EMOTION RECOGNITION OF EEG DATA USING TENSOR LOGISTIC REGRESSION

by

İbrahim Cansu

B.S., in Electrical-Electronics Engineering, Istanbul Sehir University, 2017

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
2022

EMOTION RECOGNITION OF EEG DATA USING TENSOR LOGISTIC REGRESSION

APPROVED BY:

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

Prof. Dr. Burak Güçlü .....

Assist. Prof. Dr. İpek Şen .....

DATE OF APPROVAL: 26 January 2022

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my advisor Prof. Dr. Ahmet Ademoğlu for his generous support and guidance during my study. I also would like to thank Prof. Dr. Burak Güçlü and Assist. Prof. Dr. İpek Şen for their invaluable comments and time. I want to thank my colleagues in Neurosignal Analysis Laboratory. Last but not least, I would like to thank my family for their unconditional support.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, İbrahim Cansu, 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

EMOTION RECOGNITION OF EEG DATA USING TENSOR LOGISTIC REGRESSION

Emotion recognition is a research area gaining momentum in the last three decades with a strong impact on our daily life. One of the most widely used methods to study emotion recognition is using physiological signals such as EEG data. However, using physiological signals requires using feature extraction and selection methods. Moreover, there is no gold standard for choosing the best methods. Therefore, this study aims to compare the sensor space and source space EEG data for emotion recognition using tensor based methods. In order to achieve that, different frequency bands were used as features of EEG data. In addition, support vector machine (SVM) as a conventional method, and logistic tensor regression (LTR), which was a tensor-based method, were used as two different classification methods. The results showed that the gamma was the most discriminating frequency band. Also, source space data improved the accuracy rates when compared with sensor space data. Moreover, TLR was superior in the source space than SVM. In the sensor space, both methods performed similarly.

Keywords: Emotion Recognition, Tensor, Source Space, Sensor Space, Tensor Logistic Regression, Support Vector Machine, EEG

ÖZET

TENSÖR LOJİSTİK REGRESYON KULLANARAK EEG VERİLERİNDEN DUYGU TANIMA

Duygu tanıma, son otuz yılda ivme kazanan ve günlük yaşamımıza güçlü etkileri olan bir araştırma alanıdır. Duygu tanımayı incelemek için en yaygın kullanılan yöntemlerden birisi EEG verileri gibi fizyolojik sinyalleri kullanmaktır. Ancak fizyolojik sinyallerin kullanılması birçok öznitelik çıkarma ve seçme yöntemini gerektirmektedir. Üstelik bu yöntemler için bir altın standart da bulunmamaktadır. Bu nedenle, bu çalışma, tensör tabanlı yöntemler kullanarak duygu tanıma için sensör uzay ve kaynak uzay EEG verilerini karşılaştırmayı amaçlamaktadır. Bunu sağlamak için EEG verilerinin özniteliği olarak farklı frekans bantları kullanılmıştır. Ayrıca geleneksel yöntem olarak destek vektör makinesi (DVM) ve tensör tabanlı bir yöntem olan lojistik tensör regresyonu (LTR) olmak üzere iki farklı sınıflandırma yöntemi kullanılmıştır. Sonuçlar, gama bandının en ayırt edici frekans bandı olduğunu göstermiştir. Ayrıca, kaynak uzay verileri sensör uzay verileriyle karşılaştırıldığında doğruluk oranları artmaktadır. Üstelik TLR, kaynak uzayında DVM'den daha üstün gözükmektedir. Sensör uzayında, her iki yöntem de benzer sonuçlar vermiştir.

Anahtar Sözcükler: Duygu Tanıma, Tensor, Kaynak Uzayı, Sensor Uzayı, Tensor Lojistik Regresyon, Destek Karar Makineleri, EEG.

TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . iii
ACADEMIC ETHICS AND INTEGRITY STATEMENT . . . iv
ABSTRACT . . . v
ÖZET . . . vi
LIST OF FIGURES . . . viii
LIST OF TABLES . . . ix
LIST OF SYMBOLS . . . x
LIST OF ABBREVIATIONS . . . xii
1. INTRODUCTION . . . 1
2. BACKGROUND . . . 4
2.1 Emotion . . . 4
2.1.1 Emotion Models . . . 4
2.1.2 Emotion in the Brain . . . 5
2.2 Source Imaging . . . 6
2.3 Tensors . . . 8
2.3.1 Basic Tensor Product Operations . . . 9
2.3.2 Mode-n Unfolding . . . 11
2.3.3 Tensor Decomposition . . . 12
2.4 Machine Learning . . . 13
2.4.1 Tensor Logistic Regression . . . 13
2.4.2 Support Vector Machine . . . 15
3. MATERIAL AND METHODS . . . 16
3.1 Data . . . 16
3.2 Data Processing . . . 17
4. RESULTS . . . 23
5. DISCUSSION . . . 32
6. CONCLUSION AND FUTURE WORK . . . 35
APPENDIX A. BLOCK RELAXATION ALGORITHM . . . 36
REFERENCES . . . 37

LIST OF FIGURES

Figure 1.1	Article and citation numbers since 1970 [5].	3
Figure 2.1	Circumplex model of Russell in 2 dimensions [13].	5
Figure 2.2	Illustration of the forward and inverse problem of EEG [25].	8
Figure 2.3	Illustration of tensor fiber and slices [27] a) column fibers of tensor $\boldsymbol{x}_{:jk}$ b) row fibers of tensor $\boldsymbol{x}_{i:k}$ c) tube fibers of tensor $\boldsymbol{x}_{ij:}$ d) horizontal slice of tensor $\boldsymbol{X}_{i::}$ e) lateral slices of tensor $\boldsymbol{X}_{:j:}$ f) frontal slices of tensor $\boldsymbol{X}_{::k}$.	9
Figure 2.4	Example of 3-way tensor matricization [28].	12
Figure 2.5	CP decomposition of a three-way array [26].	13
Figure 2.6	Classification of two different groups which are filled and empty circles using support vector machine a) There are candidates of the hyperplane b) Calculated best hyperplane and support vectors which are surrounded with circles [38].	15
Figure 3.1	Example of stimulus representation [39].	16
Figure 3.2	Pipeline of analysis of reconstructed sources.	18
Figure 3.3	Example of reconstructed sources in the a) delta, b) theta, c) alpha, d) beta, e) gamma1, and f) gamma2 bands, respectively.	19
Figure 3.4	Pipeline of analysis of topographic maps.	21
Figure 3.5	Example of topographic maps in the a) delta, b) theta, c) alpha, d) beta, e) gamma1, and f) gamma2 bands, respectively.	22
Figure 4.1	Accuracy results and standard deviations according to frequency bands.	25

LIST OF TABLES

Table 4.1	Accuracy in delta band (0-4 Hz).	26
Table 4.2	Accuracy in theta band (4-8 Hz).	27
Table 4.3	Accuracy in alpha band (8-12 Hz).	28
Table 4.4	Accuracy in beta band (12-25 Hz).	29
Table 4.5	Accuracy in gamma1 band (25-40 Hz).	30
Table 4.6	Accuracy in gamma2 band (40-55 Hz).	31

LIST OF SYMBOLS

x	Scalar
$\boldsymbol{x}$	Vector
$\boldsymbol{X}$	Matrix
$\boldsymbol{\mathcal{X}}$	Tensor
$\boldsymbol{L}$	Leadfield Matrix
$\boldsymbol{S}$	Source space
$\boldsymbol{\varepsilon}$	Noise
C	Number of channels
T	Time points
N	Number of cortical dipole sources
$\mathcal{N}$	Normal distribution
$\boldsymbol{e}$	Error term
$\boldsymbol{C_e}$	Covariance matrix of the error
λ	Regularization parameter
$\boldsymbol{W}$	Weight matrix
$\boldsymbol{x}_{:jk}$	Mode-1 fiber
$\boldsymbol{x}_{i:k}$	Mode-2 fiber
$\boldsymbol{x}_{ij:}$	Mode-3 fiber
$\boldsymbol{X}_{i::}$	Horizontal slices
$\boldsymbol{X}_{:j:}$	Lateral slices
$\boldsymbol{X}_{::k}$	Frontal slices
$\langle ., . \rangle$	Inner product
$\circ$	Outer product
$\otimes$	Kronecker product
$\odot$	Khatri-Rao product
$*$	Hadamard product
$\boldsymbol{A}_{(n)}$	Unfolded tensor on the nth Dimension
$\boldsymbol{\beta}^{(r)}$	Rank r vector $\boldsymbol{\beta}$

$\boldsymbol{\beta}^T$	Transpose of $\boldsymbol{\beta}$
$\boldsymbol{z}$	Vector valued covariates
1_n	Vector of all ones
Y	Response variable
$\boldsymbol{\alpha}$	Intercept
$\boldsymbol{\gamma}$	Weight of vector valued covariates
$g(.)$	Link function

LIST OF ABBREVIATIONS

fMRI	Functional Magnetic Resonance Imaging
iEEG	Intracranial Electroencephalography
EEG	Electroencephalography
PET	Positron Emission Tomography
SVM	Support Vector Machine
KNN	K-Nearest Neighbor
CNN	Convolutional Neural Networks
FD	Fractial Dimension
HOC	Higher-order Crossing
HOS	Higher-order Spectrum
DWT	Discrete Wavelet Transform
mRMR	Maximum Relevance Minimum Redundancy
PCA	Principle Component Analysis
SBS	Sequential Backward Selection
SFS	Sequential Forward Selection
TLR	Tensor Logistic Regression
ERP	Event Related Potenrial
SCP	Slow Cortical Potential
WMN	Weighted Minimum Norm
MN	Minimum Norm
MNI	Montreal Neurological Institute
SPM	Statistical Parameter Modeling
SVD	Singular Value Decomposition
CP	Canonical Polyadic Decomposition
CANDECOMP	Canonical Decomposition
PARAFAC	Parallel Factors
GLM	General Linear Model
IAPS	International Affective Picture System

1-D	One Dimensional
2-D	Two Dimensional
3-D	Three Dimensional
N-D	N Dimensional
IID	Independent and Identically Distributed
sLORETA	Standardized Low-Resolution Brain Electromagnetic Tomography
DTI	Diffusion Tensor Imaging
MRI	Magnetic Resonance Imaging
MSP	Multiple Sparse Priors Algorithm
LORETA	Low-Resolution Brain Electromagnetic Tomography
RBF	Radial Basis Function
WD	Wavelet Decomposition
HHS	Hilbert-Huang Spectrum
FC	Functional Connectivity
DE	Differential Entropy

1. INTRODUCTION

Emotion recognition is a survival attribute for the human brain. It reflects in many areas such as mental health, education, decision making, and robotics [1]. Some of these areas are directly related to human mental health such as early detection of mental disorder [2] and monitoring their progress [3]. Moreover, the total cost of mental disorders was calculated to be around $ 2.5 trillion in 2010 and is expected to be $ 6.1 trillion in 2030 [4]. Using emotion recognition, it is possible to decrease these costs. The importance of emotion recognition has gained attention since the 1990s [5]. The momentum of emotion recognition in publications and citations is shown in Figure 1.1.

Emotion recognition can be performed in three different ways i.e. by using some self-assessment, by physical signals, or by physiological signals. Backward self-assessment as a recognition system is based on a subjective statement and is hard to generalize for all people. For example, it is hard for too old people to recognize their own emotions and also for people who have been using some medication. Another way of classification of emotion is to employ physical signals like human voice, gesture, or facial expression. The disadvantage of this method is that the signals can be manipulated by the person which may give bias to the examiners' current mode. The final method is to use physiological signals to recognize emotions which seems to be the most prominent method so far. The physiological signals from the central, and peripheral nervous system require some devices to collect the data. For this purpose, functional magnetic resonance imaging (fMRI), intracranial electroencephalography (iEEG), electroencephalography (EEG), or positron emission tomography (PET) have been used to register the response of the brain in the temporal and spatial domains.

Source estimation performs the source localization on the cortical brain surface noninvasively. Due to the limited spatial resolution of EEG, several mathematical models have been developed to find a source distribution of the collected data from the human scalp. Source localization methods can be divided into two groups namely, the

equivalent current dipole models and distributed source models. The main differences between these methods are the number of dipoles and their distribution. In the dipole model the fundamental assumption is a limited number of dipoles in a small region. However, in the distributed model dipoles spread to the entire cortical surface which requires a much higher number of dipoles, and the number of dipoles is much higher than that in the equivalent current dipole model.

Machine learning is used to make a classification or prediction in affective computing. Machine learning methods can be divided into two parts which are supervised and unsupervised. Support vector machine (SVM) is one of the most popular methods which has a rate of usage higher than 50% among 26 different classifiers in emotion recognition studies [6]. Also, k-nearest neighbor (KNN), logistic regression, convolutional neural networks (CNN), are widely used.

To improve the performance of machine learning algorithms, feature selection and extraction stages have crucial roles. Until now, numerous methods were implemented for extraction and selection, but we do not have a gold standard. It is possible to classify feature extraction methods in four groups: time-domain features such as fractal dimension (FD), higher-order crossing (HOC); frequency domain features such as band power, higher-order spectra (HOS); time-frequency domain features such as discrete wavelet transform (DWT); spatial domain features such as brain region selection. After many features are extracted from the data, a feature selection method is required to choose efficient features to reduce the dimension of feature space. Many feature selection methods are developed and utilized up to the present. Some of the examples for feature selection methods are maximum relevance minimum redundancy (mRMR), principal component analysis (PCA), relief algorithm, sequential backward selection (SBS), sequential forward selection (SFS).

The main motivation of this thesis is to perform emotion recognition from reconstructed EEG sources with a minimum level of feature extraction and selection. In this study, we use reconstructed source space (3-D) as well as sensor space (2-D) EEG data in different frequency bands which are treated as features. Then, we use tensor logistic

regression (TLR) and support vector machine (SVM) to classify our preprocessed data to identify the emotional differences.

This thesis is organized as follows: Chapter 2 gives information about emotion recognition, source imaging, tensors and machine learning methods. In Chapter 3, data and analysis procedure are explained. The results and discussions are given in Chapter 4 and 5, respectively. In Chapter 6, conclusions and future work are presented.

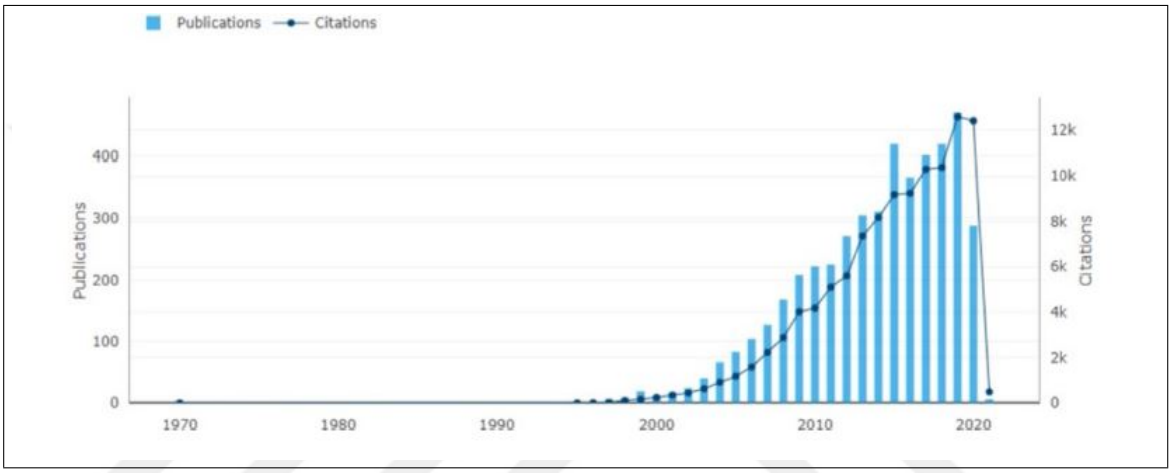

Figure 1.1 Article and citation numbers since 1970 [5].

2. BACKGROUND

2.1 Emotion

There is no clear definition of emotion. However, a widely used definition is that emotion is a cognitive process that assesses stimulus with various criteria such as possible outcomes, the urgency of the event according to the cognitive theory of emotions [7],[8].

2.1.1 Emotion Models

There are two main emotion models; which are discrete and multidimensional emotion models. In the discrete model, there are basic emotions from which the rest of the emotions are produced. However, there is inconsistency about basic emotions. Researchers have disagreement on basic emotions. According to Ekman [9], there are six basic emotions which are anger, disgust, fear, happiness, sadness, and surprise. On the other hand, Plutchik [10] states that there are eight different basic emotions which are anger, fear, sadness, disgust, surprise, curiosity, acceptance, joy.

Furthermore, in the multidimensional emotion model, emotions are described as valence and arousal levels in the circumplex model of affect [11]. The valence domain represents the feelings that are pleasant and unpleasant from good to bad and the arousal domain represents the level of emotion activation and deactivation from high to low. Russell's circumplex model of affect is the most widely used emotion model [12], are given in Figure 2.1.

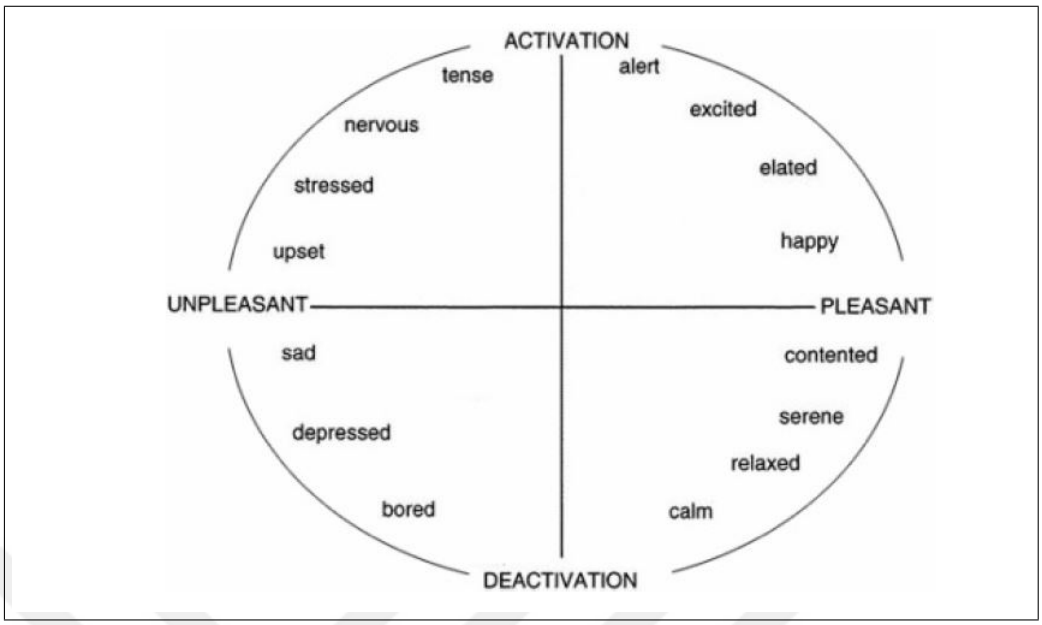

Figure 2.1 Circumplex model of Russell in 2 dimensions [13].

2.1.2 Emotion in the Brain

According to physiological studies, the emotion process shows a correlation with EEG signals. Correlation can be performed in the time domain as well as in frequency domain. Time domain features of event related potentials (ERP) are associated with the emotion process in the time domain. Also, in the frequency domain, the frequency band powers are related to the emotion process in the brain.

It is possible to use time domain features to distinguish emotions. One of the EEG paradigms is ERP. In this protocol, when participants encounter the stimuli, the brain responds to the stimuli. Time domain features of these responses are used in affective science. P/N represent positive/negative peaks and the number after letter represents latency from stimuli onset. Moreover, the features are divided into three groups which are short (P100, N100), middle (P200, N200), and long (P300, slow cortical potentials (SCP)) latencies [14]. Emotion processing is associated with the ERP output of EEG data. P100, N100, P200, and N200 latencies are associated with valence [15]. Moreover, P200, N200, P300, and N300 latencies are associated with the arousal dimension of emotion [16].

According to the literature, delta band (0-4 Hz) is associated with unconscious thought and revealed during deep sleep. Also, the theta band (4-8 Hz) is related to the subconscious mind during light sleep or daydreaming. Alpha band (8-12 Hz) appears in relax mental state. Furthermore, the beta band (12-25 Hz) is related to the intense mental activity such as problem-solving. Finally, the gamma band (>25 Hz) is associated with hyper brain activity such as cognition, or learning process [6],[12]. Also, studies find out the relationship between frequency bands and emotion processing in the brain. Changes in the gamma band allow distinguishing sadness from happiness [17]. Higher frequency bands, which are gamma and beta bands, are more discriminative to valence differences than the low-frequency bands [18]. Also, there is an asymmetry of emotion processing in the brain. For valence discrimination, temporal asymmetry in the alpha band, and prefrontal and parietal asymmetry in the alpha band are found [19]. In the arousal domain, positive affect film clips are related to the left anterior, and negative affect film clips are associated with the right anterior [20].

In addition to time and frequency domains, brain region selection is used in emotion recognition. The major part of the brain contributes to the emotion process. However, according to the literature, the amygdala and frontal lobe have a significant role in emotion recognition in the brain. Amygdala is associated with negative emotions [7]. Also, left frontal activation is associated with positive emotions such as happiness, and right frontal lobe activation is associated with negative emotions such as disgust [21].

2.2 Source Imaging

Source imaging allows to model source distribution on the cortical surface of the brain noninvasively using mathematical methods optimization based on physiological, and anatomical constraints from collected noisy EEG data. An illustration of source imaging using EEG data is given in Figure 2.2. The inverse problem can be performed using equivalent dipole model or distributed source model. The main differences between the two approaches are the number of dipoles and distribution of them on the

brain. In the dipole localization, there are limited number of dipoles located in the small region of the brain. On the other hand, distributed source model assumes that each surface element acts like a source on the entire cortex. The relation between the sensor space EEG data $\mathbf{X}$ and its source space representation $\mathbf{S}$ is given as

$$\boldsymbol{X} = \boldsymbol{LS} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\mathbf{X} \in R^{C \times T}$ is the scalp EEG signal with C being the number of channels and T, the time points. $\mathbf{S} \in R^{N \times T}$ is a source space signal with N being the number of cortical dipole sources. $\mathbf{L} \in R^{C \times N}$ is the lead-field/transfer matrix. Also, $\boldsymbol{\varepsilon} \in R^{C \times T}$ represents the measurement noise on the data.

The inverse problem is an ill-posed problem due to the limited number of sensors and too many current dipoles. Solving this problem requires extra constraints and assumptions in the formula. For solving this problem, the weighted minimum norm (WMN) solution is employed in this study. WMN solution is a member of the minimum norm (MN) solution family which has been introduced as a linear solution for distributed source localization [22],[23]. Eq. 2.1 can be solved by formulating the problem as a least squares optimization of

$$\min_{\boldsymbol{x}} \|\boldsymbol{LS} - \boldsymbol{X}\| + \lambda \|\boldsymbol{WS}\|^2 \tag{2.2}$$

where $\mathbf{W}$ is the covariance matrix of the error $\boldsymbol{e}$ which is assumed to be normally distributed as $\mathcal{N}(\boldsymbol{e}|0, \boldsymbol{C_e})$. λ is the regularization parameter and our optimally chosen λ value is $\lambda = 1.25$. The analytical solution of this problem is given as

$$\hat{\boldsymbol{x}} = \boldsymbol{L}^T(\boldsymbol{L}\boldsymbol{L}^T + \lambda \boldsymbol{W})^{-1}\boldsymbol{X} \tag{2.3}$$

The lead field matrix is obtained from the Montreal Neurological Institute (MNI) atlas brain using tesellation based realistic head modeling and forward modeling algorithms available in Statistical Parameter Modeling (SPM) [24].

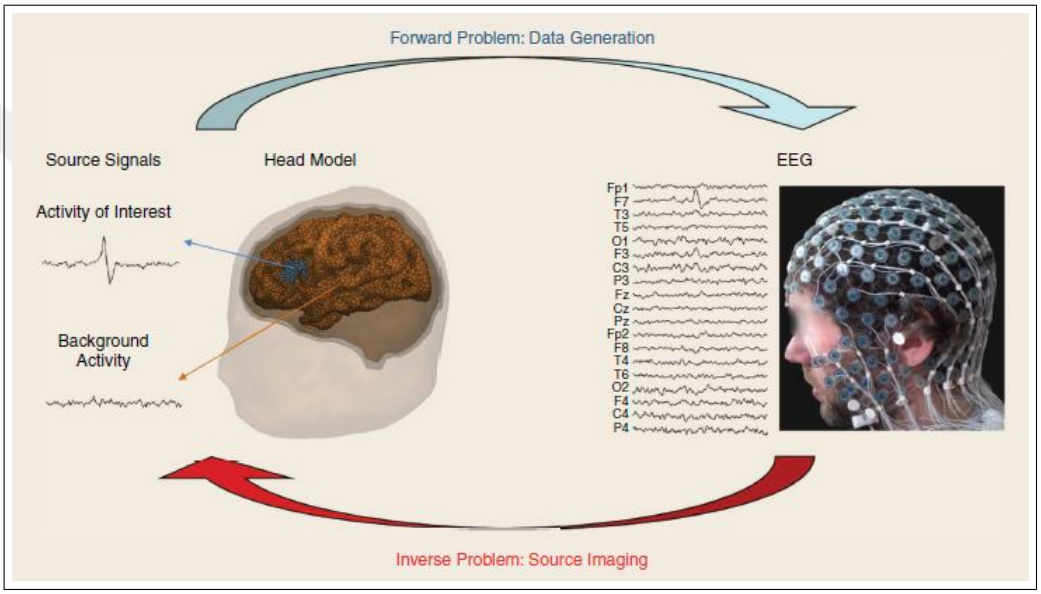

Figure 2.2 Illustration of the forward and inverse problem of EEG [25].

2.3 Tensors

In this section, basic tensor definitions, operations, and some decompositions will be introduced based on tensor notations used in [26].

A time series can be represented as one-dimensional (1-D) data which is a vector denoted as $\boldsymbol{x} \in R^{D_1}$. A two-dimensional (2-D) data which is a matrix denoted as $\boldsymbol{X} \in R^{D_1 \times D_2}$. A three (3-D) or more-dimensional (N-D) data which is tensor denoted by $\boldsymbol{\mathcal{X}} \in R^{D_1 \times D_2 \times \cdots \times D_N}$.

Analogous to the rows and columns in matrices, we have fibers in the 3-dimensional world, called as fiber. A 3-way tensor $\mathcal{X} \in R^{D_1 \times D_2 \times D_3}$ has 3 modes. Mode-1 fiber denotes matrix column $\boldsymbol{x}_{:jk}$, mode-2 fiber represents matrix row $\boldsymbol{x}_{i:k}$ and mode-3 fiber represents tube $\boldsymbol{x}_{ij:}$. Also, it is possible to divide a tensor slice by slice. The third-order tensor $\mathcal{X}$ can be divided as horizontal ($\boldsymbol{X}_{i::}$), lateral ($\boldsymbol{X}_{:j:}$), and frontal ($\boldsymbol{X}_{::k}$) slices. Fibers and slices are visualized in Figure 2.3.

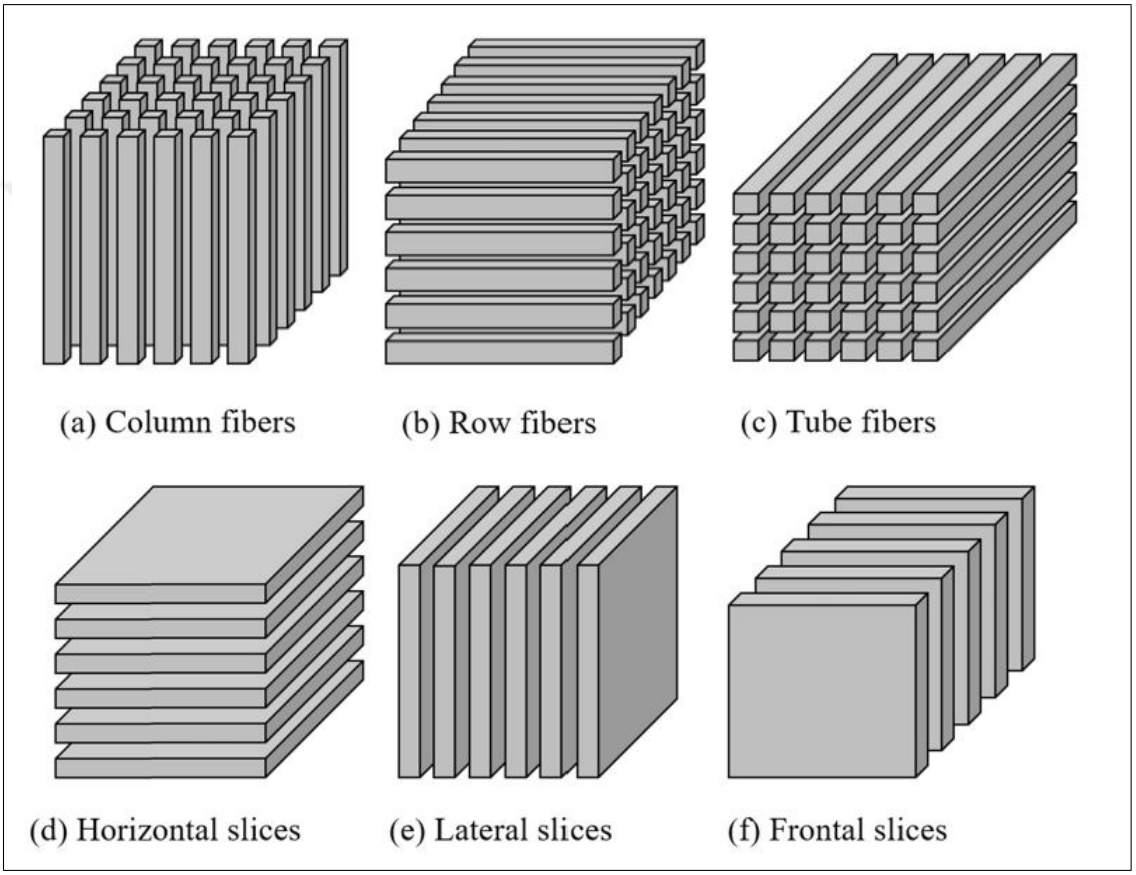

Figure 2.3 Illustration of tensor fiber and slices [27] a) column fibers of tensor $\boldsymbol{x}_{:jk}$ b) row fibers of tensor $\boldsymbol{x}_{i:k}$ c) tube fibers of tensor $\boldsymbol{x}_{ij:}$ d) horizontal slice of tensor $\boldsymbol{X}_{i::}$ e) lateral slices of tensor $\boldsymbol{X}_{:j:}$ f) frontal slices of tensor $\boldsymbol{X}_{::k}$.

2.3.1 Basic Tensor Product Operations

If $\mathcal{X}, \mathcal{Y} \in R^{N_1 \times N_2 \times \cdots \times N_D}$ are the same size tensors, the inner product of $\mathcal{X}$ and $\mathcal{Y}$ is defined as the sum of products of their entries. The formula is defined as follow:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \cdots \sum_{n_D=1}^{N_D} x_{n_1 n_2 n_3 \ldots n_D} y_{n_1 n_2 n_3 \ldots n_D} \tag{2.4}$$

The outer product of N vectors are defined as $(\boldsymbol{b}_1 \circ \boldsymbol{b}_2 \circ \cdots \circ \boldsymbol{b}_N)$ the multiplication of vectors element and is calculated as follow:

$$(\boldsymbol{b}_1 \circ \boldsymbol{b}_2 \circ \cdots \circ \boldsymbol{b}_N)_{i_1 \cdots i_n} = \prod_{n=1}^{N} b_{n i_n} \tag{2.5}$$

Kronecker product is the generalized form of outer product for matrices. $\boldsymbol{A} \in R^{K \times L}$ and $\boldsymbol{B} \in R^{M \times N}$ are matrices and Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$ denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$. The size of the result is $(KM) \times (LN)$ and calculated as follow:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \ldots & a_{1L}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \ldots & a_{2L}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}\boldsymbol{B} & a_{K2}\boldsymbol{B} & \ldots & a_{KL}\boldsymbol{B} \end{bmatrix}$$

Khatri-Rao product is a columnwise Kronecker product. Khatri-Rao product of two matrices, is denoted by $\boldsymbol{A} \odot \boldsymbol{B}$ where $\boldsymbol{A} \in R^{K \times L}$ and $\boldsymbol{B} \in R^{M \times L}$. The overall size of the $\boldsymbol{A} \odot \boldsymbol{B}$ is $(KM) \times (L)$ and calculation of Khatri-Rao product as follows:

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{b_{:1}} & a_{12}\boldsymbol{b_{:2}} & \dots & a_{1L}\boldsymbol{b_{:l}} \\ a_{21}\boldsymbol{b_{:1}} & a_{22}\boldsymbol{b_{:2}} & \dots & a_{2L}\boldsymbol{b_{:l}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}\boldsymbol{b_{:1}} & a_{K2}\boldsymbol{b_{:2}} & \dots & a_{KL}\boldsymbol{b_{:l}} \end{bmatrix}$$

Hadamard product is an elementwise Kronecker product. Hadamard product of two matrices is denoted by $\boldsymbol{A} * \boldsymbol{B}$ where $\boldsymbol{A} \in R^{K \times L}$ and $\boldsymbol{B} \in R^{K \times L}$. The overall size of the $\boldsymbol{A} * \boldsymbol{B}$ is $(K) \times (L)$ and calculation of Hadamard product as follows:

$$\boldsymbol{A} * \boldsymbol{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1L}b_{1L} \\ a_{21}b_{11} & a_{22}b_{22} & \dots & a_{2L}b_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}b_{K1} & a_{K2}b_{K2} & \dots & a_{KL}b_{KL} \end{bmatrix}$$

2.3.2 Mode-n Unfolding

Tensor unfolding is also known as matricization/flattening, is the production of the matrix from an N-way tensor. Mode-n unfolding is a special case of this process. $\boldsymbol{\mathcal{A}} \in R^{I_1 \times I_2 \times \cdots \times I_D}$ is denoted as $\boldsymbol{A}_{(n)}$ shown in Figure 2.4. From tensor elements $(a_{i_1}, a_{i_2}, \cdots, a_{i_D})$ to the production of matrix elements (m_{i_d}, j), mathematical representation is as follows,

$$j = 1 + \sum_{\substack{k=1, \\ k \neq n}}^{D} \left((i_k - 1) \prod_{\substack{m=1, \\ m \neq n}}^{k-1} I_m \right) \tag{2.6}$$

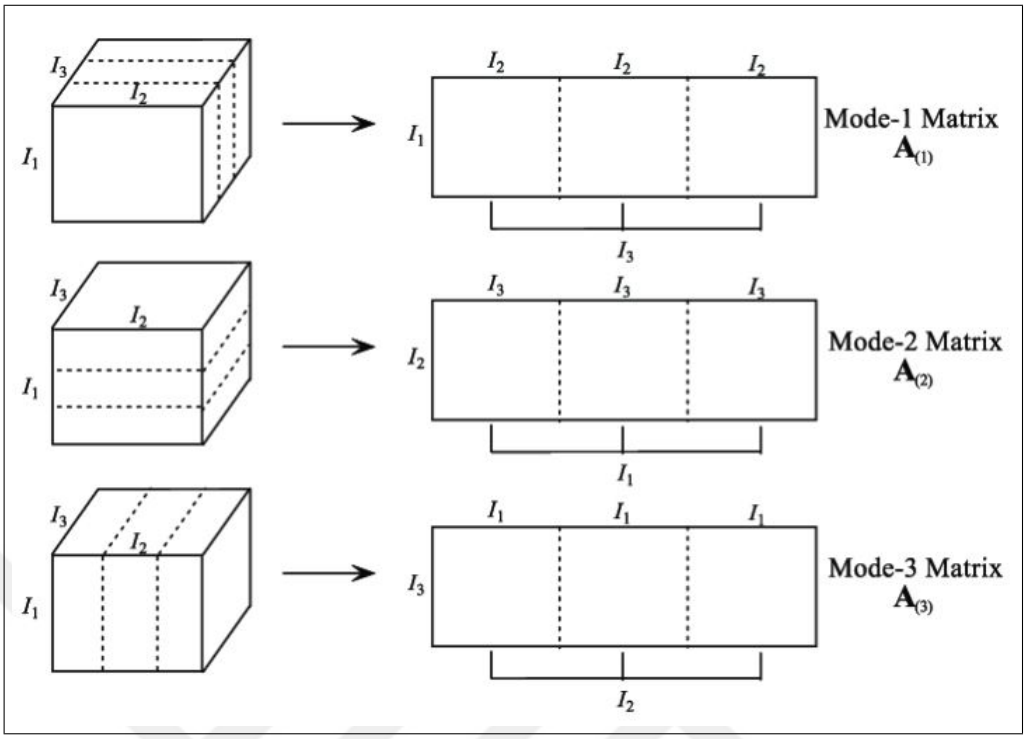

Figure 2.4 Example of 3-way tensor matricization [28].

2.3.3 Tensor Decomposition

Tensor decomposition is used to represent a tensor in a more compact way to process data with lower computational complexity. There are several decomposition methods but the most popular two are Canonical Polyadic Decomposition (CP) [29],[30], otherwise known as canonical decomposition (CANDECOM) and parallel factor analysis (PARAFAC), and Tucker Decomposition [31]. In this study, only CP [32] is used. CP is a generalized form of the singular value decomposition (SVD). CP is implemented many areas as data mining [33], neuroscience [34], image compression [35]. In this method, the tensor is represented as a sum of the low-rank tensors. In Figure 2.5 is shown a 3-way tensor decomposition with rank-1. The mathematical formulation of $\boldsymbol{\mathcal{B}} \in R^{N_1 \times N_2 \times \cdots \times N_D}$ decomposition is as follow,

$$\boldsymbol{\mathcal{B}} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)} \tag{2.7}$$

where $\boldsymbol{\beta}_K^{(r)} \in R^{N_d}$, $d = 1, 2, \cdots, D$ and $r = 1, 2, \cdots, R$ are vectors.

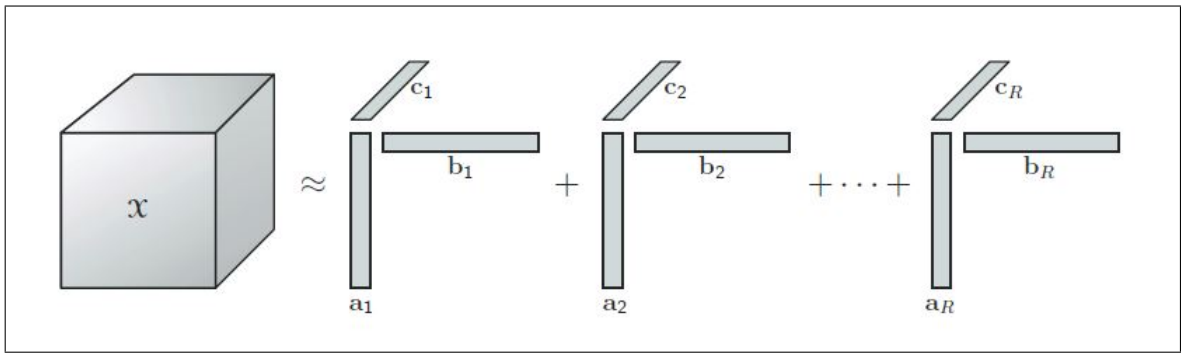

Figure 2.5 CP decomposition of a three-way array [26].

2.4 Machine Learning

In this chapter, logistic tensor regression and support vector machine are introduced which are the two algorithms used in this thesis.

2.4.1 Tensor Logistic Regression

Logistic tensor regression (LTR) [36] is a generalized form general linear model (GLM). The main difference between GLM and LTR is that GLM is using vectors as covariates but LTR is using tensors as covariates. This reduces the computational complexity. It is so important for big data processing. In a typical MRI, data size is $256 \times 256 \times 256$. Moreover, 256^3 parameters are required to perform a classical GLM. However, in rank-3 LTR, 3×256 parameters are enough to model the data. The formulation of LTR is as follows,

The method is a multilinear generalization of the basic GLM model as given in Eq. 2.8. In the formula $\boldsymbol{x} \in R^p$ is the model input related to mean of with the response Y via $\mu = E(Y/\boldsymbol{x})$ where $g(.)$ is a link function given as

$$g(\boldsymbol{\mu}) = \alpha + \beta^T \boldsymbol{x} \tag{2.8}$$

Y is a member of exponential family with probability mass density

$$p(Y \mid \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \tag{2.9}$$

which can be generalized for Eq. 2.8 for higher-order data as

$$g(\boldsymbol{\mu}) = \boldsymbol{\alpha} + \boldsymbol{\gamma}^T \boldsymbol{z} + \langle \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{X}} \rangle \tag{2.10}$$

Also,

$$g(\boldsymbol{\mu}) = \boldsymbol{\alpha} + \boldsymbol{\gamma}^T \boldsymbol{z} + \langle \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \boldsymbol{\beta}_2^{(r)} \circ \cdots \circ \boldsymbol{\beta}_K^{(r)}, \boldsymbol{\mathcal{X}} \rangle \tag{2.11}$$

$$g(\boldsymbol{\mu}) = \boldsymbol{\alpha} + \boldsymbol{\gamma}^T \boldsymbol{z} + \langle (\boldsymbol{B}_K \otimes \cdots \otimes \boldsymbol{B}_1) 1_n, vec(\boldsymbol{\mathcal{X}}) \rangle \tag{2.12}$$

where $\boldsymbol{\mathcal{X}} \in R^{p_1 \times \cdots \times p_K}$, $\mathbf{z}$ represents the vector-valued covariates with weight, $\boldsymbol{\gamma} \in R^{p_0}$, $\mathbf{1}_n$ is an array of N ones, and $\boldsymbol{B}_K = [\boldsymbol{\beta}_K^{(1)}, \boldsymbol{\beta}_K^{(2)}, \boldsymbol{\beta}_K^{(3)}, \cdots, \boldsymbol{\beta}_K^{(r)}] \in R^{p_K \times r}$.

Given n independent and identically distibuted (iid) data $((yi, xi, zi), i = 1, \cdots, n)$, to estimate the $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and multilinear parameters $\boldsymbol{B}_k$ in Eq. 2.12 maximum likelihood optimization of

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{B}_1, \cdots, \boldsymbol{B}_K) = \sum_{i=1}^{n} \frac{y\theta - b(\theta)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \tag{2.13}$$

is performed where $\theta = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{B_1}, \cdots, \boldsymbol{B_K})$. To achieve the maximum value of $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{B}_1, \cdots, \boldsymbol{B}_K)$ a block relaxation algorithm [36],[37] is employed. The algorithm is given in the Appendix A.

2.4.2 Support Vector Machine

SVM is a popular supervised classification method for emotion recognition. It is possible to implement SVM on linear and nonlinear datasets. The method classifies the data using a hyperplane. The algorithm defines many hyperplanes and decides the best hyperplane (see Figure 2.6) using the support vectors, which are the closest data points to the hyperplane. Also, the distance between support vectors are called margins. The method tries to maximize the margins which affect the accuracy rate of the algorithm.

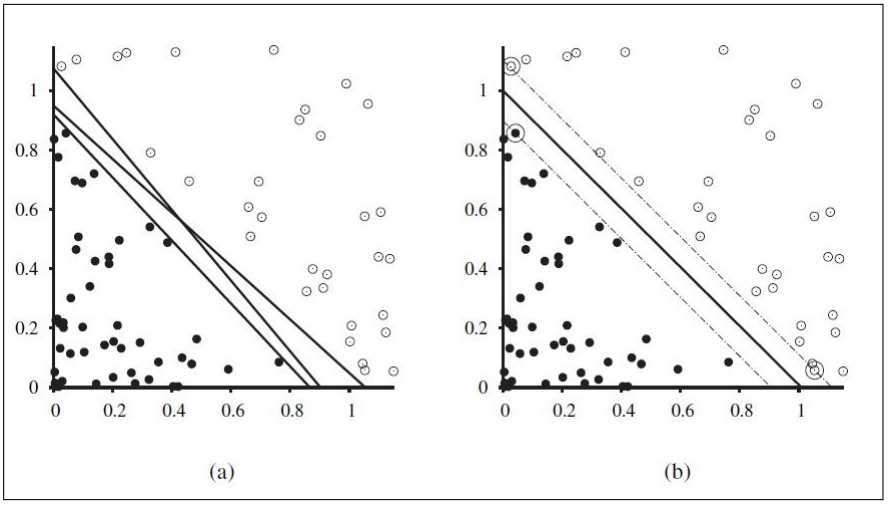

Figure 2.6 Classification of two different groups which are filled and empty circles using support vector machine a) There are candidates of the hyperplane b) Calculated best hyperplane and support vectors which are surrounded with circles [38].

3. MATERIAL AND METHODS

3.1 Data

The dataset was provided by the medical school of Istanbul University. This dataset consisted of 13 health subjects whose mean age was 27.4 ($\pm$ 2.96). The EEG data were collected using the event-related potential method with a sampling frequency of 250 Hz. The EEG data occurred from 30 electrodes which were Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz and O2. The dataset consisted of two groups of stimuli which were the pleasant group and the unpleasant group. Subjects encountered each group of stimuli separately. The experimental protocol was as follows; each picture was presented for 1 second on the computer screen, followed by 2 seconds of inter-stimulus interval. Pictures were chosen from the International Affective Picture System (IAPS) datasets. Mean valence level of groups were discriminative which were 7.13 and 2.96. Also, the mean arousal levels were similar which were 4.99 and 5.02.

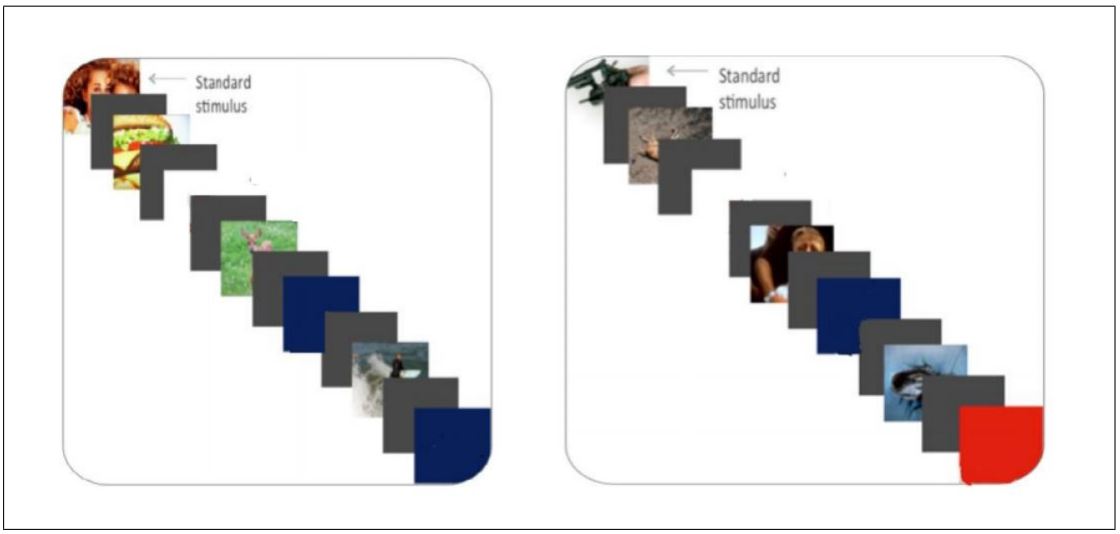

Figure 3.1 Example of stimulus representation [39].

3.2 Data Processing

In this study, we analyzed each subject individually. Here, we used two different approaches for data representation; which are cortical source reconstructed EEG data and topographic scalp EEG maps.

The first part of the analysis was the estimation of source reconstructed EEG data and its classification. The pipeline of the work is given in Figure 3.2. Raw data consisted of pleasant and unpleasant groups which were represented as $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in R^{250\times30\times275}$. The dimensions of data were $time \times electrode \times trials$. Next, we extracted frequency bands which were delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), gamma1 (25-40Hz), and gamma2 (40-55 Hz) bands using an EEGLAB filter function named eegfilt [40]. Then, the inverse problem was solved using a weighted minimum norm solution and sources were reconstructed on tesselated cortical surface of 5124 vertices (Figure 3.3). The vertices mapped into an empty MRI data (MNI averaged brain [41]) from SPM with $91 \times 109 \times 91$ dimensions. This mapping process generated a large data size which would increase the computational complexity. Spatial downsizing by a factor of 0.1 was performed using Matlab function (imresize3) and the data were converted to tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in R^{10\times11\times10\times250\times275}$. Subsequently, the data were divided into windows in the time domain, each consisting of 25-time points. Moreover, mean squared signals in each window were calculated and concatenated in the time domain; yielding tensors with dimensions $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in R^{10\times11\times10\times10\times275}$. This process was implemented for both pleasant and unpleasant datasets. Then, from the pleasant and unpleasant groups, training and test data were produced. For this purpose, 200 trials for training dataset and 50 trials for test dataset were chosen randomly from two groups. Consequently, $\boldsymbol{\mathcal{X}_{train}} \in R^{10\times11\times10\times10\times400}$ and $\boldsymbol{\mathcal{X}_{test}} \in R^{10\times11\times10\times10\times100}$ data were produced with a response variable $Y \in \{1, 0\}^{100\times1}$. The classification algorithms were implemented using SVM with linear kernel and TLR. Finally, repeated random subsampling validation was used as a validation method. This experiment was repeated 100 times and results were then averaged.

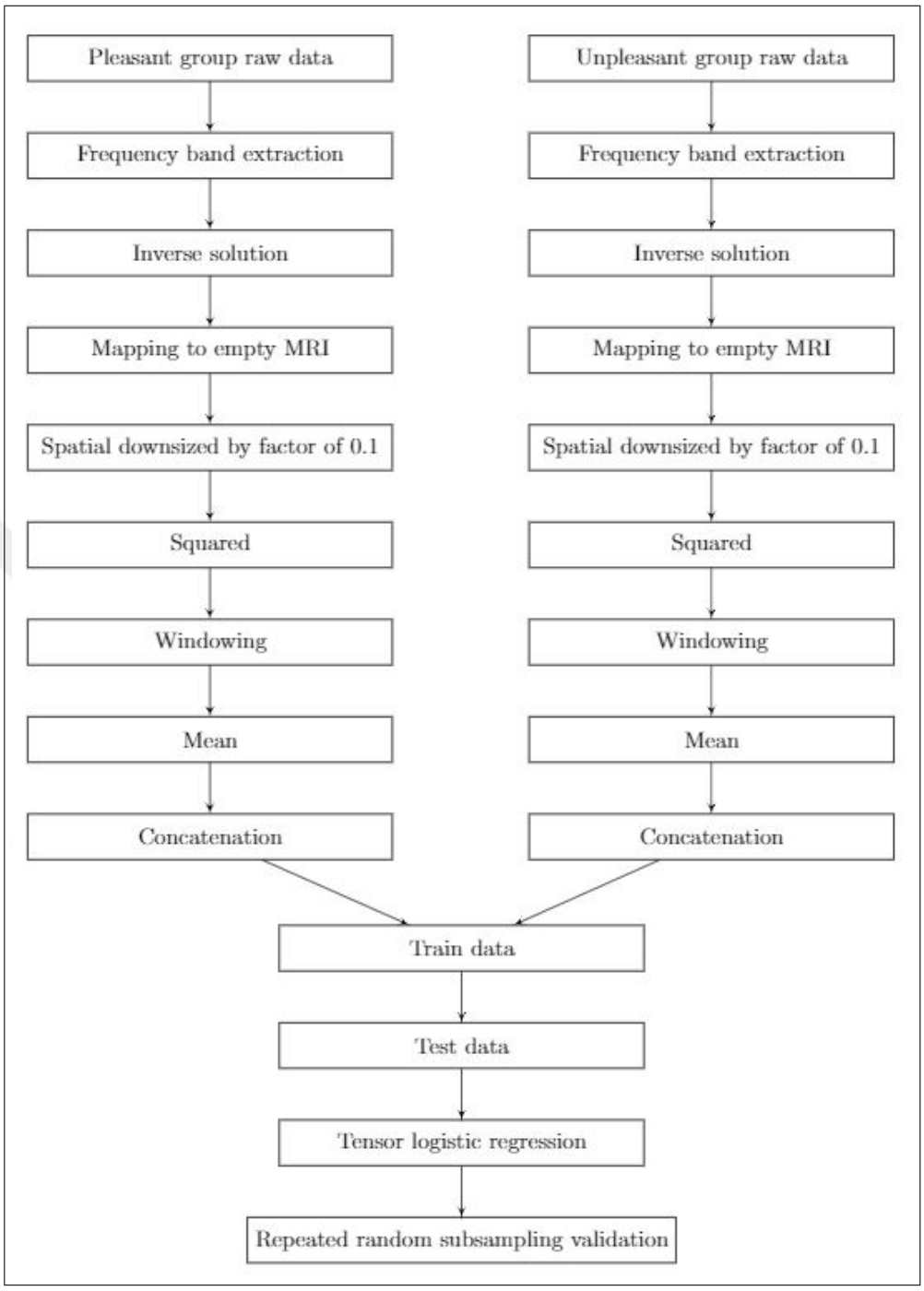

Figure 3.2 Pipeline of analysis of reconstructed sources.

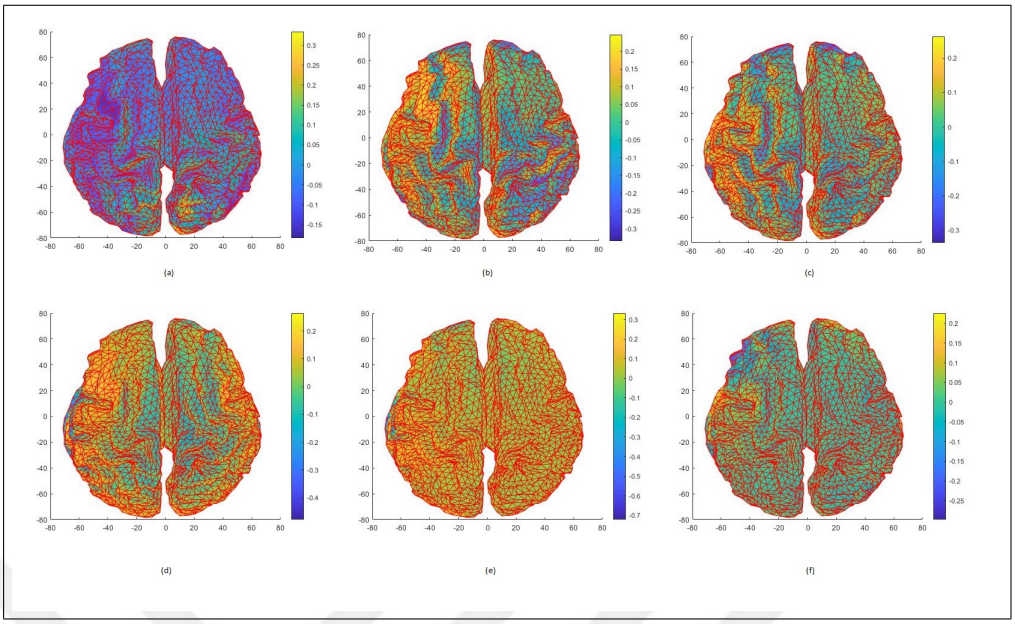

Figure 3.3 Example of reconstructed sources in the a) delta, b) theta, c) alpha, d) beta, e) gamma1, and f) gamma2 bands, respectively.

In the second part of the analysis, the same data were used. The pipeline of the work is given in Figure 3.4. First, we extracted delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), gamma1 (25-40 Hz), and gamma2 (40-55 Hz) bands using the EEGLAB filter function named eegfilt. Then, topographic maps (Figure 3.5) were generated using the topoplot function of EEGLAB [40]. Moreover, data were converted to tensors with dimensions $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in R^{32\times32\times250\times275}$ and then divided into windows in the time domain, each consisting of 25-time points. Moreover, mean squared signals in each window were calculated and concatenated in the time domain; yielding tensors with dimensions $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in R^{32\times32\times10\times275}$. This process was implemented for both pleasant and unpleasant datasets. Then, from the pleasant and unpleasant groups, training and test data were selected. For this purpose, 200 trials for training dataset, and 50 trials for test dataset were chosen randomly from two groups. Consequently, $\boldsymbol{\mathcal{X}}_{\boldsymbol{train}} \in R^{32\times32\times10\times400}$ and $\boldsymbol{\mathcal{X}}_{\boldsymbol{test}} \in R^{32\times32\times10\times100}$ data were produced with a response variable $Y \in \{1, 0\}^{100\times1}$. The classification algorithms were implemented using SVM with linear kernel and TLR. Finally, repeated random subsampling validation was used as a validation method. This experiment was repeated 100 times and results were then averaged.

Analysis was processed in Matlab R2019b. SVM was implemented using libsvm toolbox version 3.24 [42]. TLR was processed using tensorreg toolbox version 1.0.3 [36].

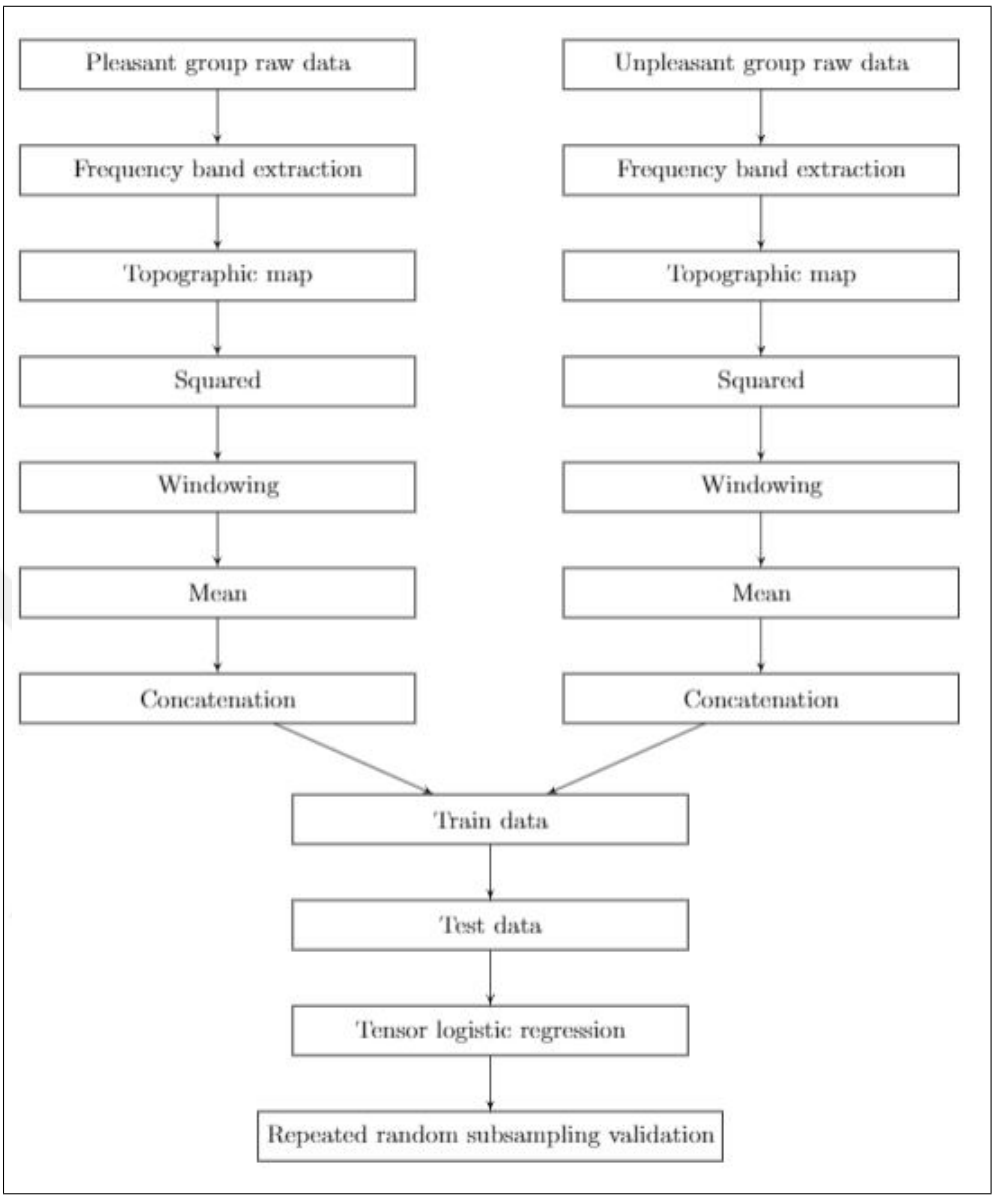

Figure 3.4 Pipeline of analysis of topographic maps.

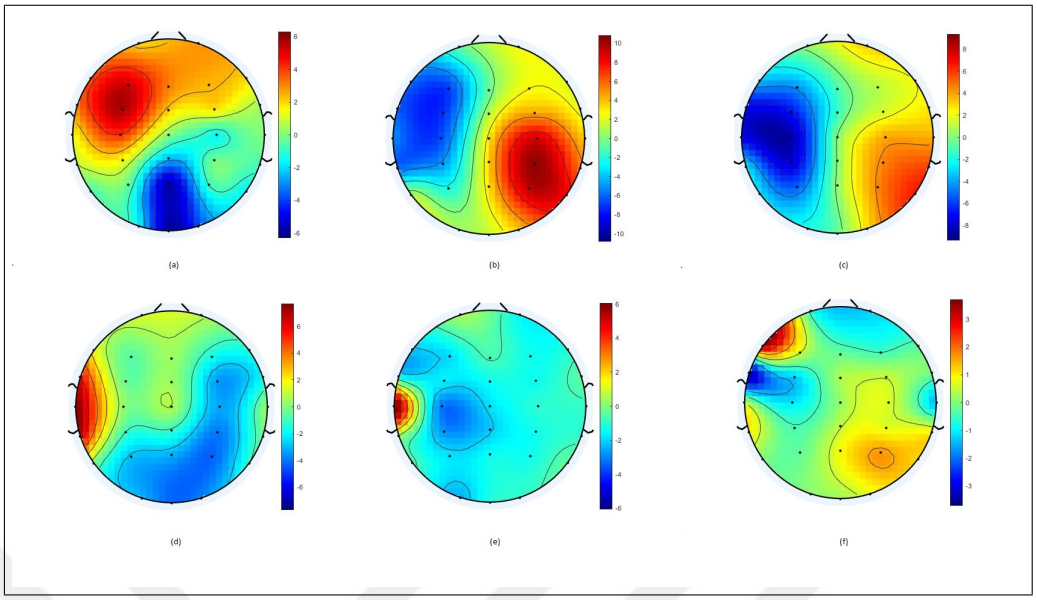

Figure 3.5 Example of topographic maps in the a) delta, b) theta, c) alpha, d) beta, e) gamma1, and f) gamma2 bands, respectively.

4. RESULTS

In this research, the maximum accuracy rate was obtained as 91.96% in the gamma1 band in the source space using TLR (see Figure 4.1). The accuracy rates of gamma2 and beta bands were 91.42% and 89.96% which were quite similar. Rest of the accuracy rates were ordered as the alpha band (80.07%), theta band (79.52%) and delta band (69.66%). Moreover, the highest accuracies of the sensor space data using TLR were 86.89%, 86.59%, and 84.36% at the gamma1, gamma2, and beta bands, respectively. Lower accuracy bands were alpha band (73.34%), theta band (73.06%), and delta band (66.83%). On the other hand, we achieved the highest accuracy rate of 92.68% in the gamma2 band in with sensor space data using SVM. The minimum accuracy rate was 66.53% in the delta band. In other bands accuracy rates were as follows: theta band (71.40%), alpha band (71.68%), beta band (83.25%), and gamma1 band (89.33%). The source space data accuracy rates were between 50-57% for all frequency bands using SVM.

When the data were converted from sensor space to source space the accuracy rate increased from 73.34% to 80.07% at the alpha band which was the maximum increment (6.73%). Averaged improvement for all bands was 5.25%. Also, there were improvements in delta (2.83%), theta (6.46%), beta (5.60%), gamma1 (5.07%), gamma2 (4.83%) bands using TLR. Contrarily, we could not obtain a similar enhancement in performance in SVM.

When we focused on the individual results, subject-5 had maximum accuracy (99.70%) in the gamma1 band in the sensor space using SVM. Furthermore, subject-13 achieved the best performance in the source space using TLR which was 98.19%. The minimum accuracy rate was observed in subject-4 in almost all frequency bands. For example, the accuracy rates of subject-4 was 74.59%, 71.34% and 58.33% in the source space in the gamma2, gamma1 and beta bands, respectively using TLR. Similar accuracies were also obtained in source space. The details of the classification results

are given in Tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 with respect to different frequency bands.

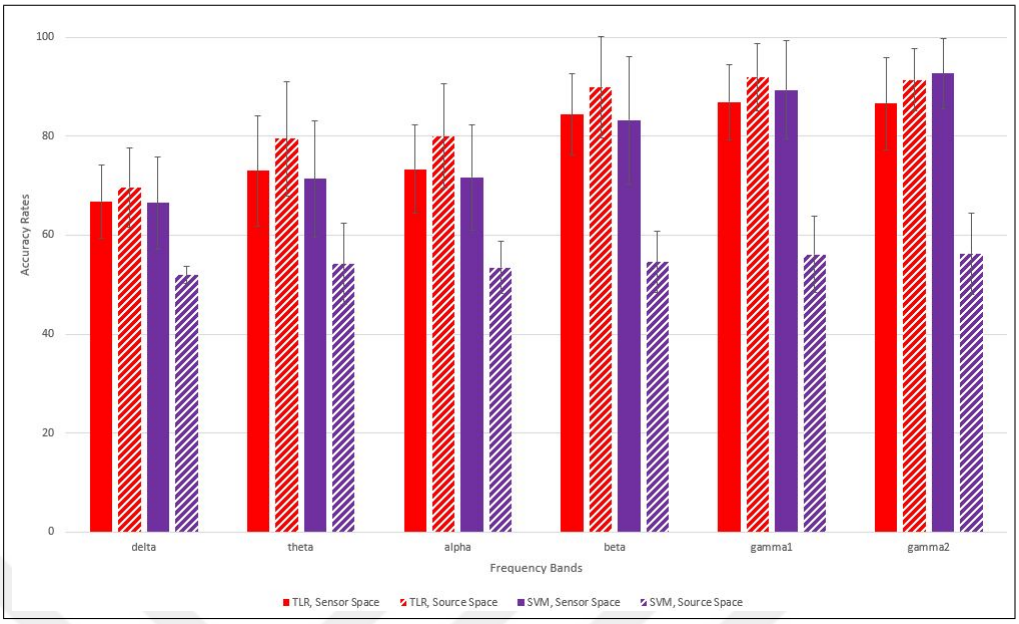

Figure 4.1 Accuracy results and standard deviations according to frequency bands.

Table 4.1
Accuracy in delta band (0-4 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	70.58%	67.40%	73.83%	51.30%
Person2	56.69%	59.99%	55.47%	52.54%
Person3	61.47%	66.36%	60.20%	50.05%
Person4	56.33%	55.75%	52.13%	51.41%
Person5	77.87%	82.12%	80.95%	54.09%
Person6	66.81%	70.24%	66.54%	51.28%
Person7	78.06%	80.45%	81.25%	51.50%
Person8	59.52%	65.28%	61.58%	56.26%
Person9	61.41%	63.21%	60.01%	50.99%
Person10	70.93%	75.62%	69.40%	50.40%
Person11	65.79%	66.50%	59.84%	50.30%
Person12	73.13%	78.66%	68.58%	53.03%
Person13	70.32%	74.08%	75.22%	52.18%
Mean	66.83%	69.66%	66.53%	51.94%

Table 4.2
Accuracy in theta band (4-8 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	77.38%	79.55%	78.07%	52.09%
Person2	55.99%	65.55%	55.38%	50.99%
Person3	61.12%	74.50%	61.31%	50.28%
Person4	54.08%	53.97%	52.20%	50.76%
Person5	90.00%	95.08%	90.96%	80.05%
Person6	69.09%	77.44%	64.56%	59.29%
Person7	85.24%	90.64%	86.03%	52.36%
Person8	67.45%	69.19%	63.00%	54.34%
Person9	71.46%	79.59%	69.78%	51.79%
Person10	80.53%	88.76%	78.59%	50.05%
Person11	74.94%	83.59%	71.84%	50.30%
Person12	82.91%	90.54%	80.78%	50.95%
Person13	79.69%	85.45%	75.71%	51.17%
Mean	73.06%	79.52%	71.40%	54.18%

Table 4.3
Accuracy in alpha band (8-12 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	74.82%	75.70%	78.87%	51.21%
Person2	66.40%	71.65%	61.26%	50.81%
Person3	62.31%	80.95%	64.87%	50.21%
Person4	54.62%	52.83%	51.53%	50.84%
Person5	85.45%	94.06%	92.54%	68.74%
Person6	72.74%	82.67%	68.35%	54.67%
Person7	82.42%	89.25%	77.26%	53.43%
Person8	66.00%	74.84%	60.47%	50.47%
Person9	74.81%	85.59%	74.05%	51.00%
Person10	81.94%	87.65%	82.82%	50.11%
Person11	73.22%	74.73%	70.50%	50.91%
Person12	79.75%	86.20%	74.87%	59.56%
Person13	79.03%	84.91%	74.54%	52.62%
Mean	73.34%	80.07%	71.68%	53.42%

Table 4.4
Accuracy in beta band (12-25 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	91.73%	93.34%	93.65%	72.65%
Person2	70.45%	85.26%	62.88%	54.00%
Person3	85.20%	94.02%	90.84%	50.29%
Person4	66.79%	58.33%	56.52%	50.97%
Person5	87.74%	96.93%	98.15%	55.55%
Person6	77.19%	87.09%	73.02%	58.86%
Person7	94.16%	96.79%	94.13%	57.76%
Person8	86.08%	88.70%	82.49%	55.21%
Person9	87.79%	94.56%	85.18%	50.01%
Person10	84.09%	90.98%	95.13%	50.10%
Person11	84.38%	91.06%	75.84%	52.40%
Person12	92.20%	97.05%	86.98%	51.06%
Person13	88.95%	95.45%	87.53%	51.27%
Mean	84.36%	89.96%	83.25%	54.62%

Table 4.5
Accuracy in gamma1 band (25-40 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	90.87%	92.13%	97.43%	71.31%
Person2	79.30%	89.42%	74.61%	54.16%
Person3	84.14%	93.76%	94.73%	49.93%
Person4	67.74%	71.34%	65.65%	51.24%
Person5	81.77%	96.93%	99.70%	72.76%
Person6	84.24%	89.95%	88.33%	57.37%
Person7	95.43%	96.08%	98.11%	57.99%
Person8	91.83%	94.84%	90.73%	59.08%
Person9	87.70%	92.87%	87.00%	50.15%
Person10	87.25%	89.19%	96.17%	49.91%
Person11	91.04%	94.47%	83.26%	54.49%
Person12	95.97%	97.24%	95.56%	50.54%
Person13	92.40%	97.36%	90.09%	50.56%
Mean	86.89%	91.96%	89.33%	56.11%

Table 4.6
Accuracy in gamma2 band (40-55 Hz).

Subjects	TLR		SVM	
	Sensor Space	Source Space	Sensor Space	Source Space
Person1	93.23%	94.28%	97.71%	76.10%
Person2	69.96%	86.70%	79.86%	55.97%
Person3	85.47%	93.92%	95.31%	49.98%
Person4	64.89%	74.59%	76.45%	51.08%
Person5	82.26%	90.68%	99.54%	69.78%
Person6	87.06%	88.88%	92.42%	55.33%
Person7	92.92%	96.08%	98.34%	56.66%
Person8	94.40%	96.70%	95.92%	60.80%
Person9	91.60%	92.16%	91.83%	50.18%
Person10	85.87%	86.62%	96.73%	50.24%
Person11	90.36%	92.56%	88.41%	54.83%
Person12	94.80%	97.11%	96.64%	50.60%
Person13	92.97%	98.19%	95.68%	50.17%
Mean	86.59%	91.42%	92.68%	56.28%

5. DISCUSSION

The main goal of this thesis was emotion classification using both sensor level and source level data using multilinear tensor regression. We used two different data structures, which were the topographic scalp EEG maps and cortical reconstructions of EEG sources. We also compared the tensor based classification method with the SVM.

We achieved an accuracy of 91.96% in the gamma1 band in the source space using TLR. In a similar study [43] it was found that using source space data yielded better results than using the sensor space data. With WMN as a source reconstruction and SVM as a classification algorithm, they reached to a 75% accuracy rate in the source space data using connectivity features from the emotion dataset which was collected by the same group. Another study, [44] had a similar approach, but the inverse problem was solved using standardized low-resolution brain electromagnetic tomography (sLORETA). Brain region selection and some feature selection methods were implemented. The highest accuracy was 89.64% using functional connectivity features in the gamma band from the DEAP dataset. [45] used multiple sparse priors algorithm (MSP), which was developed by the same group, and low-resolution brain electromagnetic tomography (LORETA). Binary classification was used in four different group classifiers which were arousal, valence, liking, dominance using the DEAP dataset. The classification was done by SVM. The maximum accuracy was achieved 58.6% in the study.

In the sensor space, we reached to an accuracy of 86.89% and 92.68% using TLR and SVM, respectively. Also, in a similar study with the same dataset [39], the performance of TLR, and SVM with radial basis function (RBF) were 89.93%, 75.59%, respectively at the beta band. The data was produced with wavelet decomposition (WD). Also, in [43],[44], which were mentioned in previous paragraph, sensor space accuracy rates were reported. Their sensor space accuracies were 70% and 88.49%,

respectively. In a study conducted in 2009 [17], one of the highest accuracy rates was 93.5% using common spatial patterns (SCP) as features. However, there was no fixed frequency band, it required a frequency band search for the best classification rate. [46] achieved 82% accuracy using higher-order spectra using IAPS with the genetic algorithm as a feature selection method. Moreover, SVM was used as a classification algorithm with RBF. In [47], Hilbert-Huang Spectrum (HHS) was employed for a feature extraction method and four different classification methods were used in classification namely, SVM, KNN, quadratic distance, and Mahalanobis distance. The best performance of this analysis was 86.2% using KNN at the beta-gamma band combination. The results showed that TLR could be a reliable alternative to SVM in sensor space data analysis.

When the data were converted from sensor space to source space, the performance of classification was increased on the average by 5.25% with a maximum of 6.73% using TLR. In [44], 3.12% improvement was achieved with two features. Moreover, [43] found 5% improvement in the accuracy. They mentioned that sample size of the EEG dataset and the number of channels were also very critical to obtain better results.

Frequency band results were shown in Figure 4.1, there were similar patterns for all conditions except the source space data using SVM. When data were transferred from sensor space to source space, the classification performance was declined by 22.73% on the average, with a maximum of 36.4%. SVM did not work properly due to the complexity of the source space data of the entire brain. [43],[44] showed that it was possible to increase the performance of SVM using additional implementation such as brain region selection.

The gamma band had the best performance in our results. There were two main reasons for the high accuracy rate of emotion recognition in the gamma band. First, higher frequency bands have emerged during high level processing such as emotion recognition. For example, valence recognition was associated with the gamma band [17]. The second reason was that, the lateralization of cortical activity has happened

during the emotion recognition process in the gamma band. Asymmetric cortical activity has been shown for emotion recognition by neuropsychological studies [48],[49]. [50] showed that right frontal activation was increased with positive emotions, and left frontal lobe activation was increased with negative emotions in the gamma band. Also, they reported that gamma band power was improved on the right frontal electrodes during emotion recognition. Moreover, valence recognition was associated with temporal asymmetry in the gamma band [19].

6. CONCLUSION AND FUTURE WORK

In this study, we showed that emotion recognition was possible using specific frequency band among which the gamma2 band yielded the highest accuracy rates. Our results are promising in emotion recognition studies using EEG data. Also, we compared SVM and TLR as classifier algorithms and they performed similarly in sensor space, but TLR was better than SVM in source space. Finally, converting data from sensor space to source space increased the accuracy rate using TLR.

In this study, we used a WMN estimation for the inverse problem. Other methods which employ more refined optimization algorithms may also be tested such to solve the inverse problems. Their comparison may lead to choosing more appropriate source localization methods. In a further study, a multi-class classification task may be conducted using valence, arousal, liking, and dominance differences. Finally, TLR can be used in other high-dimensional datasets such as fMRI and DTI or their multimodal combinations with EEG data.

APPENDIX A. BLOCK RELAXATION ALGORITHM

Algorithm 1 Block Relaxation Algorithm [36]

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

repeat

  for $k = 1, 2, \cdots, K$ **do**

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

  end for

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

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

REFERENCES

1. Soroush, M. Z., K. Maghooli, S. K. Setarehdan, and A. M. Nasrabadi, "Emotion classification through nonlinear EEG analysis using machine learning methods," *International Clinical Neuroscience Journal*, Vol. 5, pp. 135–149, 2018.

2. Turetsky, B. I., C. G. Kohler, T. Indersmitten, M. T. Bhati, D. Charbonnier, and R. C. Gur, "Facial emotion recognition in schizophrenia: when and why does it go awry?," *Schizophrenia Research*, Vol. 94, no. 1-3, pp. 253–263, 2007.

3. Gu, Y., K. J. Wong, and S. L. Tan, "Analysis of physiological responses from multiple subjects for emotion recognition," in *2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 178–183, IEEE, 2012.

4. Trautmann, S., J. Rehm, and H. U. Wittchen, "The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders?," *EMBO Reports*, Vol. 17, no. 9, pp. 1245–1249, 2016.

5. Arya, R., J. Singh, and A. Kumar, "A survey of multidisciplinary domains contributing to affective computing," *Computer Science Review*, Vol. 40, p. 100399, 2021.

6. Alarcao, S. M., and M. J. Fonseca, "Emotions recognition using EEG signals: a survey," *IEEE Transactions on Affective Computing*, Vol. 10, no. 3, pp. 374–393, 2017.

7. Chanel, G., J. J. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *International Journal of Human-Computer Studies*, Vol. 67, no. 8, pp. 607–627, 2009.

8. Sander, D., D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks*, Vol. 18, no. 4, pp. 317–352, 2005.

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

10. Plutchik, R., "The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, Vol. 89, no. 4, pp. 344–350, 2001.

11. Russell, J. A., "A circumplex model of affect.," *Journal of Personality and Social Psychology*, Vol. 39, no. 6, p. 1161, 1980.

12. García-Martínez, B., A. Martinez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, "A review on nonlinear methods using electroencephalographic recordings for emotion recognition," *IEEE Transactions on Affective Computing*, 2019.

13. Posner, J., J. A. Russell, and B. S. Peterson, "The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, Vol. 17, no. 3, pp. 715–734, 2005.

14. Kim, M. K., M. Kim, E. Oh, and S. P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Computational and Mathematical Methods in Medicine*, Vol. 2013, 2013.

15. Olofsson, J. K., and J. Polich, "Affective visual event-related potentials: arousal, repetition, and time-on-task," *Biological Psychology*, Vol. 75, no. 1, pp. 101–108, 2007.

16. Cuthbert, B. N., H. T. Schupp, M. M. Bradley, N. Birbaumer, and P. J. Lang, "Brain potentials in affective picture processing: covariation with autonomic arousal and affective report," *Biological Psychology*, Vol. 52, no. 2, pp. 95–111, 2000.

17. Li, M., and B. L. Lu, "Emotion classification based on gamma-band EEG," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1223–1226, IEEE, 2009.

18. 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, IEEE, 2013.

19. Huang, D., C. Guan, K. K. Ang, H. Zhang, and Y. Pan, "Asymmetric spatial pattern for EEG-based emotion detection," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, IEEE, 2012.

20. Davidson, R. J., "Anterior cerebral asymmetry and the nature of emotion," *Brain and Cognition*, Vol. 20, no. 1, pp. 125–151, 1992.

21. Balconi, M., and G. Mazza, "Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues.: ERS/ERD and coherence measures of alpha band," *International Journal of Psychophysiology*, Vol. 74, no. 2, pp. 158–165, 2009.

22. Hämäläinen, M. S., and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Medical & Biological Engineering & Computing*, Vol. 32, no. 1, pp. 35–42, 1994.

23. Michel, C. M., and B. He, *EEG source localization*, Vol. 160 of *Handbook of Clinical Neurology*, pp. 85–101. Elsevier, 2019.

24. Wellcome Trust Centre for Neuroimaging, "Statistical parametric mapping software." Available at `www.fil.ion.ucl.ac.uk/spm/`.

25. Becker, H., L. Albera, P. Comon, R. Gribonval, F. Wendling, and I. Merlet, "Brain-source imaging: from sparse to tensor models," *IEEE Signal Processing Magazine*, Vol. 32, no. 6, pp. 100–112, 2015.

26. Kolda, T. G., and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, Vol. 51, no. 3, pp. 455–500, 2009.

27. You, R., Y. Yao, and J. Shi, "Tensor-based ultrasonic data analysis for defect detection in fiber reinforced polymer (FRP) composites," *Chemometrics and Intelligent Laboratory Systems*, Vol. 163, pp. 24–30, 2017.

28. Qiao, X., R. Xu, Y. W. Chen, T. Igarashi, K. Nakao, and A. Kashimoto, "Generalized n-dimensional principal component analysis (GND-PCA) based statistical appearance modeling of facial images with multiple modes," *Information and Media Technologies*, Vol. 4, no. 4, pp. 999–1009, 2009.

29. 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, no. 3, pp. 283–319, 1970.

30. Harshman, R. A., "Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, Vol. 16, pp. 1–84, 1970.

31. Tucker, L. R., "Some mathematical notes on three-mode factor analysis," *Psychometrika*, Vol. 31, no. 3, pp. 279–311, 1966.

32. Kiers, H. A., "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics: A Journal of the Chemometrics Society*, Vol. 14, no. 3, pp. 105–122, 2000.

33. Acar, E., S. A. Çamtepe, M. S. Krishnamoorthy, and B. Yener, "Modeling and multiway analysis of chatroom tensors," in *Intelligence and Security Informatics*, pp. 256–268, Springer Berlin Heidelberg, 2005.

34. Mocks, J., "Topographic components model for event-related potentials and some biophysical considerations," *IEEE Transactions on Biomedical Engineering*, Vol. 35, no. 6, pp. 482–484, 1988.

35. Shashua, A., and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. I–I, IEEE, 2001.

36. Zhou, H., L. Li, and H. Zhu, "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, Vol. 108, no. 502, pp. 540–552, 2013.

37. De Leeuw, J., "Block-relaxation algorithms in statistics," in *Information Systems and Data Analysis*, pp. 308–324, Springer, 1994.

38. Russell, S. J., and P. Norvig, *Artificial Intelligence: a Modern Approach*, Pearson, 3 ed., 2014.

39. Akgün-Demir, A., "EEG data classification using multilinear regression model," Master's thesis, Bogazici University, Istanbul, Turkey, 2019.

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

41. Holmes, C. J., R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *Journal of Computer Assisted Tomography*, Vol. 22, no. 2, pp. 324–333, 1998.

42. Chang, C. C., and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, pp. 27:1–27:27, 2011.

43. Becker, H., J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, and L. Albera, "Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources," *IEEE Transactions on Affective Computing*, Vol. 11, no. 2, pp. 244–257, 2017.

44. Chen, G., X. Zhang, Y. Sun, and J. Zhang, "Emotion feature analysis and recognition based on reconstructed EEG sources," *IEEE Access*, Vol. 8, pp. 11907–11916, 2020.

45. Padilla-Buritica, J. I., J. D. Martinez-Vargas, and G. Castellanos-Dominguez, "Emotion discrimination using spatially compact regions of interest extracted from imaging EEG activity," *Frontiers in Computational Neuroscience*, Vol. 10, p. 55, 2016.

46. Hosseini, S. A., M. A. Khalilzadeh, M. B. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," in *2010 Second International Conference on Information Technology and Computer Science*, pp. 60–63, IEEE, 2010.

47. Hadjidimitriou, S. K., and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Transactions on Biomedical Engineering*, Vol. 59, no. 12, pp. 3498–3510, 2012.

48. Kringelbach, M. L., J. O'Doherty, E. T. Rolls, and C. Andrews, "Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness," *Cerebral Cortex*, Vol. 13, no. 10, pp. 1064–1071, 2003.

49. Harmon-Jones, E., P. A. Gable, and C. K. Peterson, "The role of asymmetric frontal cortical activity in emotion-related phenomena: a review and update," *Biological Psychology*, Vol. 84, no. 3, pp. 451–462, 2010.

50. Müller, M. M., A. Keil, T. Gruber, and T. Elbert, "Processing of affective pictures modulates right-hemispheric gamma band EEG activity," *Clinical Neurophysiology*, Vol. 110, no. 11, pp. 1913–1920, 1999.