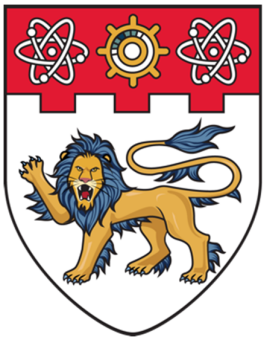


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

EEG-BASED EMOTION RECOGNITION USING MACHINE LEARNING TECHNIQUES

LAN ZIRUI

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

2018

EEG-BASED EMOTION RECOGNITION USING MACHINE LEARNING TECHNIQUES

LAN ZIRUI

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

A THESIS SUBMITTED TO THE NANYANG TECHNOLOGICAL UNIVERSITY
IN FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

2018

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

2018 July 01	LAN ZIRUI
Date	Name

Abstract

Electroencephalography (EEG)-based emotion recognition attempts to detect the affective states of humans directly via spontaneous EEG signals, bypassing the peripheral nervous system. In this thesis, we explore various machine learning techniques for EEG-based emotion recognition, and focus on the three research gaps outlined as follows.

1. Stable feature selection for recalibration-less affective Brain-Computer Interfaces
2. Cross-subject transfer learning for calibration-less affective Brain-Computer Interfaces
3. Unsupervised feature learning for affective Brain-Computer Interfaces

We propose several novel methods in this thesis to address the three research gaps and validate our proposed methods by experiments. Extensive comparisons between our methods and other existing methods justify the advantages of our methods.

Keywords: Electroencephalography (EEG), affective Brain-Computer Interface (aBCI), stable affective features, feature selection, classification, recalibration-less aBCI, domain adaptation, cross-subject transfer learning, cross-dataset transfer learning, unsupervised feature extraction, neural networks, autoencoder.

Acknowledgements

I wish to thank my supervisors, Dr. Olga Sourina, Assoc. Prof. Lipo Wang, Prof. Gernot R. Müller-Putz and Assoc. Prof. Reinhold Scherer, for their valued advice on my research topic.

I wish to thank the Nanyang Technological University, the Fraunhofer Singapore, and the Graz University of Technology. This research project is carried out under the NTU-TU Graz joint PhD degree program coordinated by Fraunhofer Singapore. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centers in Singapore Funding Initiative.

I'm never alone on my academic journey. I am grateful to my family members for their unconditional support and love. Many thanks to my mom, my late dad, my late grandma and grandpa for having fostered a loving and supportive family for me. Many thanks to my uncles, aunties and cousins for the sweet childhood memories. Many thanks to the PhD candidate, Miss Lin Xueling for the spiritual support whenever I feel lost, and the heartening encouragement when I was on the brink of giving up. I wish you the very best in your academic career.

Thanks to the NTU Lifeguard Corps for the swimming training, which enables me to overcome physical and academic challenges with improved endurance and stamina. Thanks to Dr. Olga Sourina and Dr. Wolfgang Müller Wittig from Fraunhofer Singapore for the generous financial support, which greatly eases my financial burden in my last year of PhD study. Thanks to Dr. Liu Yisi for her inspiring discussions with me on my research topic.

Thanks to the Institute of Neural Engineering, Graz University of Technology for making my joint education experience fruitful and unforgettable. Thanks to Dr. Gernot Müller-Putz, Dr. Reinhold Scherer for the valued input to my research topic. Thanks to Gernot for organizing the wonderful institutional hiking trips to the Schöckl and Hochlantsch. Thanks to Reini for the awesome football watching and BBQ session for the UEFA Euro 2017 Final. Thanks to TU Graz for the lunch with President Bertil Andersson and President Harald Kainz. Thanks to ESN Graz for the fantabulous ski trip to Kaprun, Zell am See. Thanks to Miss Mary Chen for organizing various Chinese cultural events, which sort of soothes my homesickness. Thanks to the Karl Franzens Universität Graz for the fun-filled fitness training courses, wake boarding course, windsurfing course and German language course. Thanks to the PhD-wanna-be, Mr. Yam Kah Meng for the memorable trip to Germany and Switzerland, and the useful tips for learning German, and many more.

Special thanks are owed to the participants who contributed to the EEG data collection for my research, and, to those kind souls who have helped me in one way or another. A BIG THANK YOU!

Zirui

July 2018

Table of Contents

Abstract ..... **I**
Acknowledgements ..... **III**
Table of Contents ..... **V**
List of Figures ..... **IX**
List of Tables ..... **XI**
Abbreviations ..... **XIII**
Notations ..... **XVII**
Summary ..... **XIX**
Chapter 1 Introduction **1**
1.1 Background 2
1.2 Motivation 3
1.3 Objectives 4
1.4 Contributions 7
1.5 Organization of Thesis 8
Chapter 2 Related Works **11**
2.1 Human Emotion Models 12
2.1.1 Discrete Emotion Model 12
2.1.2 3D Emotion Model 12
2.2 Self-Assessment Manikin 14
2.3 Affective Stimuli 15
2.3.1 International Affective Digitized Sounds 16
2.3.2 International Affective Picture System 17
2.4 EEG Signal Basics 19
2.4.1 EEG Channel Positions 20
2.4.2 EEG Frequencies 23
2.5 Correlation between EEG and Human Emotion 26

2.6 Fundamentals of EEG-based Emotion Recognition 28
2.6.1 System Overview 28
2.6.2 A Comparison of Existing Studies and the State of the Art 32
2.7 Chapter Conclusion 35
Chapter 3 Stable Feature Selection for EEG-based Emotion Recognition 49
3.1 Problem Statement 50
3.2 Methods 51
3.2.1 Feature Extraction 51
3.2.2 Feature Stability Measurement 56
3.2.3 Stable Feature Selection 58
3.3 Experiments 60
3.3.1 Data Collection 60
3.3.2 Simulation 1: With Re-Calibration 62
3.3.3 Simulation 2: Without Re-Calibration 64
3.3.4 Simulation 3: Stable Feature Selection 64
3.4 Results and Discussions 66
3.4.1 Simulation 1: With Re-Calibration 66
3.4.2 Simulation 2: Without Re-Calibration 68
3.4.3 Simulation 3: Stable Feature Selection 71
3.5 Chapter Conclusions 80
Chapter 4 Subject-Independent EEG-based Emotion Recognition with Transfer Learning 83
4.1 Problem Statement 84
4.2 Datasets 87
4.3 Methods 89
4.3.1 EEG Data Preparation 89
4.3.2 Feature Extraction 90
4.3.3 Domain Adaptation Method 92

4.4 Experiments 100
4.4.1 Within Dataset Domain Adaptation 101
4.4.2 Cross Dataset Domain Adaptation 104
4.5 Results and Discussions 107
4.5.1 Within Dataset Domain Adaptation 107
4.5.2 Cross Dataset Domain Adaptation 113
4.6 Chapter Conclusion 114
Chapter 5 Unsupervised Feature Extraction with Autoencoder 117
5.1 Problem Statement 118
5.2 Dataset 118
5.3 Methods 119
5.3.1 EEG Data Preparation 119
5.3.2 Power Spectral Density 120
5.3.3 Autoencoder 121
5.3.4 Proposed Unsupervised Band Power Feature Extraction 122
5.4 Experiments 123
5.4.1 Using Spectral Power Features 123
5.4.2 Using Features Learned by Autoencoder with the Proposed Structure 124
5.5 Results and Discussions 125
5.5.1 Comparison with Spectral Power Features 125
5.5.2 Comparison with Other Neural Networks 128
5.6 Chapter Conclusion 132
Chapter 6 Applications 135
6.1 Emotion Avatar 136
6.2 CogniMeter 140
6.3 Chapter Conclusion 141
Chapter 7 Conclusions and Future Work 143
7.1 Conclusions 144

7.2 Future Work 148
Reference 153
Author's Publication List 175
Appendix A Experiment Materials 177
I Affective Sound Clips 177
II Self-Assessment Questionnaire 177
III Stable Feature Ranking 180

List of Figures

Figure 2.1 3D emotion model. Image adapted from [42]. 13

Figure 2.2 The self-assessment manikin questionnaire. (a) Valence ratings; (b) Arousal ratings; (c) Dominance ratings. Image adapted from [24]. 15

Figure 2.3 Distribution of IADS audio stimuli. 16

Figure 2.4 Distribution of IAPS visual stimuli. 17

Figure 2.5 Pictorial examples and their targeted emotions. Target emotion: (a) pleasant, (b) (sad), (c) frightened, and (d) excited. 18

Figure 2.6 International 10-20 system [27]. 21

Figure 2.7 Extended EEG electrode position nomenclature by American Electroencephalographic Society [28]. 22

Figure 2.8 Illustration of 2-second long EEG waveforms of different frequency bands. 25

Figure 2.9 EEG-based emotion recognition algorithm framework. (a) Calibration session; (b) Real-time emotion recognition session. 29

Figure 3.1 Emotiv EEG device [2]. 60

Figure 3.2 Protocol of emotion induction experiment. 61

Figure 3.3 Division of the EEG trial. 62

Figure 3.4 Comparison of recognition accuracy between Simulation 1 and Simulation 2. 70

Figure 3.5 Feature ranking in descending order of stability measured by mean ICC scores over all subjects. 72

Figure 3.6 The classification accuracy of inter-session leave-one-session-out cross-validation for each subject and each classifier using the top n stable features selected on a subject-independent basis. 72

Figure 3.7 ICC scores of each feature and the inter-session leave-one-session-out cross-validation accuracy using the top n stable features, $1 \leq n \leq 255$. 75

Figure 4.1 Data sample distribution (feature level) from four subjects from DEAP dataset. 85

Figure 4.2 Illustration of transductive domain adaptation. 101

Figure 4.3 Illustration of data sample distribution at feature level. 109
Figure 4.4 Classification accuracy with varying latent subspace dimension on (a) DEAP; (b) SEED I. 110
Figure 4.5 Classification accuracy with varying number of source domain samples on (a) DEAP; (b) SEED I 112
Figure 5.1 An example of an autoencoder with one hidden layer. 121
Figure 5.2 Proposed network structure. 123
Figure 5.3 Plots of averaged weights of connection between hidden neurons within the same cluster and input neurons. 127
Figure 6.1 Screenshot of the training session [129]. 138
Figure 6.2 Screenshot of the classifier training menu [129]. 138
Figure 6.3 Subject wearing an Emotiv headset while his emotion was being recognized in real time. 139
Figure 6.4 Screenshot of the real-time emotion recognition application "Emotion Avatar" [129]. (a). pleasant, (b). angry, (c). frightened, (d). happy. 139
Figure 6.5 The block diagram of CogniMeter [130]. 140
Figure 6.6 Screenshot of CogniMeter [130]. 141
Figure A.1 Self-assessment questionnaire. 179

List of Tables

Table 2-1 Frequency band range definition of common EEG waves 23

Table 2-2 An outline of existing studies of EEG-based emotion recognition algorithm in terms of the number of recognized emotions, the number of EEG channels used, and the accuracy reported. Upper half: subject-dependent algorithm; lower half: subject-independent algorithm. 36

Table 2-3 A summary of review of EEG-based emotion recognition algorithms. Upper half: subject-dependent algorithms; lower half: subject-independent algorithms. 38

Table 3-1 Review on EEG feature stability. 57

Table 3-2 The analysis of variance table. 59

Table 3-3 Referenced state-of-the-art affective EEG features 63

Table 3-4 Four-emotion recognition accuracy of Simulation 1, simulating the use case where re-calibration is permitted during the long-term use of the affective BCI. Mean accuracy (%) ± standard deviation (%). 67

Table 3-5 Four-emotion recognition accuracy of Simulation 2, simulating the use case where no re-calibration is permitted during the long-term use of the affective BCI. Mean accuracy (%) ± standard deviation (%). 69

Table 3-6 The accuracy (%) ± standard deviation (%) of inter-session leave-one-session-out cross-validation using stable features selected on a subject-independent basis (SISF). 73

Table 3-7 The best mean accuracy of inter-session leave-one-session-out cross-validation evaluation using the top n stable features. Mean accuracy (%) ± standard deviation (%) (# of stable features) . 76

Table 3-8 Comparison of inter-session leave-one-session-out cross-validation accuracy on the *test data* between using referenced state-of-the-art feature set and stable feature set selected by our proposed algorithm. Mean accuracy ± standard deviation. 79

Table 4-1 Technical comparisons between DEAP and SEED. 90

Table 4-2 Comparison of different domain adaptation techniques. 100

Table 4-3 Details of hyperparameters..... 103

Table 4-4 Within-dataset leave-one-subject-out cross-validation accuracy, mean % (std %). 104

Table 4-5 Cross-dataset leave-one-subject-out cross-validation accuracy, mean % (std %). 106

Table 4-6 Computation time (s) of each domain adaptation method on both datasets. 113

Table 5-1 Overall mean classification accuracy (%) classifying three emotions (positive, neutral and negative) using different features. 126

Table 5-2 Comparison of three-emotion recognition accuracy between our method and MLP, RBF NN and SNN. 130

Table A-1 Selected IADS sound clips for emotion induction experiment. 178

Table A-2 Subject-independent ICC score-based feature stability ranking. 181

Table A-3 Subject-dependent ICC score-based feature stability ranking. 182

Abbreviations

aBCI	Affective Brain-Computer Interface
ADHD	Attention Deficit Hyperactive Disorder
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BCI	Brain-Computer Interface
BN	Bayesian Network
CC	Cross Correlation
CPU	Central Processing Unit
CSP	Common Spatial Pattern
DE	Differential Entropy
DEAP	Database for Emotion Analysis using Physiological signals
DFT	Discrete Fourier Transform
DT	Decision Tree
ECoG	Electrocorticogram
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
ERP	Event Related Potential
ERS	Event Related Synchronization
FD	Fractal Dimension
FFT	Fourier Transform
GB	Gigabyte
GFK	Geodesic Flow Kernel
HA	High Arousal
HAHV	High Arousal High Valence

HALV	High Arousal Low Valence
HOC	Higher Order Crossings
HOS	Higher Order Spectra
HSIC	Hilbert-Schmidt Independence Criterion
HV	High Valence
IADS	International Affective Digitized Sounds
IAPS	International Affective Picture System
ICC	Intra-class Correlation Coefficient
IIR	Infinite Impulse Response
ITL	Information-Theoretical Learning
K-NN	K-Nearest Neighbor
KPCA	Kernel PCA
LA	Low Arousal
LAHV	Low Arousal High Valence
LALV	Low Arousal Low Valence
LDA	Linear Discriminant Analysis
LinSVM	Linear Support Vector Machine
LPP	Late Positive Potential
LR	Logistic Regression
LV	Low Valence
MD	Mahalanobis Distance
MIDA	Maximum Independence Domain Adaptation
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
NIMH	National Institute of Mental Health
NN	Neural Network
PC	Personal Computer
PCA	Principal Component Analysis
PSD	Power Spectrum Density

QDA	Quadratic Discriminant Analysis
RAM	Random Access Memory
RBF	Radial Basis Function
RT	Regression Tree
RVM	Relevance Vector Machine
SA	Subspace Alignment
SAM	Self-Assessment Manikin
SDSF	Subject-Dependent Stable Feature
SEED	Shanghai Jiao Tong University Emotion EEG Dataset
SISF	Subject-Independent Stable Feature
SSVEP	Steady-State Visually Evoked Potential
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TCA	Transfer Component Analysis

Notations

x	Scalar variable
$\boldsymbol{x}$	Vector variable
x_i	The ith element of $\boldsymbol{x}$
$\boldsymbol{x}(i)$	The ith element of $\boldsymbol{x}$
n	Cardinal number
N	Cardinal constant
$\boldsymbol{X}$	Matrix variable
x_{ij}	The matrix element of the ith row and the jth column
$\boldsymbol{X}(i,j)$	The matrix element of the ith row and the jth column
$x_{i\cdot}$	The sum of matrix elements of the ith row
$\bar{x}_{i\cdot}$	The mean of matrix elements of the ith row
$\boldsymbol{X}_{i:}$	The ith row of matrix $\boldsymbol{X}$
$\boldsymbol{X}(i,:)$	The ith row of matrix $\boldsymbol{X}$
$\boldsymbol{X}_{:j}$	The jth column of matrix $\boldsymbol{X}$
$\boldsymbol{X}(:,j)$	The jth column of matrix $\boldsymbol{X}$

Summary

Human emotions are complex states of feelings that result in physical and psychological changes, which can be reflected by facial expressions, gestures, intonation in speech etc. Electroencephalogram (EEG) directly measures the changes of brain activities, and emotion recognition from EEG has the potential to assess the true inner feelings of the user. Now, EEG-based emotion recognition draws attention because it is desirable that a machine can recognize human emotions during task performance and interact with us in a more humanized way. EEG can be added as an additional input to a computer during the human-machine interaction. The state-of-the-art EEG-based emotion recognition algorithms are subject-dependent and require a training session prior to real-time emotion recognition. During the training session, stimuli (audio/video) are presented to the subject to induce certain targeted emotions and meanwhile, the EEG of the subject is recorded. The recorded EEG data are subject to feature extraction to extract numerical feature parameters, and the extracted features are fed into a classifier to learn the association with their labels. However, it was found that even for the same subject, the affective neural patterns could vary over time, hence degrading the recognition accuracy in the long run. This phenomenon is termed "intra-subject variance". Due to the existence of intra-subject variance, an EEG-based emotion recognition algorithm needs frequent re-calibration, as frequent as almost every time before running the recognition algorithm. Therefore, stable features are desired, so that re-calibration could possibly be reduced. A stable EEG feature should ideally give consistent measurements of the same emotion on the same subject over the course of time. An affective EEG database that contains multiple EEG recordings on the same subject is needed for such investigation, preferably recordings on different days. In order to establish an affective EEG database

intended for stability investigation, we designed and carried out experiments to collect EEG signals from multiple subjects across eight days. Two sessions were recorded within one day per subject. In each session, four emotions were induced by audio stimuli chosen from International Affective Digitized Sounds (IADS). We examined the stability of the state-of-the-art EEG features across all sessions by Intra-class Correlation Coefficient (ICC). We hypothesized that features with high ICC measures are more stable than those with lower ones. As such, by selecting the more stable features, we optimize the recognition performance for the long run. We proposed a stable feature selection algorithm based on ICC score ranking. The proposed algorithm selects features that maximize the inter-session recognition accuracy, which simulates the performance of an EEG-based emotion recognition system in the long run. Experiments on our dataset established our hypothesis and the effectiveness of our proposed algorithm. The proposed algorithm selects features that yields better accuracy than the best-performing of the state-of-the-art features by 0.62 % – 8.47 %. on the training set, and by 0.23 % – 6.16 % on the test set.

It is also known that subject-independent emotion recognition algorithms, which construct the classifier with training data from other subjects instead of the test subject in question, generally yield inferior accuracies compared to subject-dependent algorithms. However, subject-independent algorithms could make a great practical sense as it frees the user from the initial and subsequent calibrations of the system. We investigated the effectiveness of transfer learning techniques in improving the performance of subject-independent emotion recognition algorithms. We hypothesize that the higher discrepancy of data distribution (at feature level) between different subjects is the cause of the lower accuracy of subject-independent emotion recognition algorithm. Transfer learning techniques or specifically, domain adaptation techniques, have been introduced to help bridge the discrepancy. Simulations of subject-independent three-emotion recognition and extensive comparisons between different transfer

learning techniques were carried out on two publicly available datasets: DEAP and SEED. By leveraging various domain adaptation techniques, the recognition accuracy can be improved by as much as 9.88 % on DEAP and 20.66 % on SEED, respectively. In the scenario of cross-dataset transfer learning, the training data were from DEAP and test data from SEED and vice versa. The recognition accuracy can be improved by up to 13.40 % using Maximum Independence Domain Adaptation (MIDA) when training data were contributed by DEAP and test data by SEED, and by up to 9.41 % by Transfer Component Analysis (TCA) vice versa. Experimental results confirmed our hypothesis that transfer learning technique could help improve the accuracy of subject-independent emotion recognition algorithms.

Lastly, we investigated unsupervised feature learning from EEG using deep learning techniques. In contrast to hand-engineered feature extraction, which requires abundant expert knowledge, deep learning techniques like autoencoder are able to automatically extract features from raw EEG signals or relatively low-level features. We hypothesize that the most discriminative spectral components with respect to emotion recognition may likely be subject-dependent and differ from standard spectral bands such as delta, theta, alpha, and beta bands. We leveraged autoencoder to automatically learn salient frequency components from the power spectral density of the raw EEG signals on an unsupervised basis. We proposed to cluster the hidden units into several groups based on the similarity of their weights, with added pooling neuron on top of each cluster. We hypothesize that neurons that carry similar weights have learned similar components, and different clusters of neurons have learned different components. A pooling neuron is then added to each cluster to aggregate the output of all neurons within the same cluster. The proposed autoencoder structure extracts feature similar to spectral band power features, but without predefining the frequency bands (delta, theta, alpha or beta). The recognition accuracy using the proposed autoencoder was benchmarked against

that using standard power features on SEED dataset and experimental results showed that with features extracted by the proposed autoencoder, the recognition accuracy could outperform that using standard spectral power features by 4.37 % to 18.71 %. We also compare our proposed structure with other neural networks such as multilayer perceptron (MLP), radial basis function neural network (RBF NN) and spiking neural network (SNN). Extensive comparisons show that our method outperforms MLP by 12.73 % – 13.82 %, RBF NN by 3.33 % – 5.66 %, and SNN by 5.64 % – 11.35 %.

With the emotion recognition algorithms, EEG-enabled human-computer interfaces can be adapted to the user's internal feelings and can be driven by the user's emotions. The affective interfaces can be applied in many applications such as 1) games where the flow can be changed according to user's emotions, 2) medical applications to monitor emotions of the patient, 3) neuromarketing, 4) human factors evaluation, etc.

Chapter 1 Introduction

Chapter 1 begins with the background introduction to and the motivation for EEG-based emotion recognition. We then outline three research questions and address the objectives of our research to each of the questions. We then proceed to summarize the contribution of this thesis and conclude the chapter with the organization of the thesis.

1.1 Background

Electroencephalogram (EEG) is the recording of the electric potential of human brain. The first effort to capture and record human EEG was by a German physiologist and psychiatrist Hans Berger in 1924 [1]. Since then, EEG was gradually adopted in clinical environment to facilitate the diagnosis of certain brain diseases, such as Epileptic Seizure, Attention Deficit Hyperactive Disorder (ADHD) and Alzheimer's Disease etc. For over decades, EEG has been limited to medical environment, as the EEG device was costly and immobile, and the acquisition of EEG required professional help. However, in recent years, the advancement of manufacturing technology has introduced to the market new EEG devices which are wearable, portable, wireless and easy to use. Such new devices greatly simplify the acquisition of EEG. Subjects can easily access their EEG even without the help of medical professional. This paves the way for the application of EEG to expand from medical use to personal entertainment use. On the other hand, technological alternatives to EEG such as Functional Magnetic Resonance Imaging (FMRI), Magnetoencephalography (MEG) and functional Near-Infrared Spectroscopy (fNIR) remains expensive and limited to medical use.

Apart from its key role as a diagnostic tool for the brain in the healthcare sector, the current applications of EEG include but are not limited to a) entertainment, b) cognitive training, c) rehabilitation, d) human factors investigation, e) marketing, and f) brain-computer interfaces. For example, EEG-driven games have been developed and distributed by companies like Emotiv [2] and NeuroSky [3] to entertain the users. EEG-based neurofeedback training can be used to improve the cognitive ability of healthy subjects and help to recover part of the brain functions for patients who suffer from stroke attack [4-6] or substance addiction [7, 8]. Researchers have also used EEG to analyze human factors for workplace optimization [9, 10]. Economists have adopted EEG to

detect the brain process that drives the decisions on purchasing [11, 12]. In this thesis, we focus on EEG-based Brain-Computer Interfaces (BCI, [13]) and specifically, affective Brain-Computer Interfaces (aBCI, [14]). A BCI is a direct communication pathway between the human brain and the computer. Using a brain-computer interface, a user can potentially issue a command to a computer by thinking, bypassing the peripheral nervous system. An affective brain-computer interface further introduces the affective factors into the interaction between the user and the computer. An ideal affect-enabled BCI can detect the affective state felt by the user without explicit user input but via spontaneous EEG signals, and respond to different affective states accordingly.

EEG-based emotion recognition lies at the core of aBCI. Human emotions are complex states of feelings that result in physical and psychological changes, which are usually accompanied with gestures, facial expressions, changes in intonation in speech etc. Traditionally, efforts to recognize human emotions were based on such surface features like speech and facial expressions [15-17]. However, as human can deliberately change the intonation or disguise the facial expressions, emotion recognition based on such features may not be reliable. EEG-based emotion recognition, when compared to its traditional counterparts, has the potential to assess true inner feelings of the user because EEG directly measures the changes of brain activities. Besides, the high temporal resolution provided by EEG device makes it possible to monitor the user's emotion in real time. In the case when real-time emotion recognition is necessary, EEG-based emotion recognition may be preferable.

1.2 Motivation

Currently, the state-of-the-art EEG-based emotion recognition algorithms are subject-dependent [14], which means that the algorithm is tailored to a particular user and the classifier is trained on the training data obtained from

the user prior to running the real-time emotion recognition application. However, it is known that affective neural patterns are volatile, and that a classifier trained at an early time could perform rather poorly at a later time, even on the very same subject. To maintain satisfactory recognition accuracy, the user needs to frequently re-calibrate the classifier. As the calibration process is laborious, we are motivated to research the stable EEG-based emotion recognition algorithm which may alleviate the burden of frequent re-calibration.

Subject-independent EEG-based emotion recognition algorithm, which constructs the classifier based on training data from a pool of users, generally yields recognition accuracy inferior to subject-dependent algorithm [14]. However, from a practical viewpoint, subject-independent algorithm makes great practical and applicational sense as it completely frees the user from the initial and the subsequent calibrations. We are motivated to research on improving the recognition accuracy of subject-independent algorithm.

EEG feature extraction is one key step to successful EEG-based emotion recognition. For a long time, affective EEG features have been hand-engineered by domain experts. While pertinent to the classification task, hand-engineered feature requires great amount of expertise. Recently, the renaissance of neural networks has reintroduced the possibility of unsupervised feature learning from raw data or relatively low-level feature. It has proven that deep neural networks can effectively learn discriminative features from raw image pixels or speech signals. We are motivated to research unsupervised EEG feature extraction given raw EEG signals or relatively low-level features.

1.3 Objectives

We pursue research to the answer of the following questions.

1. How can we reduce the need of frequent re-calibrations of the classifier on the same subject?

Ideally, a stable affective EEG feature should give consistent measurement of the same emotion on the same subject over the course of time. Given a bag of EEG features, some may be more stable than the others. We hypothesize that by using stable features, the EEG-based emotion recognition system is optimized for the long run, especially when no re-calibration is allowed. The expected research outcomes are

a) A review of the state-of-the-art subject-dependent EEG-based emotion recognition algorithms.
b) An affective EEG dataset that is suitable for feature stability investigation. Such a dataset should contain multiple affective EEG recordings of the same subject over a course of time.
c) A model to quantify the stability score of affective EEG features.
d) A stable feature selection algorithm.

2. How can we improve the recognition accuracy of subject-independent EEG-based emotion recognition algorithm?

Since neural patterns are subject-specific, subject-independent algorithm is known to yield inferior recognition accuracy to subject-dependent algorithm. However, subject-independent algorithm has the advantage of being "plug-and-play", which makes great practical sense as it removes the burden of calibration from the user of interest. We hypothesize that the EEG data of each subject constitute an independent domain, and that domain discrepancies exist between different domains. By leveraging transfer learning techniques, specifically domain adaptation techniques, discrepancies between different domains could be reduced and the recognition accuracy could be improved when the classifier

is trained on EEG data from a pooled of subjects. The expected research outcomes are

a) A review of the state-of-the-art of the subject-independent EEG-based emotion recognition algorithm.
b) A review of domain adaptation algorithms.
c) Extensive comparisons of the effectiveness of different domain adaptation algorithms on affective EEG data.
d) Proposal of subject-independent EEG-based emotion recognition algorithm with integration of domain adaptation technique.

3. Can we extract discriminative affective EEG features on an unsupervised basis?

The state-of-the-art EEG-based emotion recognition algorithms heavily rely on discriminative features hand-engineered by domain experts. While pertinent to the emotion classification task, the engineering of feature requires rich expert knowledge. Neural networks have proven effective in learning features from image and speech data on an unsupervised basis. We explore unsupervised feature learning from EEG and compare the recognition performance using unsupervisedly-extracted features against the hand-engineered counterparts. The expected research outcomes are

a) A review of neural network and specifically, auto encoder.
b) Proposal of a novel architecture for feature extraction from EEG.
c) Extensive comparisons of recognition performance between hand-engineered EEG features and features learned by neural network on an unsupervised basis.

1.4 Contributions

In the pursuit of answers to the above research questions, we make contributions to the following areas.

1. We establish an affective EEG dataset for the purpose of feature stability investigation. The dataset contains multiple recordings over a long course of time for each subject. The dataset will be made available to peer researchers in this and related fields.

2. We carry out a pilot study on feature stability of affective EEG features, and propose a stable feature selection algorithm to choose the optimal feature set for better recognition accuracy where the BCI (Brain-Computer Interface) operates without re-calibrations in the long run. We fill the research gap on the investigation of the long-term recognition performance of an affective BCI, which has rarely been reported in the current literature.

3. We propose a subject-independent emotion recognition algorithm with integration of the state-of-the-art domain adaptation techniques, and carry out a preliminary study on applying domain adaptation techniques for cross EEG dataset emotion recognition. A conventional BCI generally constrains the EEG data collection under the same experimental protocol at training time and test time. A cross EEG dataset recognition paradigm posts a more challenging task as the data collection protocol, affective stimuli, technical specifications of the EEG devices, etc. are different, which introduces considerable variance between datasets. We demonstrate that domain adaptation techniques can mitigate the variance and improve the recognition performance.

4. We propose a neural network structure specially for unsupervised feature extraction from the power spectral density of EEG. Traditionally, spectral band ranges need to be defined explicitly. We propose to leverage neural network to

learn salient frequency components on an unsupervised basis. Spectral features extracted by the proposed method yield accuracy better than that by standard spectral power features.

1.5 Organization of Thesis

The remainder of the thesis is organized as follows.

In Chapter 2, related works on human emotion models, affective stimuli and the EEG signal basics are introduced. Overviews of the correlates between human emotions and EEG signals are given. We then review the fundamentals of EEG-based emotion recognition algorithms. The paradigm of EEG-based emotion recognition system is introduced, followed by a survey of and a comparison between the state of the arts.

In Chapter 3, we simulate and analyze the performance of an EEG-based emotion recognition system in the long run. To alleviate the burden of frequent re-calibration, we propose a stable feature selection algorithm.

In Chapter 4, we revisit the idea of the subject-independent EEG-based emotion recognition, and investigate how different transfer learning techniques can improve the recognition accuracy on a subject-independent basis.

In Chapter 5, we explore unsupervised feature extraction from EEG with deep learning techniques. We propose a novel network architecture specially for EEG feature extraction, and extensively compare the recognition performance between unsupervisedly-extracted features and hand-engineered features.

In Chapter 6, we showcase some EEG-based emotion recognition applications that are integrated with our proposed algorithms.

In Chapter 7, we conclude the thesis and envision some future directions.

Chapter 2 Related Works

Chapter 2 presents the literature review on EEG-based emotion recognition. We begin by introducing two human emotion models that are extensively used in emotion-related studies: discrete emotion model and 3D emotion model. Then, a standardized, non-verbal emotion assessment tool and two affective stimulus libraries are reviewed, followed by EEG signal basics and the correlation between human emotion and EEG signals. The fundamentals of EEG-based emotion recognition are covered under two categories: subject-dependent recognition and subject-independent recognition. Lastly, we present a survey of the state-of-the-arts to conclude the chapter.

2.1 Human Emotion Models

Human emotions are complex states of feelings that result in physical and psychological changes, which can be reflected by facial expressions, gestures, intonation in speech etc. Emotion models are necessary in the study of human emotions. Two emotion models are introduced in this section.

2.1.1 Discrete Emotion Model

The discrete emotion model defines some basic emotions and considers other emotions as the mixture, combination or compound of the basic emotions. The most influential discrete emotion model is proposed by Plutchik [18]. He contended that emotions in animals (including human) are strongly relevant to evolution and argued for the primacy of these eight basic emotions: fear, anger, joy, sadness, acceptance, disgust, expectation and surprise. In his hypothesis, these basic emotions are evolutionally primitive and have emerged to increase the survival and reproductivity of animals [18]. For example, when threatened, the animal feels fearful (basic emotion) and manages to escape, leading to a safety state (increase the survival). Under this model, other emotions can be identified as combination or mixture of the basic emotions. For instance, contempt can be defined as a mixture of disgust and anger emotions.

2.1.2 3D Emotion Model

The 3D psychological emotion model was jointly developed by Mehrabian and Russell [19, 20]. Under this model, emotions can be broken down to and quantified by three orthogonal dimensions, namely valence dimension, arousal dimension, and dominance dimension (see Figure 2.1). The valence dimension measures how pleasant an emotion is, ranging from negative (unpleasant) to positive (pleasant). For example, both fear and sad are unpleasant emotions

and score low in valence level, whereas joyful and surprised are pleasant emotions and score high in valence level. The arousal dimension quantifies how intense an emotion is, ranging from low arousal (deactivated) to high arousal (activated). For instance, sad is a lowly activated emotion and scores low in arousal level, whereas fear is a highly activated emotion and scores high in arousal level. The dominance dimension assesses the controlling and dominating nature of an emotion, ranging from low dominance (being under control/submissive) to high dominance (controlling/dominating). For instance, when a person feels fear, he is submissive to the surroundings and in a low dominance level. When a person feels angry, he stands in a dominating position, tends to aggress and is in a high dominance level.

In comparison with the discrete emotion model, the dimensional emotion model is preferable in emotion-related studies. The dimensional emotion model has

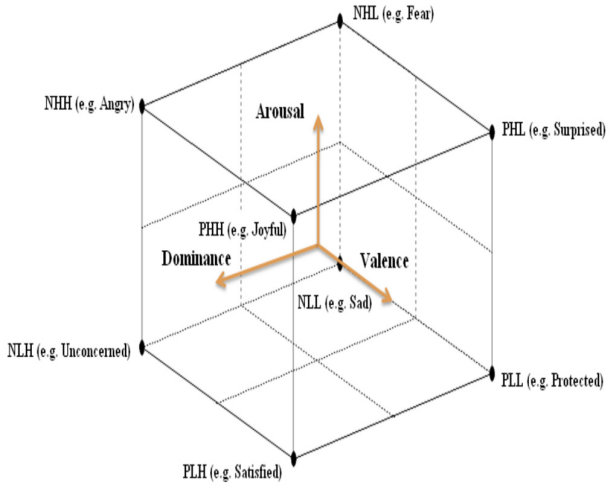
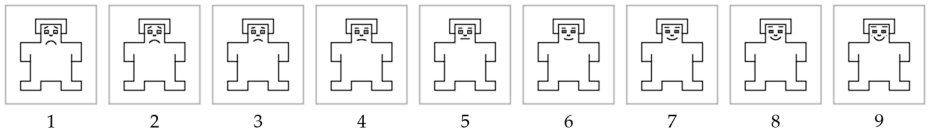


Figure 2.1 3D emotion model. Image adapted from [42]. The 3-letter alphabetic naming (e.g., NHH) follows such a convention: the first letter indicates the valence, either positive (P) or negative(N); the second and the third letter indicate the arousal and dominance, respective, either high (H) or low(L).

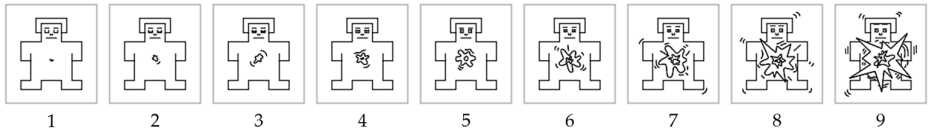
been extensively used in studies [21-23]. It could be true that sometimes we human cannot express or express too vaguely in adjective the emotion we are experiencing. By breaking down the emotion into three dimensions and evaluating each dimension in a quantitative way, the assessment of emotion is made easier and more reliable. In fact, the discrete emotion labels in [18] can be converted to 3D representation without too much effort. For instance, the fear emotion, one of the eight basic emotions in Plutchik's discrete emotion model (see Section 2.1.1), can be broken down to negative valence, high arousal and low dominance in 3D emotion model. Furthermore, the 3D emotion model has the potential to assess the emotion that may not even have an adjective to properly describe it.

2.2 Self-Assessment Manikin

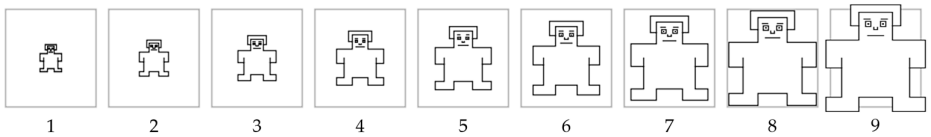
The self-assessment manikin (SAM) [24] is a non-verbal, pictorial tool for the self-assessment of the emotion being experienced by the subject. The SAM questionnaire covers the assessment of the three dimensions of the 3D emotion model: valence, arousal, and dominance. The symbols in the questionnaire are designed to be intuitive and self-explanatory, minimizing the influence of different cultural backgrounds of the subjects. The SAM is illustrated in Figure 2.2. In Figure 2.2a, the curvature of the mouth indicates the pleasantness level. The valence rating ranges from 1 (most unpleasant) to 9 (most pleasant). In Figure 2.2b, the size of the "blasted heart" of the manikin suggests the excitation level. The arousal rating ranges from 1 (calmest/most inactivated) to 9 (most activated). In Figure 2.2c, the size of the manikin is directly proportional to the dominating level of the subject. The dominance level ranges from 1 (submissive) to 9 (dominating). By using a SAM questionnaire, we avoid the vagueness of languages and the emotion assessment is made more reliable. After the subject



(a) Valence ratings



(b) Arousal ratings



(c) Dominance ratings

Figure 2.2 The self-assessment manikin questionnaire. (a) Valence ratings; (b) Arousal ratings; (c) Dominance ratings. Image adapted from [24].

has completed the questionnaire, an emotion can be located within the 3D emotion model given the coordinates of the three dimensions.

2.3 Affective Stimuli

One of the key steps in emotion-related studies is to elicit the desired emotions on the subjects. There exist standard emotion stimuli libraries intending to provide normative emotional stimuli for emotion induction experiment, such as International Affective Digitized Sounds (IADS) [21] and International Affective Picture System (IAPS) [22], which are well-acknowledged and have been used extensively by the research community. Both libraries are based on the 3D emotion model introduced in Section 2.1.2.

2.3.1 International Affective Digitized Sounds

The International Affective Digitized Sounds (IADS) [21] was contributed by the NIMH (National Institute of Mental Health) Center for Emotion and Attention at the University of Florida. IADS contains 167 sound clips, each lasts for 6 seconds. The valence, arousal and dominance levels of each sound clip was rated by and averaged over 100 subjects using the SAM on a scale of 1 to 9. The audio stimuli in IADS cover a good range of emotions within the 3D emotion model. The distribution of the IADS stimuli in the Valence-Arousal-Dominance coordinate is depicted in Figure 2.3. Furthermore, the audio stimuli in IADS, excerpted from real-life scenarios, were carefully chosen in order to minimize the variance of response from subjects coming from

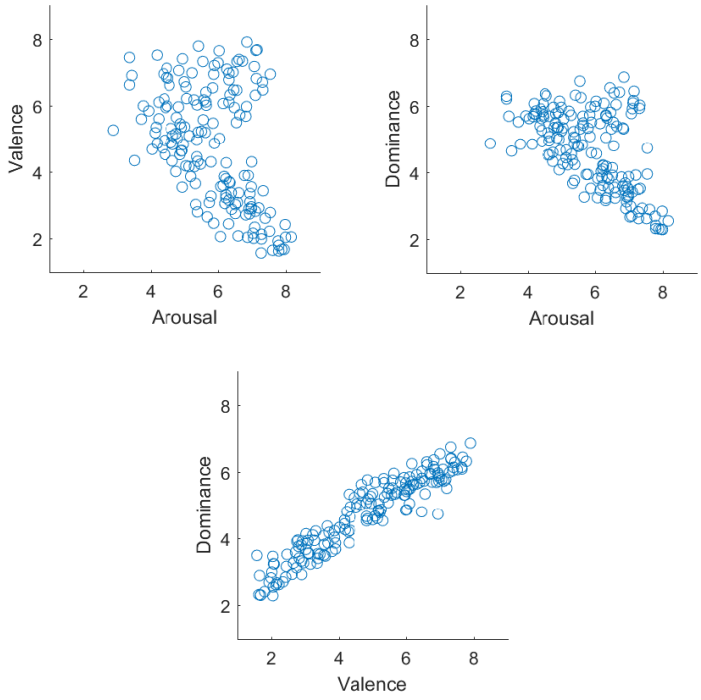


Figure 2.3 Distribution of IADS audio stimuli.

different cultural backgrounds. For example, typical audio stimuli to induce a pleasant emotion include the sound of birds' merry chirping, stream flowing, and kid's laughter etc. Stimuli to elicit a frightened emotion include the sound of woman's screaming and crying, gun shot, and car crash etc.

2.3.2 International Affective Picture System

The International Affective Picture System (IAPS) [22] was also developed by the NIMH Center for Emotion and Attention at the University of Florida, aiming at providing standardized visual stimuli for the research in human emotion. The IAPS contains 700 color photographs collected over the course of 10 years. The valence, arousal and dominance levels of each photo was rated

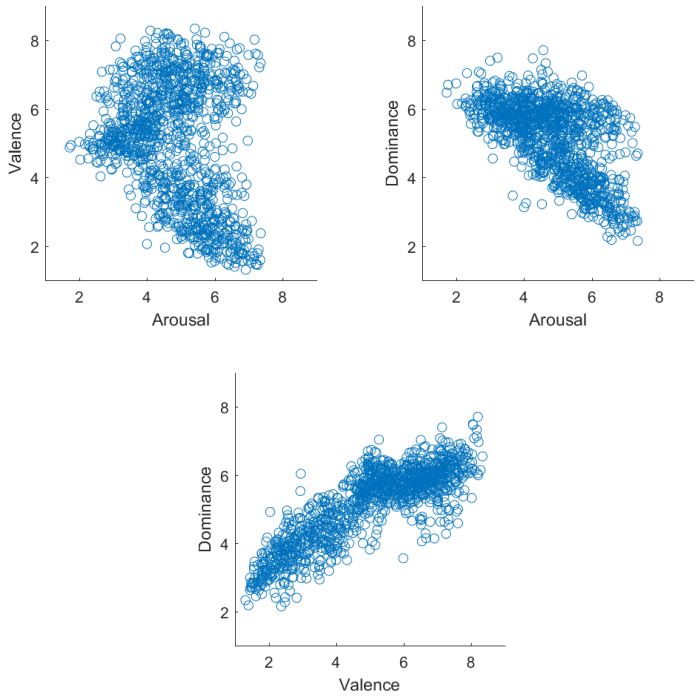


Figure 2.4 Distribution of IAPS visual stimuli.

by and averaged over a large number of subjects coming from different ages, genders and cultural background. The pictures in IAPS were chosen to cover a broad range in each of the three dimensions and at the same time minimize the culture-specific or religion-specific influence. The distribution of the IAPS stimuli in the Valence-Arousal-Dominance coordinate is depicted in Figure 2.4. Some pictures analogous to IAPS stimuli are shown in Figure 2.5. (According to the terms of use, real pictures from IAPS shall not be made known to the public in order to preserve the evocative efficacy.)

(a) Pleasant **(b) Sad**

(c) Frightened **(d) Excited**

Figure 2.5 Pictorial examples and their targeted emotions. Target emotion: (a) pleasant, (b) (sad), (c) frightened, and (d) excited. According to the user agreement with IAPS authors, original pictures from IAPS shall not be published in order to preserve the evocative efficacy. Pictures shown here are analogous to pictures from IAPS. Images sourced from the Internet[1].

[1] The figures were accessed from
(a) https://s3.amazonaws.com/boatbound_production/city_template_photos/city_photos/000/000/470/scaled_down_1440/Lake-Pleasant-Arizona.jpg
(b) https://conteudo.imguol.com.br/c/entretenimento/da/2016/01/15/o-dramatico-filme-o-campeao-de-1979-conta-a-historia-do-ex-campeao-de-boxe-billy-flynn-que-virou-treinador-de-cavalos-para-nao-perder-a-custodia-do-filho-o-lutador-decide-voltar-aos-ringues-1452892818815_956x500.jpg
(c) http://www.garret-dillahunt.net/wp-content/uploads/2009/09/Garret-Dillahunt-in-Lie-to-Me.jpg
(d) http://www.ronaldo7.net/news/archive/news274f.jpg

2.4 EEG Signal Basics

EEG is the recordings of electrical potentials along the scalps. The recorded electrical potentials originate from the intrinsic electrophysiological properties of the neurons lying beneath the skull. EEG recorded at the surface of the scalp comprises the summated postsynaptic potentials created by numerous nerve cells underlying the scalp in the cerebral cortex [25]. Electrical activity generated in sub-cortex is too subtle when propagated to the scalp, to pick up which requires the implantation of electrodes into the brain through craniotomy, and is termed ECoG (Electrocorticography). ECoG is not within the scope of our studies.

EEG signals are generated when electrical charges move within the central nervous system. When charges are moving, electrical current is formed. Neurons are constantly exchanging ions to fulfill the neural function, e.g. to maintain resting potential, to transmit action potentials, etc. A resting membrane potential is maintained by the efflux of positive ions (e.g. potassium) from intracellular to extracellular, establishing an electrochemical equilibrium of around -70 microvolts [26]. To depolarize, an influx of positive ions (e.g. sodium) occurs. Conduction to neighboring neuron membrane results in an action potential provided that the depolarization threshold is exceeded. However, it is the synaptic potentials that are the most important source contributing to EEG [26]. Excitatory postsynaptic potentials flow inwardly, from extracellular to intracellular, via sodium or calcium ions. Inhibitory postsynaptic potentials, carried by chloride or potassium ions, flow outwardly in an opposite way. The electrical potential generated by an individual neuron is never strong enough to be detected. Therefore, EEG always reflects the pooled synchronous activity of numerous neurons that have common spatial orientation. If the neurons do not line up in similar orientation, the EEG current generated by different neurons are likely to cancel out each other and the current can hardly reach the scalp.

Pyramidal neurons in the cerebral cortex are the main contributor of the synaptic potentials that constitute EEG [26]. These neurons are well-aligned in a perpendicular orientation to the cortical surface, and fire synchronously.

2.4.1 EEG Channel Positions

In order to ensure the standardized placement of EEG electrodes on the scalp, systems that regulate the positions as well as the names of EEG electrodes have been devised and widely adopted in EEG research. This section introduces the two well-recognized nomenclature systems, namely the international 10 – 20 system [27] and its extended version guided by American Electroencephalographic Society [28].

2.4.1.1 International 10-20 System

The international 10 – 20 system [27] is the first well-acknowledged method that regulates the locations and the names of scalp electrodes. The "10" and "20" are derived from the distance between any two adjacent electrodes: it is either 10 % or 20 % of the surficial distance on the skull from front to back or right to left (see Figure 2.6) This system associates the location of an electrode with an alphanumeric name (e.g., T3, F4 etc.) based on the cerebral cortex that underlies the electrode and the sequence of the electrode. The letters P, F, C, T and O stand for parietal, frontal, central, temporal and occipital respectively. Pg, Fp and A refer to the nasopharyngeal, frontal polar and earlobe respectively. Following the letter, an even number (2, 4, 6, 8) indicates electrode placements on the right hemisphere, while odd number (1, 3, 5, 7) indicates those on the left. "z" (zero) indicates the positions along the midline.

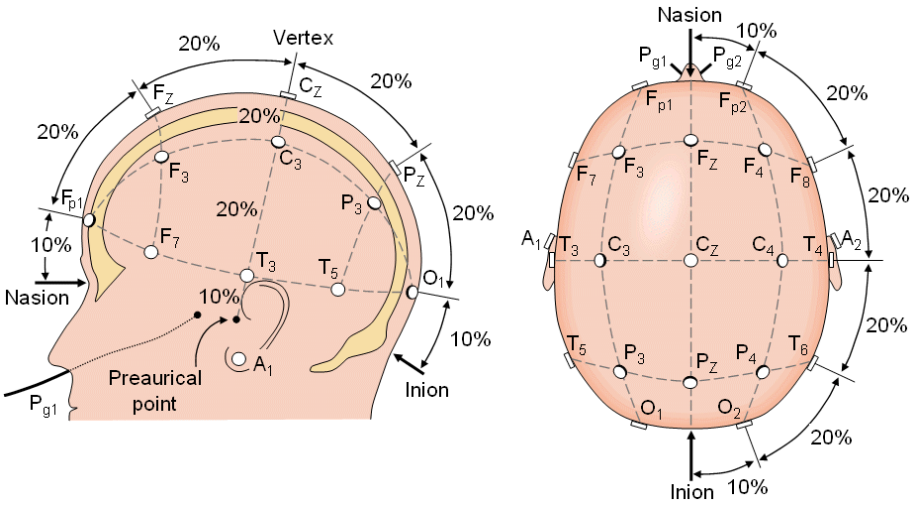


Figure 2.6 International 10-20 system [27].

2.4.1.2 Extended EEG Electrode Position Nomenclature

The international 10-20 system can identify up to 19 electrode positions. With the advancement of manufacturing technology, the new EEG devices come with more channels that beyond the 10-20 system's appellation. In 1991, the American Electroencephalographic Society introduced an extended EEG electrode position nomenclature [28] based on the old 10-20 system. This extended nomenclature can identify up to 73 electrodes (see Figure 2.7). Major revisions as compared to the old system are as follows:

1. The T3/T4 and T5/T6 in the old system are renamed T7/T8 and P7/P8 respectively.
2. Introduce new alphabetic prefix AF (Anterior Frontal), FT (Frontal Temporal), FC (Frontal Central), TP (Temporal Posterior-temporal),

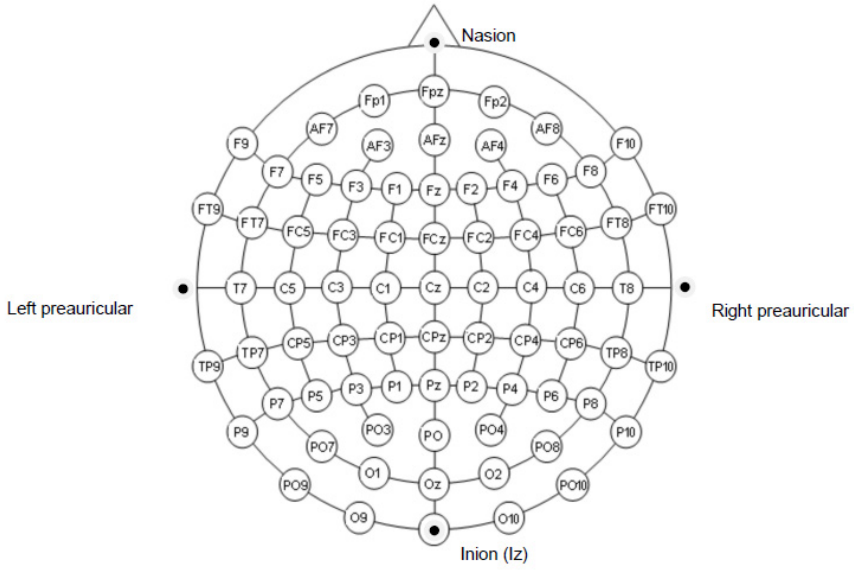


Figure 2.7 Extended EEG electrode position nomenclature by American Electroencephalographic Society [28].

CP (Central Parietal) and PO (Parietal Occipital) to name more electrodes.

With these revisions, the extended system can locate more electrodes on the scalp. Furthermore, the revised system is more internally consistent: all electrodes that lie on the same sagittal line have the same post-scripted number (except for Fp1/Fp2 and O1/O2), and all electrodes that started with the same letters lie on the same coronal line. For example, both T7 and T8 electrodes are on the same coronal line and started with the same letter T. Electrodes AF7, F7, FT7, T7, TP7, P7 and PO7 are on the same sagittal line and post-scripted with the same number 7.

Table 2-1 Frequency band range definition of common EEG waves

Spectral Band	Delta	Theta	Alpha	Beta	Gamma
Frequency Range	1 – 4 Hz	4 – 8 Hz	8 – 12 Hz	12 – 30 Hz	> 30 Hz

2.4.2 EEG Frequencies

EEG signals comprise oscillations at various frequencies. Some of these oscillations demonstrate characteristic frequency ranges and spatial distributions. As a common practice, EEG can be divided into several frequency bands, such as alpha band, beta band, delta band, gamma band, theta band etc. The correlates between EEG rhythms and brain states have been studied [29-40]. However, it must be pointed out that there is no absolute consensus on the band range definition of these frequency bands—different studies could respect different band range definition. In our work, unless otherwise stated, we refer to the definition listed in Table 2-1. The same definition is also respected in [23, 41, 42].

Delta Band

Delta waves (see Figure 2.8a) are the slowest waves generated in brain, and have been related to deep sleep [29]. Delta waves are not dominant brain waves in awoken adults but are seen normally in babies [30]. If observed in awoken adults, it may imply subcortical lesions [31].

Theta Band

Theta waves (see Figure 2.8b) may be associated with drowsiness/somnolence states [32]. They can also be observed during meditation. Some studies report a correlates between theta waves and cognitive mental task [33]. Lesions to the

subcortical region and some instances of hydrocephalus may cause abnomal theta wave pattern [34].

Alpha Band

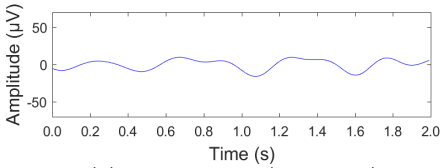
Three brain regions are thought to be associated with alpha waves (see Figure 2.8c). The alpha waves that are prominent in the posterior regions (e.g. visual cortex) of the brain are referred to as alpha rhythm [35]. It occurs when one closes eyes and diminishes when one opens eyes. It is considered as a reflection of relaxed states in rest. The alpha waves that originate from somatosensory cortex are referred to as mu rhythm [36], which is believed to be engaged in motor actions, motor action preparation and motor imagery. The tau rhythm refers to the alpha waves that source from audio cortex, which is triggered by audio stimuli [37].

Beta Band

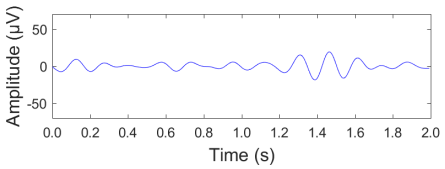
Beta waves (see Figure 2.8d) are evident in frontal and central region of the brain, and associated with motor movement, active thinking, active concentration, high alertness or anxiety. Abnormal beta band activities may indicate addiction to substance [38] such as alcohol and drug etc.

Gamma Band

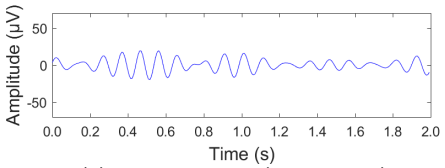
Originating from somatoseneory cortex, gamma waves (see Figure 2.8e) are observed during multi-modal sensory processing [39], i.e., perception that involves together visual, audio and tactile sensory inputs. Gamma waves are also related to short-term memory matching and cognitive tasks [40].



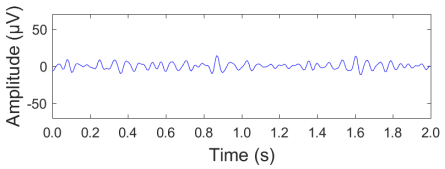
(a) Delta wave (1 – 4 Hz)



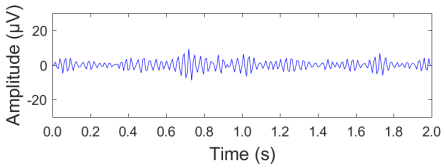
(b) Theta wave (4 – 8 Hz)



(c) Alpha wave (8 – 12 Hz)



(d) Beta wave (12 – 30 Hz)



(e) Gamma wave (> 30 Hz)

Figure 2.8 Illustration of 2-second long EEG waveforms of different frequency bands. EEG signals recoded from the author with an Emotiv Epoc device [2] at 128 Hz sampling rate, and filtered by 20-ordered IIR (Infinite Impulse Response) bandpass filters designed in MATLAB [112]. (a) Delta wave (1 – 4 Hz); (b) Theta wave (4 – 8 Hz); (c) Alpha wave (8 – 12 Hz); (d) Beta wave (12 – 30 Hz); (e) Gamma wave (> 30Hz).

2.5 Correlation between EEG and Human Emotion

Many studies have reported that human emotional states can be traced in EEG [43-64]. In [43], Sammler *et al.* have studied the correlation between EEG theta power at frontal midline and emotions induced by pleasant and unpleasant music. They reported that emotions induced by pleasant music were accompanied by an increase of theta activity at frontal midline. The investigation by Kroupi *et al.* [44] has revealed that arousal was positively correlated with logarithmic mean theta power on the right hemisphere. In [45], Aftanas *et al.* reported that valence was associated with theta ERS (Event-related Synchronization) in anterior temporal regions. Negative valence provoked greater right hemisphere ERS activities while positive valence triggered prominent left hemisphere ERS activities. The studies by Keil *et al.* [46] showed that arousal was modulated by gamma band ERP (Event-related Potentials). 30 – 45 Hz mid gamma band activity at 80 ms after the onset of low arousal stimuli was observed. 46 – 65 Hz high gamma band activity at 500 ms emerged after the onset of high arousal stimuli. Park *et al.* [47] presented in their study the correlates between emotions and different band power activities. Results showed that a significant decrease of alpha wave at the left temporal lobe was observed for negative emotions. An increase of beta activity at the left temporal lobe was associated with the fear emotion, whereas a drop in the alpha activity at C4 was accompanied by the happy emotion. In [48], it was demonstrated that the amplitude of ERP at 100 ms after the onset of happy stimuli was greater than that of sad stimuli. In [49] and [50], Trainor and Schmidt revealed that the asymmetry pattern of frontal activities were associated with valence perception. Positive emotions (joy and happiness) triggered greater left frontal alpha activities, while negative emotions (fear and sadness) provoked greater right alpha activities. This finding was supported by

[51-55], which also reported that left hemisphere was more active when processing positive emotions and right hemisphere was more dominant in processing negative emotions. In [56], Heller claimed that arousal was modulated by the right parietal-temporal lobe, and that a higher parietal-temporal activity indicated a higher arousal level experienced. Choppin in [57] analyzed the correlation between EEG and each of the three emotional dimensions. It was found out that valence was correlated with the frontal coherence in alpha power and right parietal beta power, that arousal was associated with beta power and coherence in parietal lobe, and that dominance was modulated by beta/alpha power ratio at frontal lobe or beta power at parietal lobe. In Bos's review [58], it was documented that alpha waves were more dominant in relaxed states while beta waves were prevailing in alerted states, and the beta/alpha ratio was suggested as a measure of arousal level. In [55], Niemic argued that left frontal was inactive during withdrawal response (i.e., perception of negative emotions except anger) while right frontal inactivation was relevant to approach response (i.e., perception of positive emotions or anger). In [59], an increase of gamma activity during the viewing of unpleasant pictures was observed, and strong temporal-frontal synchronization was present in the 38 – 45 Hz gamma interval at 250 ms from the onset of unpleasant stimulus. Miskovic *et al.* reported in [60] that significant increase of EEG coherence between prefrontal and posterior lobes was accompanied by viewing high arousing stimuli. In [61], Kemp *et al.* examined the relation between valence and SSVEP (Steady-State Visually Evoked Potential). Both positive and negative valences were found to trigger bilateral frontal SSVEP latency reductions, while negative valence has been associated with bilateral frontal amplitude decrease. In [62], Balconi and Mazza reported that increased right frontal activities were observed when negative emotions were being perceived. Stenberg's investigation [63] revealed that increased temporal beta activity was in association with positive emotions while decreased with negative emotions. Moreover, he further revealed that this pattern could

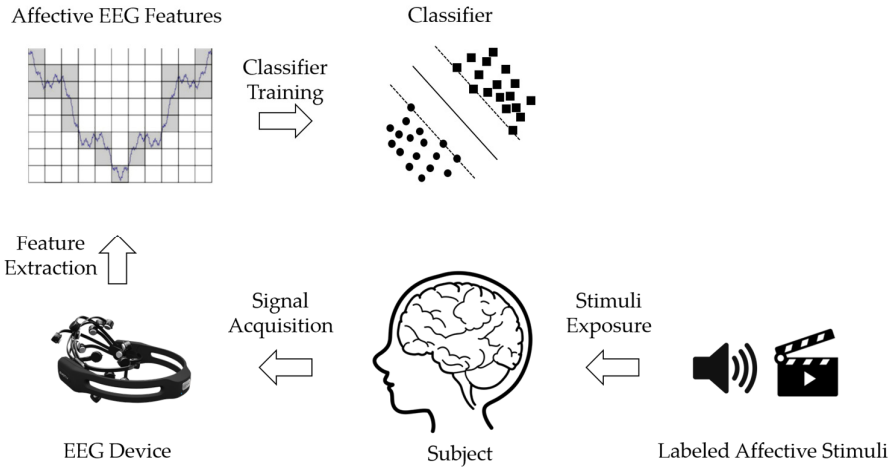
be amplified for subjects who were possessed of impulsive or anxious personality. In [64], correlation between EEG and emotional intelligence (the ability to manipulate one's emotions) were established. It was found out that subjects with higher emotional intelligence demonstrated greater left frontal activities. In [50], the intensity of emotions induced by music has been associated with overall frontal activities. The frontal activities decreased from fear to joy to happy to sad. Kawano *et al.* [65] studied the similarity in EEG between perceiving facial expression and mimicking facial expression, and reported a high similarity in EEG when the subject is perceiving or mimicking sadness.

Even though different studies focus on different EEG components, it is evident that emotions can be traced in EEG. Moreover, the asymmetry property of EEG has been consistently reported to be relevant to emotion states or emotion changes in these studies [49-55]. Therefore, it is feasible to employ EEG as an emotion indicator.

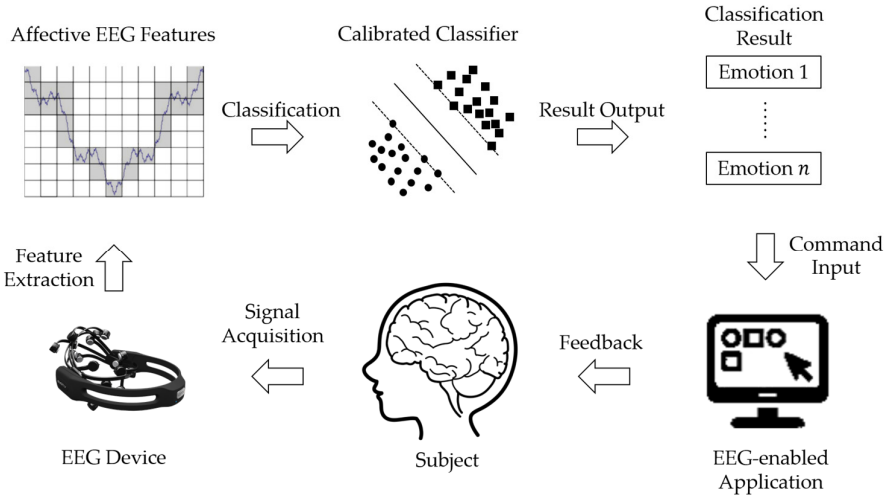
2.6 Fundamentals of EEG-based Emotion Recognition

2.6.1 System Overview

A typical EEG-based emotion recognition system generally consists of four core steps as depicted in Figure 2.9b: signal acquisition, feature extraction, classification, and feedback to the user. To obtain better recognition accuracy, the system needs calibration before use. The calibration process is illustrated in Figure 2.9a. During the calibration, the subject is exposed to the affective stimuli chosen from the stimulus library (e.g., IADS, IAPS), and meanwhile the EEG is being recorded. The recorded EEG data are subject to feature extraction, and then fed to a classifier, which learns the association between the neural patterns and the emotion labels. After calibration, the system is ready to use.



(a) Calibration session



(b) Real-time emotion recognition session

Figure 2.9 EEG-based emotion recognition algorithm framework. (a) Calibration session; (b) Real-time emotion recognition session

In the real-time emotion recognition session, the subject generates spontaneous EEG signals, which are captured by the EEG device. The EEG signals are fed to feature extractor, and then to the calibrated classifier. The classifier outputs the emotion label based on the association it has learned during the calibration session. The application program executes subroutines according to the recognized emotion results. Feedback is then given to the user.

In general, EEG-based emotion recognition algorithms can be categorized into two types: subject-dependent algorithms and subject-independent algorithms. Subject-dependent algorithms tailor the classifier to the subject-of-interest, requiring labeled training data pertinent to the subject to calibrate the classifier. Subject-independent algorithms construct the classifiers with training data from other subjects. Subject-dependent algorithms generally yield better recognition accuracy than subject-independent algorithms.

2.6.1.1 Classifiers

Feature extraction and classification are two key steps in the systems illustrated in Figure 2.9. Here, we present a brief survey of classifiers used in EEG-related studies, including but not limited to EEG-based emotion recognition, and leave the review of feature extraction to Section 3.2.1 in the next chapter. A more elaborative survey of classifiers used in BCI can be found in [66, 67].

Linear Discriminant Analysis (LDA): LDA [68] seeks a separation hyperplane that maximizes the distance between the means of the two classes' projected samples and minimizes the intraclass variance of the projected samples of both classes. LDA can be generalized to solve multiclass classification problem by using several separation hyperplanes and the "one versus the rest" strategy. This classifier has low computational requirement, few hyperparameters to tune and simple to use. LDA is said to be stable in that small variations in the training set do not considerably affect their performance

[69]. LDA has been used with success in many BCI such as P300 speller [70], motor-imagery BCI [71], and affective BCI [72-74].

Support Vector Machine (SVM): SVM [68, 75] seeks a separation hyperplane that maximizes the margin between two classes, that is the distance between the two parallel plane boundaries of the two classes. Samples on the boundaries are called support vectors, hence the name support vector machine. SVM can deal with nonlinearly separable data by leveraging the "kernel trick", which implicitly maps the data from their original feature space to a higher dimensional space, where the data become linearly separable. SVM handles multiclass problem by using "one versus the rest" strategy, similar to LDA. Thanks to its capability of classifying nonlinear patterns, SVM has been extensively used in BCI [76-78] and aBCI [72, 79-98].

Naïve Bayes (NB): NB [68] assumes that each feature is generated by a presumed probability distribution model, and further assumes a "naïve" conditional independence that each feature is conditionally independent of every other feature, given the class category. The training data are used to estimate the class probability (a priori) and the distribution parameters for each feature (e.g., if Gaussian distribution is assumed, then the mean and variance are the parameters to be estimated). Then, NB uses the Bayes rule to calculate the class probability given the features (posteriori probability). NB can naturally handle multiclass problem. This classifier is not widely used in aBCI but has been applied with success to motor-imagery BCI [99, 100].

Logistic Regression (LR): LR [101] estimates the probabilities of the binary classes by calculating a linear combination of the features and then passing the intermediate result through a logistic sigmoidal function to produce the final output. The output of the logistic sigmoidal function is bounded by (0, 1), thus can be interpreted as the probabilities of classes. LR is extended to handle

multiclass problem by similar strategy as SVM and LDA. LR has been applied to motor imagery BCI [102], seizure detection [103], and aBCI [74, 89, 97, 98].

K-Nearest Neighbor (K-NN): K-NN [68] classifies an unseen sample the major classes among its k nearest neighbors within the training set. These nearest neighbors are usually obtained by a distance metric, e.g., Euclidean distance. K-NN naturally handles multiclass problem, and given a reasonably large k and a large enough training set, K-NN can produce nonlinear separation boundaries. K-NN has been applied to many aBCI studies, such as [73, 74, 84, 88, 89, 91].

2.6.2 A Comparison of Existing Studies and the State of the Art

In this section, we present a survey of the state-of-the-arts specifically on EEG-based emotion recognition. Ignoring the implementation details, different algorithms mainly differ in the proposed features, classifiers, target emotions, and EEG channels, etc. [104] used wavelet transform, Fast Fourier Transform (FFT) and statistics such as mean and variance as feature and employed Neural Network (NN) to classify four emotions. An accuracy of 67.70 % was achieved with 3 channels. [79] utilized power differences at symmetric electrode pairs and SVM to classify four emotions and obtained an accuracy of 90.72 % with 32 channels. Recognizing three emotions, Schaaff [80] and Chanel *et al.* [72] made use of Short Time Fourier Transform (STFT) to extract power feature and SVM as classifier and achieved 62.07 % (16 channels) and 63.00 % (64 channels) accuracy respectively. In another work, Shaaff [81] reduced the number of channels to 4 and adopted the statistical features, power features and cross correlation features together with an SVM classifier. The recognition accuracy was 47.11 % for 3 emotion classes. In [82], four emotions were recognized with differential asymmetry of hemispheric EEG power spectra as feature and SVM

as classifier, and obtained an accuracy of 82.29 % using 32 channels. [73] recognized five emotions at an accuracy of 83.04 % using statistical features from different EEG bands from 62 channels and K-NN classifier. [83] used SVM to differentiate two emotions featured by logarithmic variances from 62 channels and achieved 93.50 % accuracy. [87] employed Fractal Dimension (FD) feature together with statistical and Higher Order Crossings (HOC) features, and a SVM classifier was used. Up to eight emotions were recognized with four channels. The average accuracy obtained ranged from 53.75 % (for eight emotions) to 87.02 % (for two emotions). Using HOC and Cross Correlation (CC) features extracted from 4 channels and an SVM classifier, [86] recognized 3 emotions with accuracies varying from 70.00 % to 100.00 % for different subjects. Kwon *et al.* [105] proposed to use as features the power differences between left and right hemispheres in alpha and gamma bands derived from 12 channels and Adaptive Neuro-Fuzzy Inference System (ANFIS) as classifier and achieved an accuracy of 64.78 % differentiating 2 emotions. By using Higher Order Spectra (HOS), zero-skewness-test parameters and linearity-test parameters as features and SVM as classifier, Hosseini *et al.* was able to identify 2 different emotions at an accuracy of 82.00 % with 5 EEG electrodes used. In [85], Petrantonakis explored the HOC features and CC features acquired from 4 channels in an attempt to classify 3 emotions. Reported accuracies varied from 43.71 % to 62.58 % for different subjects when an SVM classifier was used.

In [88], Sohaib *et al.* evaluated the performance of different classifiers, and reported the best accuracy of 56.10 % for 3 emotions, obtained by using SVM as classifier and statistical features sourced from 6 electrodes. Similarly, in another work [106], Quadratic Discriminant Analysis (QDA) and SVM were compared when 6 emotions were to be recognized using HOC features derived from 4 channels. SVM was reported to have better accuracy, 83.30 %, as compared to 62.30 % obtained by using QDA. Wang *et al.* reported a similar finding in their work [84]. Adopting the statistical features and power features,

Wang compared the classification performance between K-NN, SVM and MLP. SVM, reported an accuracy of 66.51 % for identifying 4 emotion classes with 62 channels, was the highest among all. Brown *et al.* [107] employed the power ratio features and band power features derived from 8 channels, and 3 different classifiers (QDA, SVM and K-NN) to evaluate recognition performance for 3 emotion classes. In this study, K-NN was reported to give the best accuracy, varying from 50.00 % to 64.00 % for different subjects. Frantzidis [108] exploited the Event Related Potential (ERP) and Event Related Oscillation (ERO) properties of EEG and proposed to use the ERP amplitude, ERP latency and ERO amplitude as features. MD classifier and SVM were chosen and compared with each other. Recognizing 4 emotion classes, SVM outperformed MD by 1.80 %, achieving 81.30 % accuracy.

A summary of the existing studies of EEG-based emotion recognition algorithms is given in Table 2-3, and a brief outline of the reviewed studies is given in Table 2-2 in terms of the number of emotions recognized, the number of EEG channels used, and the accuracy achieved. The first half of both tables summarizes the works of subject-dependent emotion recognition algorithms, while the second half reviews the subject-independent algorithms. It must be pointed out that a direct comparison between different algorithms is not appropriate, as the dataset, preprocessing, features, classifier, number of channels and number of recognized emotions are all different between different studies. That is, the accuracies are obtained under different experiment settings. Nevertheless, some conclusions can be drawn without over-generalization. As can be seen from Table 2-3, the accuracies are generally higher when more EEG channels are involved. SVM has been extensively used in these studies [72, 79-88, 106-109]. Moreover, controlled experiments have been conducted in [72, 82, 84, 88, 106, 108] in order to evaluate the performance of different classifiers. [85, 86, 88] have based their experiments on both subject-dependent and subject-

independent settings and established that subject-independent emotion recognition algorithms yield inferior accuracy to subject-dependent algorithms.

2.7 Chapter Conclusion

In this chapter, we cover the related works on EEG-based emotion recognition. We review two human emotion models that are extensively used in emotion-related studies. The self-assessment manikin, a standardized, non-verbal tool for emotion assessment, and the two established affective stimulus libraries are introduced. We then review the EEG signal basics and the correlates between EEG and emotions. We present the systematic overview of a typical EEG-based emotion recognition system and conclude this review chapter with a survey of the state-of-the-arts.

Table 2-2 An outline of existing studies of EEG-based emotion recognition algorithm in terms of the number of recognized emotions, the number of EEG channels used, and the accuracy reported. Upper half: subject-dependent algorithm; lower half: subject-independent algorithm.

Subject-dependent emotion recognition algorithm			
# of recognized emotions	# of channels	Accuracy	Reference
2	4	87.02 %	[87]
	12	51.92 % – 78.85 %	[105]
	32	~ 92.00 %	[110]
	62	93.50 %	[83]
	62	85.85 % – 88.40 %	[111]
3	4	47.11 %	[81]
	4	45.60 % – 94.40 %	[85]
	4	~ 80.00 %	[86]
	4	74.44 %	[87]
	6	36.36 % – 83.33 %	[88]
	16	62.07 %	[80]
	62	72.60 % – 76.08 %	[89]
	64	56.00 % – 63.00 %	[72]
4	3	54.50 % – 67.70 %	[104]
	4	67.08 %	[87]
	14	56.20 % – 57.90 %	[91]
	32	90.72 %	[79]
	32	81.52 % – 82.29 %	[82]
	32	62.59 %	[90]
	32	31.99 % – 54.34 %	[92]
	62	59.84 % – 66.51 %	[84]
5	4	61.67 %	[87]
	62	80.52 % – 83.04 %	[73]
6	4	59.30 %	[87]
7	4	56.24 %	[87]
8	4	53.70 %	[87]

Table 2-2 An outline of existing studies of EEG-based emotion recognition algorithm in terms of the number of recognized emotions, the number of EEG channels used, and the accuracy reported. Upper half: subject-dependent algorithm; lower half: subject-independent algorithm. (cont.)

Subject-independent emotion recognition algorithm			
# of recognized emotions	# of channels	Accuracy	Reference
2	21	50.00 % – 71.20 %	[74]
	32	~ 52.00 %	[110]
	32	65.13 % – 65.33 %	[94]
3	4	43.71 % – 62.58 %	[85]
	4	31.03 % – 57.76 %	[86]
	6	47.78 % – 56.10 %	[88]
	62	56.82 % – 77.96 %	[93]
	62	56.73 % – 76.31 %	[96]
	62	45.19 % – 77.88 %	[98]
	62	57.29 % – 80.46 %	[97]
4	32	57.30 % – 61.80 %	[95]
6	4	62.30 % – 83.30 %	[106]

Table 2-3 A summary of review of EEG-based emotion recognition algorithms. Upper half: subject-dependent algorithms; lower half: subject-independent algorithms.

Subject-dependent EEG-based emotion recognition							
Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Ishino *et al.* [104]	2003	Self-collected EEG data on 1 subject. Music, television, video and a puzzle game as stimuli	Alpha, beta, gamma power spectrums, mean and variance	NN	3	4 (Joy, anger, sorrow, relaxation)	54.50 % – 67.70 %
Schaaff [80]	2008	Self-collected EEG data on 23 subjects. IAPS as stimuli	Absolute values of STFT amplitude.	SVM	16	3 (Pleasant, neutral, unpleasant)	62.07 %

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Lin *et al.* [79]	2009	Self-collected EEG data on 26 subjects. Pre-labeled emotional music as stimuli	Power differences at symmetric electrode pairs	SVM	32	4 (Joy, anger, sadness, pleasure)	90.72 %
Schaaff *et al.* [81]	2009	Self-collected EEG data on 5 subjects. IAPS as stimuli.	Peak alpha frequency, alpha power, cross-correlation, 6 statistical features	SVM	4	3 (Pleasant, neutral, unpleasant)	47.11 %
Chanel *et al.* [72]	2009	Self-collected EEG data on 10 subjects. Subject elicited (i.e., recall of past emotional life episodes)	Mutual information, power features extracted by STFT	LDA, SVM, RVM	64	3 arousal levels (Positively excited, negatively excited, calm neutral)	LDA: 56.00 % SVM: 63.00 % RVM: 58.00 %

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Li *et al.* [83]	2009	Self-collected EEG data on 10 subjects. Smiley and crying face pictures as stimuli.	CSP on gamma band	SVM	62	2 (Happy, sad)	93.50 %
Lin *et al.* [82]	2010	Self-collected EEG data on 26 subjects. Pre-labeled emotional music as stimuli	Power of each electrode, power difference/division at symmetric electrode pairs, power of symmetric electrodes	MLP, SVM	32	4 (Joy, anger, sadness, pleasure)	SVM: 82.29 % MLP: 81.52 %
Murugappan *et al.* [73]	2010	Self-collected EEG data on 20 subjects. Video as stimuli	Statistical features from theta, delta, alpha, beta, gamma bands	K-NN, LDA	62	5 (Happy, surprise, fear, disgust, neutral)	K-NN: 83.04 % LDA: 80.52 %

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Wang *et al.* [84]	2011	Self-collected EEG data on 5 subjects. Oscar winning film clips as stimuli.	Statistical features, logarithmic power spectrum in delta, theta, alpha, beta, gamma bands	K-NN, SVM, MLP	62	4 (Joy, relax, sad, fear)	K-NN: 59.84 % SVM: 66.51 % MLP: 63.07 %
Petrantonakis [85]	2011	Self-collected EEG data on 16 subjects. IAPS as stimuli.	HOC, CC	SVM	4	3 (6 emotion sets: LA, HA, relax; LV, HV, relax; LALV, HALV relax; LAHV, HAHV relax; LALV, LAHV relax; HALV, HAHV relax)	45.60 % – 94.40%

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Petrantonakis *et al.* [86]	2012	Self-collected EEG data on 16 subjects. IAPS as stimuli.	HOC, CC	SVM	4	3 (6 emotion sets: LA, HA, relax; LV, HV relax; LALV, HALV relax; LAHV, HAHV relax; HALV, HAHV relax)	Around 80.00 %
Liu *et al.* [87]	2013	DEAP Database and self-collected EEG data on 14 subjects, IADS and IAPS as stimuli.	FD, statistical features, HOC	SVM	4	2 – 8 (Protected, satisfied, surprise, happy, sad, unconcerned, frightened, angry)	53.70 % (for eight emotions) – 87.02 % (for two emotions)
Kwon *el at.* [105]	2013	Self-collected EEG data on 6 subjects. Video clips as stimuli.	power difference between left and right hemispheres in alpha and gamma band	ANFIS	12	2 (Positive emotion, negative emotion)	51.92 % – 78.85 %

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Sohaib *et al.* [88]	2013	Self-collected EEG data on 15 subjects. IAPS as stimuli.	minimum value, maximum value, mean value and standard deviation	K-NN, SVM, ANN, BN, RT	6	3 (2 emotion sets: positive arousal, negative arousal, neutral; positive valence, negative valence, neutral)	K-NN: 54.54 % – 83.33 % RT: 36.36 % – 54.54 % BN: 36.36 % – 72.72 % SVM: 45.45 % – 50.00 % ANN: 45.45 % – 50.00 %
Zheng *et al.* [111]	2014	Self-collected EEG data on 6 subjects. Movie clips as stimuli.	DE	SVM, DBN, DBN-HMM	62	2 (Positive emotion, negative emotion)	SVM: 85.85 % DBN: 88.40 % DBN-HMM: 87.62 %
Zheng *et al.* [89]	2015	Self-collected EEG data on 15 subjects. Movie clips as stimuli.	DE	SVM, DBN, LR, K-NN	62	3 (Positive emotion, neutral, negative emotion)	SVM: 83.99 % DBN: 86.08 % LR: 82.70 % K-NN: 72.60 %

Subject-dependent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Zhang *et al.* [90]	2016	DEAP, selected 16 out of 32 subjects	Spectral band power (theta, alpha, beta, gamma)	SVM	32	4 (Joy, fear, sadness, relaxation)	Use selected 30 channels: 62.59 %
Mehmood *et al.* [91]	2016	Self-collected, 21 subjects, IAPS as stimuli.	ERP-based LPP, statistics, spectral band power	SVM, K-NN	14	4 (Happy, calm, sad, scared)	SVM: 57.90 % K-NN: 56.20 %
Zheng *et al.* [92]	2017	DEAP	DE	SVM	32	4 (LALV, LAHV, HALV, HAHV)	31.99 % – 54.34 %
Zoubi *et al.* [110]	2018	DEAP	Raw EEG signals	LSM, ANN, SVM, K-NN, LDA, DT	32	2 (HA and LA, or HV and LV)	Arousal: ~93.00 % Valence: ~92.00 %

Subject-independent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Petrantonakis *et al.* [106]	2010	Self-collected EEG data on 16 subjects. Ekman's pictures as stimuli.	HOC	QDA, SVM	4	6 (Happiness, surprise, anger, fear, disgust, sadness)	QDA: 62.30 % SVM: 83.30 %
Petrantonakis *et al.* [85]	2011	Self-collected EEG data on 16 subjects. IAPS as stimuli.	HOC, CC	SVM	4	3 (6 emotion sets: LA, HA, relax; LV, HV, relax; LALV, HALV relax; LAHV, HAHV relax; LALV, LAHV relax; HALV, HAHV relax)	43.71 % – 62.58%

Subject-independent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Petrantonakis *et al.* [86]	2012	Sefl-collected EEG data on 16 subjects. IAPS as stimuli.	HOC, CC	SVM	4	3 (6 emotion sets: LA, HA, relax; LV, HV relax; LALV, HALV relax; LAHV, HAHV relax; HALV, HAHV relax)	With selected EEG data: 43.71 % – 82.91 %. With all EEG data: 31.03 % – 57.76 %.
Sohaib *et al.* [88]	2013	Self-collected EEG data on 15 subjects. IAPS as stimuli.	minimum value, maximum value, mean value and standard deviation	K-NN, SVM, ANN, BN, RT	6	3 (2 emotion sets: positive arousal, negative arousal, neutral; positive valence, negative valence, neutral)	K-NN: 52.44 % RT: 52.44 % BN: 52.44 % SVM: 56.10 % ANN: 47.78 %
Zheng *et al.* [93]	2015	Self-collected EEG data on 15 subjects. Movie clips as stimuli.	DE	SVM	62	3 (Positive emotion, neutral, negative emotion)	56.82 % – 77.96 %

Subject-independent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Candra *et al.* [94]	2015	DEAP	Wavelet entropy	SVM	32	2 (high/low arousal; high/low valence)	Arousal: 65.33 % Valence: 65.13 %
Candra *et al.* [95]	2015	DEAP	Wavelet energy, wavelet entropy	SVM	32	4 (Angry, happy, sad, relaxed)	Wavelet energy: 57.30 % – 59.60 % Wavelet Entropy: 59.00 % – 61.80 %
Bozhkov *et al.* [74]	2015	Self-collected EEG data on 26 subjects. IAPS as stimuli.	ERP amplitude and latency	ANN, LR, LDA, K-NN, NB, SVM, DT	21	2 (high/low valence)	ANN: 71.20 % LR: 67.31 % LDA: 71.20 % K-NN: 59.60 % NB: 69.20 % SVM: 50.00 % DT: 69.20 %

Subject-independent EEG-based emotion recognition

Author	Year	Dataset	Feature	Classifier	# of channels	# of recognized emotions	Accuracy
Zheng *et al.* [96]	2016	Self-collected EEG data on 15 subjects. Movie clips as stimuli.	DE	SVM	62	3 (Positive emotion, neutral, negative emotion)	56.73 % – 76.31 %
Chai *et al.* [98]	2016	SEED, 15 subjects	DE	SVM, LR	62	3 (Positive emotion, neutral, negative)	45.19 % – 77.88 %
Chai *et al.* [97]	2017	SEED, 15 subjects	DE	SVM, LR	62	3 (Positive emotion, neutral, negative)	57.29 % – 80.46 %
Zoubi *et al.* [110]	2018	DEAP	Raw EEG signals	LSM, ANN, SVM, K-NN, LDA, DT	32	2 (HA and LA, or HV and LV)	Arousal: ~56.00 % Valence: ~52.00 %

Abbreviation

ANFIS: Adaptive Neuro-Fuzzy Inference System; **ANN**: Artificial Neural Network; **CC**: Cross Correlation; **CSP**: Common Spatial Pattern; **DE**: Differential Entropy; **DEAP**: Database for Emotion Analysis using Physiological signals; **DT**: Decision Tree; **ERP**: Event Related Potential; **FD**: Fractal Dimension; **HOC**: Higher Order Crossing; **IADS**: International Affective Digitized Sounds; **IAPS**: International Affective Picture System; **K-NN**: K-Nearest Neighbor; **LDA**: Linear Discriminant Analysis; **LPP**: Late Positive Potential; **LR**: Logistic Regression; **MLP**: Multi-Layer Perceptron; **NB**: Naïve Bayes; **NN**: Neural Network; **QDA**: Quadratic Discriminant Analysis; **RT**: Regression Tree; **RVM**: Relevance Vector Machine; **SEED**: Shanghai jiaotong university Emotion EEG Data.; **SFTF**: Short-term Fourier Transform; **SVM**: Support Vector Machine.

Chapter 3 Stable Feature Selection for EEG-based Emotion Recognition

In Chapter 3, we address our first research question outlined in Section 1.3: How can we reduce the need of frequent re-calibrations of the classifier on the same subject? We begin by stating the problem of a current EEG-based emotion recognition system in Section 3.1. We then introduce the methodologies and our proposed stable feature selection algorithms in Section 3.2. Experiments to validate our proposed methods are documented in Section 3.3. Extensive comparisons and analyses are discussed in Section 3.4.

3.1 Problem Statement

The state-of-the-art EEG-based emotion recognition algorithms tailor the classifier to each individual user, requiring a calibration session before the subject starts to use the system. However, due to the volatile affective neural patterns, frequent re-calibrations are needed during use to maintain satisfactory recognition accuracy. A great amount of the existing studies [72, 73, 79-84, 87, 104, 105] investigate and report only the short-term performance of an affective BCI. In these studies, affective EEG data are collected within a short period of time, usually in one single day within one experiment session. K-fold cross validation are carried out to assess the system performance. The recognition accuracy assessed in this way is over-optimistic and can hardly represent the system performance in the long run, especially when no re-calibration is allowed during use. On the other hand, there is little study of the long-term affective BCI performance. We devote this chapter to the investigation of affective BCI performance over a long course of time. As the (re)calibration process may be time-consuming, tedious and laborious, we are motivated to alleviate the burden of frequent re-calibrations on the user of interest. Ideally, a stable affective EEG feature should give consistent measurement of the same emotion on the same subject over a long course of time. We hypothesize that unstable features may worsen the recognition performance of the BCI in the long run. By using stable EEG features, recognition accuracy may be improved. In this chapter, we introduce an ANOVA-based method to quantify the stability score of the state-of-the-art affective EEG features. We then propose a stable feature selection algorithm to choose the optimal set of stable features that maximize the recognition accuracy of the system in the long run.

This chapter is organized as follows. Section 3.2 explains the methodologies. Section 3.3 documents the experiments. Section 3.4 presents the results with discussions. Section 3.5 concludes this chapter.

3.2 Methods

In this section, we describe the approaches to our proposed feature selection algorithm. We firstly review EEG feature extraction methods in Section 3.2.1. Then, we introduce an ANOVA-based stability measurement model called Intra-class Correlation Coefficient (ICC) in Section 3.2.2. Our proposed feature selection algorithm is presented in Section 3.2.3.

3.2.1 Feature Extraction

3.2.1.1 Fractal Dimension

Let $\boldsymbol{x}$ denote a column vector of n EEG time series samples (raw signals) from one channel. Construct k new time series by re-sampling $\boldsymbol{x}$ as follows.

$$\boldsymbol{x}_k^m = \left[\boldsymbol{x}(m), \boldsymbol{x}(m+k), \ldots, \boldsymbol{x}\left(m + \left[\frac{n-m}{k}\right]k\right)\right]^\top, m = 1, 2, \ldots, k, \tag{3-1}$$

where $[\cdot]$ denotes the floor function, m the initial time series sample and k the interval. For example, when $n = 100$ and $k = 3$, we construct three-time series as follows.

$\boldsymbol{x}_3^1 = [\boldsymbol{x}(1), \boldsymbol{x}(4), \boldsymbol{x}(7), \ldots, \boldsymbol{x}(97), \boldsymbol{x}(100)]^\top$,

$\boldsymbol{x}_3^2 = [\boldsymbol{x}(2), \boldsymbol{x}(5), \boldsymbol{x}(8), \ldots\ \boldsymbol{x}(98)]^\top$,

$\boldsymbol{x}_3^3 = [\boldsymbol{x}(3), \boldsymbol{x}(6), \boldsymbol{x}(9), \ldots, \boldsymbol{x}(99)]^\top$.

We compute the length of the curve for each new series as follows.

$$l_k^m = \frac{1}{k}\left\{\left(\sum_{i=1}^{\left[\frac{n-m}{k}\right]} |\boldsymbol{x}(m+ik) - \boldsymbol{x}(m+(i-1)k)|\right)\right\}\left(\frac{n-1}{\left[\frac{n-m}{k}\right]k}\right). \tag{3-2}$$

Let l_k denote the mean of l_k^m for $m = 1, 2, \ldots k$, the fractal dimension of time series $\boldsymbol{x}$ is computed as [113]

$$FD = -\lim_{k\to\infty} \frac{\log(l_k)}{\log(k)}. \tag{3-3}$$

Apparently, in numerical evaluation, it is not possible for k to be infinite. It has proven [42, 114] that the computed fractal value approximates the true, theoretical fractal value reasonably well given a reasonably large k. Based on the study in [42], $k = 32$ yields a good balance between accuracy and computational resources required. In this study, we follow the same parameter setting.

3.2.1.2 Statistics

A set of six statistical features were adopted in [115] for EEG-based emotion recognition, which, in combination with fractal dimension feature, have been demonstrated to improve classification accuracy [115]. Six statistical features are computed as follows.

Mean of the raw signals:

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}(i). \tag{3-4}$$

Standard deviation of the raw signals:

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}(i) - \mu_x)^2}. \tag{3-5}$$

Mean of the absolute values of first order difference of the raw signals:

$$\delta_x = \frac{1}{n-1}\sum_{i=1}^{n-1}|\boldsymbol{x}(i+1) - \boldsymbol{x}(i)|. \tag{3-6}$$

Mean of the absolute values of the first order difference of the normalized signals:

$$\tilde{\delta}_x = \frac{1}{n-1}\sum_{i=1}^{n-1}|\tilde{\boldsymbol{x}}(i+1) - \tilde{\boldsymbol{x}}(i)| = \frac{\delta_x}{\sigma_x}. \tag{3-7}$$

Mean of the absolute values of second order difference of the raw signals:

$$\gamma_x = \frac{1}{n-2}\sum_{i=1}^{n-2}|\boldsymbol{x}(i+2) - \boldsymbol{x}(\mathrm{i})|. \tag{3-8}$$

Mean of the absolute values of second order difference of the normalized signals:

$$\tilde{\gamma}_x = \frac{1}{n-2}\sum_{i=1}^{n-2}|\tilde{\boldsymbol{x}}(i+2) - \tilde{\boldsymbol{x}}(i)| = \frac{\gamma_x}{\sigma_x}. \tag{3-9}$$

In (3-4) – (3-9), $\tilde{\boldsymbol{x}}$ denotes the normalized (zero mean, unit variance) signals, i.e., $\tilde{\boldsymbol{x}} = \frac{\boldsymbol{x}-\mu_x}{\sigma_x}$.

3.2.1.3 Spectral Band Power

Spectral band power, or simply "power", is one of the most extensively used features in EEG-related research [79, 82, 84, 104, 105]. In EEG study, there is common agreement on partitioning the EEG power spectrum into several sub-bands (though the frequency range may slightly differ from case to case, see Section 2.4.2): alpha band, theta band, beta band etc. In our study, the EEG power features from theta band (4 – 8 Hz), alpha band (8 – 12 Hz), and beta band (12 – 30 Hz) are computed.

The power features are obtained by first computing the Fourier transform on the EEG signals. The discrete Fourier transform transforms a time-series $\boldsymbol{x} = [\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(N)]^\mathsf{T}$ to another series $\boldsymbol{s} = [\boldsymbol{s}(1), \boldsymbol{s}(2), \ldots, \boldsymbol{s}(N)]^\mathsf{T}$ in frequency domain. $\boldsymbol{s}$ is computed as

$$\boldsymbol{s}(k) = \sum_{n=0}^{N-1} \boldsymbol{x}(n) e^{-\frac{j2\pi kn}{N}}, \tag{3-10}$$

where N is the number of sampling points. Then, the power spectrum density is computed as

$$\hat{\boldsymbol{s}}(k) = \frac{1}{N} |\boldsymbol{s}(k)|^2. \tag{3-11}$$

Lastly, the spectral band power features are computed by averaging the power spectrum density $\hat{\boldsymbol{s}}(k)$ over the targeted sub-band. E.g., the alpha band power is computed by averaging $\hat{\boldsymbol{s}}(k)$ over 8 – 12 Hz.

3.2.1.4 Higher Order Crossing

Higher Order Crossings (HOC) was proposed in [116] to capture the oscillatory pattern of EEG, and used in [85-87, 106, 115] as features to recognize human emotion from EEG signals. The HOC is computed by first zero-meaning the time-series $\boldsymbol{x}$ as

$$\boldsymbol{z}(i) = \boldsymbol{x}(i) - \mu_x, \tag{3-12}$$

where $\boldsymbol{z}$ is the zero-meaned series of $\boldsymbol{x}$ and μ_x the mean of $\boldsymbol{x}$ computed as per (3-4). Then, a sequence of filter ∇ is successively applied to $\boldsymbol{z}$, where ∇ is the backward difference operator, $\nabla \equiv \boldsymbol{z}(i) - \boldsymbol{z}(i-1)$. Denote the kth-order filtered sequence of $\boldsymbol{z}$ as $\boldsymbol{\xi}_k(\boldsymbol{z})$, $\boldsymbol{\xi}_k(\boldsymbol{z})$ is iteratively applying ∇ on $\boldsymbol{z}$, as

$$\boldsymbol{\xi}_k(\boldsymbol{z}) = \nabla^{k-1}\boldsymbol{z},$$

$$\nabla^0 \boldsymbol{z} = \boldsymbol{z}. \tag{3-13}$$

Then, as its name suggests, the feature consists in counting the number of zero-crossing, which is equivalent to the times of sign changes, in sequence $\boldsymbol{\xi}_k(\boldsymbol{z})$. We follow [115] and compute the HOC feature of order $k = 1, 2, 3, \ldots 36$.

3.2.1.5 Signal Energy

The signal energy is the sum of squared amplitude of the time-series signal [117], computed as

$$\varepsilon = \sum_i |\boldsymbol{x}(i)|^2. \tag{3-14}$$

3.2.1.6 Hjorth Feature

Hjorth [118] proposed three features of a time-series, which have been used as affective EEG features in [119, 120].

Activity:

$$a(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}(i) - \mu_x)^2, \tag{3-15}$$

where μ_x is the mean of $\boldsymbol{x}$ computed as per (3-6).

Mobility:

$$m(\boldsymbol{x}) = \sqrt{\frac{\mathrm{var}(\dot{\boldsymbol{x}})}{\mathrm{var}(\boldsymbol{x})}}, \tag{3-16}$$

where $\dot{\boldsymbol{x}}$ is the time derivative of the time-series $\boldsymbol{x}$, and $\mathrm{var}(\cdot)$ is the variance operator.

Complexity:

$$c(\boldsymbol{x}) = \frac{m(\dot{\boldsymbol{x}})}{m(\boldsymbol{x})}, \tag{3-17}$$

which is the mobility of the time derivative of $\boldsymbol{x}$ over the mobility of $\boldsymbol{x}$.

3.2.2 Feature Stability Measurement

The stability issue of EEG features was firstly brought up under medical application environments. A feature must demonstrate high stability in order to be accepted for clinical use. A stable feature should exhibit consistency among repeated EEG measurements of the same condition on the same subject. Stability of several EEG features such as band power, coherence, and entropy has been studied. In [121] and [122], 26 subjects were involved in a 10-month experiment. Absolute power feature and relative power feature were reported to have similar stability while coherence was less stable than the former two. Power feature obtained from alpha band is the most stable, followed by theta band, delta band, and beta band. Salinsky [123] recruited 19 subjects and recorded their EEG in closed-eye state in an interval of 12 – 16 weeks. No significant difference was found between the stability of absolute power and relative power. Peak alpha frequency and median frequency were reported to be the most stable. Kondacs [124] investigated spectral power features and coherence features of the resting, closed-eye EEG of 45 subject in 25 – 62 months' interval. Total power from 1.5 – 25 Hz frequency range is found to be the most stable, followed by alpha mean frequency, absolute alpha and beta power, absolute delta power and alpha coherence. Gudmundsson [125] studied the spectral power features, entropy and coherence features. EEG data were from 15 elderly subjects, each recorded 10 sessions within two months. Spectral power parameters were reported to be more stable than entropy, while coherence feature was the least stable. Among the spectral power features, theta band was the most stable, followed by alpha, beta, delta and gamma band. Admittedly, parallels cannot be drawn easily between these studies, as subjects, features, data processing techniques, test-retest interval were all different. However, some common findings can be drawn: absolute power features and relative power features have similar stability performance; power features are more stable than coherence feature.

Table 3-1 Review on EEG feature stability.

Author	Feature	Subjects	Measurement interval	Findings
Gasser *et al.* [121, 122]	Spectral band power, coherence	26 children	10 months	Absolute power ≈ relative power Power > coherence Power: $\alpha > \theta > \delta > \beta$ Coherence: $\alpha > \theta$
Salinsky *et al.* [123]	Spectral band power, peak α frequency	19 adults	12 – 16 weeks	Peak α frequency > Absolute power ≈ relative power
Kondacs *et al.* [124]	Spectral band power, coherence	45 adults	25 – 62 months	α mean frequency > absolute α > $\beta > \delta > \alpha$ coherence
Gudmundsson *et al.* [125]	Spectral band power, entropy, coherence	15 elderlies	10 sessions within 2 months	Spectral band power > entropy > coherence, $\theta > \alpha > \beta > \delta > \gamma$

Note: (1). ≈ means "have similar stability"; (2) > means "more stable than".

Table 3-1 summarizes the stability of EEG features reviewed on existing studies. As we can see, spectral features are mostly studied, while many applicable affective features reviewed in Section 3.2.1 are not yet investigated. In the following sections, we present a systematic study of stability on a broad spectrum of applicable features to fill the gap, as well as the investigation of their performance during the long-term usage of the BCI.

3.2.2.1 Intra-class Correlation Coefficient

The stability of feature parameters was quantified by the Intra-class Correlation Coefficient (ICC). ICC allows for the assessment of similarity in grouped data. It describes how well the data from the same group resemble each other. ICC was often used in EEG stability study [125, 126]. ICC is derived from a one-way ANOVA model and defined as [127]

$$\mathrm{ICC} = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W}, \tag{3-18}$$

where MS_B, MS_W and k denote the mean square error between groups, the mean square error within group, and the number of samples in each group,

respectively. A larger ICC value indicates higher similarity among group data. ICC tends to one when there is absolute agreement among the grouped data, i.e., $MS_W = 0$. A smaller ICC value suggests a lower similarity level. ICC value can drop below zero in the case when MS_W is larger than MS_B, accounting for dissimilarity among the grouped data.

3.2.3 Stable Feature Selection

A stable affective EEG feature should give consistent measurements of the same emotion on the same subject over the course of time, therefore there is the possibility to reduce the need of re-calibration by using the more stable features. To this end, we propose a stable feature selection algorithm based on ICC score ranking [128-131]. The proposed algorithm consists of three steps: ICC assessment, ICC score ranking, and iterative feature selection.

We assess the long-term stability of different EEG features with ICC. Let

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

be the matrix of feature parameters of a specific kind of feature, rows of $\boldsymbol{X}$ correspond to different emotions, and columns of $\boldsymbol{X}$ correspond to different repeated measurements over the course of time. In this example, there are n emotions in consideration, and k repeated measurement per emotion. Intuitively, we want the feature parameters to be consistent when measuring the same emotion repeatedly over the course of time. Therefore, we want the parameters within the same row to be similar to each other. Moreover, we want the parameters measuring different affective states to be discriminative, so that different affective states are distinguishable. Therefore, we want different rows to be dissimilar to each other. The ICC measurement takes both considerations

into account. The ICC is computed as per (3-18), which is based on ANOVA. For clarity, we display $\boldsymbol{X}$ in the ANOVA table as in Table 3-2. In Table 3-2, we refer treatment to different emotions induced by specific affective stimuli. x_{ij} is the feature parameter of the jth measurement of emotion i. $x_{i\cdot}$ is the sum of all measurements of emotion i, $x_{i\cdot} = \sum_{j=1}^{k} x_{ij}$. $\bar{x}_{i\cdot}$ is the average of all measurements of emotion i, $\bar{x}_{i\cdot} = \frac{1}{k}\sum_{j=1}^{k} x_{ij}$. $x_{\cdot\cdot}$ is the sum of all measurements over all emotions, $x_{\cdot\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}$. $\bar{x}_{\cdot\cdot}$ is the average of all measurements over all emotions, $\bar{x}_{\cdot\cdot} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}$.

We can obtain the stability score of each feature by computing the ICCs, thereafter, we rank the feature according to the stability score in descending order. Features with higher ICC are more stable over the course of time, and exhibit better discriminability among different emotions. Our proposed feature selection algorithm consists in iteratively selecting the top stable features and validating the inter-session emotion recognition accuracy. The feature subset that yields the best accuracy is retained.

Table 3-2 The analysis of variance table.

Treatment (emotion)	Measurement				Total	Average
1	x_{11}	x_{12}	$\cdots$	x_{1k}	$x_{1\cdot}$	$\bar{x}_{1\cdot}$
2	x_{21}	x_{22}	$\cdots$	x_{2k}	$x_{2\cdot}$	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$x_{n\cdot}$	$\bar{x}_{n\cdot}$
					$x_{\cdot\cdot}$	$\bar{x}_{\cdot\cdot}$

Source of variance	Sum of squares	Degree of freedom	Mean square
Between treatment	$SS_B = k\sum_{i=1}^{n}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$	$n-1$	$MS_B = SS_B/(n-1)$
Within treatment	$SS_W = SS_T - SS_B$	$nk-n$	$MS_W = SS_E/(nk-n)$
Total	$SS_T = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(x_{ij} - \bar{x}_{\cdot\cdot}\right)^2$	$nk-1$	

3.3 Experiments

We design and carry out several experiments to validate the effectiveness of our proposed stable feature selection algorithm. In Section 3.3.1, we explain the EEG data collection process. Based on our dataset, we carry out three simulations of affective BCI under different paradigms, covered in Section 3.3.2, 3.3.3 and 3.3.4, respectively.

3.3.1 Data Collection

The stability of affective EEG features is of our interest of investigation. In contrast to existing affective EEG benchmark dataset such as the DEAP dataset [23], which includes a relatively large number of subjects but only one EEG recording session within one day for each subject, we designed and conducted an experiment to collect the affective EEG data from multiple sessions during the course of several days. This preliminary study included six subjects, five males and one female, aged 24 – 28. All subjects reported no history of mental diseases or head injuries. Two sessions were recorded per day for each subject for eight consecutive days, i.e. 16 sessions were recorded for each subject. An Emotiv EEG device [2] (see Figure 3.1), which can be worn

Figure 3.1 Emotiv EEG device [2].

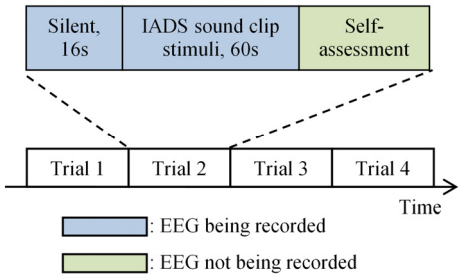


Figure 3.2 Protocol of emotion induction experiment.

for hours without significant discomfort [132], was used to record the EEG data at a sampling rate of 128Hz. Each session consisted of four trials, with each trial corresponding to one induced emotion, i.e., four emotions were elicited in one session, so totally each subject has $4 \times 2 \times 8 = 64$ trials. There are standard affective stimuli libraries such as International Affective Picture System (IAPS) [22] and International Affective Digitized Sounds (IADS) [21]. In our study, the IADS was chosen for the experiment design as during the exposure of the subjects to the audio stimuli, the subjects can keep their eyes closed and hence avoid possible ocular movements which could contaminate the EEG signals. The emotion induction experiment protocol followed work [87]. Sound clips from the same category of the IADS were chosen and appended to make a 76 seconds audio file, with the first 16 seconds silent to calm the subject down. Four audio files were used as stimuli to evoke four different emotions, namely pleasant, happy, angry and frightened. During each session of the experiment only one subject was invited to the lab and was well-instructed about the protocol of the experiment. The subject wore the Emotiv EEG device and a pair of earphones with volume properly adjusted, and he/she was required to sit still with eyes closed and avoided muscle movements as much as possible to reduce possible artifacts from eyeballs movement, teeth clenching, neck movement etc. Following each trial, the subject was required to complete a self-

assessment to describe his emotion (happy, frightened etc.). This self-assessment was used as ground truth to assess the real emotion of the subject. The protocol of this emotion induction experiment is depicted in Figure 3.2.

3.3.2 Simulation 1: With Re-Calibration

In this experiment, we simulate the recognition performance of an affective BCI where re-calibration of the system can be carried out each time before the subject uses the system. Specifically, we evaluate the within-session cross-validation recognition accuracy using the state-of-the-art affective EEG features referenced in Table 3-3.

We base the simulation on the EEG data we collected in Section 3.3.1. Each EEG trial lasts for 76 seconds. We discard both ends of the EEG trial and retain the middle part of the EEG trial for the subsequent processing, based on the assumption that emotions are better elicited in the middle of the trial. The division of the EEG trial is illustrated in Figure 3.3. EEG features are extracted out of the valid segments of the EEG trials on a sliding-windowed basis. The final feature vector is a concatenation of the feature vectors from channel AF3, F7, FC5, T7, and F4 (see Section 2.4.1), as was justified in [42]. The width of the window is 4-second, and the step of the move is 1-second [42]. Thus, each valid segment yields 7 samples.

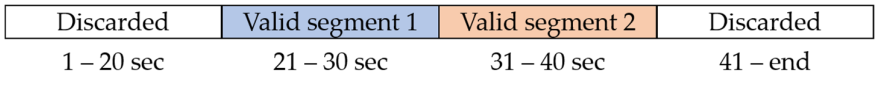


Figure 3.3 Division of the EEG trial. EEG data at both ends are discarded. The middle part is retained and divided into two valid segments of the same length. Only valid segments are used for the subsequent processing.

Table 3-3 Referenced state-of-the-art affective EEG features

Feature (dimension, abbreviation)	Reference
6 statistics (30, STAT)	[84, 87, 115, 136, 137]
36 higher order crossings (180, HOC)	[85-87, 106, 115]
Fractal dimension + 6 statistics + 36 higher order crossings (215, FD1)	[87, 115]
Fractal dimension + 6 statistics (35, FD2)	[87, 115]
3 Hjorth (15, HJORTH)	[118, 119]
Signal energy (5, SE)	[117]
Spectral power of δ, θ, α, and β bands (20, POW)	[81, 84, 104, 138]

In this within-session cross-validation evaluation, the training data and test data are from the EEG trials within the same session. As the time gap between the acquisition of training and test data is minimal, the evaluation can approximate the performance of the BCI where calibration is carried out shortly before use. We use one valid segment as the training data and the other as the test data, and repeat the process until each segment has served as the test data for once. The per-session recognition accuracy is averaged across all possible runs. In this very case, the evaluation is repeated twice per session, which is referred to as a two-fold cross validation. As we recognize four emotions in each session, the training data comprise $7 \times 4 = 28$ samples for four emotions, totally. Likewise, the test data consist of 28 samples for four emotions. We adopt Logistic Regression (LR) [101], Linear Discriminant Analysis (LDA) [68], 1-Nereast Neighbor (1-NN) [68], Linear Support Vector Machine (LinSVM) [68, 75], and Naïve Bayes (NB) [68] as classifiers. The simulation is implemented in MATLAB R2017a, where we use the MATLAB built-in classification toolboxes of the said classifiers with their default hyperparameters. The evaluation is carried out for each of the subjects on a session-by-session basis. The mean classification accuracy over 16 sessions and the standard deviations are displayed in Table 3-4 and Figure 3.4.

3.3.3 Simulation 2: Without Re-Calibration

In this experiment, we simulate the recognition performance where no re-calibration is allowed during the long-term use of the BCI. We evaluate the inter-session leave-one-session-out cross-validation accuracy of the system for this purpose. Recall that in our dataset, we have 16 recording sessions per subject throughout the course of eight days. In this evaluation, we reserve one session as the calibration session whose EEG data are used to train the classifier, and pool together the data from the remaining 15 sessions as test data. We repeat the evaluation until each session has served as calibration session for once. In this very case, the process will be repeated 16 times per subject, and the reported recognition accuracy is the mean accuracy of 16 runs. This evaluation is to simulate the system performance in the long run, since there is a longer time gap between the training session and testing sessions—up to eight days. We adopt the features referenced in Table 3-3 in this simulation, in the same sliding-windowed manner as in Section 3.3.2. We use only the valid segment 1 (see Figure 3.3) of each EEG trial and reserve the valid segment 2 for the testing purpose in Simulation 3 introduced in the following section. The sliding-windowed feature extraction yields 7 samples per valid segment. The training data consist of $7 \times 4 = 28$ samples for four emotions recorded in the same session. The test data comprise $7 \times 4 \times 15 = 420$ samples pooled together from the remaining 15 sessions. The mean classification accuracy over 16 runs and the standard deviations are displayed in Table 3-5.

3.3.4 Simulation 3: Stable Feature Selection

In this experiment, we validate the effect of our proposed stable feature selection algorithm based on the simulation of emotion recognition where no re-calibration is allowed during the long-term use of the BCI. This simulation is

similar to simulation 2, with the focus on the comparison between the state-of-the-art feature set and the stable feature set we propose.

We propose to find the stable features first on a subject-independent basis, then on a subject-dependent basis. The subject-independent evaluation aims at finding a generic stable feature set that can be applied to every subject. To find such generic features, we quantify the long-term feature stability by computing the ICC scores on the training set consisting of the valid segment 1 (see Figure 3.3) from all available trials (16 trials per subject), average the stability scores across all subjects, rank the feature according to the mean stability scores, and retain the optimal subset of features that maximizes the mean recognition accuracy over all subjects when iteratively evaluating the inter-session leave-one-session-out cross-validation accuracy using the top n stable features. The results are shown in Figure 3.5, Figure 3.6 and Table 3-6. After we find the stable features, we evaluate the performance of the stable features on the test set comprising the valid segment 2 from all available trials. The results are shown in Table 3-8, under SISF (Subject-Independent Stable Feature).

The subject-dependent evaluation intends to find subject-specific stable features for each subject. The methodologies are similar to subject-independent evaluation, but without the score averaging operation. We quantify the long-term feature stability by computing the ICC scores on the training set consisting of the valid segment 1 (see Figure 3.3) from all available trials (16 trials per subject), rank the feature according to the stability scores, and retain the optimal subset of features pertinent to the subject in question that maximizes the recognition accuracy when iteratively evaluating the inter-session leave-one-session-out cross-validation accuracy using the top n stable features. The results are shown in Figure 3.7 and Table 3-7. The recognition performance on the test set is shown in Table 3-8, under SDSF (Subject-Dependent Stable Feature).

3.4 Results and Discussions

3.4.1 Simulation 1: With Re-Calibration

Table 3-4 shows the mean accuracy ± standard deviation per subject based on 2-fold cross-validation evaluation, which simulates the use case where re-calibration is allowed each time before a subject uses the BCI. The recognition accuracies vary between subjects, features and classifiers of choice, ranging from 24.89 % (Subject 2, HOC with 1-NN classifier) to 76.90 % (Subject 6, FD1 with LDA classifier). HOC is found to be inferior to other referenced features on all subjects. The best performing feature varies between subjects. For subject 1, 3, 5, and 6, referenced feature set FD1 and FD2 yield better recognition accuracy than other referenced features in most cases. For subject 2, FD2, POW and HJORTH features give similar performance, outperforming other referenced features. For subject 4, STAT, FD1, FD2 and HJORTH features yield comparable results, being better than other referenced features. In general, FD2 performs well on all subjects in this simulation, which may suggest that FD2 is good for the use case where re-calibration is allowed from time to time.

For a four-class classification task, the theoretical chance level for random guess is 25.00 %. However, it is known that the real chance level is dependent on the classifier as well as the number of test samples. For an infinite number of test samples, the real chance level approaches the theoretical value. For a finite number of test samples, the real chance level is computed based on repeated simulations of classifying samples with randomized class label, as is suggested in [133, 134]. We carry out such simulation and present also in Table 3-4 the upper bound of the 95 % confidence interval of the simulated chance level for the best performing feature (in bold) for each classifier. Results show that the best-performing features yield recognition accuracy higher than the upper bound of the chance level, except for subject S3 coupled with LDA, 1-NN, and

Table 3-4 Four-emotion recognition accuracy of Simulation 1, simulating the use case where re-calibration is permitted during the long-term use of the affective BCI. Mean accuracy (%) ± standard deviation (%).

Subject	Feature	Classifier				
		LR	LDA	1-NN	LinSVM	NB
S1	STAT	56.81 ± 10.52	48.66 ± 11.19	55.80 ± **11.17**	56.14 ± 13.02	**52.57** ± 12.35
	HOC	32.25 ± 10.50	37.28 ± 11.29	28.57 ± 11.99	32.37 ± 13.80	32.86 ± 12.71
	FD1	43.08 ± 13.98	**53.01** ± 15.60	34.71 ± 11.92	37.95 ± 14.25	43.10 ± 12.05
	FD2	**57.14** ± **9.93**	52.46 ± **10.73**	**56.36** ± 11.77	**58.71** ± **11.90**	52.23 ± **12.00**
	HJORTH	53.24 ± 11.81	52.34 ± 12.47	50.89 ± 14.03	52.12 ± 13.35	50.00 ± 12.86
	SE	45.54 ± 15.95	43.08 ± 15.62	41.74 ± 15.96	46.54 ± 15.80	42.63 ± 16.57
	POW	48.66 ± 12.21	43.19 ± 12.25	44.08 ± 12.26	47.99 ± 13.77	44.31 ± 14.70
	Chance Lvl	42.79	42.80	42.79	42.70	39.07
S2	STAT	44.75 ± 16.66	33.37 ± 14.41	42.19 ± 17.83	46.21 ± 16.39	44.53 ± 19.74
	HOC	30.25 ± **10.05**	32.59 ± 12.17	24.89 ± **8.08**	27.90 ± **10.57**	25.71 ± **9.73**
	FD1	37.39 ± 12.58	39.62 ± 13.57	30.02 ± 10.00	36.05 ± 10.68	37.38 ± 16.09
	FD2	**46.88** ± 17.25	34.93 ± 12.71	42.97 ± 18.39	45.98 ± 15.63	**46.32** ± 19.30
	HJORTH	46.65 ± 14.30	**42.08** ± 14.24	42.63 ± 15.33	45.65 ± 14.55	42.52 ± 15.36
	SE	40.63 ± 12.67	35.94 ± 12.89	37.39 ± 12.77	41.07 ± 11.88	39.40 ± 16.39
	POW	46.88 ± 17.72	33.37 ± **8.06**	**45.20** ± 15.65	**46.32** ± 16.58	45.65 ± 14.51
	Chance Lvl	42.80	39.41	42.93	42.79	39.07
S3	STAT	43.64 ± 13.89	36.50 ± 12.32	40.85 ± 13.78	39.29 ± 13.54	**37.50** ± 12.89
	HOC	28.46 ± 10.24	30.02 ± **11.34**	30.47 ± **10.86**	28.35 ± **11.81**	29.46 ± 12.27
	FD1	33.59 ± **8.12**	37.72 ± 11.53	35.27 ± 12.10	34.26 ± 12.17	33.93 ± 11.39
	FD2	**45.76** ± 13.01	37.83 ± 13.67	**42.52** ± 12.12	**41.41** ± 12.65	37.50 ± 13.15
	HJORTH	41.41 ± 14.39	37.05 ± 17.12	36.83 ± 13.83	40.74 ± 12.76	35.60 ± 12.24
	SE	41.96 ± 17.57	**42.52** ± 18.55	39.62 ± 13.67	41.07 ± 16.04	36.83 ± 14.30
	POW	36.05 ± 14.70	34.93 ± 13.71	35.38 ± 11.78	34.38 ± 12.74	36.83 ± **11.23**
	Chance Lvl	42.79	42.71	42.80	39.16	39.23
S4	STAT	71.43 ± 14.32	65.40 ± 14.10	69.64 ± 15.55	**71.09** ± 13.97	69.64 ± 13.88
	HOC	43.53 ± **12.20**	52.57 ± 13.35	37.61 ± **12.15**	40.29 ± 13.66	41.96 ± 13.62
	FD1	58.59 ± 13.40	**72.21** ± 12.69	53.57 ± 13.60	49.33 ± **12.96**	69.08 ± **11.91**
	FD2	72.54 ± 14.49	69.20 ± 13.21	69.08 ± 14.49	69.98 ± 14.26	**70.09** ± 12.72
	HJORTH	**72.77** ± 17.82	71.21 ± 13.95	**70.42** ± 16.81	70.09 ± 18.44	64.62 ± 14.83
	SE	59.49 ± 16.23	55.02 ± 17.55	56.36 ± 14.51	53.79 ± 14.46	56.14 ± 12.30
	POW	69.20 ± 15.83	60.60 ± **11.30**	63.62 ± 13.12	62.17 ± 15.65	64.62 ± 13.70
	Chance Lvl	39.36	42.82	39.36	42.93	39.20
S5	STAT	47.92 ± 15.44	42.96 ± 12.48	48.51 ± 15.13	48.21 ± 13.26	**47.82** ± 15.63
	HOC	28.37 ± **10.95**	31.65 ± **12.38**	27.28 ± **8.37**	29.56 ± **9.47**	31.09 ± **9.13**
	FD1	39.58 ± 12.05	44.35 ± 13.27	35.71 ± 11.29	40.87 ± 11.36	41.39 ± 11.05
	FD2	**48.91** ± 15.42	41.27 ± 14.39	**48.91** ± 15.82	**49.21** ± 14.46	47.32 ± 15.17
	HJORTH	47.92 ± 15.67	**45.83** ± 16.94	48.12 ± 16.44	47.42 ± 16.11	47.22 ± 16.31
	SE	41.96 ± 18.90	39.98 ± 17.70	41.96 ± 15.14	41.57 ± 13.43	45.54 ± 14.94
	POW	42.26 ± 18.03	36.41 ± 14.70	43.35 ± 18.26	42.96 ± 15.24	42.56 ± 14.14
	Chance Lvl	42.70	39.38	39.32	42.93	39.20
S6	STAT	73.88 ± 15.29	60.60 ± 12.84	**76.79** ± 16.55	75.22 ± 15.86	73.10 ± 17.32
	HOC	36.61 ± 12.29	49.67 ± **11.50**	34.60 ± **11.01**	36.27 ± **12.05**	36.94 ± **9.00**
	FD1	54.58 ± **11.03**	**76.90** ± 15.86	50.67 ± 13.52	54.46 ± 14.48	67.97 ± 16.64
	FD2	**76.23** ± 15.51	66.29 ± 15.00	76.67 ± 16.02	**76.00** ± 14.71	**73.66** ± 17.70
	HJORTH	72.54 ± 18.78	74.55 ± 14.34	71.99 ± 21.16	74.11 ± 17.42	71.76 ± 19.17
	SE	62.83 ± 20.02	59.60 ± 17.21	60.94 ± 21.38	64.73 ± 19.49	59.93 ± 19.52
	POW	62.72 ± 16.00	59.26 ± 12.39	60.60 ± 18.23	60.60 ± 12.80	62.61 ± 19.77
	Chance Lvl	42.79	42.82	42.70	42.82	39.46

NB classifiers. We assert that the best-performing features perform significantly better than chance level at a 5 % significance level, except for subject S3 with LDA, 1-NN, and NB classifiers.

3.4.2 Simulation 2: Without Re-Calibration

Table 3-5 shows the mean accuracy ± standard deviation per subject based on inter-session leave-one-session-out cross-validation evaluation, which simulates the long-term recognition performance of the BCI when no re-calibration is permitted during use. Notable accuracy drop can be observed, compared to when re-calibration is allowed at each new session. An illustrative comparison can be found in Figure 3.4. This experiment establishes that intra-subject variance of affective feature parameters does exist and does have a negative impact on the recognition performance, though the severity varies from subject to subject. For subject S2 and S3, the recognition performance is severely affected by the variance—the best recognition performance has dropped and fallen within the 95 % confidence interval of the simulated chance level. We therefore assert that subject S2 and S3 are performing at random guess level. For subject S1, S4 and S6, the best performance remains significantly better than the chance level at 5 % significance level (except for S6 with NB classifier), which seems to suffer from the variance problem to a lesser extent. Subject S5 gives mediocre performance. We loosely categorize subject S1, S4 and S6 as good performer, S5 as moderate performer and S2 and S3 as weak performer.

Table 3-5 Four-emotion recognition accuracy of Simulation 2, simulating the use case where no re-calibration is permitted during the long-term use of the affective BCI. Mean accuracy (%) ± standard deviation (%).

Subject	Feature	Classifier				
		LR	LDA	1-NN	LinSVM	NB
S1	STAT	**37.95** ± 5.01	31.65 ± 5.37	32.40 ± 4.24	**34.33** ± 3.46	**31.03** ± 5.91
	HOC	26.55 ± 4.27	28.76 ± 4.39	25.65 ± 2.53	26.32 ± 3.39	26.15 ± **2.58**
	FD1	28.93 ± 3.98	**35.31** ± 4.27	26.73 ± 2.72	29.81 ± **2.25**	30.24 ± 4.37
	FD2	37.38 ± 6.05	34.66 ± 5.47	**32.47** ± 3.42	33.84 ± 3.00	30.98 ± 6.04
	HJORTH	31.77 ± 6.05	32.22 ± 5.98	31.61 ± 4.10	30.70 ± 3.09	28.07 ± 3.99
	SE	28.07 ± **2.83**	27.54 ± **2.92**	27.22 ± 3.05	27.81 ± 3.47	25.79 ± 2.93
	POW	30.49 ± 4.30	31.68 ± 3.83	27.53 ± **1.91**	28.88 ± 3.83	28.29 ± 3.21
	Chance Lvl	29.33	28.86	28.30	29.05	28.34
S2	STAT	24.79 ± **1.77**	23.76 ± 2.69	25.60 ± 2.40	24.11 ± 2.80	25.77 ± 2.40
	HOC	24.78 ± 2.72	25.22 ± 1.98	24.00 ± 1.72	24.64 ± **1.91**	24.40 ± **1.73**
	FD1	24.52 ± 2.27	24.61 ± **1.79**	23.75 ± **1.70**	24.15 ± 2.96	24.46 ± 2.89
	FD2	25.25 ± 2.68	25.19 ± 3.02	25.57 ± 2.60	24.39 ± 2.55	25.52 ± 2.57
	HJORTH	25.85 ± 3.33	25.04 ± 3.22	25.42 ± 2.68	26.00 ± 3.73	**26.79** ± 4.02
	SE	25.80 ± 3.04	**26.00** ± 2.20	25.86 ± 3.58	25.28 ± 3.60	26.24 ± 2.16
	POW	**28.41** ± 4.25	25.58 ± 2.20	**28.81** ± 3.74	**27.69** ± 3.85	26.64 ± 2.55
	Chance Lvl	29.09	29.79	28.55	28.59	28.85
S3	STAT	25.61 ± **1.65**	**27.23** ± 3.69	25.61 ± **1.65**	**27.23** ± 3.69	24.88 ± 4.24
	HOC	25.51 ± 2.63	25.27 ± 3.29	25.51 ± 2.63	25.27 ± 3.29	24.88 ± **1.46**
	FD1	25.13 ± 2.83	25.80 ± 2.81	25.13 ± 2.83	25.80 ± 2.81	25.61 ± 3.54
	FD2	25.16 ± 2.62	27.07 ± **2.35**	25.16 ± 2.62	27.07 ± **2.35**	25.36 ± 3.93
	HJORTH	27.05 ± 3.84	26.04 ± 3.54	27.05 ± 3.84	26.04 ± 3.54	27.43 ± 4.79
	SE	26.99 ± 2.79	26.96 ± 2.45	26.99 ± 2.79	26.96 ± 2.45	26.88 ± 4.35
	POW	**28.01** ± 3.55	26.19 ± 3.43	**28.01** ± 3.55	26.19 ± 3.43	**30.71** ± 5.08
	Chance Lvl	28.83	28.29	29.25	28.79	29.07
S4	STAT	39.49 ± 6.95	38.44 ± 6.18	41.12 ± 6.14	38.45 ± 7.06	41.95 ± 7.68
	HOC	28.68 ± **4.01**	32.47 ± 5.30	27.34 ± **3.06**	27.46 ± **3.72**	29.52 ± **3.84**
	FD1	33.68 ± 5.58	**43.45** ± **4.59**	31.67 ± 5.01	31.79 ± 5.62	40.67 ± 8.25
	FD2	**39.70** ± 7.10	41.67 ± 6.23	40.89 ± 6.09	**38.69** ± 7.01	**41.96** ± 7.25
	HJORTH	35.19 ± 8.13	37.77 ± 9.39	36.01 ± 8.25	38.63 ± 7.52	38.04 ± 10.03
	SE	38.35 ± 5.97	36.73 ± 6.90	40.55 ± 5.70	35.46 ± 6.32	36.01 ± 9.03
	POW	39.42 ± 6.44	39.21 ± 6.09	**41.47** ± 5.84	37.65 ± 6.06	37.02 ± 6.82
	Chance Lvl	28.30	28.80	29.00	28.57	28.58
S5	STAT	27.00 ± 3.98	26.18 ± 2.78	27.14 ± 3.51	27.92 ± 4.10	28.54 ± 2.85
	HOC	25.68 ± **2.78**	25.58 ± 2.57	25.58 ± 2.87	26.12 ± **2.02**	24.98 ± 2.86
	FD1	25.82 ± 3.01	**28.30** ± **1.95**	26.21 ± **2.85**	28.86 ± 2.39	27.15 ± 3.23
	FD2	27.52 ± 3.88	26.95 ± 2.46	27.03 ± 3.71	28.13 ± 4.22	27.64 ± 2.99
	HJORTH	26.32 ± 3.96	27.33 ± 3.52	27.03 ± 3.95	26.98 ± 2.92	27.14 ± **2.69**
	SE	**27.96** ± 4.37	27.50 ± 3.99	28.66 ± 4.10	**28.97** ± 3.50	**28.61** ± 3.74
	POW	27.63 ± 4.53	27.30 ± 3.09	**28.89** ± 5.30	27.88 ± 3.23	26.82 ± 3.73
	Chance Lvl	27.75	29.03	28.18	28.11	28.10
S6	STAT	30.39 ± 6.24	27.80 ± 6.23	**30.15** ± 5.24	**31.62** ± 5.72	29.09 ± 6.77
	HOC	27.01 ± **3.05**	28.20 ± **3.32**	26.50 ± **2.11**	26.38 ± **2.55**	27.32 ± 3.63
	FD1	28.45 ± 3.67	**32.35** ± 6.95	28.47 ± 3.35	29.88 ± 2.99	29.49 ± 6.31
	FD2	29.61 ± 6.25	28.17 ± 6.38	29.66 ± 5.36	31.13 ± 6.25	**29.73** ± 6.84
	HJORTH	28.18 ± 4.82	28.54 ± 6.25	27.87 ± 3.94	31.19 ± 5.89	27.53 ± 5.14
	SE	28.53 ± 3.84	28.21 ± 4.87	26.77 ± 3.93	29.17 ± 4.93	26.21 ± **3.57**
	POW	**31.49** ± 6.94	28.15 ± 6.17	29.61 ± 4.04	28.63 ± 4.25	27.29 ± 3.91
	Chance Lvl	28.85	29.77	28.84	28.57	29.77

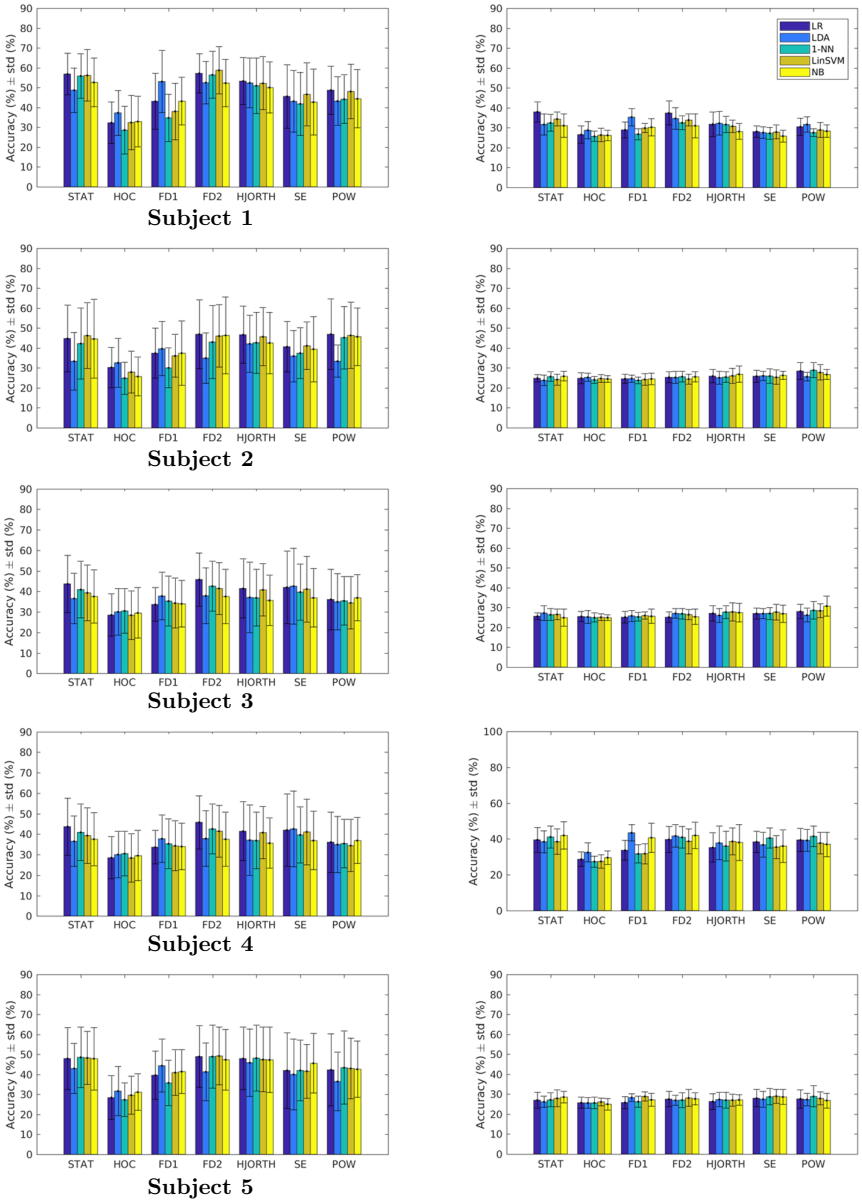
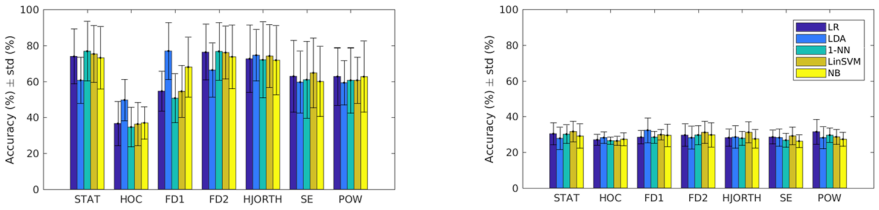


Figure 3.4 Comparison of recognition accuracy between Simulation 1 and Simulation 2. Simulation 1 simulates the use case where re-calibration is allowed from time to time during the long-term use of the BCI. Simulation 2 simulates the use case where re-calibration is not allowed during the long-term use of the BCI. Left: Simulation 1; Right: Simulation 2.



Subject 6

Figure 3.4 Comparison of recognition accuracy between Simulation 1 and Simulation 2. Simulation 1 simulates the use case where re-calibration is allowed from time to time during the long-term use of the BCI. Simulation 2 simulates the use case where re-calibration is not allowed during the long-term use of the BCI. Left: Simulation 1; Right: Simulation 2. (cont.)

3.4.3 Simulation 3: Stable Feature Selection

To improve the long-term recognition accuracy, we propose to use stable features to mitigate the intra-subject variance of the affective feature parameters. Ideally, stable feature should give consistent measurement of the same affective state over the course of time, therefore there is the possibility to mitigate the variance among repeated sessions on different days. We propose a feature selection algorithm that consists in quantifying the long-term stability of features with ICC model, ranking the features according to stability scores and iteratively selecting the topmost stable feature for inclusion in the stable feature subset. We propose to find subject-independent stable features and subject-dependent stable features.

3.4.3.1 Subject-Independent Stable Features

The results of the feature stability ranking are presented in Figure 3.5. The stability scores are ranked in descending order. The feature indices and the respective feature names and scores are given in Table A-2 in the Appendix. It can be seen that only a small portion of the investigated features exhibit desired stability, relatively. Nearly half of the features are unstable, indicated by a

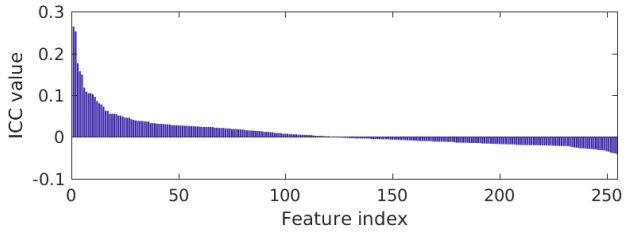
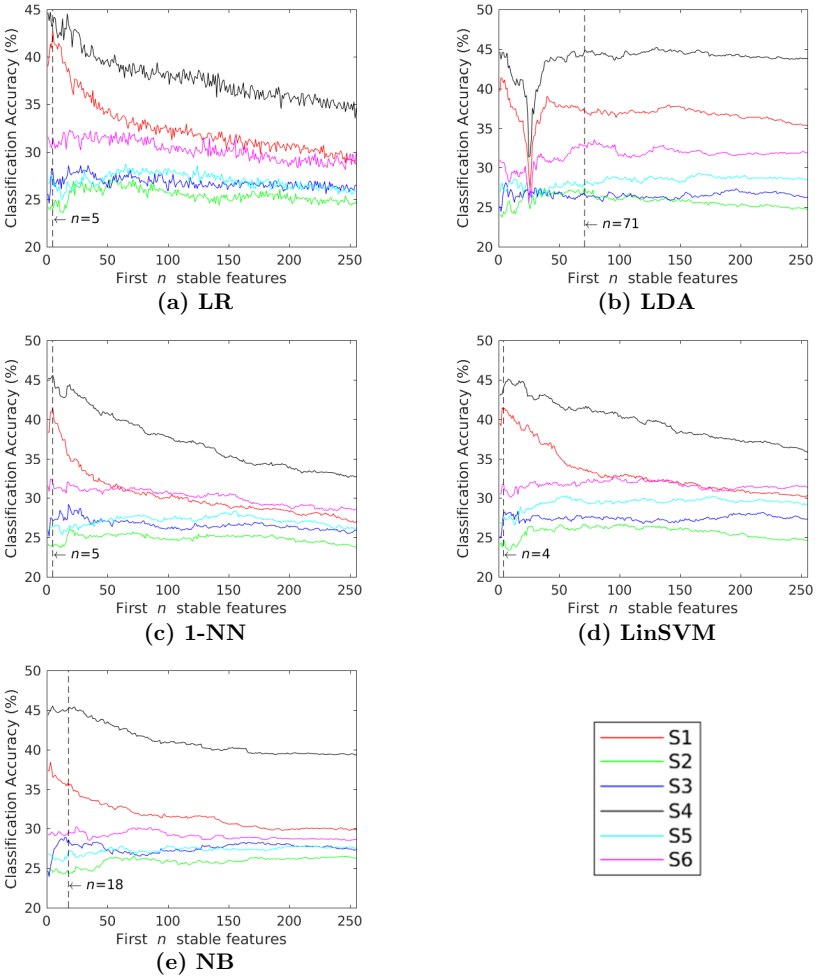


Figure 3.5 Feature ranking in descending order of stability measured by mean ICC scores over all subjects. The feature indices and their respective feature names and exact ICC scores are referred to Table A-2 in the Appendix.



(a) LR **(b) LDA**

(c) 1-NN **(d) LinSVM**

(e) NB

Figure 3.6 The classification accuracy of inter-session leave-one-session-out cross-validation for each subject and each classifier using the top n stable features selected on a subject-independent basis.

negative ICC score which accounts for larger intra-class variance than inter-class variance. Figure 3.6 presents the four-emotion recognition accuracy per subject per classifier using the first n stable features, with n varying from 1 to 255, in the inter-session leave-one-session-out cross-validation evaluation. We see that the curves exhibit similar trend in all subplots except LDA. For subject S1 and S4, it is evident that the accuracy decreases when more unstable features are selected. For the other subjects, the curves are relatively flat, which may suggest insensitivity to stable feature selection. The vertical dashed lines indicate where the mean accuracy over subjects achieves the maximum. As hypothesized, the maximum mean accuracy occurs when a relatively small number of stable features are selected. The number of stable features used and the respective recognition accuracy for each subject and each classifier are presented in Table 3-6. Compared with the best-performing state-of-the-art in Table 3-5, our proposed method has improved the recognition accuracy by 4.90 %, 1.70 %, 8.99 %, 6.88 %, and 4.35 % for subject S1, 3.71 %, 1.37 %, 4.11 %, 5.34 %, and 3.20 % for subject S4, and 0.03 %, 0.40 %, 1.82 %, 0.21 % and -0.09 % for subject S6, for LR, LDA, 1-NN, LinSVM and NB classifier, respectively. It seems that the proposed method works more effectively for good performers than for moderate and weak performers. For weak performer S2 and

Table 3-6 The accuracy (%) ± standard deviation (%) of inter-session leave-one-session-out cross-validation using stable features selected on a subject-independent basis (SISF).

Subject	Classifier (# of stable features)				
	LR (5)	LDA (71)	1-NN (5)	LinSVM (4)	NB (18)
S1	42.84 ± 4.03	37.01 ± 4.10	41.46 ± 4.34	41.21 ± 4.76	35.37 ± 5.93
S2	23.93 ± 2.37	27.47 ± 2.40	23.78 ± 3.29	24.73 ± 2.99	24.32 ± 3.22
S3	28.02 ± 4.11	26.58 ± 2.83	27.14 ± 3.74	28.07 ± 2.82	28.62 ± 4.53
S4	43.41 ± 6.85	44.82 ± 4.81	45.58 ± 5.15	44.03 ± 6.38	45.16 ± 5.65
S5	26.68 ± 4.39	27.88 ± 3.70	26.48 ± 4.27	27.16 ± 4.07	27.16 ± 3.66
S6	31.52 ± 7.60	32.75 ± 7.59	31.96 ± 6.97	31.83 ± 8.12	29.64 ± 8.21

S3, both the proposed method and the referenced state-of-the-art perform at chance level. For moderate performer S5, the state-of-the-art performs slightly better. It may be due to the stable feature ranking being dominated by the good performers, thus not selecting features truly beneficial to other subjects. To clarify this, we present the results of subject-dependent stable feature selection next.

3.4.3.2 Subject-Dependent Stable Feature

Figure 3.7 presents the results of subject-dependent stable feature selection. The bar plot in Figure 3.7 indicates the stability score given in ICC values. The higher the stability score, the less variance the feature exhibits. The stability scores are ranked in descending order. The feature indices and the respective feature names and scores for each subject can be found in Table A-3 in the Appendix. As we can see, the feature stability varies from subject to subject. For subject 1 and 4, the stability scores of the topmost stable features are notably higher than that of the other subjects. Generally, we observe that only a fraction of the features carries positive stability scores. For those with negative stability score, it suggests that the variance of the feature parameters over the course of time is even larger than the variance of the feature parameters between different emotions. Intuitively, these unstable features contribute to the deterioration of long-term recognition performance.

The curves superimposed on the bar plots indicate the inter-session leave-one-session-out cross-validation accuracy for classifying four emotions using only the first n stable features, with n varying from 1 to 255. As we can see, the curves exhibit similar trend among all subjects. The accuracy peaks at a small subset of stable features, then deteriorates when more and more unstable features are included into the feature subset being examined as n increases. For subject 2, 3, 4, 5, and 6, we can clearly see that the accuracy quickly deteriorates

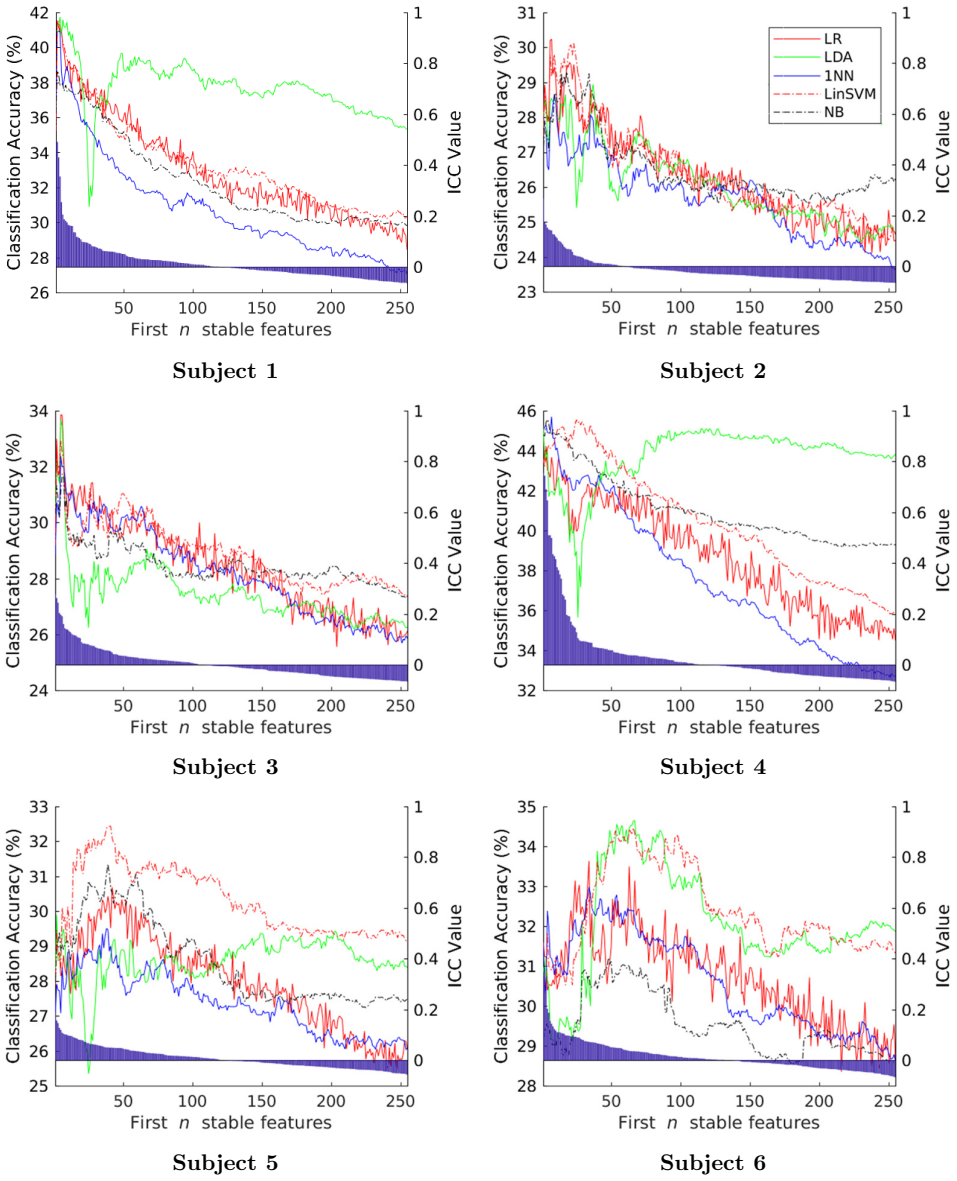


Figure 3.7 ICC scores of each feature and the inter-session leave-one-session-out cross-validation accuracy using the top n stable features, $1 \leq n \leq 255$. The features are ranked by the ICC score in descending order.

as features that carry negative stability scores are included into the feature subset being examined. This experiment shows the advantage of stable features over unstable features when the long-term performance is the utmost concern, and establishes the effectiveness of our proposed feature selection algorithm. The peak recognition accuracy (peak of the accuracy curves in Figure 3.7) and the number of stable features needed to achieve the peak performance is given in Table 3-7. Comparing Table 3-7 with Table 3-5, we can see that stable features selected by our algorithm have outperformed nearly all referenced features. Comparing our features to the best-performing referenced features in Table 3-5 (bold values), our features improve the accuracy by 3.60 %, 6.43 %, 8.47 %, 7.17 %, and 7.65 % for subject 1 for LR, LDA, 1-NN, LinSVM, and NB classifier, respectively. Likewise, the accuracy gains are 1.83 %, 2.94 %, -0.15 %, 2.43 %, and 2.51 % for subject 2; 5.86 %, 6.45 %, 4.36 %, 5.70 % and 1.03 % for subject 3; 5.52 %, 1.71 %, 4.23 %, 6.86 % and 3.61 % for subject 4; 2.72 %, 1.59 %, 0.62 %, 3.46 % and 2.72 % for subject 5; and 2.14 %, 2.29 %, 2.83 %, 2.84 % and 1.48 % for subject 6 for LR, LDA, 1-NN, LinSVM and NB classifier, respectively. The only case when our features perform slightly worse (-0.15 %) than the referenced state-of-the-art features is observed on subject 2 with the 1-NN classifier. For all others, our stable features outperform the referenced state-of-the-art features by 0.62 % – 8.47 %. Moreover, our selected features

Table 3-7 The best mean accuracy of inter-session leave-one-session-out cross-validation evaluation using the top n stable features. Mean accuracy (%) ± standard deviation (%) (# of stable features)

Subject	Classifier				
	LR	LDA	1-NN	LinSVM	NB
S1	41.55 ± 4.31 (2)	41.74 ± 5.44 (4)	40.94 ± 4.34 (3)	41.50 ± 5.02 (3)	38.68 ± 5.76 (2)
S2	30.24 ± 5.14 (7)	28.94 ± 3.31 (37)	28.66 ± 3.53 (9)	30.12 ± 4.16 (23)	29.30 ± 3.86 (18)
S3	33.87 ± 3.55 (5)	33.68 ± 3.61 (5)	32.37 ± 3.92 (5)	32.93 ± 3.80 (5)	31.74 ± 3.86 (6)
S4	45.22 ± 4.57 (1)	45.16 ± 5.01 (1)	45.70 ± 5.07 (7)	45.55 ± 6.75 (25)	45.57 ± 5.03 (4)
S5	30.68 ± 3.43 (42)	29.89 ± 3.64 (2)	29.51 ± 3.68 (38)	32.43 ± 4.77 (39)	31.33 ± 3.89 (39)
S6	33.63 ± 7.99 (34)	34.64 ± 8.87 (66)	32.98 ± 5.44 (34)	34.46 ± 7.67 (66)	31.21 ± 8.63 (48)

have a smaller dimension than the referenced state-of-the-art features, mitigating the burden of classifier training.

In addition, we observe that ICC value is in direct correlation with the long-term recognition performance, which validates our hypothesis that using stable features improves the accuracy. As can be seen from Figure 3.7 (and also Table A-3), the stability scores of the top stable features for subject 1 and subject 4 are notably higher than that for the other subjects. The long-term recognition performance of selected stable features of subject 1 and subject 4 are also notably higher than that of the other subjects. Generally, the higher the stability score, the better the recognition accuracy.

Looking at the subject-dependent feature ranking in Table A-3, we can see that the feature ranking exhibits similar pattern among subject S1, S4 and S6. Statistic features top the stability ranking, together with Hjorth features and some HOCs. This pattern is generally consistent with the subject-independent feature ranking presented in Table A-2. However, for subject S2, S3 and S5, different ranking patterns are observed. HOCs are found to be more stable, mixed with some power features and Hjorth features. This may explain why the subject-independent stable feature set is not so effective for subject S2, S3, and S5 as for subject S1, S4, and S6. We can also observe that the stability scores are higher for subject S1, S4, and S6 than for subject S2, S3, and S5. Thus, when we consider subject-independent stable feature selection, the resultant feature ranking is indeed dominated by these good performers, failing to cater for the other subjects who demand a different set of stable features.

Interestingly, HOC features have been frequently selected given their relatively high stability scores, despite their mediocre performance in Simulation 1 in Table 3-4. It may suggest that HOC features exhibit good stability and are suitable for the use case where the long-term recognition performance shall be

put into consideration. However, it is not the optimal features if re-calibration is allowed before using the BCI from time to time.

3.4.3.3 Comparison on the Test Data

We further examine the performance of the stable features on unseen test data comprising Segment 2 (see Figure 3.3) of all available trials. To simulate the long-term recognition performance, the same inter-session leave-one-session-out cross-validation evaluation scheme is applied. The stable feature set remains the same as was found on the training data on both a subject-independent and a subject-dependent basis. The recognition accuracy using our proposed stable features as well as the referenced state-of-the-art features is presented in Table 3-8. The results are principally consistent with the findings based on training data set. Our stable features outperform the best-performing of the referenced state-of-the-art features by 2.96 %, 3.04 %, 6.16 %, 1.61 %, and 5.23 % for subject S1; -1.43 %, 1.69 %, -2.59 %, 0.70 %, and 0.21 % for subject S2; 0.23 %, 0.71 %, 0.81 %, 0.62 %, and 0.46 % for subject S3; 3.13 %, 0.34 %, 1.29 %, 4.11 %, and 2.35 % for subject S4; 1.92 %, 2.66 %, -5.04 %, 1.22 %, and 0.73 % for subject S5; 1.62 %, 2.62 %, 1.95 %, 2.56 %, and 1.62 % for subject S6, for LR, LDA, 1-NN, LinSVM, and NB classifier, respectively. Our stable features yield the best accuracy on subject S1, S3, S4 and S6 irrespective of the classifier used, where our proposed features outperform the state-of-the-art by 0.23 % – 6.16 %.

As with every machine learning algorithm, our proposed algorithm does come with some limitations. The subject-independent stable feature set is prone to domination by good performers and may not be as effective on other subjects. The effective stable feature set is subject-dependent, to find which requires ample labeled affective EEG data recorded over a long course of time. The acquisition of such data may post a burden to the subjects. Although the stable

Table 3-8 Comparison of inter-session leave-one-session-out cross-validation accuracy on the *test data* between using referenced state-of-the-art feature set and stable feature set selected by our proposed algorithm. Mean accuracy ± standard deviation.

Subject	Feature	Classifier				
		LR	LDA	1-NN	LinSVM	NB
S1	STAT	36.79 ± 6.04	32.40 ± 6.87	33.97 ± 6.18	36.71 ± 6.29	31.98 ± 6.77
	HOC	28.68 ± **3.11**	31.25 ± 5.06	27.95 ± 3.28	29.00 ± 3.95	29.22 ± 4.02
	FD1	30.92 ± 3.58	37.46 ± 6.40	30.33 ± **2.78**	31.61 ± 3.98	32.10 ± 5.43
	FD2	35.61 ± 5.47	34.32 ± 5.36	33.41 ± 6.04	36.70 ± 5.56	31.46 ± 6.51
	HJORTH	31.65 ± 5.86	33.54 ± 6.13	32.77 ± 6.23	31.65 ± 4.63	29.52 ± 4.75
	SE	26.28 ± 3.97	25.46 ± **2.76**	25.04 ± 3.01	25.43 ± **2.68**	24.30 ± **2.66**
	POW	33.41 ± 7.11	31.74 ± 7.06	29.42 ± 5.39	32.19 ± 6.20	29.35 ± 5.61
	Ours (SISF)	**39.75** ± 5.33	37.99 ± 4.19	**40.13** ± 4.56	37.66 ± 3.52	33.76 ± 4.89
	Ours (SDSF)	39.33 ± 6.13	**40.49** ± 5.03	39.97 ± 4.27	**38.32** ± 4.75	**37.32** ± 6.04
S2	STAT	26.80 ± 3.87	25.24 ± 3.36	25.63 ± 4.55	25.70 ± 4.41	27.41 ± 4.10
	HOC	24.51 ± 2.84	24.73 ± 3.11	25.04 ± 2.64	24.97 ± 2.52	25.01 ± **2.33**
	FD1	24.64 ± 3.56	24.93 ± **3.09**	25.25 ± **2.59**	25.40 ± **2.51**	25.28 ± 3.99
	FD2	26.44 ± 4.22	24.91 ± 3.39	25.85 ± 4.45	25.57 ± 4.18	27.02 ± 3.96
	HJORTH	26.62 ± **2.80**	25.67 ± 4.54	25.89 ± 4.02	25.40 ± 3.35	26.53 ± 4.99
	SE	26.61 ± 5.40	25.71 ± 4.89	26.98 ± 3.52	26.71 ± 5.77	26.22 ± 4.99
	POW	**27.95** ± 3.66	25.54 ± 3.63	**29.36** ± 4.15	27.16 ± 3.56	27.05 ± 4.48
	Ours (SISF)	25.76 ± 3.12	26.59 ± 3.56	25.10 ± 3.27	25.10 ± 2.35	24.82 ± 2.47
	Ours (SDSF)	26.52 ± 4.23	**27.40** ± 4.20	26.77 ± 3.62	**27.86** ± 5.03	**27.62** ± 4.13
S3	STAT	26.88 ± 3.97	25.28 ± 3.89	26.58 ± 3.55	26.65 ± 3.29	26.68 ± 4.00
	HOC	25.55 ± 3.87	26.58 ± 3.66	25.28 ± 3.67	25.89 ± 3.27	26.00 ± 2.89
	FD1	25.95 ± 4.43	27.23 ± 3.96	25.89 ± **3.37**	26.58 ± 3.69	26.22 ± 4.51
	FD2	27.50 ± 3.57	25.49 ± 3.21	26.59 ± 3.59	26.82 ± 3.37	26.58 ± 3.69
	HJORTH	26.82 ± 3.15	26.06 ± 3.54	25.65 ± 3.87	27.14 ± **2.65**	27.25 ± 3.67
	SE	26.64 ± **2.93**	27.68 ± 3.11	27.37 ± 4.24	27.89 ± 3.41	27.75 ± 3.68
	POW	28.04 ± 3.14	25.89 ± **2.93**	27.54 ± 3.47	27.01 ± 3.09	26.47 ± 4.20
	Ours (SISF)	26.47 ± 2.43	26.06 ± 2.60	25.94 ± 3.04	26.76 ± 2.21	27.35 ± 2.81
	Ours (SDSF)	**28.27** ± 3.72	**28.39** ± 3.15	**28.35** ± 4.15	**28.51** ± 4.96	**28.21** ± **2.60**
S4	STAT	38.68 ± 5.92	39.15 ± 9.20	41.35 ± 5.38	38.18 ± 7.02	41.80 ± 4.37
	HOC	28.62 ± **3.74**	34.02 ± **5.30**	28.13 ± **3.41**	28.18 ± 3.76	30.82 ± **3.12**
	FD1	35.51 ± 5.57	44.35 ± 5.39	33.14 ± 5.01	33.57 ± 4.93	42.47 ± 3.81
	FD2	40.54 ± 5.89	39.66 ± 7.40	41.13 ± 5.52	37.31 ± 6.30	41.82 ± 4.38
	HJORTH	38.47 ± 5.85	41.10 ± 7.88	38.65 ± 6.84	39.42 ± 6.08	41.13 ± 6.74
	SE	36.98 ± 8.46	40.21 ± 5.50	40.40 ± 6.56	38.75 ± 5.46	39.51 ± 4.51
	POW	38.85 ± 8.02	39.30 ± 6.76	41.71 ± 5.43	37.49 ± **3.16**	36.90 ± 5.12
	Ours (SISF)	43.44 ± 5.90	43.21 ± 7.79	**43.01** ± 5.23	42.57 ± 6.98	**44.82** ± 4.22
	Ours (SDSF)	**43.66** ± 6.09	**44.69** ± 6.77	42.80 ± 5.84	**43.53** ± 7.04	44.51 ± 5.57
S5	STAT	28.38 ± 4.06	25.98 ± 3.67	27.36 ± 2.82	28.03 ± 5.14	27.36 ± 3.00
	HOC	25.90 ± **2.67**	26.37 ± 2.70	25.35 ± 2.64	26.20 ± **2.81**	26.31 ± 3.45
	FD1	26.41 ± 2.87	27.95 ± 3.89	25.44 ± 3.00	28.36 ± 3.67	27.27 ± 3.57
	FD2	27.47 ± 3.49	26.65 ± **2.58**	26.87 ± 2.45	27.85 ± 4.73	27.23 ± 3.47
	HJORTH	26.76 ± 2.84	25.91 ± 3.82	26.17 ± **2.44**	27.22 ± 4.25	27.91 ± 3.65
	SE	28.89 ± 3.40	27.38 ± 2.75	**31.21** ± 4.30	27.88 ± 4.04	28.80 ± 3.63
	POW	27.65 ± 3.94	26.35 ± 3.95	29.33 ± 3.69	27.72 ± 3.25	28.31 ± **2.76**
	Ours (SISF)	26.25 ± 3.45	27.96 ± 4.24	25.54 ± 2.96	26.52 ± 3.12	26.45 ± 3.21
	Ours (SDSF)	**30.81** ± 5.11	**30.61** ± 5.02	26.17 ± 3.24	**29.58** ± 4.55	**29.53** ± 4.39
S6	STAT	31.29 ± 7.76	29.30 ± 5.62	31.29 ± 5.89	29.61 ± 6.42	29.84 ± 6.63
	HOC	27.23 ± 4.30	30.09 ± 4.71	27.43 ± **3.76**	28.54 ± **3.00**	27.50 ± 4.30
	FD1	29.99 ± 5.22	32.57 ± 8.03	28.71 ± 4.81	29.66 ± 4.54	30.60 ± 5.88
	FD2	31.82 ± 7.93	30.48 ± 7.61	30.88 ± 5.55	30.52 ± 5.61	30.31 ± 5.86
	HJORTH	29.64 ± **3.78**	30.31 ± 6.25	31.19 ± 4.65	29.90 ± 5.25	29.94 ± 5.18
	SE	27.49 ± 5.36	28.11 ± **3.95**	27.08 ± 4.28	27.02 ± 3.68	25.22 ± 3.95
	POW	31.92 ± 7.68	27.90 ± 4.95	31.09 ± 7.36	28.20 ± 3.86	27.44 ± **2.42**
	Ours (SISF)	32.40 ± 5.96	33.63 ± 7.52	**33.24** ± 7.53	**33.08** ± 7.40	**32.22** ± 4.22
	Ours (SDSF)	**33.54** ± 6.93	**35.19** ± 10.12	32.77 ± 5.43	32.34 ± 8.95	32.19 ± 5.11

features perform relatively better than the state-of-the-art in the long run, the absolute recognition accuracy is still admittedly low. It remains an open question as to how can we effectively mitigate or even eliminate the need of frequent re-calibrations of the BCI. We have taken the approach to finding the stable components pertinent to affective states over the course of time. It assumes the existence of stationary components and relies on such features to build a static classifier. From the opposite point of view, we could consider a dynamic or incremental classifier to accommodate the nonstationary feature parameters. A dynamic, incremental classifier may suggest a future direction to work on.

3.5 Chapter Conclusions

An EEG-based affective BCI needs frequent re-calibrations as the affective neural patterns are volatile over the course of time even for the same subject, and intra-subject variance exist in the affective feature parameters. The volatility of affective neural patterns presented in the EEG signals has arguably been the major hindrance of voluntary adoption of such an affective interface among healthy users. To maintain satisfactory recognition accuracy, the users need to calibrate the BCI from time to time before they start to use it, or even need to be interrupted and re-calibrate it halfway when they are using it if the usage prolongs. The frequent re-calibrations can have two major impacts on the users. Firstly, it may further lower the users' interest in using such a system. Secondly, during re-calibration, the user need to be presented the affective stimuli to arouse the affective states. But after several times of presentation of the affective stimuli, the user may develop habituation effect [135], which refers to the phenomenon of decreased response to a stimulus after the subject has been repeatedly exposed to it. In other words, repeated stimulus presentation leads to ineffective emotion inducement.

On the other hand, a "plug-and-play" BCI without re-calibration can greatly ease the burden from the user, as the user is free from the time-consuming re-calibration process. The major problem of a BCI without re-calibration is arguably the lower recognition accuracy compared to a re-calibrated BCI. In this chapter, we propose a stable feature selection algorithm [128-131] to select the optimal feature set that maximize the recognition accuracy for the long run of an affective BCI without re-calibration. The proposed method consists in modeling the feature stability by ICC, feature ranking and iterative selection of stable features. We hypothesize that unstable features contribute to the accuracy deterioration when the BCI operates without re-calibration over the course of time, and by using stable features, the recognition accuracy can be improved. We carry out extensive comparison between our stable features and the state-of-the-art features. In Simulation 1, we show the recognition accuracy of an affective BCI using the state-of-the-art features, where the BCI is allowed to be re-calibrated from time to time. In Simulation 2, we simulate the long-term usage of an affective BCI and establish that substantial accuracy deterioration occurs when the BCI operates without re-calibration. By comparing the results of Simulation 1 and Simulation 2, we establish that the decrease in recognition accuracy is inevitable given the current technologies and computation methods. In Simulation 3, we analyze the performance of stable features selected by our proposed method. We demonstrate the accuracy trajectory when we iteratively include features into the selected feature subset. Experimental results show that recognition accuracy peaks at a small subset of stable features, and as more unstable features are included, the recognition accuracy quickly deteriorate. The experiment results validate our hypothesis. Comparisons between our stable features and the state-of-the-art features show that our stable features yield better accuracy than the best-performing of the state-of-the-art by 0.62 % – 8.47 % on the training set, and by 0.23 % – 6.16 % on the test set. We stress that the benefit of using stable features is to mitigate the accuracy decrease to a lesser extent compared to not using stable features.

Chapter 4 Subject-Independent EEG-based Emotion Recognition with Transfer Learning

In Chapter 4, we address our second research question outlined in Section 1.3: How can we improve the recognition accuracy of subject-independent EEG-based emotion recognition algorithm? We begin by stating the problem of subject-independent emotion recognition in Section 4.1, then proceed to introduce in Section 4.2 the two datasets based on which we validate the effectiveness of different transfer learning techniques reviewed in Section 4.3. The experiments are documented in Section 4.4, followed by extensive discussions and analyses of the experiment results in Section 4.5.

4.1 Problem Statement

The best-performing aBCIs are subject-dependent algorithms that adopt machine learning techniques and rely on discriminative features [14, 41]. A subject-dependent aBCI paradigm operates as follows. In a calibration session, affective stimuli targeting specific emotions are presented to the user to induce the desired emotions while recording the EEG signals. A classifier is then trained using the chosen features extracted out of the recorded EEG data and the emotion labels. In a live BCI session that immediately follows the training session, the incoming EEG data are fed to the feature extractor then to the already-trained classifier for real-time emotion classification. Satisfactory classification performance has been reported by many researchers under this paradigm [14]. However, the need to calibrate the classifier also posts a hindrance to the adoption of the BCI system, as the (re-)calibration process can be tedious, tiresome and time-consuming for the subject-of-interest. On the other hand, a subject-independent algorithm constructs the classifier with labeled training data from other subjects, which has the advantage of being "plug-and-play" and eliminates the training/calibration process. The applicational value of a subject-independent algorithm is highly appreciated in this regard. Yet, subject-independent algorithms are known to yield inferior recognition accuracy to subject-dependent algorithms, due to the subject-specific affective neural pattern (see Figure 4.1 for an illustration of how the data distribute differently among different subjects).

In the related field such as motor-imagery BCI, an early attempt to tackle the volatility of the EEG signals was to train the subjects to modulate the EEG signals in a way that complies with the classification rule [139-142]. For example, Wolpaw et. al. [139] proposed to train the subject to manipulate the mu rhythm power and a movement direction was classified by thresholding the mu power amplitude. The thresholding rule was fixed for the subjects, and the subjects

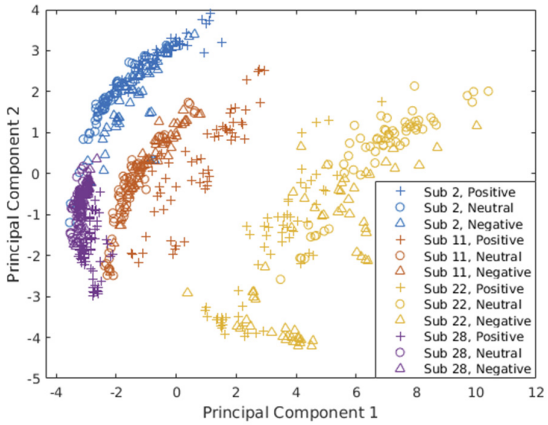


Figure 4.1 Data sample distribution (feature level) from four subjects from DEAP dataset. Original feature vectors are reduced to 2-dimensional by principal component analysis for visualization. The plot shows how differently samples are distributed among different subjects. This is one of the main reasons why a classifier trained on a particular subject does not generalize well to other subjects.

needed to generate control signals in compliance with the classification rule. They reported high classification accuracy, at the expense of prolonged training time—several weeks. Other attempts involve those adopting transfer learning in a BCI setting [93, 96-98, 143-146]. Transfer learning is a machine learning technique that aims to extract common knowledge from one or more source tasks and apply the knowledge to a related target task [147]. Speaking in a BCI context, we can either attempt to find some common feature representations that are invariant across different subjects, or we can try to uncover how the classification rules differ between different subjects. The two methods are denoted as domain adaptation and rule adaptation [148], respectively. Domain adaptation approach has almost exclusively dominated the current BCI-related literature [148]. Krauledat et. al. [143] proposed to find prototypical filters of Common Spatial Pattern (CSP) from multiple recording sessions and apply the said filters to follow-up session without recalibrating the classifier. Fazli et. al. [144] proposed to construct an ensemble of classifiers derived from subject-

specific temporal and spatial filters from 45 subjects, and chose a sparse subset of the ensemble that is predictive for a BCI-naïve user. Kang et. al. [145] developed composite CSP that is a weighted sum of covariance matrices of multiple subjects to exploit the common knowledge shared between the subjects. Lotte et. al. [146] proposed a unifying framework to design regularized CSP that enables subject-to-subject transfer. In aBCI studies, [93, 96-98] explore various domain adaptation methods based on the SEED dataset. In these studies, domain adaptation amounts to finding a domain-invariant space where the inter-subject/inter-session discrepancies of the EEG data are reduced and discriminative features across subjects/sessions are preserved.

Though inter-subject or inter-session transfer and adaptation have been extensively studied in the current literature, the said transfer and adaptation have been restricted within the SEED dataset. That is, the source and target EEG data are from the same dataset in these studies. One question that has not been addressed in the current studies is the efficacy of knowledge transfer and adaptation across different EEG datasets. One could expect that a cross-dataset adaptation sets a more challenging task. Different EEG datasets can be collected using different EEG devices, different experiment protocols, different stimuli etc. These technical differences could add to the discrepancies that are already existing between different subjects/sessions. However, we believe that an ideal, robust BCI should function independently of the device of choice, stimuli used, subjects and experiment context etc. This also makes great practical and applicational sense as it relaxes the constraints in a conventional BCI context. Therefore, in this study, we carry out also a preliminary study to investigate the effectiveness of domain adaptation techniques in a cross-dataset setting, which stands in contrast to existing studies.

Specifically, in this chapter, we first investigate the performance of subject-independent emotion recognition with and without domain adaptation

techniques in a within-dataset leave-one-subject-out cross-validation setting. We hypothesize that each subject constitutes a domain himself/herself, and that EEG data distribute differently across different domains. We apply a recent domain adaptation technique MIDA [149] and compare it to several state-of-the-art domain adaptation methods on DEAP and on SEED datasets. We then propose a cross-dataset emotion recognition scheme to testify the effectiveness of different domain adaptation methods. Under the cross-dataset emotion recognition scheme, the training (source) data are from one dataset and the test (target) data are from the other. Besides the inter-subject variance that is known to exist between different subjects, under a cross-dataset scheme, there also exist technical discrepancies underlying two datasets, hence a more challenging task.

This chapter is organized as follows. Section 4.2 reviews the two datasets we use in this study. Section 4.3 documents data processing methods, including data preparation, feature extraction, and domain adaptation method. Section 4.4 explains the experiment in detail. Section 4.5 analyzes and discusses the experiment results. The chapter is concluded in Section 4.6.

4.2 Datasets

There are a few established EEG datasets for affective states investigation. In this study, we use two of the publicly available datasets, DEAP [23] and SEED [89]. Domain adaptation on SEED has been extensively studied [93, 96-98]. However, little is known about the effectiveness of domain adaptation on DEAP. Moreover, we are also interested in the efficacy of an aBCI in a cross-dataset evaluation setting, especially when two datasets are heterogeneous in many technical aspects. The purpose of cross-dataset evaluation is to attest whether it is possible to maintain satisfactory recognition accuracy when the training data and test data are from different subjects, recorded with different EEG

devices, and have the affective states induced by different stimuli, and whether domain adaptation technique can potentially enhance the performance in a cross-dataset evaluation setting.

The DEAP dataset [23] consists of 32 subjects. Each subject was exposed to 40 one-minute long music video as affective stimuli while having the physiological signals recorded. The resultant dataset comprises 32-channel[2] EEG signals, 4-channel Electrooculography (EOG), 4-channel Electromyography (EMG), respiration, plethysmograph, Galvanic Skin Response (GSR) and body temperature. There are 40 EEG trials recorded per subject, each trial corresponding to one emotion elicited by one music video. Immediately after watching each video, the subject was required to rate their truly-felt emotion assessed from five dimensions: valence (associated with pleasantness level), arousal (associated with excitation level), dominance (associated with control power), liking (associated with preference) and familiarity (associated with the knowledge of the stimulus). The rating ranges from one (weakest) to nine (strongest), except familiarity which rates from one to five. The EEG signals were recorded by Biosemi ActiveTwo devices at a sampling rate of 512 Hz and down sampled to 128 Hz by the dataset authors.

The SEED dataset [89] contains 15 subjects. Movie excerpts were chosen to elicit three emotions: positive, neutral and negative emotions, with five movie excerpts assigned to each emotion. All subject underwent three EEG recording sessions, with an interval of two weeks between two successive recording sessions. Within each session, each subject was exposed to fifteen approximately four-minute-long movie excerpts to induce the desired emotions. The same fifteen movie excerpts were used in all recording sessions. The resultant dataset

[2] The 32 EEG channels include AF3, AF4, C3, C4, CP1, CP2, CP5, CP6, CZ, F3, F4, F7, F8, FC1, FC2, FC5, FC6, FP1, FP2, FZ, O1, O2, OZ, P3, P4, P7, P8, PO3, PO4, PZ, T7 and T8.

contains 15 EEG trials recorded per subject per session, each emotion having 5 trials. Trial #1, 6, 9, 10, and 14 correspond to positive emotions. Trial #2, 5, 8, 11, and 13 target at neutral emotions. Trial #3, 4, 7, 12, and 15 provoke negative emotions. The EEG signals were recorded by a 62-channel[3] ESI NeuroScan device, at a sampling rate of 1000 Hz and down sampled to 200 Hz by the dataset authors. A table summarizing the technical specifications of the two datasets is given in Table 4-1.

4.3 Methods

4.3.1 EEG Data Preparation

As shown in Table 4-1, the two datasets are quite different in every technical aspect. Given that emotions are rated on five numeric scales in DEAP, we discretize and partition the dimensional emotion space in accordance with SEED as follows: emotions are considered positive if valence rating is greater than 7; neutral if valence rating is smaller than 7 and greater than 3; negative if valence rating is smaller than 3. We then look in DEAP for the trial that has the most participants reported to successfully induce positive, neutral and negative emotion, respectively. These trials are: trial #18 for positive emotion, #16 for neutral emotion, and #38 for negative emotion, having 27, 28 and 19 subjects reported that the desired emotion has been induced, respectively.

[3] The 62 EEG channels include AF3, AF4, C1, C2, C3, C4, C5, C6, CB (cerebellum)1, CB2, CP1, CP2, CP3, CP4, CP5, CP6, CPZ, CZ, F1, F2, F3, F4, F5, F6, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCZ, FP1, FP2, FPZ, FT7, FT8, FZ, O1, O2, OZ, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO5, PO6, PO7, PO8, POZ, PZ, T7, T8, TP7 and TP8.

Table 4-1 Technical comparisons between DEAP and SEED.

Item	DEAP [23]	SEED [89]
EEG device	Biosemi ActiveTwo	ESI NeuroScan
# of channels	32 for EEG, 8 for peripheral physiological signals	62 for EEG
Sampling rate	Originally 512 Hz, down-sampled to 128 Hz	Originally 1000 Hz, down-sampled to 200 Hz
# of subjects	32	15
Affective stimuli	Music videos	Chinese movie excerpts
Emotions	Valence, liking, arousal, dominance on a 1 (weakest) to 9 (strongest) scale. Familiarity on 1 to 5 scale.	Positive, neutral, negative
# of recording sessions per subject	1	3
# of trials per session	40	15
Trial length	63 seconds	Approx. 4 minutes

Subjects that commonly reported successful emotion induction with these three trials (#18, #16 and #38) are subjects 2, 5, 10, 11, 12, 13, 14, 15, 19, 22, 24, 26, 28 and 31. The three trials from these fourteen subjects are used for subsequent processing. Each trial lasts for 63 seconds. Since the first 3 seconds are baseline recording without emotion elicitation, we only use the segment from the 4th second to the end. Thus, a valid trial lasts for 60 seconds. For SEED, the trial length varies from 185 seconds to 265 seconds, depending on the length of the affective stimulus used to elicit the desired emotion. We truncate all trials to 185-second-long so as to balance the data of different classes.

4.3.2 Feature Extraction

In this study, we adopt differential entropy (DE) as features for emotion recognition. DE features have been extensively used in the current literature

[93, 96-98] studying the application of transfer learning techniques in EEG-based emotion recognition. Before feature extraction, each EEG trial is divided into multiple 1-second-long segments. Let $\boldsymbol{T}$ denote one EEG segment, $\boldsymbol{T} \in \mathbb{R}^{s \times w}$, where s is the number of channels, $s = 32$ for DEAP or $s = 62$ for SEED, and w is the number of sampling points per channel per segment, $w = 128$ for DEAP or $w = 200$ for SEED. Each valid trial in DEAP lasts for 60 seconds, and thus yields 60 segments per trial. Similarly, each valid trial in SEED yields 185 segments. The DE feature is extracted out of each EEG segment.

4.3.2.1 Differential Entropy

Let $\boldsymbol{t} \in \mathbb{R}^{w}$ denote the time series of EEG signal from one channel, the DE of $\boldsymbol{t}$ is calculated by [150]

$$\mathrm{DE} = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(t-\mu)^2}{2\sigma^2}\right) \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(t-\mu)^2}{2\sigma^2}\right) dt = \frac{1}{2} \log 2\pi e\sigma^2, \qquad (4\text{-}1)$$

where the random variable t follows the Gaussian distribution $N(\mu, \sigma^2)$, and $\boldsymbol{t}$ is the time-series observation of t. The EEG signal, of course, does not follow the Gaussian distribution. It has proven [150] that after $\boldsymbol{t}$ has been band-pass filtered, the time-series of sub-band signal approximately follow the Gaussian distribution. According to [93, 96-98, 150], five sub-bands are defined: delta (1 – 3 Hz), theta (4 – 7 Hz), alpha (8 – 13 Hz), beta (14 – 30 Hz) and gamma (31 – 50 Hz). As such, five DE features can be extracted from $\boldsymbol{t}$. The final feature vector is a concatenation of features from all channels. For DEAP, the final feature vector is of $5 \times 32 = 160$ dimensions, and each trial yields 60 samples. For SEED, the final feature vector is of $5 \times 62 = 310$ dimensions, and each trial yields 185 samples.

4.3.3 Domain Adaptation Method

In the following, we assume that we have a set of labeled data $\boldsymbol{X_s} \in \mathbb{R}^{m \times n_s}$ and a set of unlabeled data $\boldsymbol{X_t} \in \mathbb{R}^{m \times n_t}$, where m is the dimension of the feature, n_s and n_t are the number of samples in the respective set. Let $\boldsymbol{Y_s}$ be the labels associated with $\boldsymbol{X_s}$, we refer to $\mathcal{D}_s = \{(\boldsymbol{X_s}, \boldsymbol{Y_s})\}$ as the source domain, and $\mathcal{D}_t = \{\boldsymbol{X_t}\}$ the target domain. In many use cases, $\boldsymbol{X_s}$ and $\boldsymbol{X_t}$ are differently distributed. That said, domain discrepancies exist between the source and the target domain. Usually, a classifier trained in $\mathcal{D}_s$ can perform rather poorly when directly applied to $\mathcal{D}_t$. The task of domain adaptation is to find a latent, domain-invariant subspace to project $\boldsymbol{X} = [\boldsymbol{X_s}\, \boldsymbol{X_t}] \in \mathbb{R}^{m \times n}$ to be $\boldsymbol{X'} = [\boldsymbol{X'_s}\, \boldsymbol{X'_t}] \in \mathbb{R}^{h \times n}$, where h is the desired dimension of the latent subspace, and $n = n_s + n_t$. In the domain invariant subspace, the discrepancies between $\boldsymbol{X'_s}$ and $\boldsymbol{X'_t}$ have been reduced. Subsequently, we can train a classifier in $\mathcal{D}'_s = \{(\boldsymbol{X'_s}, \boldsymbol{Y_s})\}$ and apply it to $\mathcal{D}'_t = \{\boldsymbol{X'_t}\}$. This is a typical unsupervised transductive transfer learning setting [147].

4.3.3.1 Maximum Independence Domain Adaptation

Maximum Independence Domain Adaptation (MIDA) [149] seeks to maximize the independence between the projected samples and their respective domain features measured by the Hilbert-Schmidt Independence Criterion (HSIC) [151]. Domain feature captures the background information of a specific sample, for example, which domain the sample belongs to. The domain feature $\boldsymbol{d} \in \mathbb{R}^{m_d}$ of a specific sample $\boldsymbol{x} \in \mathbb{R}^m$ is defined using one-hot encoding scheme as $d_i = 1$ if the sample is from subject i, and 0 otherwise, where m_d is the number of subjects considered, d_i the ith element of $\boldsymbol{d}$. In a cross-dataset scheme, $d_i = 1$ if subject i is from DEAP, or $d_{i+14} = 1$ if subject i is from SEED, and 0 otherwise. The first fourteen bits of $\boldsymbol{d}$ are attributed to subjects from DEAP dataset, and the remaining fifteen bits attributed to subjects from SEED dataset. The feature vector is augmented with its domain feature by

concatenation $\hat{\boldsymbol{x}} = [\boldsymbol{x}^\top \boldsymbol{d}^\top]^\top \in \mathbb{R}^{m+m_d}$. By augmenting the feature vector with domain feature, we need not distinguish which domain a specific sample is from, and such information is encoded in the augmented feature vector.

Let $\hat{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{D} \end{bmatrix} \in \mathbb{R}^{(m+m_d)\times n}$ be the matrix of the augmented feature where source data and target data are pooled together, we project $\hat{\boldsymbol{X}}$ to the desired subspace by applying a mapping $\boldsymbol{\phi}$ followed by a linear transformation matrix $\widetilde{\boldsymbol{W}}$ to $\hat{\boldsymbol{X}}$, denoted by $\boldsymbol{X}' = \widetilde{\boldsymbol{W}}^\top \boldsymbol{\phi}(\hat{\boldsymbol{X}})$. Like other kernel dimensionality reduction methods [152, 153], the key idea is to construct $\widetilde{\boldsymbol{W}}$ as a linear combination of all samples in $\boldsymbol{\phi}(\hat{\boldsymbol{X}})$, namely $\widetilde{\boldsymbol{W}} = \boldsymbol{\phi}(\hat{\boldsymbol{X}})\boldsymbol{W}$. Hence, $\boldsymbol{X}' = \boldsymbol{W}^\top \boldsymbol{\phi}(\hat{\boldsymbol{X}})^\top \boldsymbol{\phi}(\hat{\boldsymbol{X}})$. Using the kernel trick, we need not compute $\boldsymbol{\phi}(\hat{\boldsymbol{X}})^\top \boldsymbol{\phi}(\hat{\boldsymbol{X}})$ explicitly in the $\boldsymbol{\phi}$ space, but in the original feature space via a proper kernel function $\mathsf{ker}(\cdot)$. Let $\boldsymbol{K}_{\hat{\boldsymbol{X}}} = \boldsymbol{\phi}(\hat{\boldsymbol{X}})^\top \boldsymbol{\phi}(\hat{\boldsymbol{X}}) \in \mathbb{R}^{n\times n}$ denote the kernel matrix of $\hat{\boldsymbol{X}}$, $\boldsymbol{K}_{\hat{\boldsymbol{X}}} = [k_{ij}]$, where k_{ij} is computed by $k_{ij} = \mathsf{ker}(\hat{\boldsymbol{X}}_{:i}, \hat{\boldsymbol{X}}_{:j})$, where $\hat{\boldsymbol{X}}_{:i}$ is the ith column of $\hat{\boldsymbol{X}}$, and $\mathsf{ker}(\boldsymbol{u}, \boldsymbol{v})$ is a proper kernel function that can take the form of linear function ($\mathsf{ker}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^\top \boldsymbol{v}$), polynomial function ($\mathsf{ker}(\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}^\top \boldsymbol{v} + c)^d$), or radial basis function (RBF, $\mathsf{ker}(\boldsymbol{u}, \boldsymbol{v}) = \exp(-\frac{\|\boldsymbol{u}-\boldsymbol{v}\|}{2\sigma^2})$) etc.

$\boldsymbol{W} \in \mathbb{R}^{n\times h}$ is the actual projection matrix we wish to find, and such matrix should bear the desired property so that after projection, $\boldsymbol{X}'$ is independent of domain feature $\boldsymbol{D}$. Intuitively, when $\boldsymbol{X}'$ is independent of domain features $\boldsymbol{D}$, we cannot distinguish from which domain a specific sample $\boldsymbol{X}'_{:i}$ comes, suggesting that the difference of distribution among different domains is reduced in $\boldsymbol{X}'$. The HSIC [151] is used as a convenient method to quantify the level of independence. $\mathsf{HSIC}(\boldsymbol{X}', \boldsymbol{D}) = 0$ if and only if $\boldsymbol{X}'$ and $\boldsymbol{D}$ are independent [154]. The larger the HSIC value is, the stronger dependence. HSIC has a convenient but biased empirical estimate given by $(n-1)^{-2}\mathrm{tr}(\boldsymbol{K}_{\boldsymbol{X}'}\boldsymbol{H}\boldsymbol{K}_{\boldsymbol{D}}\boldsymbol{H})$ [151], where $\boldsymbol{K}_{\boldsymbol{D}} = \boldsymbol{D}^\top \boldsymbol{D} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{K}_{\boldsymbol{X}'} = (\boldsymbol{W}^\top \boldsymbol{K}_{\hat{\boldsymbol{X}}})^\top (\boldsymbol{W}^\top \boldsymbol{K}_{\hat{\boldsymbol{X}}}) \in \mathbb{R}^{n\times n}$ are the kernel

matrices of $\boldsymbol{X'}$ and $\boldsymbol{D}$, respectively, $\boldsymbol{H} = \boldsymbol{I} - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top \in \mathbb{R}^{n\times n}$ is the centering matrix, and $\mathbf{1}_n$ is an all-one vector of dimension n.

Besides maximizing the independence between the projected samples and the domain features, it is also important to preserve the statistical property of the data in the latent space, such as the variance [155]. This can be done by maximizing the trace of the covariance matrix $\mathrm{cov}(\boldsymbol{X'}) = \frac{1}{n}(\boldsymbol{X'} - \overline{\boldsymbol{X'}})(\boldsymbol{X'} - \overline{\boldsymbol{X'}})^\top$ of the projected samples, where $\overline{\boldsymbol{X'}}$ denotes the mean of $\boldsymbol{X'}$. Assembling the HSIC (dropping the scalar) and the covariance objectives, and further adding an orthogonal constraint on W, the final objective function to be maximized is

$$\begin{aligned} &\max_{W} -\mathrm{tr}(\boldsymbol{W}^\top \boldsymbol{K}_{\widehat{\boldsymbol{X}}}\boldsymbol{H}\boldsymbol{K}_{\boldsymbol{D}}\boldsymbol{H}\boldsymbol{K}_{\widehat{\boldsymbol{X}}}\boldsymbol{W}) + \mu\mathrm{tr}(\boldsymbol{W}^\top \boldsymbol{K}_{\widehat{\boldsymbol{X}}}\boldsymbol{H}\boldsymbol{K}_{\widehat{\boldsymbol{X}}}\boldsymbol{W}), \\ &\text{s.t.}\ \ \boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}, \end{aligned} \tag{4-2}$$

where $\mu > 0$ is a trade-off parameter between optimizing the HSIC and the covariance. The solution of $\boldsymbol{W}$ is given by the h eigenvectors of $\boldsymbol{K}_{\widehat{\boldsymbol{X}}}(-\boldsymbol{H}\boldsymbol{K}_{\boldsymbol{D}}\boldsymbol{H} + \boldsymbol{\mu}\boldsymbol{H})\boldsymbol{K}_{\widehat{\boldsymbol{X}}}$ corresponding to the h largest eigenvalues.

4.3.3.2 Transfer Component Analysis

Transfer Component Analysis (TCA) [156] attempts to mitigate the distribution mismatch by minimizing the Maximum Mean Discrepancy (MMD) in a reproducing kernel Hilbert space (RKHS) [157], which measures the distance between the empirical means of the source domain and the target domain. Intuitively, when the distance between the means of both domains is small, the data tend to distribute similarly in both domains. It has proven that when the RKHS is universal, MMD will asymptotically approach zero if and only if the two distributions are identical [158]. Using the kernel trick, the distance measured in terms of MMD between the means of the projected source data $\boldsymbol{X}_{\boldsymbol{s}}'$ and target data $\boldsymbol{X}_{\boldsymbol{t}}'$ in the latent subspace evaluates to

$$\mathrm{Dist}(\boldsymbol{X}_s', \boldsymbol{X}_t') = \mathrm{tr}\big((\boldsymbol{K}\boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{K})\boldsymbol{L}\big) = \mathrm{tr}(\boldsymbol{W}^\top\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}\boldsymbol{W}), \tag{4-3}$$

where $\boldsymbol{W} \in \mathbb{R}^{n\times h}$ is the projection matrix, $\boldsymbol{K} = [k_{ij}] \in \mathbb{R}^{n\times n}$ the kernel matrix defined on $\boldsymbol{X}$, and $\boldsymbol{L} = [L_{ij}]$ where $\boldsymbol{L_{ij}} = 1/n_s^2$ if $\boldsymbol{X_{:i}}, \boldsymbol{X_{:j}} \in \boldsymbol{X_s}$, else $L_{ij} = 1/n_t^2$ if $\boldsymbol{X_{:i}}, \boldsymbol{X_{:j}} \in \boldsymbol{X_t}$, otherwise $\boldsymbol{L_{ij}} = -(1/n_s n_t)$.

The cost function comprises the distance and a regularization term we wish to minimize, and is subjected to a variance constraint [156]:

$$\begin{aligned} &\min_{W} \mathrm{tr}(\boldsymbol{W}^\top\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}\boldsymbol{W}) + \mu\mathrm{tr}(\boldsymbol{W}^\top\boldsymbol{W}), \\ &\text{s.t. } \boldsymbol{W}^\top\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}\boldsymbol{W} = I, \end{aligned} \tag{4-4}$$

where $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is the same centering matrix as in MIDA, and μ the trade-off parameter. Solving (4-4) for $\boldsymbol{W}$ analytically yields the h eigenvectors of $(\boldsymbol{K}\boldsymbol{L}\boldsymbol{K} + \boldsymbol{\mu}\boldsymbol{I})^{-1}\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}$ corresponding to the h leading eigenvalues.

4.3.3.3 Subspace Alignment

Subspace alignment (SA) [159] attempts to align the principal component analysis (PCA)-induced bases of the subspace of the source and the target domains. We generate the bases of the h-dimensional subspaces of the source domain and the target domain by applying PCA to $\boldsymbol{X_s}$ and $\boldsymbol{X_t}$ and taking the h eigenvectors corresponding to the h leading eigenvalues. Let $\boldsymbol{Z_s}$ and $\boldsymbol{Z_t}$ denote the bases of the subspaces of the source and the target domain, respectively, $\boldsymbol{Z_s} \in \mathbb{R}^{m\times h} = \mathrm{PCA}(\boldsymbol{X_s}, h)$, $\boldsymbol{Z}_t \in \mathbb{R}^{m\times h} = \mathrm{PCA}(\boldsymbol{X}_t, h)$. To align $\boldsymbol{Z_s}$ with $\boldsymbol{Z_t}$, a linear transformation matrix $\boldsymbol{W} \in \mathbb{R}^{h\times h}$ is applied to $\boldsymbol{Z_s}$. The desired $\boldsymbol{W}$ is to minimize the Bregman matrix divergence:

$$\boldsymbol{W} = \min_{W} (\|\boldsymbol{Z_s}\boldsymbol{W} - \boldsymbol{Z_t}\|_{\mathcal{F}}^2), \tag{4-5}$$

where $\|\cdot\|_F^2$ is the Frobenius norm. It follows that the closed-form solution of $\boldsymbol{W}$ is given by $\boldsymbol{W} = \boldsymbol{Z_s^\top Z_t}$ [159]. The source and target data can then be projected to the aligned subspaces, respectively, by $\boldsymbol{X_s'} = (\boldsymbol{Z_s Z_s^\top Z_t})^\top \boldsymbol{X_s}$ and $\boldsymbol{X_t'} = \boldsymbol{Z_t^\top X_t}$.

4.3.3.4 Information Theoretical Learning

Information theoretical learning (ITL) [160] hypothesizes discriminative clustering and consists in optimizing two information-theoretical quantities: (4-7) and (4-8).

Let $\boldsymbol{W} \in \mathbb{R}^{h \times m}$ be the projection matrix to the domain-invariant subspace. The squared distance between two points $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ in the subspaces is expressed as $d_{ij}^2 = \|\boldsymbol{W x_i} - \boldsymbol{W x_j}\|_2^2 = (\boldsymbol{x_i} - \boldsymbol{x_j})^\top \boldsymbol{M} (\boldsymbol{x_i} - \boldsymbol{x_j})$, where $\boldsymbol{M} = \boldsymbol{W}^\top \boldsymbol{W}$ is the Mahalanobis distance metric in the original m-dimensional feature space. Given a point $\boldsymbol{x_i}$ and a set of points $\{\boldsymbol{x_j}\}$, the conditional probability of having $\boldsymbol{x_j}$ as the nearest neighbor of $\boldsymbol{x_i}$ is parametrized by $p_{ij} = e^{-d_{ij}^2} / \sum_{j \neq i} e^{-d_{ij}^2}$. Thus, if the labels of $\{\boldsymbol{x_j}\}$ are known (e.g., $\{\boldsymbol{x_j}\}$ are from the source data), it follows that the posterior probability $\hat{p}(y_i = k | \boldsymbol{x_i})$ for labeling $\boldsymbol{x_i}$ as class k is

$$\hat{p}_{ik} = \sum_{j \neq i} p_{ij} \delta_{jk}, \tag{4-6}$$

where δ_{jk} is 1 if $\boldsymbol{x_j}$ is labeled as class k, and 0 otherwise. Given c classes, a c-dimensional probability vector can be formed: $\hat{\boldsymbol{p}}_i = [\hat{p}_{i1}, \hat{p}_{i2}, \ldots, \hat{p}_{ic}]^\top$. We wish to maximize the mutual information between the target data X_t and their estimated labels $\hat{Y}$ parametrized by $\hat{\boldsymbol{p}}$:

$$I_t(X_t; \hat{Y}) = H[\hat{\boldsymbol{p}}_0] - \frac{1}{n_t} \sum_i H[\hat{\boldsymbol{p}}_i], \tag{4-7}$$

where $H[\boldsymbol{p}] = -\sum_i p_i \log p_i$ denotes the entropy of the probability vector $\boldsymbol{p}$, $\hat{\boldsymbol{p}}_0$ is the prior distribution given by $1/n_t \sum_i \hat{\boldsymbol{p}}_i$.

Since $\hat{\boldsymbol{p}}_i$ is estimated based on the principle of nearest neighbors, the validity of $\hat{\boldsymbol{p}}_i$ hinges on the assumption that the source data and target data are close to each other in the latent subspace. That said, given a sample $\boldsymbol{x}_i$ and a binary probability vector $\boldsymbol{q}_i$ denoting its domain label, if the assumption holds, we cannot determine $\boldsymbol{q}_i$ given $\boldsymbol{x}_i$ well above the chance level. To achieve this, we minimize the mutual information between domain label Q and data samples X, expressed as

$$I_{st}(X;Q) = H[\hat{\boldsymbol{q}}_0] - \frac{1}{n}\sum_i H[\hat{\boldsymbol{q}}_i], \tag{4-8}$$

where $\hat{\boldsymbol{q}}_i = [\hat{q}_{i1}\ \hat{q}_{i2}]^\top$ is estimated via $\hat{q}_{ik} = \sum_{j \neq i} p_{ij}\delta_{jk}$ similar to (4-6), except that δ_{jk} now indicates domain label. The prior probability $\hat{\boldsymbol{q}}_0$ is computed as $\frac{1}{n}\sum_i \hat{\boldsymbol{q}}_i$.

Assembling the two information-theoretical quantities, we derive the cost function as

$$\min_{\boldsymbol{W}} -I_t(X_t;\hat{Y}) + \lambda I_{st}(X;Q), \tag{4-9}$$

where λ is a trade-off parameter. The cost function (4-9) is parametrized by $\boldsymbol{W}$ and is a non-convex function. We resort to iterative gradient descend methods to optimize (4-9). $\boldsymbol{W}$ can be heuristically initialized as being the PCA of the target domain [160].

4.3.3.5 Geodesic Flow Kernel

The subspaces of the source and the target domains are represented by two points on a Grassmann manifold, where geometric, differential, and probabilistic structures can be defined [161]. Authors of Geodesic Flow Kernel (GFK) domain adaptation [161] proposed to construct a geodesic flow linking the subspaces of the source and the target domain via an infinite number of interpolating subspaces in-between on a Grassmann manifold. Then, they

project the source and the target data into each of the infinitely many interpolating subspaces and concatenate the resultant, infinitely many feature vectors to form a super feature vector. To avoid explicitly manipulating on this infinite dimensional feature space, they leverage geodesic flow kernel representing the inner products between any two points in the infinite space, known as the kernel trick. Let $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ be two points in the original m-dimensional feature space, the GFK between them is defined as

$$\mathrm{GFK}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top G \boldsymbol{x}_j, \tag{4-10}$$

To derive G, we need some more math definitions. Let $\boldsymbol{P_s}$, $\boldsymbol{P_t} \in \mathbb{R}^{m \times h}$ be the bases of the subspaces induced by PCA for the source data and the target data, $\boldsymbol{R_s} \in \mathbb{R}^{m \times (m-h)}$ be the orthogonal complement to $\boldsymbol{P_s}$, namely $\boldsymbol{R_s}^\top \boldsymbol{P_s} = 0$. Let $\boldsymbol{U_1} \in \mathbb{R}^{h \times h}$, $\boldsymbol{U_2} \in \mathbb{R}^{(m-h) \times h}$ be the components of the following pair of singular value decomposition (SVD),

$$\boldsymbol{P_s}^\top \boldsymbol{P_t} = \boldsymbol{U_1} \boldsymbol{\Gamma} \boldsymbol{V}^\top, \ \boldsymbol{R_s}^\top \boldsymbol{P_t} = -\boldsymbol{U_2} \boldsymbol{\Sigma} \boldsymbol{V}^\top \tag{4-11}$$

$\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are $h \times h$ diagonal matrices consisting of $\cos\theta_i$ and $\sin\theta_i$ for $i = 1, 2, \ldots, h$, where θ_i are the principal angles between the ith bases of $\boldsymbol{P_s}$ and $\boldsymbol{P_t}$. Then, $\boldsymbol{G}$ is defined as

$$\boldsymbol{G} = [\boldsymbol{P_s U_1} \ \boldsymbol{R_s U_2}] \begin{bmatrix} \boldsymbol{\Lambda_1} & \boldsymbol{\Lambda_2} \\ \boldsymbol{\Lambda_2} & \boldsymbol{\Lambda_3} \end{bmatrix} \begin{bmatrix} \boldsymbol{U_1}^\top \boldsymbol{P_s}^\top \\ \boldsymbol{U_2}^\top \boldsymbol{R_s}^\top \end{bmatrix}, \tag{4-12}$$

where $\boldsymbol{\Lambda_1}$, $\boldsymbol{\Lambda_2}$ and $\boldsymbol{\Lambda_3}$ are diagonal matrices consisting of diagonal elements

$$\lambda_{1i} = 1 + \frac{\sin(2\theta_i)}{2\theta_i}, \ \lambda_{2i} = \frac{\cos(2\theta_i) - 1}{2\theta_i}, \ \lambda_{3i} = 1 - \frac{\sin(2\theta_i)}{2\theta_i}. \tag{4-13}$$

4.3.3.6 Kernel Principal Component Analysis

Kernel-based principal component analysis (KPCA) [152] is the kernelized extension of PCA exploiting the kernel trick. Strictly speaking, KPCA was not originally developed for domain adaptation purpose, but has been included for comparison with domain adaptation methods in the literature [93, 96, 149, 156], citing the denoising and dimension reduction effect of KPCA. Let $\boldsymbol{\phi}$ be the mapping that maps $\boldsymbol{X}$ to a possibly very high dimensional space, the kernel matrix $\boldsymbol{K_X} = [k_{ij}] = \boldsymbol{\phi}(\boldsymbol{X})^\top \boldsymbol{\phi}(\boldsymbol{X})$ for $\boldsymbol{X}$ is computed with a proper kernel function $\ker(\cdot)$, $k_{ij} = \ker(\boldsymbol{X}_{:i}, \boldsymbol{X}_{:j})$. We then center the kernel matrix $\boldsymbol{K}$ by computing $\tilde{\boldsymbol{K}} = \boldsymbol{K} - \boldsymbol{HK} - \boldsymbol{KH} + \boldsymbol{HKH}$, where $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is the centering matrix with all elements equal to $1/n$. Next, we solve the eigendecomposition problem $\lambda \boldsymbol{V} = \tilde{\boldsymbol{K}} \boldsymbol{V}$ for $\boldsymbol{V}$ and λ, where $\boldsymbol{V}$ and λ denote the eigenvectors and eigenvalues, respectively.

Let h be the desired latent subspace dimension, the projection matrix $\boldsymbol{W}$ is constructed with the h eigenvectors from $\boldsymbol{V}$ corresponding to the h largest eigenvalues. The projected samples $\boldsymbol{X'} = [\boldsymbol{X'_s}\, \boldsymbol{X'_t}]$ are computed by $\boldsymbol{X'} = \boldsymbol{K_X W}$. It has proven [152] that KPCA is equivalent to performing standard PCA in the $\boldsymbol{\phi}$ space, directly manipulating which can be prohibitively expensive.

Before we proceed to the next section, we briefly discuss the distinctions of the methods, which are summarized in Table 4-2. MIDA is the only method that can handle multiple source domains, thanks to its domain feature augmentation. MIDA and TCA are closely related in that both try to optimize statistics in RKHS. MIDA, TCA, GFK, and KPCA employ kernel methods and transform the data into kernel representation. GFK and SA have closed-form solutions, which gives them advantages in speed. ITL is based on iterative gradient optimization and may be slower than other methods. It is worth pointing out

Table 4-2 Comparison of different domain adaptation techniques.

Method	Objective	Multi source/target domain	Optimal solution
TCA [156]	Minimize the distance between the empirical means of the source domain and the target domain	No	Analytical
SA [159]	Minimize the Frobenius norm of the difference between the bases of source and target domain	No	Analytical
ITL [160]	Minimize the mutual information between domain label and data samples, and maximize the mutual information between target data and their estimated labels	No	Iterative optimization
GFK [161]	Link the source and target domain with infinite many interpolating subspaces in-between on a Grassmann manifold	No	Analytical
KPCA [152]	Perform PCA in the kernel space	No	Analytical
MIDA [149]	Maximize the independence between data and their domain labels	Yes	Analytical

that the label information Y_s is not used in any method, and the transfer learning is carried out on an unsupervised basis.

4.4 Experiments

In the following experiments, we first evaluate the effectiveness of the domain adaptation techniques in a within-dataset leave-one-subject-out cross-validation setting. We then focus on the evaluation of cross-dataset domain adaptation performance. Both evaluations are based on a standard transductive transfer learning scheme [147] as was used in [93, 96-98]. Figure 4.2 illustrates the transductive transfer learning scheme. We adopt the Logistic Regression classifier [101] throughout all experiments.

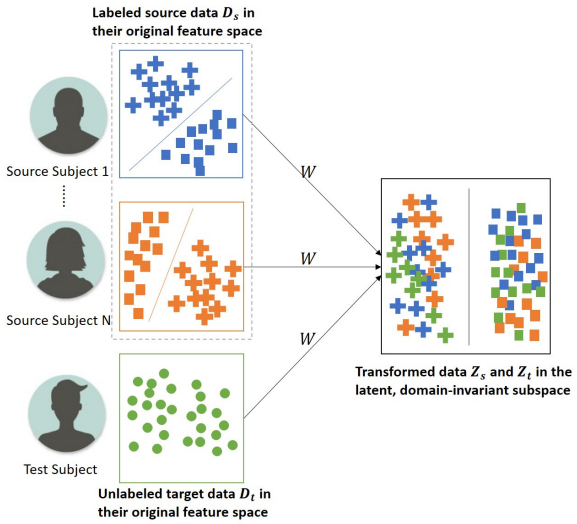


Figure 4.2 Illustration of transductive domain adaptation. Labels of data are represented by the shape (plus: class 1; square: class 2; circle: unlabeled). Source data and target data are pooled together to learn the transformation matrix $\boldsymbol{W}$ to the latent, domain-invariant subspace. A classifier can be trained on the transformed source data, and predict the class labels on the transformed target data. In a within-dataset setting, the training data and the test data are from different subjects within the same dataset. In a cross-dataset setting, the training data and the test data are from two different datasets.

4.4.1 Within Dataset Domain Adaptation

In this experiment, we evaluate the classification accuracy on a leave-one-subject-out cross-validation basis. Specifically, one subject from the dataset in question is left out as the test subject, and the remaining subjects are viewed as training subjects who contribute training data. In DEAP, one subject contributes 180 samples (60 samples/class). As such, the training set consists of $180 \times 13 = 2340$ samples from 13 subjects, and the test set 180 samples from the test subject. In SEED, one subject contributes 2775 samples (925 samples/class/session). The training set comprises $2775 \times 14 = 38850$ samples from 14 subjects and the test set 2275 samples from the test subject. We adopt

unsupervised transductive domain adaptation scheme to jointly project the training data and test data to the latent, domain-invariant subspace. It has to be pointed out that for SEED, due to the large number of training samples, it is infeasible to include all training samples into the domain adaptation algorithm given the limited computer memory [93, 96]. Therefore, for SEED, we randomly sample 1/10 of the training data, equaling to 3885 samples, as actual training data for the domain adaptation algorithms and the subsequent classifier training. We repeat the procedure 10 times for SEED, so that the randomly sampled training data covers a good range of the whole training data set. The classification performance is averaged over 10 runs. We compare the performance of several state-of-the-art domain adaptation techniques to each other, as well as to the baseline performance where no domain adaptation method is adopted. We also compare the domain adaptation performance on two established affective EEG datasets. We stress that domain adaptation techniques have been applied to SEED with success in [93, 96-98]. However, there is little study looking into the performance of the said domain adaptation techniques on DEAP. Chai *et al.* [97] mentioned briefly without presenting results that it is difficult to successfully apply domain adaptation techniques on DEAP, and that negative transfer has been observed, where the classification performance is actually degraded when domain adaptation techniques are applied.

As with other machine learning algorithms, domain adaptation algorithms require that certain hyperparameters be set. One such common hyperparameter is the dimension of the latent subspace. We find the best latent dimension h by searching $\{5, 10, \ldots, 100\}$ for each domain adaptation algorithm, respectively. For other hyperparameters, we set to the default values recommended by their authors. Table 4-3 gives the details of the hyperparameters used in this experiment.

Table 4-3 Details of hyperparameters.

Method	Hyperparameters
TCA [156]	Kernel = linear, $\mu = 1$, $h = \{5, 10, \ldots 100\}$
SA [159]	$h = \{5, 10, \ldots, 100\}$
ITL [160]	$\lambda = 1$, $h = \{5, 10, \ldots, 100\}$
GFK [161]	$h = \{5, 10, \ldots, 100\}$
KPCA [152]	Kernel = linear, $h = \{5, 10, \ldots, 100\}$
MIDA [149]	Kernel = linear, $\mu = 1$, $h = \{5, 10, \ldots, 100\}$

Table 4-4 presents the classification accuracy of different methods on DEAP and SEED. For DEAP, the mean classification accuracy (std) of the baseline method is 39.05 % (8.36 %). Note that the theoretical chance level for random guessing is 33.33 %, and the baseline accuracy is seemingly close to random guess. The real chance level is dependent on the classifier and the number of test samples [133, 134]. When there are infinitely many samples, the real chance level approaches the theoretical value. For a finite number of samples, the chance level is computed based on repeated simulations of classifying samples with randomized class labels, as is suggested in [133, 134]. We carry out the chance level simulation and present also in Table 4-4 the upper bound of the 95 % confidence interval of the accuracy of simulated random guessing. As we can see, the baseline accuracy exceeds the upper bound of chance level, which leads to the assertion that the baseline is significantly better than chance at a 5 % significance level. Nonetheless, the low absolute accuracy still suggests that there are substantial discrepancies between the sample distributions of different subjects, without handling which would adversely affect the classification accuracy. SA yields an accuracy slightly inferior to the baseline (38.73 % vs. 39.05 %), falling below the upper bound of chance level. It suggests that negative transfer may have happened. Other domain adaptation methods yield improved classification performance over baseline performance. MIDA sees a 9.88 % improvement over the baseline and is the best-performing method,

Table 4-4 Within-dataset leave-one-subject-out cross-validation accuracy, mean % (std %).

Method	DEAP	SEED			
		Session I	Session II	Session III	Session Average
Baseline	39.05 (**8.36**)	57.96 (10.85)	48.79 (14.47)	57.45 (13.09)	54.73 (12.80)
TCA [156]	47.22 (15.59)	**73.56** (8.37)	68.89 (14.43)	72.57 (11.38)	71.67 (11.3)
SA [159]	38.73 (9.39)	66.03 (**7.49**)	64.14 (**10.47**)	66.67 (10.59)	65.61 (**9.52**)
ITL [160]	40.56 (11.92)	65.82 (11.86)	64.00 (15.09)	69.08 (14.77)	66.30 (13.91)
GFK [161]	46.51 (13.48)	65.75 (12.06)	64.15 (12.12)	72.62 (12.87)	67.51 (12.35)
KPCA [152]	39.84 (11.37)	63.56 (11.01)	58.34 (11.51)	65.58 (**9.80**)	62.49 (10.77)
MIDA [149]	**48.93** (15.5)	72.31 (10.86)	**69.45** (15.18)	**75.64** (11.37)	**72.47** (12.47)
Acc Diff (Best)	9.88	15.6	20.66	18.19	17.74
Chance Lvl	38.85	34.58	34.65	34.60	34.61

closely followed by TCA. Though the relative improvement is significant (t-test, $p < 0.05$), the absolute accuracy is still rather low.

On SEED, the baseline accuracies are noticeably higher than that on DEAP, and much higher than the upper bound of the chance level. The mean accuracy (std) of the baseline method on SEED is 54.73 % (12.80 %) on average over three sessions. The introduced domain adaptation methods can effectively enhance the mean classification accuracy to up to 72.47 % (t-test, $p < 0.05$). The best-performing methods are MIDA and TCA.

4.4.2 Cross Dataset Domain Adaptation

So far, current works [93, 96-98] investigating domain adaptation methods in EEG-based emotion recognition have based their studies on one dataset: SEED. In the previous section, we present the study of domain adaptation methods on SEED as well as on another established, more extensively-used dataset DEAP, and focus on the comparison between different domain adaptation techniques on the two datasets. In this section, we present a preliminary study of domain adaptation methods in a cross-dataset setting. Cross EEG dataset domain adaptation has not been addressed in the existing studies, and little is known about the performance of cross-dataset emotion classification. Conventionally,

EEG studies have some constraints on the experiment settings. Notably, the training and test sessions adopt the same experiment paradigm, the same device or devices with the same technical specification. In our cross-dataset emotion classification experiment, the training data are contributed by one dataset and the test data by the other dataset. This experiment setting simulates the use case when the conventional setting could not be satisfied, and that the training data and test data are collected under different experimental paradigms using different EEG devices and affective stimuli. We stress that such investigation has been lacking thus far, and that it could make great practical sense as it relaxes the constraints on a conventional BCI.

We carry out six experiments to analyze the performance of cross-dataset emotion classification with and without using domain adaptation techniques: DEAP→SEED I, DEAP→SEED II, DEAP→SEED III, SEED I→DEAP, SEED II→DEAP, and SEED III→DEAP. The notation A→B denotes that dataset A contributes the training (source) data and dataset B contributes the test (target) data. SEED I, II and III denotes the data of session I, II and III from SEED, respectively. Since the domain adaptation methods require that the feature space of the source and target domain be the same, we use only the 32 channels in common between DEAP and SEED, and the DE features are 160-dimensional for both datasets. In the first three experiments, DEAP contributes the source training data containing $180 \times 14 = 2520$ samples. The classification performance is evaluated on a per-subject basis on SEED. The target test data contain 2775 samples. The mean classification performance is the average over 15 subjects in SEED. In the last three experiments, SEED contributes the source training data amounting to $2775 \times 15 = 41625$ samples. The classification performance is evaluated on a per-subject basis on DEAP. The test data contain 180 samples. The mean classification performance is the average over 14 subjects in DEAP. Due to limited computer memory, it is infeasible to include all source domain samples into the domain adaptation

algorithms [93, 96-98]. Therefore, we randomly sample 1/10 of the source data, amounting to 4162 samples, as actual source data for the domain adaptation algorithms. We then repeat the experiments 10 times and average the mean classification performance over 10 runs.

Table 4-5 presents the results of the six cross-dataset experiments with different domain adaptation methods as well as with baseline method where no domain adaptation technique is used. As is shown in Table 4-5, the baseline accuracies range from 32.15 % to 34.71 %, all falling below the upper bound of the 95 % confidence interval of the chance level. We therefore assert that the baseline performance is no significantly different from random guess at a 5 % significance

Table 4-5 Cross-dataset leave-one-subject-out cross-validation accuracy, mean % (std %).

Method	DEAP→SEED I	DEAP→SEED II	DEAP→SEED III
Baseline	34.42 (**2.82**)	34.71 (**3.91**)	33.71 (**3.94**)
TCA [156]	**46.95** (11.77)	**47.68** (10.12)	45.83 (11.01)
SA [159]	46.74 (9.47)	42.35 (8.71)	40.65 (12.37)
ITL [160]	42.69 (8.77)	45.19 (7.05)	44.94 (9.05)
GFK [161]	39.53 (4.57)	38.98 (4.76)	39.67 (6.43)
KPCA [152]	42.98 (12.41)	44.26 (11.49)	39.99 (11.29)
MIDA [149]	43.75 (10.04)	46.68 (8.01)	**47.11** (10.60)
Acc Diff (Best)	12.53	12.97	13.40
Chance Lvl	34.68	34.72	34.74
Method	SEED I→DEAP	SEED II→DEAP	SEED III→DEAP
Baseline	34.57 (7.98)	32.99 (**3.44**)	32.51 (**6.73**)
TCA [156]	**42.60** (14.69)	**42.40** (14.56)	**39.76** (15.15)
SA [159]	36.73 (10.69)	37.36 (7.90)	37.27 (10.05)
ITL [160]	34.50 (13.17)	34.10 (9.29)	33.62 (10.53)
GFK [161]	41.91 (11.33)	40.08 (11.53)	39.53 (11.31)
KPCA [152]	35.60 (**6.97**)	34.69 (4.34)	35.11 (10.05)
MIDA [149]	40.34 (14.72)	39.90 (14.83)	37.46 (13.11)
Acc Diff (Best)	8.03	9.41	7.25
Chance Lvl	38.35	38.38	38.44

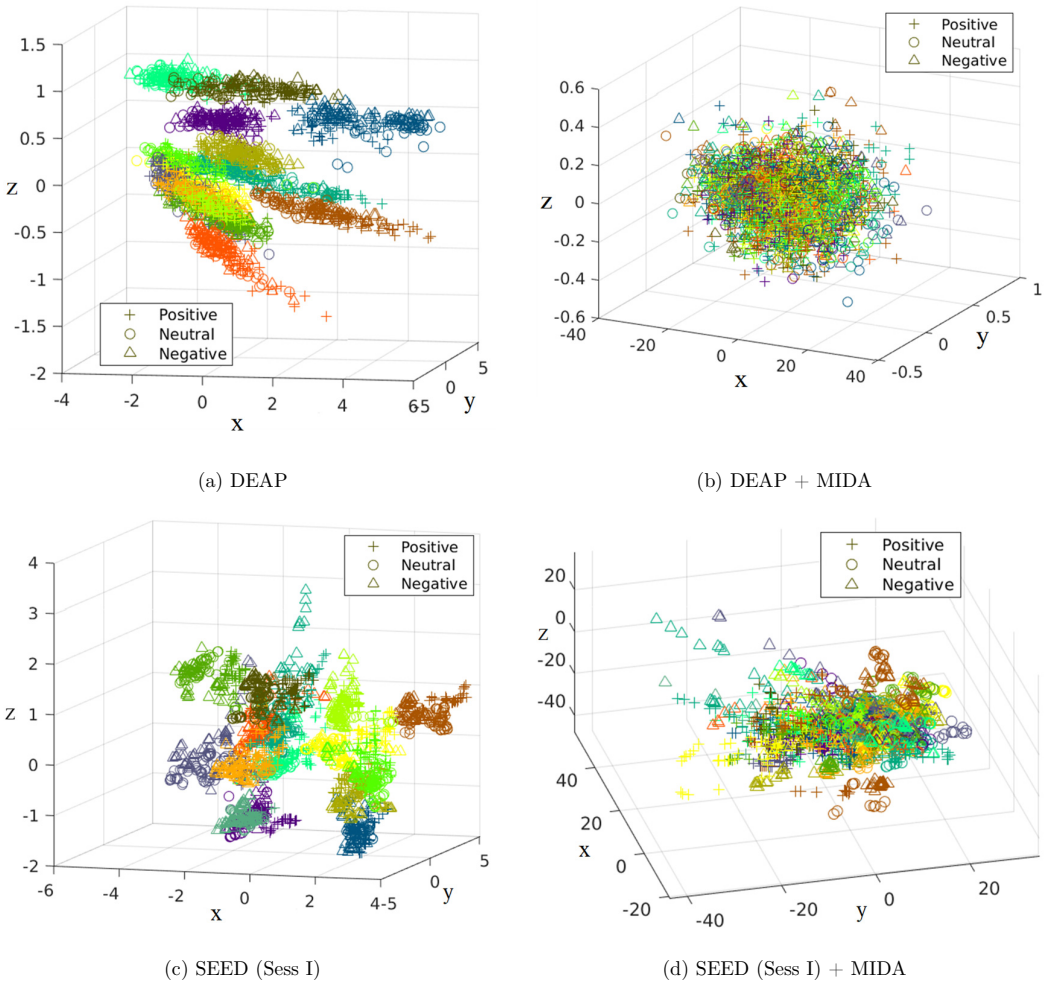
level. It hints that the technical differences between the two datasets may have introduced large discrepancies between the sample distributions of the source and target domains, besides the inter-subject variance. Domain adaptation methods can effectively improve the accuracies over the baseline performance. TCA and MIDA are found to be the best-performing methods in the cross-dataset experiment settings: we observe 7.25 % – 13.40 % accuracy gains over the baseline performance.

4.5 Results and Discussions

4.5.1 Within Dataset Domain Adaptation

We present the study on the effectiveness of domain adaptation methods in a within-dataset leave-one-subject-out cross-validation setting [162]. In this setting, each subject is hypothesized to constitute a domain by himself/herself, and domain discrepancy exists between different subjects. MIDA and several domain adaptation methods have been introduced to bridge the discrepancy between different subjects, so as to enhance the classification accuracy. In our study, domain adaptation methods work effectively on SEED, which coincides with the findings of [93, 96-98]. MIDA and TCA are found to be the more effective methods, gaining an improvement of up to 20.66 % over the baseline accuracy. On DEAP, domain adaptation could, to a less significant extent, improve the accuracy by up to 9.88 %. We observe that domain adaptation methods work less effectively on DEAP than on SEED, which partially coincides with [97], which briefly mentioned that negative transfer had hindered the successful application of domain adaptation methods on DEAP. Despite that it is important to know how to avoid negative transfer, little research work has been published on this topic [147]. It remains an open question as to what determines the effectiveness of domain adaptation methods on a specific dataset. Rosenstein *et al.* [163] empirically showed that if two tasks are too dissimilar,

then brute-force transfer may hurt the performance in the target task. Here, we try to address this question with some empirical evidence. Figure 4.3 presents the sample distribution of both datasets with and without using domain adaptation. As we can see in Figure 4.3a and Figure 4.3c, originally, the samples distribute differently between different subjects—each subject forms a cluster in the space by himself/herself. This suggests that large discrepancies between different subjects exist in the original feature space. We observe that samples are distributed more "orderly" on SEED than on DEAP. For example, on SEED, positive samples tend to locate at the right-hand side of each cluster. However, on DEAP we do not observe similar patterns. In fact, samples belonging to different classes overlap substantially in each cluster on DEAP, making it difficult to discriminate between different classes. Figure 4.3b and Figure 4.3d show the data sample distribution after applying MIDA. Clearly, the discrepancies between different subjects have been reduced, as the clusters are closer to each other. We observe that samples are better aligned on SEED then on DEAP. For example, on SEED, negative samples from different subjects tend to cluster in the upper space, while positive samples from different subjects tend to cluster in the lower space. However, on DEAP, samples of different classes are not well-aligned in the projected space. It might suggest that samples that are more "orderly" distributed in its original feature space tend to be better aligned in the domain-invariant subspace, and it might explain why the baseline performance on SEED is superior to that on DEAP, and why domain adaptation methods give better performance on SEED than on DEAP.



(a) DEAP

(b) DEAP + MIDA

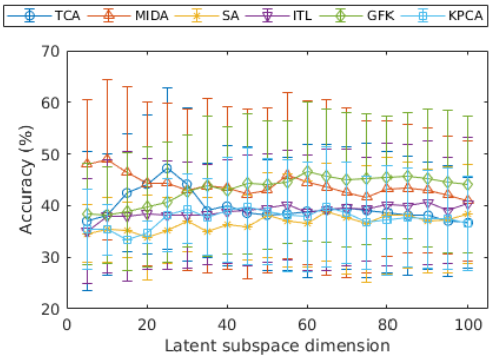
(c) SEED (Sess I)

(d) SEED (Sess I) + MIDA

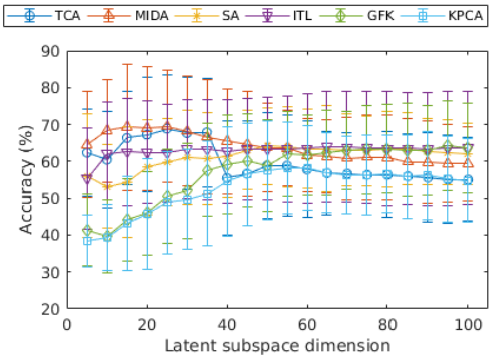
Figure 4.3 Illustration of data sample distribution at feature level. Samples are reduced to 3 dimensions via Principal Component Analysis. The x, y and z axis corresponds to the first, second and third principal components, respectively. +, O, and Δ denotes positive, neutral and negative samples, respectively. Colors represent different subjects. The perspective is adjusted to the best viewing angle for each plot. (a) DEAP. (b) DEAP + MIDA. (c) SEED (Sess I). (d) SEED (Sess I) + MIDA.

4.5.1.1 Latent Dimension

The dimension of the latent, domain-invariant subspace is one common hyperparameter among different domain adaptation methods. However, there is no analytically optimal value and it has to be tuned based on trial-and-error or cross-validation, as was used in the [93, 96, 149, 156]. Figure 4.4 presents the trajectories of mean classification accuracy with varying latent subspace dimension h. We observe that on both datasets, the performance of ITL is not so sensitive to varying h, but the other methods are. The best accuracies are



a. DEAP.



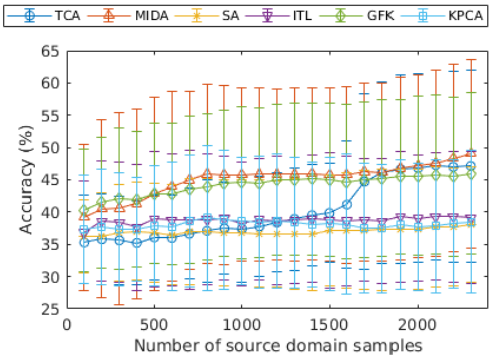
b. SEED I.

Figure 4.4 Classification accuracy with varying latent subspace dimension on (a) DEAP; (b) SEED I

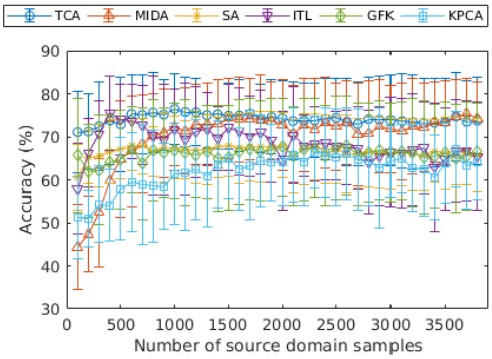
obtained in a low-dimensional subspace, generally under 40. The optimal latent subspace dimension is considerably smaller than the original feature space dimension. It suggests that domain-invariant information may exist in a low dimensional manifold. Besides the benefits of domain invariance, a low dimensional latent space also reduces the burden of classifier training. Based on the finding, $10 - 30$ is an empirically suggested range for the latent subspace.

4.5.1.2 Number of Source Samples

Figure 4.5 presents the effect of varying number of source samples. Due to prolonged computation time and limited computer memory, it is not practicable to include all available source data samples into the calculation. Therefore, we have to randomly sample a subset of source data for inclusion into the calculation. As a reference, [93, 96, 149, 156, 160, 161] have restricted their source datasets to be under 6000 samples. In our experiment on DEAP, the source dataset size varies from 100 to 2300. SA, ITL, and KPCA are less sensitive to varying number of source dataset size. For TCA, MIDA, and GFK, accuracies could be improved with growing number of source data. MIDA maintains a better accuracy than other methods at above 500 source domain samples. On SEED, the source dataset size varies from 100 to 3800. Similarly, accuracies are improved with more available source data. The accuracy flattens at above 1000 source domain samples. From that point onwards, MIDA and TCA perform similarly and are superior to the other methods. From this finding, we empirically conclude that if we have sufficient source data, MIDA and TCA tend to outperform other methods and hence the preferred techniques in terms of accuracy.



a. DEAP



b. SEED I

Figure 4.5 Classification accuracy with varying number of source domain samples on (a) DEAP; (b) SEED I

4.5.1.3 Computation Time

The domain adaptation methods introduce extra computational overhead. Table 4-6 shows the computation time for each domain adaptation method on both datasets. All experiments are simulated on MATLAB R2017a on a desktop PC equipped with one Intel Xeon E5-2630 CPU @ 2.20 GHz, 64 GB RAM, 512 GB SSD.

Table 4-6 Computation time (s) of each domain adaptation method on both datasets.

	TCA	MIDA	SA	ITL	GFK	KPCA
DEAP	50.20	14.28	0.08	213.36	0.31	44.30
SEED	950.61	268.18	0.51	1348.70	1.33	992.14

Let $|\boldsymbol{D_s}|$ and $|\boldsymbol{D_t}|$ denote the size of the source and target datasets, respectively. On DEAP, $|\boldsymbol{D_s}| = 2340$ and $|\boldsymbol{D_t}| = 180$. On SEED, $|\boldsymbol{D_s}| = 3885$ and $|\boldsymbol{D_t}| = 2775$. The computation time is highest for ITL in both cases, due to its gradient-based iterative optimization. The two best-performing methods in terms of accuracy, TCA and MIDA, introduce considerable overheads. The major overheads of TCA, MIDA, and KPCA can be attributed to the eigendecomposition operation, which has a time complexity of $O(hn^2)$ [156]. This can become expensive when n grows to a large value. In existing studies [93, 96, 149, 156, 160, 161] simulated on averaged-specced PCs, n has been restricted to under 6000. Thus, they are more suitable for offline processing. The computation time of SA and GFK are almost negligible. SA and GFK might be used for online processing, but at the cost of lower accuracy performance.

4.5.2 Cross Dataset Domain Adaptation

We present a preliminary study of cross-dataset EEG-based emotion classification task [162]. Conventionally, EEG-based applications have been constrained to using the same experiment protocol and device in the training and testing sessions. Clearly, it makes great practical sense if such constraints can be relaxed. In one scenario, for example, we could unite the high-quality datasets published by different research groups, and adapt those datasets to cater for our applicational need, instead of collecting and labeling new data from scratch. We set out to investigate the performance of cross-dataset

emotion classification, where the two datasets are heterogeneous in various technical specifications, such as EEG devices, affective stimuli, and experiment protocol etc. We observe that the baseline accuracies without applying any domain adaptation method are below the upper bound of random guess, which hints that TCA and MIDA can effectively improve the classification performance over the baseline by 7.25 % – 13.40 %, suggesting that they could potentially reduce the technical discrepancies between datasets. However, though the accuracy improvements are significant (t-test, $p < 0.05$), the absolute accuracies remain below that of within-dataset training and testing. Considering the applicational values, more future studies on this topic are needed.

4.6 Chapter Conclusion

In this chapter, we present a comparative study on domain adaptation techniques on two affective EEG datasets, and a pilot study on cross-dataset emotion recognition [162]. We use two publicly available affective EEG datasets — DEAP and SEED. Though successful application of domain adaptation has been reported on SEED, little is known about the effectiveness of domain adaptation on other EEG datasets. We found that domain adaptation methods work more effectively on SEED than on DEAP. It remains an open question as to what determines the effectiveness of transfer learning techniques. The "orderliness" of the samples in the original feature space might have an impact on the effectiveness of adaptation.

The cross-dataset scheme simulates the use case where a conventional BCI paradigm cannot be satisfied. We demonstrate the effectiveness of MIDA and TCA in coping with domain discrepancy introduced by different subjects and the technical discrepancies with respect to the EEG devices, affective stimuli, experiment protocols etc. We stress that this is of great practical sense as it

relaxes the constraint of a conventional BCI, but has been lacking sufficient investigation thus far. More future studies are needed on this topic.

Chapter 5 Unsupervised Feature Extraction with Autoencoder

In Chapter 5, we address our third research question outlined in Section 1.3: Can we extract discriminative affective EEG features on an unsupervised basis? We begin by stating the problem in Section 5.1 of ad hoc definitions of spectral band widths on using spectral band power features. We then recapitulate the dataset used for our experiment in Section 5.2. Section 5.3 presents the methodologies and the proposed autoencoder structure for unsupervised feature extraction. Section 5.4 documents the experimental procedures. Section 5.5 discusses the experiment results with comparisons to using hand-engineered spectral band power features and using other neural network structures.

5.1 Problem Statement

A closed-loop affective BCI generally consists of signal acquisition, feature extraction, neural pattern classification and feedback to the user (e.g., see Figure 2.9). Feature extraction and neural pattern classification are arguably the most crucial parts in the loop. Spectral band power features have been one of the most widely used features [41, 72, 79, 81, 82, 84, 104, 105] in BCI studies and EEG-based applications. Despite their popular use, however, there lacks a consensus on the definition of frequency ranges—different studies respect different definitions. On the other hand, we argue that the most discriminative spectral components with respect to the task in question are subject-specific, that is, it is difficult to find a common definition of frequency ranges that could perform equally well on all subjects. In view of this, we propose to use autoencoder to learn from each subject the subject-specific, salient frequency components from the power spectral density of EEG signals. Building upon the trained autoencoder, we propose a network architecture especially for EEG feature extraction, one that adopts hidden neuron clustering with added pooling neuron per cluster. The classification performance using features extracted by our proposed method is benchmarked against that using band power features.

The chapter is organized as follows. Section 5.2 introduces the dataset based on which we carry out the experiment. Section 5.3 explains the methodologies. Section 5.4 documents the experiments. Section 5.5 presents the experimental results with discussions. Section 5.6 concludes the chapter.

5.2 Dataset

In this study, we use a publicly available affective EEG dataset SEED contributed by Zheng *et al.* [89], which has been reviewed in Section 4.2. To briefly recapitulate, the dataset contains 15 subjects, each subject taking three

recording sessions during one month at an interval of two weeks between successive sessions. In each session, each subject was presented fifteen movie clips to induce the desired emotional states: positive, neutral and negative, with five movie clips assigned to each emotion. Sixty-two-channeled EEG signals were simultaneously recorded when the subject was exposed to the affective stimuli, at a sampling rate of 1000 Hz. The EEG signals were then down-sampled to 200 Hz and post-processed by a 75 Hz low-pass filter by the authors. The same affective stimuli were used for all three sessions. The resultant dataset contains fifteen EEG trials corresponding to fifteen movie clips per subject per session. Each trial lasts for three to five minutes, depending on the length of the movie clip. Trial #1, 6, 9, 10, and 14 correspond to positive emotions. Trial #2, 5, 8, 11, and 13 target at neutral emotions. Trial #3, 4, 7, 12, and 15 provoke negative emotions.

5.3 Methods

5.3.1 EEG Data Preparation

All EEG trials except the shortest one are truncated at the end to have the same length as the shortest trial, which is 185-second. Each EEG trial is then segmented into multiple 4-second-long sections (each section equaling to 800 sampling points) without overlapping between any two successive sections. As such, each trial yields 46 sections. Features are extracted out of each section. In this study, we use 41 EEG channels [4] out of the 62 recorded.

[4] The 41 channels include: AF3, C2, C3, C4, C5, C6, CB1 (cerebellum), CB2, CP1, CP2, CP3, CP4, CP5, F1, F4, F6, F8, FC1, FC3, FC4, FC6, FP1, FP2, FPZ, FT7, FT8, FZ, O1, O2, OZ, P2, P4, P6, P8, PO4, PO5, PO7, PO8, PZ, TP7 and TP8

5.3.2 Power Spectral Density

Spectral band power features are one of the most widely used features in the context of EEG-based emotion recognition [41, 72, 79, 81, 82, 84, 104, 105]. Though based on neuroscientific findings, the frequency band ranges of interest are somewhat defined on an ad-hoc basis and vary between studies. In our study, we follow such definition [23, 33, 41]: delta band (1 – 4 Hz), theta band (4 – 8 Hz), alpha band (8 – 12 Hz), and beta band (12 – 30 Hz).

Let $\boldsymbol{X} \in \mathbb{R}^{s \times t}$ be one section of EEG signals, where $s = 41$ is the number of channels and $t = 800$ the number of points sampled. The power spectral density of $\boldsymbol{X}(i,:)$ is estimated as periodogram by Fast Fourier Transform (FFT), where $\boldsymbol{X}(i,:)$ is the ith row of $\boldsymbol{X}$. Since one row comprises 800 points sampled at a rate of 200 Hz, the resolution of the periodogram is 0.25 Hz. The power features are computed by averaging the periodogram over the target frequency ranges defined above. The final feature vector is a concatenation of the features of the same frequency band derived from all 41 channels. The dimension of the feature vector is 41 when using delta band, theta band, alpha band, or beta band alone. In addition, we also combine all power bands at feature level by concatenating the feature vectors of four bands. The feature vector is of $41 \times 4 = 164$ dimensions. Each trial yields 46 samples per feature.

5.3.3 Autoencoder

An autoencoder is a neural network that is trained to produce outputs approximating to its inputs [164]. The structure of a simple feedforward, nonrecurrent autoencoder with one hidden layer is shown in Figure 5.1. The objective of the autoencoder is to make $\hat{\boldsymbol{x}}$ resemble $\boldsymbol{x}$. Autoencoder can be trained using the backpropagation algorithm [164]. The training does not involve the class labels of the data and is on an unsupervised basis. However, we are not particularly interested in the output $\hat{\boldsymbol{x}}$. Instead, we are more interested in the output of hidden layer, $\boldsymbol{h}$. When the hidden layer has fewer neurons than the input layer, $\boldsymbol{h}$ is a compressed representation of $\boldsymbol{x}$ and has to capture the most salient feature of $\boldsymbol{x}$ [164] in order to be able to reproduce it at the output layer. $\boldsymbol{h}$ could then be used for further feature learning or as the feature vector of $\boldsymbol{x}$ for classification or regression.

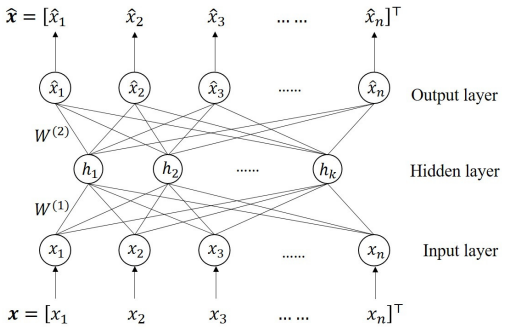


Figure 5.1 An example of an autoencoder with one hidden layer. $\boldsymbol{x}$ is the input to the network, $\boldsymbol{h} = \boldsymbol{f}(\boldsymbol{W^{(1)}x} + \boldsymbol{b^{(1)}})$ and $\hat{\boldsymbol{x}} = \boldsymbol{g}(\boldsymbol{W^{(2)}h} + \boldsymbol{b^{(2)}})$, where $\boldsymbol{h}$ is the output of hidden neurons (also known as *code*), $\boldsymbol{W^{(1)}}$ is the weights between hidden layer and input layer, $\boldsymbol{b^{(1)}}$ is the bias vector of hidden neurons (not drawn in the figure), $\boldsymbol{f}(\cdot)$ is the activation function of hidden neurons (also known as transfer function), $\boldsymbol{W^{(2)}}$ is the weights between hidden layer and output layer, $\boldsymbol{b^{(2)}}$ is the bias vector of output neurons, $\boldsymbol{g}(\cdot)$ is the activation function of output neurons. The network is trained to reproduce input $\boldsymbol{x}$ at the output layer.

5.3.4 Proposed Unsupervised Band Power Feature Extraction

In this study, we leverage autoencoder to automatically learn the salient frequency components from the periodogram instead of predefining the frequency ranges such as delta, theta, alpha, and beta. The input to the autoencoder is the raw periodogram from 1 to 30 Hz with a resolution of 0.25 Hz. The dimension of the periodogram is 117-D, thus the input layer and output layer both consist of 117 neurons. The hidden layer consists of k neurons. After the autoencoder has been trained, the hidden neurons have learned the salient frequency components over $1 - 30$ Hz. Such information is encoded in the weight vectors of the hidden neurons. We hypothesize that hidden neurons carrying similar weights have learned similar frequency components. We propose to cluster the hidden units into several groups by their weight vectors [165]. A mean pooling neuron is added on top of each group to aggregate the outputs from all hidden neurons that are in the same cluster. The outputs of the mean pooling neurons are considered features learned from the raw periodogram, which are essentially weighted power features, but without predefinition of band ranges. The final feature vector is the concatenation of features derived from 41 channels. The dimension of the final feature vector is, therefore, $41m$, where m is the number of clusters of hidden neurons. The proposed network structure is illustrated in Figure 5.2.

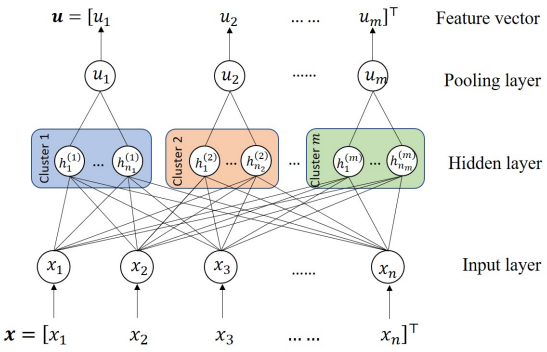


Figure 5.2 Proposed network structure. After an autoencoder has been trained, k hidden neurons are clustered into m groups based on the similarity of their weights. Neurons within the same groups carry weights similar to each other, thus have learned similar components from the input. $h_i^{(c)}$ is the output of the ith hidden neuron in cluster c. n_c is the number of neurons belonging to cluster c, $\sum_{c=1}^{m} n_c = k$. u_c is the pooling neuron added to cluster c. $u_c = (1/n_c)\sum_{i=1}^{n_c} h_i^{(c)}$. $u = [u_1, u_2, \ldots, u_m]^\top$ is viewed as the feature extracted out of x.

5.4 Experiments

We benchmark the performance of standard power features against features that are automatically learned by autoencoder under our proposed structure. The performance is measured by accuracy discriminating the three emotion states.

5.4.1 Using Spectral Power Features

We evaluate the classification accuracy on a per-subject basis by five-fold cross-validation. Within one session, the fifteen trials from the subject in question are partitioned into five folds as follows. Fold 1 = {trial #1,2,3}, fold 2 = {trial #4, 5, 6}, fold 3 = {trial #7, 8, 9}, fold 4 = {trial #10, 11, 12} and fold 5 = {trial #13, 14, 15}. Each fold contains one trial for each emotion. We train the classifier with four folds and test the classifier with the remaining fold. As such,

the training set comprises $46 \times 3 \times 4 = 552$ training samples and the test set consists of $46 \times 3 = 138$ test samples. The process is repeated five times until each fold has served as the test set for once. The per-subject classification accuracy is averaged over five runs. The overall mean accuracy is the average per-subject accuracy over fifteen subjects. In this experiment, we adopt a Logistic Regression classifier [101]. The classifier training process stops at maximum 100 iterations.

5.4.2 Using Features Learned by Autoencoder with the Proposed Structure

Firstly, we need to train the autoencoder to reconstruct the input data (periodograms). Based on the same partition scheme as is used in Section 5.4.1, we set aside one fold as the test set, and the remaining four folds are pooled together as the training set. The training set comprises $46 \times 41 \times 3 \times 4 = 22632$ periodograms. The test set consists of $46 \times 41 \times 3 = 5658$ periodograms. Eighty-five percent of the data randomly sampled from the training set are used as actual training data by the autoencoder, and the rest fifteen percent of the data in the training set are used as validation data to select the best weights, that is, the weight parameters that lead to the minimum reconstruction error on the validation data. In this experiment, we use one hidden layer with $k = 100$ hidden neurons. Input data are 117-D raw periodogram covering 1 – 30 Hz frequency range. Thus, the autoencoder architecture is 117 (input neurons)-100 (hidden neurons)-117 (output neurons). The linear activation function is used in all layers. The reconstruction error between input $\boldsymbol{x}$ and output $\hat{\boldsymbol{x}}$ is measured by mean squared error. The whole network is trained using backpropagation and batch gradient descent with minibatch size equal to 256. Training stops at maximum 50 epochs. The weight parameters that minimize the reconstruction error on validation data are retained. After the autoencoder

has been trained, the output layer is removed from the network. We then employ k-means algorithm to cluster the hidden units into m groups based on the similarity of their weight vectors, m varies from 1 to 10. A mean pooling neuron is added on top of each group to aggregate the outputs, as is shown in Figure 5.2. The outputs of the mean pooling neurons are viewed as features extracted out of the periodogram. The training data, validation data, and test data are fed to the trained network with added pooling layers to extract features. The final feature vector is a concatenation of features from 41 channels. The classifier (same configuration as what is used in Section 5.4.1) is trained on training data pooled with validation data, and tested on the test data. As such, the training data and validation data together contribute 552 training samples to the classifier. The test data contribute 138 test samples. The procedures (autoencoder training, hidden unit clustering, feature extraction and classifier training and testing) are repeated five times per subject, until each fold has served as the test set for once. The per-subject classification accuracy is averaged over five runs. The overall mean accuracy is the average per-subject accuracy over fifteen subjects.

5.5 Results and Discussions

5.5.1 Comparison with Spectral Power Features

The accuracy results classifying three emotion states using different features are tabulated in Table 5-1. Among the four spectral band power features, beta power performs the best. Theta and alpha powers give similar performance, both being inferior to beta and delta power. The fusion of all power features (combined power) does not lead to improved accuracy compared to beta power feature.

Table 5-1 Overall mean classification accuracy (%) classifying three emotions (positive, neutral and negative) using different features.

Feature		Session 1	Session 2	Session 3	Average
Delta		43.68	40.54	41.86	42.03
Theta		43.09	40.11	39.06	40.75
Alpha		41.11	40.02	39.84	40.32
Beta		50.16	50.01	48.51	49.56
Combined power		44.03	40.55	41.89	42.16
Proposed method	$m = 1$	44.70	43.14	46.22	44.69
	$m = 2$	47.23	46.65	48.66	47.51
	$m = 3$	50.45	49.95	50.48	50.29
	$m = 4$	52.37	50.87	53.60	52.28
	$m = 5$	53.77	54.25	55.68	54.57
	$m = 6$	57.07	56.92	58.41	57.47
	$m = 7$	56.37	56.88	59.14	57.46
	$m = 8$	56.62	57.98	58.66	57.75
	$m = 9$	56.97	58.12	59.10	58.06
	$m = 10$	58.19	59.24	59.66	59.03

The results of the proposed feature extraction method are displayed at the lower half of Table 5-1, with a varying number of clusters of hidden neurons m from 1 to 10. When $m = 1$ and 2, the accuracy is better than that of delta, theta, alpha powers but below beta power. Starting from $m = 3$, the accuracy of the proposed feature exceeds standard power features. When $m = 4$, the feature vector dimension is the same as combined power features. The accuracy of the proposed method sees a 10.12 % increase over combined power feature. There is also a tendency that the classification accuracy increases with growing number of clusters of hidden neurons. The best accuracy is attained by the proposed method when $m = 10$, a nearly 20 % increment over theta and alpha power. However, when m exceeds 6, the improvement is only marginal. It is also worth noting that a large m value may not always be favorable, especially when the size of the training set is small. A larger m value results in a large-dimensional feature vector, which requires more training data to fit the classifier. A limited training set increases the risk of overfitting when using large feature vectors.

To see what frequency components may have been chosen by the autoencoder, we visualize the weights of clustered hidden units in Figure 5.3. The plots show the weights of connection between the input layer and the hidden layer of the trained autoencoder of subject 1 in session 1 when $m = 2$. We average the weights within the same cluster and display the positive averaged weights for each cluster. Generally, a connection with a positive weight between input neuron i and hidden neuron j suggests that hidden neuron j favors the input from neuron i, whereas negative weight implies that hidden neuron j opposes the input from neuron i. The first cluster of hidden neurons has three weight peaks at 5.5 Hz, 13.75 Hz, and 24 Hz, respectively, suggesting that this cluster of hidden neurons may favor theta and beta components. The second cluster show relatively evenly distributed weights over the spectrum, peaking at 8.75 Hz within the alpha band. Some delta and higher beta components are also selected by the second cluster, contrary to the first cluster.

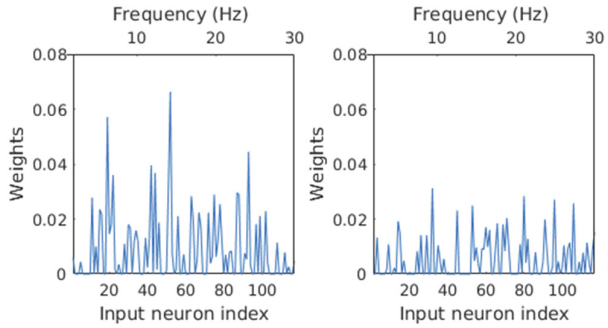


Figure 5.3 Plots of averaged weights of connection between hidden neurons within the same cluster and input neurons. Left: cluster 1; right: cluster 2. Bottom horizontal axis represents the index of input neuron (117 in total). Each input neuron receives the magnitude of periodogram at a specific frequency. The frequency is noted at the top horizontal axis (1 – 30 Hz, corresponding to the 117-D periodogram at a resolution of 0.25 Hz). The weight indicates to what extent a specific frequency component is favored by the hidden neuron. The left cluster has a strong preference for theta and beta components. The right cluster has a preference for delta and higher theta components as compared to the left cluster.

5.5.2 Comparison with Other Neural Networks

In this section, we compare our proposed network with three other neural networks, namely Multilayer Perceptron (MLP) [166], Radial Basis Function Neural Network (RBF NN) [167], and Spiking Neural Network (SNN) [168, 169]. We adopt the same five-fold cross-validation method as was used in Section 5.4.2 to evaluate the performance of these neural networks.

MLP: MLP is a feedforward neural network composed of one input layer, one or several hidden layers and one output layer. Each neuron's input is fully connected to the output of the preceding layer, while the last output layer decides the class of the input. In this experiment of comparison, the input layer consists of 4797 (41×117) neurons which receive the input of concatenated PSDs from 41 channels, and the hidden layer consists of 410 neurons which transform the PSD input into a 410-dimensional feature vector, to be consistently compared with our method when $m = 10$ in our proposed network. The output layer consists of 3 neurons, each corresponding to 1 emotion class. The transfer function of the hidden neurons is set to hyperbolic tangent function and the output neurons adopt softmax transfer function. The connection weights of the network are established using backpropagation algorithm.

RBF NN: RBF NN is a three-layer feedforward neural network. It employs Radial Basis Function (RBF) in its hidden layers as activation functions. The output of the network is a linear combination of the RBF output. The RBFs in the hidden layers are parameterized by the center of cluster $\boldsymbol{c}$ and the standard deviation σ. By nature, a RBF hidden neuron responds to input that is close to its center, and suppress input that is far from its center. Thus, different RBF hidden neurons capture different clusters of information from the input. The centers of RBF neurons are determined by either randomly selecting some samples from the training data and using their values as the centers of the RBF

neurons, or by performing clustering (e.g., use the k-means algorithm) on the training data to find the centers of the clusters and using these values as the centers of the RBF neurons. The weights between neurons are adjusted by using the typical backpropagation algorithm, similar to other neural networks. In this experiment of comparison, for the RBF NN, the input layer consists of 4797 neurons, the hidden layer 410 neurons and output layers 3 neurons, similar to MLP. To initialize the center and width of the hidden neurons, we adopt k-means algorithm to cluster the training data into 410 groups, and heuristically assign the empirical center (mean) of each cluster to the centers of the RBFs, and assign the empirical mean Euclidean distance from the empirical center of each cluster to the width of the RBFs. The connection weights are adjusted by backpropagation algorithm.

SNN: SNN falls into the third generation of neural network models [170], which uses spikes to encode data in the network, and takes into account the time delay in activating a neuron. Thus, it is said to be more faithful to the real biological neurons. For the SNN, we use the NeuCube v1.3 toolkit authored by Kasabov *et al.* [168, 169]. Unlike the other two structures for which we flatten the input, the input to NeuCube is a 41×117 matrix containing PSDs from 41 channels. We then adopt Thresholding Representation (TR) method to encode the real-valued input to spike chains, with the spike threshold set to its default value 0.5. A reservoir of neurons is then initialized with 1471 spiking neurons, as was adopted in [171], where the coordinates of the spiking neurons follow the Talairach system [172] The connection weights between the initialized spiking neurons are adjusted by the Spike-Timing Dependent Plasticity (STDP) learning method [168, 169], with its hyperparameters set to their default values, namely Potential Leak Rate = 0.002, Firing Threshold = 0.5, Refractory Time = 6, STDP Rate = 0.01, Training Iteration = 1, and LDC Probability = 0. By this step, the SNN has learned to extract features from PSDs on an unsupervised basis. We then train a dynamic evolving SNN (deSNN) classifier

to learn the association between the extracted features and their class labels. For this supervised training step, we adopt the hyperparameters as was suggested by the user manual [173] that comes with the NeuCube v1.3, namely Mod = 0.4, Drift = 0.25, K = 3, and Sigma = 1.

As Table 5-2 shows, our proposed method has yielded the best recognition accuracy, closely followed by RBF NN, SNN and MLP. Though the rationale behind the two methods are different, RBF NN and our method both adopt clustering at some point of the feature extraction process, and both methods perform relatively better than other methods. MLP and RBF NN are architecturally similar (both being 4797 – 410 – 3), however, MLP does not adopt any form of explicit clustering in handling the data. It turns out that the performance of MLP is inferior to that of RBF NN. It may suggest that clustering is helpful in improving the quality of the features extracted in both methods, possibly by aggregating similar components within the same cluster and contrasting dissimilar components among different clusters. SNN-based NeuCube is found to be underperforming in this comparison possibly due to two reasons. 1) We adopt the default hyperparameters given by NeuCube or its manual, without finetuning the hyperparameters. 2) In this experiment, we maintain that the input into different networks be the same, that is the power spectral density. PSD loses the temporal information of the signals, which may hinder the holistic spatial-temporal learning of NeuCube. In the future, we will

Table 5-2 Comparison of three-emotion recognition accuracy between our method and MLP, RBF NN and SNN.

Method	Session 1	Session 2	Session 3	Average
MLP	44.40%	45.42%	46.93%	45.58%
RBF NN	54.86%	53.58%	55.24%	54.56%
SNN	52.55%	47.89%	48.98%	49.81%
Ours	58.19%	59.24%	59.66%	59.03%

revisit the idea of spatial-temporal learning from raw EEG signals using NeuCube.

Our proposed network structure is conceptually similar to the RBF NN in that we both adopt the concept of clustering in the hidden layer of our proposed network. However, the following differences between our proposed network structure and the RBF NN can be highlighted.

1. In RBF NN, the clustering is performed directly on the training data, and the resultant cluster centers are used to parametrize the hidden neurons. In our proposal, the clustering is performed on the weights of the hidden neurons of a trained autoencoder. Clustering in RBF network is for the purpose of initializing the neural network, while in our proposal is for finalizing the neural network.
2. RBF NN involves both unsupervised learning (e.g., adjusting the RBFs of the hidden neurons by applying a statistical clustering method) and supervised learning (e.g., adjusting the output neurons' connection based on the supervised labels to produce the desired outputs, be it a classification task or a regression task) [174]. In our proposal, the neural network learns solely on an unsupervised basis. The goal of our proposed network is to extract features from the input, and such features are supplied to any generic classifier for the actual classification task. The neural network, during training, does not utilize any label information from the training data. The classifier training is not a part of our proposed neural network, while in RBF NN, the classifier training is a part of the holistic network.
3. The training of RBF NN can be done holistically via backpropagation, while in our proposed structure, only the first step (training the generic autoencoder) is done via backpropagation. The second step (hidden units clustering) requires no adjustment of the connection weights, but

involves merely grouping the hidden units and aggregating their outputs, which are then used as features for subsequent classification done by any generic classifier.

5.6 Chapter Conclusion

Spectral band power features have been one of the most widely used features in BCI studies and EEG-based applications. On the one hand, the definition of the frequency range, though based on neuroscientific findings, is somewhat on an ad-hoc basis and varying between studies. On the other hand, it is arguable that one definition of band ranges could perform equally well on all subjects. In this study, we propose to find the subject-specific salient frequency components using autoencoder. We propose a network architecture especially for power feature extraction out of raw periodograms of EEG signals [165]. The proposed architecture consists in clustering the hidden neurons of a trained autoencoder with added pooling neuron per each cluster. The proposed method essentially extracts features similar to power features, but without predefinition of band ranges. We benchmark the proposed methods against standard power feature extraction method. Experimental results show that our proposed method yields better accuracy than standard power features when the number of hidden unit clusters $m \geq 3$. When $m = 4$, the proposed method yields feature of the same dimension as combined power features, but performs better than the latter by 10.12 %. The classification accuracy may be further improved given a larger value of m, but we also see that the accuracy increment is only marginal when m exceeds 6. We also compare our proposed structure with other neural networks such as multilayer perceptron (MLP), radial basis function neural network (RBF NN) and spiking neural network (SNN). Extensive comparisons show that our method outperforms MLP by 12.73 % – 13.82 %, RBF NN by 3.33 % – 5.66 %, and SNN by 5.64 % – 11.35 %. We conclude that the proposed method, which automatically learns the salient frequency components, could

potentially outperform standard band power features, whose frequency components are explicitly defined.

Chapter 6 Applications

EEG-based emotion recognition algorithm lies at the core of affective BCIs and can be integrated with different applications that need to be driven by the affective states of the user. In Chapter 6, we showcase two EEG-based emotion-driven applications that we developed: Emotion Avatar and CogniMeter.

6.1 Emotion Avatar

Emotion Avatar [129] is designed to visualize the emotional feeling of the subject with an animated avatar. This application enables the real-time monitoring of human emotion, even if the subject does not explicitly express his/her emotion via facial or vocal expression. It has the potential to reveal the inner feeling of the subject and could be applied in cases such as marketing survey etc. The subject can choose to use a generic recognition model, or to personalize the recognition model by calibrating the program. During the training session, the subject is exposed to affective stimuli to evoke certain emotions and the EEG data are recorded simultaneously. After each stimulus, a self-assessment dialogue is prompted to the subject to evaluate his/her real elicited emotion. The feedback from subject includes the arousal level, dominance level, valence level, liking level and familiarity level (all are discrete and scale from one to nine, with one being the weakest level and nine the strongest level) and a word describing his/her emotion. Figure 6.1 shows the screenshot of the calibrating session. The feature extraction is done in the background and, upon clicking the "Train SVM" button as shown in Figure 6.2, a SVM classifier is trained using the recorded EEG data and the emotion labels from the subject's self-assessment. After training, a personalized emotion recognition application is ready for real-time use. In the real-time recognition phase, as shown in Figure 6.3, the EEG signals are collected by the Emotiv device and wirelessly transmitted to the emotion recognition application where the feature extraction and classification are done. As a result, an emotion label is assigned as the current emotional state of the subject. Simultaneously, the recognized emotion of the subject is visualized via the animated facial expression of a 3D Haptek avatar [175]. Some examples of the animated visualization are presented in Figure 6.6. In Figure 6.6a, the current recognized emotional state of the subject is "pleasant", and the avatar shows a smile on

her face to visualize the pleasant emotion. In Figure 6.6b, the current recognized emotional state of the subject is “angry”, and the avatar shows a face in rage to visualize the angry emotion. Figure 6.6c and Figure 6.6d visualize the frightened and happy emotion, respectively. This application could potentially be helpful for patients with cognitive impairment, who may not be able to express their emotions. In marketing, this application could potentially be used to monitor the feeling of the consumer evaluating the product-to-launch.

We must clarify that we cannot assess the accuracy during the real-time usage, as we would not know the true emotions felt by the user when the he is freely using the program. However, we can reasonably expect that the recognition accuracy of this program is in line with the performance of the proposed algorithm which has been validated on our labeled EEG datasets in the previous Chapter 3. That is, we can reasonably expect that if the subject uses the system with re-calibration, the recognition accuracy for 4 emotions are around 27.90 % – 76.00 % (see Table 3-4, referring to the SVM results as the program is implemented with SVM). If the subject uses the system without re-calibration but uses his previously trained classifier model, we can reasonably expect that the recognition accuracy for 4 emotions are around 27.86 % – 43.53 % (see Table 3-8, referring to the SVM results).

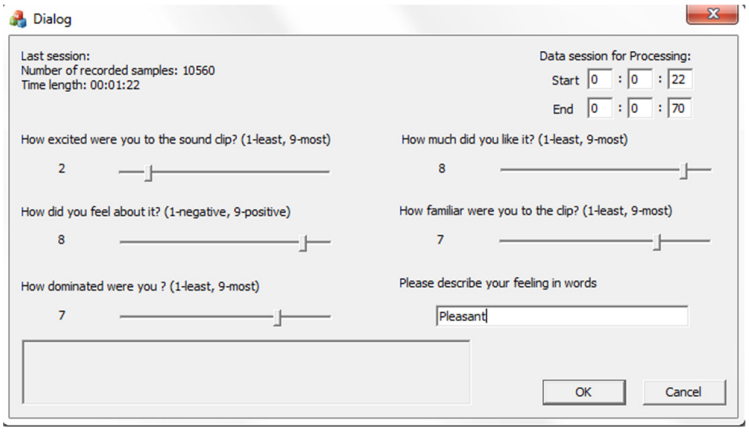


Figure 6.1 Screenshot of the training session [129].

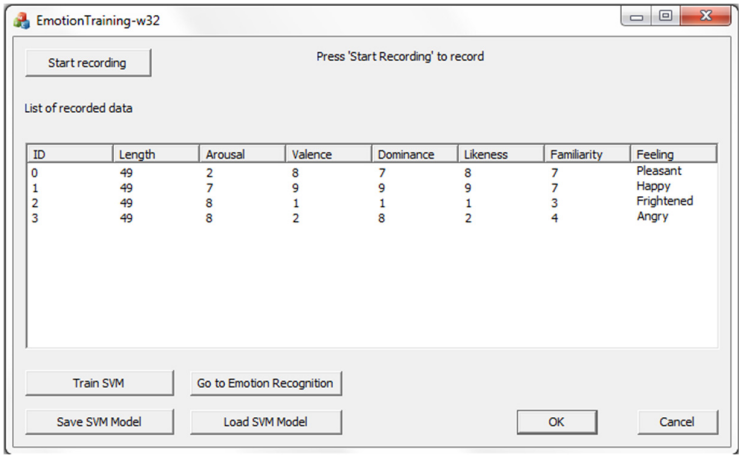


Figure 6.2 Screenshot of the classifier training menu [129].

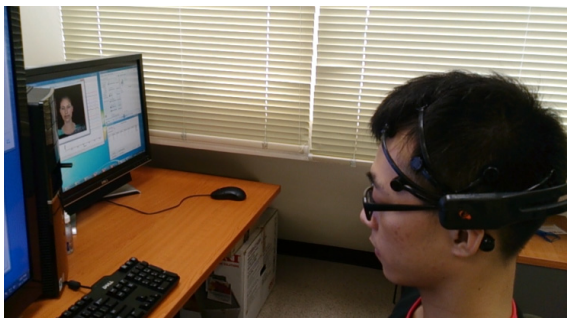


Figure 6.3 Subject wearing an Emotiv headset while his emotion was being recognized in real time.

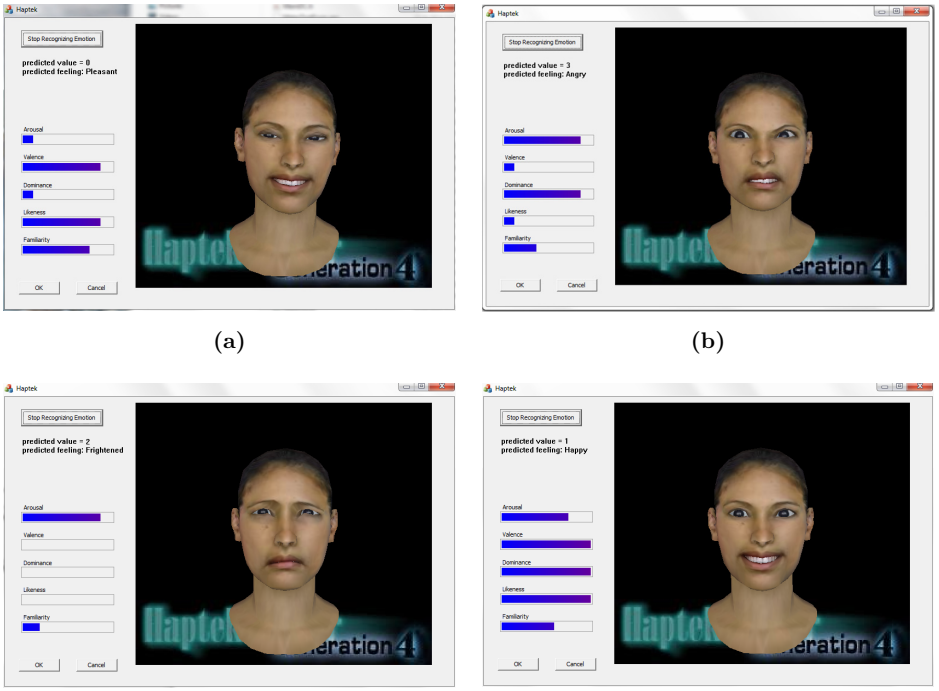


Figure 6.4 Screenshot of the real-time emotion recognition application "Emotion Avatar" [129]. (a). pleasant, (b). angry, (c). frightened, (d). happy.

6.2 CogniMeter

CogniMeter [130] is designed to integrate the monitoring of three mental states into one web-based interface: the emotion [176], mental workload [177] and stress [178]. The application can provide real time visualization of the emotion, mental workload and stress level recognized from the EEG signals, and could potentially benefit the investigation of human factors if deployed in user study [9]. Figure 6.5 shows the block diagram of the CogniMeter. The application comprises two parts: recognition and visualization. In the recognition part, the EEG data are streamed to three recognizers, which recognize the emotion, mental workload and stress states independently. The recognized states are sent to a server which consolidates the results. In the visualization part, a web browser is connected to the server and continuously render the recognized brain

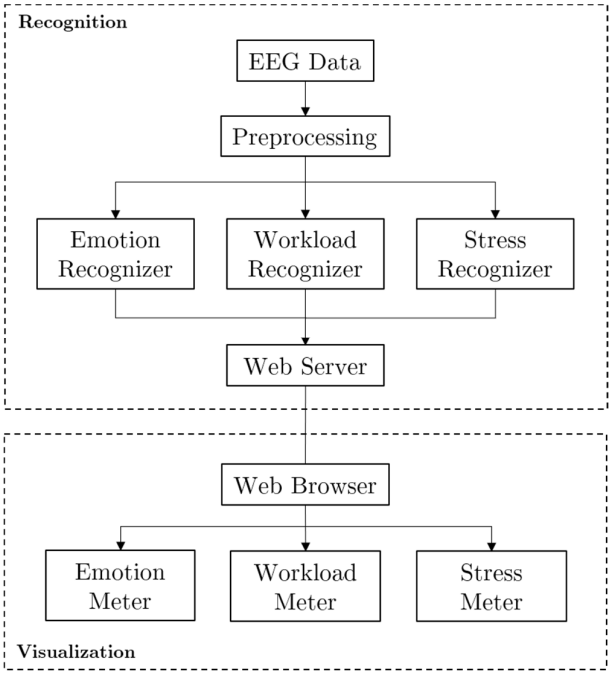


Figure 6.5 The block diagram of CogniMeter [130].

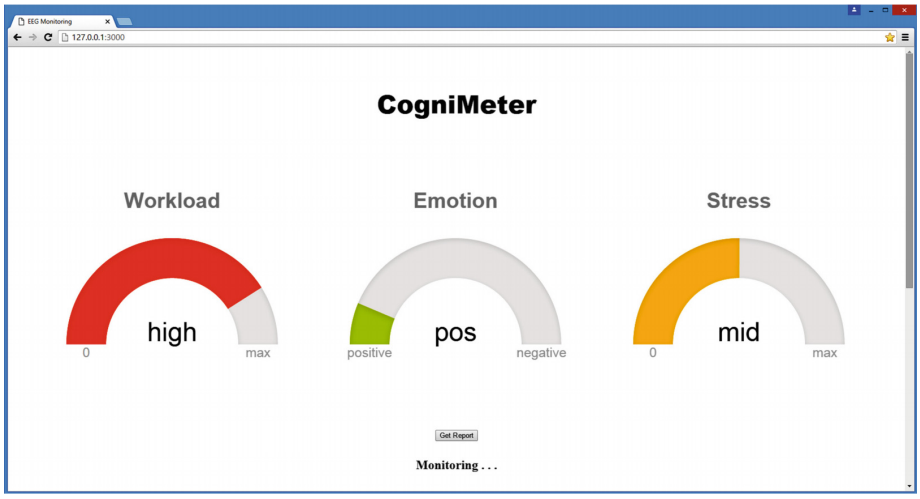


Figure 6.6 Screenshot of CogniMeter [130].

states locally. The advantage of a web-based interface is that the investigator may remotely monitor the user. Figure 6.6 shows the web interface which visualizes the mental workload, emotion and stress level. The meters reflect the current mental states of the user in real time. The color of the gauge is changed in accordance with the recognized brain states. For workload or stress, a green bar indicates a low level of workload or stress, a red bar a high level, and an amber bar a median level. For emotion, a green bar indicates a positive emotion, whereas a red bar indicates a negative emotion. Descriptive words are displayed in the center of the meters, indicating the current recognized mental states.

6.3 Chapter Conclusion

In this chapter, we showcase two EEG-based emotion-driven applications that we developed, at the core of which lies the EEG-based emotion recognition algorithm. The potential application areas of the EEG-based emotion recognition algorithm are wider than what we have demonstrated here—the EEG-based emotion recognition algorithm can be integrated into many different

applications that need to be driven by emotions. The recognized emotion states can be the sole or additional input to other program that controls the flow of program executions. EEG-based emotion recognition also sees applicational potentials in fields like human factor study [179], emotion therapy [180], maritime [10, 179], shooter performance analysis [181], air traffic controller monitoring [9, 182], etc.

Chapter 7 Conclusions and Future Work

In Chapter 7, we summarize the thesis and envision some future directions to work on.

7.1 Conclusions

In this thesis, we study and investigate approaches to EEG-based emotion recognition using machine learning techniques. In Chapter 1 and Chapter 2, we cover the introduction, review the related works and introduce the fundamentals of EEG-based emotion recognition. After extensive literature review, we identify three research gaps listed below that have been lacking sufficient investigation in the field of EEG-based emotion recognition thus far, and try to answer these questions in the three subsequent chapters.

1. How can we reduce the need of frequent re-calibrations of the classifier on the same subject?

The state-of-the-art EEG-based emotion recognition algorithms personalize the classifier for each subject to achieve satisfactory recognition performance. Due to the volatile affective neural patterns, a classifier trained at an earlier time would perform rather poorly at a later time even on the same subject. Thus, the emotion recognition system needs frequent re-calibration over time. Existing studies [72, 73, 79-84, 87, 104, 105] focus on the best possible performance of the short-term emotion recognition and rarely address the issue of intra-subject variance or the long-term recognition performance. We devote Chapter 3 to the study of stable affective EEG features to address problems of intra-subject variance and how to improve the long-term recognition performance. We established that the recognition performance of a BCI will inevitably drop when it operates without re-calibration given the current technologies and methods. We proposed to use stable features to mitigate the degradation of the recognition performance to a lesser extent. Stable affective EEG features are defined as EEG features that gives consistent measurement of the same emotion on the same subject over the course time. Considering

the long run of an EEG-based emotion recognition system, when no re-calibration is possible or allowed, we hypothesize that such stable features should give superior recognition accuracy to unstable features. We have carried out an experiment to collect affective EEG data from multiple subjects over days, focusing on repeated measurements of the same emotions in the long course. To the best of our knowledge, such an affective EEG dataset is the first of its kind. We propose a novel stable feature selection algorithm to select the most stable features based on the intra-class correlation coefficient (ICC). Extensive comparisons of the recognition accuracy have been made between using the stable features selected by our proposed algorithm and using the state-of-the-art EEG features that are extensively used in the current literature [81, 84-87, 104, 106, 115, 117-119, 136, 138]. Experimental results have established the advantage of our proposed algorithm in the long run when no re-calibration is allowed on a conventional EEG-based emotion recognition system. Our experimental results show that stable features selected by our proposed algorithm outperform the best-performing of the state-of-the-art features by 0.62 % – 8.47 % on the training set, and by 0.23 % – 6.16 % on the test set in a four-emotion recognition task.

2. How can we improve the recognition accuracy of subject-independent EEG-based emotion recognition algorithm?

It has also been known that affective neural patterns are subject dependent, and that a classifier trained on one specific subject or a pool of subjects does not generalize well to other subjects. The state-of-the-art EEG-based emotion recognition systems circumvent this issue by tailoring the classifier to the subject of interest [72, 73, 79-84, 87, 104, 105]. However, there are occasions that the subject may not want to undergo the training/calibrating sessions, which could be tedious and laborious. From

a practical viewpoint, a subject-independent EEG-based emotion recognition system has the potential to eliminate user training. The question is how we can improve the recognition accuracy of a subject-independent system. We devote Chapter 4 to the investigation of applying transfer learning techniques to improve the performance of subject-independent EEG-based emotion recognition. Specifically, we apply domain adaptation methods [149, 152, 156, 159-161] to reduce the discrepancies of data distribution between different subjects. We hypothesize that each subject constitutes a domain by himself/herself. By applying domain adaptation methods, discrepancies between domains could be reduced and data are distributed more similarly between subjects, yielding improved performance in a subject-independent emotion recognition setting. We have carried out simulations of subject-independent emotion recognition on two public EEG datasets: DEAP and SEED. The effectiveness of different domain adaptation methods has been confirmed. MIDA and TCA are found to be the best performing methods, improving the classification performance by up to 9.88 % on DEAP and 20.66 % on SEED. We carry out a preliminary research in cross-dataset emotion recognition. Under this setting, the subjects that contribute training data to the classifier and the test subjects are from different datasets. The training data and the test data are collected under different experimental paradigms using different affective stimuli and EEG devices. These technical discrepancies make the subject-independent emotion recognition task even more challenging. However, we believe that an ideal EEG-based emotion recognition system should function independently of the subjects, EEG devices, and affective stimuli etc. Domain adaptation methods, specifically speaking, MIDA and TCA are found to be effective in improving the performance of subject-independent emotion recognition in the cross-dataset setting. Compared to baseline performance where no

adaptation method is used, MIDA and TCA can improve the recognition accuracy by 7.25 % – 13.40 %.

3. Can we extract discriminative affective EEG features on an unsupervised basis?

EEG feature extraction is one key step to successful emotion recognition. For a long time, affective EEG features have been hand-engineered by domain experts. While pertinent to the classification task, hand-engineered feature requires great amount of expertise. Recently, the renaissance of neural networks has reintroduced the possibility of unsupervised feature learning from raw data or relatively low-level feature. We devote Chapter 5 to the study of unsupervised EEG feature extraction using deep learning techniques. We hypothesize that the most discriminative spectral components pertinent to emotion recognition are subject-dependent and different from standard spectral bands such as delta, theta, alpha, and beta, etc., which have been extensively used in various EEG studies[41, 72, 79, 81, 82, 84, 104, 105]. We leverage autoencoder to learn salient frequency components from the PSD of the raw EEG signals on an unsupervised basis. We propose a novel feature extraction architecture that consists in clustering the hidden units into several groups based on the similarity of their weights, with added pooling neuron on top of each cluster. The proposed network structure extracts feature similar to spectral band power features, but without predefining the frequency band (delta, theta, alpha or beta). The recognition performance using features extracted by our proposed methods are benchmarked against that using standard power features. Experimental results show that our proposed method yields better accuracy than standard spectral power features by 4.37 % – 18.71 %. We also compare our proposed structure with other neural networks such as multilayer

perceptron (MLP), radial basis function neural network (RBF NN) and spiking neural network (SNN). Extensive comparisons show that our method outperforms MLP by 12.73 % – 13.82 %, RBF NN by 3.33 % – 5.66 %, and SNN by 5.64 % – 11.35 %.

After addressing the three research gaps, we then showcase some EEG-based emotion recognition applications in Chapter 6, with highlight in the "Emotion Avatar" application. The application could potentially be used for real-time emotion monitoring and may be beneficial to marketing surveyors, caregivers, etc. We stress that the potential application areas of EEG-based emotion recognition are wider than what we have demonstrated here, including but not limited to human factor study [179], emotion therapy [180], maritime [10, 179], shooter performance analysis [181], air traffic controller monitoring [9, 182], etc.

7.2 Future Work

We envision some future directions from three perspectives: multimodality, EEG hardware development and computational methods.

1. Multimodality

In our current work, we explore emotion recognition relying on the sole modality of EEG. While having excellent temporal resolution, EEG has relatively poor spatial resolution. The diameter of EEG electrodes is orders of magnitude larger than one single neuron, and that the area of an electrode covers more than 250000 neurons [183]. EEG reflects the summed electric activities propagated to the scalp and the source of the electric activities are not exactly from beneath where the electrode is placed. The limited number of electrodes that can be placed on the scalp also limits the spatial sampling resolution. The human skull may distort the EEG signals when being propagated through. Muscle movement may

produce also electric activities that contaminate the signal of EEG, resulting in low signal-to-noise ratio. To achieve more robust detection of emotion, it is well worth the investigation of multimodal fusion of several technologies, producing a **hybrid BCI** [184]. For example, Zheng *et al.* [185] have tried to combined EEG with eye tracking data to detect emotion. Multimodal approaches have also been used with success in other nonaffective BCI such as [186] for wheelchair control, [187] for orthosis control, and [188] for vigilance detection, etc. In the future, we will work on the development of a hybrid affective BCI.

2. EEG hardware development

In the current presentation of the thesis, we mainly focus on algorithmic design and proposal. We did not look into developing EEG equipment or hardwares in the present study. We reckon that off-the-shelf Emotiv EEG devices are affordable and sufficient for the present study, which are also used in many other EEG-related studies [189-193]. However, we also reckon that the form factor of the current EEG devices can be further improved, and we will look into the development of integrating EEG electrodes into the wearables, such as a safety helmet equipped with EEG electrodes to monitor machine operator's brain states.

3. Computational methods

We proposed a novel stable feature selection algorithm to tackle the intra-subject variance of neural pattern. We assume that stationary components pertinent to emotion recognition exist in EEG signals throughout a long course of time. By quantifying the EEG feature stability with ICC models and ranking the feature stability index, we greedily select the topmost stable feature. However, the stationary assumption barely holds true for EEG signals. This might explain why the improvement is slim. From

another viewpoint, we may try to adapt the classifier to the changing neural patterns, which suggests one future direction. For example, we may train a classifier with an initial dataset collected at the very first calibration session. In the subsequent real-time recognition session, if no calibration is carried out, we may use the data collected during real-time recognition session and the predicted/pseudo emotional labels to retrain the classifier and makes it adaptive to the changing neural patterns. The correctness of the predicted/pseudo labels are the key to a successful adaptation. We may use a classifier that outputs probabilities (e.g., logistic regression) and opt for the predicted/pseudo labels with a high confidence for inclusion in the recalibration set.

We investigated unsupervised EEG feature learning from PSD seen as a relatively lower-level feature using autoencoder. One future direction is to learn useful EEG features from raw EEG signals. It has proven effective to use deep learning techniques to directly learn higher-level feature from raw data in other fields such as image processing (e.g. [194-196], learn from raw pixels) and speech recognition (e.g., [197, 198], learn from time-series of speech). However, for EEG signals, unsupervised feature learning directly from raw EEG data remains a challenge. The low signal-to-noise nature of EEG signals may pose a hindrance to successfully learning useful representation on an unsupervised basis. Nevertheless, an initial study in [199] demonstrates the possibility to learn higher-level feature from raw EEG signals using deep neural network.

SNN, as the third generation of neural network [170], uses spikes to encode data in the network and takes into account of time delay in activating a neuron. Thus, it is said to be more faithful to the real biological neurons. Recently, SNN has shown a broad spectrum of applicational potentials in a number of EEG studies. For example, SNN has been widely applied to

recognize from EEG the familiarity/unfamiliarity for neuromarketing [200], attentional bias towards marketing stimuli [201, 202], substance addiction [171], Alzheimer disease [203, 204], Absence Epilepsy [205], treatment-resistant schizophrenia [206], mental tasks (letter composition / multiplication / counting / rotation) [207, 208], upper-limb activities of daily living (ADL) [209], music vs. noise perception [210], and so on. Although the referenced works here do not deal with emotion recognition directly, they can be borrowed and applied to EEG-based emotion recognition.

Reference

[1] T. J. La Vaque, "The History of EEG Hans Berger: Psychophysiologist. A Historical Vignette," *Journal of Neurotherapy,* vol. 3, no. 2, pp. 1-9, 1999.

[2] Emotiv. Available: http://www.emotiv.com

[3] Neurosky. Available: www.neurosky.com

[4] S. E. Kober *et al.*, "Specific effects of EEG based neurofeedback training on memory functions in post-stroke victims," *Journal of NeuroEngineering and Rehabilitation,* vol. 12, no. 1, pp. 107-119, 2015.

[5] G. R. Rozelle and T. H. Budzynski, "Neurotherapy for stroke rehabilitation: A single case study," *Biofeedback and Self-regulation,* vol. 20, no. 3, pp. 211-228, 1995.

[6] L. A. Nelson, "The role of biofeedback in stroke rehabilitation: past and future directions," *Topics in stroke rehabilitation,* vol. 14, no. 4, pp. 59-66, 2007.

[7] D. L. Trudeau, "EEG Biofeedback for Addictive Disorders—The State of the Art in 2004," *Journal of Adult Development,* vol. 12, no. 2, pp. 139-146, 2005.

[8] T. M. Sokhadze, R. L. Cannon, and D. L. Trudeau, "EEG Biofeedback as a Treatment for Substance Use Disorders: Review, Rating of Efficacy, and Recommendations for Further Research," *Applied Psychophysiology and Biofeedback,* vol. 33, no. 1, pp. 1-28, 2008.

[9] X. Hou *et al.*, "EEG-Based Human Factors Evaluation of Conflict Resolution Aid and Tactile User Interface in Future Air Traffic Control Systems," in *Advances in Human Aspects of Transportation*, 2017, pp. 885-897.

[10] Y. Liu *et al.*, "EEG-based Mental Workload and Stress Recognition of Crew Members in Maritime Virtual Simulator: A Case Study," in *2017 International Conference on Cyberworlds (CW)*, 2017, pp. 64-71.

[11] R. Ohme, D. Reykowska, D. Wiener, and A. Choromanska, "Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures," *Journal of Neuroscience, Psychology, and Economics,* vol. 2, no. 1, pp. 21-31, 2009.

[12] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. Prosad Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimedia Tools and Applications,* vol. 76, no. 18, pp. 19087-19111, 2017.

[13] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology,* vol. 113, no. 6, pp. 767-791, 2002.

[14] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces,* vol. 1, no. 2, pp. 66-84, 2014.

[15] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing,* vol. 13, no. 2, pp. 293-303, 2005.

[16] K. P. Truong and D. A. van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Communication,* vol. 49, no. 2, pp. 144-158, 2007.

[17] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 10, pp. 1683-1699, 2007.

[18] R. Plutchik, "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.

[19] A. Mehrabian, "Framework for a comprehensive description and measurement of emotional states," *Genetic, social, and general psychology monographs,* vol. 121, no. 3, pp. 339-361, 1995.

[20] A. Mehrabian, "Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology,* vol. 14, no. 4, pp. 261-292, 1996.

[21] M. M. Bradley and P. J. Lang, "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual," University of Florida2007.

[22] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," University of Florida2008.

[23] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE Transactions on Affective Computing,* vol. 3, no. 1, pp. 18-31, 2012.

[24] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry,* vol. 25, no. 1, pp. 49-59, 1994.

[25] S. R. Benbadis *et al.*, *Handbook of EEG interpretation.* Demos Medical Publishing, 2007.

[26] B. J. Fisch and R. Spehlmann, *Fisch and Spehlmann's EEG primer : basic principles of digital and analog EEG.* Elsevier, 1999.

[27] G. H. Klem, H. O. Lüders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation," *Electroencephalography and clinical neurophysiology,* vol. 52, pp. 3-6, 1999.

[28] A. E. Society, "American electroencephalographic society guidelines for standard electrode position nomenclature," *Journal of Clinical Neurophysiology,* vol. 8, no. 2, pp. 200-202, 1991.

[29] F. Amzica and M. Steriade, "Electrophysiological correlates of sleep delta waves," *Electroencephalography and clinical neurophysiology,* vol. 107, no. 2, pp. 69-83, 1998.

[30] F. Torres and C. Anderson, "The normal EEG of the human newborn," *Journal of Clinical Neurophysiology,* vol. 2, no. 2, pp. 89-104, 1985.

[31] P. Gloor, G. Ball, and N. Schaul, "Brain lesions that produce delta waves in the EEG," *Neurology,* vol. 27, no. 4, pp. 326-333, 1977.

[32] A. M. Strijkstra, D. G. M. Beersma, B. Drayer, N. Halbesma, and S. Daan, "Subjective sleepiness correlates negatively with global alpha (8–

12 Hz) and positively with central frontal theta (4–8 Hz) frequencies in the human resting awake electroencephalogram," *Neuroscience Letters,* vol. 340, no. 1, pp. 17-20, 2003.

[33] Y. Liu, O. Sourina, and X. Hou, "Neurofeedback Games to Improve Cognitive Abilities," in *2014 International Conference on Cyberworlds (CW)*, 2014, pp. 161-168.

[34] K. Hashi, S. Nishimura, A. Kondo, K. Nin, and S. Jac-Hong, "The EEG in normal pressure hydrocephalus," *Acta Neurochirurgica,* vol. 33, no. 1, pp. 23-35, 1976.

[35] O. N. Markand, "Alpha rhythms," *Journal of Clinical Neurophysiology,* vol. 7, no. 2, pp. 163-190, 1990.

[36] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage,* vol. 31, no. 1, pp. 153-159, 2006.

[37] L. Lehtelä, R. Salmelin, and R. Hari, "Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex," *Neuroscience letters,* vol. 222, no. 2, pp. 111-114, 1997.

[38] M. Rangaswamy *et al.*, "Beta power in the EEG of alcoholics," *Biological psychiatry,* vol. 52, no. 8, pp. 831-842, 2002.

[39] N. Kanayama, A. Sato, and H. Ohira, "Crossmodal effect with rubber hand illusion and gamma - band activity," *Psychophysiology,* vol. 44, no. 3, pp. 392-402, 2007.

[40] C. Tallon-Baudry, A. Kreiter, and O. Bertrand, "Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans," *Visual Neuroscience,* vol. 16, no. 3, pp. 449-459, 1999.

[41] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing,* vol. 5, no. 3, pp. 327-339, 2014.

[42] Y. Liu, "EEG-based Emotion Recognition for Real-time Applications," Ph.D. Thesis, Nanyang Technological University, 2014.

[43] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology,* vol. 44, no. 2, pp. 293-304, 2007.

[44] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Affective Computing and Intelligent Interaction*, 2011, pp. 457-466.

[45] L. I. Aftanas, A. A. Varlamov, S. V. Pavlov, V. P. Makhnev, and N. V. Reva, "Affective picture processing: Event-related synchronization within individually defined human theta band is modulated by valence dimension," *Neuroscience Letters,* vol. 303, no. 2, pp. 115-118, 2001.

[46] A. Keil, M. M. Müller, T. Gruber, C. Wienbruch, M. Stolarova, and T. Elbert, "Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials," *Clinical neurophysiology,* vol. 112, no. 11, pp. 2057-2068, 2001.

[47] K. S. Park, H. Choi, K. J. Lee, J. Y. Lee, K. O. An, and E. J. Kim, "Emotion recognition based on the asymmetric left and right activation," *International Journal of Medicine and Medical Sciences,* vol. 3, no. 6, pp. 201-209, 2011.

[48] R. Degabriele, J. Lagopoulos, and G. Malhi, "Neural correlates of emotional face processing in bipolar disorder: an event-related potential study," *Journal of affective disorders,* vol. 133, no. 1, pp. 212-220, 2011.

[49] L. J. Trainor and L. A. Schmidt, "Processing emotions induced by music," *The cognitive neuroscience of music,* pp. 310-324, 2003.

[50] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition and Emotion,* vol. 15, no. 4, pp. 487-500, 2001.

[51] N. A. Jones and N. A. Fox, "Electroencephalogram asymmetry during emotionally evocative films and its relation to positive and negative affectivity," *Brain and Cognition,* vol. 20, no. 2, pp. 280-299, 1992.

[52] K. Trochidis and E. Bigand, "EEG-based emotion perception during music listening," in *12th International Conference on Music Perception and Cognition*, 2012, pp. 1018-1021.

[53] R. J. Davidson, "Cerebral asymmetry and emotion: Conceptual and methodological conundrums," *Cognition and Emotion,* vol. 7, no. 1, pp. 115-138, 1993.

[54] E. Harmon-Jones, "Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity," *Biological psychology,* vol. 67, no. 1, pp. 51-76, 2004.

[55] C. P. Niemic, "Studies of Emotion: A Theoretical and Empirical Review of Psychophysiological Studies of Emotion.," *Journal of Undergraduate Research,* vol. 1, pp. 15-18, 2004.

[56] W. Heller, "Neuropsychological Mechanisms of Individual Differences in Emotion, Personality, and Arousal," *Neuropsychology,* vol. 7, no. 4, pp. 476-489, 1993.

[57] A. Choppin, "EEG-based human interface for disabled individuals: Emotion expression with neural networks," Master Thesis, Tokyo Institute of Technology, 2000.

[58] D. O. Bos, "EEG-based emotion recognition," *The Influence of Visual and Auditory Stimuli,* pp. 1-17, 2006.

[59] N. Martini *et al.*, "The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity," *NeuroImage,* vol. 60, no. 2, pp. 922-932, 2012.

[60] V. Miskovic and L. A. Schmidt, "Cross-regional cortical synchronization during affective image viewing," *Brain research,* vol. 1362, pp. 102-111, 2010.

[61] A. Kemp, M. Gray, P. Eide, R. Silberstein, and P. Nathan, "Steady-state visually evoked potential topography during processing of emotional valence in healthy subjects," *NeuroImage,* vol. 17, no. 4, pp. 1684-1692, 2002.

[62] M. Balconi 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.

[63] G. Stenberg, "Personality and the EEG: Arousal and emotional arousability," *Personality and Individual Differences,* vol. 13, no. 10, pp. 1097-1113, 1992.

[64] M. Mikolajczak, K. Bodarwé, O. Laloyaux, M. Hansenne, and D. Nelis, "Association between frontal EEG asymmetries and emotional intelligence among adults," *Personality and Individual Differences,* vol. 48, no. 2, pp. 177-181, 2010.

[65] H. Kawano, A. Seo, Z. G. Doborjeh, N. Kasabov, and M. G. Doborjeh, "Analysis of similarity and differences in brain activities between perception and production of facial expressions using EEG data and the NeuCube spiking neural network architecture," in *International conference on neural information processing*, 2016, pp. 221-227.

[66] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 4, no. 2, pp. R1-R13, 2007.

[67] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of neural engineering,* vol. 15, no. 3, p. 031005, 2018.

[68] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley, New York, 1973.

[69] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks,* vol. 10, no. 5, pp. 988-999, 1999.

[70] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Transactions on Biomedical engineering,* vol. 51, no. 6, pp. 1057-1061, 2004.

[71] G. Pfurtscheller and F. L. Da Silva, "EEG-Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS)," in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*: Lippincott Williams and Wilkins, 2005.

[72] G. Chanel, J. J. M. 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, Aug 2009.

[73] M. Murugappan, R. Nagarajan, and S. Yaacob, "Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals," *Journal of Medical and Biological Engineering,* vol. 31, no. 1, pp. 45-51, 2011.

[74] L. Bozhkov, P. Georgieva, I. Santos, A. Pereira, and C. Silva, "EEG-based Subject Independent Affective Computing Models," *Procedia Computer Science,* vol. 53, pp. 375-382, 2015.

[75] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology,* vol. 2, no. 3, p. 27, 2011.

[76] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado, "Ensemble of SVMs for improving brain computer interface P300 speller performances," in *International conference on artificial neural networks*, 2005, pp. 45-50.

[77] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Transactions on neural systems and rehabilitation engineering,* vol. 11, no. 2, pp. 141-144, 2003.

[78] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Advances in neural information processing systems*, 2002, pp. 157-164.

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

[80] K. Schaaff, "EEG-based Emotion Recognition," Master Thesis, Institute of Algorithms and Cognitive Systems, University of Karlsruhe, 2008.

[81] K. Schaaff and T. Schultz, "Towards an EEG-based emotion recognizer for humanoid robots," in *IEEE International Workshop on Robot and Human Interactive Communication*, 2009, pp. 792-796.

[82] Y. P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering,* vol. 57, no. 7, pp. 1798-1806, 2010.

[83] M. Li and B. Lu, "Emotion classification based on gamma-band EEG," in *IEEE International Conference on Engineering in Medicine and Biology Society*, 2009, pp. 1223-1226.

[84] X.-W. Wang, D. Nie, and B.-L. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *Neural Information Processing*, 2011, pp. 734-743.

[85] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, no. 5, pp. 737-746, 2011.

[86] P. C. Petrantonakis and L. J. Hadjileontiadis, "Adaptive Emotional Information Retrieval From EEG Signals in the Time-Frequency Domain," *IEEE Transactions on Signal Processing,* vol. 60, no. 5, pp. 2604-2616, 2012.

[87] Y. Liu and O. Sourina, "EEG Databases for Emotion Recognition," in *2013 Internation Conference on Cyberworlds (CW)*, Yokohama, 2013, pp. 302-309.

[88] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić, "Evaluating Classifiers for Emotion Recognition Using EEG," in *International Conference on Augmented Cognition*, 2013, pp. 492-501.

[89] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development,* vol. 7, no. 3, pp. 162-175, 2015.

[90] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-Based EEG Sensor Selection Methods for Emotion Recognition," *Sensors,* vol. 16, no. 10, p. 1558, 2016.

[91] R. M. Mehmood and H. J. Lee, "A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns," *Computers and Electrical Engineering,* vol. 53, pp. 444-457, 2016.

[92] W. L. Zheng, J. Y. Zhu, and B. L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," *IEEE Transactions on Affective Computing,* pp. 1-15, 2017. In Press.

[93] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu, "Transfer components between subjects for EEG-based emotion recognition," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 917-922.

[94] H. Candra *et al.*, "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," in *2015 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 7250-7253.

[95] H. Candra, M. Yuwono, A. Handojoseno, R. Chai, S. Su, and H. T. Nguyen, "Recognizing emotions from EEG subbands using wavelet analysis," in *2015 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 6030-6033.

[96] W.-L. Zheng and B.-L. Lu, "Personalizing EEG-based affective models with transfer learning," in *25th International Joint Conference on Artificial Intelligence*, 2016, pp. 2732-2738.

[97] X. Chai *et al.*, "A Fast, Efficient Domain Adaptation Technique for Cross-Domain Electroencephalography (EEG)-Based Emotion Recognition," *Sensors,* vol. 17, no. 5, p. 1014, 2017.

[98] X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, and Y. Li, "Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition," *Computers in biology and medicine,* vol. 79, pp. 205-214, 2016.

[99] J. Machado, A. Balbinot, and A. Schuck, "A study of the Naive Bayes classifier for analyzing imaginary movement EEG signals using the Periodogram as spectral estimator," in *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, 2013, pp. 1-4.

[100] J. Machado and A. Balbinot, "Executed movement using EEG signals through a Naive Bayes classifier," *Micromachines,* vol. 5, no. 4, pp. 1082-1105, 2014.

[101] F. C. Pampel, *Logistic regression: A primer*. Sage Publications, 2000.

[102] R. Tomioka, K. Aihara, and K.-R. Müller, "Logistic regression for single trial EEG classification," in *Advances in neural information processing systems*, 2007, pp. 1377-1384.

[103] A. Alkan, E. Koklukaya, and A. Subasi, "Automatic seizure detection in EEG using logistic regression and artificial neural network," *Journal of Neuroscience Methods,* vol. 148, no. 2, pp. 167-176, 2005.

[104] K. Ishino and M. Hagiwara, "A feeling estimation system using a simple electroencephalograph," in *IEEE International Conference on Systems, Man and Cybernetics*, 2003, vol. 5, pp. 4204-4209.

[105] M. Kwon, J.-S. Kang, and M. Lee, "Emotion classification in movie clips based on 3D fuzzy GIST and EEG signal analysis," in *International Winter Workshop on Brain-Computer Interface (BCI)*, 2013, pp. 67-68.

[106] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 2, pp. 186-197, 2010.

[107] L. Brown, B. Grundlehner, and J. Penders, "Towards wireless emotional valence detection from EEG," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 2188-2191: IEEE.

[108] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 3, pp. 589-597, 2010.

[109] S. A. Hosseini, 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*, 2010, pp. 60-63.

[110] O. Al Zoubi, M. Awad, and N. K. Kasabov, "Anytime multipurpose emotion recognition from EEG data using a Liquid State Machine based framework," *Artificial intelligence in medicine,* vol. 86, pp. 1-8, 2018.

[111] W. L. Zheng, J. Y. Zhu, Y. Peng, and B. L. Lu, "EEG-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1-6.

[112] Mathworks. Available: www.mathworks.com

[113] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena,* vol. 31, no. 2, pp. 277-283, 1988.

[114] Y. Liu and O. Sourina, "Real-Time Fractal-Based Valence Level Recognition from EEG," *Transactions on Computational Science XVIII,* vol. 7848, pp. 101-120, 2013.

[115] Y. Liu and O. Sourina, "Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm," *Transactions on Computational Science XXIII,* pp. 199-223, 2014.

[116] B. Kedem and E. Slud, "Time series discrimination by higher order crossings," *The Annals of Statistics,* pp. 786-794, 1982.

[117] F. Feradov and T. Ganchev, "Ranking of EEG time-domain features on the negative emotions recognition task," *Annual Journal of Electronics,* vol. 9, pp. 26-29, 2015.

[118] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and clinical neurophysiology,* vol. 29, no. 3, pp. 306-310, 1970.

[119] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *15th European Signal Processing Conference*, 2007, pp. 1241-1245.

[120] R. Horlings, D. Datcu, and L. J. Rothkrantz, "Emotion recognition using brain activity," in *International Conference on Computer Systems and Technologies*, 2008, pp. 1-6.

[121] T. Gasser, P. Bächer, and H. Steinberg, "Test-retest reliability of spectral parameters of the EEG," *Electroencephalography and clinical neurophysiology,* vol. 60, no. 4, pp. 312-319, 1985.

[122] T. Gasser, C. Jennen-Steinmetz, and R. Verleger, "EEG coherence at rest and during a visual task in two groups of children," *Electroencephalography and clinical neurophysiology,* vol. 67, no. 2, pp. 151-158, 1987.

[123] M. Salinsky, B. Oken, and L. Morehead, "Test-retest reliability in EEG frequency analysis," *Electroencephalography and clinical neurophysiology,* vol. 79, no. 5, pp. 382-392, 1991.

[124] A. Kondacs and M. Szabó, "Long-term intra-individual variability of the background EEG in normals," *Clinical Neurophysiology,* vol. 110, no. 10, pp. 1708-1716, 1999.

[125] S. Gudmundsson, T. P. Runarsson, S. Sigurdsson, G. Eiriksdottir, and K. Johnsen, "Reliability of quantitative EEG features," *Clinical Neurophysiology,* vol. 118, no. 10, pp. 2162-2171, 2007.

[126] J. J. Allen, H. L. Urry, S. K. Hitt, and J. A. Coan, "The stability of resting frontal electroencephalographic asymmetry in depression," *Psychophysiology,* vol. 41, no. 2, pp. 269-280, 2004.

[127] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological methods,* vol. 1, no. 1, pp. 30-46, 1996.

[128] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Stability of features in real-time EEG-based emotion recognition algorithm," in *2014 International Conference on Cyberworlds (CW)*, 2014, pp. 137-144.

[129] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer,* vol. 32, no. 3, pp. 347-358, 2016.

[130] X. Hou *et al.*, "CogniMeter: EEG-based brain states monitoring," *Transactions on Computational Science XXVIII,* pp. 108-126, 2016.

[131] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Stable feature selection for EEG-based emotion recognition," in *2018 Internation Conference on Cyberworlds (CW)*, 2018, pp. 1-8. In Press.

[132] J. I. Ekandem, T. A. Davis, I. Alvarez, M. T. James, and J. E. Gilbert, "Evaluating the ergonomics of BCI devices for research and experimentation," *Ergonomics,* vol. 55, no. 5, pp. 592-598, 2012.

[133] G. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than random: a closer look on BCI results," *International Journal of Bioelectromagnetism,* vol. 10, no. 1, pp. 52-55, 2008.

[134] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *Journal of Neuroscience Methods,* vol. 250, pp. 126-136, 2015.

[135] M. E. Bouton, *Learning and behavior: A contemporary synthesis*. Sinauer Associates, 2007.

[136] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence,* vol. 23, no. 10, pp. 1175-1191, 2001.

[137] K. Takahashi and A. Tsukaguchi, "Remarks on emotion recognition from multi-modal bio-potential signals," in *IEEE International Conference on Systems, Man and Cybernetics*, 2003, vol. 2, pp. 1654-1659.

[138] Y. Liu and O. Sourina, "EEG-based Dominance Level Recognition for Emotion-enabled Interaction," in *IEEE International Conference on Multimedia and Expo*, Melbourne, 2012, pp. 1039-1044.

[139] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalography and clinical neurophysiology,* vol. 78, no. 3, pp. 252-259, 1991.

[140] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain-computer communication," *Electroencephalography and clinical Neurophysiology,* vol. 90, no. 6, pp. 444-449, 1994.

[141] T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer, "Biofeedback of slow cortical potentials. I," *Electroencephalography and Clinical Neurophysiology,* vol. 48, no. 3, pp. 293-301, 1980.

[142] N. Birbaumer *et al.*, "A spelling device for the paralysed," *Nature,* vol. 398, no. 6725, p. 297, 1999.

[143] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PloS one,* vol. 3, no. 8, p. e2967, 2008.

[144] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural networks,* vol. 22, no. 9, pp. 1305-1312, 2009.

[145] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters,* vol. 16, no. 8, pp. 683-686, 2009.

[146] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on biomedical Engineering,* vol. 58, no. 2, pp. 355-362, 2011.

[147] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[148] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine,* vol. 11, no. 1, pp. 20-31, 2016.

[149] K. Yan, L. Kou, and D. Zhang, "Learning Domain-Invariant Subspace Using Domain Features and Independence Maximization," *IEEE transactions on cybernetics,* vol. 48, no. 1, pp. 288-299, 2017.

[150] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2013, pp. 6627-6630.

[151] A. Gretton, O. Bousquet, A. Smola, and B. Scholkopf, "Measuring statistical dependence with Hilbert-Schmidt norms," in *International Conference on Algorithmic Learning Theory*, 2005, vol. 16, pp. 63-78.

[152] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation,* vol. 10, no. 5, pp. 1299-1319, 1998.

[153] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 1999, pp. 41-48.

[154] L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt, "Feature selection via dependence maximization," *Journal of Machine Learning Research,* vol. 13, no. 1, pp. 1393-1434, 2012.

[155] C.-W. Seah, Y.-S. Ong, and I. W. Tsang, "Combating negative transfer from predictive distribution differences," *IEEE transactions on cybernetics,* vol. 43, no. 4, pp. 1153-1165, 2013.

[156] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks,* vol. 22, no. 2, pp. 199-210, 2011.

[157] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in neural information processing systems*, 2007, pp. 513-520.

[158] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A Hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*, 2007, pp. 13-31.

[159] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2960-2967.

[160] Y. Shi and F. Sha, "Information-theoretical learning of discriminative clusters for unsupervised domain adaptation," *arXiv:1206.6438,* 2012.

[161] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2066-2073.

[162] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain Adaptation Techniques for EEG-based Emotion Recognition: A Comparative Study on Two Public Datasets," *IEEE Transactions on Cognitive and Developmental Systems,* pp. 1-10, 2018. In Press.

[163] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich, "To transfer or not to transfer," in *NIPS 2005 workshop on transfer learning*, 2005, vol. 898, pp. 1-4.

[164] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[165] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. Muller-Putz, "Unsupervised Feature Learning for EEG-based Emotion Recognition," in *2017 International Conference on Cyberworlds (CW)*, 2017, pp. 182-185.

[166] C. Bishop and C. M. Bishop, *Neural networks for pattern recognition*. Oxford University Press, 1995.

[167] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks," Royal Signals and Radar Establishment,1988.

[168] N. K. Kasabov, "NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data," *Neural Networks,* vol. 52, pp. 62-76, 2014.

[169] N. Kasabov *et al.*, "Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: design methodology and selected applications," *Neural Networks,* vol. 78, pp. 1-14, 2016.

[170] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural networks,* vol. 10, no. 9, pp. 1659-1671, 1997.

[171] M. G. Doborjeh, G. Y. Wang, N. K. Kasabov, R. Kydd, and B. Russell, "A spiking neural network methodology and system for learning and comparative analysis of EEG data from healthy versus addiction treated versus addiction not treated subjects," *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 9, pp. 1830-1841, 2016.

[172] L. Koessler *et al.*, "Automated cortical projection of EEG sensors: anatomical correlation via the international 10–10 system," *Neuroimage,* vol. 46, no. 1, pp. 64-72, 2009.

[173] *NeuCube Manual*. Available: https://kedri.aut.ac.nz/neucube_manual

[174] N. K. Kasabov, *Evolving connectionist systems: the knowledge engineering approach*. Springer, 2007.

[175] Haptek. Available: http://www.haptek.com

[176] Z. Lan, Y. Liu, O. Sourina, and L. Wang, "Real-time EEG-based user's valence monitoring," in *2015 International Conference on Information, Communications and Signal Processing (ICICS)*, 2015, pp. 1-5.

[177] W. L. Lim, O. Sourina, Y. Liu, and L. Wang, "EEG-based mental workload recognition related to multitasking," in *2015 International Conference on Information, Communications and Signal Processing (ICICS)*, 2015, pp. 1-4.

[178] X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang, and W. Mueller-Wittig, "EEG Based Stress Monitoring," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 3110-3115.

[179] Y. Liu, O. Sourina, H. P. Liew, H. S. Salem, and E. Ang, "Human Factors Evaluation in Maritime Virtual Simulators Using Mobile EEG-Based Neuroimaging," *Transdisciplinary Engineering: A Paradigm Shift,* vol. 5, pp. 261-268, 2017.

[180] O. Sourina, Y. Liu, and M. K. Nguyen, "Real-time EEG-based emotion recognition for music therapy," *Journal on Multimodal User Interfaces,* vol. 5, no. 1, pp. 27-35, 2012.

[181] Y. Liu, S. C. H. Subramaniam, O. Sourina, E. Shah, J. Chua, and K. Ivanov, "Neurofeedback Training for Rifle Shooters to Improve Cognitive Ability," in *2017 International Conference on Cyberworlds (CW)*, 2017, pp. 186-189.

[182] F. Trapsilawati *et al.*, "Perceived and Physiological Mental Workload and Emotion Assessments in En-Route ATC Environment: A Case Study," *Transdisciplinary Engineering: A Paradigm Shift,* vol. 5, pp. 420-427, 2017.

[183] R. Nusslock, C. B. Young, N. Pornpattananangkul, and K. S. Damme, "Neurophysiological and Neuroimaging Techniques," *The Encyclopedia of Clinical Psychology,* pp. 1-9, 2014.

[184] G. Pfurtscheller *et al.*, "The hybrid BCI," *Frontiers in Neuroscience,* vol. 4, no. 30, pp. 1-11, 2010.

[185] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *2014 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5040-5043.

[186] J. Li, H. Ji, L. Cao, R. Gu, B. Xia, and Y. Huang, "Wheelchair Control Based on Multimodal Brain-Computer Interfaces," in *International Conference on Neural Information Processing*, 2013, pp. 434-441.

[187] G. Pfurtscheller, T. Solis-Escalante, R. Ortner, P. Linortner, and G. R. Muller-Putz, "Self-Paced Operation of an SSVEP-Based Orthosis With and Without an Imagery-Based "Brain Switch": A Feasibility Study Towards a Hybrid BCI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 18, no. 4, pp. 409-414, 2010.

[188] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of neural engineering,* vol. 14, no. 026017, pp. 1-14, 2017.

[189] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the Emotiv Epoc headset for P300-based applications," *Biomedical engineering online,* vol. 12, no. 1, p. 56, 2013.

[190] Y. Liu *et al.*, "Implementation of SSVEP based BCI with Emotiv EPOC," in *2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems*, 2012, pp. 34-37.

[191] W. A. Jang, S. M. Lee, and D. H. Lee, "Development BCI for individuals with severely disability using EMOTIV EEG headset and robot," in *2014 International Winter Workshop on Brain-Computer Interface (BCI)*, 2014, pp. 1-3.

[192] S. Grude, M. Freeland, C. Yang, and H. Ma, "Controlling mobile Spykee robot using Emotiv neuro headset," in *2013 32nd Chinese Control Conference (CCC)*, 2013, pp. 5927-5932.

[193] A. S. Elsawy, S. Eldawlatly, M. Taher, and G. M. Aly, "Performance analysis of a Principal Component Analysis ensemble classifier for Emotiv headset P300 spellers," in *2014 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5032-5035.

[194] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation,* vol. 18, no. 7, pp. 1527-1554, 2006.

[195] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[196] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[197] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine,* vol. 29, no. 6, pp. 82-97, 2012.

[198] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 6645-6649.

[199] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Advances in Artificial Neural Systems,* vol. 2012, pp. 1-9, 2012.

[200] Z. G. Doborjeh, N. Kasabov, M. G. Doborjeh, and A. Sumich, "Modelling peri-perceptual brain processes in a deep learning spiking neural network architecture," *Scientific reports,* vol. 8, no. 8912, pp. 1-13, 2018.

[201] Z. G. Doborjeh, M. G. Doborjeh, and N. Kasabov, "Efficient recognition of attentional bias using EEG data and the NeuCube evolving spatio-temporal data machine," in *International Conference on Neural Information Processing*, 2016, pp. 645-653.

[202] Z. G. Doborjeh, M. G. Doborjeh, and N. Kasabov, "Attentional bias pattern recognition in spiking neural networks from spatio-temporal EEG data," *Cognitive Computation,* vol. 10, no. 1, pp. 35-48, 2018.

[203] E. Capecci, Z. G. Doborjeh, N. Mammone, F. La Foresta, F. C. Morabito, and N. Kasabov, "Longitudinal study of alzheimer's disease degeneration through EEG data analysis with a NeuCube spiking neural network model," in *2016 International Joint Conference on Neural Networks*, 2016, pp. 1360-1366.

[204] E. Capecci, F. C. Morabito, M. Campolo, N. Mammone, D. Labate, and N. Kasabov, "A feasibility study of using the neucube spiking neural network architecture for modelling alzheimer's disease eeg data," in *Advances in neural networks: Computational and theoretical issues*: Springer, 2015, pp. 159-172.

[205] E. Capecci *et al.*, "Modelling absence epilepsy seizure data in the neucube evolving spiking neural network architecture," in *2015 International Joint Conference on Neural Networks*, 2015, pp. 1-8.

[206] C. McNabb *et al.*, "Classification of people with treatment-resistant and ultra-treatment-resistant schizophrenia using resting-state EEG and the NeuCube," in *Schizophrenia bulletin*, 2015, vol. 41, pp. S233-S234.

[207] N. Kasabov and E. Capecci, "Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes," *Information Sciences,* vol. 294, pp. 565-575, 2015.

[208] S. Schliebs, E. Capecci, and N. Kasabov, "Spiking neural network for on-line cognitive activity classification based on EEG data," in *International Conference on Neural Information Processing*, 2013, pp. 55-62.

[209] J. Hu, Z.-G. Hou, Y.-X. Chen, N. Kasabov, and N. Scott, "EEG-based classification of upper-limb ADL using SNN for active robotic rehabilitation," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2014, pp. 409-414.

[210] N. Kasabov, J. Hu, Y. Chen, N. Scott, and Y. Turkova, "Spatio-temporal EEG data classification in the NeuCube 3D SNN environment: methodology and examples," in *International Conference on Neural Information Processing*, 2013, pp. 63-69.

Author's Publication List

Journal

[1] **Z. Lan**, O. Sourina, L. Wang, R. Scherer, G. R. Müller-Putz, "Domain adaptation techniques for cross EEG dataset emotion classification: A comparative study on two public datasets," *IEEE Transaction on Cognitive and Developmental Systems,* pp. 1-10, 2018. In Press.

[2] **Z. Lan**, O. Sourina, L. Wang and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer*, vol. 32, no. 3, pp. 347-358, 2016.

[3] X. Hou, Y. Liu, W. L. Lim, **Z. Lan**, O. Sourina, W. Mueller-Wittig and L. Wang, "CogniMeter: EEG-Based brain states monitoring," *Transaction on Computational Science XXVIII*, vol. 28, no. 9590, pp. 108-126, 2016.

Conference

[4] **Z. Lan**, O. Sourina, L. Wang and Y. Liu, "Stability of Features in Real-Time EEG-based Emotion Recognition Algorithm," in *2014 International Conference on Cyberworlds (CW),* 2014, pp. 137-144.

[5] **Z. Lan**, Y. Liu, O. Sourina and L. Wang, "Real-time EEG-based user's valence monitoring," in *2015 10th International Conference on Information, Communications and Signal Processing (ICICS)*, 2015, pp. 1-5.

[6] **Z. Lan**, G. R. Müller-Putz, L. Wang, Y. Liu, O. Sourina and R. Scherer, "Using Support Vector Regression to estimate valence level from EEG," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 2558-2563.

[7] **Z. Lan**, O. Sourina, L. Wang, R. Scherer, G. R. Müller-Putz, "Unsupervised Feature Extraction for EEG-based Emotion Recognition," in *2017 International Conference on Cyberworlds (CW)*, 2017, pp. 182-185.

[8] **Z. Lan**, O. Sourina, L. Wang, Y. Liu, "Stable feature selection for EEG-based Emotion Recognition," in *2018 International Conference on Cyberworlds (CW)*, 2018, pp. 1-8. In Press.

[9] Y. Liu, **Z. Lan**, G. H. H. Khoo, H. K. H. Li, O. Sourina, "EEG-based evaluation of mental fatigue using machine learning algorithms," in *2018 International Conference on Cyberworlds (CW)*, 2018, pp. 1-4. In Press.

[10] Y. Liu, **Z. Lan**, O. Sourina, S. P. H. Liu, G. Krishnan, D. Konovessis and H. E. Ang, "EEG-based Cadets Training and Performance Assessment System in Maritime Virtual Simulator," in *2018 International Conference on Cyberworlds (CW)*, 2018, pp. 1-8. In Press.

Appendix A Experiment Materials

I Affective Sound Clips

In our emotion induction experiment introduced in 0, affective sound stimuli were used to induce targeted emotions on subjects. Each audio file was 76 seconds in length, composed by 16 seconds silent part followed by 60 seconds audio part. The 60 seconds audio part comprised 10 sound clips from IADS. The selected sound clips from IADS to create the audio files are tabulated in Table A-1.

II Self-Assessment Questionnaire

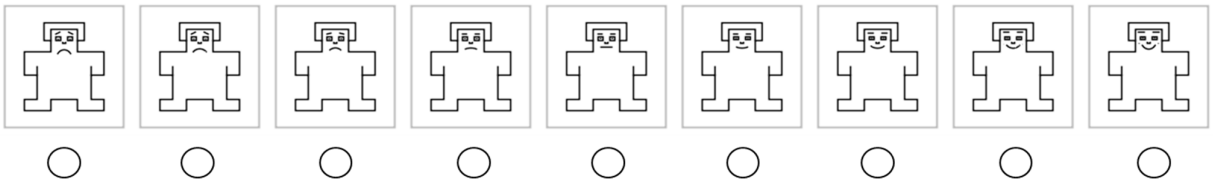
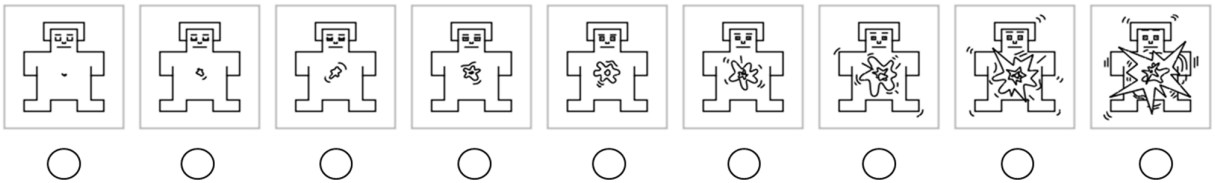
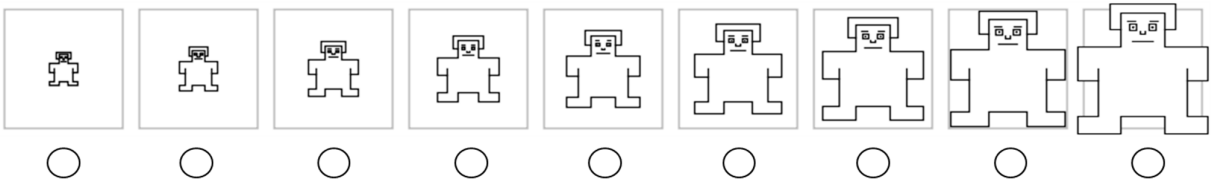
The self-assessment questionnaire used in the data collection experiment in Section 3.3.1 is displayed in Figure A.1.

Table A-1 Selected IADS sound clips for emotion induction experiment.

Targeted Emotion	Sound Clip ID in IADS	Sound Clip Description	Targeted Emotion	Sound Clip ID in IADS	Sound Clip Description
	150	Seagull		275	Screaming
	151	Robin's Chirping		276	Female Screaming 2
	171	Country Night		277	Female Screaming 3
	172	Brook		279	Attack 1
PLH	377	Rain	NHL	284	Attack 3
(pleasant)	809	Harp	(frightened)	285	Attack 2
	810	Beethoven's Music		286	Victim
	812	Choir		290	Fight
	206	Shower		292	Male Screaming
	270	Whistling		422	Tire Skids
	109	Carousel		116	Buzzing
	254	Video Game		243	Couple Sneeze
	315	Applause		251	Nose Blow
	716	Slot Machine		380	Jack Hammer
PHH	601	Colonial Music	NHH	410	Helicopter 2
(happy)	367	Casino 2	(angry)	423	Injury
	366	Casino 1		702	Belch
	815	Rock & Roll Music		706	War
	817	Bongos		729	Paper 2
	820	Funk Music		910	Electricity

Note: Please indicate ONLY ONE choice in each row. Please indicate your choice by shading the corresponding circle.

Name: **Gender:** **Age:** **Date of Experiment:**

Valence	
Arousal	
Dominance	
Preference/ Liking	

Familiarity	1	2	3	4	5	6	7	8	9
	Very Unfamiliar ○	In between ○	Unfamiliar ○	In between ○	Neutral ○	In between ○	Familiar ○	In between ○	Very Familiar ○

Word Description	Pleasant ○	Surprised ○	Happy ○	Protected ○	Frightened ○	Sad ○	Angry ○	Unconcerned ○	If none applies, please indicate your word description here:

Figure A.1 Self-assessment questionnaire.

III Stable Feature Ranking

Table A-2 and Table A-3 below present the subject-independent and subject-dependent feature stability ranking, respectively, for the experiment introduced in Section 3.3.4. In both tables, the feature name consists of two parts joined by an underscore, e.g., x_y, which denotes that feature x is extracted out of EEG channel y. The notation hoc$i$ denotes HOC feature of order i, and stat$i$ denotes the ith statistic feature. The notation fd, se, actvt, mblty and cpxty denotes fractal dimension, signal energy, activity, mobility and complexity, respectively. Delta, theta, alpha and beta denotes the spectral band power feature from the named frequency range, respectively.

Table A-2 Subject-independent ICC score-based feature stability ranking.

Rank	Feature	Score	Rank	Feature	Score	Rank	Feature	Score	Rank	Feature	Score
1	stat3_T7	0.2642	63	hoc26_T7	0.0236	125	hoc20_FC5	-0.0015	187	hoc15_F4	-0.0145
2	stat5_T7	0.2527	64	beta_F7	0.0235	126	hoc16_AF3	-0.0018	188	theta_F7	-0.0146
3	hoc1_T7	0.1761	65	hoc27_AF3	0.0235	127	hoc8_AF3	-0.0018	189	hoc18_F7	-0.0146
4	beta_T7	0.1568	66	hoc30_T7	0.0234	128	delta_FC5	-0.0023	190	delta_F7	-0.0149
5	stat2_T7	0.1494	67	beta_AF3	0.0219	129	hoc8_F7	-0.0025	191	hoc29_F7	-0.015
6	mblty_T7	0.118	68	hoc29_T7	0.0219	130	hoc12_F7	-0.003	192	hoc13_F7	-0.0159
7	stat4_T7	0.1083	69	stat5_FC5	0.0215	131	hoc10_F4	-0.0032	193	hoc18_T7	-0.0159
8	actvt_T7	0.1046	70	theta_F4	0.0215	132	hoc5_FC5	-0.0032	194	hoc25_FC5	-0.0162
9	se_T7	0.1046	71	hoc19_AF3	0.0214	133	hoc9_F4	-0.0037	195	stat1_T7	-0.0163
10	stat6_T7	0.1019	72	hoc25_T7	0.0202	134	hoc32_F4	-0.0037	196	hoc10_F7	-0.0165
11	hoc4_T7	0.0955	73	hoc2_AF3	0.0201	135	hoc14_F7	-0.0038	197	hoc12_FC5	-0.0168
12	cpxty_T7	0.0856	74	beta_FC5	0.0199	136	stat2_F7	-0.0038	198	hoc13_F4	-0.0169
13	hoc5_T7	0.0803	75	actvt_F4	0.0188	137	actvt_AF3	-0.0039	199	hoc3_F7	-0.017
14	hoc2_T7	0.0778	76	se_F4	0.0188	138	se_AF3	-0.0039	200	hoc15_T7	-0.0172
15	alpha_T7	0.0721	77	hoc22_AF3	0.0185	139	stat4_FC5	-0.0044	201	hoc30_F4	-0.0174
16	hoc3_T7	0.0627	78	fd_F4	0.0181	140	mblty_F7	-0.005	202	hoc26_FC5	-0.0174
17	hoc2_FC5	0.0624	79	hoc35_F4	0.018	141	hoc17_AF3	-0.005	203	hoc36_AF3	-0.0175
18	alpha_F7	0.0553	80	hoc23_T7	0.0177	142	hoc6_F7	-0.005	204	hoc20_F7	-0.0178
19	hoc2_F4	0.0552	81	hoc7_F7	0.0164	143	stat4_AF3	-0.0056	205	hoc16_FC5	-0.0179
20	hoc8_T7	0.0552	82	hoc11_T7	0.0159	144	stat4_F7	-0.0056	206	hoc11_F4	-0.018
21	hoc9_T7	0.0546	83	hoc20_AF3	0.0155	145	stat6_F7	-0.0057	207	hoc35_FC5	-0.0181
22	alpha_F4	0.0512	84	theta_T7	0.0153	146	cpxty_FC5	-0.0059	208	hoc21_F7	-0.0191
23	hoc29_AF3	0.0507	85	hoc34_AF3	0.015	147	hoc4_AF3	-0.0059	209	hoc23_FC5	-0.0192
24	beta_F4	0.049	86	hoc24_T7	0.0146	148	hoc9_AF3	-0.006	210	hoc31_FC5	-0.0193
25	hoc7_T7	0.0464	87	cpxty_F4	0.0146	149	theta_FC5	-0.0065	211	fd_AF3	-0.0195
26	hoc6_T7	0.0457	88	hoc13_AF3	0.0135	150	hoc33_FC5	-0.0066	212	hoc22_FC5	-0.0196
27	mblty_F4	0.0454	89	hoc5_F7	0.0131	151	hoc12_T7	-0.0067	213	hoc24_FC5	-0.0196
28	hoc28_AF3	0.0422	90	hoc21_T7	0.0123	152	stat1_FC5	-0.0073	214	hoc17_F7	-0.0198
29	hoc31_T7	0.0404	91	hoc10_FC5	0.0122	153	cpxty_F7	-0.0074	215	hoc34_FC5	-0.0198
30	fd_T7	0.0393	92	hoc18_AF3	0.0118	154	hoc3_AF3	-0.0074	216	hoc17_FC5	-0.0199
31	stat5_F7	0.0385	93	hoc5_F4	0.0112	155	theta_AF3	-0.0075	217	hoc9_F7	-0.02
32	hoc30_AF3	0.0384	94	delta_F4	0.0109	156	hoc15_AF3	-0.0079	218	hoc31_F4	-0.02
33	stat4_F4	0.0381	95	stat2_FC5	0.0095	157	hoc29_F4	-0.008	219	hoc15_FC5	-0.0201
34	stat3_F7	0.0374	96	hoc7_AF3	0.0092	158	hoc21_FC5	-0.0083	220	hoc16_T7	-0.0203
35	hoc24_AF3	0.0372	97	hoc1_AF3	0.0092	159	hoc6_F4	-0.0086	221	hoc14_F4	-0.0206
36	hoc32_T7	0.0366	98	hoc36_F4	0.0085	160	stat6_FC5	-0.0088	222	hoc35_F7	-0.0206
37	hoc28_T7	0.0327	99	stat2_AF3	0.0077	161	hoc7_FC5	-0.0092	223	hoc10_AF3	-0.0207
38	delta_T7	0.0326	100	hoc21_AF3	0.0076	162	hoc30_FC5	-0.0097	224	hoc36_FC5	-0.0208
39	hoc23_AF3	0.0325	101	hoc12_AF3	0.0076	163	hoc18_FC5	-0.0099	225	hoc27_F4	-0.0211
40	hoc36_T7	0.0315	102	hoc2_F7	0.0067	164	hoc7_F4	-0.01	226	hoc5_AF3	-0.0213
41	stat6_F4	0.0312	103	hoc34_F4	0.0064	165	actvt_F7	-0.0101	227	hoc14_FC5	-0.0215
42	hoc32_AF3	0.0312	104	hoc1_FC5	0.0064	166	se_F7	-0.0101	228	hoc11_AF3	-0.0215
43	hoc26_AF3	0.0308	105	hoc4_F7	0.0058	167	hoc29_FC5	-0.0101	229	hoc16_F4	-0.0218
44	hoc1_F4	0.0302	106	mblty_AF3	0.0056	168	hoc3_F4	-0.0103	230	hoc4_F4	-0.0218
45	alpha_FC5	0.0299	107	hoc4_FC5	0.0053	169	hoc11_F7	-0.0112	231	hoc15_F7	-0.022
46	hoc10_T7	0.0298	108	hoc33_F4	0.0049	170	hoc27_FC5	-0.0114	232	hoc13_T7	-0.0221
47	stat3_F4	0.0282	109	hoc3_FC5	0.0047	171	hoc13_FC5	-0.0115	233	hoc34_F7	-0.0231
48	stat3_AF3	0.0281	110	mblty_FC5	0.0046	172	hoc14_T7	-0.0115	234	hoc16_F7	-0.0239
49	stat2_F4	0.0279	111	hoc14_AF3	0.0043	173	hoc30_F7	-0.0116	235	hoc17_T7	-0.0253
50	alpha_AF3	0.0279	112	hoc22_T7	0.0041	174	delta_AF3	-0.0118	236	hoc18_F4	-0.0255
51	stat5_F4	0.0274	113	stat1_F7	0.0032	175	stat1_F4	-0.0118	237	hoc27_F7	-0.0268
52	stat5_AF3	0.0272	114	hoc6_FC5	0.0026	176	hoc19_FC5	-0.0119	238	hoc25_F7	-0.0275
53	hoc34_T7	0.0266	115	se_FC5	0.0024	177	stat6_AF3	-0.012	239	hoc33_F7	-0.0275
54	hoc33_T7	0.0265	116	actvt_FC5	0.0024	178	hoc12_F4	-0.0122	240	hoc22_F7	-0.0278
55	hoc9_FC5	0.026	117	hoc6_AF3	0.0024	179	hoc28_F4	-0.0126	241	hoc28_F7	-0.0286
56	hoc35_T7	0.0258	118	stat1_AF3	0.0016	180	cpxty_AF3	-0.0138	242	hoc36_F7	-0.0289
57	hoc8_F4	0.0256	119	hoc35_AF3	0.0014	181	fd_F7	-0.0138	243	hoc23_F7	-0.0292
58	hoc31_AF3	0.0251	120	hoc28_FC5	0.0013	182	fd_FC5	-0.0138	244	hoc32_F7	-0.0297
59	hoc25_AF3	0.0249	121	hoc11_FC5	0.0004	183	hoc31_F7	-0.0141	245	hoc26_F7	-0.0299
60	stat3_FC5	0.0246	122	hoc8_FC5	0.0004	184	hoc19_F7	-0.0142	246	hoc17_F4	-0.0312
61	hoc33_AF3	0.0236	123	hoc1_F7	-0.0001	185	hoc19_T7	-0.0142	247	hoc24_F4	-0.0317
62	hoc27_T7	0.0236	124	hoc20_T7	-0.0005	186	hoc32_FC5	-0.0145	248	hoc24_F7	-0.0318

Rank	Feature	Score	Rank	Feature	Score	Rank	Feature	Score	Rank	Feature	Score
249	hoc26_F4	-0.0326	251	hoc25_F4	-0.0356	253	hoc21_F4	-0.039	255	hoc20_F4	-0.0415
250	hoc23_F4	-0.0343	252	hoc19_F4	-0.0377	254	hoc22_F4	-0.0396			

Table A-3 Subject-dependent ICC score-based feature stability ranking.

Rank	Subject S1		Subject S2		Subject S3		Subject S4		Subject S5		Subject S6	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
1	hoc1_T7	0.4921	beta_F4	0.2671	hoc9_FC5	0.2771	stat5_T7	0.7548	hoc2_FC5	0.1909	stat3_T7	0.3430
2	stat3_T7	0.4913	hoc31_T7	0.1751	alpha_F7	0.2630	stat3_T7	0.7443	hoc32_AF3	0.1570	stat5_T7	0.3331
3	stat5_T7	0.4302	hoc33_T7	0.1651	hoc10_FC5	0.2445	beta_T7	0.6914	hoc28_AF3	0.1469	beta_T7	0.2726
4	hoc2_T7	0.3557	hoc34_T7	0.1523	hoc11_FC5	0.2040	stat2_T7	0.6473	hoc29_AF3	0.1279	hoc1_T7	0.2030
5	mblty_T7	0.2547	hoc32_T7	0.1450	hoc3_T7	0.1973	se_T7	0.5098	hoc30_AF3	0.1194	stat2_T7	0.1872
6	stat4_T7	0.2057	hoc2_F4	0.1428	hoc4_T7	0.1911	actvt_T7	0.5098	mblty_F4	0.1086	mblty_T7	0.1544
7	cpxty_T7	0.1874	beta_F7	0.1413	hoc8_FC5	0.1607	hoc1_T7	0.4992	hoc8_F4	0.1048	stat4_T7	0.1438
8	hoc13_AF3	0.1815	beta_AF3	0.1269	mblty_F4	0.1439	hoc5_T7	0.4353	hoc19_AF3	0.1015	hoc25_T7	0.1334
9	hoc29_AF3	0.1749	hoc2_FC5	0.1249	alpha_AF3	0.1398	alpha_T7	0.4272	hoc33_AF3	0.1011	stat6_T7	0.1233
10	stat6_T7	0.1652	hoc35_T7	0.1173	stat4_F4	0.1385	hoc4_T7	0.4161	hoc34_F4	0.0980	hoc26_T7	0.1173
11	hoc26_AF3	0.1633	hoc28_T7	0.1113	stat6_F4	0.1326	mblty_T7	0.3839	mblty_FC5	0.0972	hoc36_T7	0.1106
12	hoc23_AF3	0.1618	hoc3_F4	0.1011	alpha_F4	0.1284	stat4_T7	0.3724	hoc35_F4	0.0955	hoc23_T7	0.1105
13	hoc28_AF3	0.1606	hoc7_F7	0.0984	hoc4_FC5	0.1213	stat6_T7	0.3559	hoc23_AF3	0.0948	hoc32_T7	0.1102
14	hoc22_AF3	0.1364	hoc20_FC5	0.0943	cpxty_F4	0.1175	hoc6_T7	0.3393	alpha_AF3	0.0908	cpxty_T7	0.1069
15	hoc27_AF3	0.1357	hoc36_T7	0.0868	hoc12_FC5	0.1163	hoc9_T7	0.2768	hoc31_AF3	0.0891	hoc28_T7	0.1062
16	hoc25_AF3	0.1181	hoc21_FC5	0.0780	hoc6_FC5	0.1150	hoc7_T7	0.2702	hoc33_F4	0.0873	hoc30_T7	0.1049
17	hoc31_AF3	0.1176	hoc30_T7	0.0731	hoc5_T7	0.1094	hoc8_T7	0.2567	stat5_AF3	0.0849	hoc24_T7	0.0988
18	hoc30_AF3	0.1134	hoc18_FC5	0.0684	fd_F4	0.1071	cpxty_T7	0.1966	hoc20_AF3	0.0807	hoc31_T7	0.0977
19	hoc14_AF3	0.1086	hoc26_T7	0.0679	hoc1_F4	0.1050	hoc10_T7	0.1925	hoc27_AF3	0.0799	hoc2_T7	0.0973
20	fd_T7	0.0993	hoc29_T7	0.0676	hoc2_F4	0.0890	stat3_F7	0.1907	alpha_F4	0.0798	hoc8_F4	0.0964
21	hoc14_F7	0.0978	hoc19_FC5	0.0665	hoc33_FC5	0.0842	stat5_F7	0.1694	stat3_AF3	0.0778	actvt_T7	0.0941
22	hoc17_F7	0.0976	hoc27_T7	0.0571	hoc7_FC5	0.0839	hoc11_T7	0.1569	hoc2_AF3	0.0729	se_T7	0.0941
23	hoc32_AF3	0.0965	stat1_T7	0.0448	hoc5_FC5	0.0825	delta_T7	0.1517	hoc36_F4	0.0724	hoc2_FC5	0.0932
24	hoc12_AF3	0.0951	hoc23_T7	0.0431	hoc4_F7	0.0818	beta_F7	0.1466	stat4_F4	0.0677	hoc29_T7	0.0931
25	hoc3_T7	0.0913	hoc2_AF3	0.0407	stat2_F4	0.0810	hoc5_F4	0.1285	stat5_F7	0.0675	hoc9_F4	0.0912
26	hoc24_AF3	0.0893	hoc12_F7	0.0394	hoc24_AF3	0.0769	hoc2_T7	0.1048	hoc34_AF3	0.0641	hoc33_T7	0.0896
27	hoc18_AF3	0.0822	stat5_F4	0.0389	hoc13_FC5	0.0766	hoc21_T7	0.0973	hoc25_AF3	0.0641	hoc35_T7	0.0893
28	hoc8_T7	0.0810	hoc8_F7	0.0369	stat1_AF3	0.0747	hoc28_FC5	0.0938	hoc26_AF3	0.0638	hoc22_T7	0.0845
29	hoc9_T7	0.0796	cpxty_T7	0.0354	cpxty_FC5	0.0741	fd_T7	0.0929	stat3_F7	0.0620	stat3_FC5	0.0814
30	hoc12_F7	0.0783	hoc3_T7	0.0319	alpha_FC5	0.0691	hoc3_T7	0.0929	alpha_F7	0.0615	hoc27_T7	0.0794
31	hoc17_AF3	0.0756	hoc11_F7	0.0278	mblty_AF3	0.0670	hoc20_T7	0.0919	hoc24_AF3	0.0609	stat1_AF3	0.0752
32	hoc35_F4	0.0728	stat3_F4	0.0272	hoc30_FC5	0.0647	beta_FC5	0.0914	hoc6_F4	0.0590	hoc19_AF3	0.0751
33	hoc16_F7	0.0723	hoc24_T7	0.0259	hoc3_F7	0.0639	theta_T7	0.0899	stat5_F4	0.0576	alpha_F7	0.0731
34	hoc16_AF3	0.0697	alpha_FC5	0.0235	hoc1_AF3	0.0630	hoc2_F4	0.0881	hoc10_F4	0.0570	hoc7_T7	0.0729
35	theta_F4	0.0636	alpha_AF3	0.0208	stat1_F7	0.0605	hoc16_T7	0.0875	hoc7_AF3	0.0548	hoc30_F7	0.0714
36	actvt_F4	0.0623	hoc2_F7	0.0203	stat1_FC5	0.0590	hoc27_FC5	0.0865	hoc22_AF3	0.0532	hoc31_F7	0.0708
37	se_F4	0.0623	hoc25_T7	0.0166	hoc33_F4	0.0567	stat3_AF3	0.0705	hoc1_AF3	0.0514	stat5_FC5	0.0705
38	delta_F4	0.0620	hoc34_AF3	0.0137	stat5_F7	0.0567	hoc12_T7	0.0690	stat2_AF3	0.0503	hoc34_T7	0.0697
39	hoc15_F7	0.0610	hoc21_T7	0.0135	hoc35_F4	0.0566	hoc26_FC5	0.0688	hoc5_F7	0.0503	hoc9_T7	0.0685
40	hoc33_AF3	0.0607	hoc17_FC5	0.0129	delta_T7	0.0552	stat3_FC5	0.0675	hoc12_F4	0.0498	stat4_F7	0.0660
41	hoc20_AF3	0.0605	hoc22_FC5	0.0125	hoc2_AF3	0.0525	hoc1_F4	0.0668	stat3_F4	0.0496	hoc4_T7	0.0657
42	alpha_F4	0.0599	beta_FC5	0.0124	alpha_T7	0.0491	hoc14_T7	0.0656	hoc30_T7	0.0495	stat6_F7	0.0641
43	hoc13_F7	0.0595	hoc5_F7	0.0123	hoc33_AF3	0.0419	stat5_FC5	0.0653	stat2_F7	0.0493	hoc21_T7	0.0637
44	hoc18_F7	0.0576	actvt_FC5	0.0101	actvt_T7	0.0417	hoc15_T7	0.0651	hoc32_F4	0.0487	hoc8_T7	0.0619
45	hoc21_AF3	0.0557	se_FC5	0.0101	se_T7	0.0416	hoc19_T7	0.0634	stat5_FC5	0.0485	hoc29_F7	0.0609
46	beta_F4	0.0548	delta_FC5	0.0099	theta_F4	0.0397	hoc29_FC5	0.0620	hoc29_F4	0.0484	hoc1_F7	0.0578
47	hoc15_AF3	0.0547	stat6_T7	0.0097	theta_T7	0.0373	beta_AF3	0.0619	hoc7_F4	0.0483	mblty_F7	0.0576
48	hoc36_F4	0.0539	hoc12_T7	0.0096	hoc34_F4	0.0362	hoc25_FC5	0.0582	stat4_FC5	0.0483	stat3_AF3	0.0562
49	stat2_F4	0.0514	hoc27_F4	0.0087	hoc34_AF3	0.0357	fd_F4	0.0580	hoc30_F4	0.0477	hoc5_T7	0.0530
50	hoc6_AF3	0.0513	hoc18_F4	0.0086	actvt_F4	0.0351	hoc20_FC5	0.0576	hoc6_F7	0.0462	hoc13_AF3	0.0513
51	hoc19_AF3	0.0501	theta_FC5	0.0072	se_F4	0.0351	stat3_F4	0.0573	hoc9_AF3	0.0439	alpha_FC5	0.0508
52	hoc32_T7	0.0443	hoc16_FC5	0.0064	stat2_AF3	0.0351	hoc17_T7	0.0500	hoc1_FC5	0.0438	hoc30_AF3	0.0503
53	beta_T7	0.0417	hoc14_F7	0.0058	stat1_F4	0.0328	hoc22_T7	0.0485	stat3_FC5	0.0422	beta_FC5	0.0496
54	hoc34_F4	0.0397	stat1_F7	0.0056	stat4_AF3	0.0319	hoc2_FC5	0.0482	stat2_F4	0.0415	hoc8_AF3	0.0491

Rank	Subject S1		Subject S2		Subject S3		Subject S4		Subject S5		Subject S6	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
55	stat2_FC5	0.0388	hoc26_F4	0.0031	fd_T7	0.0316	stat5_AF3	0.0474	stat3_T7	0.0386	hoc10_T7	0.0484
56	hoc7_T7	0.0388	hoc1_FC5	0.0024	fd_FC5	0.0307	hoc30_FC5	0.0450	hoc9_F4	0.0377	hoc11_T7	0.0481
57	hoc11_F7	0.0377	fd_T7	0.0019	hoc14_FC5	0.0301	stat4_F4	0.0436	hoc2_F4	0.0375	stat5_AF3	0.0480
58	hoc30_F4	0.0348	stat1_AF3	0.0016	hoc27_T7	0.0292	theta_F4	0.0433	hoc28_T7	0.0373	hoc7_F7	0.0453
59	hoc14_F4	0.0340	hoc23_FC5	0.0013	stat2_T7	0.0291	mblty_F4	0.0428	stat2_FC5	0.0366	stat1_FC5	0.0414
60	hoc35_AF3	0.0330	hoc8_F4	-0.0003	stat3_F7	0.0288	delta_F4	0.0408	hoc21_AF3	0.0354	hoc8_F7	0.0413
61	stat3_F4	0.0320	hoc4_T7	-0.0028	hoc2_F7	0.0272	hoc1_AF3	0.0399	stat5_T7	0.0349	hoc32_F7	0.0388
62	hoc19_F7	0.0320	hoc5_F4	-0.0028	hoc36_F4	0.0257	hoc6_F4	0.0383	hoc4_AF3	0.0344	hoc12_AF3	0.0385
63	stat5_F4	0.0304	hoc20_T7	-0.0032	hoc35_FC5	0.0255	stat2_F4	0.0380	hoc8_F7	0.0335	hoc7_AF3	0.0368
64	hoc31_F4	0.0303	hoc8_T7	-0.0040	hoc31_FC5	0.0254	se_F4	0.0376	fd_F7	0.0330	cpxty_F7	0.0361
65	fd_F4	0.0303	alpha_F4	-0.0043	stat5_AF3	0.0245	actvt_F4	0.0376	delta_T7	0.0323	hoc10_F4	0.0358
66	actvt_FC5	0.0300	hoc9_T7	-0.0065	hoc4_AF3	0.0241	hoc13_T7	0.0375	hoc19_F7	0.0307	hoc25_F7	0.0357
67	se_FC5	0.0300	hoc35_AF3	-0.0070	hoc19_F7	0.0239	hoc24_FC5	0.0375	hoc30_F7	0.0289	hoc14_AF3	0.0352
68	cpxty_F7	0.0298	hoc5_FC5	-0.0076	hoc34_FC5	0.0226	stat6_F4	0.0370	hoc3_AF3	0.0289	hoc17_AF3	0.0348
69	stat2_T7	0.0283	hoc1_F4	-0.0078	hoc28_F4	0.0225	stat2_F7	0.0363	alpha_FC5	0.0285	hoc16_AF3	0.0347
70	delta_FC5	0.0280	hoc15_FC5	-0.0081	hoc32_F4	0.0222	hoc4_F4	0.0339	hoc18_AF3	0.0278	hoc18_AF3	0.0344
71	hoc10_T7	0.0275	hoc3_AF3	-0.0087	hoc18_F7	0.0219	hoc21_AF3	0.0331	hoc31_T7	0.0277	beta_AF3	0.0335
72	hoc7_AF3	0.0274	stat2_FC5	-0.0089	delta_F4	0.0207	alpha_F4	0.0329	stat2_T7	0.0276	fd_T7	0.0325
73	hoc32_F4	0.0273	actvt_F4	-0.0093	hoc30_AF3	0.0206	hoc22_FC5	0.0329	hoc1_F4	0.0263	stat4_F4	0.0323
74	hoc33_F4	0.0272	se_F4	-0.0093	stat3_T7	0.0199	hoc18_AF3	0.0327	hoc29_T7	0.0263	hoc27_F7	0.0315
75	hoc29_F4	0.0270	hoc33_AF3	-0.0094	hoc10_F4	0.0195	hoc25_AF3	0.0325	hoc23_T7	0.0255	hoc28_F7	0.0314
76	hoc28_FC5	0.0265	hoc4_F4	-0.0098	hoc29_AF3	0.0193	hoc18_T7	0.0311	stat6_F4	0.0255	hoc1_F4	0.0309
77	hoc9_FC5	0.0263	hoc14_T7	-0.0101	stat3_AF3	0.0189	hoc24_AF3	0.0297	hoc3_FC5	0.0252	stat6_F4	0.0292
78	hoc34_AF3	0.0245	hoc28_F4	-0.0104	stat5_T7	0.0184	stat5_F4	0.0290	hoc11_F4	0.0244	hoc2_AF3	0.0269
79	hoc20_F7	0.0242	hoc25_F4	-0.0113	hoc16_F4	0.0180	actvt_F7	0.0267	hoc2_F7	0.0243	hoc26_F7	0.0268
80	hoc8_AF3	0.0242	stat4_T7	-0.0117	hoc36_FC5	0.0179	se_F7	0.0267	stat6_FC5	0.0211	hoc36_AF3	0.0249
81	theta_FC5	0.0233	theta_F4	-0.0128	actvt_AF3	0.0173	hoc28_AF3	0.0265	hoc31_F4	0.0206	hoc20_AF3	0.0247
82	hoc6_T7	0.0227	hoc11_T7	-0.0130	se_AF3	0.0172	cpxty_F4	0.0262	hoc21_F7	0.0190	hoc7_F4	0.0245
83	hoc3_FC5	0.0221	hoc10_T7	-0.0133	hoc23_AF3	0.0161	stat2_FC5	0.0252	hoc21_T7	0.0186	hoc6_T7	0.0244
84	alpha_FC5	0.0220	hoc15_T7	-0.0133	stat2_F7	0.0159	hoc32_FC5	0.0244	mblty_AF3	0.0185	mblty_F4	0.0238
85	hoc36_T7	0.0217	hoc22_T7	-0.0139	stat5_F4	0.0151	hoc24_T7	0.0244	hoc25_F7	0.0183	hoc9_AF3	0.0235
86	stat6_AF3	0.0211	hoc3_F7	-0.0144	stat6_AF3	0.0150	hoc1_F7	0.0244	hoc25_T7	0.0180	hoc6_F7	0.0214
87	beta_FC5	0.0197	theta_T7	-0.0148	hoc3_FC5	0.0143	hoc25_T7	0.0243	hoc20_T7	0.0176	hoc2_F7	0.0210
88	hoc34_T7	0.0195	hoc7_FC5	-0.0151	hoc15_F4	0.0137	hoc23_FC5	0.0227	hoc35_AF3	0.0174	fd_F7	0.0203
89	stat4_AF3	0.0174	stat2_F4	-0.0152	hoc35_AF3	0.0135	hoc26_AF3	0.0226	delta_AF3	0.0174	stat2_AF3	0.0201
90	stat3_FC5	0.0170	hoc9_F7	-0.0157	hoc26_AF3	0.0132	hoc31_FC5	0.0217	hoc15_F4	0.0171	hoc21_AF3	0.0194
91	hoc5_F7	0.0165	hoc3_FC5	-0.0158	hoc6_AF3	0.0128	hoc16_AF3	0.0202	hoc7_F7	0.0164	delta_AF3	0.0175
92	fd_AF3	0.0155	hoc6_F7	-0.0164	stat6_FC5	0.0111	hoc19_FC5	0.0189	hoc28_F4	0.0164	hoc15_AF3	0.0162
93	hoc13_F4	0.0154	hoc12_AF3	-0.0169	hoc23_F7	0.0110	fd_AF3	0.0185	hoc31_F7	0.0159	se_AF3	0.0161
94	hoc6_FC5	0.0145	stat5_AF3	-0.0170	hoc5_F7	0.0107	hoc23_T7	0.0183	stat1_F7	0.0158	actvt_AF3	0.0161
95	fd_FC5	0.0144	hoc19_T7	-0.0171	hoc12_F4	0.0106	delta_F7	0.0168	hoc36_T7	0.0157	hoc29_AF3	0.0156
96	hoc26_T7	0.0143	alpha_F7	-0.0182	hoc32_FC5	0.0103	beta_F4	0.0123	theta_AF3	0.0150	stat3_F7	0.0155
97	hoc27_T7	0.0132	hoc6_FC5	-0.0186	hoc11_F4	0.0102	hoc33_FC5	0.0115	hoc26_T7	0.0149	hoc24_AF3	0.0151
98	stat6_FC5	0.0120	hoc19_F4	-0.0191	hoc29_FC5	0.0096	theta_F7	0.0111	hoc20_F7	0.0143	hoc34_F7	0.0135
99	hoc35_T7	0.0109	delta_T7	-0.0199	hoc32_AF3	0.0079	se_FC5	0.0107	hoc24_T7	0.0143	alpha_T7	0.0131
100	hoc31_T7	0.0106	hoc21_F4	-0.0200	hoc34_T7	0.0076	actvt_FC5	0.0107	hoc6_AF3	0.0136	hoc20_T7	0.0130
101	stat5_FC5	0.0100	hoc17_F4	-0.0211	hoc10_F7	0.0058	theta_FC5	0.0106	hoc26_F7	0.0125	hoc23_FC5	0.0130
102	stat6_F4	0.0098	mblty_T7	-0.0213	hoc31_AF3	0.0047	hoc22_AF3	0.0102	se_T7	0.0125	hoc21_FC5	0.0105
103	hoc8_FC5	0.0093	hoc6_F4	-0.0223	hoc28_AF3	0.0042	mblty_AF3	0.0100	actvt_T7	0.0125	mblty_FC5	0.0103
104	hoc33_T7	0.0091	stat2_T7	-0.0232	hoc15_FC5	0.0021	hoc21_FC5	0.0096	hoc15_F7	0.0117	alpha_F4	0.0102
105	cpxty_AF3	0.0091	hoc36_F4	-0.0239	hoc5_F4	0.0005	hoc1_FC5	0.0073	hoc22_T7	0.0116	hoc13_F4	0.0096
106	hoc7_F7	0.0087	hoc29_F4	-0.0241	hoc17_F4	0.0004	hoc27_AF3	0.0064	hoc32_T7	0.0103	hoc33_F7	0.0095
107	hoc6_F7	0.0085	hoc10_FC5	-0.0252	cpxty_AF3	0.0003	hoc20_AF3	0.0063	mblty_F7	0.0098	hoc11_F7	0.0088
108	hoc4_F7	0.0083	hoc8_AF3	-0.0254	hoc1_FC5	-0.0008	hoc17_AF3	0.0061	hoc8_AF3	0.0089	hoc24_F7	0.0087
109	stat4_FC5	0.0057	hoc13_F7	-0.0255	hoc25_AF3	-0.0009	hoc5_F7	0.0056	hoc5_F4	0.0086	hoc4_AF3	0.0087
110	stat1_F4	0.0053	hoc9_AF3	-0.0256	hoc35_T7	-0.0012	hoc29_AF3	0.0031	se_AF3	0.0081	hoc36_F7	0.0086
111	hoc5_FC5	0.0048	hoc4_FC5	-0.0259	stat3_F4	-0.0013	hoc11_AF3	0.0023	actvt_AF3	0.0081	hoc10_AF3	0.0085
112	hoc4_FC5	0.0047	hoc18_T7	-0.0268	beta_F4	-0.0014	hoc7_FC5	0.0020	hoc12_AF3	0.0070	hoc12_F4	0.0082
113	stat4_F4	0.0027	hoc14_FC5	-0.0270	cpxty_T7	-0.0023	stat2_AF3	0.0017	theta_T7	0.0064	stat4_FC5	0.0075
114	stat6_F7	0.0023	stat3_FC5	-0.0281	hoc29_F4	-0.0032	hoc5_FC5	0.0006	cpxty_FC5	0.0064	hoc35_F7	0.0072
115	hoc32_FC5	0.0019	cpxty_F7	-0.0282	hoc20_F7	-0.0037	hoc7_AF3	-0.0001	hoc33_F7	0.0063	theta_AF3	0.0070
116	hoc22_F7	0.0016	hoc23_F4	-0.0283	theta_AF3	-0.0044	hoc4_FC5	-0.0001	hoc1_F7	0.0060	mblty_AF3	0.0064
117	cpxty_FC5	0.0015	hoc24_F4	-0.0284	hoc29_T7	-0.0045	hoc29_F7	-0.0007	stat4_F7	0.0050	stat5_F7	0.0063
118	hoc10_F7	0.0014	hoc1_F7	-0.0284	actvt_F7	-0.0053	theta_AF3	-0.0008	hoc27_T7	0.0044	hoc18_FC5	0.0050

Rank	Subject S1		Subject S2		Subject S3		Subject S4		Subject S5		Subject S6	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
119	stat2_AF3	0.0013	hoc27_F7	-0.0293	se_F7	-0.0053	hoc15_AF3	-0.0008	hoc18_F7	0.0029	hoc27_FC5	0.0044
120	delta_T7	0.0007	se_T7	-0.0297	hoc3_AF3	-0.0054	hoc34_FC5	-0.0009	stat6_F7	0.0000	stat3_F4	0.0042
121	hoc28_F4	0.0006	actvt_T7	-0.0297	hoc31_T7	-0.0054	stat4_AF3	-0.0010	hoc19_T7	-0.0002	hoc2_F4	0.0041
122	cpxty_F4	0.0004	hoc13_T7	-0.0297	delta_FC5	-0.0054	hoc23_AF3	-0.0010	se_F7	-0.0003	hoc14_T7	0.0039
123	hoc21_F7	-0.0002	hoc36_AF3	-0.0308	mblty_FC5	-0.0055	cpxty_AF3	-0.0011	actvt_F7	-0.0003	hoc13_T7	0.0031
124	hoc28_T7	-0.0006	hoc24_FC5	-0.0308	hoc18_F4	-0.0057	hoc19_AF3	-0.0013	stat1_F4	-0.0009	hoc19_FC5	0.0030
125	se_T7	-0.0010	stat5_FC5	-0.0308	hoc28_FC5	-0.0059	stat6_AF3	-0.0018	hoc13_F4	-0.0010	hoc1_FC5	0.0029
126	actvt_T7	-0.0010	hoc4_F7	-0.0310	hoc27_AF3	-0.0062	actvt_AF3	-0.0019	hoc10_AF3	-0.0013	stat1_F7	0.0027
127	theta_T7	-0.0025	hoc9_FC5	-0.0310	hoc9_F7	-0.0066	se_AF3	-0.0019	cpxty_AF3	-0.0019	hoc32_AF3	0.0024
128	hoc29_T7	-0.0037	hoc16_T7	-0.0314	hoc21_F7	-0.0068	delta_FC5	-0.0024	hoc13_F7	-0.0022	hoc9_F7	0.0021
129	mblty_AF3	-0.0040	delta_F4	-0.0315	beta_T7	-0.0078	hoc10_AF3	-0.0036	theta_F4	-0.0027	hoc14_F4	0.0020
130	hoc16_F4	-0.0041	hoc1_T7	-0.0316	theta_F7	-0.0085	hoc36_FC5	-0.0037	se_FC5	-0.0027	hoc29_FC5	0.0019
131	delta_AF3	-0.0050	hoc20_AF3	-0.0317	hoc28_T7	-0.0093	hoc4_F7	-0.0040	actvt_FC5	-0.0028	stat4_AF3	0.0018
132	mblty_F4	-0.0051	stat5_F7	-0.0317	hoc36_T7	-0.0096	hoc6_FC5	-0.0041	theta_F7	-0.0029	hoc24_FC5	0.0013
133	hoc15_F4	-0.0057	hoc20_F4	-0.0326	hoc17_F7	-0.0096	hoc7_F4	-0.0044	hoc28_F7	-0.0033	hoc23_F4	0.0005
134	theta_AF3	-0.0064	hoc25_FC5	-0.0326	hoc36_AF3	-0.0112	hoc31_AF3	-0.0061	theta_FC5	-0.0034	hoc35_AF3	0.0004
135	delta_F7	-0.0068	hoc13_AF3	-0.0326	hoc30_T7	-0.0116	delta_AF3	-0.0063	hoc17_T7	-0.0034	hoc16_FC5	-0.0001
136	se_AF3	-0.0070	cpxty_F4	-0.0329	hoc6_T7	-0.0125	hoc31_F7	-0.0063	hoc18_T7	-0.0037	hoc11_F4	-0.0003
137	actvt_AF3	-0.0070	hoc6_AF3	-0.0329	hoc22_F7	-0.0127	fd_FC5	-0.0073	cpxty_F7	-0.0044	hoc5_AF3	-0.0006
138	hoc13_FC5	-0.0077	hoc29_F7	-0.0333	hoc8_F4	-0.0134	hoc5_AF3	-0.0074	hoc16_F4	-0.0048	hoc22_FC5	-0.0007
139	alpha_T7	-0.0083	hoc11_AF3	-0.0334	stat6_T7	-0.0135	hoc35_FC5	-0.0081	hoc14_F7	-0.0048	hoc35_F4	-0.0021
140	hoc24_F4	-0.0087	hoc32_F4	-0.0343	hoc14_F4	-0.0139	hoc3_FC5	-0.0081	se_F4	-0.0049	theta_F4	-0.0022
141	se_F7	-0.0103	hoc5_T7	-0.0344	hoc13_F4	-0.0146	hoc2_F7	-0.0082	actvt_F4	-0.0049	hoc10_F7	-0.0023
142	actvt_F7	-0.0103	stat3_AF3	-0.0345	hoc27_F4	-0.0147	hoc18_FC5	-0.0089	hoc35_F7	-0.0050	hoc22_F4	-0.0027
143	hoc9_AF3	-0.0105	hoc7_F4	-0.0347	hoc31_F4	-0.0148	mblty_FC5	-0.0094	delta_F7	-0.0051	hoc15_F4	-0.0032
144	hoc16_FC5	-0.0105	stat1_F4	-0.0347	mblty_F7	-0.0161	hoc14_AF3	-0.0105	hoc11_AF3	-0.0052	hoc18_T7	-0.0057
145	hoc3_F4	-0.0105	hoc32_AF3	-0.0347	hoc9_F4	-0.0167	hoc2_AF3	-0.0124	hoc36_AF3	-0.0054	delta_F7	-0.0060
146	fd_F7	-0.0117	hoc15_F7	-0.0349	hoc1_F7	-0.0168	cpxty_FC5	-0.0128	hoc29_F7	-0.0058	delta_F4	-0.0063
147	hoc3_AF3	-0.0125	hoc22_F4	-0.0363	fd_AF3	-0.0169	stat1_T7	-0.0129	beta_T7	-0.0061	hoc12_T7	-0.0064
148	hoc4_F4	-0.0130	hoc29_AF3	-0.0366	hoc7_AF3	-0.0171	hoc17_FC5	-0.0130	beta_F4	-0.0065	stat5_F4	-0.0064
149	hoc15_FC5	-0.0138	stat6_F7	-0.0371	theta_FC5	-0.0173	alpha_FC5	-0.0143	beta_AF3	-0.0082	hoc31_AF3	-0.0065
150	hoc3_F7	-0.0143	hoc10_F7	-0.0371	stat2_FC5	-0.0173	alpha_F7	-0.0145	delta_FC5	-0.0093	cpxty_F4	-0.0067
151	hoc27_F4	-0.0145	alpha_T7	-0.0377	actvt_FC5	-0.0178	stat4_FC5	-0.0149	hoc34_F7	-0.0095	hoc34_AF3	-0.0073
152	hoc8_F4	-0.0147	stat1_FC5	-0.0382	se_FC5	-0.0178	mblty_F7	-0.0154	fd_FC5	-0.0100	hoc4_F7	-0.0080
153	hoc17_FC5	-0.0152	hoc30_F4	-0.0389	stat1_T7	-0.0191	hoc3_AF3	-0.0164	cpxty_T7	-0.0105	actvt_F4	-0.0081
154	hoc30_T7	-0.0152	hoc2_T7	-0.0393	stat4_FC5	-0.0211	hoc6_AF3	-0.0165	beta_FC5	-0.0106	se_F4	-0.0081
155	stat4_F7	-0.0155	hoc35_F7	-0.0393	stat4_F7	-0.0238	hoc9_F4	-0.0166	alpha_T7	-0.0107	hoc25_AF3	-0.0092
156	mblty_FC5	-0.0156	hoc22_AF3	-0.0394	hoc19_F4	-0.0239	hoc7_F7	-0.0172	hoc27_F4	-0.0114	hoc33_AF3	-0.0096
157	hoc12_FC5	-0.0157	hoc18_AF3	-0.0411	hoc7_T7	-0.0240	hoc9_AF3	-0.0173	hoc18_FC5	-0.0115	hoc3_FC5	-0.0096
158	theta_F7	-0.0168	hoc13_FC5	-0.0411	hoc33_T7	-0.0240	hoc8_F4	-0.0190	beta_F7	-0.0117	theta_F7	-0.0099
159	hoc34_FC5	-0.0170	hoc16_F4	-0.0413	hoc25_F4	-0.0245	stat4_F7	-0.0205	hoc4_F7	-0.0120	stat6_AF3	-0.0100
160	hoc1_FC5	-0.0171	mblty_F4	-0.0417	hoc8_AF3	-0.0250	stat1_FC5	-0.0215	hoc12_F7	-0.0133	hoc1_AF3	-0.0102
161	mblty_F7	-0.0174	hoc21_F7	-0.0418	hoc27_FC5	-0.0251	stat6_FC5	-0.0220	hoc5_AF3	-0.0136	hoc3_T7	-0.0104
162	hoc14_FC5	-0.0176	hoc19_AF3	-0.0421	hoc24_F4	-0.0253	hoc30_F7	-0.0224	hoc9_F7	-0.0137	se_F7	-0.0105
163	hoc33_FC5	-0.0183	hoc10_F4	-0.0422	hoc2_T7	-0.0255	hoc28_F7	-0.0224	hoc15_T7	-0.0137	actvt_F7	-0.0105
164	stat2_F7	-0.0183	hoc14_AF3	-0.0424	hoc8_T7	-0.0257	hoc21_F4	-0.0225	hoc26_FC5	-0.0140	hoc20_FC5	-0.0107
165	hoc10_FC5	-0.0190	stat3_F7	-0.0426	hoc2_FC5	-0.0264	hoc13_AF3	-0.0252	hoc32_F7	-0.0151	hoc19_T7	-0.0113
166	hoc25_T7	-0.0194	hoc16_F7	-0.0431	hoc11_T7	-0.0267	hoc12_AF3	-0.0273	hoc13_FC5	-0.0167	hoc22_AF3	-0.0127
167	hoc24_T7	-0.0195	hoc20_F7	-0.0438	hoc16_FC5	-0.0270	hoc30_AF3	-0.0283	hoc27_F7	-0.0167	hoc28_FC5	-0.0129
168	stat3_AF3	-0.0200	hoc11_FC5	-0.0439	cpxty_F7	-0.0275	hoc13_F7	-0.0285	cpxty_F4	-0.0170	hoc27_AF3	-0.0138
169	hoc4_AF3	-0.0222	hoc30_F7	-0.0441	stat4_T7	-0.0277	hoc3_F4	-0.0292	hoc15_AF3	-0.0174	stat6_FC5	-0.0139
170	hoc7_FC5	-0.0223	hoc15_AF3	-0.0442	stat6_F7	-0.0289	hoc4_AF3	-0.0303	hoc17_F4	-0.0179	hoc6_AF3	-0.0139
171	hoc23_F7	-0.0229	hoc10_AF3	-0.0443	hoc32_T7	-0.0292	hoc29_F4	-0.0305	stat4_AF3	-0.0182	fd_F4	-0.0141
172	hoc35_F7	-0.0242	hoc5_AF3	-0.0449	hoc35_F7	-0.0293	hoc34_F7	-0.0311	hoc17_AF3	-0.0194	hoc36_F4	-0.0147
173	stat5_AF3	-0.0247	hoc30_AF3	-0.0449	hoc26_F4	-0.0298	hoc9_FC5	-0.0314	hoc18_F4	-0.0194	se_FC5	-0.0156
174	hoc26_FC5	-0.0249	stat4_F7	-0.0449	hoc30_F4	-0.0303	hoc14_F7	-0.0323	hoc16_AF3	-0.0196	actvt_FC5	-0.0156
175	hoc34_F7	-0.0252	fd_F7	-0.0451	hoc10_T7	-0.0303	hoc33_F7	-0.0329	hoc22_F7	-0.0198	hoc5_F7	-0.0166
176	hoc27_FC5	-0.0256	hoc32_F7	-0.0454	hoc26_T7	-0.0306	hoc10_F7	-0.0331	hoc14_AF3	-0.0199	alpha_AF3	-0.0170
177	hoc5_AF3	-0.0260	hoc23_AF3	-0.0458	hoc25_FC5	-0.0321	hoc35_F7	-0.0331	hoc12_T7	-0.0201	stat2_FC5	-0.0176
178	hoc18_T7	-0.0264	hoc36_F7	-0.0458	hoc7_F4	-0.0324	stat1_F7	-0.0332	fd_F4	-0.0203	hoc34_F4	-0.0176
179	hoc36_AF3	-0.0267	hoc21_AF3	-0.0461	stat3_FC5	-0.0326	hoc3_F7	-0.0336	delta_F4	-0.0205	hoc11_AF3	-0.0176
180	hoc25_F4	-0.0270	hoc17_T7	-0.0462	delta_F7	-0.0330	hoc20_F4	-0.0337	hoc35_FC5	-0.0208	hoc12_F7	-0.0179
181	hoc11_FC5	-0.0274	hoc7_AF3	-0.0464	hoc24_F7	-0.0341	hoc36_F7	-0.0337	hoc35_T7	-0.0217	hoc26_AF3	-0.0180
182	hoc35_FC5	-0.0275	stat6_F4	-0.0468	stat5_FC5	-0.0344	stat6_F7	-0.0345	hoc33_T7	-0.0220	hoc32_FC5	-0.0181

Rank	Subject S1		Subject S2		Subject S3		Subject S4		Subject S5		Subject S6	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
183	hoc12_F4	-0.0282	hoc8_FC5	-0.0470	hoc26_FC5	-0.0348	fd_F7	-0.0362	fd_T7	-0.0223	stat1_F4	-0.0186
184	alpha_AF3	-0.0282	hoc33_F7	-0.0471	hoc5_AF3	-0.0351	hoc36_T7	-0.0362	hoc16_F7	-0.0226	hoc13_FC5	-0.0186
185	stat1_T7	-0.0292	hoc34_F7	-0.0480	hoc8_F7	-0.0358	hoc8_FC5	-0.0367	mblty_T7	-0.0227	hoc36_FC5	-0.0199
186	hoc26_F4	-0.0296	hoc31_AF3	-0.0481	hoc4_F4	-0.0360	hoc22_F4	-0.0381	hoc34_FC5	-0.0234	hoc28_AF3	-0.0210
187	hoc23_F4	-0.0297	mblty_F7	-0.0483	hoc16_F7	-0.0364	alpha_AF3	-0.0387	hoc16_T7	-0.0237	hoc15_FC5	-0.0216
188	hoc11_AF3	-0.0299	hoc24_F7	-0.0483	hoc6_F4	-0.0365	hoc35_T7	-0.0399	hoc14_FC5	-0.0241	hoc33_FC5	-0.0239
189	stat3_F7	-0.0301	hoc24_AF3	-0.0483	hoc22_AF3	-0.0368	hoc12_F7	-0.0401	hoc24_F7	-0.0246	delta_T7	-0.0241
190	hoc2_F4	-0.0303	hoc33_F4	-0.0486	hoc14_F7	-0.0370	hoc34_AF3	-0.0406	hoc23_F7	-0.0249	hoc17_FC5	-0.0244
191	hoc10_AF3	-0.0306	hoc36_FC5	-0.0486	hoc34_F7	-0.0382	hoc8_F7	-0.0408	hoc36_F7	-0.0253	theta_T7	-0.0244
192	hoc36_F7	-0.0307	hoc35_FC5	-0.0490	mblty_T7	-0.0408	hoc6_F7	-0.0411	stat1_FC5	-0.0255	hoc14_FC5	-0.0254
193	hoc5_T7	-0.0309	mblty_FC5	-0.0492	delta_AF3	-0.0419	hoc28_F4	-0.0413	hoc17_FC5	-0.0255	hoc32_F4	-0.0266
194	hoc25_FC5	-0.0312	hoc1_AF3	-0.0494	hoc23_F4	-0.0425	hoc27_T7	-0.0417	hoc2_T7	-0.0261	hoc5_F4	-0.0272
195	beta_AF3	-0.0316	hoc31_F7	-0.0499	beta_FC5	-0.0428	hoc32_AF3	-0.0422	hoc17_F7	-0.0267	hoc25_FC5	-0.0277
196	stat1_F7	-0.0322	hoc4_AF3	-0.0499	fd_F7	-0.0431	hoc26_T7	-0.0424	hoc14_F4	-0.0267	hoc35_FC5	-0.0285
197	hoc30_FC5	-0.0330	hoc7_T7	-0.0506	beta_F7	-0.0436	hoc8_AF3	-0.0425	hoc3_T7	-0.0269	stat2_F4	-0.0291
198	hoc36_FC5	-0.0332	beta_T7	-0.0507	hoc21_FC5	-0.0438	hoc33_AF3	-0.0429	stat1_AF3	-0.0278	hoc3_AF3	-0.0304
199	alpha_F7	-0.0334	hoc28_F7	-0.0509	hoc11_AF3	-0.0452	hoc24_F4	-0.0429	hoc7_T7	-0.0286	hoc23_AF3	-0.0306
200	beta_F7	-0.0352	hoc22_F7	-0.0512	hoc14_AF3	-0.0454	hoc22_F7	-0.0439	hoc15_FC5	-0.0290	hoc8_FC5	-0.0314
201	hoc11_F4	-0.0357	hoc31_F4	-0.0513	hoc23_T7	-0.0457	hoc11_F7	-0.0442	stat6_T7	-0.0291	hoc33_F4	-0.0315
202	hoc11_T7	-0.0358	stat4_FC5	-0.0516	hoc36_F7	-0.0467	hoc10_FC5	-0.0444	hoc16_FC5	-0.0296	hoc23_F7	-0.0321
203	hoc24_FC5	-0.0361	hoc31_FC5	-0.0517	hoc30_F7	-0.0472	hoc32_F7	-0.0458	hoc11_FC5	-0.0305	beta_F4	-0.0323
204	stat5_F7	-0.0370	stat3_T7	-0.0517	hoc3_F4	-0.0474	hoc11_F4	-0.0460	hoc9_T7	-0.0307	hoc31_FC5	-0.0325
205	hoc31_FC5	-0.0371	hoc34_FC5	-0.0519	hoc20_AF3	-0.0476	hoc16_FC5	-0.0466	hoc34_T7	-0.0309	hoc16_T7	-0.0327
206	hoc12_T7	-0.0378	delta_AF3	-0.0522	hoc20_F4	-0.0477	hoc9_F7	-0.0470	stat6_AF3	-0.0316	hoc24_F4	-0.0336
207	hoc6_F4	-0.0387	fd_F4	-0.0522	hoc20_FC5	-0.0482	hoc27_F7	-0.0470	hoc25_FC5	-0.0317	fd_AF3	-0.0339
208	hoc9_F7	-0.0389	hoc12_FC5	-0.0537	hoc6_F7	-0.0487	hoc11_FC5	-0.0473	hoc14_T7	-0.0319	cpxty_AF3	-0.0343
209	hoc1_AF3	-0.0396	hoc25_F7	-0.0539	hoc13_T7	-0.0496	hoc29_T7	-0.0475	hoc29_FC5	-0.0326	delta_FC5	-0.0349
210	hoc1_F4	-0.0400	cpxty_AF3	-0.0546	hoc14_T7	-0.0498	hoc18_F4	-0.0482	stat4_T7	-0.0328	hoc4_FC5	-0.0350
211	hoc24_F7	-0.0400	hoc23_F7	-0.0549	hoc21_F4	-0.0502	hoc13_F4	-0.0484	hoc4_FC5	-0.0329	hoc21_F7	-0.0350
212	hoc5_F4	-0.0406	fd_FC5	-0.0550	hoc9_AF3	-0.0502	hoc28_T7	-0.0487	stat1_T7	-0.0330	hoc26_FC5	-0.0354
213	hoc10_F4	-0.0406	delta_F7	-0.0550	hoc27_F7	-0.0504	hoc35_AF3	-0.0487	hoc19_F4	-0.0331	hoc13_F7	-0.0364
214	hoc23_FC5	-0.0414	hoc16_AF3	-0.0551	hoc12_AF3	-0.0507	hoc10_F4	-0.0489	hoc10_F7	-0.0336	hoc16_F4	-0.0375
215	hoc29_FC5	-0.0417	stat5_T7	-0.0551	beta_AF3	-0.0511	hoc15_F4	-0.0496	hoc33_FC5	-0.0340	hoc17_T7	-0.0377
216	hoc33_F7	-0.0432	hoc9_F4	-0.0552	hoc25_T7	-0.0515	hoc21_F7	-0.0497	hoc3_F4	-0.0342	hoc15_T7	-0.0378
217	hoc1_F7	-0.0434	hoc25_AF3	-0.0552	hoc21_AF3	-0.0517	hoc15_FC5	-0.0500	hoc11_T7	-0.0344	hoc30_FC5	-0.0383
218	hoc15_T7	-0.0438	theta_AF3	-0.0553	hoc24_FC5	-0.0518	cpxty_F7	-0.0501	fd_AF3	-0.0362	hoc6_FC5	-0.0383
219	hoc4_T7	-0.0444	hoc26_F7	-0.0554	hoc11_F7	-0.0527	hoc17_F4	-0.0508	hoc13_T7	-0.0368	hoc31_F4	-0.0390
220	hoc2_F7	-0.0446	actvt_AF3	-0.0559	hoc10_AF3	-0.0529	hoc23_F7	-0.0513	hoc12_FC5	-0.0371	hoc9_FC5	-0.0391
221	hoc23_T7	-0.0452	se_AF3	-0.0559	hoc7_F7	-0.0534	hoc27_F4	-0.0516	hoc19_FC5	-0.0372	stat2_F7	-0.0407
222	hoc18_F4	-0.0457	stat4_F4	-0.0560	hoc22_F4	-0.0536	hoc15_F7	-0.0517	hoc13_AF3	-0.0372	hoc22_F7	-0.0408
223	hoc14_T7	-0.0470	hoc28_FC5	-0.0561	hoc17_FC5	-0.0541	hoc26_F7	-0.0523	hoc36_FC5	-0.0373	hoc3_F4	-0.0415
224	hoc22_FC5	-0.0482	cpxty_FC5	-0.0577	hoc29_F7	-0.0544	hoc24_F7	-0.0523	hoc20_F4	-0.0375	hoc18_F4	-0.0425
225	hoc20_F4	-0.0484	hoc35_F4	-0.0581	hoc12_T7	-0.0546	hoc17_F7	-0.0529	hoc28_FC5	-0.0377	hoc10_FC5	-0.0430
226	hoc27_F7	-0.0490	hoc30_FC5	-0.0583	hoc22_T7	-0.0546	hoc19_F4	-0.0531	hoc26_F4	-0.0378	hoc21_F4	-0.0431
227	hoc8_F7	-0.0502	hoc33_FC5	-0.0590	hoc19_AF3	-0.0549	hoc23_F4	-0.0536	hoc24_FC5	-0.0379	hoc27_F4	-0.0433
228	hoc17_F4	-0.0504	hoc6_T7	-0.0590	hoc15_AF3	-0.0561	hoc25_F7	-0.0536	hoc30_FC5	-0.0386	hoc20_F7	-0.0434
229	hoc32_F7	-0.0505	hoc12_F4	-0.0590	hoc24_T7	-0.0562	hoc25_F4	-0.0537	hoc21_F4	-0.0387	hoc19_F7	-0.0436
230	hoc26_F7	-0.0508	hoc15_F4	-0.0595	hoc15_F7	-0.0563	hoc12_F4	-0.0545	hoc8_T7	-0.0388	hoc26_F4	-0.0443
231	hoc19_F4	-0.0510	hoc29_FC5	-0.0599	hoc13_AF3	-0.0568	hoc20_F7	-0.0546	hoc4_F4	-0.0396	hoc25_F4	-0.0455
232	hoc22_T7	-0.0515	hoc26_AF3	-0.0602	hoc21_T7	-0.0569	hoc18_F7	-0.0546	hoc10_FC5	-0.0400	hoc19_F4	-0.0459
233	hoc17_T7	-0.0516	hoc14_F4	-0.0602	hoc33_F7	-0.0573	stat1_F4	-0.0547	hoc6_T7	-0.0409	cpxty_FC5	-0.0466
234	hoc25_F7	-0.0533	theta_F7	-0.0603	hoc18_FC5	-0.0573	hoc34_F4	-0.0550	hoc1_T7	-0.0414	hoc3_F7	-0.0469
235	hoc18_FC5	-0.0549	hoc11_F4	-0.0606	hoc25_F7	-0.0580	hoc36_AF3	-0.0554	hoc31_FC5	-0.0416	hoc17_F4	-0.0476
236	hoc31_F7	-0.0551	se_F7	-0.0610	hoc19_T7	-0.0584	hoc16_F7	-0.0558	hoc5_FC5	-0.0420	hoc34_FC5	-0.0481
237	hoc22_F4	-0.0560	actvt_F7	-0.0610	hoc16_T7	-0.0592	hoc35_F4	-0.0569	hoc21_FC5	-0.0421	stat1_T7	-0.0482
238	hoc30_F7	-0.0562	stat6_FC5	-0.0610	hoc15_T7	-0.0597	hoc26_F4	-0.0570	hoc11_F7	-0.0442	hoc7_FC5	-0.0482
239	stat1_AF3	-0.0562	hoc32_FC5	-0.0611	hoc31_F7	-0.0599	stat1_AF3	-0.0578	hoc32_FC5	-0.0444	hoc20_F4	-0.0491
240	hoc2_FC5	-0.0563	hoc27_FC5	-0.0611	hoc26_F7	-0.0602	hoc34_T7	-0.0584	hoc20_FC5	-0.0444	hoc6_F4	-0.0513
241	hoc13_T7	-0.0569	hoc27_AF3	-0.0614	hoc9_T7	-0.0603	hoc14_F4	-0.0585	hoc9_FC5	-0.0456	hoc18_F7	-0.0515
242	hoc29_F7	-0.0570	hoc17_AF3	-0.0614	hoc16_AF3	-0.0604	hoc12_FC5	-0.0586	hoc10_T7	-0.0461	hoc14_F7	-0.0520
243	hoc20_T7	-0.0575	hoc19_F7	-0.0617	hoc32_F7	-0.0605	hoc33_T7	-0.0589	hoc27_FC5	-0.0472	hoc11_FC5	-0.0522
244	hoc20_FC5	-0.0579	stat2_AF3	-0.0623	hoc23_FC5	-0.0613	hoc32_F4	-0.0593	hoc23_FC5	-0.0493	hoc12_FC5	-0.0523
245	stat1_FC5	-0.0590	hoc13_F4	-0.0624	hoc28_F7	-0.0616	hoc30_T7	-0.0602	hoc5_T7	-0.0505	fd_FC5	-0.0558
246	hoc21_F4	-0.0596	hoc34_F4	-0.0626	hoc13_F7	-0.0620	hoc30_F4	-0.0602	hoc22_F4	-0.0507	beta_F7	-0.0565

Rank	Subject S1		Subject S2		Subject S3		Subject S4		Subject S5		Subject S6	
	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
247	hoc19_FC5	-0.0598	hoc18_F7	-0.0637	hoc22_FC5	-0.0624	hoc32_T7	-0.0609	hoc25_F4	-0.0514	hoc30_F4	-0.0573
248	hoc2_AF3	-0.0601	hoc28_AF3	-0.0637	hoc17_T7	-0.0630	hoc16_F4	-0.0612	hoc22_FC5	-0.0516	hoc5_FC5	-0.0578
249	hoc7_F4	-0.0613	hoc26_FC5	-0.0639	hoc19_FC5	-0.0631	hoc13_FC5	-0.0613	hoc24_F4	-0.0516	hoc16_F7	-0.0579
250	hoc19_T7	-0.0617	fd_AF3	-0.0639	hoc18_T7	-0.0639	hoc33_F4	-0.0619	hoc23_F4	-0.0523	theta_FC5	-0.0598
251	hoc21_FC5	-0.0621	hoc17_F7	-0.0644	hoc1_T7	-0.0645	hoc36_F4	-0.0623	hoc8_FC5	-0.0526	hoc15_F7	-0.0615
252	hoc21_T7	-0.0623	mblty_AF3	-0.0645	hoc12_F7	-0.0647	hoc31_T7	-0.0635	hoc6_FC5	-0.0528	hoc17_F7	-0.0627
253	hoc9_F4	-0.0623	stat6_AF3	-0.0646	hoc20_T7	-0.0649	hoc14_FC5	-0.0649	hoc4_T7	-0.0529	hoc28_F4	-0.0633
254	hoc16_T7	-0.0624	stat4_AF3	-0.0653	hoc18_AF3	-0.0651	hoc31_F4	-0.0659	hoc7_FC5	-0.0553	hoc29_F4	-0.0658
255	hoc28_F7	-0.0645	stat2_F7	-0.0653	hoc17_AF3	-0.0658	hoc19_F7	-0.0662	hoc3_F7	-0.0566	hoc4_F4	-0.0665