test is shown in Figure 3.1.

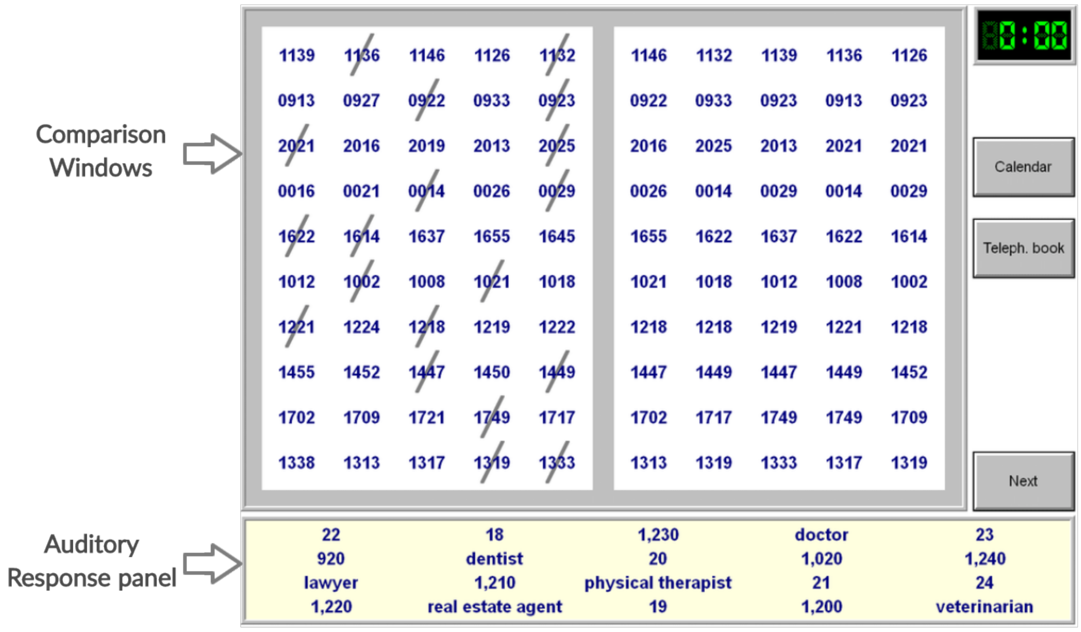


Figure 3.1: SIMKAP test GUI. There are two segments: a comparison window and and an auditory response panel. A timer is displayed on the upper right corner [2]

The GUI shows two segments of the test. First is a comparison window in which subjects are required to select the identical terms or items from one comparison window and mark that term in other comparison window. Second segment of the test is auditory response panel in which subjects are to select the answer of auditory question. Subjects are to perform both tasks simultaneously for a duration of 18 minutes.

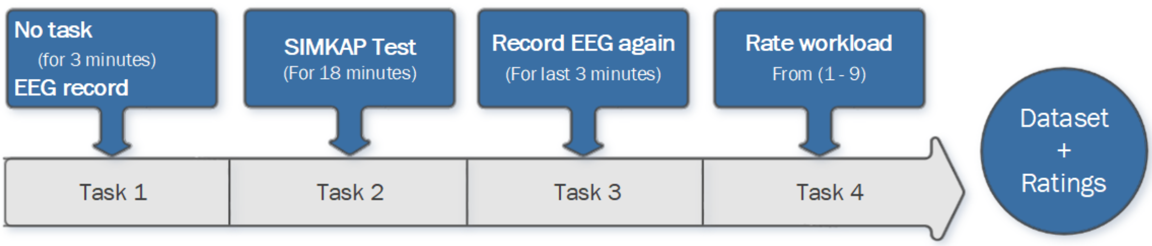


Figure 3.2: EEG data recording pipeline.

The process of data recording is shown in Figure 3.2. Firstly, the subjects are asked to perform no task for 3 minutes meanwhile EEG is recorded for those 3 minutes. Then, SIMKAP test is conducted for 18 minutes and in the last 3 minutes of SIMKAP test EEG has been recorded again for those last 3 minutes. After completion of test and recording workload rating has been taken on the rating scale of 1 - 9 from subjects. This is pipeline for one session and this pipeline has been repeated for all 50 subjects. At the end, EEG recordings as well as workload ratings of each subject are obtained [9].

3.2 Workload measurement

The acquired data is not as much usable in its raw form as it is full of noise and artifacts. A lot of preprocessing is required to rely on it. The workload measurement pipeline used is divided into four parts. A) is data collection which is described in Section 3.1, B) is signals processing, C) is classification and D) is workload assessment. All parts are further subdivided into 2 or 3 sub-parts each as illustrated in Figure 3.3.

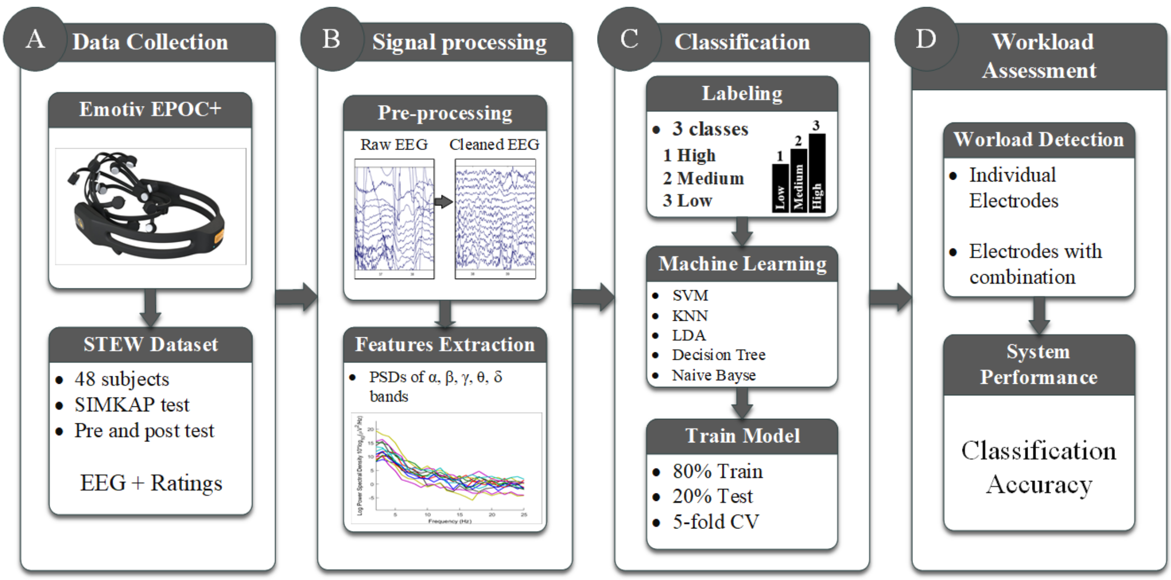


Figure 3.3: Workload measurement pipeline

A Data collection

We have used a freely available dataset called STEW - Simultaneous Task EEG Workload dataset. The data collection tools, test and procedures are already discussed in Section 3.1.

B **Signal processing**

(a) Pre-processing

i. Data filtering

The data acquired from subjects' EEG recordings, is useful in its raw form. This may have a lot of noise and unwanted artifacts. It has multiple frequencies in it but we need data only between 1 – 50 Hz frequencies. The frequencies below 1 Hz and above 50Hz are useless because the meaningful information is within five frequency bands and frequency range of these bands is from 1 to 50 Hz. So, first the data is high pass filtered on frequency 1 Hz. High pass filter stops all the frequencies below the threshold and let all frequencies pass above the threshold.

ii. Line noise cleaning

Since signals are affected by electrical line noise that occurs because of electrical equipment operation. So, line noise generates misinformation in EEG signals and this needs to be removed. Frequency of line noise is 60 Hz so for line noise cleaning data is filtered with a notch/band reject filter at 60 Hz.

iii. Artifact Subspace Reconstruction (ASR)

The very important task of this segment is artifact subspace reconstruction. This is a functionality used to remove bad channels, bad periods and high amplitude artifacts and reconstruct them. High amplitude artifacts occur because of muscular movement or sometimes if an electrode is fixed poorly. These high amplitude artifacts increase the power of signals. Removal of bad channels and bad periods changes the dimension of dataset because it removes bad channels or bad periods completely. So, we only need to remove high amplitude artifacts for getting signals cleaned.

iv. Referencing

The voltage values recorded from one electrode is relative to the voltage values recorded on other electrodes. Therefore, EEG data needs to be re-referenced. Sometimes the data is referenced with respect to some reference electrode and other times it is average re-referenced offline. In average referencing, the data of each signal is first averaged, and subtracted from the signal at all electrodes at all sampled times. Average referencing minimizes the influence of activities of all electrodes in each other.

(b) Features extraction

Since this is machine learning problem, so first we need some unique identifiers of each signal for the entire dataset. As discussed in Chapter 2, the power of delta, theta, alpha, beta and gamma frequency bands can be considered as identifiers of presence of workload, so these bands are used as the unique features of the signals. First, signals are filtered and segmented into these bands and then power of all bands of each signal is calculated. For band power, the power spectral density (PSD) is calculated. PSD is the density of power spectrum that describes how much of the total signal power a signal component is holding.

PSD is calculated with discrete Fourier transform that converts a time-domain series $x(n) = x(1), x(2), x(3) \ldots x(N)^T$ into another series in frequency domain $s(n) = s(1), s(2), s(3) \ldots (N)^T$. The equation of s is Eq. 3.1.

$$s(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{jk2\pi n}{N}}, \quad k = 0, \ldots, N-1. \tag{3.1}$$

After computing Fourier transform, power is calculates as.

$$\hat{s}(k) = \frac{1}{N} |s(k)|^2, \quad k = 0, \ldots, N-1 \tag{3.2}$$

For calculating power of each band, average power of frequencies of respective band will be calculated.

C **Classification**

(a) Data labeling

After extracting the power features the data needs to be labeled. Subjects' ratings are used as labels. Although ratings are on the rating scale from 1 – 9, we consider this as a 3-class classification problem. This means ratings from 1-3 are considered as class one i.e. low workload, ratings 4-6 are considered as class 2 i.e. medium workload and 7-9 ratings are considered as class 3 i.e. high workload. So, the data is labeled in 3 classes.

(b) Machine learning

After labeling, this labeled data will be fed to different machine learning classifiers such as SVM, KNN, NB, LDA and Decision Tree for training.

(c) Training model

Model will be trained on 5-fold cross validation that means 80% training data and 20 % test data for all 5 folds.

D **Workload assessment**

(a). Workload detection/recognition

After training of models we will predict the workload. Workload assessment will be done in all brain regions such as frontal region, parietal region, occipital region individually as well as combined regions such as frontal and parietal region, frontal and occipital region.

(b). System performance

After workload assessment we will evaluate the system's performance in terms of assessment accuracy of individual region as well as combined region analysis and computational time.

3.3 Problem description

The first step after acquisition of EEG data is cleaning data from noise and unnecessary artifacts. Then for features extraction, sliding window technique is most commonly used while taking Fourier transform. Sliding window has some window length and some step size. In previous studies, people mostly use n window length and $n * 3/4$ overlapping length for spectral analysis [9, 30]. In [8], two analysis are reported for stress detection on an EEG dataset of construction workers. In the first analysis, the authors used $n = 128$ window and 0 overlapping window length for spectral analysis and in the second analysis, $n = 128$ window length and $n-1 = 127$ overlapping window length were used. They reported that the first analysis provides better accuracy than second. The data set used consisted of 7 subjects which is very small for making a conclusion of better results. We hypothesize that, an increase in overlapping length may improve classification accuracy if the dataset is not very small. According to literature, workload related activities are more present in frontal and parietal regions of brain and θ and α frequency bands [14]. So hypothetically, use of dataset acquired from the electrodes representing these two brain region and these two frequency bands may provides better results.

3.4 Methodology

First, we implement the workload assessment pipeline and validate this baseline for two and three level mental workload assessment. Next, we evaluate the factor that effect the assessment accuracy. This experiment will focus on factors such as sliding window size, sliding window step size and artifact subspace reconstruction (ASR) threshold. To assess mental workload from each region of brain we will perform region-based analysis to find out the brain regions in which presence of information related to workload is higher. To assess mental workload from each frequency bands we will perform frequency bands analysis to find out the frequency bands in which more workload related information is present. In our last experiment, a comparison of workload classification will be done with multiple machine learning classifiers. This comparison will show the performance of various classifiers on the basis of workload assessment accuracy. All experiments are summarized in Table 4.10.

No.	Experiment Name	Aim	Independent Variable(s)	Performance Metric
0	Baseline pipeline validation	To validate the baseline pipeline weather it matches the reported results	Number of classes	Assessment accuracy
1	Factors affecting workload assessment accuracy	To select the suitable threshold of parameters	Window size Overlap size ASR threshold	Assessment accuracy Computation time
2	Region-based Analysis	To find out the brain regions that are more responsive for Workload	Brain regions	Assessment accuracy
3	Frequency bands Analysis	To find out the frequency bands that are more responsive for Workload	Frequency bands	Assessment accuracy
4	Classifiers Comparison	To analyze the performance of ML classifiers w.r.t accuracy	Classifiers	Assessment accuracy

Table 3.1: Description of experiments performed

3.4.1 Baseline Pipeline Validation

To prepare dataset for first experiment, data is perprocessed. First step in preprocessing is data filtration (data is high pass filtered at 1 Hz), second step is line noise removal (line noise is removed at 60 Hz), third step is ASR (Artifact Subspace Reconstruction), for ASR standard deviation cutoff value is used 5 and in fourth and last preprocessing step data is re-referenced at average. Output of preprocessing segment is fed to feature extraction segment. For this purpose EEGLAB toolbox's spectopo() method is used that uses fft() Fast Fourier Transform method of MATLAB signal processing toolbox. It first segments the data and then slides the window to take Fourier transform from starting value to end ending value of signal. The window size used in this experiment is 512 and step size used is 128. After this, data is segmented into 4 frequency bands that are θ, α, β and γ, then PSD (Power Spectral Density) of each band is calculated. These PSD values are considered are unique features. Some parameters we have used in this experiment are listed below, see table 3.2

NO	PARAMETERS	VALUES
1	High pass filter cutoff	1 Hz
2	Line noise removal value	60 Hz
3	ASR Standard deviation cutoff	5
4	Re-referencing value	average
5	Sliding window size	512
6	Sliding window step size	128

Table 3.2: Different variable parameters and the values we have used

Two level workload assessment

This pipeline is first tested for two level workload classification. The Level one is for no workload. All the recordings that came under first recording session (before test recording) are considered as class 1. The level two is for high workload. All the recordings that came under second recording session (after test recording) are considered as class 2. These features are then labeled and then fed to classifier for training model.

Three level workload assessment

The extracted features are then labeled in three classes with respect to the subject ratings. The recordings that are rated from 1 - 3 are labeled for class 1 and considered as low workload, the recordings that are rated from 4 - 6 are labeled for class 2 and considered as medium workload and The recordings that are rated from 7 - 9 are labeled for class 3 and considered as high workload. After labeling this, dataset fed to classifier for training model.

3.4.2 Experiment 1: Factors influencing workload assessment accuracy

Apart from the classifier used, the accuracy of the workload assessment also depends on a number of parameters. A list of such parameters that exist throughout the processing pipeline is given in Table 3.3.

NO	PARAMETERS/VARIABLES
1	High pass filter cutoff
2	ASR Standard deviation cutoff
3	Sliding window size
4	Sliding window step size

Table 3.3: List of variable parameters effecting workload assessment accuracy

ASR Cut-off

ASR is a technique used to remove bad channels, bad periods and high amplitude artifacts and reconstruct them. We limited our analysis to high amplitude artifacts removal and reconstruction only. ASR standard deviation cutoff threshold needed to be set first and ASR algorithm compares the signals' standard deviation with this threshold and reject all those periods which have standard deviation more then the decided threshold. We have used 10 different ASR standard deviation threshold values which are 0.5, 1, 2, 3, 5, 7, 9, 11, 13 and 15.

Window size

For spectral analysis of EEG data, sliding window technique is most commonly used as Fourier transform of the whole recording at a time is unable to capture variations in the electric potential. If number of samples in a window increases the estimation of spectrum will be more smooth and less fine detailed and it loses some meaningful information too. If number of samples in window will be fewer, the estimation of spectrum will be less smooth but more fine detailed and it retains maximum meaningful information. Minimum acceptable window size in EEGLAB is equal to sample rate of dataset. Sample rate of our data set is 128 samples per second so, we have analyzed four different window sizes i.e. f_s, $2f_s$, $4f_s$ and $8f_s$ where f_s is sampling frequency. For the STEW dataset $f_s = 128$ so the window sizes we applied are 128, 256, 512 and 1024 respectively. In Figure 3.4, a window size of 128 used.

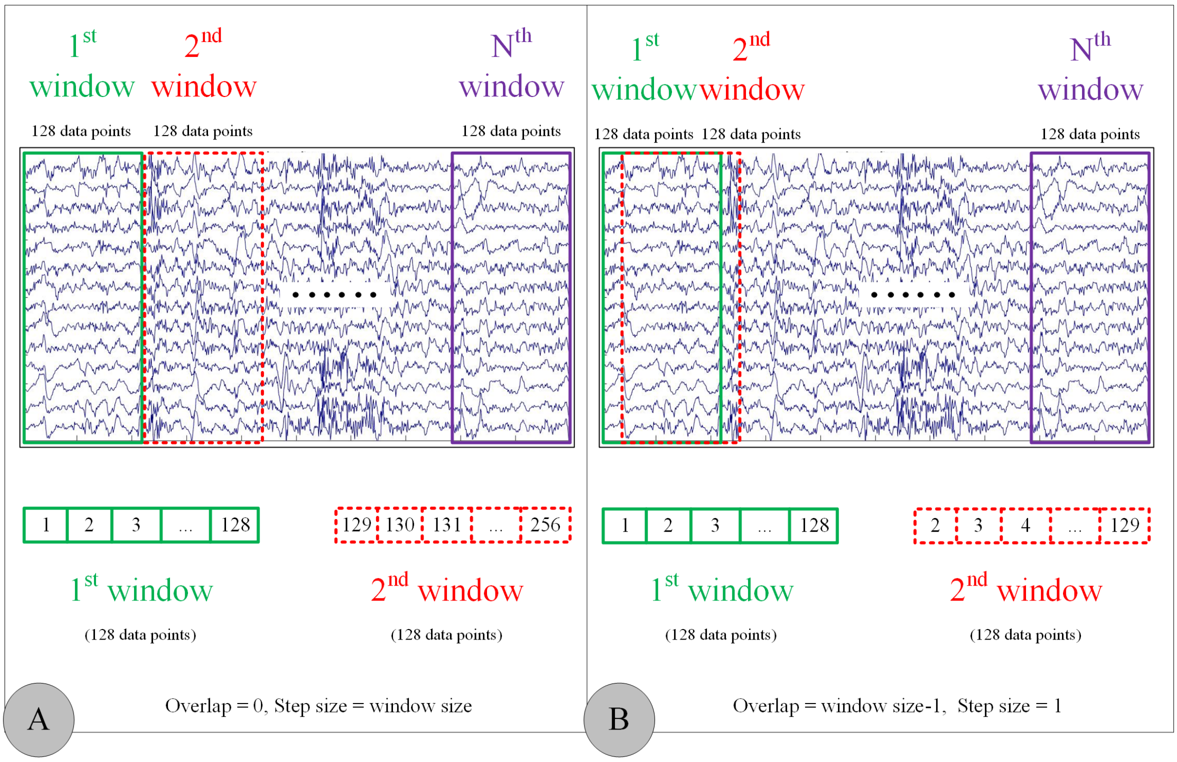


Figure 3.4: Visual description of window size and step size.

Overlap size

Overlapping technique is used to get more accurate results. Step size of the sliding window is equal to the difference of window size and overlap $Step\ Size = Window\ Size - Overlap\ Size$. Step size also effects the meaningful information. Larger step size loses more meaningful information and retains less, and smaller step size retains more meaningful information. We have used 4 different overlap sizes which are $w/4,\ 2w/4,\ 3w/4\ and\ w-1$ respectively where w is the window size. In Figure 3.4, the window size is 128 and in segment (A) overlap size is 0 and step size is 128 and in segment (B) step size is 1 and overlap size is 127.

Combined effect of window size and overlap size

We also study the combined effect of both window size and overlap size on the accuracy of workload assessment. We simulate our pipeline on four different window sizes which are $f_s,\ 2f_s,\ 4f_s\ and\ 8f_s$, where f_s is sampling frequency or sample rate. In our case f_s is 128 so the windows we used for simulation are 128, 256, 512 and 1024. Then for each window size, we set overlap size at 4 different sizes which are $w/4,\ 2w/4,\ 3w/4\ and\ w-1$ where w is the window size. Therefore, the stated experiment has 16 different combinations of window size and overlap size.

3.4.3 Experiment 2: Region-based analysis

As discussed in Chapter 2, it has been reported that some brain regions correspond to mental fatigue and workload. These regions are frontal, parietal and occipital lobes of the brain. In this experiment, we studied the effect of workload in all these regions individually as well as combined.

Analysis of individual region

First we performed analysis on the frontal region in which we used data from the frontal region electrodes (AF4, F3, F4, F7, F8, FC5 and FC6) individually. Then we labeled it according to the rating scale and fed it to the classifier for training the model. Second, we analyzed the parietal region in which we used parietal region electrodes (P7 and P8) and fed this data to the classifier after labeling. Third, we use (O1 and O2) electrodes data for training the model. The workflow for preparing dataset for individual region based analysis is illustrated in Figure 3.5.

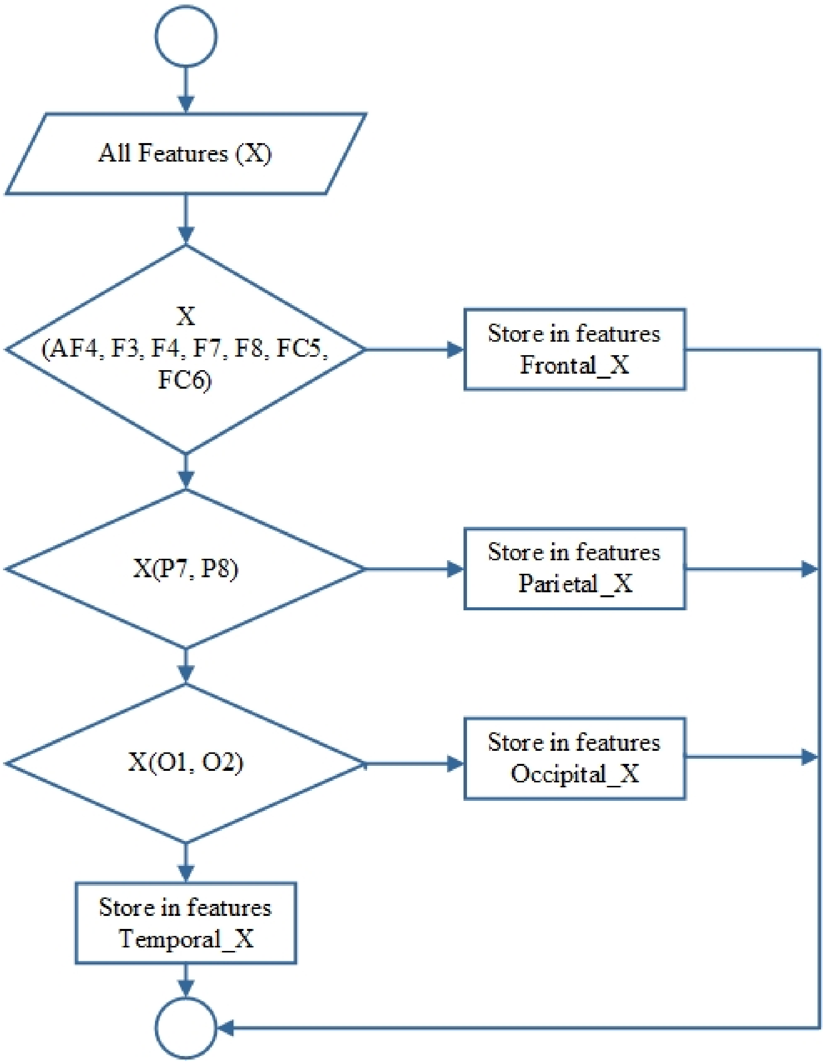


Figure 3.5: Individual region data extraction workflow

Frontal and parietal region-based analysis

According to the 10-20 electrodes placement system AF3, AF4, F3, F4, F7, F8, FC5 and FC6 electrodes of the EMOTIV EPOC+ lie on frontal region of brain and P7 and P8 lie on parietal region. So, the data recorded recorded from these electrodes is used in frontal and parietal region analysis. We imported the data of only these electrodes from overall preprocessed data and then labeled it and fed this to the classifier for training.

Frontal and occipital region-based analysis

Similarly for frontal region we used AF3, AF4, F3, F4, F7, F8, FC5 and FC6 electrodes and for occipital region we used O1 and O2 electrodes. The recorded data from these electrodes is used for frontal and occipital region analysis. From overall preprocessed data first we imported data of these electrodes and then trained a classifier on this dataset.

Analysis of individual channel

Further we have analyzed the effect of individual channel location on workload assessment. For this analysis, we trained the classifier on dataset acquired from each individual channel and we analyzed whether workload effects the electric potential at individual channel locations. Some electrodes datasets provide higher accuracy than rest of others.

3.4.4 Experiment 3: Frequency-band analysis

Since it has been reported in literature that information about mental workload is present in some frequency bands, therefore in this experiment we have studied the information in all θ $(4-8)Hz$, α $(8-13)Hz$, β $(13-30)Hz$ and γ $(>30)Hz$ frequency bands individually. The workflow for frequency band extraction is shown in Figure 3.6.

Analysis of individual frequency bands

In this part of experiment we have extracted all frequency bands individually and formed five datasets consisting of individual frequency bands. Each frequency band

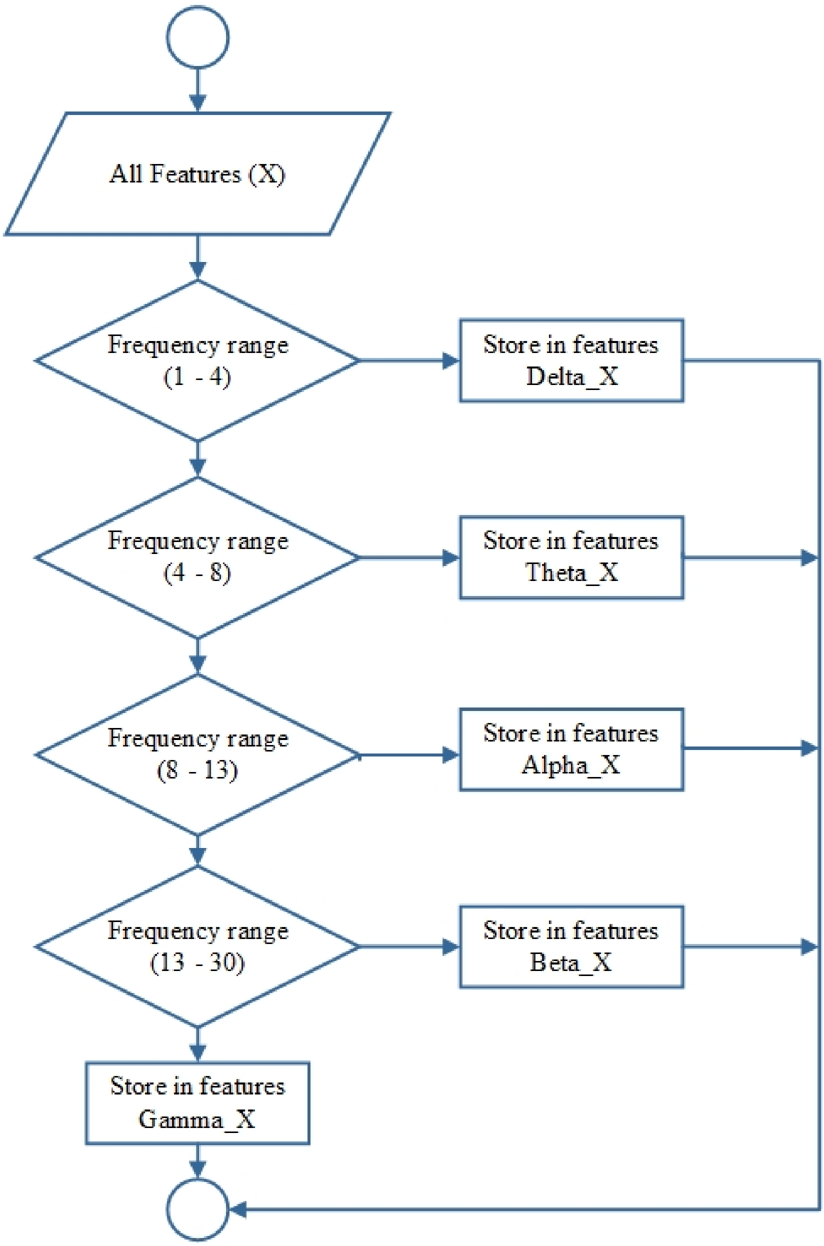


Figure 3.6: Individual band extraction flowchart

dataset is then labeled and fed to the machine learning classifier for training. Theta and Alpha band datasets provides accuracy higher than rest of three band datasets.

Analysis of combined frequency bands

From the above analysis we have noticed that prediction accuracy of model trained on theta and alpha frequency bands dataset is higher then rest of two. So, we combined these two datasets, 1) theta band dataset having frequency range (4 - 8) Hz and, 2) alpha band dataset having frequency range (8 - 13) Hz. After labeling we fed this combined dataset to the classifier for training the model.

Region based frequency-band analysis

On the basis of results from the previous experiments, we concluded that theta and alpha band are the bands that carry maximum information regarding workload so, we explore these two bands for further analysis.
First we extracted theta band as features from the data of the electrodes which are located on frontal and parietal lobe. Second we extracted theta band as features from the data of electrodes located on frontal and occipital lobe. Third we extracted two bands theta and alpha from frontal and occipital lobes. After preparing these three datasets we labeled these and fed one by one to the classifier for training models.

3.4.5 Experiment 4: Classifier analysis

There are multiple machine learning classifiers that are being commonly used in brain computer interface (BCI) applications such as, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Linear Discrimination Analysis (LDA) and Naive Bayes (NB), etc. In this simulation we applied these five classifiers on STEW dataset for 3-level workload assessment.

3.5 Chapter Summary

We have described the dataset used in this research and the workload assessment pipeline implemented in this thesis. We have also explained the methodology consisting of four experiments used to analyze the performance of workload assessment techniques. In the next chapters, we will present and discuss the results along with the conclusions that can be drawn from these results.

Chapter 4

Results and Discussion

4.1 Validation experiment

In this experiment we simulated the workload assessment pipeline for two analysis. First analysis was for two level of workload assessment in which all of the recordings in idle state i.e. before taking the SIMKAP test are considered as class 1 (no or low workload), and all recordings after taking the test are considered as class 2 (some or high workload). The parameter values we have used are displayed in Table 4.1. The classification accuracy achieved is 82.2%.

Second analysis was for three levels of workload assessment. In this analysis data is labeled with respect to the given rating scale where 1 - 3 rated data is labeled as class 1 (for low workload), 4 - 6 rated data is labeled as class 2 (for medium workload) and 7 - 9 rated data is labeled as class 3 (for high workload). We have used SVM classifier and other parameter values used are displayed in third column of Table 4.1. The classification accuracy we have achieved is 69.9% which matches with the results mentioned in [9].

NO	PARAMETERS/VARIABLES	TWO LEVEL WORKLOAD ASSESSMENT	THREE LEVEL WORKLOAD ASSESSMENT
1	High pass filter cutoff	1 Hz	1 Hz
2	Line noise removal value	60 Hz	60 Hz
3	ASR Standard deviation cutoff	3	3
4	Re-referencing value	average	average
5	Sliding window size w	512	512
6	Sliding window step size	128	128
7	Labeling Classes	2	3
8	ML classifier	SVM	SVM
9	SVM kernel	default (linear)	default (linear)
	Assessment Accuracy	82.2%	69.9%

Table 4.1: Parameter values used for pipeline validation

4.2 Experiment 1: Analysis of parameter influence on workload assessment accuracy

We analyzed the influence of various factors on the assessment accuracy. These include ASR threshold, window size and overlap size.

First, we used 10 different ASR threshold values and [512,384] window and overlaps size respectively to observe how the assessment accuracy varies. In Table 4.2 it can be seen that effect of ASR threshold on accuracy is random while the maximum accuracy achieved is 70.01% at ASR threshold of 5 and 9.

ASR VALUES	0.5	1	2	3	5	7	9	11	13	15
ACCURACY	65.56	67.79	65.56	65.56	70.01	65.56	70.01	63.34	64.45	65.56

Table 4.2: Effect of different ASR values on accuracy

Next, we simulated the system for 4 window sizes f_s, $2f_s$, $4f_s$ and $8f_s$ where f_s is sampling frequency. For the dataset we have used, f_s is 128 so the window sizes we applied are 128, 256, 512 and 1024 respectively and 4 overlap sizes $w/4,\ 2w/4,\ 3w/4\ and\ w-1$ respectively where w is the window size. Table 4.3 shows the 16 different combinations of window and overlap sizes along with the workload prediction accuracy for each. The maximum accuracy we achieved is 72.2% by using [256,256-1] window and overlap size.

The results drawn from these two analyses are inconclusive. Both experiments are

Window	Overlap	Accuracy
128	$w/4$	36.67
128	$2w/4$	34.44
128	$3w/4$	65.56
128	$w-1$	66.68
256	$w/4$	66.68
256	$2w/4$	66.68
256	$3w/4$	65.56
256	$w-1$	72.23
512	$w/4$	67.79
512	$2w/4$	68.89
512	$3w/4$	65.56
512	$w-1$	64.45
1024	$w/4$	66.68
1024	$2w/4$	64.45
1024	$3w/4$	66.68
1024	$w-1$	67.79

Table 4.3: Effect of different window and overlap sizes on accuracy for ASR cut-off = 3

regarding effect of individual parameter. To get more conclusive results we need to calculated the combined effect of parameters mentioned in these two analyses.

4.2.1 Evaluation of combined effect of ASR, window size and overlap size on accuracy

On the basis of previous two analyses, we have simulated the pipeline to analyze the combined effect of ASR threshold, window size and overlap size. In Table 4.4, it can be seen that there are 16 different combinations of window and overlap size across 10 different values of ASR threshold. We have simulated the pipeline for all 160 combinations. The maximum workload assessment accuracy of all combinations we have achieved is 72.2% on [256,256-1],[512,512-1] window and overlap sizes with 3 and 0.5 ASR thresholds respectively. And minimum workload assessment accuracy we have achieved is 34.4% on [128,128*1/4],[128,128*2/4] window and overlap sizes with 2 and 3 ASR thresholds respectively. But these results does not lead to the conclusion whether we should choose these parameter values as baseline or not.

Window	Overlap	0.5	1	2	3	5	7	9	11	13	15	Average
128	128*1/4	36.67	35.56	34.44	36.67	36.67	36.67	36.67	35.56	37.78	40.00	36.67
128	128*2/4	36.67	35.56	36.67	34.44	35.56	35.56	36.67	37.78	40.00	37.78	36.67
128	128*3/4	66.68	64.45	64.45	65.56	64.45	65.56	62.23	63.34	62.23	64.45	64.34
128	128-1	68.89	65.56	67.79	66.68	66.68	64.45	64.45	65.56	62.23	62.23	65.45
256	256*1/4	64.45	67.79	66.68	66.68	65.56	66.68	67.79	65.56	62.23	63.34	65.67
256	256*2/4	68.89	**70.00**	65.56	66.67	68.89	68.89	67.79	62.23	63.34	65.56	66.79
256	256*3/4	67.79	67.79	66.68	65.56	**71.12**	**70.00**	64.45	63.34	68.89	66.68	67.23
256	256-1	67.79	66.68	65.56	**72.23**	**70.01**	66.68	65.56	68.89	65.56	64.45	67.34
512	512*1/4	66.68	66.68	67.79	67.79	**70.01**	68.89	67.79	62.23	63.34	66.68	66.79
512	512*2/4	66.68	68.89	67.79	68.89	67.79	67.79	64.45	64.45	65.56	66.68	66.89
512	512*3/4	65.56	67.79	65.56	65.56	**70.01**	65.56	**70.01**	63.34	64.45	65.56	66.34
512	512-1	65.56	67.79	66.68	64.45	66.68	67.79	65.56	62.23	63.34	65.56	65.56
1024	1024*1/4	66.68	63.34	66.68	66.68	66.68	68.89	68.89	68.89	67.79	66.68	67.12
1024	1024*2/4	66.68	68.89	67.79	64.45	66.68	67.79	66.68	66.68	63.34	67.79	66.67
1024	1024*3/4	68.89	65.56	67.79	66.68	67.79	65.56	67.79	62.23	65.56	64.45	66.23
1024	1024-1	65.56	65.56	68.89	67.79	**70.01**	65.56	68.89	67.79	66.68	65.56	67.23
	Average	63.12	62.99	62.92	62.92	64.03	63.26	62.85	61.25	61.39	62.08	

Table 4.4: Effect of different window and overlap size on accuracy.

In the last column of Table 4.4 it can be seen that the maximum average accuracy of all ASR threshold is 67.34% achieved on window and overlap combination [256,256-1], and in last row it can be seen that maximum average accuracy of all combinations of window and overlap size is 64.03% is achieved on ASR threshold 5. In Figure 4.1, it can be observed that constellation of maximum accuracy is also in 5th column and 7th, 8th and 11th row where ASR threshold is 5 and window and overlap sizes are [256, 256*3/4], [256, 256-1] and [512,512-1]. So, on the basis of these result the most reliable window and overlap size is [256, 256-1] and ASR threshold is 5.

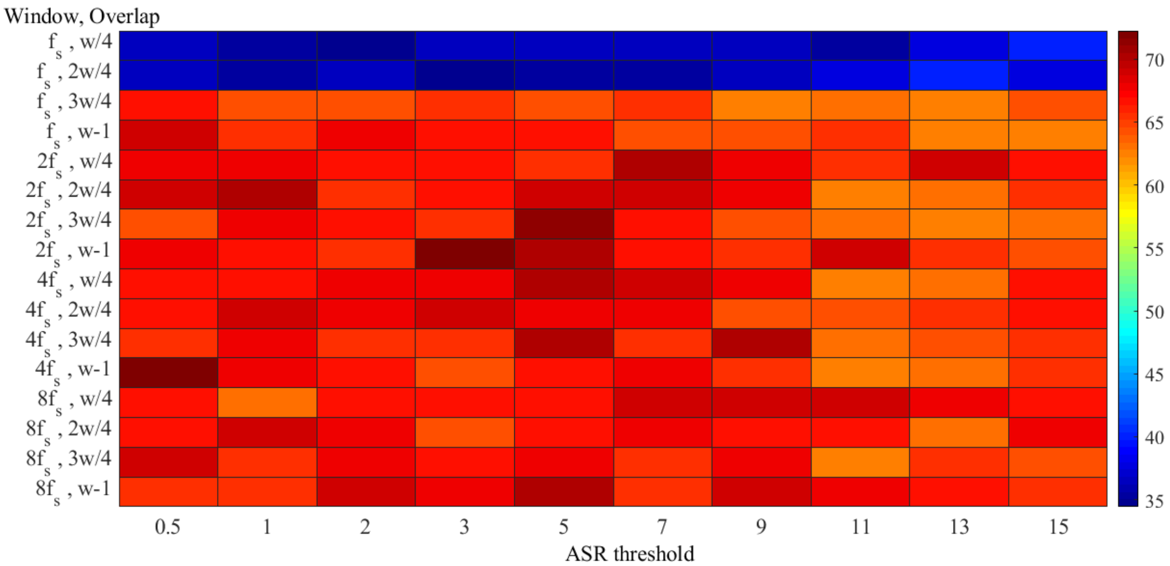


Figure 4.1: Heatmap of combined effect of ASR, Window and Overlap size on accuracy.

4.2.2 Evaluation of combined effect of window size and overlap size on time

In this experiment we have calculated the computation time for preprocessing for four different window size and 9 different overlap sizes. These window sizes are f_s, $2f_s$, $4f_s$ and $8f_s$ and overlap sizes are $w/4$, $2w/4$, $3w/4$, $w-2^5$, $w-2^4$, $w-2^3$, $w-2^2$, $w-2^1$ and $w-2^0$ respectively.

Smaller window sizes took less time to process one window but took more time to process all data because if we decrease the size of window, the number of windows to be processed increase while if we increase the size of window the number of windows to be processed decrease. Therefore, there is an inverse relation between the size of window and number of windows to be processed. Moreover, an inverse relation between size of window and time exists while there is a direct relation between number of windows to be processed and time, see equation 4.1.

$$
\begin{aligned}
w &\propto \frac{1}{no.of windows} \\
w &\propto \frac{1}{time} \\
no.of windows &\propto time
\end{aligned}
\tag{4.1}
$$

In Figure 4.2, the x-axis shows overlap size and y-axis shows the computation time in hours, for 9 different overlap sizes and curve shows the mean time of 4 windows. As expected, with an increase in the window and overlap sizes the computation time also increases.

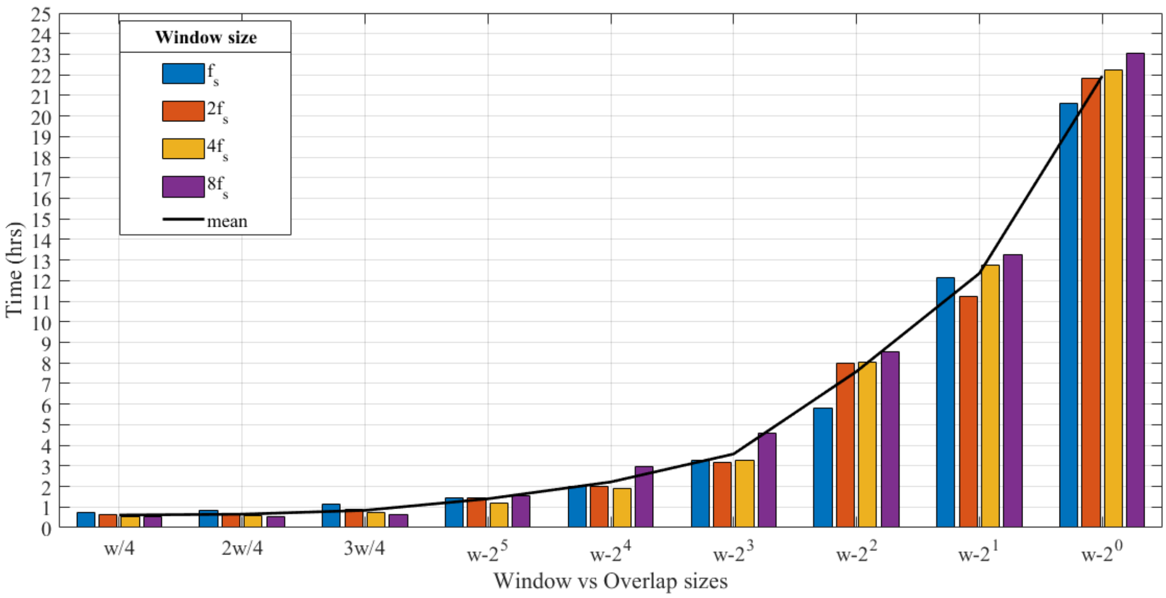


Figure 4.2: Effect of Window and Overlap size on time. x-axis represents the computation time and y-axis represents 9 different overlaps.

4.3 Evaluation Experiment 2: Region-based Analysis

In this experiment we have extracted features from frontal region (AF3, AF4, F3, F4, F7, F8), parietal region (P7, P8), occipital region (O1, O2), together from both frontal and parietal regions (AF3, AF4, F3, F4, F7, F8, P7, P8) and then together from both frontal and occipital regions' (AF3, AF4, F3, F4, F7, F8, O1, O2) electrodes. We performed workload assessment using these datasets with 4 window and 2 overlap sizes and an ASR threshold value of 5.

In Table 4.5, it can be seen that maximum average workload classification accuracy achieved in frontal region is 65.69%, in parietal region is 50.41%, in occipital region is 56.38%, in combined frontal and parietal region is 64.58% and in frontal and occipital region is 65.83%.

WINDOW	OVERLAP	FRONTAL ACCURACY	PARIETAL ACCURACY	OCCIPITAL ACCURACY	F AND P ACCURACY	F AND O ACCURACY
128	$3w/4$	66.68	50.00	56.67	64.45	64.45
128	w-1	65.56	50.00	54.45	63.34	64.45
256	$3w/4$	66.68	51.12	55.56	65.56	65.56
256	w-1	64.45	51.12	56.67	65.56	65.56
512	$3w/4$	65.56	50.00	58.89	63.34	66.68
512	w-1	66.68	51.12	55.56	65.56	68.89
1024	$3w/4$	65.56	51.12	56.67	64.45	65.56
1024	w-1	64.45	48.89	56.67	64.45	65.56
	Average	65.69	50.41	56.38	64.58	65.83

Table 4.5: Effect of region based analysis on accuracy with 8 different windows and overlaps.

Figure 4.3 illustrates that maximum information about workload activities is present in frontal and occipital regions and minimum in parietal region. Our experiment results have contradict with prior studies in which frontal and parietal regions are reported for workload.

According to the results it can be concluded that frontal and occipital regions are more responsive for mental workload. Our accuracy is hovering between 60-70% by using less number of electrodes where as 69% accuracy is reported in baseline paper [9]. So, by using less number of electrodes our results are comparable.

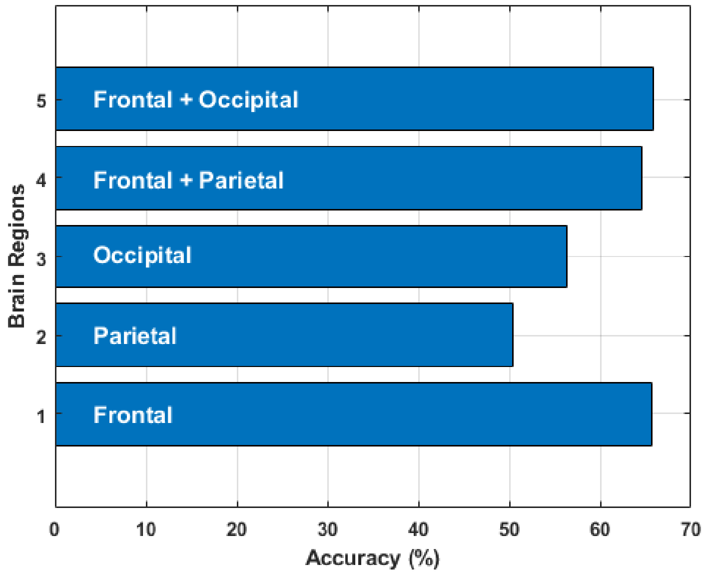


Figure 4.3: Effect of workload in different brain regions

Next, we analyse the results of individual channels with 2 different window and overlap sizes. In Figure 4.4, it can be seen that there are 6 electrodes that are red circled. These 6 electrodes provide maximum accuracy, see in Table 4.6. Four of these electrodes are located on frontal lobe and 2 are located on occipital lobe.

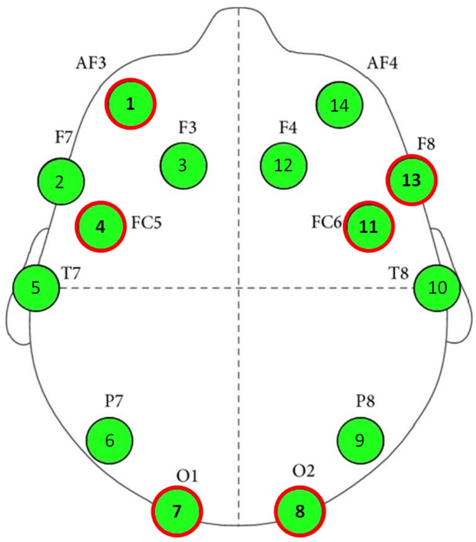


Figure 4.4: Red circles shows the electrodes that are more responsive for mental workload.

Electrodes	$2f_s$, $3w/4$	$2f_s$, w-1	$4f_s$, $3w/4$	$8f_s$, w-1	Average
AF3	50.0000	52.2222	48.8889	54.4444	**51.3889**
F7	47.7778	48.8889	47.7778	50.0000	48.6111
F3	47.7778	48.8889	47.7778	47.7778	48.0556
FC5	53.3333	52.2222	53.3333	48.8889	**51.9444**
T7	47.7778	48.8889	47.7778	48.8889	48.3333
P7	46.6667	46.6667	47.7778	47.7778	47.2222
O1	51.1111	50.0000	51.1111	51.1111	**50.8333**
O2	51.1111	50.0000	52.2222	51.1111	**51.1111**
P8	46.6667	46.6667	46.6667	46.6667	46.6667
T8	47.7778	47.7778	47.7778	48.8889	48.0556
FC6	57.7778	55.5556	57.7778	51.1111	**55.5556**
F4	48.8889	47.7778	50.0000	46.6667	48.3333
F8	55.5556	54.4444	54.4444	56.6667	**55.2778**
AF4	46.6667	48.8889	47.7778	48.8889	48.0556

Table 4.6: Analysis of individual channel

4.4 Evaluation of Experiment 3: Frequency band analysis

We further evaluated the performance of workload assessment for individual frequency bands. In this experiment we have prepared six datasets in which first dataset has only δ band powers as unique features, second dataset has θ band powers as unique features, third dataset has α band powers as unique features, fourth dataset has β band powers as unique features, fifth dataset has γ band powers as unique features and sixth dataset has θ and α band powers as unique features. These dataset are then fed to the classifiers one by one for training the models.

In Table 4.7, it can be seen that the maximum assessment accuracy is achieved from individual theta band dataset and combined theta and alpha band dataset. Theta band dataset accuracy is 71.1% and combined theta and alpha bands dataset accuracy is 72.2%. So, it can be concluded that workload related information is present more in theta and alpha bands than other bands.

Next, we have analyzes the pipeline for brain regions based frequency bands. In Table 4.8, average workload assessment accuracy for frontal α and parietal θ is 58.47%, for frontal α and occipital θ is 58.89%, for frontal and parietal α , θ is 59.58% and for

Bands	delta δ	theta θ	alpha α	beta β	gamma γ	theta θ and alpha α
Accuracy	55.56	71.12	67.79	55.56	51.12	72.23

Table 4.7: Workload assessment results in individual δ, θ, α, β and γ and combined θ and α frequency bands

frontal and occipital α , θ is 60.28%. These results show that workload related activities are mostly present in frontal and occipital region of brain and both theta and alpha frequency bands.

Window	Overlap	Frontal α Parietal θ	Frontal α Occipital θ	F and P $\alpha + \theta$	F and O $\alpha + \theta$
128	128*3/4	57.78	58.89	58.89	58.89
128	128-1	57.78	58.89	58.89	60.00
256	256*3/4	58.89	58.89	58.89	61.12
256	256-1	57.78	58.89	60.00	60.00
512	512*3/4	58.89	58.89	61.12	62.23
512	512-1	58.89	60.00	61.12	58.89
1024	1024*3/4	58.89	57.78	58.89	60.00
1024	1024-1	58.89	58.89	58.89	61.12
	Average Accuracy	58.47	58.89	59.58	60.28

Table 4.8: Effect of region based frequency bands on accuracy for 8 different windows and overlaps.

4.5 Evaluation of different classifiers and their effect on accuracy

Further we examine different machine learning classifiers to evaluate the performance of these classifiers for mental workload assessment for three levels of workload classification.

Figure 4.5 shows 5 different classifiers SVM, Decision Tree, KNN, LDA and Naive Bayes and there assessment accuracy.

In Table 4.9, all 5 classifiers are trained on 10 different ASR threshold values for [512,512-1] window and overlap size. Average assessment accuracy of 66.22% is achieved with SVM, 59.78% with decision tree, 58.22% with KNN where $k = 4$, 52.33% with

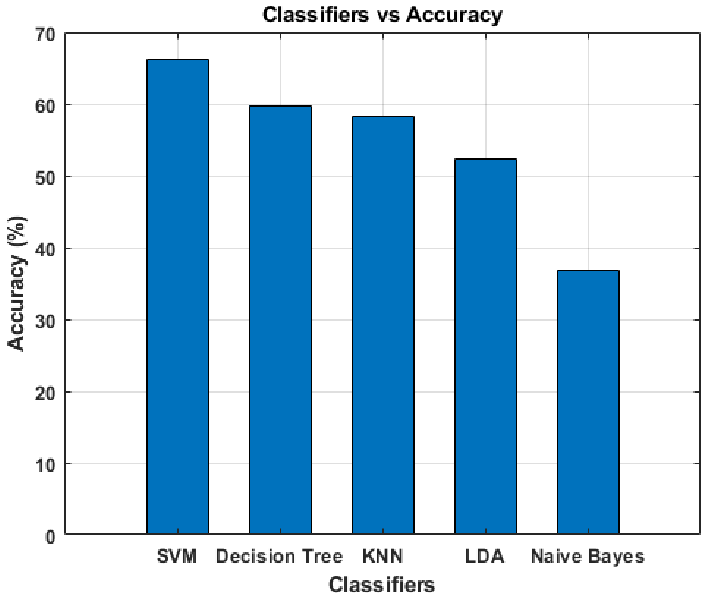


Figure 4.5: Effect of Classifiers on accuracy

LDA and 36.89% with Naive Bayes. SVM provides maximum average accuracy among these classifiers.

ASR Threshold	SVM	Decision tree	KNN	LDA	Naive Bayes
0.5	67.79	64.44	63.33	58.89	36.67
1	67.79	62.23	54.45	53.34	36.67
2	66.68	61.12	57.78	48.89	35.56
3	64.45	58.89	60.00	51.12	34.44
5	72.23	60.00	62.23	51.12	36.67
7	66.68	61.12	61.12	51.12	34.44
9	65.56	57.78	58.89	51.12	36.67
11	62.23	57.78	55.56	51.12	38.89
13	63.34	57.78	55.56	53.34	40.00
15	65.56	56.67	53.34	53.34	38.89
Average	66.22	59.78	58.22	52.33	36.89

Table 4.9: Effect of different classifiers on accuracy

SVM employs large margin hyperplane technique. Conventional classifiers that are based on hyperplane such as, LDA, require the training examples lie on the right side of the hyperplane and find optimal value of θ and θ_0 so that $\theta^T x_i + \theta_0 > 0$ if belongs to positive class and $\theta^T x_i + \theta_0 < 0$ if belongs to negative class, where θ is an n-dimensional *Weight vector* of feature space and θ_0 is a *Bias term*. But, SVM require the training examples lie on right side as well as it require *Safety margin* for better generalization capability. The objective function is $J(\theta) = \dfrac{1}{2}||\theta||$, for maximization of margin, $||\theta||$ needs to be minimized. SVM also allows misclassification for non-linearly separable data i.e. point lie on wrong side or within the margin. A *Slack variable* ξ_i is introduced which is set to the distance of x_i from the decision boundary (hyperplane) for this kind of scenarios. Slack variables summed up in objective function as *Penalty factor*. So, our objective function looks like $J(\theta) = \dfrac{1}{2}||\theta|| + C(\Sigma_{i=1}^{N}\xi_i)$ where C is regularization parameter [35–38].

In case of EEG signals it is more likely to occur misclassification because of several reasons such as, poor placement of electrodes, subjective assessment bias and difference of ability to withstand workload in subjects. Because of such reasons the dataset may not linearly separable and training examples of one class may lie in other class. So, SVM is more suitable in this case and provides good results than rest of the classifiers.

No.	Experiment Name	Aim	Result
0	Baseline pipeline validation	To validate the baseline pipeline weather it matches the reported results	Matches with literature
1	Factors affecting workload assessment accuracy	To select the suitable threshold of parameters	Window size = 2 sec Overlap size = 1.5 sec ASR threshold = 3-7
2	Region-based Analysis	To find out the brain regions that are more responsive for Workload	Frontal + Occipital
3	Frequency bands Analysis	To find out the frequency bands that are more responsive for Workload	Theta (θ) + $Alpha(\alpha)$
4	Classifiers Comparison	To analyze the performance of ML classifiers w.r.t accuracy	SVM

Table 4.10: Result of experiments performed

Chapter 5

Conclusion

5.1 Conclusions

Review of existing literature on EEG-based mental workload assessment using machine learning has shown that an analysis of the factors that influence workload assessment accuracy is needed. Moreover, the effect of different regions and frequency bands on workload is required. Finally, a comparison of different machine leaning classifiers also needs to be performed.

The recent research of EEG-based mental workload assessment present two and three level workload assessment with different assessment accuracy [8, 9, 24, 31]. In order to analyze the effect of different factors that influence of workload assessment methods, we investigated various parameters such as sliding window size, sliding window step size and Artifact subspace reconstruction (ASR) threshold. We have analyzed individual as well as combined effect of these parameters. While analyzing effect of individual parameters, results are inconclusive where as analysis of combined effect of these parameters show that most suitable window and step sizes are 256 and 256*3/4 respectively. Moreover, an ASR threshold of 5 provides maximum average accuracy of 64.03%.

In previous studies, different regions of brain are theoretically reported for different mental states such as, frontal and parietal regions are more responsive for fatigue and mental workload. To validate this we have analyzed individual region of brain for presence of workload related activities. Our investigation contradicts prior studies in which frontal and parietal regions are reported for mental workload. In our results, frontal and occipital region provides maximum average accuracy of 65.83% which is higher than rest of other brain regions. We have evaluated the response of individual channel and the best performance is obtained from the channels that are located on frontal and occipital region.

Theta θ and Alpha α frequency bands are reported to have maximum workload related information. We have analyzed the presence of mental workload activities in each frequency band. Our results indicate that indeed maximum mental workload related information is present in theta θ and alpha α frequency bands.

Lastly, we have compared the performance of various machine learning classifiers with respect to workload assessment accuracy. Our results indicate that SVM classifier performs best among SVM, Decision Tree, kNN, LDA and Naive Bayes, with maximum average assessment accuracy of 66.22%.

5.2 Limitations

In this study we have used workload dataset called STWE:(Simultaneous task EEG workload dataset). The workload is induced in this dataset with SIMKAP test which is based on simultaneous tasks so, it may not be used for single task based workload detection. Single task based workload can better be detected with some individual task based workload dataset. In subjective rating, there may be self assessment bias because, a subject cannot correctly assess himself, which level of workload is he suffering from. The current implementation is for offline data processing so, it may not be used for real-time workload detection in its present form. We may use this implementation for real-time workload detection after making some changes. Further more, the employed dataset is skewed and all 3 classes do not have equal number of examples (data samples), class 1 holds 42 data samples, class 2 holds 23 data samples and class 3 holds 25 data samples so, the assessment accuracy may not be same of all classes.

5.3 Future directions

In our current research we have worked on EEG based workload detection based on an offline dataset in which workload is induced with a benchmark SIMKAP test. We foresee future directions for real-time workload detection using the proposed pipeline.

Next, the real-time workload detection pipeline can be validated for tasks carried out in the natural environment. This will allow for estimation of mental workload of operators performing critical tasks in the context of their working environment.

Real-time monitoring is more challenging than offline monitoring. Since the proposed pipeline is for offline processing and may not be used for real-time workload detection in its present form. For real-time processing we need to record EEG activity for at least equal to the length of STEW dataset recording time which is 2.5 minutes. This 2.5 minutes data requires some time for processing and cannot be processed as a whole so, we may employ parallel processing for faster data processing. We can also use this implementation in real-world environment. Real-world scenarios are more sensitive to external noise such as, muscular and head movement, eye blinks etc. So, the real-world monitoring may require a lot more processing for noise reduction.

References

[1] Richard W Homan, John Herman, and Phillip Purdy. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4):376–382, 1987.

[2] O Bratfisch and E Hagman. Simultankapazität/multi-tasking (simkap) version 24.00: Handanweisung (simultaneous capacity/multi-tasking (simkap) release 24.00: Manual). *Mödling, Austria: Schuhfried*, 2003.

[3] T. J. La Vaque. The history of eeg hans berger: psychophysiologist. a historical vignette. *Journal of Neurotherapy*, 3(2):1–9, 1999.

[4] Z. Lan. *EEG-based emotion recognition using machine learning techniques*. Doctoral thesis, Nanyang Technological University, Singapore, 2018.

[5] Fares Al-Shargie, Tong Boon Tang, and Masashi Kiguchi. Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: an fNIRS-EEG study. *Biomedical Optics Express*, 8(5):2583, 2017. ISSN 2156-7085. doi: 10.1364/boe.8.002583.

[6] Sami Elzeiny and Marwa Qaraqe. Machine Learning Approaches to Automatic Stress Detection: A Review. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2018-Novem:1–6, 2019. ISSN 21615330. doi: 10.1109/AICCSA.2018.8612825.

[7] Zhendong Mu, Jianfeng Hu, and Jianliang Min. Driver fatigue detection system using electroencephalography signals based on combined entropy features. *Applied Sciences (Switzerland)*, 7(2), 2017. ISSN 20763417. doi: 10.3390/app7020150.

[8] Houtan Jebelli, Mohammad Mahdi Khalili, and Sang Hyun Lee. A Continuously Updated, Computationally Efficient Stress Recognition Framework Using Electroencephalogram (EEG) by Applying Online Multi-Task Learning Algorithms (OMTL). *IEEE Journal of Biomedical and Health Informatics*, 23(5):1928–1939, 2018. ISSN 21682194. doi: 10.1109/JBHI.2018.2870963.

[9] W. L. Lim, O. Sourina, and L. P. Wang. STEW: Simultaneous task EEG workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2106–2114, 2018. ISSN 15344320. doi: 10.1109/TNSRE.2018.2872924.

[10] Rebecca L. Charles and Jim Nixon. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74(September 2016):221–232, 2019. ISSN 18729126. doi: 10.1016/j.apergo.2018.08.028. URL https://doi.org/10.1016/j.apergo.2018.08.028.

[11] Mike X Cohen. Analyzing Neural Time Series Data. *Analyzing Neural Time Series Data*, 2019. doi: 10.7551/mitpress/9609.001.0001.

[12] Herbert H Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, 10:370–375, 1958.

[13] Robert Oostenveld and Peter Praamstra. The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, 112(4):713–719, 2001.

[14] Budi Thomas Jap, Sara Lal, Peter Fischer, and Evangelos Bekiaris. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, 36(2 PART 1):2352–2359, 2009. ISSN 09574174. doi: 10.1016/j.eswa.2007.12.043.

[15] John Duncan and Adrian M Owen. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences*, 23(10):475–483, 2000.

[16] Moon Kyoung Choi, Seung Min Lee, Jun Su Ha, and Poong Hyun Seong. Development of an eeg-based workload measurement method in nuclear power plants. *Annals of Nuclear Energy*, 111:595–607, 2018.

[17] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195, 1999.

[18] Tetsuya Takahashi, Tetsuhito Murata, Toshihiko Hamada, Masao Omori, Hirotaka Kosaka, Mitsuru Kikuchi, Haruyoshi Yoshida, and Yuji Wada. Changes in eeg and autonomic nervous activity during meditation and their association with personality traits. *International Journal of Psychophysiology*, 55(2):199–207, 2005.

[19] Michael Doppelmayr, Wolfgang Klimesch, Waltraud Stadler, Doris Pöllhuber, and Cornelia Heine. Eeg alpha power and intelligence. *Intelligence*, 30(3):289–302, 2002.

[20] Feng Wan, Wenya Nan, Mang I Vai, and Agostinho Rosa. Resting alpha activity predicts learning ability in alpha neurofeedback. *Frontiers in human neuroscience*, 8:500, 2014.

[21] Dae-Jin Kim, Amanda R Bolbecker, Josselyn Howell, Olga Rass, Olaf Sporns, William P Hetrick, Alan Breier, and Brian F O'Donnell. Disturbed resting state eeg synchronization in bipolar disorder: a graph-theoretic analysis. *NeuroImage: Clinical*, 2:414–423, 2013.

[22] Matthias M Müller, Thomas Gruber, and Andreas Keil. Modulation of induced gamma band activity in the human eeg by attention and visual information processing. *International Journal of Psychophysiology*, 38(3):283–299, 2000.

[23] R. D. Dias, M. C. Ngo-Howard, M. T. Boskovski, M. A. Zenati, and S. J. Yule. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of Surgery*, 105(5):491–501, 2018. ISSN 13652168. doi: 10.1002/bjs.10795.

[24] Xiyuan Hou, Yisi Liu, Olga Sourina, and Wolfgang Mueller-Wittig. CogniMeter: EEG-based Emotion, Mental Workload and Stress Visual Monitoring. *Proceedings - 2015 International Conference on Cyberworlds, CW 2015*, pages 153–160, 2016. doi: 10.1109/CW.2015.58.

[25] Budi Thomas Jap, Sara Lal, Peter Fischer, and Evangelos Bekiaris. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, 36(2 PART 1):2352–2359, 2009. ISSN 09574174. doi: 10.1016/j.eswa.2007.12.043. URL http://dx.doi.org/10.1016/j.eswa.2007.12.043.

[26] Leonard J. Trejo, Karla Kubitz, Roman Rosipal, Rebekah L. Kochavi, and Leslie D. Montgomery. EEG-Based Estimation and Classification of Mental Fatigue. *Psychology*, 06(05):572–589, 2015. ISSN 2152-7180. doi: 10.4236/psych.2015.65055.

[27] Sanay Muhammad Umar Saeed, Syed Muhammad Anwar, Humaira Khalid, Muhammad Majid, and Ulas Bagci. Electroencephalography based Classification of Long-term Stress using Psychological Labeling. *Sensors 20, no. 7: 1886*, pages 1–8, 2019. URL http://arxiv.org/abs/1907.07671.

[28] Sanay Muhammad Umar Saeed, Syed Muhammad Anwar, Muhammad Majid, Muhammad Awais, and Majdi Alnowami. Selection of Neural Oscillatory Features for Human Stress Classification with Single Channel EEG Headset. *BioMed Research International*, 2018, 2018. ISSN 23146141. doi: 10.1155/2018/1049257.

[29] Yisi Liu and Olga Sourina. EEG databases for emotion recognition. *Proceedings - 2013 International Conference on Cyberworlds, CW 2013*, pages 302–309, 2013. doi: 10.1109/CW.2013.52.

[30] Xiyuan Hou, Yisi Liu, Wei Lun Lim, Zirui Lan, Olga Sourina, Wolfgang Mueller-Wittig, and Lipo Wang. CogniMeter: EEG-based brain states monitoring. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9590:108–126, 2016. ISSN 16113349. doi: 10.1007/978-3-662-53090-0_6.

[31] Yisi Liu, Zirui Lan, Han Hua Glenn Khoo, King Ho, Holden Li, Olga Sourina, and Wolfgang Mueller-Wittig. EEG-based evaluation of mental fatigue using machine learning algorithms. *Proceedings - 2018 International Conference on Cyberworlds, CW 2018*, pages 276–279, 2018. doi: 10.1109/CW.2018.00056.

[32] Hafeez Ullah Amin, Wajid Mumtaz, Ahmad Rauf Subhani, Mohamad Naufal Mohamad Saad, and Aamir Saeed Malik. Classification of EEG signals based on pattern recognition approach. *Frontiers in Computational Neuroscience*, 11 (November):1–12, 2017. ISSN 16625188. doi: 10.3389/fncom.2017.00103.

[33] Rifai Chai, Yvonne Tran, Ganesh R. Naik, Tuan N. Nguyen, Sai Ho Ling, Ashley Craig, and Hung T. Nguyen. Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob:4654–4657, 2016. ISSN 1557170X. doi: 10.1109/EMBC.2016.7591765.

[34] Lipo Wang Lim, Wei Lun, Olga Sourina. STEW: Simultaneous Task EEG Workload Dataset, 2018. URL http://dx.doi.org/10.21227/44r8-ya50.

[35] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[36] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. *expert systems with applications*, 36(10): 11994–12000, 2009.

[37] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[38] LR Quitadamo, F Cavrini, L Sbernini, F Riillo, L Bianchi, S Seri, and G Saggio. Support vector machines to detect physiological patterns for eeg and emg-based human–computer interaction: a review. *Journal of neural engineering*, 14(1): 011001, 2017.