EEG DATA CLASSIFICATION USING MULTILINEAR REGRESSION MODEL

by

Ayşe Akgün Demir

B.S., in Mathematical Engineering, Yıldız Technical University, 2013

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2019

EEG DATA CLASSIFICATION USING MULTILINEAR REGRESSION MODEL

APPROVED BY:

Prof. Ahmet Ademoğlu .....
(Thesis Advisor)

Assoc. Prof. Esin Öztürk Işık .....

Asst. Prof. Adil Deniz Duru .....

DATE OF APPROVAL: 3 December 2019

ACKNOWLEDGMENTS

First of all I would like to express my gratitude to my thesis advisor Prof. Dr. Ahmet Ademoğlu whom I will always take an example with his knowledge, life experience and personality, for his help and for his guidance in completing this thesis.

I would like to express my respects to my managers at work, Yakup Gülmez and Turgay Özkan, for supporting me during my graduate studies. I would also like to thank Sencer Melih Deniz for sharing academic knowledge in my study and my friends in the Neurosignal Analysis Laboratory where we had pleasant conversations.

I am greatful to my dear life partner Mehmet Demir who always stands by me, supports me all the time and patiently endured my stressful education life.

Finally, I would like to thank my dear mother Melek Akgün and father Sami Akgün for their greatest efforts in coming to these days, my dear sister Ebru Akgün, who has always supported me and has been with me during my hard times, my dear sisters Cihane and Habibenur Akgün whose love I feel with me.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Ayşe Akgün Demir, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

EEG DATA CLASSIFICATION USING MULTILINEAR REGRESSION MODEL

Brain Computer Interfaces (BCI) are systems that facilitate people to use a computer, to control an electromechanical or a neuroprosthetic device without using their motor nervous system. It is possible to obtain an information about the brain tissues with electrodes placed on the skull which record the electrical activity called electroencephalogram (EEG) . The electrodes placed in different regions capture the activity in their neighborhood. BCI systems combining the electrical signals from these electrodes use signal processing and machine learning algorithms to identify the motor or the cognitive activity that is embedded in the brain signals so as to mobilize the peripheral devices according to the information gathered. Emotion estimation is often used in brain computer interface applications to improve and control the communication between man and machine. In recent years, emotion estimation studies based on brain electrical activity, which is the most widespread method used for accurate emotion analysis, have gained momentum. In this thesis study, multichannel EEG data taken from normal subjects who encountered emotionally pleasant and unpleasant pictures were classified with a multilinear regression algorithm. The results were compared with those of the Support Vector Machine (SVM) and proved to be better in accuracy.

Keywords: Tensors, Multilinear Regression Model, Brain Computer Interface, Emotion Detection.

ÖZET

ÇOKLU DOĞRUSAL REGRESYON MODELİ KULLANARAK EEG VERİLERİNİN SINIFLANDIRMASI

Beyin Bilgisayar Arayüzleri (BCI), kişilerin, sinir sistemlerini kullanmaksızın, elektromekanik veya nöroprostetik bir cihazı bilgisayar yardımıyla denetlemelerini sağlayan sistemlerdir. Kafaderisi üzerine yerleştirilen elektrotlardan elde edilen ve Elektroansefalogram (EEG) adı verilen elektriksel kayıtlar sayesinde beyin dokuları hakkında bilgi edinebilmek mümkündür. Farklı bölgelere yerleştirilen elektrotlar çevrelerindeki elektriksel aktiviteyi yakalayabilmektedirler. BCI sistemleri, beyin sinyallerinin barındırdığı motor veya bilişsel etkinlik bilgisini tanımlayıp çevrede bulunan cihazları harekete geçirebilmek için, bu elektrotlardan elde edilen elektriksel sinyalleri birleştirerek, sinyal işleme ve yapay öğrenme algoritmaları kullanırlar. Duygu tahmini, genellikle beyin bilgisayar arayüzü uygulamalarında insan ve makine arasındaki iletişimi geliştirmek ve denetlemek için kullanılır. Doğru duygu analizi için en yaygın kullanılan yöntem olan beyin elektriksel aktivitesine dayalı duygu tahmini çalışmaları son yıllarda ivme kazanmıştır. Bu tez çalışmasında, duygusal anlamda hoş ve nahoş resimlerle karşılaşan normal deneklerden alınan çok kanallı EEG verileri çoklu doğrusal regresyon algoritması ile sınıflandırılmıştır. Sonuçlar, Destek Vektörü Makinesi (SVM) sonuçları ile karşılaştırılarak doğrulukta daha iyi olduğu kanıtlanmıştır.

Anahtar Sözcükler: Tensörler, Çoklu Doğrusal Regresyon Modeli, Beyin Bilgisayar Arayüzü, Duygu Sezim.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

x	Vector
X	Matrix
$\mathcal{X}$	Tensor
$\otimes$	Kronecker Product
$\odot$	Khatri-Rao Product
vec($\mathcal{B}$)	Vectorization Operator of a Multidimensional Array B
$\langle a, b \rangle$	Inner Product of the Vectors a and b
$a \circ b$	Outer Product of the Vectors a and b
β^T	Transpose of β
$\beta^{(r)}$	Rank r vector β
$\mathcal{X}_{(n)}$	Unfolded Tensor on the nth dimension
$x(i)$	(i)th element of vector x
$X(i, j)$	(i,j) th element of matrix X
$\mathcal{X}(i, j, k)$	(i,j,k) th element of tensor $\mathcal{X}$

LIST OF ABBREVIATIONS

BCI	Brain Computer Interface
MRI	Magnetic Resonance Imaging
fMRI	Functional Magnetic Resonance Imaging
EEG	Electroencephalography
DTI	Diffusion Tensor Imaging
PET	Positron Emission Tomography
ALS	Amyotrophic Lateral Sclerosis
GLM	General Linear Model
ECOG	Electrocorticography
BIC	Bayesian Information Criterion
IAPS	International Affective Picture System
EDA	Electrodermal Activity
ECG	Electrocardiography
BVP	Blood Volume Pulse
RSP	Respiration
GSR	Galvanic Skin Response
EMG	Electromyography
CP	Canonical Polyadic Decomposition
1-D	One Dimensional
2-D	Two Dimensional
3-D	Three Dimensional
N-D	N Dimensional
ANN	Artificial Neural Network
SVR	Support Vector Regression
GP	Gaussian Process
GLM	General Linear Model
ML	Maximum Likelihood
SVM	Support Vector Machine

DWPT	Discrete Wavlet Packet Transform
ERP	Event Related Potential
PCA	Priciple Component Analysis
ICA	Independent Component Analysis
HMM	Hidden Markov Model
fNIRS	Functional Near Infrared Spectroscopy

1. INTRODUCTION

To fully understand the inner workings of the human brain and to establish its connections with neuropsychiatric and neurodegenerative diseases is one of the most popular topics of recent scientific studies. Various neuroimaging technologies such as anatomical Magnetic Resonance Imaging (MRI) , functional Magnetic Resonance Imaging (fMRI), Electroencephalogram (EEG),Diffusion Tensor Imaging (DTI) and Positron Emission Tomography (PET) are the mainstay of studies in the field of neuroscience.However, the high size of the medical imaging data and the complexity of the data make it difficult to accurately analyze the data.

It is possible to obtain an information about the brain tissues with electrodes placed on the skull an electrical activity called EEG is recorded. The electrodes placed in different regions capture the activity in their neighborhood. BCI systems combining the electrical signals from these electrodes use signal processing and machine learning algorithms to identify the motor or the cognitive activity that is embedded in the brain signals so as to mobilize the peripheral devices according to the information gathered. These systems can be used for rehabilitation of people having medical problems such as ALS or Tetraplegia who are unable to perform motor functions [2]. They can also provide hands-off capability to control their environment in the context of entertainment or neuro-feedback applications [3].

Tensor methods allow for classifying multidimensional images into different clinical outcomes using a regression based model incorporating the pertinent information such as age, gender, treatment and other possible covariates or factors. This study describes a classification method based on a tensor regression algorithm using General Linear Model (GLM) [4]. An approach to classification of emotion from EEG signals is described. Subjects are given visual stimuli from IAPS database library and EEG signals are recorded for analysis.The signals are subjected to wavelet decomposition and divided into several frequency bands. Emotional states in various bands are classified

as having a positive or negative valence. The same classification is compared using Support Vector Machines (SVM), as well.

The organization of the thesis is as follows: In Chapter 2, the BCI systems, its history and application fields are explained. Chapter 3 gives general information about emotion estimation, emotion analysis and emotion models. In Chapter 4, tensor notation and operations that are used throughout the thesis are introduced. In Chapter 5 the tensor regression model which is used in this thesis is described and several simulations are conducted. In Chapter 6, two tensorial data analysis results and their comparison with those of the SVM are given. In Chapter 7, the results of the data analysis are discussed and compared with those of similar studies. Finally, in Chapter 8, conclusion and future work are presented.

2. BRAIN-COMPUTER-INTERFACE (BCI) SYSTEMS

A BCI is a communication system that enables people to interact with their environment using control signals generated from their brain activity instead of their peripheral nerves and muscles. This is done by measuring and interpreting the electrical or hemodynamic activity of the brain using computer algorithms and generating appropriate control signals. The idea of BCI was originally created in 1929 by Hans Berger [5], who developed a device capable of measuring electrical brain activity known as EEG.

A study of instrumental and operational conditioning by Eberhard E. Fetz at the University of Seattle in 1969 showed that monkeys first learned to control the displacement of a biofeedback arm with neuronal activity. It was found that monkeys were able to learn single control as well as control of multiple neurons in the primary motor cortex very quickly. Subsequently, these studies were also performed with rats. It was found that the activity of newly isolated cells increased in both species [6]. In 1970, the US Defense Advanced Research Projects Agency (DARPA) launched a program to explore brain communications using EEG [7]. In the 1980s, it was found that multiple neuron groups in different parts of the brain could send motor commands together. In the 1990s, several groups of independent researchers were able to record and analyze the motor signals of neuron groups. Thus, it was possible to have a substantially better understanding of the electrical brain activity which led to the BCI to develop in real sense [5]. Together with the Department of Neuroscience at Brown University, Cyberkinetics Corporation launched its first publicly available BCI game in 2003. In the 2000's, Wireless BCI ENOBIO was developed by a Spanish Company called Starlab, "Silent Talk" BCI research was funded by DARPA and worldwide Annual BCI competitions were organized by various organizations [7].

2.1 Basic Prnciples of BCI Systems

BCI communication is made possible by measuring the brain electrical activity. This interface can be invasive, non-invasive, or partially invasive depending on the source from which it recieves the neural signals [8].

2.1.1 Invasive Systems

In the invasive method, the microelectrodes are implanted directly into the grey matter into specific neuronal regions in the brain. Thus, electrical signals can be captured in the brain and transferred to the computer. The invasive method is the most accurate means, but always carries some risks and requires a painstaking surgical procedure.

2.1.2 Partially-Invasive Systems

Partially invasive BCI devices are placed inside the skull but outside the brain tissue. in this method, Electrocorticography, (ECOG) which measures the electrical activity of the brain from the outer surface of the cortex.

2.1.3 Non-Invasive Systems

In non-invasive techniques, the EEG [9] functional near infrared spectrum (fNIRS) activity are recorded from the scalp can be measured[10]. In some studies, fMRI is also used [11] .Another non-invasive technique called Electromyogram (EMG) is also used to record the electrical signal reaching the skeletal muscles from peripheral nerves [12]. EMG provides both high signal resolution and specificity as it is directly based on the activity of motor neurons. For example, for an electromechanical arm control, EMG can pick up the signals directly from the nerves that control arm and palm movements

in this way [13].

Among all the noninvasive methods, EEG is the most commonly used technique in BCI. The electrical signal measured as the EEG passes through the tissues and the skull as it reaches to the electrodes. This phenomena is called the volume conduction effect. Therefore, its spatial resolution is poorer than that of fMRI or invasive methods. However, the time resolution of EEG is high in the sense that it can capture the events within the milisecond range [14]. When these systems are used, the subject performs some mental task such as imagining to move an arm that leads to a change in the brain activity, One of the most important objectives of the development of these systems is to allow patients to control their environment who are restrained or completely immobilized due to a neurological disorder, such as paralysis, or a neurodegenerative problem such as ALS.

BCI systems are divided into three categories according to the way they record signals from subjects [15]:

1. Active BCI: The brain signals are modified consciously by the user. The purpose of the movement, motor imageries and mental tasks are in this type of BCI systems. [16].

2. Reactive BCI: The brain signals are generated in response to an external stimulation such as audio, video, pain stimulations producing as a reaction [17].

3. Passive BCI : The brain signals are derived from spontaneous brain activity without any purpose of voluntary control. This type of BCI is used for applications that require assessment of mental states such as attention level, stress or workload [17].

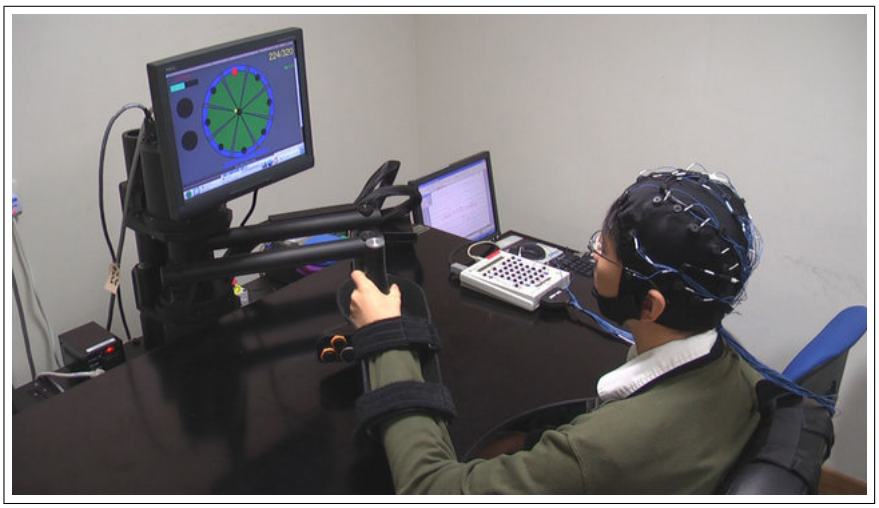

Figure 2.1 EEG based BCI [1].

2.2 BCI Application Areas

BCI is used in several industrial sectors, including medical, educational entertainment and defence. Primary objectives in the medical field can be divided into three sub-areas; rehabilitation, diagnosis and treatment.

2.2.1 Rehabilitation

In neuro-rehabilitation, the BCI basically aims to decode the brain signals while the patient is attempting to perform a specific mental or motor task. The main objective is to reactivate selected brain areas and to facilitate neural plasticity [18].

2.2.2 Diagnosis

In diagnostics, BCI is used for data collection. The data is then checked for abnormalities. In epileptic patients, for instance, who are resistant to medication, epicortical electrodes measuring the field potentials on the cortical surface are analyzed to localize the epilepsy focus and to treat it surgically [19].

2.2.3 Treatment

Therapy with BCIs is achieved by mainly by sending an impulse to a particular organ. A commonly used therapy is Deep Brain Stimulation (DBS). As an example, in Parkinson's disease, the electrodes are implanted under the thalamus, which can suppress the vibration and thus cancel the movement stiffness. These procedures can be used to treat obsessive-compulsive behavior and depression as well as some forms of epilepsy [20].

3. EEG-BASED EMOTION RECOGNITION

3.1 Emotion Analysis

Although it has no clear definition, the emotion can be expressed as the situation or reaction of people against external stimuli while they are thinking, communicating, learning or making decisions. Emotions play an important role in daily life and people with positive emotions are thought to be more successful in society. Therefore, emotion affects human life both psychologically and physically. BCIs are used to measure emotion-based responses and allow for communicaton between human and machines.

Different methods such as facial expressions and voice signals were used in the prediction of emotion and initially sound and face based prediction was performed in 1990s. Although data collection was easier in these methods, distortions due to microphone and camera quality could occur during recording. This sometimes led to incorrect interpretations. Because of that, physiological signals such as Electrodermal Activity (EDA), Electrocardiography (ECG), photopletismography, EEG, Blood Volume Pulse (BVP), Respiration (RSP), Galvanic Skin Response (GSR) and EMG have been used recently in emotion prediction applications [21].

EEG is one of the most powerful methods in emotion prediction studies [22], [23]. There are different approaches to emotion recognition in the literature. Both low frequency [24] and high frequency bands [25] can be used to determine the emotional states. In addition, some researchers claim that important information to detect emotional status is found below 30 Hz frequency band [26].

In this thesis, five different frequency bands (Figure 3.1) [27] were examined for emotion recognition :

- 0-4 Hz frequency band (Delta Waves)

- 4-8 Hz frequency band (Theta Waves)
- 8-12 Hz frequency band (Alpha Waves)
- 12-25 Hz frequency band (Beta Waves)

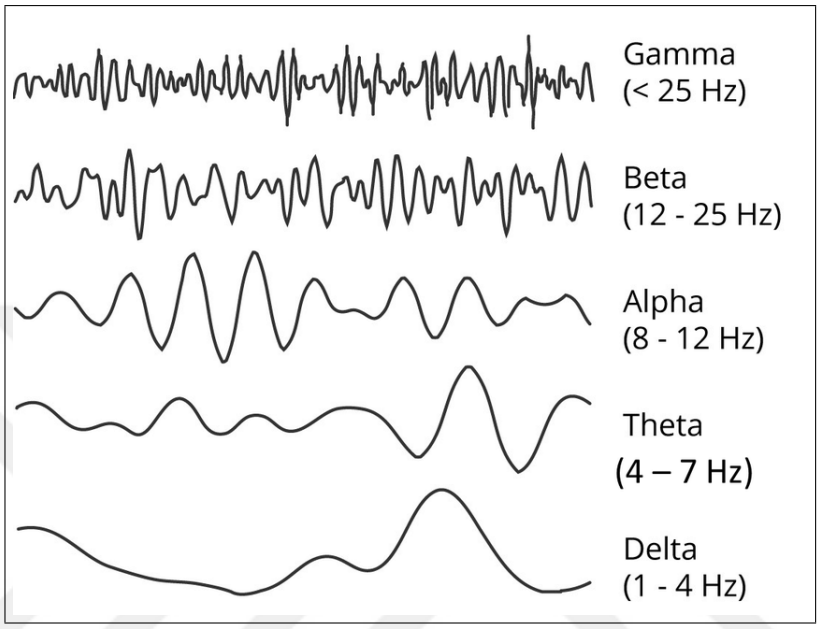

Figure 3.1 EEG frequency bands.

3.1.1 Emotion Models

Emotion estimation studies have two different models: discrete model and dimensional model. In the discrete model structure [28], there are six main emotions such as fear, nerve, happiness, hate, sadness and surprise, whereas in the dimensional model structure [29] there is an arousal-valence plane. In this model structure, emotions in arousel coordinates are defined from calm to intense state, whereas in valence coordinate system, emotions are expressed as negative to positive. Valence represents an individual's love or hatred towards a particular situation or event. Arousal, on the other hand, indicates the physiological state of individuals and represents the passive or active state shown against a stimulus (Figure 3.2).

Although the application of a discrete emotion model is easier, the fact that the arousal valence model is more universal has made this model more attractive [30]. In

this model, a particular emotion is expressed not on a discrete basis but on a coordinate system. For example, in this model, the sense of happiness is expressed as high arousal positive valence, and the sense of sadness is expressed as low arousal negative valence.

Different stimuli are used in the stage of obtaining emotions. Stimulants are visual, auditory, memory based and both visual and auditory. In this study, images to visual stimuli from IAPS database is used [31]. There are so many images in this database that different emotions are revealed through these visuals. The pictures are divided into four different regions on the arousel-valence plane. The positive/negative and intense/calm state of the pictures is expressed with the values given to the pictures. If the valence value is greater than 5, the emotion is positive, if smaller, then it is negative. An excitation value greater than 5 indicates that the emotion is experienced intensively, while a smaller value indicates that emotion is experienced in a calm way [32].

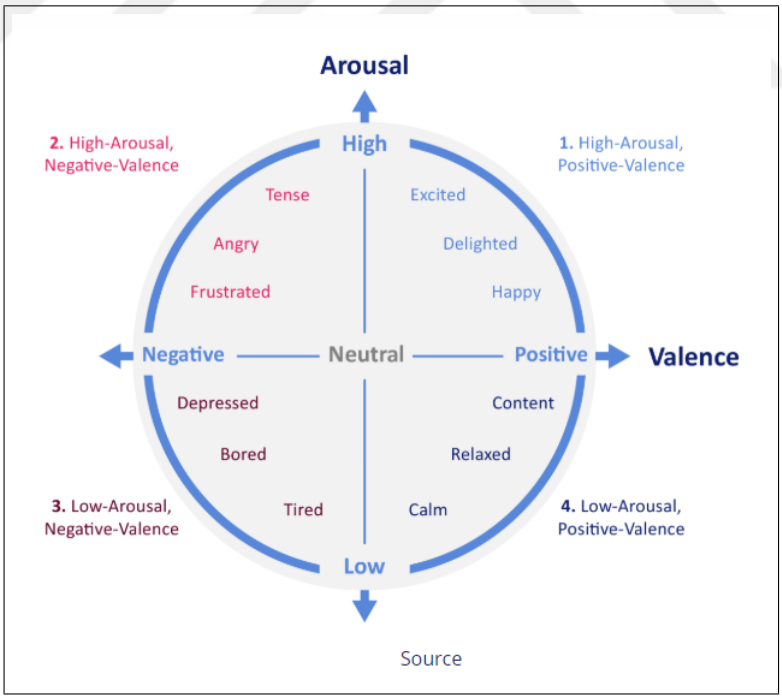

Figure 3.2 Valance-Arousel model for basic emotions.

4. TENSOR NOTATION AND OPERATIONS

In this chapter, the basic concepts of tensor algebra and some of the decomposition methods will be reviewed. The basic notations, definitions and operations will be introduced [33]. As a tensor decomposition method, widely used Canonical Polyadic Decomposition (CP) will be explained.

4.1 Definitions

Tensors, which are multidimensional arrays, are generalizations of vectors and matrices to higher dimensions. The number of dimensions of a tensor denotes the order of the tensor. From this, vectors are $1-D$ tensors which are denoted as $\boldsymbol{x} \in R^{P_1}$ and matrices are $2-D$ tensors which are denoted as $\mathbf{X} \in R^{P_1 \times P_2}$. $N-D$ tensors are represented by $\boldsymbol{\mathcal{X}} \in R^{P_1 \times P_2 \times \ldots \times P_D}$.

In the tensor jargon, the rows and columns in the matrices are called *fibers* and formed by fixing all indices except one indice in tensor. A $3-D$ tensor $\boldsymbol{\mathcal{X}} \in R^{P_1 \times P_2 \times P_3}$ have 3 mode fibers: The mode-1 fiber which is denoted by $\boldsymbol{\mathcal{X}}(\mathrm{i}, :, \mathrm{k})$, the mode-2 fiber which is denoted by $\boldsymbol{\mathcal{X}}(:, \mathrm{j}, \mathrm{k})$ and the mode-3 fiber which is denoted by $\boldsymbol{\mathcal{X}}(\mathrm{i}, \mathrm{j}, :)$.

Slices are two-dimensional sections formed by fixing all indices except two indices in tensor. A $3-D$ tensor $\boldsymbol{\mathcal{X}} \in R^{P_1 \times P_2 \times P_3}$. For example, jth vertical slice of $\boldsymbol{\mathcal{X}}$ is denoted by $\boldsymbol{\mathcal{X}}(:, \mathrm{j}, :)$.

4.2 Tensor Operations

4.2.1 Inner Product , Outer Product , Kronecker Product and Khatri-Rao Product

The Inner product of two tensor which have same size $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in R^{P_1 \times P_2 \times \ldots \times P_D}$ is the sum of products of their entries and is calculated as;

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} \cdots \sum_{p_n=1}^{P_N} x_{p_1 p_2 \cdots p_N} y_{p_1 p_2 \cdots p_N} \tag{4.1}$$

The outer product of K vectors is denoted as $b_1 \circ b_2 \cdots \circ b_K$ where $b_k \in R^{P_k}$. The over all size of the product is a $p_1 \times p_1 \times \cdots p_K$ and calculated as;

$$(b_1 \circ b_2 \cdots \circ b_K)_{i_1 \cdots i_K} = \prod_{k=1}^{K} b_{k i_k} \tag{4.2}$$

The Kronecker product of two matrices, denoted by $\mathbf{I} \otimes \mathbf{J}$ where $\mathbf{I} \in R^{K \times L}$ and $\mathbf{J} \in R^{M \times N}$. The over all size of the $\mathbf{I} \otimes \mathbf{J}$ is KL $\times$ MN and calculated as;

$$\mathbf{I} \otimes \mathbf{J} = \begin{bmatrix} I(1,1)J & I(1,2)J & \cdots & I(1,L)J \\ I(2,1)J & I(2,2)J & \cdots & I(2,L)J \\ \vdots & \vdots & \ddots & \vdots \\ I(K,1)J & I(K,2)J & \cdots & I(K,L)J \end{bmatrix}$$

The Khatri-Rao Product is a special form of kronecker product. Khatri-Rao product of I and J (with the condition of L=N) denoted as $\mathbf{I} \odot \mathbf{J}$. The over all size of

the $\mathbf{I}\odot\mathbf{J}$ is KM $\times$ L and calculated as;

$$\mathbf{I}\odot\mathbf{J} = \begin{bmatrix} I(:,1)\otimes J(:,1) & I(:,2)\otimes J(:,2) & \cdots & I(:,L)\otimes J(:,L) \end{bmatrix}$$

4.2.2 Rank-R CP Decomposition

The basis of tensor modeling lies in the decomposition that can represent a high-dimensional data with low-dimensioanal factors. Nowadays, high-dimensional tensor decomposition is often applied to various fields such as chemometrics, neuroscience, image, video and signal processing [34]. There are basically two kinds of decomposition methods: i) CP Decomposition and ii) Tucker Decomposition [35]. Only CP decomposition is used in this study.

The CP decomposition [36],[37],[38] of a tensor generalizes the bilinear factor models to multilinear data. Tensor rank decompositon can be considered as a generalization of the matrix singular value decompositon (SVD) [39]. The rank of a tensor is the smallest number of components in a decomposition.

Rank-R decomposition of $N-D$ tensor as visualize in Figure 4.1 $\boldsymbol{\mathcal{B}} \in R^{P_1\times P_2\times\ldots\times P_D}$ is as follows:

$$\boldsymbol{\mathcal{B}} = \sum_{r=1}^{R} \beta_1^{(r)} \circ \cdots \circ \beta_D^{(r)} \tag{4.3}$$

where $\beta_D^{(r)} \in R^{P_d}$,d= $1,\cdots,D$, r= $1,\cdots,R$ are vectors.

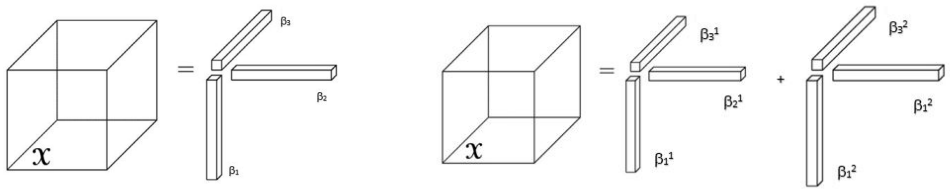

Figure 4.1 A 3-D tensor $\boldsymbol{\mathcal{X}}$ rank-1 decomposition (Left) , A 3-D tensor $\boldsymbol{\mathcal{X}}$ rank-2 decomposition (Right).

4.2.3 Tensor Matricization

Matrixization, also known as unfolding or flattening, is the process of reordering elements of an $N-D$ tensor into a matrices or as mentioned above into slices. In this study, only the mode-n matricization is discussed. The mode-n matricization of a tensor $\boldsymbol{\mathcal{X}} \in R^{P_1 \times P_2 \times \ldots \times P_D}$. is denoted as $X_{(n)}$ as shown in Figure 4.2. The mode-n fibers of the tensor $\boldsymbol{\mathcal{X}}$ will be the coloumns of the resulting matrix. Notationally, the tensor element $(p_1, p_2, \cdots, p_N)$ maps to the matrix element (p_n, j), where $j = 1 + \sum_{k=1}^{N} (p_k - 1) J_k, (k \neq n)$ and $J_k = \prod_{m=1}^{k-1} I_m$, $(m \neq n)$.

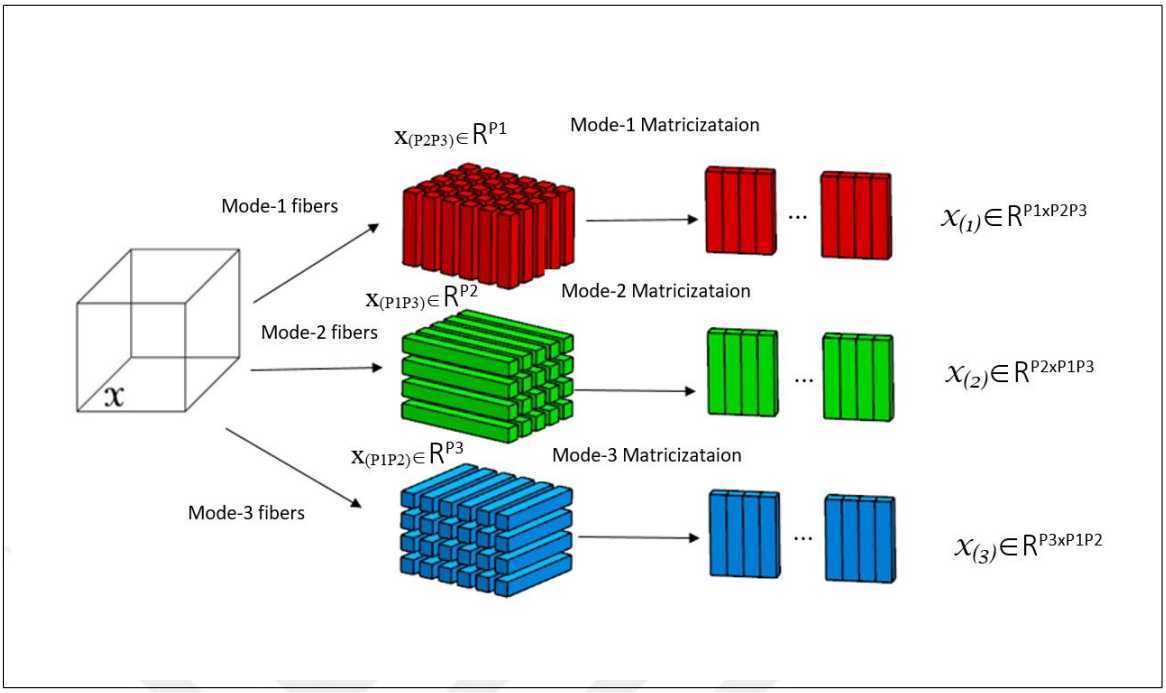

Figure 4.2 Tensor matricization.

5. TENSOR REGRESSION

Tensor regression has several advantages over vector or matrix regression. Tensor models preserve the multidimensional interdependence among variables. In terms of learning, tensor models call for a lower sample complexity. The most commonly used regression models are typically classified as linear and nonlinear regression. Linear regression models describe the relationship between dependent and independent variables using a linear function in the parameters. GLM, as an extension of linear regression models, can model response variables by sampling through a certain distribution from the exponential family. Unlike linear regression models, nonlinear regression characterizes nonlinear dependencies in data. The non-linear parameters generally take the form of an exponential, trigonometric or a power function. Some of the nonparametric nonlinear regression models frequently used in machine learning are Gaussian processes (GP) [40] artificial neural networks (ANN) [41], decision trees [42] and support vector regression (SVR) [43].

5.1 Low-Rank Tensor Regression :CP Tensor Regression

Tensor regression takes advantage of high degree of correlation in the data and learns a multi-linear function whose parameters generate the tensor.

The classical GLM is related to a multidimensional $\mathcal{X} \in R^p$ with the observation Y via $\mu = E(Y|\mathcal{X})$ which is embedded in a link function $g(.)$;

$$g(\mu) = \eta = \alpha + \beta^T \mathcal{X} \tag{5.1}$$

Y is a member of exponential family with probability mass function

$$p(y|\theta, \phi) = exp\Big\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\Big\} \tag{5.2}$$

Generalization of Eq. 5.1 to high dimensional matrices can be expressed as in Eq. 5.3.

$$g(\mu) = \alpha + \gamma^T Z + \langle\ B, \mathcal{X}\rangle \tag{5.3}$$

$$\begin{aligned} g(\mu) &= \alpha + \gamma^T Z + \langle \sum_{r=1}^{R} \beta_1^{(r)} \circ \beta_2^{(r)} \circ \cdots \circ \beta_K^{(r)}, \mathcal{X} \rangle \\ &= \alpha + \gamma^T Z + \langle (B_K \odot \ldots \odot B_1) 1_n, vec(\mathcal{X}) \rangle \end{aligned} \tag{5.4}$$

where $\mathcal{X} \in R^{p_1 \times \ldots \times p_K}$ is a K-dimensional array variate, Z is a vector-valued covariate with weights, 1_n is an array off N ones. $\gamma \in R^{p_0}$, and $B_K = [\beta_K^{(1)} \beta_K^{(2)} \ldots \beta_K^{(r)}] \in R^{p_D \times r}$ is the array of multilinear regression coefficients.

The Maximum Likelihood (ML) method for estimating the parameters B_K is employed. The likelihood function for given data $(y_i, \boldsymbol{\mathcal{X}}_i, Z_i)$, $i = 1, \cdots, n$ is;

$$\mathcal{L}(\alpha, \gamma, B_1, \cdots .B_K) = \sum_{i=1}^{n} \frac{y\theta - b(\theta)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \tag{5.5}$$

Where θ_i is related to regression parameters $(\alpha, \gamma, B_1, \cdots .B_K)$ through Eq. 5.4. For maximizing $\mathcal{L}(\alpha, \gamma, B_1, \cdots .B_K)$ a block relaxation algorithm is used (Appendix A.1).

The algorithm uses a known rank when estimating B. It is very important to use the correct rank in tensor models. Finding the right rank is considered as a model selection problem so Bayesian Information Criterion (BIC) can be used for this purpose.BIC is a criterion for model selection among a finite set of models and formulated as follows;

$$BIC = -2\log(\theta) + \log(n)p_i \tag{5.6}$$

p_i is the effective number of parameters. For the model (Eq. 5.4): $p_i = R \times (p_1 + p_2) - R^2 for K = 2$ and $p_i = R(\sum_k p_k - K + 1)$ for $K > 2$ (K=dimension).

In statistics, BIC is a criterion for model selection among a finite set of models which is based on the likelihood function. When fitting models, the likelihood by adding parameters is increased, but this usually results in overfitting. The BIC tries to solve this problem by introducing a penalty term for the number of parameters in the model. The smaller the BIC, the better will be the fitting.

5.1.1 Simulations for CP Tensor Regression

The simulations were performed for two shapes: Ellipse and 'L' shaped figure. For each shape, two simulations were conducted ; i) 500 and ii) 1000 univariate responses y_i with a normal distribution where the mean $\mu_i = \gamma^T z_i + \langle \mathbf{B}, x_i \rangle$ where $\gamma = 1_5$. The inner product of two arrays is defined as $\langle \mathbf{B}, \mathbf{X} \rangle = \langle vec\mathbf{B}, vec\mathbf{X} \rangle = \sum_{i_1 \cdots i_K} \beta_{i_1 \cdots i_K} x_{i_1 \cdots i_K}$. The coefficient array $\mathbf{B}$ is binary, where the actual signal area is equal to one and the remaning area is zero. The regular covariate z_i and image covariate x_i are randomly generated and all elements are independent standard normals. The aim is to identify the actual signal area in $\mathbf{B}$ using data y_i, x_i, and z_i. The accuracies of the estimates in the simulations were checked with BIC values.

It is clearly seen that the larger sample size is always an advantage and yields a better estimate (Figure 5.1 compared with Figure 5.2 and Figure 5.3 compared with Figure 5.4). When each simulation is evaluated in itself, the elliptical shape gives better results in high rank estimates. BIC values are smaller at rank=3 estimations (Figure 5.1,Figure 5.2). The 'L' shaped figure gives better results in rank=2 estimates . BIC values are smaller at rank=2 estimations (Figure 5.3 and Figure 5.4).

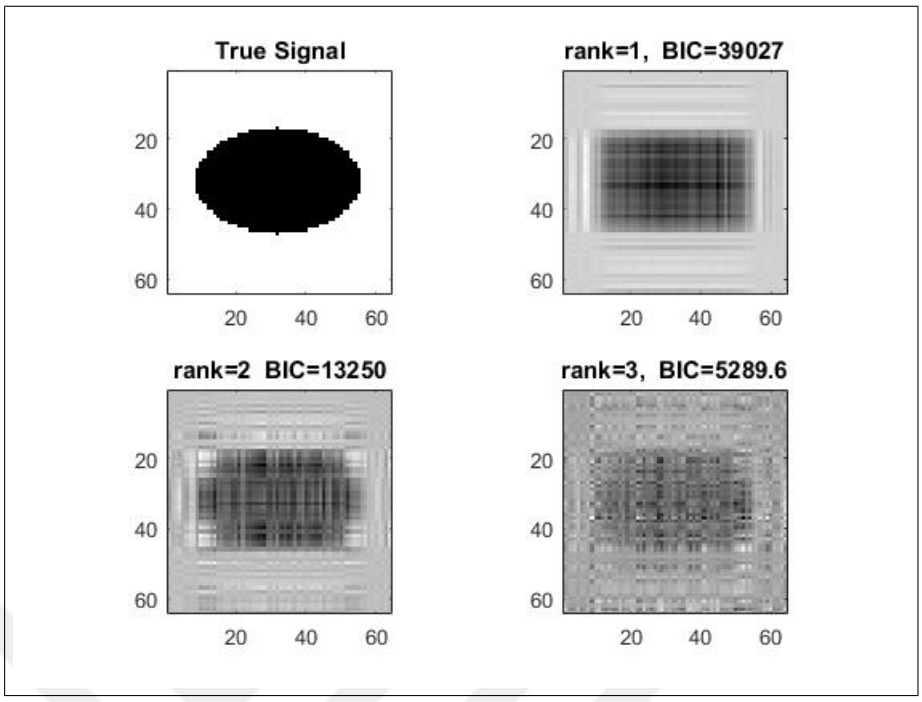

Figure 5.1 Simulation 1: Tensor regression of ellipse shape (Sample size n=500, level= 1,2,3 rank estimations.)

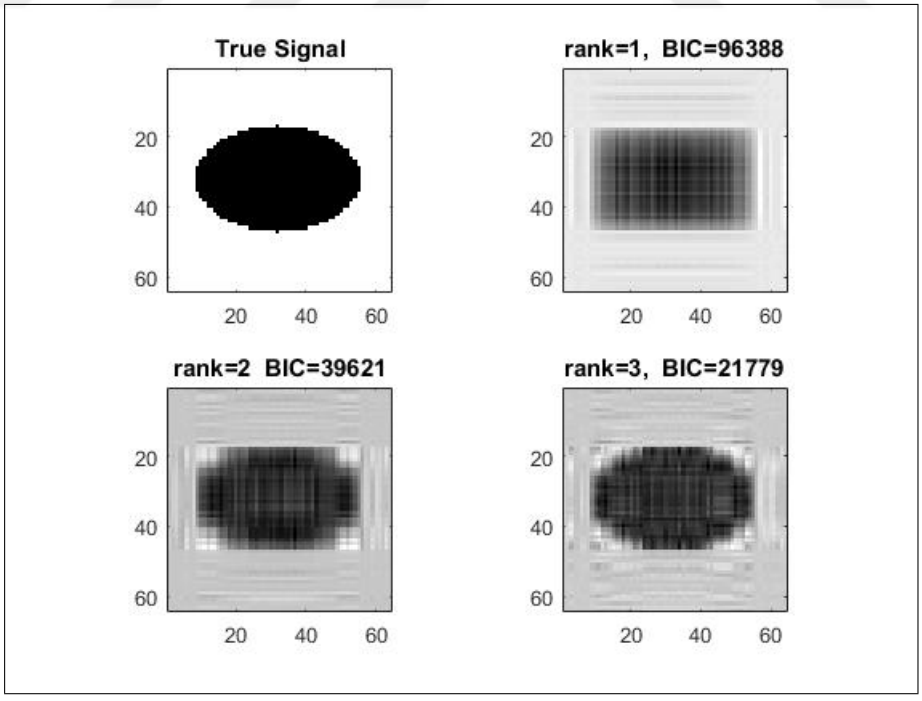

Figure 5.2 Simulation 2 :Tensor regression of ellipse shape (Sample size n=1000, level= 1,2,3 rank estimations).

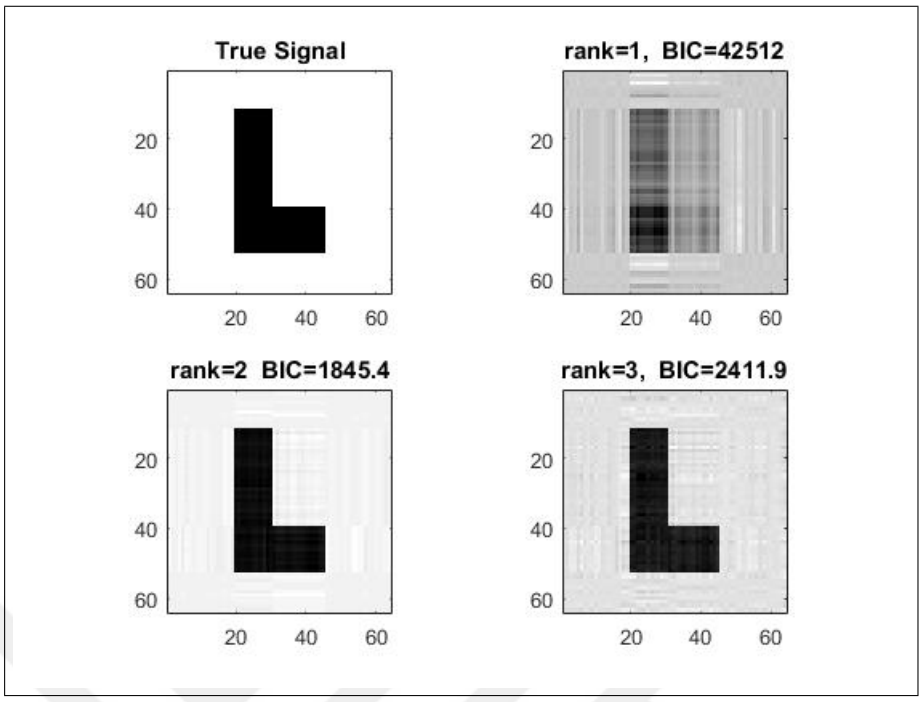

Figure 5.3 Simulation 3 :Tensor regression of 'L' Shape (Sample size n=500, level= 1,2,3 rank estimations).

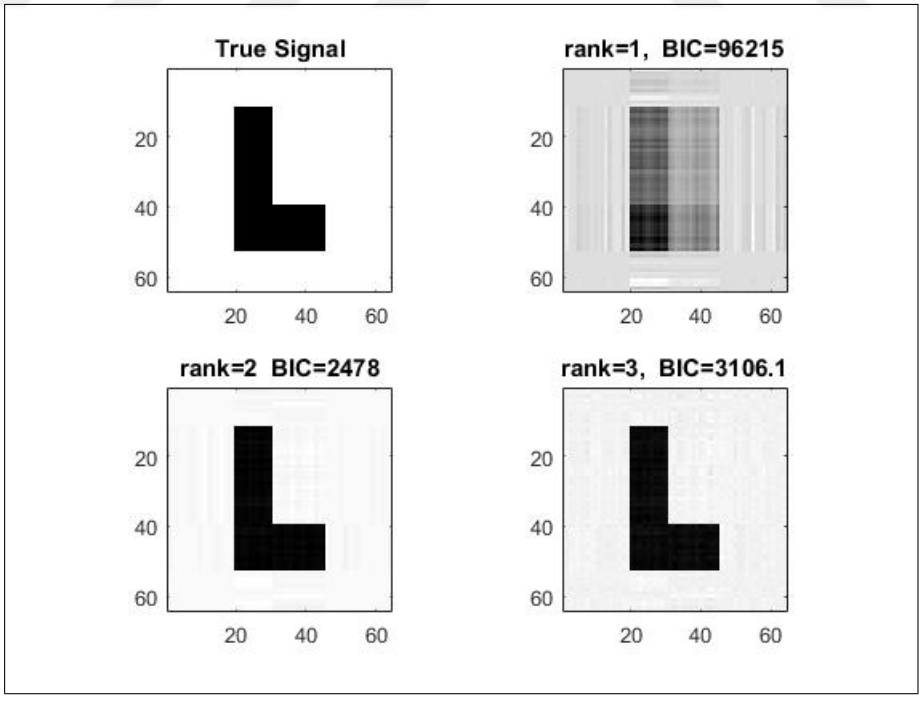

Figure 5.4 Simulation 4 :Tensor regression of 'L' shape (Sample size n=1000, level= 1,2,3 rank estimations).

5.2 Logistic Tensor Regression for Classification

Logistic regression is one of the most frequently used classification methods commonly used in neuroscience applications which aims to model a logit transformation as a linear combination of covariates. Tensor Logistic Regression has several advantages over vector or matrix logistic regression as it preserves the spatial information of data and its high dimensional structure.

Logistic tensor regression allows for classifying multidimensional images into different clinical categories by incorporating the pertinent information such as age, gender, treatment and other possible covariates or factors. As an example, a typical 3-D Magnetic Resonance Image (MRI) require $256 \times 256 \times 256 = 16,777,216$ regression parameters in a classication regression formulation. In contrast, a tensor based multilinear model requires only $r \times (256 + 256 + 256) = r \times 768$ regression parameters for a rank r which is typically chosen as a small integer like $r = 1, 2, 3,$ or 4. This leads to a substantial reduction in dimension and computational cost.

The logistic tensor regression model is expressed at Appendix A.2.

5.2.1 Simulations for Logistic Tensor Regression

The simulations were performed for two shapes: Ellipse and 'L' shaped figures. For each shape, two set of simulation data were generated; i) 500 and ii) 1000 univariate responses y_i according to a binomial model with mean $\mu_i = \gamma^T z_i + \langle \mathbf{B}, x_i \rangle$ where $\gamma = 1_5$. The inner product between two arrays is defined as $\langle \mathbf{B}, \mathbf{X} \rangle = \langle vec\mathbf{B}, vec\mathbf{X} \rangle = \sum_{i_1 \cdots i_D} \beta_{i_1 \cdots i_D} x_{i_1 \cdots i_D}$. The coefficient array $\mathbf{B}$ is binary, where the actual signal area is equal to one and the the remaining area is zero. The regular covariate z_i and image covariate x_i are randomly generated and all elements are independent standard normals. The aim is to identify the actual signal area in $\mathbf{B}$ using data y_i , x_i, and z_i . The accuracy of the estimates in the simulations were checked with BIC values.

It is clearly seen that the larger sample size is always an advantage and yields a better estimate (Figure 5.5 compared with Figure 5.6 and Figure 5.7 compared with Figure 5.8). When each simulation is evaluated in itself, the elliptical shape gives better results in low rank estimates. BIC values are smaller at rank=1 estimations (Figure 5.5,Figure 5.6). The 'L' shaped figure gives better results in rank=1 estimates. BIC values are smaller at rank=2 estimations (Figure 5.7, Figure 5.8).

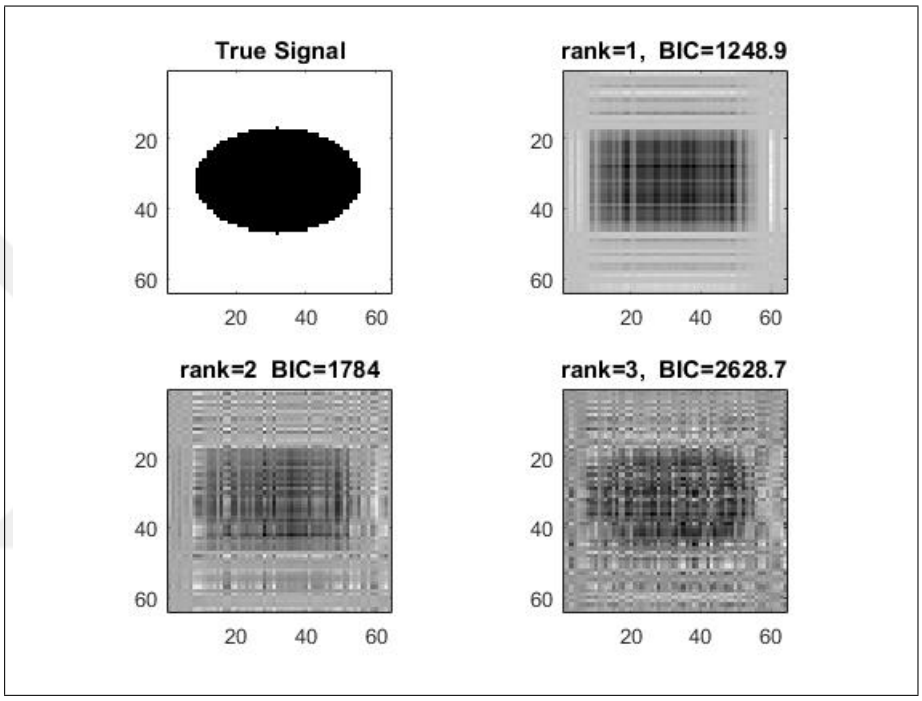

Figure 5.5 Simulation 1: Tensor logistic regression of ellipse (Sample size n=500, level= 1,2,3 rank estimations).

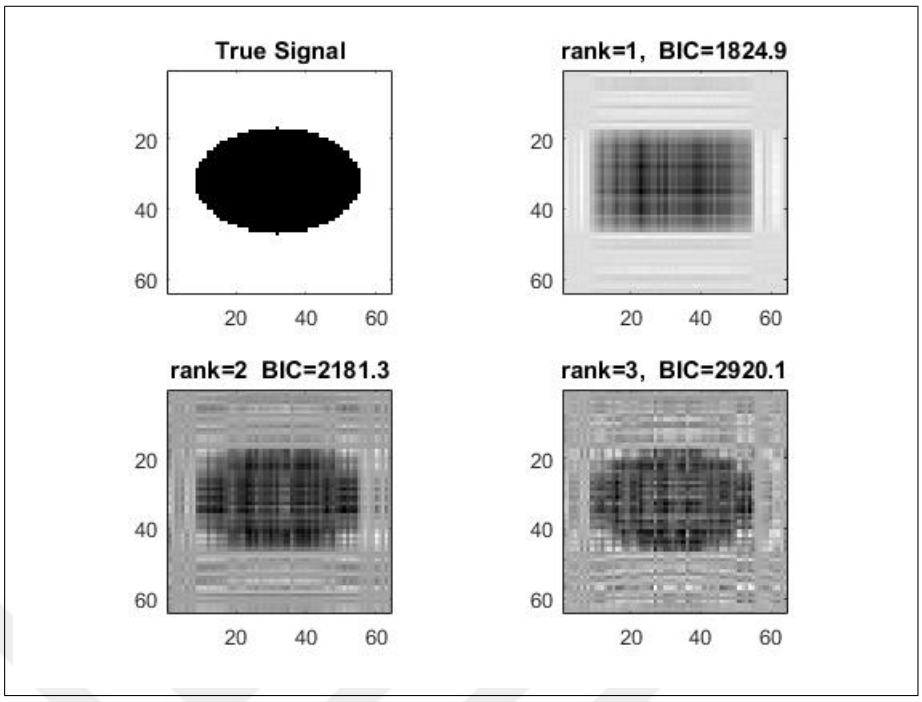

Figure 5.6 Simulation 2: Tensor logistic regression of ellipse (Sample size n=1000, level= 1,2,3 rank estimations).

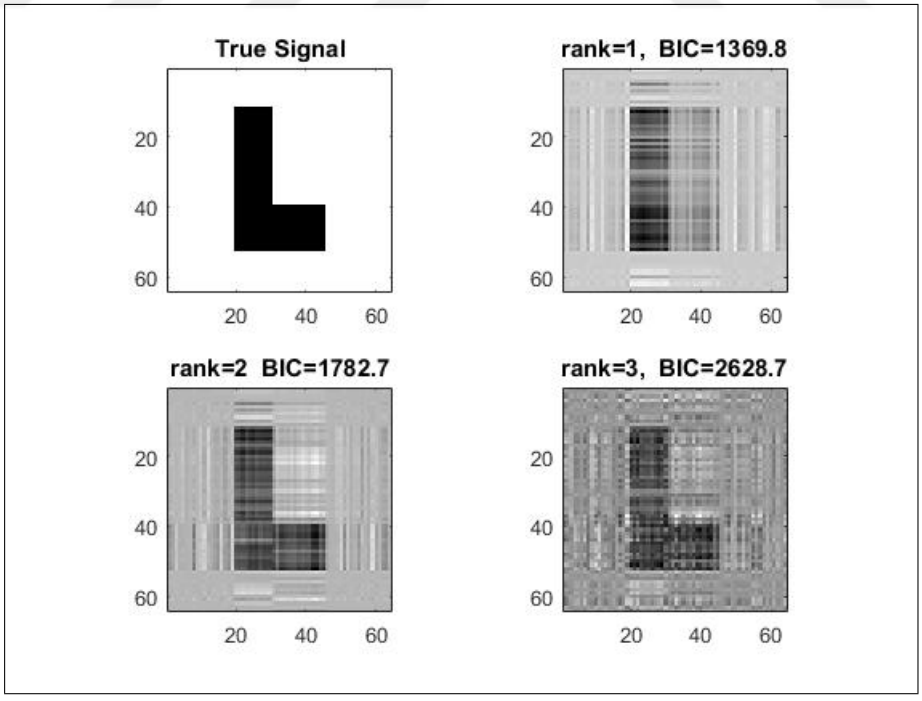

Figure 5.7 Simulation 3 :Tensor logistic regression of 'L' shape (Sample size n=500,level= 1,2,3 rank estimations).

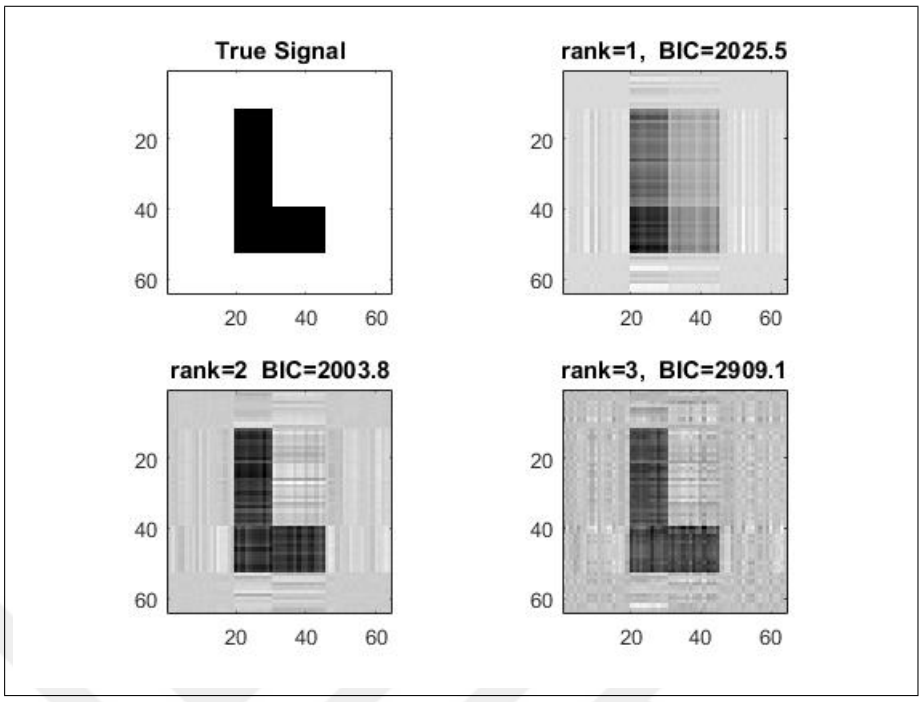

Figure 5.8 Simulation 4 :Tensor logistic regression of 'L' shaped Figure (Sample size n=1000, level= 1,2,3 rank estimations).

6. TENSORIAL DATA ANALYSIS

6.1 EEG Motor Movement/Imagery BCI Data Analysis

This dataset is taken from PhysioNet [44] which was created by BCI2000 developers [45] The dataset is obtained from 109 volunteers, as explained below:

64-channel EEG signals recorded via BCI2000 system with 160 Hz sample rate while subjects were performed four motor/imagery tasks . The electrode locations are schematised in Figure 6.1. Each subject performed 14 experimental runs: two one-minute baseline runs (one with eyes open, one with eyes closed), and three two-minute runs of each of the four following tasks:

1. Move the right or left hand against the target respectively that appears on the right or left side of the screen.
2. Imagine moving the right or left hand against the target respectively that appears on the right or left side of the screen.
3. Open or close your fists or move your feet against the target respectively that appears at the top or bottom of the screen .
4. Imagine opening and closing your fists against the target respectively that appears at the top or bottom of the screen, or imagining moving your feet.

Here, the analysis was made on a subject-based basis. For each subject, the EEG records of the first two of the above-numbered tasks were matricized. In total, x-dimensional matrix of task 1 and y-dimensional matrix of task 2 were created for each subject. This was repeated for 109 subjects.

For each subject, we obtained a data structure $\boldsymbol{\mathcal{X}} \in R^{196\times64\times656}$ with dimensions defining the $trial \times electrode \times time$ and an associated binary output $Y \in$

$\{-1,+1\}^{196\times1}$before feeding it into the algorithm. The test $\boldsymbol{\mathcal{X}}_{test}(i,:,:) \in R^{64\times656}$ (%25) and the train $\boldsymbol{\mathcal{X}}_{train}(i,:,:) \in R^{64\times656}$ (%75) portions of the data were used to classify them into a binary status $Y(i)$ indicating whether the task is the first or the second. A rank-3 tensor was fit for $B_D \in R^{64\times656}$. The vector-covariates Z corresponding to the linear part of the model was not included in the model.

The number of parameters in the multilinear model is only $p = 3\times(64+656) = 2160$. This is much smaller than $p_0 = 64 \times 656 = 41984$ parameters in the classical GLM [4].

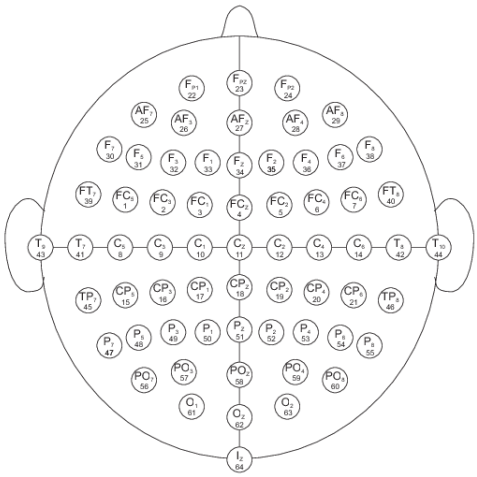

Figure 6.1 EEG Electrode locations of 64 channel recording.

6.1.1 Performance Comparison on Classification for BCI Data

The analyses were performed using MATLAB 2018a. For the tensor regression model "Matlab TensorReg Toolbox Version 1.0" developed by Zhou [4] was used. SVM was implemented using the MATLAB routines. The results are taken from a standard notebook with 2.6 GHz Intel i5 CPU. Total analysis took about 65 minutes for the multilinear regression. For the SVM classsification MATLAB was also used. For the purpose of training, Matlab code '$fitcsvm$' was used, while for classification, Matlab code '*predict*' was used. The analysis took 20 minutes for the SVM classification.

6.1.1.1 Tensor Regression Model Results . 100 different runs of the same experiment were performed by randomly selecting %75 of the 196 trials for training and %25 for test for each subject. The ratio of the pleasant trials to unpleasant ones were identically %50 for both train and test sets. A rank-3 tensor $B_D \in R^{(64\times656)}$ was fit for each train set and is used for classification with the $\boldsymbol{\mathcal{X}}$ and Y in the test set. The maximum accuracy, obtained for subject #13, was 89,1% and the average accuracy determined over 100 runs for the 109 subjects was 79,6%.

6.1.1.2 Support Vector Machine Classification Results . The classification was repeated using the SVM algorithm using the same data set for a comparison with the multilinear regression. The maximum accuracy, obtained for subject #56, was 73,9% and the average accuracy determined over 100 runs for the 109 subjects was 62,3%.

The 109 subjects' classification summary results using the multilinear regression and SVM were given in (Table 6.1).

Table 6.1
EEG motor movement/imagery BCI data clasification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	% 81,3	% 62,9
Subject 2	% 79,4	% 68,7
Subject 3	% 77,6	% 60,9
Subject 4	% 83,5	% 61,3
⋮	⋮	⋮
Subject 109	% 85,7	% 64,11
Average	% 79,6	% 62,3

6.2 EEG Emotion BCI Data Analysis

6.2.1 Dataset

The EEG data was obtained from the Electrophysiology laboratory at the Istanbul Medical School in Istanbul University. Event Related Potential (ERP) data was recorded with a sampling frequency of 250 Hz from 13 healthy subjects, whose mean age was 27.4 ($\pm$2.96). Approximately 265 pleasant and 265 unpleasant pictures with positive or negative emotional valence were presented as standard stimuli in two separate sessions (Figure 6.2). Each picture was displayed in the monitor for one second and inter-stimulus interval was two seconds. Pictures were selected from the IAPS dataset with the mean valence level differing (7.13/2.96), but the mean arousal level being equal for the two sessions (4.99/5.02).

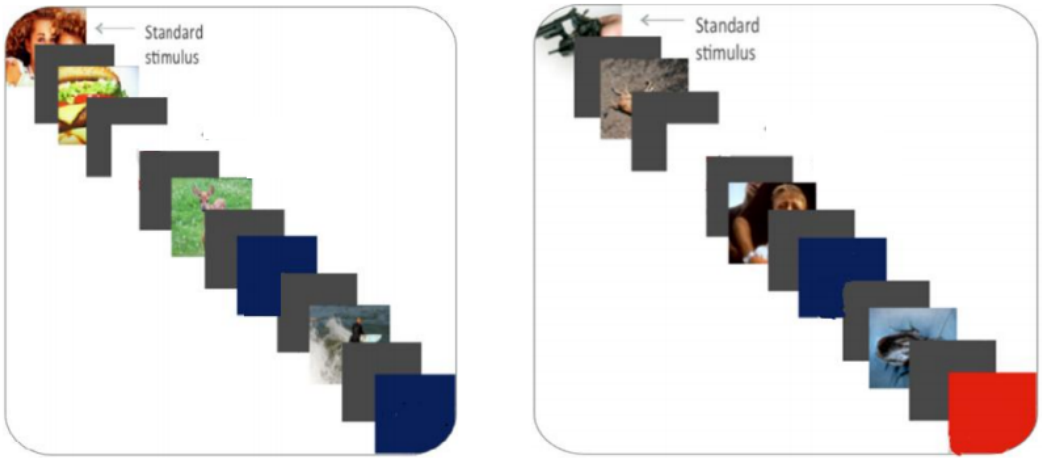

Figure 6.2 Emotional stimuli - pleasent and unpleasent pictures.

Labels and locations of recorded 30 channels are shown in Figure 6.3.

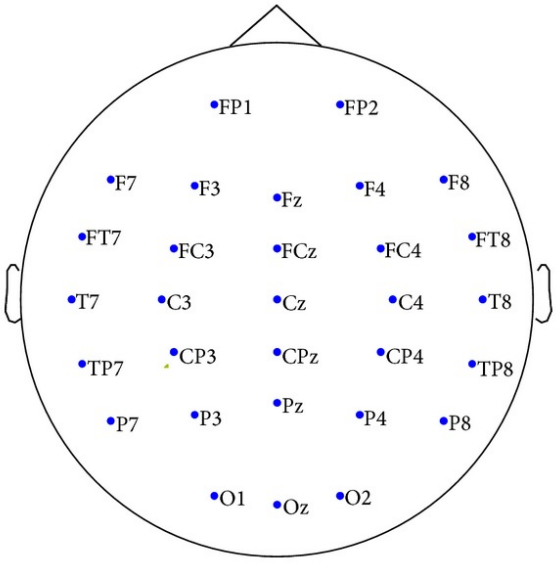

Figure 6.3 EEG electrode locations of the 30 channel recording.

6.2.2 Wavelet Decomposition

The data was decomposed into five frequency bands using a Daubechies-4 wavelet basis in the form of wavelet packet transform (DWPT): 0 Hz-4 Hz frequency band (Delta), 4 Hz-8 Hz frequency band (Theta), 8 Hz-12 Hz frequency band (Alpha), 0 Hz-8 Hz frequency band and 12 Hz-25 Hz frequency band (Beta).

For each subject, we obtained a data structure $\boldsymbol{\mathcal{X}} \in R^{530\times30\times250}$ with dimensions defining the *trial* $\times$ *electrode* $\times$ *time*. Before feeding it into the algorithm with the associated binary output $Y \in \{-1,+1\}^{530\times1}$, this data structure was converted to topographic maps using "topoplot" function in EEGLAB function with dimensions $\boldsymbol{\mathcal{X}} \in R^{32\times32\times250\times530}$ for each frequency band of each subject (Figure 6.4). The test $\boldsymbol{\mathcal{X}}_{test}(i,:,:,:) \in R^{32\times32\times250}$ (%25) and the train $\boldsymbol{\mathcal{X}}_{train}(i,:,:,:) \in R^{32\times32\times250}$ (%75) portions of the data were used to classify them into a binary status $Y(i)$ indicating whether the stimulus was pleasant or unpleasant. A rank-3 tensor was fit for $B_D \in R^{32\times32\times250}$. The vector-covarites Z corresponding to the linear part of the model was not included in the model.

The same data structure $\boldsymbol{\mathcal{X}} \in R^{32\times32\times250\times530}$ and $Y \in \{-1,+1\}^{530\times1}$ was used

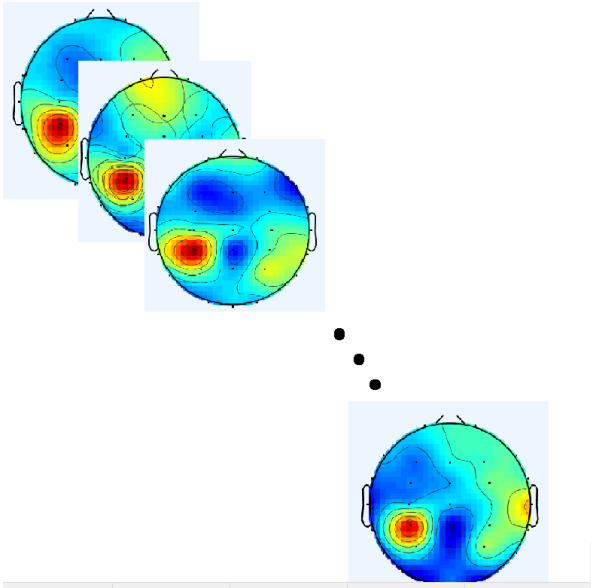

Figure 6.4 Example of a topographic graphs for one subject's one trial ($\boldsymbol{\mathcal{X}}(:,:,:,1) \in R^{32\times32\times250\times1}$) at beta frequency band.

as input of Support Vector Machine (SVM) algorithm for a comparison. The radial basis function (RBF) was chosen as the kernel and the scale parameter σ was chosen as 6 [46].

6.2.3 Performance Comparison on Classification for BCI Data

The analyses were performed using MATLAB 2018a. For the tensor regression model "Matlab TensorReg Toolbox Version 1.0" developed by Zhou [4] was used. SVM was implemented using the MATLAB routines. The results are taken from a standard notebook with 2.6 GHz Intel i5 CPU. Total analysis took about 15 minutes for the multilinear regression. For the SVM, training was done using MATLAB function '*fitcsvm*' and classification with '*predict*'. The analysis took about 5 minutes for the SVM classification.

6.2.3.1 Tensor Regression Model Results . 100 different runs of the same experiment were performed by randomly selecting %75 of the 530 trials for training and %25 for test for each subject. The ratio of the pleasant trials to unpleasant ones were

identically %50 for both train and test sets. A rank-3 tensor $B_D \in R^{(32\times32\times250)}$ was fit for each train set and was used for classification with the $\boldsymbol{\mathcal{X}}$ and Y in the test set. The average classification accuracies determined over 100 runs for several frequency bands were;

1. 0-4Hz Average Classification Accuracy : **61,13%**
2. 0-8Hz Average Classification Accuracy : **71,21%**
3. 4-8Hz Average Classification Accuracy : **64,59%**
4. 8-12Hz Average Classification Accuracy : **80,17%**
5. 12-25Hz Average Classification Accuracy : **89,93%**.

6.2.3.2 SVM Classification Results . The classification was repeated using the SVM algorithm using the same data set for a comparison with the multilinear regression. The average classification accuracies determined over 100 runs for several frequency bands were ;

1. 0-4Hz Average Classification Accuracy : **56,05%**
2. 0-8Hz Average Classification Accuracy : **54,82%**
3. 4-8Hz Average Classification Accuracy : **53,33 %**
4. 8-12Hz Average Classification Accuracy : **58,04%**
5. 12-25Hz Average Classification Accuracy : **75,59%**.

Accuracy rates of each subject according to frequency bands are given in the Table 6.2, Table 6.3, Table 6.4, Table 6.5 and Table 6.6 with both multilinear regression algorithm and SVM.

Table 6.2
0Hz- 4Hz frequency band EEG emotion BCI data classification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	60,91 %	54,23 %
Subject 2	63,11 %	56,33 %
Subject 3	56,16%	53,85 %
Subject 4	61,47%	54,06 %
Subject 5	60,54%	58,42%
Subject 6	57,78%	58,20 %
Subject 7	59,26%	56,21 %
Subject 8	60,58%	51,14%
Subject 9	66,08%	56,34%
Subject 10	64,70%	58,80%
Subject 11	57,70 %	55,72 %
Subject 12	65,16%	58,23 %
Subject 13	61,23%	57,19 %
Average	61,13 %	56,05 %

Table 6.3
0Hz- 8Hz frequency band EEG emotion BCI data classification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	69,18%	55,61 %
Subject 2	69,33%	51,82 %
Subject 3	70,99%	53,62 %
Subject 4	73,30%	52,85%
Subject 5	69,84%	61,04%
Subject 6	71,55 %	53,92 %
Subject 7	70,66%	57,20 %
Subject 8	72,68%	54,94 %
Subject 9	69,37%	57,53%
Subject 10	71,05%	53,56 %
Subject 11	71,83%	52,03%
Subject 12	72,65%	51,99%
Subject 13	73,32%	56,57%
Average	71,21%	54,82%

Table 6.4
4Hz- 8Hz frequency band EEG emotion BCI data classification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	66,38 %	49,97 %
Subject 2	64,42 %	53,20 %
Subject 3	65,56%	52,92 %
Subject 4	63,51%	54,04 %
Subject 5	66,69%	54,02 %
Subject 6	62,82%	51,67 %
Subject 7	62,97%	53,34%
Subject 8	63,46%	53,07%
Subject 9	59,41%	54,66%
Subject 10	67,73 %	55,59 %
Subject 11	65,62%	55,62 %
Subject 12	63,57%	51,67 %
Subject 13	67,60%	53,55%
Average	64,59%	53,33%

Table 6.5
8 Hz- 12 Hz frequency band EEG emotion BCI data classification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	79,73 %	58,50%
Subject 2	78,71%	59,67%
Subject 3	83,13 %	63,21%
Subject 4	78,28 %	57,03%
Subject 5	80,29%	58,66%
Subject 6	82,38%	58,14%
Subject 7	78,79 %	54,63 %
Subject 8	81,04%	57,47%
Subject 9	81,25%	54,90%
Subject 10	79,63 %	59,91 %
Subject 11	80,32%	56,61%
Subject 12	78,25%	58,49%
Subject 13	80,44 %	57,26%
Average	80,17 %	58,04%

Table 6.6
12Hz- 25Hz frequency band EEG emotion BCI data classification accuracy rates.

Succesfully classified test data [%]	Accuracy	
	Tensor Regression	SVM
Subject 1	92,50 %	74,32%
Subject 2	85,84 %	75,69%
Subject 3	88,52%	73,88%
Subject 4	89,30%	73,32%
Subject 5	91,07%	76,69 %
Subject 6	85,82%	75,30%
Subject 7	91,18%	75,21%
Subject 8	86,97%	76,67%
Subject 9	90,17%	76,19%
Subject 10	91,63 %	75,83%
Subject 11	90,82%	77,18%
Subject 12	92,71%	76,69 %
Subject 13	92,52%	75,66%
Average	89,93%	75,59%

7. DISCUSSION

A tensor regression algorithm has been implemented for BCI and its capability is demonstrated by various simulations as well as real EEG data. Two classification studies were performed using this algorithm:

1. In the first analysis, motor imagery data were classified with tensor regression algorithm and SVM. An overwhelming advantage of tensor regression algorithm over the SVM was observed with regard to the accuracy of the methods which were 79,6% and 62,3%, respectively. When similar studies in the literature were examined, in [47] they performed a similar experiment for classifying right and left hand imagination. The data were subjected to various preprocessing steps such as Laplacian spatial filtering, time-frequency analysis and bandpass filtering then Independent Component Analysis (ICA) method was used for transforming the multi dimensional vector into statistically independent vectors. These vectors then used in source analysis for classification and it ended up with a classification performance of 80%. One superiority of our approach over theirs is that, it does not require any pre-processing step and there is no need to define a prior region of interest to select the relevant electrodes for analysis since we include all the electrodes.

   In another study [48], linear and nonlinear classification methods were applied to EEG data recorded while imagining left or right hand movement. Different feature sets extracted from EEG were used as inputs into linear, neural network and Hidden Markov Model (HMM) classifiers for classifying imagery movement mental task after certain electrodes were selected. As another advantage, we do not need to perform any feature extraction method for classification in our algorithm.

2. In the second analysis, EEG signals recorded from subjects driven by visual stimuli obtained IAPS database library were subjected to wavelet decomposition and divided into several frequency bands. Emotional states in various bands were

classified as positive/negative valence. In our low frequency emotion classification study, the performance rate was 65% for theta band, 80% for alpha band and 89% for beta band. In a similar classification study with SVM [49], the performance rate for these frequencies was 81% at most. In a study conducted in 2008 [50], SVM classifying accuracy rates for the valence from EEG signals was maximally 71%. In another emotion recognition study [51], video clips were used as visual stimuli and it was found that beta band had the highest performance rate in emotion identification with a mean classification accuray rate of 91%. Although our accuracy is slightly lower than what is found in that study, not using any feature extraction method for the EEG data prior to classification is an advantage for our method.

8. CONCLUSION AND FUTURE WORK

A key motivation of our study was to test the multilinear regression framework for a typical BCI classification problem using EEG data and to compare it to a more conventional SVM method. Our results proved to be superior to that obtained by SVM. Furthermore, the multilinear approach allows for a parsimonius regression coefficients compared with those in a general regression in a multidimensional data setting. The accuracy of the method can be improved in certain ways if the multidimensional nature of the EEG data can be exploited by using spatial and spectral transforms which will augment the dimensionality but preserve its features which are lost in the process of its vectorization in a classical feature space. The downside of the method is its computational exorbitance compared with the faster SVM. However, if more efficient algorithms are devised, the multilinear regression may be a more attractive alternative in BCI applications due its better performance.

In this study, two different analyzes were performed. In the first, the data were not subjected to any preprocessing algorithms. In the second, the raw data were divided into frequency bands by DWPT and feeded into the algorithm. Morlet Wavelet as a symmetric filter can be used [52] for preprocessing for time-frequency analysis. The topographic 2-D data can be substituted by a 3-D source reconstructed data by using EEG inverse modeling and the tensor decompositions may be repeated using source space data instead of sensor space. Also the algorithm needs to be more efficiently implemented to be comparable with other methods like SVM in terms of computational demand. The application area of the tensor classsification method is not limited to the EEG or neurosignal analysis. This algorithm may be used for all 'high dimensional' data types such as fMRI and DTI.

APPENDIX A. ALGORITHMS USED IN THESIS

A.1 Tensor Regression Algorithm

Algorithm 1 Tensor Regression

1: Initialize : $(\alpha^{(0)}, \gamma^{(0)}) = argmax_{\alpha,\gamma}\mathcal{L}(\alpha, \gamma, 0, \cdots .0)$, $\mathcal{B}_k^{(0)} \in P_k \times R$ a random matrix for $k = 1, \cdots, K$

2: **repeat**

3: **for** $k = 1, \cdots, K$ **do**

4: $\mathbf{B}_k^{(t+1)} = argmax_{\mathbf{B}_k}\mathcal{L}(\alpha^{(t)}, \gamma^{(t)}, \mathbf{B}_1^{(t+1)}, \cdots, \mathbf{B}_{(k-1)}^{(t+1)}, \mathbf{B}_{(d)}, \mathbf{B}_{(k+1)}^{(t)}, \cdots \mathbf{B}_K^{(t)})$

5: **end for**

6: $(\alpha^{(t+1)}, \gamma^{(t+1)} = argmax_{\alpha,\gamma}\mathcal{L}(\alpha, \gamma, \mathbf{B}_1^{(t+1)}, \cdots, \mathbf{B}_K^{(t+1)})$

7: **until** $\mathcal{L}(\theta^{(t+1)} - \mathcal{L}(\theta^{(t)} < \epsilon$

In the algorithm, only the terms which depend on $d = 1, 2, \cdots, D$ change, the rest of the components is fixed. This is a feature of the block relaxation algorithm. What makes this algorithm attractive is that updating the B_d block in each iteration.

A.2 Tensor Logistic Regression Algorithm

Tensor Logistic Regression :

$$\mathcal{L}(\alpha, \gamma, B_1, \cdots .B_K) = -\sum_{i=1}^{n} log(1 + exp(-y_i(\langle \mathcal{B}, \mathcal{X} \rangle) + \gamma)) \tag{A.1}$$

where $\mathcal{X} \in R^{P_1 \times P_2 \times \cdots \times P_N}$ is the tensorial data and $y_i \in \{-1, +1\}$ is its corresponding class label.

Algorithm 2 Tensor Logistic Regression

Input: The tensorial data $\mathcal{X} \in R^{P_1 \times P_2 \times \cdots \times P_N}$ and their corresponding class labels $y_i \in \{-1, +1\}$ Output: The regression coefficient tensor $\mathcal{B}$ and the intercept γ

1: Initialize : $(\alpha^{(0)}, \gamma^{(0)}) = argmax_{\alpha,\gamma}\mathcal{L}(\alpha, \gamma, 0, \cdots .0)$, $\mathcal{B}_k^{(0)} \in P_k \times R$ a random matrix for $k = 1, \cdots, K$

2: **repeat**

3: **for** $k = 1, \cdots, K$ **do**

4: $\mathbf{B}_k^{(t+1)} = argmax_{\mathbf{B}_k}\mathcal{L}(\alpha^{(t)}, \gamma^{(t)}, \mathbf{B}_1^{(t+1)}, \cdots, \mathbf{B}_{(k-1)}^{(t+1)}, \mathbf{B}_{(k)}, \mathbf{B}_{(k+1)}^{(t)}, \cdots \mathbf{B}_K^{(t)})$

5: **end for**

6: $(\alpha^{(t+1)}, \gamma^{(t+1)} = argmax_{\alpha,\gamma}\mathcal{L}(\alpha, \gamma, \mathbf{B}_1^{(t+1)}, \cdots, \mathbf{B}_K^{(t+1)})$

7: **until** $\mathcal{L}(\theta^{(t+1)} - \mathcal{L}(\theta^{(t)} < \epsilon$

REFERENCES

1. Ang, K., C. Guan, K. Chua, B. Ti Ang, C. Kuah, C. Wang, K. S. Phua, Z. Yang Chin, and H. Zhang, "Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* Vol. 2010, pp. 5549–52, Agu 2010.

2. Pfurtscheller, G., G. Müller-Putz, S. Reinhold, and C.Neuper, "Rehabilitation with brain-computer interface systems," *Computer*, Vol. 41, pp. 58 – 65, Nov 2008.

3. Nijholt, A., "BCI for games: A 'state of the art' survey," *Journal of Electromyography and Kinesiology* ,Vol. 5309, pp. 225–228, Sep 2008.

4. Zhou, H., L. Li, and H. Zhu, "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, Vol. 108, pp. 540–552, Jul 2013.

5. Lotte, F., C. Nam, and A. Nijholt, "Study of Electroencephalographic Signal Processing and Classification Techniques towards the use of Brain-Computer Interfaces in Virtual Reality Applications" *Introduction: Evolution of Brain-Computer Interfaces*, pp. 1–8. Jan 2018.

6. Fetz, E., "Operant conditioning of cortical unit activity," *Science*, Vol. 163, no. 3870, pp. 955–958, 1969.

7. Arafat, I., "Brain-computer interface: past, present & future," *International Islamic University Chittagong (IIUC)*, 2013.

8. S Kotchetkov, I., B. Y Hwang, G. Appelboom, C. Kellner, and E. Connolly, "Brain-computer interfaces: Military, neurosurgical, and ethical perspective," *Neurosurgical focus*, Vol. 28, p. E25, May 2010.

9. Galan, F., M. Nuttin, E. Lew, G. Vanacker, J. Philips, and J. d. R. Millan, "A brain-actuated wheelchair: Asynchronous and non-invasive brain-computer interfaces for continuous control of robots," *Clinical Neurophysiology*, Vol. 119, Agu 2008.

10. Coyle, S., T. Ward, and C. Markham, "Brain computer interface using a simplified functional near-infrared spectroscopy system," *Journal of Neural Engineering*, Vol. 4, pp. 219–26, Nov 2007.

11. Sitaram, R., A. Caria, R. Veit, T. Gaber, G. Rota, A. Kübler, and N. Birbaumer, "fMRI brain-computer interface: A tool for neuroscientific research and treatment," *Computational Intelligence and Neuroscience*, Vol. 2007, pp. 297 – 298, 2007.

12. Gao, Q., L. Dou, A. N. Belkacem, and C. Chao, "Noninvasive electroencephalogram based control of a robotic arm for writing task using hybrid BCI system," *BioMed Research International*, Vol. 2017, pp. 1–8, Jun 2017.

13. Fukuda, O., T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motion," *Robotics and Automation, IEEE Transactions on*, Vol. 19, pp. 210 – 222, May 2003.

14. Burle, B., L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, "Spatial and temporal resolution of EEG: Is it really black and white? a scalp current density view," *International Journal of Psychophysiology*, Vol. 8, MAy 2015.

15. Zander, T., and C. Kothe, "Towards passive brain computer interfaces applying brain computer interface technology to human-machine systems in general," *Journal of Neural Engineering*, Vol. 8, p. 025005, March 2011.

16. Nijboer, F., S. Carmien, E. Leon, F. Morin, R. Koene, and U. Hoffmann, "Affective brain-computer interfaces: Psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sclerosis," *Affective Computing and Intelligent Interaction and Workshops* , pp. 1 – 11, Oct 2009.

17. Ahn, M., M. Lee, J. Choi, and S. Jun, "A review of brain-computer interface games and an opinion survey from researchers, developers and users," *Sensors (Basel, Switzerland)*, Vol. 14, pp. 14601–14633, Agu 2014.

18. Dokkum, L., T. Ward, and I. Laffont, "Brain computer interfaces for neurorehabilitation its current status as a rehabilitation strategy post-stroke," *Annals of Physical and Rehabilitation Medicine*, Vol. 58, Jan 2015.

19. Liang, S.-F., F. Shaw, C.-P. Young, D.-W. Chang, and Y.-C. Liao, "A closed-loop brain computer interface for real-time seizure detection and control," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, Vol. 2010, pp. 4950–3, Agu 2010.

20. Little, S., A. Pogosyan, S. Neal, B. Zavala, L. Zrinzo, M. Hariz, T. Foltynie, P. Limousin, K. Ashkan, J. FitzGerald, A. L Green, T. Z Aziz, and P. Brown, "Adaptive deep brain stimulation in advanced parkinson disease," *Annals of Neurology*, Vol. 74, Apr 2013.

21. Wioleta, S., "Using physiological signals for emotion recognition," *International Conference on Human System Interactions* , pp. 556–561, Jun 2013.

22. Lahane, P., and A. K. Sangaiah, "An approach to EEG based emotion recognition and classification using kernel density estimation," *Procedia Computer Science*, Vol. 48, pp. 574 – 581, 2015. International Conference on Computer, Communication and Convergence (ICCC 2015).

23. Szwoch, W., "Emotion recognition using physiological signals," in *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*, MIDI '15, (New York, NY, USA), pp. 15:1–15:8, ACM, 2015.

24. Lin, Y.-P., C.-H. Wang, T.-L. Wu, S.-K. Jeng, and J. Chen, "EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine," *IEEE International Conference on Acoustics, Speech and Signal Processing* pp. 489–492, Apr 2009.

25. Kortelainen, J., E. Vayrynen, and T. Seppanen, "High-frequency electroencephalographic activity in left temporal area is associated with pleasant emotion induced by video clips," *Computational Intelligence and Neuroscience*, Vol. 2015, p. 762769, Apr 2015.

26. Min, Y.-K., S.-C. Chung, and B.-C. Min, "Physiological evaluation on emotional change induced by imagination," *Applied Psychophysiology and Biofeedback*, Vol. 30, pp. 137–50, Jul 2005.

27. Lotte, F., "Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications," *INSA de Rennes* Dec 2008.

28. Ekman, P., "An argument for basic emotions," *Cognition and Emotion*, Vol. 6, no. 3-4, pp. 169–200, 1992.

29. A Russell, J., "Core affect and the psychological construction of emotion," *Psychological Review*, Vol. 110, pp. 145–72, Feb 2003.

30. Soleymani, M., M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, Vol. 3, pp. 211–223, Apr 2012.

31. Lang, P. J., M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Technical manual and affective ratings," tech. rep., University of Florida, 1999.

32. Liu, Y.-H., W.-T. Cheng, Y.-T. Hsiao, C.-T. Wu, and M.-D. Jeng, "EEG-based emotion recognition based on kernel fisher's discriminant analysis and spectral powers," *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 2014, pp. 2221–2225, Oct 2014.

33. G. Kolda, T., and B. W. Bader, "Tensor Decompositions and Applications" *SIAM Review* pp. 455–500, Agu 2009.

34. Sidiropoulos, N., L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, Vol. PP, Jul 2016.

35. Li, X., H. Zhou, and L. Li, "Tucker tensor regression and neuroimaging analysis," *Statistics in Biosciences*, Apr 2013.

36. Carroll, J. D., and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, Vol. 35, pp. 283–319, Sep 1970.

37. Harshman, R. A., "Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-model factor analysis," *UCLA working papers in phonetics* 1970.

38. Kiers, H., "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics - J CHEMOMETR*, Vol. 14, pp. 105–122, May 2000.

39. Rabanser, S., O. Shchur, and S. Gunnemann ," Introduction to tensor decompositions and their applications in machine learning ",*ArXiv*, Nov 2017.

40. Rasmussen, C. E., "Gaussian processes for machine learning," *MIT Press*, 2006.

41. Haykin, S., "Neural Networks: A Comprehensive Foundation (3rd Edition)", *Prentice-Hall*, Inc., 2007.

42. Quinlan, J. R., "Induction of decision trees," *Machine Learning*, Vol. 1, pp. 81–106, Mar 1986.

43. Smola, A. J., V. Vapnik, H. Drucker, C. J. C. Burges, and L. Kaufman, "Support vector regression machines," in *Advances in Neural Information Processing Systems 9* (Mozer, M. C., M. I. Jordan, and T. Petsche, eds.), pp. 155–161, MIT Press, 1997.

44. Goldberger, A., L. Amaral, L. Glass, S. Havlin, J. Hausdorg, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, H. Stanley, and P. Physiobank, "Components of a new research resource for complex physiologic signals," *PhysioNet*, Vol. 101, Jan 2000.

45. Schalk, G., D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, Vol. 51, pp. 1034–1043, June 2004.

46. Panda, R., P. Khobragade, P. Jambhule, S. Jengthe, P. Pal, and T. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," *International Conference on Systems in Medicine and Biology*, pp. 405 – 408, Jan 2011.

47. Qin, L., L. Ding, and B. He, "Motor imagery classification by means of source analysis for brain-computer interface applications," *Journal of Neural Engineering*, Vol. 1, pp. 135-141, Oct 2004.

48. Akrami, A., S. Solhjoo, A. Motie-Nasrabadi, and M. . Hashemi-Golpayegani, "EEG-based mental task classification: Linear and nonlinear classification of movement imagery," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 4626–4629, Jan 2005.

49. Jatupaiboon, N., S. Pan-ngum, and P. Israsena, "Emotion classification using minimal EEG channels and frequency bands," in *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 21–24, May 2013.

50. Horlings, R., D. Datcu, and L. Rothkrantz, "Emotion recognition using brain activity," *9th International Conference on Computer Systems and Technologies*, p. 6, Jan 2008.

51. Murugappan, M., and S. Murugappan, "Human emotion recognition through short time electroencephalogram (EEG) signals using Fast Fourier Transform (FFT)," in *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, pp. 289–294, March 2013.

52. Cohen, M. X., "A better way to define and describe morlet wavelets for time-frequency analysis," *NeuroImage*, Vol. 199, pp. 81 – 86, 2019.