VŠB – Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Computer Science

DIPLOMA THESIS

2018 Bc. Hikmat Dashdamirov

VŠB – Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Computer Science

Estimation of Emotions and Mental Concentration using Deep Learning Techniques
Odhad emocí a duševní koncentrace pomocí technik Deep Learningu

2018 Bc. Hikmat Dashdamirov

VŠB - Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Computer Science

Diploma Thesis Assignment

Student: **Bc. Hikmat Dashdamirov**

Study Programme: N2647 Information and Communication Technology

Study Branch: 2612T025 Computer Science and Technology

Title: Estimation of Emotions and Mental Concentration using Deep Learning Techniques
Odhad emocí a duševní koncentrace pomocí technik Deep Learningu

The thesis language: English

Description:

Deep Learning paradigm has gained relevance last years in the community of Artificial Intelligence. Several Deep Learning techniques have been successfully applied for computer vision, natural language processing, image processing, time-series analysis. In this master project the student will analyze the performance of unconventional Neural Networks for scoring the mental concentration and recognizing emotions. The student will collect information of EEG signals that are associated with specific tasks. In particular, the experiments will be related with mental concentration of subjects in activities that required high control of the persons (for instance driving a car). The students will create a dataset of EGG signals following standard protocols. Besides, he will develop Deep Neural Network techniques to identify concentration levels of the person.

1. Theoretical revision of Deep Learning techniques and current state in the field of emotions recognition.
2. Theoretical study of EEG signals and techniques for signal processing.
3. Collection of the empirical data.
4. Implementation of Deep Learning tools, preprocessing of the data.
5. Evaluation phase, testing and analysis.

References:

1] Juergen Schmidhuber, Deep Learning in Neural Networks: An Overview, Neural Networks, Vol. 61, pages 85-117, 2015. Doi: 10.1016/j.neunet.2014.09.003, available at: https://arxiv.org/abs/1404.7828
[2] Yoshua Bengio, Deep Learning of Representations: Looking Forward, Department of Computer Science and Operations Research, University of Montreal, Canada, 2013. Available at: http://goo.gl/OK0WV9 .
[3] Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, Springer, February 2009.
[4] Deep Learning information: http://deep learning.net.
[5] Signal Analysis: http://neuro.felk.cvut.cz/publications/

Extent and terms of a thesis are specified in directions for its elaboration that are opened to the public on the web sites of the faculty.

Supervisor: **Sebastian Basterrech Tiscordio, PhD.**

Date of issue: 01.09.2017

Date of submission: 30.04.2018

doc. Ing. Jan Platoš, Ph.D.
Head of Department

prof. Ing. Pavel Brandštetter, CSc.
Dean

I hereby declare that this masters thesis was written by myself. I have quoted all the references I have drawn upon.

Ostrava, 30th April, 2018 ....

Acknowledgments

I want here to express my deep gratitude to the people who supported me throughout my work in this thesis. That is why, I would like to use this opportunity in order to give back something to this great people.

First of all, I would like to express my big appreciation to Dr. Sebastián Basterrech, who supported me in every part of this work. He helped me almost instantly whenever I faced with technical challenges, answered and explained every question of mine patiently, if necessary several time. A lot of great advices were given by him, which enabled me to improve the quality of this work. The motivation he gave me every time is invaluable.

I am very grateful to my family for encouraging me during this work. Priceless support was given to me by my fantastic family when i struggled. My special thanks to my little niece who is not even one year old. She increased my motivation and mood with her positive nature.

It is my pleasure to dedicate this thesis to my little niece, Humra hanim.

Abstrakt

Cílem práce je ohodnocení lidských mozkových vln s využitím metod hlubokého učení (deep learning) a evolučních výpočetních technik a pro ověření výkonu aplikovaných technik. V diplomové práci jsou využity dobře známé metaheuristiky a umělé neuronové sítě pro klasifikaci lidských mentálních aktivit za použití elektroencefalografických signálů. Bylo vyvinuto rozhraní mozek-počítač, které je schopno zpracovat elektroencefalografické signály a klasifikovat mentální soustředění v porovnání s relaxací. Systém je schopen automaticky extrahovat a naučit se reprezentaci daných dat. Na základě vědeckých protokolů byl navržen experiment pro rozhraní mozek-počítač a byla vytvořena původní a relevantní data pro průmyslovou a akademickou komunitu. Vygenerovaná pokusná data jsou přístupné pro vědeckou komunitu. V rámci experimentů bylo využito zařízení založené na encefalografii pro sběr mozkových signálů subjektu během specifických aktivit. Nasbíraná data reprezentují mozkové vlny subjektu, který byl stimulován psaním úloh .

Dále byla vybrána nejlepší kombinace vstupních vlastností (informace o mozkových vlnách) s využitím následujících dvou metaheuristických metod: simulovaného žíhání a geometrické optimalizace hejnem částic (Particle Swarm Optimization). Umělá neuronová síť, která se nazývá Echo State síť, byla aplikována pro řešení mapování mezi informacemi z mozku a aktivitami subjektu. Výsledky ukazují, že je možné odhadnout lidskou aktivitu pomocí několika encefalografických signálů. Kromě toho, navrhovaný systém je vyvinut s využitím rychlých a robustních učících technik, které mohou být jednoduše přizpůsobeny podle jednotlivých subjektů. Tento přístup navíc nevyžaduje výkonné výpočetní prostředky. V důsledku toho může být systém využit v prostředí, které jsou výpočetně omezeny a/nebo v případech, kdy výpočetní čas je důležitým hlediskem.

Klíčová slova: Echo State sítě, EEG signály, rozhraní mozek-počítač, hejnová optimalizace, simulované žíhání, rozpoznávání emocí

Abstract

The purpose of this work is to evaluate the brain waves of humans with deep learning methods and evolutionary computation techniques, and to verify the performance of applied techniques. In this thesis, we apply well–known metaheuristics and Artificial Neural Networks for classifying human mental activities using electroencephalographic signals. We developed a Brain–Computer Interface system that is able to process electroencephalographic signals and classify mental concentration versus relaxation. The system is able to automatically extract and learn representation of the given data. Based on scientific protocols we designed the Brain–Computer Interface experiments and we created an original and relevant data for the industrial and academic community. Our experimental data is available to the scientific community. In the experiments we used an electroencephalographic based device for collecting brain information form the subjects during specific activities. The collected data represents brain waves of subjects who was stimulated by writing tasks.

Furthermore, we selected the best combination of the input features (brain waves information) using the following two metaheuristic techniques: Simulated Annealing and Geometric Particle Swarm Optimization. We applied a specific type of Artificial Neural Network, named Echo State Network, for solving the mapping between brain information and subject activities. The results indicate that it is possible to estimate the human concentration using few electroencephalographic signals. In addition, the proposed system is developed with a fast and robust learning technique that can be easily adapted according to each subject. Moreover, this approach does not require powerful computational resources. As a consequence, the proposed system can be used in environments which are computationally limited and/or where the computational time is an important issue.

Keywords: Echo State Networks, EEG signals, Brain Computer Interface, Swarm Optimization, Simulating Annealing, Emotion Recognition.

Contents

List of Abbreviations

ABC Artificial Bee Colony

ACM Association for Computing Machinery

ACO Ant Colony Optimization

ANNs Artificial Neural Networks

BCI Brain Computer Interface

CNNs Convolutional Neural Networks

CO Combinatorial Optimization

DNNs Deep Neural Networks

EA Evolutionary Algorithms

ECG Electrocardiography

ECoG Electrocorticography

EEG Electroencephalographic

EP Evolutionary Programs

ESN Echo State Networks

FMRI Functional Magnetic Resonance Imaging

FNIRS Functional Near Infrared Spectroscopy

GA Genetic Algorithms

GPSO Geometrical Particle Swarm Optimization

GRASP Greedy Randomized Adaptive Search Procedure

GSR Galvanic Skin Response

HCI Human Computer Interaction

ILS Iterated Local Search

MA Memetic Algorithms

MEG Magnetoencephalography

MLP Multilayer Perceptrons

MRI Magnetic Resonance Imaging

LSM Liquid State Machine

LSTM Long Short-Term Memory

PET Position Emission Tomography

PSO Particle Swarm Optimization

RC Reservoir Computing

RNNs Recurrent Neural Networks

SA Simulation Annealing

TS Tabu Search

VND Variable Neighborhood Descent

List of Figures

List of Tables

1 Introduction

Activities of daily living require several mental operations such as attention, concentration and meditation. There are special cases where we need to measure emotional responses for making assessments about the true condition of mental focus and concentration in a specific task. Such cases could involve jobs which require high concentration, psychological tests, medical diagnostic, and educational purposes. A Brain Computer Interface (BCI) system can be useful for studying emotional responses and understanding human brain activities [10]. A BCI system interfaces the brain with an external acquisition device. It fundamentally is combination of three components [39]: an acquisition system for brain information (for example: Electroencephalographic (EEG) signals, Magnetic Resonance Imaging (MRI), Functional Magnetic Resonance Imaging (FMRI), a processing information system that contains a Machine Learning tool, and an external device (for example a robotic arm, a keyword, a wheel chair, etc.).

In a BCI system the subjects perform mental tasks. In other words, the subjects don't realize a specific activity itself, they imagine the realization of some tasks. That is why, it is useful for assisting and repairing human cognitive and sensory-motor functions. In recent years, a BCI system has become very useful for diverse areas such as clinical applications, communications devices, learning systems, human performance evaluations, entertainment [38, 39, 19]. In this thesis, we present an application of Artificial Neural Networks (ANNs) and evolutionary algorithms for classifying human mental activities using EEG signals in BCI context. Inspired by the BCI system, we developed a portable system that collects few EEG signals and is able to identify the human concentration versus relaxation actions during human activities. Instead of collecting the brain information during a specific mental task, we collect the EEG signals during real activities that requires high concentration from a subject. We are interested in developing a system for measuring the human concentration during specific activities. We study a binary situation: high concentration versus relaxation.

An ANNs solves the classification problem with a binary output variable: high focus on the task (concentration) *yes* or *not*. We use an Echo State Networks (ESN) [74] and some variations of the canonical ESN model. These ANNs are simple, fast and robust learning classification tools. The training schema of the ESN model allows to use the device in scenarios where the computational resources may be highly constrained. In addition, the ESN model obtains competitive accuracy with respect to other learning tools. This model has shown good performances in several real-world problems, and it has the following two main advantages with respect to other learning tools: the training process is fast and robust. Thus, we can easily adapt our system to process new training data, as well as the system can be specifically trained for each subject.

In addition, we apply two evolutionary algorithms for selecting the most relevant input features. A disadvantage of the EEG signals is that the signals may contain noise. The noise can be presented in each signal, as well as is created among the signals [15]. For that reason, we apply a searching algorithm for selecting the best combination of input features. We collected data which represents 12 input signals. Instead of using a brute force algorithm for selecting the best combination of features, we apply a local search algorithm and a population search method for finding the optimal solution. We selected the Simulation Annealing (SA) and Geometrical Particle Swarm Optimization (GPSO). Both methods are well–known in the community for their good results. Moreover, they contain few parameters, hence, they are suitable searching techniques for embedded systems.

The collected dataset is available to the academic research community, as well as the source code of the developed algorithms presented in this article [13].

Main Contributions

The summary of our main contributions during this thesis are the following ones. We designed a BCI system, which is able to predict mental states. Our system has several advantages, we remark at least the following four:

- **Robustness:** The applied learning algorithms are fast and robust methods
- **Portability:** The BCI is designed with a portable signal acquisition device

- **Adaptability:** We designed the BCI system with learning algorithms which are able to adjust the model parameters with a low computation cost algorithms

- **Low costs:** The designed BCI is a low-budget system which can be used as a portable device for this kind of researches

The initial part of our work was focused on the experimental part. We designed the experiments and we collected the data by following the scientific protocols in the area [38, 37]. On the next stage we focused on developing a Machine Learning tool for finding hidden patterns in the EEG signals. After getting satisfactory results, we extended our methodology. We integrated evolutionary algorithms for selecting the most interesting features of the collected data. In particular, we developed a very robust solution that is confirmed working after many examples. Our contributions were accepted on the following top-tier conferences on the area:

- Hikmat Dashdamirov and Sebastián Basterrech, "Estimation of Human Concentration using Echo State Networks," European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 25 - 27 April 2018. Article available in Arxiv: `goo.gl/Pt2bmz`.

- Hikmat Dashdamirov, Sebastián Basterrech and Pavel Kromer, "A Nature-inspired System for Mental State Recognition," 2018 IEEE World Congress on Computational Intelligence (WCCI 2018), Rio de Janeiro, Brazil, 08-13 July 2018. Article available in Arxiv: `goo.gl/Pt2bmz`.

Experimental Reproducibility

The elaborated dataset during this master thesis and the source codes are available for the scientific community in `goo.gl/KX6aMa`.

Organization of the thesis

The thesis is organized as follows:

- **Chapter 2:** Section 2.1 provides a general discussion of Human Computer Interaction (HCI) as a field in computer science. After a general introduction into HCI we introduce the BCI and applications of BCI in Section 2.2 and in Section 2.2.1, respectively. We list signal capturing techniques from brain in Section 2.2.2. With the knowledge of general introduction into the field of HCI, BCI as a subfield of HCI and capturing brainwaves, EEG based BCI systems and relevance with this thesis is given subsequently. Firstly, the insight of EEG signals is described in Section 2.2.3, followed by an introduction into EEG based concentration evaluation in Section 2.2.4 and concluded by EEG based Emotion recognition in Section 2.2.5 which is our main aim in this thesis.

- **Chapter 3:** This chapter shortly describes computational models and focuses on the ANNs. Section 3.1 describes main ideas of ANNs, single neuron and gives knowledge about multi-layered architecture. Subsequently, Recurrent Neural Networks (RNNs) and ESN are provided in Section 3.2 and in Section 3.3. Lastly, review of literature about applications of ANNs in Biomedical Analysis is provided broadly in Section 3.4.

- **Chapter 4:** In this chapter the optimization problem is introduced in Section 4.1 with several listed solutions. SA and GPSO are covered deeply with formulation and pseudo-codes in Section 4.1 and Section 4.3, respectively.

- **Chapter 5:** The detailed explanation of conducted research is given in this chapter. First of all, common classification problem is formalized in Section 5.1. In Section 5.2, insights of data collection protocol are provided in detail. Following Section 5.3 describes exact procedure for better understanding of data collection. The experimental settings are given in Section 5.4 which are used for evaluation of collected data. Finally, the main results of this thesis are provided in Section 5.5.

- **Chapter 6:** The thesis is concluded with the summarization of the major findings and discussion about the future work.

2 State of Art in Mental Concentration

Human's mental models has been a very interesting problem for scientist from fields such as psychology, biology, educational areas as well as computer science. This chapter gives up-to-date information about mental concentration in relation with EEG signals. We firstly, discuss HCI as a field in computer science which tries to study how differently persons interact with technology in Section 2.1. The main goal of HCI is to bridge gaps between users and new technologies. The discussion is given about how developing efficient and effective forms of HCI can reduce the skill levels needed to use complex devices.

The Section 2.2 introduces BCI and relevant literature about BCI as a subfield of HCI. After the general information about BCI, in Subsection 2.2.1, applications of BCI is discussed. Brain information capturing techniques are presented in Subsection 2.2.2, subsequently. The EEG signal details are given in Subsection 2.2.3 for better understanding of later chapters. Besides the general information, the focus in this chapter is on EEG signals in concentration and emotion recognition context. These topics are addressed in Subsection 2.2.4 and 2.2.5, respectively.

2.1 Human-Computer Interaction

HCI is a multidisciplinary area that involves cognitive sciences and human factors engineering. It has been started as a field in the computer science area. Nowadays, it has emerged as a very active research and practice discipline. The interdisciplinary field that now we call HCI, it has gained prominence during the 80s in in the universities of Stanford and MIT, U.S. The Association for Computing Machinery (ACM) proposed the following definition [46]:

> *Human–Computer Interaction is a discipline concerns to the design, implementation and evaluation of interactive computing systems for human use and with the study of major phenomena surrounding them.*

Today the area is composed by the knowledge coming from: Psychology, Engineering, Semiotic and Linguistics, Social Sciences, and Computer Sciences. The discipline has evolved along the years. We can see three periods: first, second and third wave of understanding the HCI [22]. The first wave of HCI (at the 80s) has put emphasis in the user and users dimensions. The center has been the machine and not the user. A second wave of HCI (at the 90s) expanded the machine to context based systems. It has been an expansion from the desktop to workspaces. In addition, it has been changed the concept from user to human. The emphasis has been in the design of the environment. A third wave of HCI has started in the 2000s, where it has been expanded the interaction between humans and their home and larger environments. The users are interpreted now like actors and participants. The research has focus also on non-tangible factors such as attention and emotions. There are a huge number of applications. It encompasses information systems, visualization and many areas of design.

Main objective of this research field is to improve HCI by improving the user-friendliness of computer interfaces [22]. Computer system can be controlled by different means of information. We can design unconventional ways of communication between the environment and the human. Furthermore, with todays technology it is possible to input some information into the computer system only just by thinking. That is possible for example with the help of EEG devices and recent advances in signal processing and pattern recognition [92]. These systems have made it possible to evaluate humans brain signals and integrate the results in human computer interfaces.

EEG devices could easily catch brain states such as concentration level, stress level etc. in real time. Brain of humans receive some stimulations such as audio, visual from computer systems and this information is processed by brain. As a result some emotions is generated and EEG devices can recognize this emotions. Based on this information some patterns can be learned and commands could be formed. Commands can represent different aspects of human mind such as pain level or concentration level depending on application. This systems could be applied to medical applications, entertainment, performance improvement.

2.2 Brain Computer Interface

One of the important sub-research field of HCI is BCI which deals with electrical activity of human brains. BCI [10] are systems those provide interaction between a wired brain and an external device. BCI captures the signals and delivers them to the computer to complete the intended task. BCI has broad usage in different areas such as industry, educational, advertising, entertainment, and smart transportation.

A BCI system fundamentally consists of three components: a brain signal acquisition system, an information processing device, and external device. A General schema of a BCI system is demonstrated in Figure 2.1.

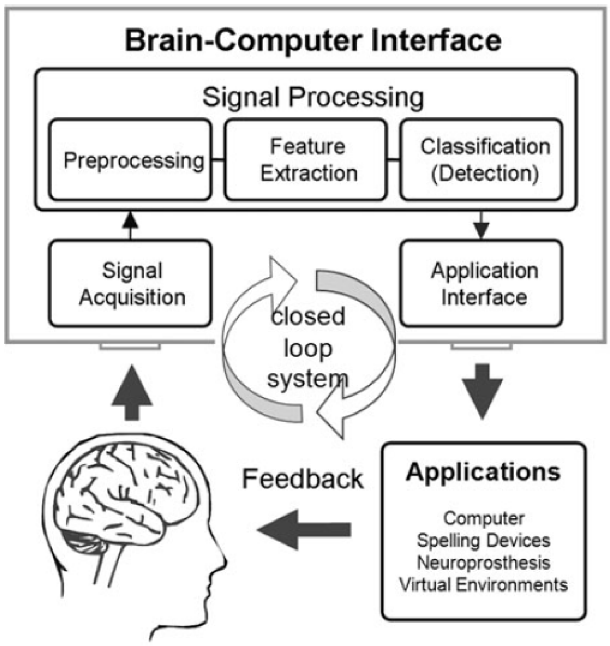

Figure 2.1: The basic structure of a BCI systems. The image was taken from [42]

The user generates brain waves patterns which are captured by appropriate signal acquisition device. Currently several signal acquisition methods are used for BCI operations such as EEG signals acquired from scalp electrodes [15, 111]. In signal processing stage, the recorded data is pre-processed, features are selected, and classified. In the subsequent stage, a control signal is created which is sent to application through well–defined connection. Finally, application gives feedback to the user.

The device used in this thesis is named NeuroSky‘s Mindset [48]. It is one of the BCI which is EEG-based and provides information on a subject‘s Delta, Theta, Alpha, Beta, and Gamma brainwave. Main research fields and applications of BCI is presented in following subsection.

2.2.1 Applications of BCI

BCI have broad range of usage in various fields of research [109]. These fields can involve from medical applications to security fields. In the following list we present an overview of the applications of BCI.

- **Medical applications:** BCI can provide crucial information about health conditions of human beings by taking advantage of brain signals. Later this data can be used in all phases of healthcare including prevention [91, 59, 7], detection [112, 71, 61], diagnosis [35, 128], rehabilitation [18, 12, 3] and restoration [96, 31].
- **Neuroergonomics and smart environment:** Safety, luxury and physiological control of human‘s daily life could be achieved by the use of BCI [89, 72, 73]. Such benefits could be applied to smart houses, transportation, workplaces or generally smart environments [126, 106, 34].
- **Neuromarketing and advertisement:** Marketing area could be exploited by BCI systems especially by analyzing EEG signals. BCI system can make assessment about generated attention during advertisement time and evaluate advertising method based on collected data [124, 134, 125].
- **Educational and self-regulation:** Brain performance could be enhanced by getting feedback with the help of BCI. For instance clearness of studied material or stress level of sport competitors could be determined by recording brain signals [116, 139, 85].
- **Games and entertainment:** Entertainment and gaming applications presents new experience by using brain controlling features. One of such example is BrainArena.

The developed game is done with two BCIs such that the players can score goals by imagining left or right hand movement [121, 107, 20].

- **Security and authentication:** Security systems mainly include knowledge based, object based and/or biometrics based authentication. These methods have some drawbacks such as insecure passwords, shoulder surfing, theft crime etc. Brain signals are explored as a solution for these vulnerabilities. Main point is that bio-signals are not accessible for external observers. Furthermore, brain signals can be used by disabled patients or users who misses the associated physical trait by the biometrics based authentication [86, 87, 88].

2.2.2 Capturing brain signals

There are several source of information from the brain. One of the most popular and used in our work are the EEG. Nevertheless, EEG signals are not the only one which are able to provide information about the brain. We can also extract the information from the brain using other devices. Here we summarize the different techniques [38, 37]:

- **Microelectrode array**: It is an invasive technique. It has a very high spatial resolution and a very high temporal resolution. An advantage is that it is a portable device.
- **Electrocorticography (ECoG)**: It is a partial-invasive. It has a very high spatial resolution, and a very high temporal resolution. It is considered a portable device.
- **Magnetoencephalography (MEG)**: It is no invasive, It has a high spatial resolution, and a very high temporal resolution. A limitation is that is not a portable tool.
- **FMRI**: It is no invasive. It has a very high spatial resolution, and a low temporal resolution. It is a no portable device.
- **Functional Near Infrared Spectroscopy (FNIRS)**: It is no invasive, and it has a high spatial resolution, and a low temporal resolution. It is a portable device.
- **EEG**: It is no invasive, it has low spatial resolution, and a very high temporal resolution. It is no portable. An example of an EEG signal is presented in Figure 2.2.

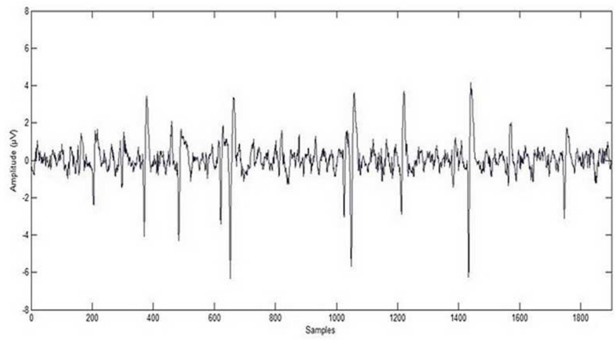

Figure 2.2: One second example of different types of EEG signal

2.2.3 EEG signal processing

EEG is an electrophysiological approach for measuring electrical activity of brain with the electrodes those are attached to the scalp. The voltage fluctuations which occurs during neurotransmission within the brain are recorded. EEG has usability advantages over other methods such as, it is easy to use, portable, inexpensive and it provides high temporal resolution. However, limitations in signal-to-noise ratio, low spatial resolution, restrictions related to measurement and identification of specific locations of brain are disadvantages of EEG. Established data collection experiment uses NeuroSkys Mindset which is EEG-based signal acquisition device and provides information about a subjects Delta, Theta, Alpha, Beta, and Gamma brainwaves. Figure 2.3 graphically represents these signals [4] and detailed description of these signals are listed below [108].

- **Delta wave:** The changing interval of delta wave is between 0.5-4 Hz. Pattern of delta is the slowest in waves and highest in amplitude. Delta wave provides information about:deep sleep, serious brain disorders and the waking state.

- **Theta wave:** The changing interval of theta wave is between 4-8 Hz, and its amplitude is commonly greater than 20 μV. Emotional stress, frustration or disappointment and unconscious material, creative inspiration and deep meditation is related to theta.

- **Alpha wave:** This signal lies between range of 8-13 Hz, and the shape is observed with 30-50 μV amplitude which mainly is seen when the subject is in a relaxation state or has eyes closed in the areas of the brain (occipital lobe). It is commonly related to the intense mental activity, stress and tension.

- **Beta wave:** This wave contains the frequency range from 13 to 30 Hz, and appears in a low amplitude. The beta wave is generated when brain actively engaged in mental activities. The beta wave are usually related to solving problems, focusing on things, paying attention to something are usually. Generally the beta waves are features of highly engaged mind.

- **Gamma wave:** The frequency of gamma wave is equal or more than 30 Hz. Sometimes frequency of this wave can go up to 80 or 100 Hz. Gamma waves are generated when subject performs cognitive or motor functions.

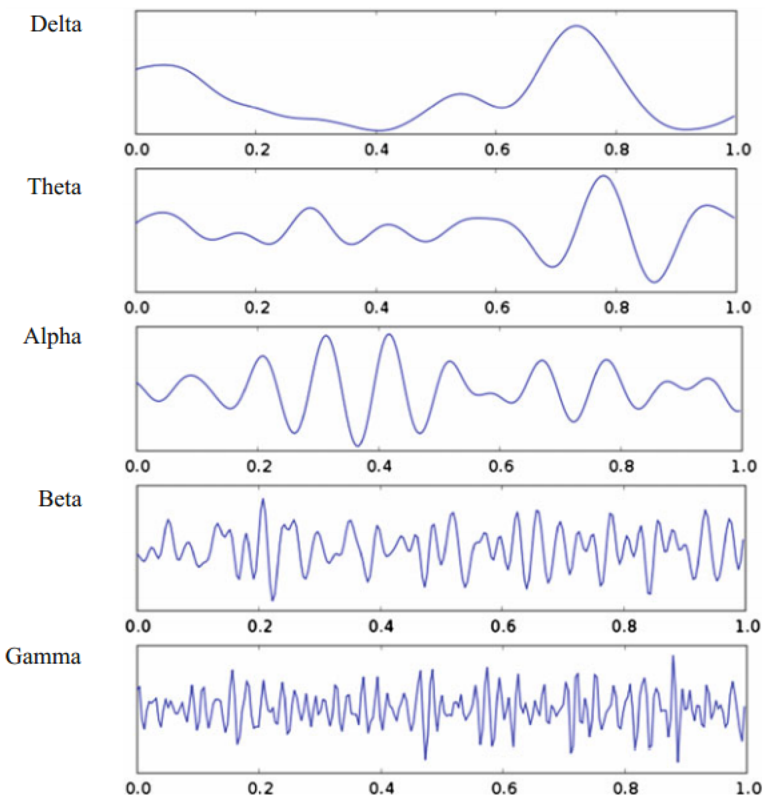

Figure 2.3: Delta, Theta, Alpha, Beta, and Gamma signals. The figure was presented in [113]

2.2.4 EEG signals in concentration

Attention states have strong relation with brainwave frequencies. Scholars has provided great deal of evidences about this phenomena. Researches approved that various states of attention or concentration can be influenced by changes in the amplitude and frequency of alpha, beta, theta waves [26]. Prinzel et al. [97] showed that increased attention reflects itself as a increased alpha waves and decreased theta waves. Another research discovered that concentrating on solving mathematical problems affects alpha waves in a negative way. Generally, attention has been classified as follows [114, 84, 63].

- **Selective attention:** Keeping attention level in some level despite of distractions.
- **Divided attention:** Focusing on more than one task at a time.
- **Sustained attention:** Maintaining a constant response in repetitive activity.
- **Focused attention:** Reacting to a stimulation in a certain task.
- **Alternating attention:** Changing response level between activities which require different reactions.

As EEG provides data about brainwave it has been used in several researches. Li et al. [70] designed several experiments for capturing EEG signals. These experimental scenarios include silent reading, mental math, conceptual understanding, playing a game. Similarly Ming et al. [79] also designed empirical tasks for subjects such as attention (playing tennis), inattention (thinking about things other than playing tennis), and rest. Xu et al. [132] designed four activities: relaxation, viewing computer images, playing a substraction game, and complex division and multiplication games. Frontal midline theta activity was studied by Sauseng et al. [110] and confirmed close relation between theta changes and attention. Doppelmayr et al. [33] examined brainwaves in the period of rifle shooting and discovered that theta waves increase before the shot.

2.2.5 EEG signals in emotion recognition

The growing topic in HCI field is being as productive as possible in user satisfaction criteria. As the users are human beings the requirements of this topic includes emotions and interaction. Several researches have been done in recognizing emotions from face and voice. However, emotions are not limited what is displayed known as expressions. That is why interpreting only expressions is not objective. Inner emotions play crucial role in our daily life and decisions.

Eight basic emotion states: anger, fear, sadness, disgust, surprise, anticipation, acceptance and joy was defined by Robert Plutchik [95]. All other emotions can be generated with the combination of these basic emotions. Different stimulations can be used for processing emotions. These stimulations involve visual, auditory and combined. With the help of this kind of stimulations various brain areas could be activated. These field is relatively new and gathered quite attention during last decade. Nowadays emotion recognition is one of the main topics of affective computing. Emotions can be captured through internal physicological signals, such as heart rate, skin conductance, Galvanic Skin Response Galvanic Skin Response (GSR), EEG, MEG, Position Emission Tomography (PET), and FMRI.

In a BCI system the EEG signals are one of the most used sources of brain information. An EEG based emotion recognition is suitable approach for researchers, the reason is that emotion are present in EEG signals. In spite of a lot of inconsistency in information about emotions some researchers showed that emotions can be recognized in some extend. A classifier with neural networks was build by Choppin et al. [27] and this system achieved 64% accuracy in classification of new unseen EEG signal samples, when using three emotion classes. Chanel et al. [24], showed that arousal dimension of emotions could be detected with 60% accuracy from EEG signals. Many researches reported that emotions could be extracted from EEG signals which are recorded from different locations in both hemispheres by looking differences in activation rate.

3 Machine Learning

Several categories of computational models have been studied during the last decades. At least five alternatives proposed and studied in the period of 1930s and '40 when the first formal definition of computability were proposed. These models include the mathematical model, the logic-operational model (Turing machines), the computer model, cellular automata and the biological model (neural networks). The main computational object studied in this thesis (neural network system) is formally a parallel distributed computational system. In Figure 3.1 different models of computation was shown.

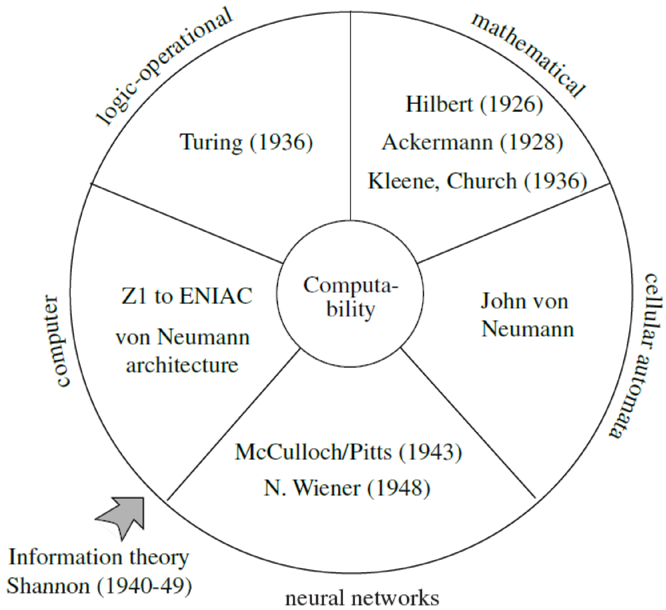

Figure 3.1: Five models of computations. The figure was taken from [102].

The short review of ANNs are presented in the following subsection since our main focus is on the biological model. ANNs are covered in this chapter as a machine learning approach for better understanding of this thesis. The literature about machine learning is extensive because of decades of research activities in this discipline [81]. However, we discuss the related topics which are necessary in order to comprehend the specific machine learning techniques which are used and to assess the results of the thesis. Afterwards, a general introduction to ANNs are given in Section 3.1 as a approach which we used in this thesis. Firstly the basic structure of single neuron is described. Additionally, this knowledge helps us to have general idea about how neurons are connected to form a neural net and understand the multi-layered structure. The main ideas of training procedure of the parameters of a network is described for better understanding the architecture. Subsequently, RNNs are covered briefly in Section 3.2. Characteristics of RNNs are discussed as a building block for ESN. ESN is described as a main approach which we evaluated in this thesis. The concept of ESN is discussed in Section 3.3. We outlined the Reservoir Computing model and the motivation for us for using this in our work. The formulation of Reservoir Computing (RC) and parameters of this model are described step by step in order to understand the main ideas of this approach. This is essential for better understanding of neural networks as a major evaluation technique. It helps to handle and realize the finding of this thesis. Lastly, the medical applications of ANNs are covered in Section 3.4.

3.1 Artificial Neural Networks

ANNs are fast growing research field in recent years. The first model of artificial neurons was presented by Warren McCulloch and Walter Pitts in 1943 [77]. Since the first model, new and sophisticated approaches have been proposed decade to decade. Artificial neural networks are based on the information processing abilities of nervous systems.

Nervous system is complex, self–organizing, but all posses basic building blocks, the neural cells or neurons. Fundamentally, neurons produce reaction or response for received signals. Biologically information is reserved and transmitted through the contact points between several neurons, the so called synapses. The information processing involves several actions such as electrical, chemical or fusion of these actions. This is the basic structure which was inspired nature of brain and applied to the artificial neural networks. Figure 3.2 demonstrates the architecture of a basic abstract neuron with n inputs [102].

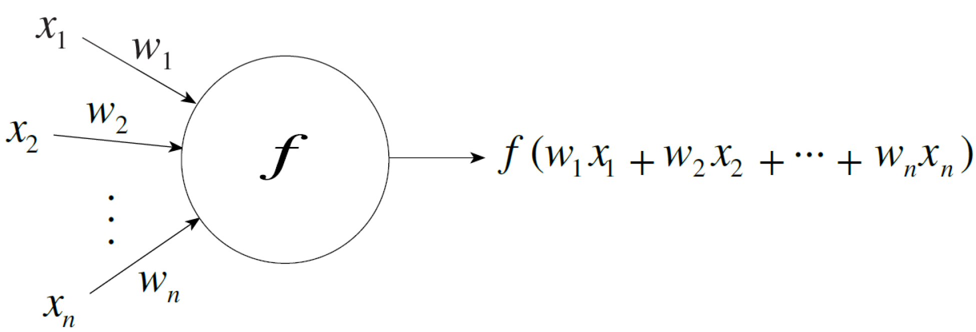

Figure 3.2: Abstract neuron with inputs $[x_1, x_2, \ldots, x_n]$ and weight connections $W_1, W_2, \ldots, W_n$. The function $f(\cdot)$ denotes the activation function of the neuron. The figure was taken from [102].

A general input pattern x_i can be transmitted through input channels which is denoted as i. The synapses is represented as an edge between input channels and neuron itself. Every edge has different transmission efficiency which was build for simulating synaptic efficiency. Data passing through the edge is multiplied by w_i constant for reflecting their efficiency. The information processing takes place in the body of neuron. The primitive function f so called activation function evaluated based on given data. This function can be selected arbitrarily depending on input data and problem formulation. If we consider each node as a unit which processes primitive function based on fed input and gives defined output, then the artificial neural networks are combination of these basic structures. Usually this is called layered architecture, because computing units are subdivided into so-called layers as presented in Figure 3.3.

The input layer covers the set of input units, the output layer covers the output units. The set of remaining units which does not have direct connections with outside are called hidden layers. This architecture does not posses cyclic connections between units. The input data is evaluated layer by layer until the final result. This first and simplest artificial neural network is known as feedforward neural network. FeedForward neural networks also known as Multilayer Perceptrons (MLP) The capabilities of feedforward neural network are limited with one time calculation, the input goes through every layer which were connected by weighted edge. That is why it could not perform efficiently when it comes to complicated and complex problems. Combination of weights of edges should be adjusted achieving satisfactory results which is known as training or learning phase.

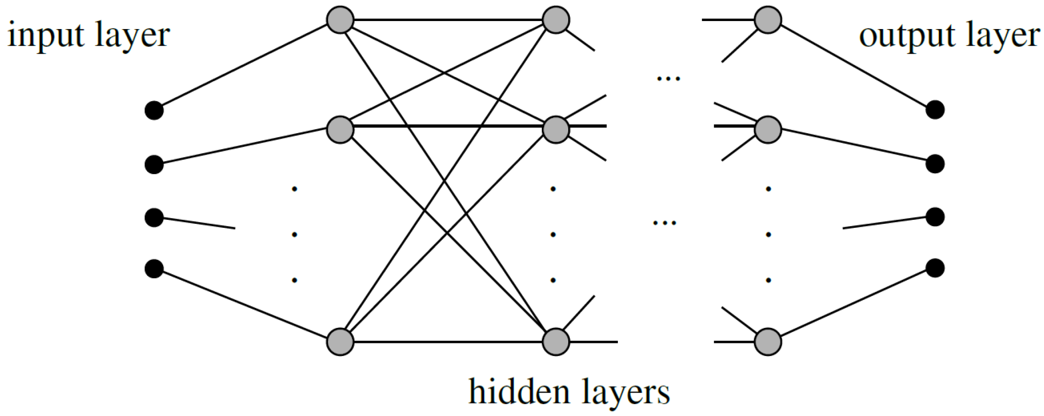

Figure 3.3: Visualization of a layered ANNs architecture. The first layer is the input layer and process the input pattern, the hidden layers performs a sequence of computations and last layer named output layer produces the ANNs results. The Figure can be found [102]

The computational effort is needed for training the neural network. The most popular learning method which is capable of handling such problem is known as backpropagation algorithm. Simply the algorithm calculates the error between desired output and produced output. By this way, algorithm tries to minimize the error function in weight space by applying the gradient descent method. Since the method calculates the gradient of error function at each step, the error function have to be continuous and differentiable. Moreover, backpropagation requires known output for every input sample. The learning process is iterative and it can be generalized as follows:

i. Weights of every edge initialized randomly.

ii. Feed-Forward computation.

iii. Backpropagation is applied until input layer.

iv. Weights are updated.

v. Termination when the error function value is small enough.

The goal of this approach is to create a MLP with the learning ability. As a result it can learn internal representation of complex problems and map the input data to the desired output as accurate as possible. However, as mentioned MLP can not involve cyclic edges which limits its power. In the next subsection, general overview about RNNs are presented those have loops (cyclic connections) in its structure of interconnections between neurons. This characteristic provides interesting properties and enables us to work with time dependent data.

3.2 Recurrent Neural Networks

RNNs also were inspired from the brain structure of living beings. Recurrent structure of brain is a known fact in biology. The basic difference between MLP and RNN is information flow. MLP has feedforward architecture, where information flows in one direction only. RNN has been organized differently from MLP in the means of information flow. Cyclic edges exist in RNN architecture which enables neurons to feed itself. This property of RNN well-suits time series data because RNN can remember and process past information.

The RNN receives an input at each time-step, adjusts its hidden state, and evaluates data. The procedure can be demonstrated as in Figure 3.4.

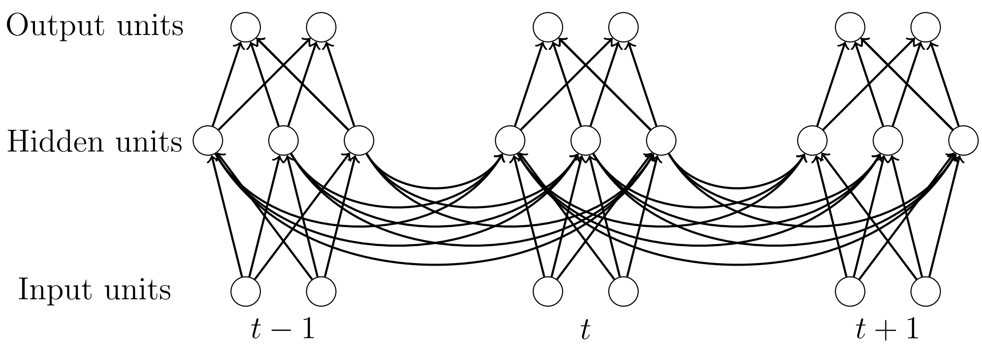

Figure 3.4: The topology of a Recurrent Neural Network. The figure was taken from [16].

One way of training RNN is backpropagation through time. Although, RNN is highly unstable due to its complex dynamics and gradient descent is ineffective [120]. As, the collected data is time-dependent (EEG signals), thus a Recurrent Neural Network (RNN) [129] was used for time-series modeling which is a powerful tool for solving this type of problems. In this thesis, we apply a particular case of RNN named Echo State Network (ESN) [74]. The ESN was proposed as a learning method. Detailed description is presented on the following subsections.

3.3 Echo State Networks

At the beginning of 2000s, a new approach for modelling and training recurrent ANNs was introduced with the names, Liquid State Machine (LSM) [75] and ESN [49]. Since 2007 the approach has become popular under the name of RC [127].

The ESN model is a Neural Network composed by a hidden recurrent structure (called reservoir) and a readout structure that is a linear regression. A RC model performs a convolution composed by at least two operations. The first one, named *reservoir*, is a dynamical system implemented using a recurrent neural network. The second operation, named *readout*, consists of a supervised learning method.

An outstanding characteristic of a RC model is the extreme efficiency of the training process [6]. The weight parameters in the reservoir are randomly initialized and kept fixed during the learning process. Only the parameters in the readout structure are adjusted using the training set. The reservoir can be seen as a kernel method that expands the input patterns in a new space. Then, the classification problem is solved by adjusting the parameters of the readout structure. The first two RC techniques (ESN and LSM) use a linear regression as supervised method in the readout part. Therefore, the learning is very fast and robust. In addition, the family of RC models has obtained very well performances in several real applications [127, 74]. The general structure of ESN is demonstrated in Figure 3.5

Subsequenly, we formalize a RC model, we use the notation of [49]. Let N_{a} be the number of input neurons, N_{x} denotes the number of neurons in the reservoir and N_{y} is the number of output neurons. In our application, the model output is binary then $N_{\mathrm{y}} = 1$. Let $\mathbf{w}^{\mathrm{in}}$ be a $N_{\mathrm{x}} \times N_{\mathrm{a}}$ matrix that collects the input-reservoir weights. Similarly, let $\mathbf{w}^{\mathrm{r}}$ be a $N_{\mathrm{x}} \times N_{\mathrm{x}}$ matrix of hidden-hidden weights and let $\mathbf{w}^{\mathrm{out}}$ be a $N_{\mathrm{y}} \times N_{\mathrm{x}}$ matrix with the projected space to the output space. The hidden-hidden weights define a RNNs that is characterized by the following recurrence:

$$\mathbf{x}(t) = f_1(\mathbf{w}^{\mathrm{in}}\mathbf{a}(t) + \mathbf{w}^{\mathrm{r}}\mathbf{x}(t-1)). \tag{3.1}$$

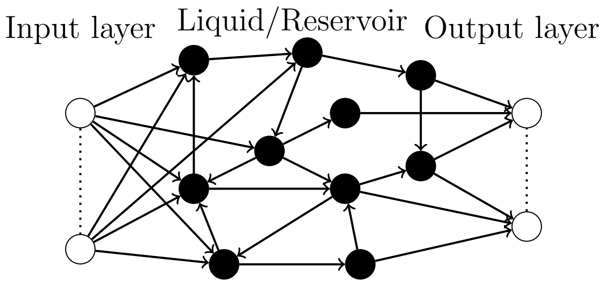

Figure 3.5: Example of a standard topology of the Reservoir Computing model. Figure was taken from [17].

The output of the model is given by:

$$\hat{y}(t) = f_2(\mathbf{w}^{\text{out}}\mathbf{x}(t)), \tag{3.2}$$

The functions $f_1(\cdot)$ and $f_2(\cdot)$ are two predefined coordinate-wise functions. In this research, we use a slight variation of the previous defined RC model consists of leaky-neurons in the reservoirs [50], the reservoir states are computed by:

$$\mathbf{x}(t) = (1-\alpha)\mathbf{x}'(t) + \alpha\mathbf{x}(t-1), \tag{3.3}$$

where the parameter $\alpha \in [0, 1)$ is called *leaky rate* and is used for controlling the update of the reservoir state. For the sake of notation simplicity, we omit the bias term (it is implicitly included in the weight matrices).

The family of RC models is very large. There are two main differences among the RC methods. One is related to the type of random projection that depends on the reservoir matrix. Another difference, is the type of supervised learning used in the readout structure. In the case of the canonical RC model (the ESN method) the function $f_1(\cdot)$ is an hyperbolic tangent and the readout function $f_2(\cdot)$ is a linear regression.

A RC model has several global parameters, which may have an impact on the model performance. In the literature the most analyzed ones are [74, 21]:

- **Dimension of projected space:** The dimension of the feature space is given by the reservoir size, a larger number of neurons in the reservoir may improve the linear separability of the data.

- **Input scaling factor:** That is a parameter for weighting the input patterns.
- **Controllability of the network:** The reservoir matrix controls the recurrent dynamics of the expression (3.1). There are several analyses about the stability of the system considering the singular value and the spectral radius of $\mathbf{w}^{\mathrm{r}}$ [74, 14, 76, 133].
- **Density of the weight matrix of the reservoir:** It is suggested to use around a 20% of non-zero values on the reservoir matrix [74].

There are several variations and extensions of the canonical ESN model. In [50] was introduced a reservoir with leaky neurons, where the reservoir neuron performs a more smooth state update that is controlled by a leaky rate. Reservoir neurons with noise have been evaluated in [101], and recently deep reservoirs have been developed [41, 40].

3.4 Applications of Neural Networks in Biomedical Analysis

Enormous amounts of biomedical data such as omics, images, signal data has been acquired through years. Several techniques were developed for capturing required information from inside of human body. These include radiography, MRI, nuclear medicine, ultrasound, elastography, tomography, EEG, FMRI, etc., which provide biomedical information image or signal format. This gives great potential to the applications in biological and healthcare research and has caught the attention of industry and academia. Figure 3.6 represents this graphically. The following paragraphs provide literature review of ANN in biomedical analysis.

Analysis of biomedical data with computational and mathematical methods provides valuable information which is crucial for diagnosing and treating diseases. One of the most critical challenges in biomedical analysis is the transformation of big biomedical data into valuable knowledge. IBM researches calculated that medical images currently take at least ninty percent of medical data, which makes it the largest data source in the healthcare field.

Evaluation of these various image or signal data requires knowledge in computer science, data science, electrical engineering, physics, mathematics and medicine. Addition-

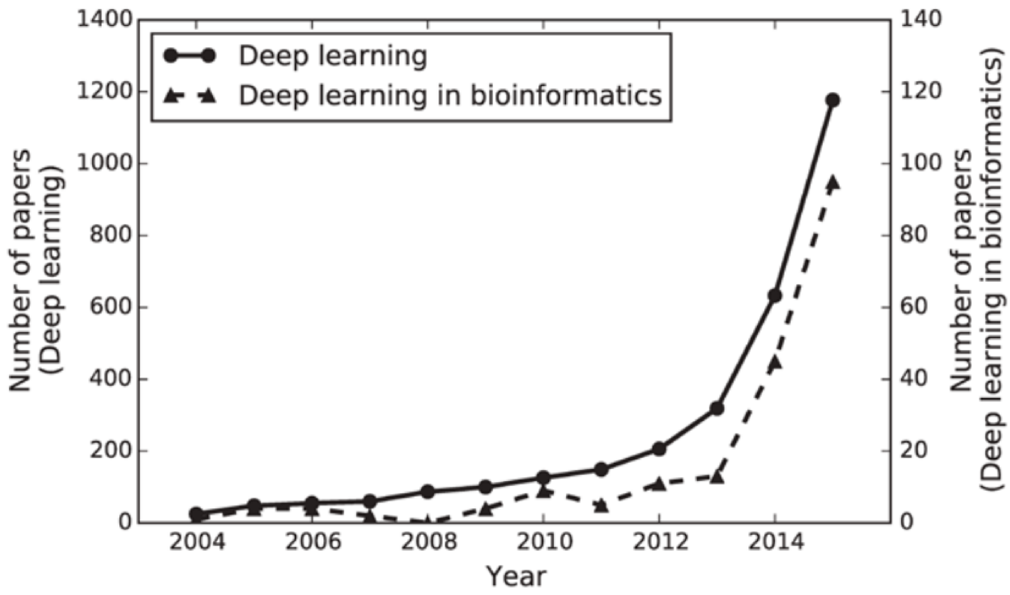

Figure 3.6: Comparison of Deep learning papers and Deep learning in bioinformatics. The figure was taken from [78].

ally, desired performance of classical machine learning algorithms mainly is dependable on data representation called features. This is why, traditionally, features of data designed and selected by expert engineers, and selection of more appropriate features for given task remains difficult. ANNs, especially Deep Neural Networks (DNNs) have solved previous restrictions and major advances has been made in diverse fields, certainly also in biomedical analysis.

In omics research, common input data is raw biological sequences(i.e., DNA, RNA, amino acid sequence) which represents genetic information such as genome, transcriptome, proteome data. DNNs have been applied on this kind of data for predicting structure of protein. For instance, Heffernan et al. [45] and Spencer et al. [117] applied neural network approach to predict protein secondary structure. DNNs have also demonstrated potential in the gene expression regulation area [69, 68, 136, 25] Asgari et al. [5] used DNNs for protein family classification and Fakoor et al. [36] evaluated DNNs to classify the different cancers such as breast cancer , ovarian cancer. However, relatively few studies have used Convolutional Neural Networks (CNNs) in biological sequence problems, they showed strong advantages of CNNs and its potential for future researches [1, 64, 98, 135, 138].

Since RNNs are appropriate deep learning architectures for sequential data has great importance in protein structure prediction [8, 55, 9], gene expression requlation [65, 66, 90], and protein classification [47, 54].

Image processing and analysis in biomedical data is another research area, which applies deep learning techniques for finding hidden patterns in the image–related problems. Main focus areas in biomedical imaging can be categorized as anomaly classification [94, 105, 119, 11], image segmentation(i.e., extracting specific informations such as cellular structures or a brain tumor) [122, 28, 44, 104], pattern recognition [131, 67, 103, 103], and brain decoding [123, 60] in order to interpret human behavior or emotions. Application of DNNs can be found in every category of bimedical imagine. Plis et al. [94] used brain MRIs to classify schizophrenia patients, Xu et al. [131] used DNNs to detect cell nuclei from histopathalogy images, and Van Gerven et al. [123] analyzed MRIs of subjects who are looking at the digital images and classified handwritten digital images based on that data. CNNs also widely used in biomedical imaging studies since they perform very well in general image–related tasks. For example, Roth et al. [105] applied CNNs for classifying anomaly, Cireşan et al. [29] used CNNs for detecting mitosis in breast cancer which is crucial for cancer diagnosis and assessment. CNNs also applied in segmentation [122, 44] and recognition [67, 103]. However, RNNs are not considered a lot in biomedical imaging studies since DNNs or CNNs performs better with image data.

ANNs have great amount of applications in biomedical signal processing that processes recorded electrical activity from human body. Nevertheless, recorded signals generally are noisy, which makes it hard to evaluate. That is why, before using the signal as input in deep learning algorithms, data is pre–processed for improving the results. Biomedical signal processing can be categorized into two groups: brain decoding [2, 51, 52] and anomaly classification [130, 137, 93, 32]. DNNs have been applied in biomedical signal processing in both brain decoding and classification. An et al. [2], Jia et al. [51], and Jirayucharoensak et al. [52] applied DNNs to EEG signals to classify left- and right-hand motor imagery skills and emotion, respectively. Raw EEG signals also have been studied in several researches. For example, Wulsin et al. [130] used both raw EEG signals and extracted features as a input. Although, Zhao et al. [137] applied DNNs over only raw EEG signals to diagnose Alzheimers disease. CNNs are also a popular approach in biomedical signal processing in brain decoding [118, 23] and anomaly classification [80]. EEG signals have been analyzed by several researches, for instance, Stober et al. [118] classified the rhythm type and genre of music which subjects listened to, and Cecotti et al. [23] made classification of characters which is viewed by participants. RNNs are common and suitable deep learning architecture to work with the data and produces promising outcomes, since, biomedical signals are sequential data. Davidson et al. [32] applied Long Short-Term Memory (LSTM) on EEG signals to detect lapses which is modification of RNNs. Additionally, Petrosian et al. [93] used RNNs over EGG signals for predicting seizures. The brain decoding studies can be found in [115].

4 Combinatorial Optimization using Nature-inpired Algorithms

In this chapter, firstly, we give introduction about general Combinatorial Optimization problems in Section 4.1. The topic is briefly described and several approaches are counted for solving such problems. Specially two approaches, SA and GPSO, are considered in detail since we used these methods in this thesis. Section 4.2 discusses motivation, procedure principles, parameters of Simulated Annealing. Finally, in Section 4.3 we cover briefly the main ideas of the Particle Swarm Optimization (PSO) in order to understand the Geometric PSO which is more general approach. Pseudo-codes of both methods are presented in related sections.

4.1 Combinatorial Optimization problems

Combinatorial optimization problems constitute a class of problems whose solutions are discrete or can be transformed into discrete ones. In more formal terms, a general Combinatorial Optimization (CO) problem, $\mathbb{P} = \{I, \{sol(i)\}_{i \in I}, m\}$, can be defined as a minimization or maximization problem that consists of a set of problem instances, I, a set of feasible solutions, $sol(i)$, for every instance $i \in I$, and a function, $m : \{(i, q) | i \in I, q \in sol(i)\} \rightarrow \mathbb{Q}_+$, where $\mathbb{Q}_+$ is the set of positive rational numbers and $m(i, q)$ is the value of solution q for the problem instance i [53]. An optimal solution to an instance of a CO problem is a solution that has maximum (or minimum) value among all other solutions. Well-known CO problems from the area of operations research include feature selection, column subset selection, and various types of p-median [82] and p-center problems.

Optimization problems can be solved by a number of algorithms including linear programming, quadratic programming, gradient methods, stochastic approximation, and by a large family of *metaheuristic* methods. Metaheuristics algorithms are robust, reusable, and adaptive by design. They are frequently employed to solve hard optimization problems that cannot be solved by exact problem–specific algorithms within reasonable time and space constraints. Many metaheuristic methods, often used to solve CO problems, are nature-inspired [99].

Optimization methods such as simulated annealing SA, Tabu Search (TS), Iterated Local Search (ILS), Evolutionary Algorithms (EA), Evolutionary Programs (EP), Greedy Randomized Adaptive Search Procedure (GRASP), Memetic Algorithms (MA), Variable Neighborhood Descent (VND), Genetic Algorithms (GA), PSO, Artificial Bee Colony (ABC), and e.g. Ant Colony Optimization (ACO), draw inspiration from various natural and biological phenomena and emulate successful natural optimization principles to solve practical problems. The rest of following sections present the two techniques applied in our system, which are the SA algorithm and a variation of PSO named as GPSO.

4.2 Simulated Annealing

SA is a stochastic search method successful in solving combinatorial optimization problems. The algorithm is inspired by the annealing process of materials such as metals, glasses, and crystals. Annealing involves heating the material above its melting point in certain time and then slowly cooling it until a firm crystalline structure is created [58]. The structure is during the process gradually re-organized into an optimal material state, identified by the minimal energy configuration.

The SA algorithm is a computational search and optimization method based on the physical/chemical annealing process. Candidate problem solutions are in this algorithm interpreted as physical material states and their costs are linked to the energy of the system. Its iterative search strategy is attempting to minimize material temperature and to find the system state with minimum energy.

The SA is a modification of the Metropolis algorithm [58] and can successfully refrain from getting stuck in the local minima due to the possibility to select uphill moves. At each iteration, current solution maintained by the SA, S^{curr}, is with a probability, p, replaced by a new random *nearby solution*, S^{new}. Let $E(\cdot)$ be the energy (cost) of a problem solution, and let ΔE denote the change in energy obtained by the newly created solution $E(S^{\text{new}}) - E(S^{\text{curr}})$. If ΔE is negative (a downhill step was performed) the change is accepted and S^{new} is used as new current solution in the next iteration. If ΔE is positive (an uphill step was performed), the new solution, S^{new}, is selected with the probability

$$P(\Delta E) = \exp\left(-\Delta E\right)/k_B T), \tag{4.1}$$

where k_B is a control parameter named Boltzmann's constant. This stochastic element allows jumps from local minima to other regions of the search space. The temperature, T, is a control parameter. The initial temperature is high and represents the heating (melting) of the system. It is in each iteration decreased until it reaches an arbitrary *frozen* value T_f. The pseudo-code 1 describes general procedure of the algorithm.

4.3 Geometric Particle Swarm Optimization

The PSO is a population–based optimization algorithm often used to search complex spaces solution spaces [57]. It is a popular optimization method with an excellent record of successful applications in several domains [30]. It is based on the social behavior of a group of particles, called the *swarm*. Each particle in the swarm represents a candidate problem solution. In each iteration, the particles exchange information and *learn* from the best particle in the swarm. In the case of the standard PSO, each particle can be interpreted as a point in multidimensional search space defined by two pieces of information (real vectors): a *position* and a *speed*. Particle positions represent candidate problem solutions and the algorithm finds optimum points by updating the position and the velocity of individual particles. Standard PSO was originally developed to solve numerical optimization problems.

The GPSO is a generalization of the standard PSO that can be used to solve both, numerical and combinatorial optimization problems [83]. Let N be the number of particles in the swarm and let M denote the dimension of the search space. Each particle, i, is denoted by $\mathbf{p}_i \in A^M$, where A is a metric space. Here, we consider $\mathbf{p}$ as a binary string ($A = \{0, 1\}$). A cost function evaluates the solution, represented by the position of each particle. The PSO is iterative and the cost function is evaluated for each (new) particle in every iteration. In a way similar to the operations of Genetic Algorithms, GPSO has a mutation operation that produces a small displacement of one particle. The

Inputs : System information, T_f,$Iter$,$T^{(0)}$, ρ
Outputs: A solution S^{curr}

$i \leftarrow 0$; $T \leftarrow T^{(0)}$;
Compute an initial solution $S^{(0)}$;
$S^{\text{curr}} \leftarrow S^{(0)}$;
while $(T \geq T_f)$ **do**
  while $(i <= Iter)$ **do**
    Select a random nearby solution S^{new};
    if $(Energy(S^{new}) \leq Energy(S^{curr}))$ **then**
      $S^{\text{curr}} \leftarrow S^{\text{new}}$;
    end
    else
      Compute p using expression (4.1);
      if $(rand(0,1) < p)$ **then**
        $S^{\text{curr}} \leftarrow S^{\text{new}}$;
      end
    end
    $i \leftarrow i + 1$;
  end
  Decrease temperature: $T \leftarrow \rho T$;
  $i \leftarrow 0$;
end
Return S^{curr};

Algorithm 1: Simulated Annealing Algorithm.

main operation of the algorithm is position update, performed as a convex combination of three feasible solutions (other particles from the swarm). This operation contains three parameters: w_1, w_2 and w_3. We denote by $\mathbf{p}_i^l(t)$ the best position of the particle i ever found until time t. We denote by $\mathbf{p}^g(t)$ the best position reached by any particle ever found until time t. The position update in a binary search space is given by the following stochastic rule [83]: $\mathbf{p}_i(t+1) = \mathbf{p}_i(t)$ with probability w_1, $\mathbf{p}_i(t+1) = \mathbf{p}_i^l(t)$ with probability w_2, and $\mathbf{p}_i(t+1) = \mathbf{p}_i^g(t)$ with probability w_3.

$$p_i(t+1) = CX((p_i(t), w_1), (p_i^l(t), w_2), (p_i^g(t), w_3)) \tag{4.2}$$

Besides, a mutation operator is applied to each particle in every algorithm iteration [83]. An advantage of the algorithm is that is extremely easy to compute, robust, and has only a small number of parameters: number of particles, mutation rate, the weights w_1, w_2 and w_3. As a consequence, the GPSO algorithm is a good option for being applied in scenarios where computational resources are highly constrained.

Inputs : System information, N,μ mutation rate, t,w_1, w_2, w_3
Outputs: The best position $\mathbf{p}_i^l(t)$

initialization;
$t^0 \leftarrow 0$;
forall *particle i* **do**
    initialise position p_i randomly in the search space
end
$S^{\text{curr}} \leftarrow S^{(0)}$;
while *($t^0 \leq t$)* **do**
    forall *particle i* **do**
        set personal best $p_i^l(t)$ as best position found so far by the particle;
        set global best $p_i^g(t)$ as best position found so far by the whole swarm
    end
    forall *particle i* **do**
        Update position of particle $p_i(t+1)$ using expression (4.2);
        mutate $p_i(t+1)$
    end
end
Return S^{curr};

Algorithm 2: Geometric Particle Swarm Optimization Algorithm.

5 Nature-Inspired system for Mental State Recognition

This chapter contains the main contributions of this thesis. We start with a formal computational-mathematical formalization of the problem studied in our thesis 5.1. As a beginning stage we designed experiments for collecting data from humans. We present the insights of data collection procedure in Section 5.2. The experimental protocol is defined for data collection stage and was explained to every subject beforehand. Details of experimental protocol are discussed for a better understanding of subsequent sections. This section also covers common pre-processing method since all collected data was pre-processed before the evaluation process.

The main technical problem during data collection procedure is precise mapping between the EEG signals and collected data. The Section 5.3 presents the solution for this problem using Openvibe software platform. We give information about Openvibe packages in detail and how modification was done to these packages with Lua scripting language in order to get perfect synchronized data. We evaluated the collected data with different combinations of parameters based on discussed approaches in Chapter 3 and Chapter 4. The outline of evaluation process presented in Section 5.4. Finally, the obtained results are covered with figures and tables in Section 5.5.

5.1 Formalization of our problem

In a classification problem, the goal is to define a model $\varphi(\cdot)$ for labeling an outcome variable based on a set of input features [43]. The parameters of the system are adjusted using a set of examples named training samples, with the condition that the generated input-output mapping should be capable of "well" mapping any other unknown samples. The intersection between the training set and testing set is empty. In the case of this thesis, the outcome variable is binary, and the input features contain information from the brainwaves.

Let $\mathbf{a}(t)$ be a N_{a}-dimensional brainwave data collected at time step t, and let $y(t)$ be a binary variable representing the mental state (concentration versus relaxation). A quantitative measure called *cost function* measures the quality of the learning model, in our case we evaluate the model over the examples in S^{test} using the Mean Squared Error (E_{MSE}) and Accuracy (E_{ACC}):

$$E_{MSE} = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}(t) - y(t))^2, \tag{5.1}$$

and

$$E_{ACC} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\hat{y}(t)==y(t)}, \tag{5.2}$$

where $\hat{y}(t)$ denotes the prediction for the input $\mathbf{a}(t)$ and $\mathbb{1}_A$ is the indicator function of a set A. At first, we do a preliminary setting of the global parameters of the classifier system. We empirically find the best values of the reservoir size N_{x}, the spectral radius $\rho(\mathbf{w}^{\mathrm{r}})$ and the leaky rate α.

The model selection schema is the usual in the area, the data is divided in two disjoint subsets: training and testing sets. For finding the *best* configuration of the classifier we used all the input features of the collected data. We consider the *best* configuration of global parameters the one that obtains the minimum testing error in average for all subjects. Once we obtain the *best* global parameters of an ESN, let denote them N_{x}^*, ρ^* and α^*, we apply the GPSO and SA methods for automatically selecting a combination of features that optimizes the learning system. Next, we apply a searching algorithm for selecting the best combination of input features.

Our original system contains 12 input signals. However, it is well-known that to use all the signals may insert noise in the training process. Hence, we can obtain a classifier with lower accuracy than a trained classifier with well-selected EEG channels. This is due to the fact that the signals produce noise among them. Therefore, we used evolutionary techniques for finding the best combinations of features for being considered in the classifier model. A similar approach was presented in [15].

The feature selection schema is as follows. Without loss of generality we enumerate the input features by $\{1, \ldots, N\}$, where N is the number of brainwave variables. Therefore, the searching space of our problem is $\{0, 1\}^N$. The optimization problem has solutions with the form $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ where $s_i = 0$ represents that the input feature i is omitted as source of the classification method, and $s_i = 1$ represents that the signal is an input of the classification method. We consider as a cost function the accuracy of the learning tool (E_{ACC}). The problem is to find the vector $\mathbf{s} \in \{0, 1\}^N$ such that E_{ACC} is maximized when the system is used for the concentration-relaxation identification.

5.2 Data collection

We designed an experiment for recording human brainwaves using EEG channels placed on the scalp. The EEG allows to build a portable system with a very high temporal resolution. In addition, the device costs are significantly lower than those of most other brain information techniques. We collect the brain information using a NeuroSky Mindset [48], which provides data in forms of the delta, theta, alpha, beta, and gamma signals. Delta signal provides information about brain disorders, theta signal gives data about meditation of subject, alpha and beta signals represents mental activity or highly engaged mind and gamma signal is related to cognitive or motor functions of subject [4].

The experimental protocol is as follows. The data was collected during the sessions with 4 healthy subjects aged from 23 to 38, all of them were right handed. A subject was sitting in a comfortable chair located one meter from a 17" monitor. The subject was performing instructions displayed on the screen. We created two type of commands that were displayed on the screen:

- **Action *A* - Mental Concentration:** We asked to the subject to read and to write in hand the displayed text on the screen. We displayed a random selected page of the Franz Kafka's book: "The trial" [56].
- **Action *B* - Relaxation:** We ask to the subject to relax until a new command appears in the monitor. We asked to the subject to do not change the head and body position during the relaxation activity. Then, the subject stay in the seat watching the monitor.

An important technical issue in this type of experiments is to create an exact relationship between the EEG signals and the subject's actions (without delays). We developed a Lua script for displaying the images of the Kafka's book. The system synchronization was made using Openvibe software platform [62], in this synchronization we are able to automatically tag the intervals of the EEG signals with the subject actions.

The experimental timeline is as follows. We do not consider the first 5 seconds. In this period, we show *GET READY* command on the screen Then, there are 20 seconds where we display a command on the screen, for subject to familiarize himself/herself,although,we do not use the collected data in the learning system. Next, the relaxation actions have a duration of 20 seconds and the mental concentration actions have a duration of 30 seconds. The actions are alternating among them. The total time of the experiment was 300 seconds. Figure 5.1 shows a general view of the whole experiment.

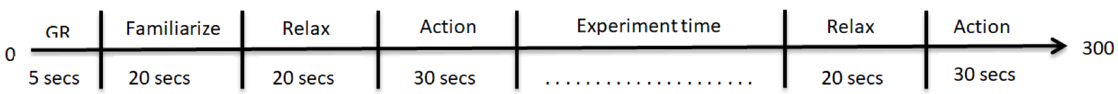

Figure 5.1: The general view of experiment period

We observe the produced brainwaves for every subject in real time for detecting any conductance failure between scalp of subjects and electrode and take notes. In no cases we intervened in experiment period even there was conductance failure. Any intervention in experiment could make subjects nervous and we wanted from them to follow the instructions those were displayed without any distraction. In Figure 5.2, an example of our experiments is presented. We can see the portable Neurosky's Mindset device on the scalp of the subject and how she performs the commands displayed in the screen.

In the next Figure 5.3, an example of the text that is displayed in the screen is presented. The figure shows an example of the random selected page from the text of the Kafka's book.

Figure 5.2: One of the subjects who goes through the experiment.

door of which was already wide open. K. knew very well that this room had recently been let to a typist called 'Miss Burstner'. She was in the habit of going out to work very early and coming back home very late, and K. had never exchanged more than a few words of greeting with her. Now, her bedside table had been pulled into the middle of the room to be used as a desk for these proceedings, and the supervisor sat behind it. He had his legs crossed, and had thrown one arm over the backrest of the chair.

In one corner of the room there were three young people looking at the photographs belonging to Miss Burstner that had been put into a piece of fabric on the wall. Hung up on the handle of the open window was a white blouse. At the window across the street, there was the old pair again, although now their number had increased, as behind them, and far taller than they were, stood a man with an open shirt that showed his chest and a reddish goatee beard which he squeezed and twisted with his fingers. "Josef K.?" asked

Figure 5.3: An example of randomly selected page, which is displayed to the subject for creating writing stimulation.

5.3 Design of the system

In the following we present some details about the Openvibe scenario and the synchronization implementation. Openvibe software platform [100] is used for working with the device, designing the BCI project related to the research and recording all data associated to the actions of subject. Openvibe provides a number of packages which can be used to create scenarios for EEG devices. An scenario is predefined stages of experiment which was designed with the help of Openvibe. We created our own scenario with the help of Openvibe packages and Lua scripting language for synchronization purpose. Our main objective was to start and end subject actions and data recording at the same time. By this way, data related to subject would be flawlessly synchronized with the time interval of each action because time both packages reference to time of Openvibe. Packages of Openvibe which are used for our experiment are shown in Figure 5.4, they were the following ones:

- **Acquisition client:** the function is to receive brain waves from the device. It can generate 5 type of data from for sending to other packages. We used signal stream output for observing incoming signals from subjects brain. The main purpose to identify electrode disconnections. Besides this output was also used for recording data. We also used stimulation output for starting image display
- **Signal display:** The package receives signal stream and displays it in real time
- **CVS file writer:** The package receives signal stream and records it to the predefined cvs file
- **Display cue image:** The package receives stimulation for starting image display. This package was modified for the synchronization. It has been assigned Lua file and selected images.

The implementation was made as follows. Lua script which was written for synchronization procedure works in simple manner. It starts to run when experiment starts and sends stimulations to Display Cue Image package of Openvibe. Stimulations are selected randomly and sent based on time flow during experiment. Due to this simplicity, it works in $O(n)$ time and does not require a lot of memory. The pseudo-code 3 describes the implementation of synchronization

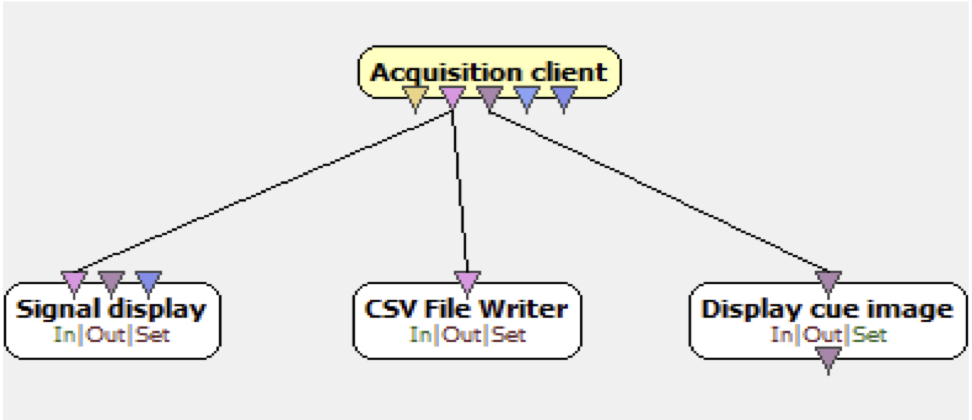

Figure 5.4: Diagram of the modules programmed in LUA.

```
Experiment time is 300 seconds;
while (time < 300) do
    if (time == 0) then
        Get Ready stimulation;
        time = time + 5;
        Get Ready stimulation;
        First activity for familiarizing stimulation;
        time = time + 20;
    end
    Relax stimulation;
    time = time + 20;
    Generate random number between 2 and 20;
    Activity stimulation based on this number;
    time = time + 30;
end
Finishing experiment stimulation;
```

Algorithm 3: Procedure for sending stimulation to Cue image package

5.4 Experimental settings

Firstly, all collected data was pre–processed in order to remove the signal noise. The sample collected EEG signals is given in Figure 5.5.

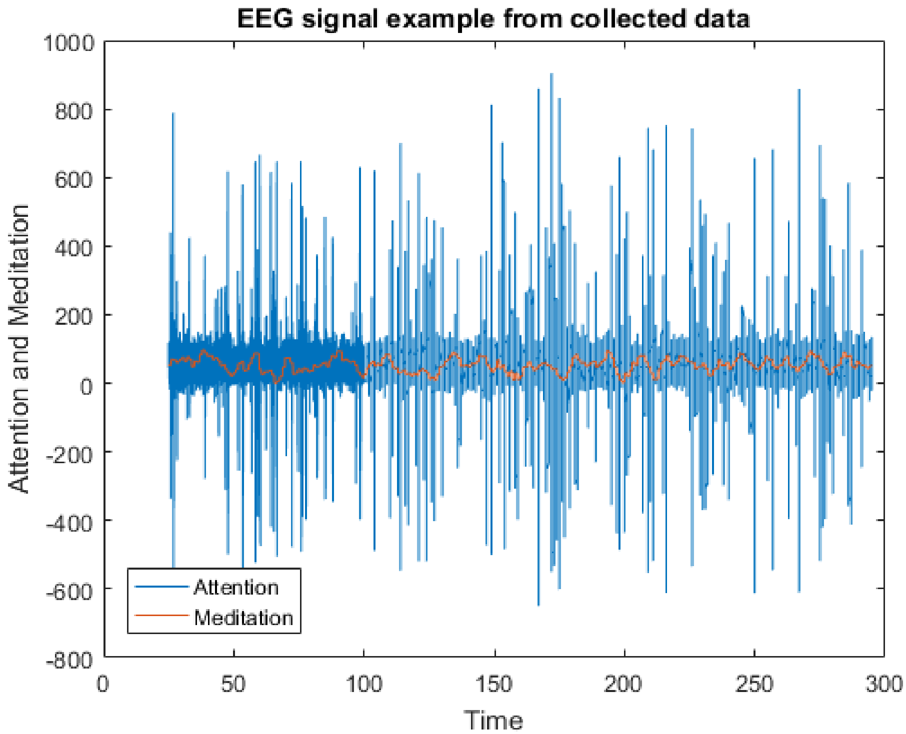

Figure 5.5: Captured raw EEG signal. Only attention and meditation signal were presented.

The first 25 seconds were cut because these time was for familiarizing subject to the tasks. We also cut last 5 seconds from data. Original data was cut by 30 seconds totally. Besides, this experiment was conveyed with the help of five subject, we decided to use only four data set. There occurred conductance failure between electrode and forehead of the last subject while experiment.

All the collected data was normalized in the range of $[0, 1]$. The normalization was made because of applied approach to this data which results in binary classification. In each column we applied the following formula, given an instance $x^{old}(k)$ we compute for each coordinate the normalized data $x^{new}(k)$:

$$x_j^{new}(k) = \frac{(x_j^{old}(k) - min_j)}{(max_j - min_j)}, \tag{5.3}$$

where max_j and min_j are the maximum and minimum of column j respectively.

The outliers are identified using the following criteria: we consider a sample $\mathbf{a}(k)$ as artefact if any coordinate of the vector $\mathbf{a}(k)$ doesn't belong to the interval: $(\hat{a_j} - 3\sigma_j, \hat{a_j} + 3\sigma_j)$, where $\hat{a_j}$ is the mean of the coordinate input variable j, and σ_j is the standard deviation of this variable j. In addition, we verified that this rule doesn't exclude more than the 7% of the total data. Then, the data without the identified outliers is normalized in the range $[0, 1]$. In total there are 12 input features from the Neurosky device, and the binary output variable that corresponds to the subject actions. The collected dataset and the source codes are available to the community in [13].

As usual in the context of learning, we divide the learning set in training and testing. We compute the output weights using the training data. We present the accuracy only obtained on the testing data. The training set contains the 70% of the data, the size of the training set is composed by 90994 samples. We sort randomly the learning set of each subject as follows. We divide the original data in mini-batches (time windows of 100 time steps) $\mathcal{L} = [\Delta_1, \ldots, \Delta_M]$ where Δ_i contains 100 instances. The new learning dataset is given by the permutation of the mini-batches, for instance a permutation example: $\mathcal{L} = [\Delta_j, \ldots, \Delta_M, \Delta_1, \ldots \Delta_i]$. According our empirical results, it is better to randomly sort the learning set of each subject using the mini-batches instead of using the original sequential data.

The predictions of the classifier system are real numbers. Thus, we discretize the predictions in $\{0, 1\}$ (if $\hat{y}(t) > 0.5$ then we assign $\hat{y}(t) = 1$, and if $\hat{y}(t) <= 0.5$ then we assign $y(t) = 0$). We evaluate the model with different global parameters. We created a grid of values for (N_{x}, ρ) and we evaluate all the combinations when the values are $N_{\mathrm{x}} \in \{25, 50, 75, 100, 150, 200, 300, 500, 1000, 1500\}$, $\rho(\mathbf{w}^{\mathrm{r}}) \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. We considered two RC models, the standard ESN model and the ESN with leaky-neurons introduced in expression 3.3 [50]. We evaluated the model with two possible leaky rates $\{0.5, 0.9\}$. The regularization factor of the linear regression was 0.01.

When we applied SA as feature selection technique we consider a *nearby solution* such that its Hamming distance with the current solution is less than or equal to 3. In other words, the binary vectors S^{curr} and S^{new} differ each other on a maximum of 3 bits. The temperature is given by the number of iterations in the SA algorithm, and we set up a Boltzmann factor empirically we used $k_B = 15$. The stop condition T_f is given by the 500 iterations. The algorithm GPSO has few parameters, we set up that parameters using the suggestions presented in [83]. We defined a swarm with 20 particles, and we evaluated the following two configurations of the GPSO weights (w_1, w_2, w_3) equal to the values $(1/6, 5/12, 5/12)$ and $(1/3, 1/3, 1/3)$.

5.5 Experimental results

In order to control the spectral radius of the reservoir, the recurrent weight matrix is randomly initialised and scaled as follows: $\mathbf{w}^{\mathrm{r}} \leftarrow (\beta/\rho(\mathbf{w}^{\mathrm{r}}))\mathbf{w}^{\mathrm{r}}$, where β is a constant in the range $(0, 1]$. The testing data of the subject 1 contains 39111 samples. We run 30 trials (different reservoir random initializations) for each of the combinations of (N_{x}, ρ). The figures 5.6 and 5.7 show the averages of the E_{MSE} values among those 30 trials obtained for the subject 1. The first figure presents the results obtained by a standard ESN and the second one the results obtained by an ESN with leaky-neurons with $\alpha = 0.5$. Figure 5.11

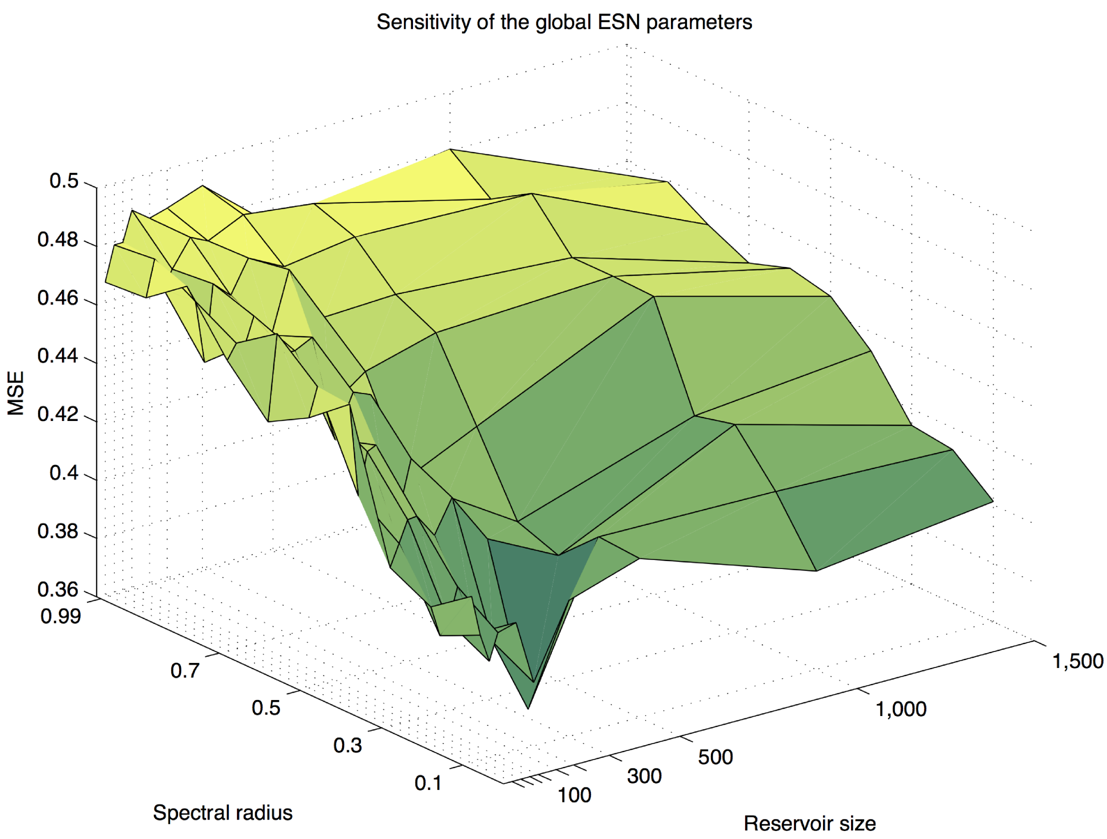

Figure 5.6: Example of the sensitive analysis of the main global parameters of the standard ESN model when the data corresponds to the subject 1.

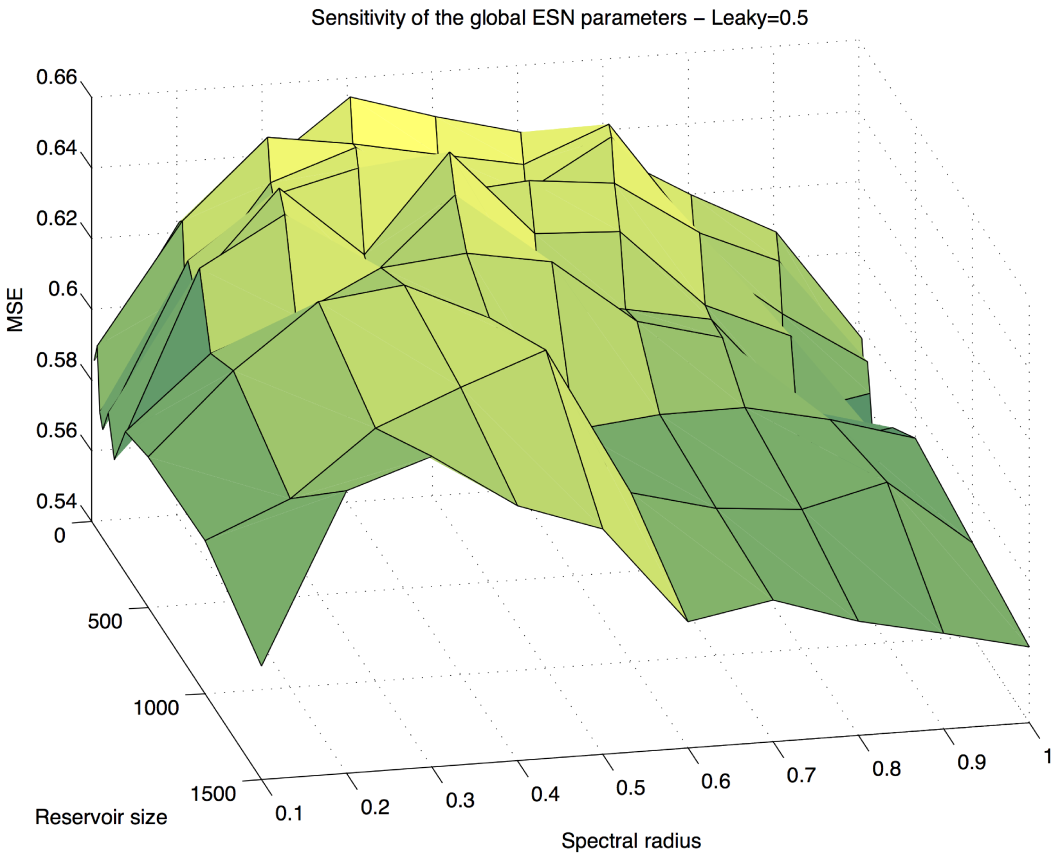

Figure 5.7: Example of the sensitive analysis of the main global parameters of the α-ESN model with leaky rate $\alpha = 0.5$ when the data corresponds to the subject 1.

shows an example of the ESN prediction using all the input features in a specific time interval with discretization in $\{0, 1\}$. A similar graphic is presented in figure 5.9, in this case the there are the ESN predictions in a specific time interval without a discretization.

Table 5.1 presents the solution of the combinatorial optimization problem. The found

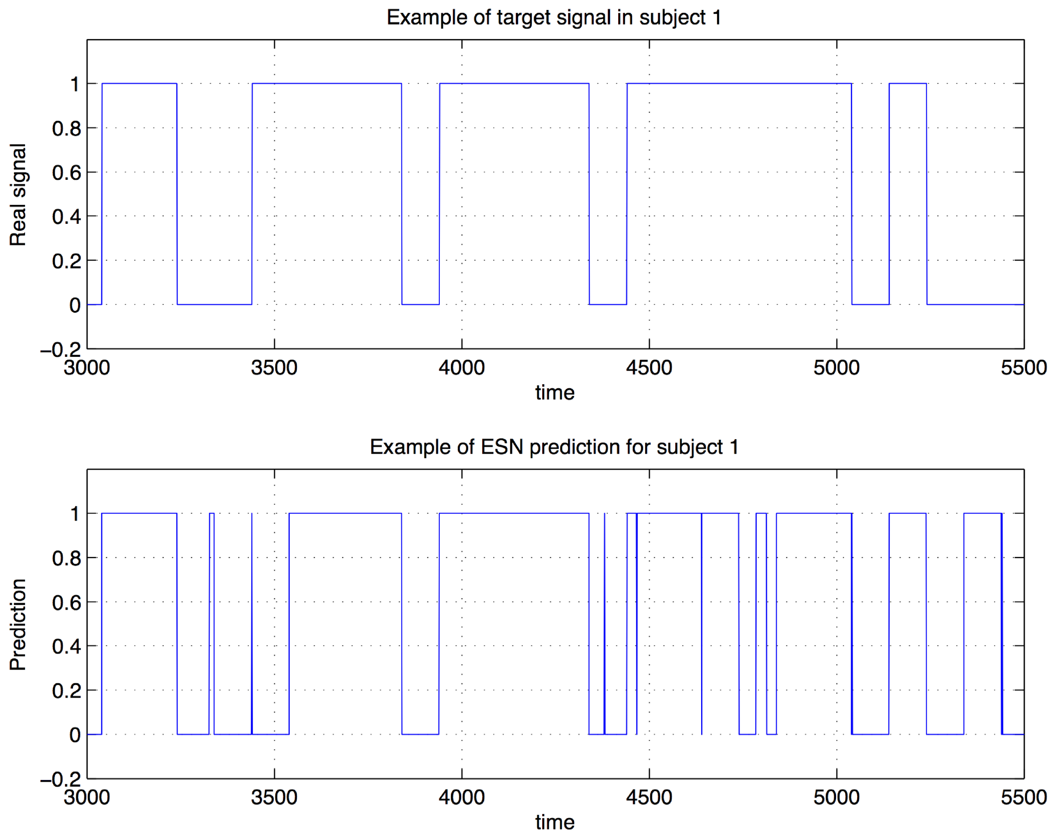

Figure 5.8: Example of the target signal and the ESN prediction in $\{0, 1\}$ using all the input features.

solution with SA obtain an accuracy of 78.46%. The solution with $GPSO_1$ has an accuracy of 79.47% and the solution obtained with $GPSO_2$ has an accuracy of 78.35%. Figure 5.10 presents the evolution of the accuracy obtained by the system using different combination of input features. The green curves show the evolution of the accuracy with SA, and the blue and red curves show the accuracy evolution using the two configurations of GPSO ($GPSO_1$ and $GPSO_2$). Note that the reservoir initialization is random, then the accuracy can be different even when the model contains a same input features. For that reasons,

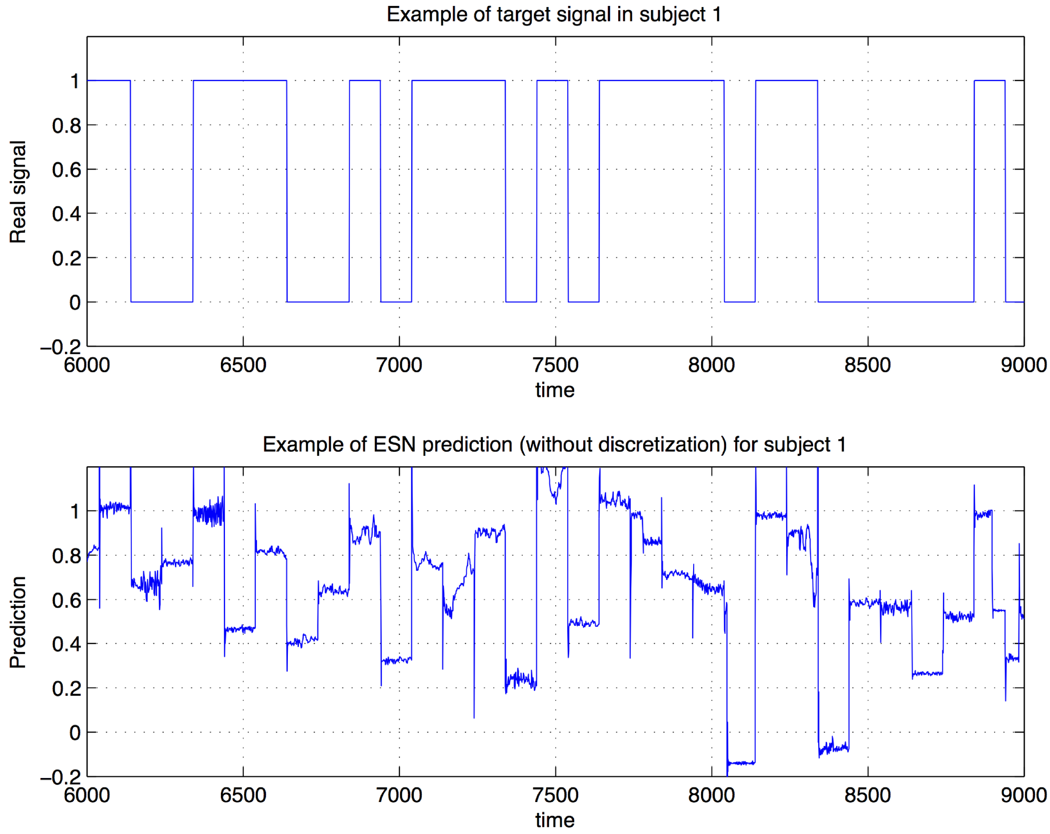

Figure 5.9: Example of the target signal and the ESN prediction in $[0, 1]$ using all the input features.

SA	0	1	1	1	1	1	1	1	0	0	1	1
$GPSO_1$	0	1	1	1	1	1	1	1	0	0	1	1
$GPSO_2$	0	1	1	1	1	0	1	1	0	0	1	1

Table 5.1: Input feature selection using SA and GPSO. Each column indicates whether the input feature is used or not. Last column present the final accuracy for the case of subject 1. $GPSO_1$ has the weights $(1/6, 5/12, 5/12)$ and $GPSO_2$ has weights $(1/3, 1/3, 1/3)$.

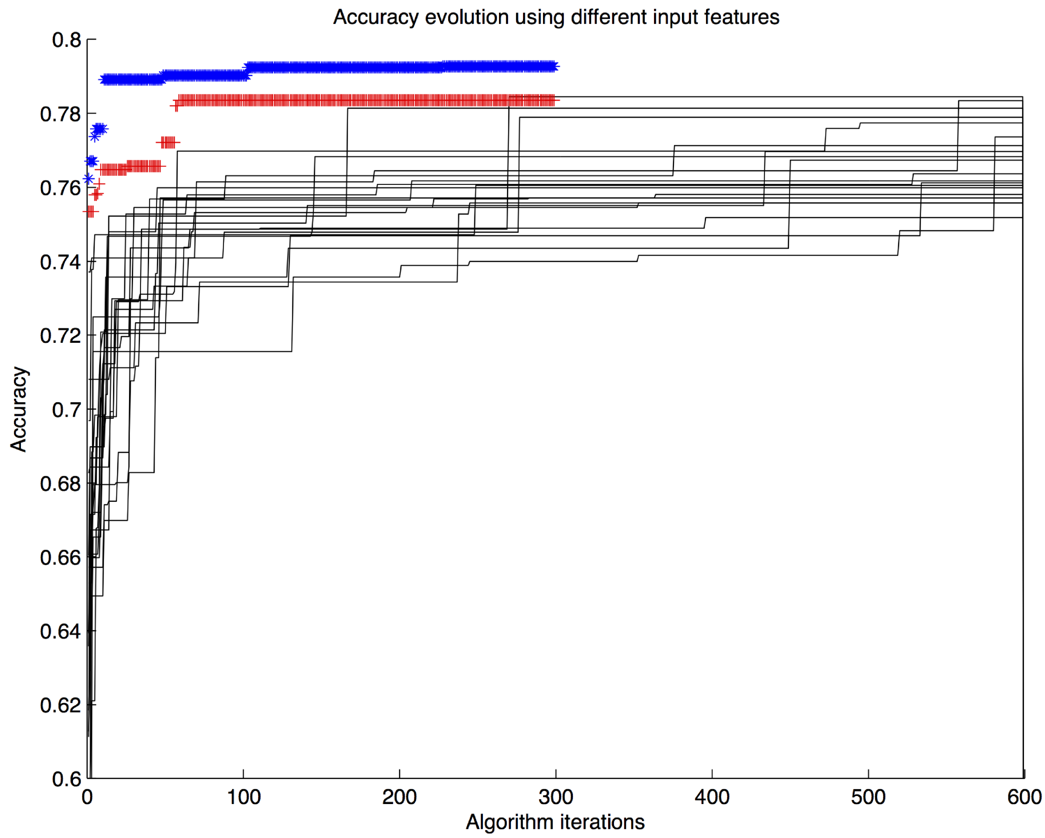

Figure 5.10: Evolution of the accuracy using different number of input features for the subject 1. The green curves show the evolution of the accuracy with SA. Blue and red curves (with starts (*) and plus (+) dots) show the accuracy evolution using $GPSO_1$ and $GPSO_2$ respectively.

the starting points in the curves are different in each case. We present 20 curves in the case of the SA due to the fact that each of the experiments starts the algorithm using a different initial solution. On the other hand, the GPSO method evaluates the searching space using a population, then several solutions are evaluated in parallel. Figure 5.10 shows the accuracy obtained by the best global solution of the swarm. Note that GPSO starts with a higher accuracy than the SA algorithm, that is due to the fact that we show the best global solution of the swarm.

The tables below represents the obtained results for each subject. We can see that most often the model with a leaky rate parameter doesn't obtain a better accuracy. The selected features according to the optimization techniques has been done using only the data of subject 1. The RC model contains a reservoir with 300 neurons and $\rho = 0.8$. The first column indicates the type of feature selection. There are three type of feature selections: All as shown in Table 5.2 (we used all the input features), Sol_1 refers to the selected features by SA and GPSO_1 as indicated in Table 5.3 and Sol_2 refers to the solution obtained by GPSO_2 as demonstrated in Table 5.4. From the tables, we can see that the optimization techniques improve the accuracy. The GPSO_1 method obtains high accuracy for the subject 1 (we used the training and testing data of subject 1 for feature selection), on the other hand it has less generalization ability than the solution presented by GPSO_2. We evaluated the selected features (according to the subject 1) in the other subjects. In that case, we found that the schema with a better generalization ability was the obtained with the GPSO_2. We can see that the obtained accuracy depends of the subject. That is common in the area due to the type of data. In spite of that, the final obtained accuracy is around 79%, what is *acceptable* due to the characteristics of the input information. The detailed description of each table can be found on the below.

In the table 5.2, the first column shows that we used all feature of collected data in the learning model. The first rows presents the accuracy for every subject‘s data which was evaluated with the learning model. We can see that the learning model has overall well performance with the data of subject 2 while it performs poorly when applied on the data of subject 3. Note that, the feature selection was done only using the data of subject 1, although, the model works better on the data of subject 2 and 3.

The first column, in the table 5.3, shows that we used only selected feature with SA algorithm (it was the same solution with GPSO_1) from collected data in the learning model. The first rows presents the accuracy for every subject‘s data which was evaluated

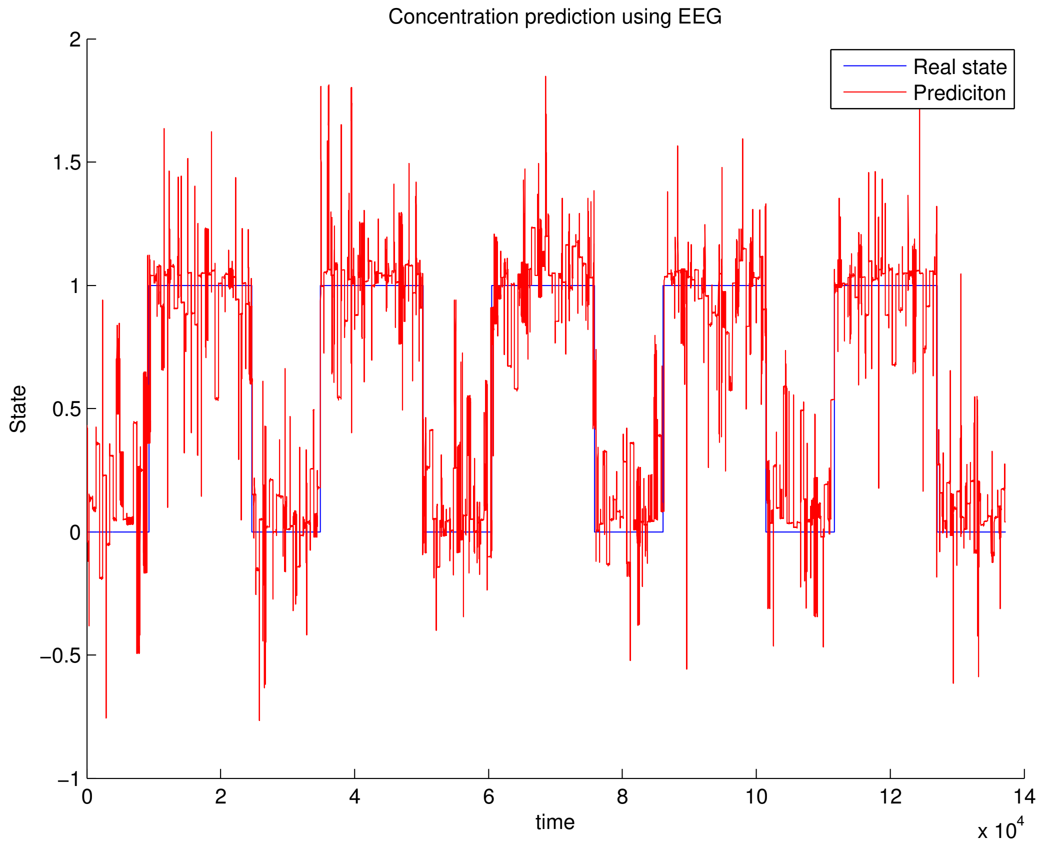

Figure 5.11: Example of the mental state prediction with application of the learning model over EEG signal data of subject 2. All the input features were used in the evaluation process.

	Model	S1	S2	S3	S4
All	ESN	0.7819	0.7911	0.7161	0.7944
All	α-ESN$_{\alpha=0.5}$	0.7898	0.8013	0.7363	0.7955
All	α-ESN$_{\alpha=0.9}$	0.7362	0.7652	0.6713	0.7291

Table 5.2: Test accuracy results for each subject with all input features. The RC model contains a reservoir with 300 neurons and $\rho = 0.8$.

with the learning model. We can see that the learning model has demonstrated well performance with the data of subject 1 , in contrast, it performs poorly when applied on the data of subject 3 and 4. Since, the feature selection was done only using the data of subject 1, the learning model corresponds very–well to the data of subject 1. It gives average results when it was applied on the data of subject 2.

	Model	S1	S2	S3	S4
Sol_1	ESN	0.8112	0.7760	**0.7524**	0.7559
Sol_1	α-$ESN_{\alpha=0.5}$	0.7849	0.7992	0.7381	0.7490
Sol_1	α-$ESN_{\alpha=0.9}$	0.7561	0.7625	0.6884	0.6977

Table 5.3: Test accuracy results for each subject with only SA selected input features. The RC model contains a reservoir with 300 neurons and $\rho = 0.8$.

Lastly, in the table 5.4, the first column shows that we used only selected feature with $GPSO_2$ algorithm from collected data in the learning model. The first rows presents the accuracy for every subject‘s data which was evaluated with the learning model. We can see that the learning model has better findings overall when we compare it with other outcomes. The learning model performance well with the data of subject 1, 2 and 4 and is satisfactory. However, it shows performance below average when applied on the data of subject 3. Note that, the feature selection was done only using the data of subject 1, the learning model corresponds very-well to the data of subject 1.

Sol_2	ESN	0.8085	0.8070	0.7269	**0.8168**
Sol_2	α-$ESN_{\alpha=0.5}$	**0.8161**	**0.8109**	0.7347	0.7950
Sol_2	α-$ESN_{\alpha=0.9}$	0.7473	0.7760	0.6934	0.8008

Table 5.4: Test accuracy results for each subject with only SA selected input features. The RC model contains a reservoir with 300 neurons and $\rho = 0.8$.

6 Conclusions and Future Work

In this thesis we develop an intelligent system for estimating the mental concentration. Mainly, our thesis cover the following areas: Brain Computer Interface, Neural Networks and metaheuristic techniques. We present an application of metaheuristics and Neural Networks (NNs) for computing the human concentration during specific activities. Our primary and basic goal was to develop a portable, non-invasive and robust device. The brain information was collected with NeuroSky Mindset, which collects few EEG signals and it has a low cost in the market. In addition, we used very fast and robust machine learning methods from the Reservoir Computing (RC) family. Specifically, we used the Echo State Network (ESN) model and an ESN variation in which the neurons are smoothed with a leaky factor. Besides, we evaluated the performance of two metaheuristics as feature selection tools. One is based on a local search (Simulated Annealing (SA)), and another one is a population based method (Geometrical Particle Swarm Optimization (GPSO)). The metaheuristics have been used for solving a combinatorial optimization problem, which was to find the best combination of input features for the classification tool. The selected optimization and learning techniques have several advantages. They have few global parameters, low computational cost, they are robust, and we obtained good performance (according to the characteristics of the device). Thus, it is possible in few seconds to adjust the model parameters according to each subject.

We were able to increase the accuracy of the system by applying the automatic feature selection. The GPSO technique reached the best combination of input features faster than the SA model. The obtained results are promising, independently of the person the reached accuracy of our system was around 80%.

The thesis open several new possibilities of research. For example, it is possible to pre-process the EEG signals before to use the Machine Learning tools. For instance, it is possible to use Independent Component Analysis (ICA) and some signal filters. Maybe, it can improve the performance. In addition, we are interesting in analyzing in more detail the generalization ability of the proposed system, we would like to train the system in a group of persons and to evaluate it in another group. Furthermore, other metaheuristic techniques can be evaluated.

The main contributions of this thesis have been already accepted in two major international conferences in the area of ANNs and Evolutionary Computation. They were: 25th European Symposium On Artificial Neural Networks (ESANN), Computational Intelligence and Machine Learning, Bruges, Belgium, 25-27 April 2018, and in the bi-annual IEEE World Congress on Computational Intelligence (IEEE WCCI), Rio de Janeiro, Brazil, 8-13 July 2018.

Bibliography

[1] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey. Predicting the sequence specificities of dna- and rna-binding proteins by deep learning. *Nature biotechnology*, 33(8):831838, August 2015.

[2] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He. A deep learning method for classification of eeg data based on motor imagery. In D.-S. Huang, K. Han, and M. Gromiha, editors, *Intelligent Computing in Bioinformatics*, pages 203–210, Cham, 2014. Springer International Publishing.

[3] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang. Clinical study of neurorehabilitation in stroke using eeg-based motor imagery brain-computer interface with robotic feedback. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 5549–5552, Aug 2010.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Ambient assisted living and home care*, pages 216–223. Springer, 2012.

[5] E. Asgari and M. R. K. Mofrad. Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLOS ONE*, 10(11):1–15, 11 2015.

[6] D. Bacciu, P. Barsocchi, S. Chessa, C. Gallicchio, and A. Micheli. An experimental characterization of reservoir computing in ambient assisted living applications. *Neural Computing and Applications*, 24(6), 2014.

[7] C. M. Bagchi S. An easy-to-adopt approach for regular and routine monitoring of the consciousness level of human brain of stayed alone sick person. *Sensing Technology (ICST)*, 6, 2012.

[8] P. Baldi, S. Brunak, P. Frasconi, G. Soda, and G. Pollastri. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics*, 15(11):937–946, 1999.

[9] P. Baldi, G. Pollastri, C. A. F. Andersen, and S. Brunak. Matching protein b-sheet partners by feedforward and recurrent neural networks. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, pages 25–36. AAAI Press, 2000.

[10] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive computing*, pages 1–17. Springer, 2004.

[11] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan. Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis*, volume 9414, page 94140V, Mar. 2015.

[12] M. M. Barbosa A.O., Achanccaray D.R. Activation of a mobile robot through a brain computer interface. *Robotics and Automation (ICRA)*, 2010.

[13] S. Basterrech. Availability of data and materials. http://cs.felk.cvut.cz/en/page/sebastian-basterrech.

[14] S. Basterrech. Empirical analysis of the necessary and sufficient conditions of the echo state property. In *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017*, pages 888–896, 2017.

[15] S. Basterrech, P. Bobrov, A. Frolov, and D. Husek. Nature-inspired Algorithms for Selecting EEG Sources for Motor Imagery Based BCI. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, editors, *Artificial Intelligence and Soft Computing*, volume 9120 of *Lecture Notes in Computer Science*, pages 79–90. Springer International Publishing, 2015.

[16] S. Basterrech and G. Rubino. Random Neural Network as Supervised Learning Tool. *Neural Network World*, September 2015. Preprint, estimated publication date: 2016.

[17] S. Basterrech, V. Snásel, and G. Rubino. An Experimental Analysis of Reservoir Parameters of the Echo State Queueing Network Model. In *Innovations in Bio-inspired*

Computing and Applications - Proceedings of the 4th International Conference on Innovations in Bio-Inspired Computing and Applications, IBICA 2013, August 22 -24, 2013 - Ostrava, Czech Republic, pages 13–22, 2013.

[18] L. Bi, X. A. Fan, and Y. Liu. Eeg-based brain-controlled mobile robots: A survey. *IEEE Transactions on Human-Machine Systems*, 43(2):161–176, March 2013.

[19] P. Bobrov, A. A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, and A. Zhavoronkov. Brain-computer interface based on generation of visual images. *PLOS ONE*, 6(6):1–12, 2011.

[20] L. Bonnet, F. Lotte, and A. Lcuyer. Two brains, one game: Design and evaluation of a multiuser bci video game based on motor imagery. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):185–198, June 2013.

[21] J. B. Butcher, D. Verstraeten, B. Schrauwen, C. R. Day, and P. W. Haycock. Reservoir Computing and Extreme Learning Machines for Non-linear Time-series Data Analysis. *Neural Networks*, 38:76–89, feb 2013.

[22] S. K. Card, A. Newell, and T. P. Moran. *The Psychology of Human-Computer Interaction.* L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1983.

[23] H. Cecotti and A. Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):433–445, March 2011.

[24] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun. Emotion assessment: Arousal evaluation using eeg's and peripheral physiological signals. In B. Gunsel, A. K. Jain, A. M. Tekalp, and B. Sankur, editors, *Multimedia Content Representation, Classification and Security*, pages 530–537, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[25] Y. Chen, Y. Li, R. Narayan, A. Subramanian, and X. Xie. Gene expression inference with deep learning. *Bioinformatics*, 32(12):1832–1839, 2016.

[26] H.-S. Chiang, K.-L. Hsiao, and L.-C. Liu. Eeg-based detection model for evaluating and improving learning attention. *Journal of Medical and Biological Engineering*, Dec 2017.

[27] A. Choppin. EEG-Based Human Interface for Disabled Individuals: Emotion Expression with Neural Networks. Master's thesis, Tokyo institute of technology, 2000.

[28] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2843–2851. Curran Associates, Inc., 2012.

[29] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 411–418, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[30] M. Clerc and J. Kennedy. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *Evolutionary Computation, IEEE Transactions on*, 6(1):58–73, Feb 2002.

[31] J. L. Contreras-Vidal, A. Presacco, H. Agashe, and A. Paek. Restoration of whole body movement: Toward a noninvasive brain-machine interface system. *IEEE Pulse*, 3(1):34–37, Jan 2012.

[32] P. R. Davidson, R. D. Jones, and M. T. R. Peiris. Eeg-based lapse detection with high temporal resolution. *IEEE Transactions on Biomedical Engineering*, 54(5):832–839, May 2007.

[33] M. Doppelmayr, T. Finkenzeller, and P. Sauseng. Frontal midline theta in the pre-shot phase of rifle shooting: Differences between experts and novices. *Neuropsychologia*, 46(5):1463 – 1467, 2008.

[34] D. G. Duru, A. D. Duru, D. E. Barkana, O. Sanli, and M. Ozkan. Assessment of surgeon's stress level and alertness using eeg during laparoscopic simple nephrectomy. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 452–455, Nov 2013.

[35] C. W. N. F. C. W. Fadzal, W. Mansor, and L. Y. Khuan. Review of brain computer interface application in diagnosing dyslexia. In *2011 IEEE Control and System Graduate Research Colloquium*, pages 124–128, June 2011.

[36] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber. Using deep learning to enhance cancer diagnosis and classification. *Proceedings of the ICML Workshop on the Role of Machine Learning in Transforming Healthcare*, 06 2013.

[37] A. A. Frolov, D. Husek, and P. Bobrov. Comparison of four classification methods for brain computer interface. *Neural Network World*, 21(2):101–115, 2011.

[38] A. A. Frolov, D. Husek, P. Bobrov, A. Korshakov, L. Chernikova, R. Konovalov, and O. Mokienko. Sources of EEG activity most relevant to performance of brain-computer interface based on motor imagery. *Neural Network World*, 22(1):21–37, 2012.

[39] A. A. Frolov, D. Husek, P. Bobrov, O. Mokienko, and J. Tintera. Sources of electrical brain activity most relevant to performance of brain-computer interface based on motor imagery. In *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, pages 175–193. InTech, 2013.

[40] C. Gallicchio and A. Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9(3):337–350, Jun 2017.

[41] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268(Supplement C):87 – 99, 2017. Advances in artificial neural networks, machine learning and computational intelligence.

[42] B. Graimann, B. Z. Allison, and G. Pfurtscheller. *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Springer Publishing Company, Incorporated, 2013.

[43] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of Statistical Learning*. Spring series in statistics. Springer-Verlag, New York, USA, 2001.

[44] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18 – 31, 2017.

[45] R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Yang, and Y. Zhou. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Scientific Reports*, 5:11476, June 2015.

[46] T. T. Hewett, R. Baecker, S. Card, T. Carey, J. Gasen, M. Mantei, G. Perlman, G. Strong, and W. Verplank. Acm sigchi curricula for human-computer interaction. Technical report, New York, NY, USA, 1992.

[47] S. Hochreiter, M. Heusel, and K. Obermayer. Fast model-based protein homology detection without alignment. *Bioinformatics*, 23(14):1728–1736, 2007.

[48] U. Huebner, N. B. Abraham, and C. O. Weiss. The Santa Fe Time Series Competition Data. Available at: http://goo.gl/6IKBb9, date of access: 12 September 2013.

[49] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology, 2001.

[50] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert. Optimization and applications of Echo State Networks with leaky-integrator neurons. *Neural Networks*, 20(3):335–352, 2007.

[51] X. Jia, K. Li, X. Li, and A. Zhang. A novel semi-supervised deep learning framework for affective state recognition on eeg signals. In *2014 IEEE International Conference on Bioinformatics and Bioengineering*, pages 30–37, Nov 2014.

[52] S. Jirayucharoensak, S. Pan-ngum, and P. Israsena. Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal*, 2014:627892, 09 2014.

[53] H. T. Jongen, K. Meer, and E. Triesch. *Optimization Theory.* Kluwer Academic Publishers, 2004.

[54] S. Kaae Sønderby, C. Kaae Sønderby, H. Nielsen, and O. Winther. Convolutional LSTM Networks for Subcellular Localization of Proteins. *ArXiv e-prints*, Mar. 2015.

[55] S. Kaae Sønderby and O. Winther. Protein Secondary Structure Prediction with Long Short Term Memory Networks. *ArXiv e-prints*, Dec. 2014.

[56] F. Kafka. *The trial.* Der Prozeß, 1925. Translated by David Wyllie.

[57] J. Kennedy and R. Eberhart. Particle Swarm Optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948, 1995.

[58] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598), 1983.

[59] L. W. Ko, H. C. Lee, S. F. Tsai, T. C. Shih, Y. T. Chuang, H. L. Huang, S. Y. Ho, and C. T. Lin. Eeg-based motion sickness classification system with genetic feature selection. In *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pages 158–164, April 2013.

[60] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii. Deep learning of fMRI big data: a novel approach to subject-transfer decoding. *ArXiv e-prints*, Jan. 2015.

[61] K. A. H. Kulasuriya and M. U. S. Perera. Forecasting epileptic seizures using eeg signals, wavelet transform and artificial neural networks. In *2011 IEEE International Symposium on IT in Medicine and Education*, volume 1, pages 557–562, Dec 2011.

[62] D. Kutscher and J. Ott. Dynamic device access for mobile users. In M. Conti, S. Giordano, E. Gregori, and S. Olariu, editors, *Personal Wireless Communications*, pages 53–64, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[63] D. LaBerge. Sustained attention and apical dendrite activity in recurrent circuits. *Brain Research. Brain Research Reviews*, 50, 2005.

[64] J. Lanchantin, R. Singh, Z. Lin, and Y. Qi. Deep Motif: Visualizing Genomic Sequence Classifications. *ArXiv e-prints*, May 2016.

[65] B. Lee, J. Baek, S. Park, and S. Yoon. deeptarget: End-to-end learning framework for microrna target prediction using deep recurrent neural networks. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '16, pages 434–442, New York, NY, USA, 2016. ACM.

[66] B. Lee, T. Lee, B. Na, and S. Yoon. DNA-Level Splice Junction Prediction using Deep Recurrent Neural Networks. *ArXiv e-prints*, Dec. 2015.

[67] S. Lee, M. Choi, H. s. Choi, M. S. Park, and S. Yoon. Fingernet: Deep learning-based robust finger joint detection from radiographs. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4, Oct 2015.

[68] T. Lee and S. Yoon. Boosted categorical restricted boltzmann machine for computational prediction of splice junctions. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 2483–2492. JMLR.org, 2015.

[69] M. K. K. Leung, H. Y. Xiong, L. J. Lee, and B. J. Frey. Deep learning of the tissue-regulated splicing code. *Bioinformatics*, 30(12):i121–i129, 2014.

[70] Y. Li, X. Li, M. Ratclife, L. Liu, Y. Qi, and Q. Liu. A real-time eeg-based bci system for attention recognition in ubiquitous environment. *The 2011 international workshop on Ubiquitous affective awareness and intelligent interaction(UAAII??11).*, 2010.

[71] S. F. Liang, F. Z. Shaw, C. P. Young, D. W. Chang, and Y. C. Liao. A closed-loop brain computer interface for real-time seizure detection and control. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 4950–4953, Aug 2010.

[72] C. T. Lin, B. S. Lin, F. C. Lin, and C. J. Chang. Brain computer interface-based smart living environmental auto-adjustment control system in upnp home networking. *IEEE Systems Journal*, 8(2):363–370, June 2014.

[73] C. T. Lin, B. S. Lin, F. C. Lin, and C. J. Chang. Brain computer interface-based smart living environmental auto-adjustment control system in upnp home networking. *IEEE Systems Journal*, 8(2):363–370, June 2014.

[74] M. Lukoševičius and H. Jaeger. Reservoir Computing Approaches to Recurrent Neural Network Training. *Computer Science Review*, 3:127–149, 2009.

[75] W. Maass. Noisy spiking neurons with temporal coding have more computational power than sigmoidal neurons. Technical Report TR–1999–037, Institute for Theorical Computer Science. Technische Universitaet Graz. Graz, Austria, 1999.

[76] G. Manjunath and H. Jaeger. Echo State Property Linked to an Input: Exploring a Fundamental Characteristic of Recurrent Neural Networks. *Neural Computation*, 25(3):671–696, March 2013.

[77] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943.

[78] S. Min, B. Lee, and S. Yoon. Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18(5):851–869, 2017.

[79] D. Ming, Y. Xi, M. Zhang, H. Qi, L. Cheng, B. Wan, and L. Li. Electroencephalograph (eeg) signal processing method of motor imaginary potential for attention level classification. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4347–4351, Sept 2009.

[80] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky. Classification of patterns of eeg synchronization for seizure prediction. *Clinical Neurophysiology*, 120(11):1927 – 1940, 2009.

[81] T. M. Mitchell. *The Discipline of Machine Learning.* Carnegie Mellon University, School of Computer Science, Machine Learning Department, July 2006.

[82] N. Mladenovic, J. Brimberg, P. Hansen, and J. A. Moreno-Pérez. The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research*, 179(3):927 – 939, 2007.

[83] A. Moraglio, C. Di Chio, J. Togelius, and R. Poli. Geometric particle swarm optimization. *Journal of Artificial Evolution and Applications*, 2008:11:1–11:14, Jan. 2008.

[84] A. Moran. Attention and concentration training in sport. *Encyclopaedia of Applied Psychology*, 9, 2004.

[85] B. Mrquez, A. Alanis, M. Angel Lopez, and J. Sergio Magdaleno-Palencia. Sport education based technology: Stress measurement in competence. *e-Learning and e-Technologies in Education (ICEEE)*, pages 247–252, 09 2012.

[86] I. Nakanishi, S. Baba, and S. Li. Evaluation of brain waves as biometrics for driver authentication using simplified driving simulator. In *2011 International Conference on Biometrics and Kansei Engineering*, pages 71–76, Sept 2011.

[87] I. Nakanishi, S. Baba, K. Ozaki, and S. Li. Using brain waves as transparent biometrics for on-demand driver authentication. *International Journal of Biometrics*, 5:288 – 305, 01 2013.

[88] I. Nakanishi, K. Ozaki, and S. Li. Evaluation of the brain wave as biometrics in a simulated driving environment. *Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 01 2012.

[89] A. Navarro, L. Ceccaroni, F. Velickovski, S. Torrellas, F. Miralles, B. Allison, R. Scherer, and J. Faller. Context-awareness as an enhancement of brain-computer interfaces. ambient assisted living. *Engineering in Medicine and Biology Society*, 2011.

[90] S. Park, S. Min, H. Choi, and S. Yoon. deepMiRGene: Deep Neural Network based Precursor microRNA Prediction. *ArXiv e-prints*, Apr. 2016.

[91] M. P. Paulraj, S. B. Yaccob, A. H. B. Adom, C. R. Hema, and K. Subramaniam. Eeg based hearing perception level estimation for normal hearing persons. In *2012 IEEE Conference on Control, Systems Industrial Informatics*, pages 160–162, Sept 2012.

[92] D. A. Peterson, J. N. Knight, M. J. Kirby, C. W. Anderson, and M. H. Thaut. Feature selection and blind source separation in an EEG-based brain-computer interface. *EURASIP J. Appl. Signal Process.*, 2005:3128–3140, jan 2005.

[93] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch. Recurrent neural network based prediction of epileptic seizures in intra- and extracranial eeg. *Neurocomputing*, 30(1):201 – 218, 2000.

[94] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, and V. D. Calhoun. Deep learning for neuroimaging: a validation study. *ArXiv e-prints*, Dec. 2013.

[95] R. Plutchik. *The Nature of Emotions: Clinical Implications*, pages 1–20. Springer US, Boston, MA, 1988.

[96] C.-V. J. Presacco A, Forrester L. Towards a noninvasive brain-machine interface system to restore gait function in humans. *Engineering in Medicine and Biology Society*, 2011.

[97] Prinzel, L. J, A. Pope, F. G. Freeman, M. Scerbo, and P. Mikulka. Empirical analysis of eeg and erps for psychophysiological adaptive task allocation. 52, 07 2001.

[98] D. R Kelley, J. Snoek, and J. Rinn. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Research*, 26:gr.200535.115, 05 2016.

[99] A. Ravindran. *Operations Research and Management Science Handbook*. Operations Research Series. CRC Press, 2016.

[100] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lcuyer. Openvibe: An open-source software platform to design, test, and use brain ;computer interfaces in real and virtual environments. *Presence*, 19(1):35–53, Feb 2010.

[101] A. Rodan and P. Tiňo. Minimum Complexity Echo State Network. *IEEE Transactions on Neural Networks*, 22:131–144, 2011.

[102] R. Rojas. *Neural Networks.A Systematic Introduction.* Springer-Verlag, Berlin, 1996.

[103] H. R. Roth, C. T. Lee, H. C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers. Anatomy-specific classification of medical images using deep convolutional nets. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 101–104, April 2015.

[104] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 556–564, Cham, 2015. Springer International Publishing.

[105] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Transactions on Medical Imaging*, 35(5):1170–1181, May 2016.

[106] R. N. Roy, S. Bonnet, S. Charbonnier, and A. Campagne. Mental fatigue and working memory load estimation: Interaction and implications for eeg-based passive bci. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6607–6610, July 2013.

[107] A. S. Royer, A. J. Doud, M. L. Rose, and B. He. Eeg control of a virtual helicopter in 3-dimensional space using intelligent control strategies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(6):581–589, Dec 2010.

[108] S. Sanei and J. Chambers. *EEG signal processing*, volume 2007. Wiley-Interscience, 01 2007.

[109] M.-S. M. M. Sarah N. Abdulkader, Ayman Atia. Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, 16(6), July 2015.

[110] P. Sauseng, J. Hoppe, W. Klimesch, C. Gerlof, and F. C. Hummel. Dissociation of sustained attention from central executive functions: local activity and interregional connectivity in the theta range. *European Journal of Neuroscience*, 25, 2007.

[111] M. Schröder, M. Bogdan, T. Hinterberger, and N. Birbaumer. Automated EEG feature selection for brain computer interfaces. In *First International IEEE EMBS Conference on Neural Engineering*, pages 626–629, March 2003.

[112] K. P. Sharanreddy M. Automated eeg signal analysis for identification of epilepsy seizures and brain tumour. *Med. Eng. Technol.*, 6, 2013.

[113] S. Siuly, Y. Li, and Y. Zhang. *Electroencephalogram (EEG) and Its Background*, pages 3–21. Springer International Publishing, Cham, 2016.

[114] M. M. Sohlberg and C. A. Mateer. Efectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology*, 9, 1987.

[115] M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu. Continuous emotion detection using eeg signals and facial expressions. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, July 2014.

[116] K. A. Sorudeykin. An Educative Brain-Computer Interface. *ArXiv e-prints*, Mar. 2010.

[117] M. Spencer, J. Eickholt, and J. Cheng. A deep learning network approach to ab initio protein secondary structure prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12(1):103–112, Jan 2015.

[118] S. Stober, D. J. Cameron, and J. A. Grahn. Using convolutional neural networks to recognize rhythm ?stimuli from electroencephalography recordings. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1449–1457. Curran Associates, Inc., 2014.

[119] H.-I. Suk and D. Shen. Deep learning-based feature representation for ad/mci classification. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 2):583590, 2013.

[120] I. Sutskever, J. Martens, and G. E. Hinton. Generating text with recurrent neural networks. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1017–1024, 01 2011.

[121] D. Tan and A. Nijholt. *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction.* Springer Springer Verlag, 01 2010.

[122] S. C. Turaga, J. F. Murray, V. Jain, F. Roth, M. Helmstaedter, K. Briggman, W. Denk, and H. S. Seung. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.*, 22(2):511–538, Feb. 2010.

[123] M. A. J. van Gerven, F. P. de Lange, and T. Heskes. Neural decoding with hierarchical generative models. *Neural Comput.*, 22(12):3127–3142, Dec. 2010.

[124] G. Vecchiato, L. Astolfi, F. D. V. Fallani, S. Salinari, F. Cincotti, F. Aloise, D. Mattia, M. G. Marciani, L. Bianchi, R. Soranzo, and F. Babiloni. The study of brain activity during the observation of commercial advertsing by using high resolution eeg techniques. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 57–60, Sept 2009.

[125] G. Vecchiato, F. Babiloni, L. Astolfi, J. Toppi, G. Vecchiato, L. Astolfi, P. Cherubino, J. Dai, W. Kong, and D. Wei. Enhance of theta eeg spectral activity related to the memorization of commercial advertisings in chinese and italian subjects. In *2011*

4th International Conference on Biomedical Engineering and Informatics (BMEI), volume 3, pages 1491–1494, Oct 2011.

[126] B. Venthur, B. Blankertz, M. F. Gugler, and G. Curio. Novel applications of bci technology: Psychophysiological optimization of working conditions in industry. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 417–421, Oct 2010.

[127] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):287–289, 2007.

[128] L. Wei, Q. Hong, H. Yue, and C. Xi. The research in a plantar pressure measuring system connected with eeg. In *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, pages 434–437, Oct 2010.

[129] R. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1:270–280, 1989.

[130] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *Journal of Neural Engineering*, 8(3):036015, June 2011.

[131] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging*, 35(1):119–130, 1 2016.

[132] L. Xu, J. Liu, G. Xiao, and W. Jin. Characterization and classifcation of eeg attention based on fuzzy entropy. *Proceedings of the 2012 Third International Conference on Digital Manufacturing and Automation (ICDMA).*, 2012.

[133] I. B. Yildiza, H. Jaeger, and S. J. Kiebela. Re-visiting the echo state property. *Neural Networks*, 35:1–9, 2012.

[134] M. Yoshioka, T. Inoue, and J. Ozawa. Brain signal pattern of engrossed subjects using near infrared spectroscopy (nirs) and its application to tv commercial evaluation. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, June 2012.

[135] H. Zeng, M. D. Edwards, G. Liu, and D. K. Gifford. Convolutional neural network architectures for predicting dnacprotein binding. *Bioinformatics*, 32(12):i121–i127, 2016.

[136] S. Zhang, J. Zhou, H. Hu, H. Gong, L. Chen, C. Cheng, and J. Zeng. A deep learning framework for modeling structural features of rna-binding protein targets. *Nucleic acids research*, 44(4):e32, February 2016.

[137] Y. Zhao and L. He. Deep learning in the eeg diagnosis of alzheimer's disease. In C. Jawahar and S. Shan, editors, *Computer Vision - ACCV 2014 Workshops*, pages 340–353, Cham, 2015. Springer International Publishing.

[138] J. Zhou and O. G. Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature methods*, 12(10):931934, October 2015.

[139] V. Zotev, R. Phillips, H. Yuan, M. Misaki, and J. Bodurka. Self-regulation of human brain activity using simultaneous real-time fmri and eeg neurofeedback. *NeuroImage*, 85:985 – 995, 2014. Neuro-enhancement.

A Review comments of our article

Review comments of our article in IEEE World Congress on Computational Intelligence (IEEE WCCI), 2018

```
Date: Thu, 15 Mar 2018 17:46:33 -0400 (EDT)
 Subject: IEEE CEC 2018 Paper #18472 Decision Notification
 Dear Author(s),

 Congratulations! On behalf of the IEEE CEC 2018 Technical Program
 Committee
 and Technical Chairs, we are pleased to inform you that your paper:

     Paper ID:   18472
     Author(s):  Hikmat Dashdamirov, Sebastian Basterrech and Pavel Kromer
     Title:      A Nature-inspired System for Mental State Recognition

 has been accepted for presentation at the IEEE CEC 2018 and for
 publication in
 the conference proceedings published by IEEE.


                                    REVIEWERS' COMMENTS

 ----------------------------------------------
 REVIEW NO. 1

 Comments to the authors:
```

The article is one of the bests I read in my reviews for this conference for 2018. The application seems useful and EA seems to be the choice fo the problem in combination for NN. We need to see more of these applications in our conferences. You only need to refer to more recent articles in the PSO area, doi 10.1162/EVCO_r_00180, as a comprehensive PSO.

REVIEW NO. 2

Comments to the authors:

Authors present an interesting application of meta-heuristics and artificial neural networks for computing the human concentration during specific activities. In particular, authors developed a very robust solution that is confirmed working after many lucid examples. Additionally, data were also collected by authors, while full data collection process is described in details. In a nutshell, I have not found any major problems in this paper. Due to the complexity of proposed solution, authors are encouraged to present their solution also in pseudocode style. Please also revise the taxonomy of featured algorithms. PSO belongs much more to the swarm algorithms than evolutionary algorithms.

REVIEW NO. 3

Comments to the authors:

The paper is good and topic is relevant for this distinguished conference audience.

However, authors should improve the following items:

1. overall figure readability must be increased

2. tables and figures caption should be properly formatted according to template.

CHAIR'S COMMENTS

B The poster of our article

Poster of our article in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2018

Estimation of Human Concentration using Echo State Networks

Hikmat Dashdamirov and Sebstián Basterrech

VSB, Ostrava and FEL, CTU, Prague, Czech Republic

Abstract

We introduce a very simple and portable device for estimating the human concentration. We developed a Brain-Computer Interface system based on EEG signals which is able to produce highly accurate prediction of the human activities. There are two types of mental activities, one requires high concentration and another one requires relaxation.

We show that it is possible to estimate the human concentration with few brain signals. The classification problem is solved using Neural Networks . In particular, we obtain a very accurate classifier using the fast and robust Echo State Network method.

Keywords: Echo State Networks, Brain Computer Interface, Signal Processing, Classification

Reservoir Computing model

- An ESN is a specific type of three layered Neural Network. The input and output layers are as in traditional feed-forward networks. The hidden layer, called reservoir, is a recurrent network. The existence of cycles in the *reservoir* has important consequences:
  - ➢ The reservoir is a dynamical system (neural networks without cycles are functions).
  - ➢ The cycles can be used to learn and memorize temporal information.
  - ➢ There are often difficulties during learning processes that use methods of the gradient descent type or in the Quasi Newton class.
- The ESN model is based on the empirical observation that under certain hypothesis, training just the output weights is often sufficient to obtain an excellent performance in many learning problems.

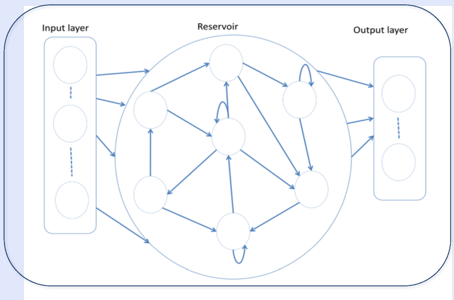

Fig. 1: Topology of Echo State Network (Figure extracted from [4]).

Brain Computer Interface

- Brain-computer interface (BCI) system is a functional interaction between the brain and an external device.
- BCI basically consists of three components: a brain signal acquisition system, an information processing device, and an external device.

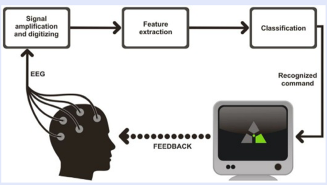

Fig. 2: Brain Computer Interface.

- In respect of first component, there are several kinds of signals that have been used for BCI. The most widespread signal is the Electroencephalography (EEG) that presents good advantages in respect to the other ones, such as: good temporal resolution, portability, and low set-up cost.
- The second BCI provides a parametric mapping between the brain signals and the mental states.
- The third component is an external device committed to receive commands from the classifier and to provide feedbacks to the subject.

Methodology:

- A total of 4 healthy subjects aged from 23 to 38.
- All the subjects were right handed.
- The subjects performed instructions that were displayed on a screen.
- Command A we asked to the subject for reading and handwriting the displayed text (Kafka's
- book:"The trial").
- Command B we asked to the subject for being relaxed.The total time of the experiment for each subject was 300 seconds.
- The actions A and B were alternating between them.
- EEG synchronization was made using Openvibe software platform and Lua language.
- The EEG signals were recorded using NeuroSky's Mindset device.
- Data provided by the BCI device: delta, theta, alpha, beta, and gamma signals.

Formalization of the Echo State Networks

Formalization of the model

- Let an ESN have N_u input neurons, N_x neurons in the reservoir and N_y output neurons, with w^{in},w^{res} and w^{out} the respective weights.
- We have a times series data set of inputs *(u(t),u(t-1),…)* and a desired output $y_{targ}(t)$.
- We denote by *X(t)* the state of the reservoir at the time *t*, produced by:

$$X(t+1)=f(W^{in}u(t+1)+W^{res}X(t)),$$

where f() most often is the sigmoid function.

The model output is performing a linear combination:

$$y(t+1)=W^{out}X(t+1),$$

the readout weights are obtained using ridge linear regression.

- **The important principle of the reservoir: its weights are deemed fixed during the learning process.**

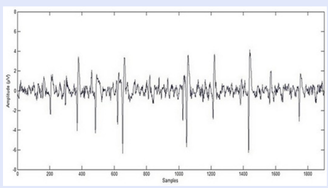

Fig. 3: Example of an EEG signal.

Numerical experiences

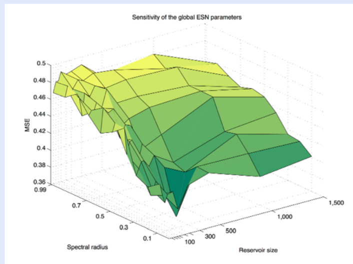

Fig. 4: Sensitive analysis of the ESN parameters.

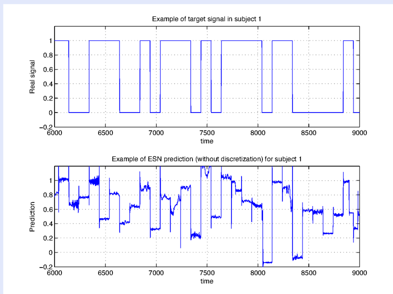

Fig. 5: Target signal and predicted signal.

Conclusion

- The article introduces an application of ESN in the BCI area for computing the human concentration during specific activities.
- We are particularly interested in developing a portable, non-invasive and robust device.
- We developed a system with the following advantages: few global parameters, low computational cost, robust and fast learning, and economic and portable device

Main references

[1] H. Jaeger, *"The echo state approach to analysing and training recurrent neural networks"*, German National Research Center for Information Technology, Tech. Rep. 148, 2001.

[2] M. Lukosevicius and H. Jaeger, "Reservoir Computing Approaches to Recurrent Neural Network Training," Computer Science Review, vol. 3, pp. 127–149, 2009.

[3] S. Basterrech, P. Bobrov, A. Frolov, and D. Husek, "Nature-inspired Algorithms for Selecting EEG Sources for Motor Imagery Based BCI," in Artificial Intelligence and Soft Computing,Springer International Publishing, 2015, vol. 9120, pp. 79–90.

[4] S. Basterrech, "Availability of data and materials."
http://aic.fel.cvut.cz/basterrech/

European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)

Bruges (Belgium), 25 - 27 April 2018