

UNIVERSIDAD DEL NORTE

**Probabilistic graphical models for ERP-based Brain
computer interfaces**

By

Jorge Humberto Cristancho Cuervo

Submitted to the Electrical and Electronics Engineering Department in partial fulfillment of
the requirements for the degree of
Doctor in Philosophy
In Electrical and Electronics Engineering

Universidad del Norte

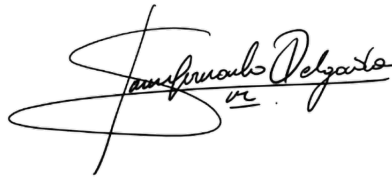
July 2022

UNIVERSIDAD DEL NORTE
AUTHORIZATIONS FOR PRESENTATION OF THE THESIS BOOK

Having reviewed the thesis project, each of the signatories authorizes its presentation to the
jury:



Student: Jorge Humberto Cristancho Cuervo



Advisor: Jaime Fernando Delgado Saa, Ph.D.



Advisor: Lácides Antonio Ripoll Solano, Ph.D.

PROBABILISTIC GRAPHICAL MODELS FOR ERP-BASED BRAIN COMPUTER INTERFACES

By Jorge Humberto Cristancho Cuervo

Electrical and Electronics Engineering, Ph.D. thesis, 2022

Thesis advisors: Jaime Fernando Delgado Saa, Ph.D, Assoc. Prof. Lácides Antonio Ripoll Solano.

Keywords: Probabilistic Graphical Model, Temporal features, Spatial features, Conditional Random Fields, Event related potentials, Brain-computer interfaces, Sensorimotor rhythms, Frequency band, Electroencephalogram, Electrocorticogram,

ABSTRACT

An event related potential (ERP) is an electrical potential recorded from the nervous system of humans or other animals. An ERP is observed after the presentation of a stimulus. Some examples of the ERPs are P300, N400, among others. Although ERPs are used very often in neuroscience, its generation is not yet well understood and different theories have been proposed to explain the phenomena. ERPs could be generated due to changes in the alpha rhythm, an internal neural control that reset the ongoing oscillations in the brain, or separate and distinct additive neuronal phenomena. When different repetitions of the same stimuli are averaged, a coherence addition of the oscillations is obtained which explain the increase in amplitude in the signals.

Two ERPs are mostly studied: N400 and P300. N400 signals arise when a subject tries to make semantic operations that support neural circuits for explicit memory. N400 potentials have been observed mostly in the rhinal cortex. P300 signals are related to attention and memory operations. When a new stimulus appears, a P300 ERP (named P3a) is generated in the frontal lobe. In contrast, when a subject perceives an expected stimulus, a P300 ERP (named P3b) is generated in the temporal – parietal areas. This implicates P3a and P3b are related, suggesting a circuit pathway between the frontal and temporal–parietal regions, whose existence has not been verified.

A Brain-Computer Interface (BCI) is a communication system in which messages or commands that a subject sends to the external world comes from some brain signals rather

than the peripheral nerves and muscles. BCI uses sensory-motor rhythms or ERP signals, so a classifier is needed to distinguish among the correct stimuli and the incorrect ones. The majority of studies about ERPs classification uses the following scheme: when brain signals are acquired, they are pre-processed to get some features that serve as inputs to a classifier. Then, a classifier is proposed for discriminating ERP signals, without taking in account the nature of the process, in most of the cases.

In this work, we propose to use probabilistic graphical models for the modeling of the temporal and spatial dynamics of the brain signals with applications to BCI. Our work will be focused on ERPs and the corresponding modeling of such phenomena. Graphical models have been selected for its flexibility and capacity to incorporate prior information. This flexibility has been used before for modeling temporal dynamics alone. We expect that the model reflects some aspects of the brain functioning related to the ERPs, by including spatial and temporal information. We might expect too an improvement in the performance of a BCI system when using the proposed model.

ACKNOWLEDGEMENTS

In the final part of the present chapter of my life, I have a lot to thank, as well as many people to thank for their support, presence, advice and even exhortations. All these contributions have been well received, even if I could not understand them at first. To God, who has been my shelter, sustenance and support at all times. To the life and the universe, who have not yet rejected me from this existential plane. To my parents, who rejoiced, suffered and celebrated with me or for me during all these years. To my sister, who many times helped me to keep calm in moments of restlessness, anxiety and uncertainty. To them: infinite thanks.

I also have a lot to thank my extended family: my grandparents Juan and Lucila, who were and have been aware of my progress, as well as my uncles, aunts, and cousins. Special mention must be made to my uncle Adriano, who has been my support in this academic adventure, started since 2015, and to my aunt Lili, who has received me with my uncle in their home for many years. I must also thank my cousin Andrés Camilo, from whom I have received valuable advice and counsel for the completion of this work. To Juan Manuel, Nicolás and Daniel, who in one way or another helped me in some administrative and form issues, along with their partners Alejandra, Linda, Yanan and Daniela. All of them have a place in my heart.

To my classmates, colleagues and professors who shared space with me at Universidad del Norte, I remember them fondly. First of all, to the professors and administrative staff of the Department of Electrical and Electronic Engineering, with whom I had the privilege and opportunity to deal on countless occasions, from whom I received valuable support, invaluable advice and exhortations that are appreciated now and in my future. Special mention must be made to professors Juan Carlos Vélez, Luis Torres, Ingrid Oliveros, José Soto, César Viloria and Jaime Delgado, with whom I had the opportunity to work as a graduate assistant and learn from their experience as teachers and professionals. To the directors of the Electrical and Electronics Department, for their management and work over the years, to ensure the proper development of the doctoral program. To the professors of CEDU, especially those with whom I shared different spaces, especially the reading clubs: thank you very much for your advice, points of view and shared experiences. From all of

them I had something to learn for my personal, academic and professional life. To the UNINORTE scholarship program, I extend a special gratitude for the opportunity through them to carry out my doctoral studies, under the Grant 2015-29364 0.

To my doctoral classmates, with whom we suffered and enjoyed many moments, I will always keep them in mind, especially those with whom I had the privilege of seeing subjects and fighting in the front line: Jamer, Liesle, Wendy, Sandra and Aura. To my colleagues in the BSPAI lab and research group, with whom I shared space and experience for years, and also received collaborations from them, I appreciate them immensely: Cristian, Catherine, Pedro, Mario, Daniela, Dayan, Harold and other colleagues. I also thank the teachers of the group immensely for their advice, appreciation and shared moments, regardless of time and current circumstances: Juan Pablo Tello, Winston Percybrooks, Norelli Schettini and Jaime Delgado. To my colleagues who were in the different masters offered by the Electrical and Electronics Department, with whom I also had the privilege of sharing time and space, I am immensely grateful to them. Special mention should be made to Loraine Navarro and Daniela Charris, with whom I had the privilege of working, being graduate assistants, and Ana Maria Lopez, from whom I learned about her dedication and effort to get her Master's degree.

To my brothers in faith and leaders, with whom I have crossed paths from my childhood until today, and who have been accompanying me in this journey of life. Their names are so many that they would not fit in this space, but I remember them all fondly and thank them for their prayers, words of encouragement, shared moments and support in difficult times. To my friends, with whom we laughed, suffered, cried and enjoyed this time. In Bucaramanga, I especially mention Yenny Carolina, who has supported me with her advice and her time, Ana Milena, whose support and advise have given me breath in different situations, Viviana, who has been aware of my work and person, and Carolina Rueda, who came as reinforcement at this stage of the road; in Barranquilla, the combo of the Festival del Guandul Chino, Luz Marina (Lumi), the Adrianas (Anaya and Margarita), Karen Looren, Anita Mevania, Analuz, Dianita.

To my psychologists and psychiatrists, who have helped me to direct my thoughts and feelings regarding my future as a researcher, professional and person. In particular, I would like to thank Milagro Villareal Bacareo and Julio Quijano, who were attentive to

my case, and gave me advice and strategies to work on my thesis while I was dealing with my mental health, and also taught me the importance of closing cycles and completing stages. stages.

A special mention must go to those people who, without sharing many genes, have become special to my life, so much so that I cannot imagine my doctoral (and personal) life without their presence as counselors, support and, in some cases, exhorters: Julieta Paola, from whom I have learned not to force and to let it flow; Diana Astrid, from whom I have learned that perseverance bears fruit; María Isabel, from whom I have learned to say yes and no; Diana Patricia, with whom I have had the opportunity to share and discuss my points of view; Ivette Rocío, whose friendship of years has led me to become a little more like her, even if she has not taken much from me. To them, a big “thank you” for everything I have received over the years, for their friendship, their appreciation.

Finally, I cannot end these lines without special thanks to the professors and engineers Jaime Delgado, María Gabriela Calle, Lácides Ripoll and Ingrid Oliveros, from whom I have received their professional, academic and personal support in the development of this doctorate. Without them, this story could not have been told from beginning to end.

AGRADECIMIENTOS

En la parte final del presente capítulo de mi vida, tengo mucho por agradecer, así como a muchos por quiénes agradecer sus apoyos, presencias, consejos y hasta exhortaciones. Todos estos aportes han sido bien recibidos, así yo no los haya podido comprender en un principio. A Dios, quien ha sido mi refugio, sustento y respaldo en todo momento. A la vida y al universo, que aún no me han rechazado de este plano existencial. A mis padres, quienes se alegraron, sufrieron y celebraron conmigo o por mí durante todos estos años. A mi hermana, quien muchas veces me ayudó a mantener la calma en momentos de inquietud, ansiedad e incertidumbre. A ellos: infinitas gracias.

A mi familia extendida, también tengo mucho por agradecer: a mis abuelos Juan y Lucila, quienes estuvieron y han estado pendientes de mi progreso, al igual que mis tíos, tías, primos y primas. Especial mención he de hacer a mi tío Adriano, quien ha sido mi respaldo en esta aventura académica, emprendida desde 2015, y a mi tía Lili, quien me ha recibido junto a mi tío en su hogar por muchos años. También he de agradecer a mi primo Andrés Camilo, de quien he recibido consejos y asesorías muy valiosas para la culminación de este trabajo. A Juan Manuel, Nicolás y Daniel, quienes de una u otra forma me ayudaron en algunas cuestiones administrativas y de forma, junto con sus parejas Alejandra, Linda, Yanan y Daniela. Tienen un lugar ganado en mi corazón.

A mis compañeros, colegas y profesores que compartieron espacio conmigo en la Universidad del Norte, los recuerdo con cariño. En primera instancia, a los docentes y personal administrativo del departamento de Ingeniería Eléctrica y Electrónica, con quienes tuve el privilegio y la oportunidad de tratar en infinitas oportunidades, de quienes recibí valioso apoyo, invaluable consejos y exhortaciones que se aprecian ahora y en mi porvenir. Especial mención tienen los profesores Juan Carlos Vélez, Luis Torres, Ingrid Oliveros, José Soto, César Viloria y Jaime Delgado, con quienes tuve la oportunidad de trabajar como asistente graduado y aprender de su experiencia como docentes y profesionales. A las directivas del departamento de Eléctrica y Electrónica, por su gestión y labor en estos años, en procura del buen desarrollo del programa del doctorado. A los docentes del CEDU, en especial con quienes compartí en diferentes espacios, en especial los clubes de lectura: muchas gracias por sus consejos, puntos de vista y experiencias

compartidas. De todos ellos tuve algo por aprender para mi vida personal, académica y profesional. Al programa de becas UNINORTE, les doy un agradecimiento especial por la la oportunidad brindada a través de ellos para sacar adelante mis estudios de doctorado, bajo la beca 2015-29364 0.

A mis compañeros de doctorado, con quienes sufrimos y gozamos muchos momentos, los tendré siempre presentes, en especial con quienes tuve el privilegio de ver asignaturas y luchar en primera línea: Jamer, Liesle, Wendy, Sandra y Aura. A mis compañeros del laboratorio y grupo de investigación BSPAI, con quienes compartí espacio y conviví por años, y también recibí colaboraciones de su parte, los aprecio inmensamente: Cristian, Catherine, Pedro, Mario, Daniela, Dayan, Harold y demás compañeros. A los docentes del grupo también agradezco inmensamente por sus consejos, aprecio y momentos compartidos, independiente del tiempo y circunstancias actuales: Juan Pablo Tello, Winston Percybrooks, Norelli Schettini y Jaime Delgado. A mis colegas que estuvieron en las diferentes maestrías ofrecidas por el departamento de Eléctrica y Electrónica, con quienes también tuve el privilegio de compartir tiempo y espacio, tengo un agradecimiento inmenso con ellos. Cabe mencionar en especial a Loraine Navarro y Daniela Charris, con quienes tuve el privilegio de trabajar, siendo ellas asistentes graduadas, y Ana María López, de quien aprendí su dedicación y esfuerzo por sacar adelante su maestría.

A mis hermanos en la fe y líderes, con quienes he cruzado camino desde mi infancia hasta hoy, y me han estado acompañando en este trayecto de la vida. Sus nombres son tantos que no cabrían en este espacio, pero a todos ellos los recuerdo con cariño y les agradezco sus oraciones, palabras de aliento, momentos compartidos y apoyo en momentos difíciles. A mis amigos, con quienes reímos, sufrimos, lloramos y nos gozamos este tiempo. En Bucaramanga, menciono especialmente a Yenny Carolina, quien me ha apoyado con sus consejos y su tiempo, Ana Milena, cuyo apoyo y consejos me han brindado aliento en diversas oportunidades, Viviana, que ha estado pendiente de mi trabajo y persona, y Carolina Rueda, quien llegó como refuerzo en esta etapa del camino; en Barranquilla, al combo del Festival del Guandul Chino, a Luz Marina (Lumi), las Adrianas (Anaya y Margarita), Karen, Anita Mevania, Analuz, Dianita.

A mis psicólogos y psiquiatras, quienes me han ayudado a encaminar mis pensamientos y sentimientos respecto a mi devenir como investigador, profesional y persona. En especial,

quiero agradecer a Milagro Villareal Bacareo y a Julio Quijano, quienes estuvieron pendientes de mi caso, y me brindaron consejos y estrategias para trabajar mi tesis lidiaba con mi salud mental, y me enseñaron también la importancia de cerrar ciclos y culminar etapas.

Mención aparte tiene aquellas personas que, sin compartir tantos genes, se han vuelto especiales para mi vida, tanto así que no imagino mi vida de doctorado (y personal) sin ellas como consejeras, apoyo y, en algunos casos, exhortadoras: Julieta Paola, de quien he aprendido a no forzar y dejar fluir; Diana Astrid, de quien he aprendido que la perseverancia brinda sus frutos; de María Isabel, de quien he aprendido a decir sí y no; de Diana Patricia, con quien he tenido la oportunidad de compartir y debatir mis puntos de vista; de Ivette Rocío, cuya amistad de años me ha llevado a parecerme un poco más a ella, así ella no haya tomado mucho de mí. A ellas, un agradecimiento enorme por todo lo recibido en estos años, por su amistad, su aprecio.

Finalmente, no puedo terminar estas líneas sin un agradecimiento especial a los profesores e ingenieros Jaime Delgado, María Gabriela Calle, Lácides Ripoll e Ingrid Oliveros, de quienes he recibido su apoyo profesional, académico y personal en el desarrollo de este doctorado. Sin ellos, esta historia no habría podido ser contada desde el principio hasta el final.

LIST OF FIGURES

Figure 1-1. Graphical models from some linear Conditional Random Fields.....	8
Figure 2-1. Place of used electrodes in the dataset.	18
Figure 2-2. Averaged results of all subjects, for LSVM.....	23
Figure 2-3. Averaged results of all subjects, for Log Reg.	25
Figure 2-4. Averaged results of all subjects, for SWLDA.....	27
Figure 2-5. Averaged results of all subjects, for BLDA.....	29
Figure 3-1. Location of primary sensorimotor cortex, indicated in red and green.	34
Figure 4-1. Electrodes location in the dataset BCI Competition IV, dataset 2b, and their interactions.	45
Figure A-1-1. A linked MVHCRF graphical model (left) and a coupled MVHCRF graphical model (right) with three timestamps and two HCRF models.	60
Figure A-1-2. A graphical model of HUCRF.....	61
Figure A-1-3. Summarization of results of CRF implementation with CSP features.	67

LIST OF TABLES

Table 2-1. Averaged metrics by subject, for LSVM.....	24
Table 2-2. Averaged metrics by subject, for Log Reg.....	26
Table 2-3. Averaged metrics by subject, for SWLDA.	28
Table 2-4. Averaged metrics by subject, for BLDA.....	30
Table 3-1. Results of Accuracy, by subject and type of data. Overall performance: 0,526	36
Table 3-2. Results of Cohen's kappa index, by subject and type of data. Average performance: 0,394.....	37
Table 3-3. Results of Accuracy, by subject and classifier. Overall performance: 0,433	38
Table 3-4. Results of Cohen's kappa index, by subject and classifier. Overall performance: 0,278.....	39
Table 3-5. Confusion Matrix across all subjects for CRF, LDA and LSVM classifiers. Grayscale indicates the degree of distribution of assigned labels.	42
Table 3-6. Results of balanced accuracy, by subject and classifier.	42
Table 4-1. Results of Accuracy in LDA model, with and without correlation. Overall performance: 0,74.....	49
Table 4-2. Results of Accuracy in LDA, with and without Jaccard Distance Overall performance: 0,741.....	49
Table 4-3. Results of Accuracy in LDA model with correlation as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,751.	50
Table 4-4. Results of Accuracy in LDA model with correlation as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,751.50	
Table 4-5. Results of Accuracy in LDA model with Jaccard Distance as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,752.....	52

Table 4-6. Results of Accuracy in LDA model with Jaccard Distance as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,752..... 52

Table 4-7. Results of Accuracy in HCRF model, with and without correlation. Overall performance: 0,757..... 53

Table 4-8. Results of Accuracy in HCRF, with and without Jaccard Distance. Overall performance: 0,758..... 53

Table 4-9. Results of Accuracy in HCRF model with correlation as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,777. ... 54

Table 4-10. Results of Accuracy in HCRF model with correlation as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,777..... 54

Table 4-11. Results of Accuracy in HCRF model with Jaccard Distance as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,778. 56

Table 4-12. Results of Accuracy in HCRF model with Jaccard Distance as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,778. 56

Table A-1-1. Results of LDA, HCRF and MVHCRF implementation, by patient. 62

Table A-1-2. Results of HUCRF implementation, by patient. 63

Table A-1-3. Results of HUCRF implementation, by patient. 64

Table A-1-4. Results of CRF implementation with Alpha and Beta power band features, by patient. 65

Table A-1-5. Results of CRF implementation with CSP features, by patient. 68

LIST OF ACRONYMS

ANOVA: analysis of variance.

BCI: brain-computer interface.

BLDA: bayesian linear discriminant analysis.

CAR: common average reference.

CRF: conditional random fields.

CSP: common spatial patterns.

d.o.f.: degrees of freedom.

ECoG: Electrocorticography.

EEG: Electroencephalography.

EOG: Electro-oculography.

ERP: event-related potential.

FBCSP: Filter bank common spatial patterns.

HCRF: hidden conditional random fields.

HG: high gamma band.

HMM: hidden Markov model.

HUCRF: hidden-unit conditional random fields.

LDA: linear discriminant analysis.

LDCRF: latent dynamic conditional random fields.

LFC: low frequency components.

Log. Reg.: logistic regression.

LSVM: linear-kernel support vector machine.

MVHCRF: multi-view hidden conditional random fields.

PGM: probabilistic graphical model

QDA: quadratic discriminant analysis.

SVM: support vector machine.

SWT: sliding window technique.

SWLDA: step-wise linear discriminant analysis.

TABLE OF CONTENT

1	Introduction	1
1.1	Brain-computer interfaces	1
1.2	Some classifiers used in BCI.....	2
1.2.1	Linear and Quadratic Discriminant Analysis.....	3
1.2.2	Step-wise LDA	3
1.2.3	Bayesian LDA.....	3
1.2.4	Linear SVM	4
1.2.5	Logistic regression	5
1.3	Conditional random fields.....	5
1.3.1	Probabilistic graphical model	5
1.3.2	Linear CRF	6
1.3.3	Latent dynamic conditional random fields	7
1.3.4	Hidden Conditional Random Fields.....	7
1.4	Some performance metrics used in BCI.....	9
1.4.1	Accuracy	9
1.4.2	Cohen's kappa index κ	9
1.4.3	Balanced accuracy	10
1.5	Instantaneous interactions	11
1.5.1	Pearson correlation.....	11
1.5.2	Jaccard distance	11
1.6	Background	12
1.7	Problem statement.....	13
1.8	Justification	13
1.9	Hypotheses	14
1.9.1	First Hypothesis	14
1.9.2	Second Hypothesis.....	14

2	A Bootstrapping Method for Improving a P300 Speller Classification Performance	15
2.1	Preliminaries	15
2.2	Methods.....	17
2.2.1	Experiment and dataset description	17
2.2.2	Data processing.....	17
2.2.3	Classifiers.....	20
2.2.4	Performance metrics	20
2.2.5	Statistical analysis.....	20
2.3	Results and discussion.....	21
2.3.1	Number of bootstrapped samples.....	21
2.3.2	Type of training samples.....	22
2.3.3	Discussion	26
2.4	Conclusions	31
3	A First Approximation to Linear CRF classifiers for Finger Movement Classification	32
3.1	Preliminaries	32
3.2	Materials and methods	33
3.2.1	Experiment and dataset description	33
3.2.2	Data pre-processing	33
3.2.3	Classifiers to be compared.....	34
3.2.4	Performance metrics	35
3.2.5	Statistical analysis.....	35
3.3	Results and Discussion.....	36
3.3.1	Type of Data: High Gamma Band vs. Low-Frequency Component	36
3.3.2	CRF and LDCRF against other classifiers.....	37
3.3.3	Discussion.....	38

3.4	Conclusions	41
4	Analysis of instantaneous brain interactions contribution to a motor imagery classification task.....	43
4.1	Preliminaries	43
4.2	Materials and methods	44
4.2.1	Experiment and dataset description	44
4.2.2	Data pre-processing	46
4.2.3	Sliding window technique	46
4.2.4	Pearson correlation.....	46
4.2.5	Jaccard distance	47
4.2.6	Performance metrics	47
4.2.7	Statistical analysis.....	47
4.3	Results and Discussion.....	48
4.3.1	Results with LDA model	48
4.3.2	Results with HCRF model	51
4.3.3	Discussion.....	55
4.4	Conclusions	57
	Appendices.....	58
	Appendix 1. Additional models and multivariate BCI datasets tested.....	59
A-1 1.	Additional models tested.....	60
A-1.1.1.	Multi-view hidden conditional random fields (MVHCRF).....	60
A-1.1.2.	Hidden-Unit CRF (HUCRF)	61
A-1 2.	Additional Results.....	61
A-1.2.1.	MVHCRF	62
A-1.2.2.	CRF.....	63
A-1.2.3.	HUCRF.....	64
A-1.2.4.	CRF – Alpha and Beta power band	65
A-1.2.5.	CRF – Filter bank Common Spatial Patterns (FBCSP).....	66

References.....	70
-----------------	----

1 INTRODUCTION

1.1 Brain-computer interfaces

A Brain–Computer Interface (BCI) is a communication and control system that sends messages and commands to the external world without the normal use of nerves and muscles [1]. A typical BCI system is composed of a brain monitoring system, a signal pre-processing stage, a stage for extracting features of the pre-processed signal, and a decision stage where features are translated into commands or messages [2]. Some methods for monitoring the brain activity, for a BCI, are electroencephalography (EEG), electrocorticography (ECoG), functional magnetic resonance imaging (fMRI), among others [1]. Pre-processing signals from the brain activity monitor allows to drop undesired components of the signal that carries non useful information [2]. After pre-processing, some features of the signal are extracted, in order to be translated into commands and messages. Features may be extracted in one or more of the following domains of the signal: temporal, frequency, or spatial [2].

In health applications, BCIs have been used for brain-controlled wheelchairs [3], spelling devices [4], attention-deficit hyperactivity disorder (ADHD) disorder rehabilitation [5], and the interaction of Locked-in Syndrome (LIS) patients [6], [7]. Some explored applications further health care are games [8]–[10], attention monitors [11], tools for mobile devices [12], [13], control of items [14], navigation in virtual environments [3], [15], [16], gesture recognition [17], among others.

Some modalities of BCI activation are by evoked potentials [18], [19], brain rhythms and motor imagery [20], [21], readiness potentials [8], [22], slow cortical potentials [23], and error potentials [17], [24]. Some works have proposed to use steady – state evoked potentials as EEG patterns for developing visual [25], auditory [26], and vibratory [27] BCIs. Other studies propose to combine BCI with other electrophysiological signals [28]–[31]. Some studies implement brain–computer interfaces to detect mental states, as mental load [32] and drowsiness [33]. Other works treat to prove associations between emotions and BCIs, to study or control them [34]–[36].

1.2 Some classifiers used in BCI

Some features are extracted from a pre-processed signal recorded from the brain monitor activity, to be translated into commands and messages [2]. In most of BCI systems, a vector data groups features extracted from the brain signals. A classifier performs mostly the translation to commands [37], [38], assigning a vector data automatically to one of a finite number of discrete categories [39].

For BCI systems, the main problem is finding an adequate classifier algorithm for EEG signals [38], taking into account their unique features: 1) Signals are noisy and have many outliers, so their signal-to-noise ratio (SNR) is poor [19], [38]. 2) Since EEG signals are recorded from multiple channels, they are concatenated, therefore their dimensionality is high. 3) Temporal [40], [41], frequency [42], [43] or spatial domain from EEG data [2] provide the adequate information. 4) Signals of interest are non-stationary. 5) Training and testing datasets are small, due to the difficulty of getting enough volunteers for the experiments, as well as the difficulty that they do the tasks to run without exhaustion or other problems.

In general terms, the classification task consists of assigning a label or *class* to a vector of features \mathbf{x} and K classes, employing some mathematical criteria. For optimal classification, is necessary to know the probability of a class $C = k$, given \mathbf{x} , namely the *posterior probability* $p(C = k | X = \mathbf{x})$. By using the Bayes theorem, we get the *posterior probability* of a class k with class-conditional density probability $f_k(\mathbf{x})$, given a vector of features \mathbf{x} , and a set of prior probabilities by class π_k [44]:

$$p(C = k | X = \mathbf{x}) = \frac{f_k(\mathbf{x})\pi_k}{\sum_{j=1}^K f_j(\mathbf{x})\pi_j} \quad k = 1, \dots, K \quad (1-1)$$

The classification task could be seen from one of two potential points of view. The first one divides the classification problem into two stages. The *inference stage* learns a probabilistic model of data given the class. Then, the *decision stage* implements the theorem of Bayes to determine the class of each sample data. A classifier implemented in this way is a *generative classifier* [39]. Based *linear discriminant analysis (LDA)* classifiers are generative because they assume Gaussian distributions in the data [39], [44]. The second viewpoint specifies that data could map directly to a class by a

classification model. The model comes from either a probabilistic *discriminant model* of the class, given data, or a deterministic *discriminant function*. A classifier that uses the latter approach is a *discriminative classifier* [39]. *Logistic regression (Log Reg)* and the *support-vector machine (SVM)* are examples of discriminative classifiers that use a probabilistic model and a discriminant function, respectively.

1.2.1 Linear and Quadratic Discriminant Analysis

A discriminant classifier is a function that allocates an input vector \mathbf{x} to one of K classes \mathcal{C}_k [39]. If we assume that $f_k(\mathbf{x})$ in Equation (1-1) is multivariate Gaussian with vector of mean μ_k and covariance matrix Σ_k , we get the following discriminant function $\delta_k(\mathbf{x})$ [44]:

$$\delta_k(\mathbf{x}) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (\mathbf{x} - \mu_k)^T \Sigma_k^{-1} (\mathbf{x} - \mu_k) + \log \pi_k \quad (1-2)$$

If each class has its covariance matrix Σ_k , the discriminant function is quadratic, by making the decision boundary between a pair of classes k and l as $\delta_k(\mathbf{x}) = \delta_l(\mathbf{x})$. However, if we suppose a shared covariance matrix Σ for all classes, the discriminant $\delta_k(\mathbf{x})$ becomes linear [39], [44]. Notice that based *Linear Discriminant Analysis (LDA)* classifiers are generative since they mostly assume Gaussian distributions in the data [40]. Also, notice that LDA and QDA are static models, since the time is not taken into account as a parameter.

1.2.2 Step-wise LDA

A *Step-wise LDA (SWLDA)* classifier is an LDA modified version where a stepwise regression is used before the classification duty [45]. Unlike other LDA-based classifiers, it chooses the coefficients of the model regression iteratively, according to a statistical standard [46]. As a result, the model obtained is more compact than the model defined in Equation (1-2). Additional details about SWLDA are can be found in [47].

1.2.3 Bayesian LDA

When the coefficients of the LDA model are chosen according to Bayesian criteria, we get an LDA classifier based on *Bayesian interpolation (BLDA)*. According to the literature, BLDA provides better results than LDA or SWLDA [48], [49]. Like the SWLDA classifier, the coefficients are achieved by iteration. However, the statistical standards for choosing

corrections are based on the Bayes rule, and they are not added or removed from the model [50]. Further details about BLDA can be obtained from [48].

1.2.4 Linear SVM

For a two-class classification problem, a training set contains N feature vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ with their corresponding label values $\{t_1, \dots, t_n, \dots, t_N\}$ where $t_n \in \{-1, 1\}$. We assume the following linear classification model [39]:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1-3)$$

Where \mathbf{w} is a vector of parameters and b is a bias term. Equation (1-3) satisfies $t_n y(\mathbf{x}_n) > 0$ for all training feature vectors. Here, the problem is to find the parameters \mathbf{w} and b that maximizes the distance of the nearest \mathbf{x}_n training vectors to a separation hyperplane that satisfies the requirements of Equation (1-3). The distance is known as the *margin*, the separation hyperplane is the *decision boundary* and the nearest \mathbf{x}_n training vectors to the decision boundary are the *vector supports* [44].

The perpendicular distance between an arbitrary point \mathbf{x} and a hyperplane is calculated by $|y(\mathbf{x})|/\|\mathbf{w}\|$. Since the interest relies in the correct classification of \mathbf{x}_n that satisfies $t_n y(\mathbf{x}_n) > 0$, the distance of any feature vector to the decision boundary becomes [39]:

$$\frac{t_n y(\mathbf{x}_n)}{\|\mathbf{w}\|} = \frac{t_n (\mathbf{w}^T \mathbf{x}_n + b)}{\|\mathbf{w}\|} \quad (1-4)$$

Since the optimization problem consists of maximizing the margin between the vector supports and the decision boundary with a minimum quantity of vector supports, the optimal solution is found by solving [39], [44]:

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_n [t_n (\mathbf{w}^T \mathbf{x}_n + b)] \right\} \quad (1-5)$$

By making the term inside the minimization equal to 1 as the margin distance, the distance restriction becomes $t_n y(\mathbf{x}_n) \geq 1$ for any vector of the training set, there will be, as minimum, two points that satisfies the minimum margin distance [39], so the optimization problem is reduced to maximize $1/\|\mathbf{w}\|$, which is equivalent to minimize $\|\mathbf{w}\|^2$, so the solution is simplified by solving [39], [44]:

$$\arg \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (1-6)$$

Support-vector machine (SVM) classifiers have been implemented in various earlier studies associated to BCIs, including P300 spellers [51]–[55], motor imagery [56]–[62] and finger flexion classification [63]–[66].

1.2.5 Logistic regression

Fewer works related to BCIs implement logistic regression classifiers [67], [68]. *Logistic regression* belongs to the *log-linear models* family, implemented in discriminative classifiers [39], [44]. More specifically, the logistic regression models the probability of K classes given a feature vector \mathbf{x} via linear functions, in terms of $K - 1$ log-odds [44]:

$$\ln \left(\frac{p(C = k | X = \mathbf{x})}{p(C = K | X = \mathbf{x})} \right) = w_{0k} + \mathbf{w}_k^T \mathbf{x} \quad k = 1, \dots, K - 1 \quad (1-7)$$

Where \mathbf{w}_k is a vector of parameters and w_{0k} is a bias term for each class $k = \{1, \dots, K - 1\}$. Ensuring that the sum of all probabilities is 1 and each one remains in $[0, 1]$, we get the *logistic regression* [44]:

$$p(C = k | X = \mathbf{x}) = \frac{\exp(w_{0k} + \mathbf{w}_k^T \mathbf{x})}{1 + \sum_{j=1}^{K-1} \exp(w_{0j} + \mathbf{w}_j^T \mathbf{x})} \quad k = 1, \dots, K - 1 \quad (1-8)$$

Further details about logistic regression are available in [39].

1.3 Conditional random fields

1.3.1 Probabilistic graphical model

A probabilistic graphical model (PGM) is a graph-based representation of a probabilistic function [69], [70]. It is based on two or more variables, represented as nodes, interacting in a one or more local functions, represented as edges. When a local function expresses a probability, it is represented as a directed edge. Conversely, if the local function represents a product of functions involving the nodes, it is represented as an undirected edge [69], [70]. In the study, we chose only probabilistic models represented with undirected edges.

The use of probabilistic graphical models (PGM) as classifiers for BCI has been proposed since 2000s [71]. The most used PGMs are hidden Markov models (HMM) and conditional

random fields (CRF). The main difference lies in HMM graphs are directed, while CRF graphs are undirected, that implies some differences in the statistical inference [72].

Some applications that use a PGM as classifier are the following:

- P300-based language models and word predictors using HMMs [40], [73]–[75].
- Motor imagery tasks (HMM [71], [76]–[78], and Hidden CRF (HCRF) [79]–[81].
- Finger movement decoding with Latent Dynamics CRF (LDCRF) [82].
- EEG rhythms with HMM and input/output HMM (IOHMM) [83].

1.3.2 Linear CRF

The conditional random field classifier is a member of the PGM family. CRF represents complex distributions through products of local factors on small subsets of variables [69]. Unlike other PGM models, CRF does not consider the dependencies among entries but models directly the conditional distribution between a set of input vectors $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t, \dots, \mathbf{x}_{T-1}, \mathbf{x}_T\}$ and its corresponding output labels $\mathbf{y} = \{y_1, y_2, \dots, y_{t-1}, y_t, \dots, y_{T-1}, y_T\}$ as $p(\mathbf{y}|\mathbf{x})$. It makes CRF a discriminative classifier. In its linear form, the CRF model is as Equation (1-9) shows.

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{1}{Z} \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, \mathbf{x}_t) \right\} \\ &= \frac{1}{Z} \exp \left\{ \sum_{t=1}^T \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, \mathbf{x}_t) \right\} \end{aligned} \quad (1-9)$$

Here, $\theta = \{\theta_k\} \in \mathbb{R}^K$ is a parameter vector, and $\{f_k(y_t, y_{t-1}, \mathbf{x}_t)\}_{k=1}^K$ is a set of real-valued feature functions. Z is the *partition function*, a normalization function along \mathbf{y} of Equation (1-9). Figure 1-1 a) illustrates the graphical representation of a linear CRF model, where undirected edges represent feature functions that link any y_t outputs with their corresponding \mathbf{x}_t inputs or their previous outputs y_{t-1} .

Feature functions does not indicate any probabilistic relationship among engaged variables, as other PGM does. However, it has an important property that comes from Markov chains: each output variable y_t depends exclusively of its predecessor y_{t-1} and the current input variables, being conditionally independent of other predecessors.

1.3.3 Latent dynamic conditional random fields

To expand the CRF models, we could add a set of hidden variables $\mathbf{h} = \{h_m\} \in \mathbb{R}^M$ not observed during the training stage [84]. Specifying the model to have disjoint sets of hidden states associated with each value y_j of output label: $h_j \in \mathcal{H}_{y_j}$, a Latent Dynamic Conditional Random Field takes the form as in Equation (1-10).

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \sum_{h_j, h_{j'} \in \mathcal{H}_{y_j}} \exp \left\{ \sum_{t=1}^T \sum_{k=1}^K \theta_k f_k(h_t, h_{t-1}, \mathbf{x}_t, y_t) \right\} \quad (1-10)$$

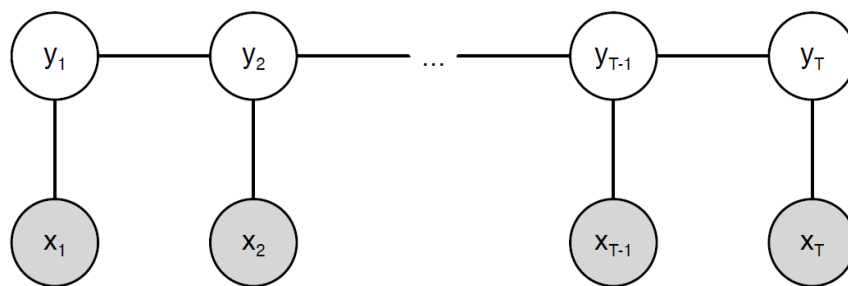
The partition function Z is the normalization function that takes all possible values of hidden variables along \mathbf{y} of Equation (1-10). As CRF, a hidden variable $h_{j,t}$ depends only on its predecessor $h_{j',t-1}$ and the current input variables. Figure 1-1 b) illustrates the graphical representation of a linear LDCRF model, where undirected edges represent feature functions that link any $h_{j,t}$ hidden states with their corresponding \mathbf{x}_t inputs and y_t outputs or their previous hidden states $h_{j',t-1}$.

1.3.4 Hidden Conditional Random Fields

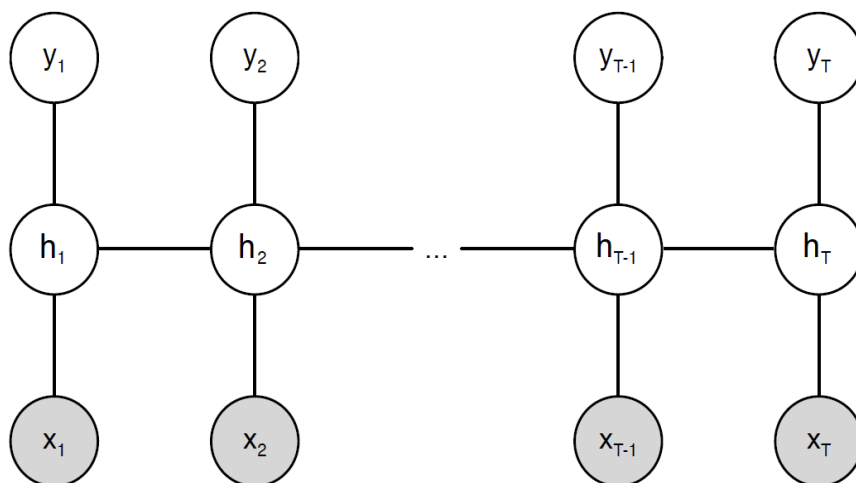
To expand CRF models, we could include a set of hidden variables $\mathbf{h} = \{h_1, h_2, \dots, h_m\}$ not observed during training stage, associated with a singular output label y [85]. Restricting the model to have disjoint sets of hidden states associated with a value y of output label: $h_j \in \mathcal{H}_y$, a Hidden Conditional Random Fields (HCRF) takes the following form:

$$p(y|\mathbf{x}) = \frac{1}{Z} \sum_{h_j, h_{j'} \in \mathcal{H}_y} \exp \left\{ \sum_{t=1}^T \sum_{k=1}^K \theta_k f_k(h_t, h_{t-1}, \mathbf{x}_t, y) \right\} \quad (1-11)$$

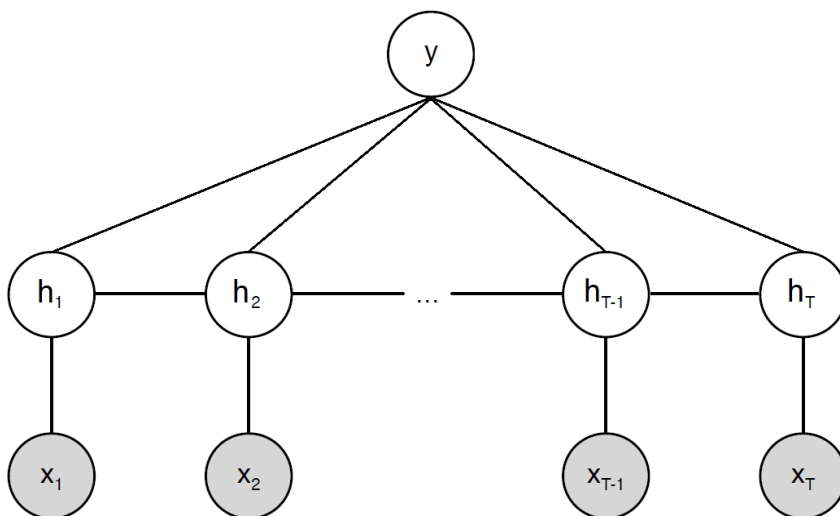
The partition function is defined as the normalization function along all possible values of hidden variables along y . As CRF, a hidden variable $h_{j,t}$ depends only of its predecessor $h_{j',t-1}$ and the corresponding input variables. Figure 1-1 c) illustrates the graphical representation of a linear LDCRF model, where undirected edges represent feature functions that link any $h_{j,t}$ hidden states with their corresponding \mathbf{x}_t inputs and the output y or their previous hidden states $h_{j',t-1}$.



a) Linear Conditional Random Field (CRF)



b) Latent Conditional Random Field (LDCRF)



c) Hidden Conditional Random Field (HCRF)

Figure 1-1. Graphical models from some linear Conditional Random Fields.

Source: Adapted from [86].

1.4 Some performance metrics used in BCI

1.4.1 Accuracy

A recurrent measure of performance for classification is *accuracy*. It estimates the closeness between measured or predicted values and their actual values [87]. A measure commonly used for the accuracy, with M_c classes, is defined as the rate between the trace of a confusion matrix H and the total number of samples N_s [88] as Equation (1-12) shows:

$$p_0 = \frac{\text{trace}(H)}{N_s} \quad (1-12)$$

Where $\text{trace}(H)$ is the number of samples correctly classified. The accuracy varies from 0 to 1, where 1 denotes a perfect classification. Since accuracy closely relates to the binomial distribution definition $\mathcal{B}(N_s, p_0)$ with success probability p_0 and number of trials N_s , p_0 could be approximated to a normal distribution with its standard deviation defined in Equation (1-13):

$$s_e(p_0) = \sqrt{\frac{p_0(1-p_0)}{N_s}} \quad (1-13)$$

However, a high accuracy does not always mean a high performance of a classifier. When the number of members by class is highly unbalanced, the classifier tends to bias toward the class with the highest number of occurrences in the dataset. It is known as the *accuracy paradox* [89].

1.4.2 Cohen's kappa index κ

A typically used estimation of precision is the *Cohen's kappa index* κ [88], [90], [91]. It is an alternative method of measuring the predictive capability of a classifier that associates the accuracy with the probability to classify by chance, as expressed in Equation (1-14)

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (1-14)$$

The numerator is the subtraction between the accuracy and the expected probability to classify correctly by chance (p_e). The denominator is the subtraction between the maximum accuracy and p_e . Consequently, κ defines the rate between the subtraction of the accuracy

and p_e and its maximum value. Possible values for κ come from -1 to 1 [92]. A value of 1 means perfect classification, whereas a value of 0 shows indicates random assignments between actual classes and predicted values. Finally, -1 indicates an opposite relationship between actual and predicted values. The expected probability p_e is defined in the Equation (1-15).

$$p_e = \frac{1}{N_s^2} \sum_{i=1}^{M_c} n_{i:} n_{:i} \quad (1-15)$$

Where the sum of all elements for the i -th row $n_{i:}$ and the sum of all elements for the i -th column $n_{:i}$ are expressed in Equations (1-16) and (1-17):

$$n_{i:} = \sum_{j=1}^{M_c} H_{ij} \quad (1-16)$$

$$n_{:i} = \sum_{j=1}^{M_c} H_{ji} \quad (1-17)$$

The standard deviation of κ is calculated using Equation (1-18):

$$S_e(\kappa) = \frac{\sqrt{p_0 + p_e^2 - \sum_{i=1}^{M_c} n_{i:} n_{:i} (n_{i:} + n_{:i}) / N_s^3}}{1 - p_e \sqrt{N_s}} \quad (1-18)$$

The standard error can be used to create confidence intervals and compute statistical significances when accuracy or kappa values are compared.

1.4.3 Balanced accuracy

The *balanced accuracy* metric is the average of the accuracy obtained by each class, when the symmetry about the class is assumed [93], [94]. If the symmetry assumption is dropped, the *balanced accuracy* is defined as the weighted average of class-specific accuracies, using a cost associated with misclassification by class as weights [94]. In the current work, the symmetry about the class is assumed.

1.5 Instantaneous interactions

1.5.1 Pearson correlation

Pearson correlation, defined for two zero-mean and real-valued random variables x, y , is the coefficient between the cross-correlation of the random variables $E[xy]$ and the product of the square root of their variances $\sigma_x \sigma_y$ [95], [96]:

$$\rho(x, y) = \frac{E[xy]}{\sigma_x \sigma_y} \quad (1-19)$$

1.5.2 Jaccard distance

By the other hand, the Jaccard distance comes from its counterpart, the Jaccard index J . The latter is a similarity measure for two arbitrary finite discrete sets A, B , defined as the coefficient between the size of the intersection $|A \cap B|$ and the size of its union $|A \cup B|$ [97]. It can be extended as the ratio between the measure of the intersection $\mu(A \cap B)$ and its union $\mu(A \cup B)$, with an arbitrary measure μ . If we define μ as the dot product between two sets, expressed as multivariate variables A, B : $\mu(A \cap B) = A \cdot B$, and $\mu(A) = \|A\|^2$, being $\|A\|$ the Euclidean norm of A , and using the relationship between the intersection and the union of two sets, the Jaccard Index J becomes:

$$J(A, B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B} = \frac{A \cdot B}{\|A - B\|^2 + A \cdot B} \quad (1-20)$$

With this definition, Jaccard Index takes values between -1/3 (when $B = -A$) and 1 (for $B = A$). However, for obtaining a non-negative metric, the Jaccard Distance $J_D = 1 - J$ is defined as follows [98]:

$$J_D(A, B) = \frac{\|A - B\|^2}{\|A\|^2 + \|B\|^2 - A \cdot B} = \frac{\|A - B\|^2}{\|A - B\|^2 + A \cdot B} \quad (1-21)$$

With this definition, Jaccard Distance takes values between 0 and 4/3, becoming a non-negative measure.

1.6 Background

The area of BCIs has had a major surge in the last 10 years. As evidence of this, in the last year there have been more than fifty-three thousand papers related to the development of BCI applications, according to the website Google Scholar [99]. A more refined search on specialized search engines gives more than 450 papers related to BCI in PubMed [100] and 850 documents in SCOPUS [101]. The majority of topics about BCI is related to new applications and changes in BCI modalities.

Another indicator of the growing importance of BCIs is the allocation of resources for the study of brain activity. In Europe, the Blue Brain Project [102] has a budget of 1.3 billion dollars, and the Brain Initiative in the United States [103], with 470 million dollars per year during the next 10 years. A recent project, named The BrainCom Project, has a budget of 8.35 million euros for the next five years [104]. In addition, there are two major BCI research projects in Europe: The Berlin Brain – Computer Interface (BBCI), supported by the Ministry of Education and Research of Germany [105], and The BNCI Horizon 2020 Project, supported by more than 10 European institutions, and coordinated by the Graz University of Technology [106], in Austria.

Other BCI projects, laboratories and research groups are:

- a) The Institute of Neural Engineering Graz Brain – Computer Interface Lab, Austria [107]
- b) Penso and Brain – Computer Interface, Australia [108]
- c) The Brain – Computer Interface project, Canada [109]
- d) Brain – Computer Interface Laboratory, Denmark [110]
- e) Tools for Brain Computer Interaction – TOBI, Europe [111]
- f) Cognitive and Social Systems – Centre of Excellence in Computational Complex Systems Research, Finland [112]
- g) Software for Brain – Computer Interfaces and Real – Time Neurosciences – OpenViBE, France [113]
- h) Riken Brain Science Institute, Japan [114]
- i) Brain Computer Interface research at NUST, Pakistan [115]
- j) Laboratory for Neurophysiology and Neuro – Computer Interfaces, Russia [116]
- k) Multimedia Signal Processing Group – MMSPG, Switzerland [117]

- l) TUBITAK BCI – Sabanci University, Turkey [118]
- m) Human Computer Interaction – University of Twente, The Netherlands [119]
- n) BCI Group – University of Essex, United Kingdom [120]
- o) Wadsworth Center – New York Department of Health, United States [121]

1.7 Problem statement

Given the growing importance of the topic, this investigation proposal is framed within the area of the BCIs. In addition, there are few published studies that propose modeling the ERPs with temporal and spatial features to develop or improve a classifier for a brain–computer interface. For this reason, the proposed study will explore the extraction of spatial and temporal features of brain signals during the execution of mental tasks. These features will be used as the input for a probabilistic graphical model whose structure may include prior physiological information about the functioning of the brain. We expect that the model to reflect some aspects of the brain functioning related to ERPs, by including spatial and temporal information. Therefore, we might expect the performance of a BCI system to improve when using the proposed model.

1.8 Justification

The importance of modeling brain signals and applying the proposed models on BCI may be observed from different activities. The first one is the provision of a non-muscular communication path with the environment for patients suffering from neuro-degenerative and muscular diseases such as amyotrophic lateral sclerosis (ALS). The development of BCIs may be extended to other activities related to rehabilitation, as in the case of development of devices for people with loss of mobility in a part of the body.

Another activity with high potential for application of brain-computer interfaces is in the development of elements handling equipment for hazardous environments, both civilian and military. The purpose is to reduce the response time in the remote control equipment, and increase the precision of the movements. These developments may be extended to designing new devices whose purpose is to enhance the capacity for humans to interact with electrical and electronic devices, both at home and industry. Other activities where it is possible to implement a BCI are the restoration of the senses of sight and hearing, the establishment of

new connections between brain regions, object recognition in surveillance footage, restoration and enhancement of memory, development of interactive games based in reality, among others.

Finally, the modeling of temporal and spatial features of the EEG signals can give insight in the functioning of neural circuits within the brain, leading to a better understanding of specific brain functions.

1.9 Hypotheses

This thesis proposal has two major hypotheses:

1.9.1 First Hypothesis

The occurrence of ERP can be predicted more accurately if the spatial and temporal relationships between the EEG signals are taken into account simultaneously. Given that the brain is highly interconnected and the processing of different aspect of the sensory information is achieved in different areas, we expect that the inclusion of spatial modeling will provide a better model for the generation of the ERPs.

1.9.2 Second Hypothesis

The ERPs are likely to show a probabilistic distribution of the times when the brain responds to an event. This approach is considerably different from what the BCI community uses. However, experiments in animals suggest this hypothesis [122]. We suggest that the averaged ERP could be used as prior information in the detection of ERP components from single trials.

2 A BOOTSTRAPPING METHOD FOR IMPROVING A P300 SPELLER CLASSIFICATION PERFORMANCE

2.1 Preliminaries

A Brain–Computer Interface (BCI) is a communication and control system that sends messages and commands to a device without the normal use of nerves and muscles [123]. An usual BCI system is comprised of a brain monitoring system, a signal pre-processing phase, a phase for extracting features of the pre-processed signal, and a decision phase where features are translated into commands or messages [2]. Some modalities of BCI activation are by evoked potentials, brain rhythms and motor imagery, among others [19], [21]. Initially focused on EEG signals [63], [124], it was moved quickly to ECoG signals [64], [125] because of the advantages provided by the latter signals.

A remarkable application for Brain–Computer Interfaces is the *P300 speller*, proposed initially by Farwell and Donchin [126], revised and improved in many other studies [127]–[130]. A basic speller consists of a character set, distributed and displayed in rows and columns on a screen. Rather than unveiling one character, the speller randomly highlights a row or column of characters. When the user watches the chosen character in a highlighted row or column, the brain generates a *P300* signal related to memory and attention processes in the brain [131].

A P300 speller takes the brain activity by electroencephalography (EEG) and attempts to distinguish between P300 and non-P300 signals. When the speller detects a P300 signal in a specific row and column, it takes the corresponding character and displays it. The described speller is one of the bases for developing online BCI applications [130], [132]–[134]. It means that the goal of the classification task is not to recognize P300 signals but identify the row and column of a specific character from the P300 recognition.

In literature, there are mainly three approaches to process input features to feed a classifier of a P300 speller. The first one consists of training and evaluating the classifier from single trials [128], [133], [135]. The second strategy makes use of averaged data over a fixed number of trials, for training and testing the system [55], [130], [132], [134], [136]–[139]. The third approach consists of training the classifier in single trials, and evaluating the

classifier with averaged ones [45], [53], [140], [141]. The last method (labeled as the *traditional approach* in the present work) is commonly used in the literature. However, the statistical properties of signals during the training phase are different from those signals used during the testing phase. It violates the assumption of any classification problem that training and testing data should come from the same population [39]. Consequently, the classifier has reduced capacity to differentiate between P300 and non-P300 trials.

Different statistical properties of training and testing features have another problem. Estimation of the posterior probabilities from probabilistic classification would not be correct. The latter issue is crucial for P300 applications that use language models [40], [73], [74], [132] since the posterior probability of the classifier output is usually combined with the probability of the presence of a symbol in a particular language to determine the most likely sequence of characters.

In addition, as the P300 speller works based on the *oddball paradigm*, where a P300 signal is a deviant stimulus, the number of events is unbalanced since the number of non-P300 trials is larger than the number of P300 trials [142]. Both unbalanced classes and small datasets could affect the performance of a classifier [143], [144]. So, the number of samples by class should be balanced to get a more confident performance. Some researchers have proposed removing samples randomly from the class with more members to reach a 1:1 proportion [53], [54], [138], [139], [141], trying to preserve as many samples as possible in the training phase [145]. It solves the problem of unbalanced classes at the expense of reducing the number of training samples. In the same way, since P300 and non-P300 classes are unbalanced, performance measures as accuracy are inclined to biasing. It occurs because the classifier assigns most samples to the class with higher prior probability [144]. Some studies have proposed the Cohen's kappa index κ as an alternative measure of performance that does not have the matters previously described [79], [88], [146].

This work presents a method for training linear classifiers to identify the presence of P300 potentials. We propose a bootstrapping approach for training and testing linear classifiers from averaged samples. Here, we show that the traditional approach could lead to misinterpretation of the actual performance of these classifiers. It happens because the

performance metric based on accuracy is not well suited for the cases of unbalanced classes. Results show a significant improvement using the proposed method for detecting P300 potentials.

2.2 Methods

2.2.1 Experiment and dataset description

The experiment comprised of stating one of 36 possible characters (26 letters and ten digits). A subject observed a six-by-six character array on a screen, focusing on the prescribed character above the array speller. The matrix was unveiled for a 2.5 s period, where all characters had the same intensity. Afterward, each column and row were randomly intensified for 100 ms, followed by a blank period of 75 ms after each intensification. Hence, there were 12 different row/column stimuli by round and 15 rounds of intensifications by character, for a total of 31.5 s. Each subject spelled 32 characters in total. Fourteen healthy subjects participated in the study.

The dataset contains EEG signals recorded by a cap embedded with 64 electrodes, according to the modified 10–20 system [147]. All electrodes were referenced to the right earlobe and grounded to the right mastoid. Raw EEG signals were band-pass filtered between 0.1 and 60 Hz and amplified with a 20000X SA Electronics amplifier [45]. Each experiment took into account only 16 EEG channels, motivated by the study presented by Krusienski et al. [45]: F3, Fz, F4, FCz, C3, Cz, C4, CPz, P3, Pz, P4, PO7, PO8, O1, O2, and Oz, as Figure 2-1. Each channel was sampled at 240 Hz for one subject and 256 Hz for the others. All aspects of data collection and experimental control were controlled by the BCI2000 system [137]. Each subject has two datasets: one is used for training, and the other one is implemented for testing. Both datasets were taken on different days and were obtained from the Wadsworth Center, NYS Department of Health [45]. Data for two subjects are available at <https://www.bbc.de/competition/iii/> [148].

2.2.2 Data processing

We band-pass filtered, separated in trials, and decimated the data. Then, a single vector concatenated all of the input features. We took data either directly as the training input of a

classifier or as the new population for obtaining averaged samples. The following subsections give additional details.

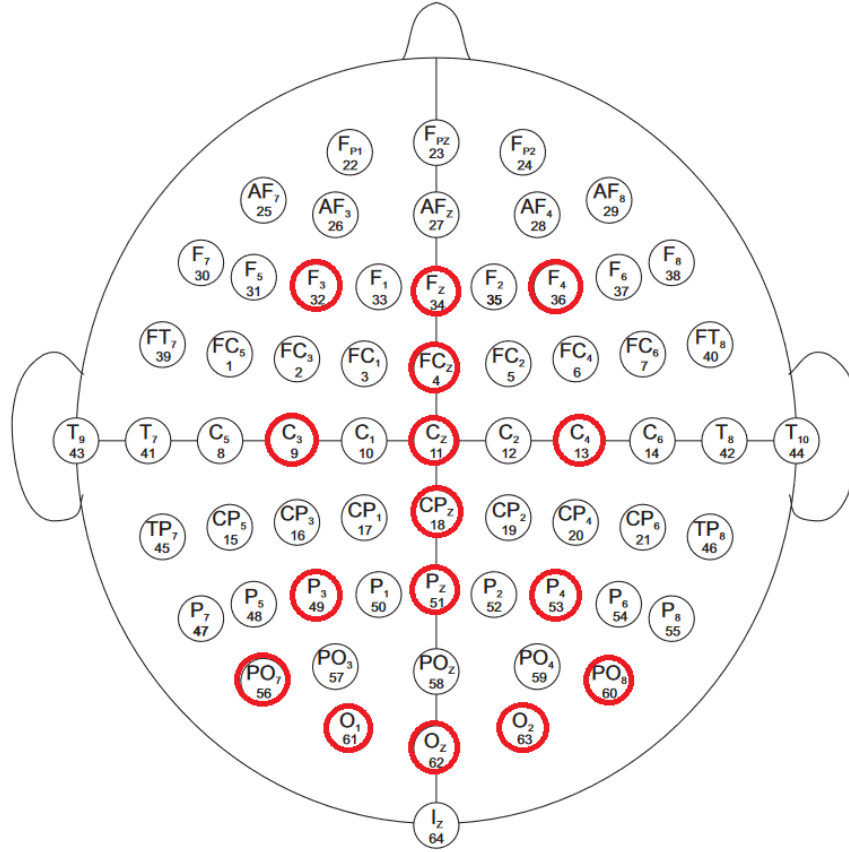


Figure 2-1. Place of used electrodes in the dataset.

Source: adapted from <https://archive.physionet.org/physiobank/database/eegmmidb/?C=M;O=A>

2.2.2.1 Pre-processing

First, we band-pass filtered the raw data between 1 and 20 Hz using a 4th fourth-order Butterworth filter. The chosen bandwidth eliminates the trend of each channel and prevents aliasing, allowing the posterior decimation of the signal. Afterward, a 600 ms window separated the data in trials after the presentation of each visual stimulus (the highlighted row or column), as proposed in a previous work [73].

Signals from all channels were decimated by a factor of 4 and concatenated in a single feature vector. We chose it because the maximum analog frequency of the EEG signal is 60 Hz [45]. In addition, frequencies higher than the beta band reflect unrelated neural processes to P300 in attention [149]. For the averaged process, signal segments were

averaged across repetitions, up to the maximum repetitions by character (15 repetitions). Then, concatenated channels were used as the inputs of the classifier, as described in [45].

2.2.2.2 *Re-sampling of training data*

In the traditional approach, the classifier is trained with single trials and tested on averaged trials to increase the *signal-to-noise ratio*. Note that besides the issue of having unbalanced data, the statistical properties of the training data do not match those of the testing data. To avoid these problems, we implement an approach based on *bootstrap* resampling (*bootstrapping*) [44], [150].

Here, a new dataset is obtained by re-sampling N trials with replacement, where N is the number of trials used to get a new averaged sample. The process is repeated M times by class to get M averaged samples for each class. The new dataset helps training a classifier such that 1) the number of samples is equal for each class in the training set, and 2) the statistical properties of training and testing data remain identical. It is worth mentioning that N may not be defined a priori in practical scenarios. However, the procedure can be performed for any value of N . Additionally, it does not imply any additional significant computational load, as the re-sampling is computationally inexpensive.

In this work, we used training dataset as a new population to implement the resampling. We changed the number of recurrences (single trials) used to get a new averaged trial, with $N = \{2, 3, \dots, 14, 15\}$, because 15 is the maximum number of repetitions available by character. Then, we replicated the process M times by class. Single trials were not used because resampling only allows acquiring repeated samples, reducing the variability of the training samples. We tested a classifier trained with one of the following types of samples: unbalanced classes with single trials (as in the traditional approach) and balanced classes by resampling $M = \{1000, 2000, 3000\}$ averaged trials by class. The value of M is chosen according to the statistical importance obtained in the results. Averaged trials were used as testing data.

2.2.3 Classifiers

In the speller, the classification problem implicates identifying the row and the column corresponding to a character. In the present study, the target of the classification is to determine whether a signal is P300 or not. We implemented four classifiers in the study: stepwise linear discriminant analysis (SWLDA), Bayesian linear discriminant analysis (BLDA), linear support-vector machine (LSVM), and logistic regression (Log Reg). While based-LDA-based algorithms are generative classifiers, LSVM and Log Reg lie in the category of linear discriminative classifiers [39], [151]. Results come from test data, unknown by the implemented classifiers during the training process.

For discriminative classifiers, it is crucial to choose the value of a regularization factor C . The training process implements four-fold cross-validation to get the best value of C . The number of values tested for C was 25, all located between 0.01 and 1. After the process, the final classifier is trained using the whole training dataset and the chosen value of C . The process is repeated by each user and each type of training sample [151].

2.2.4 Performance metrics

All performance measures described below are used rather than precision, recall, or area under the curve. It is due to the unbalanced nature in the data, even if the problem is a two-class classification task. All measures were obtained by five-fold cross-validation, being training data used to tune the model parameters.

Briefly, accuracy is the rate between the number of correct predictions over the total number of predictions [68], [135], [93]. Cohen's kappa index relates the accuracy with the classification by chance probability as a rate between their subtraction by its maximum value [87], [135], [152].

2.2.5 Statistical analysis

The statistical significance of differences among the number of bootstrapped samples for averaged training data was tested by a one-way randomized blocks ANOVA by performance index and classifier. ANOVA was chosen rather than a Student's t-test because the latter does not consider the random effects of the number of averages and subjects, whereas ANOVA does. The number of training data (M) was the design variable,

and each performance index was the output variable. Subjects and number of averaged samples by a testing trial were taken as randomized blocks.

The study also tested the statistical significance of differences between the previously described types of training data by two procedures. First, we performed a one-way randomized blocks ANOVA by metric and classifier. Type of training data (traditional approach or proposed) was the design variable. Also, we performed a paired Student's t-test for each subject, by assembling the metrics and their corresponding standard deviations. Their purpose was to estimate the significance of the differences between the traditional and the proposed method for training data.

2.3 Results and discussion

Results refer to the average performance obtained by each classifier, measured by the accuracy and Cohen's kappa index metrics. First, we tried to find an optimal value for the number of bootstrapped samples M . Then, we implemented the chosen value of M to compare the performance of classifiers with bootstrapped samples and the traditional approach. In all cases, we changed the number of trials to get an averaged trial $N = \{2, 3, \dots, 14, 15\}$. All metrics were obtained from testing data.

2.3.1 Number of bootstrapped samples

The statistical significance of differences among the number of bootstrapped samples for averaged training data was tested by a one-way randomized blocks ANOVA by each performance index and classifier. The numbers of samples used were $M = \{1,000, 2,000, 3,000\}$.

For Log Reg, the ANOVA test does not reveal any significant statistical differences among the number of bootstrapped samples for neither accuracy nor Cohen's kappa index, indicated by a p -value higher than the significance level greater than 0.05 (accuracy: $F = 1.72, p = 0.18$; Cohen's kappa index: $F = 0.51, p = 0.60$). In the case of LSVM, there is no statistical difference in the number of samples, indicated by a p -value higher than the significance level greater than $\alpha = 0.05$ (accuracy: $F = 0.92, p = 0.40$; Cohen's kappa index: $F = 0.02, p = 0.98$).

Similar conclusions are obtained by analyzing the results of ANOVA tests for SWLDA (accuracy: $F = 0.61$, $p = 0.55$; Cohen's kappa index: $F = 0.13$, $p = 0.88$) and BLDA (accuracy: $F = 0.67$, $p = 0.51$; Cohen's kappa index: $F = 0.39$, $p = 0.25$), by analyzing p-values, greater than the significance level $\alpha = 0.05$. Although there is no significant difference, 2,000 averaged bootstrapped samples allowed to get the highest performance values. Hence, the chosen number of M is 2,000 in the remaining sections of the chapter.

2.3.2 Type of training samples

2.3.2.1 Linear SVM

Figure 2-2 shows the average performance obtained to use the LSVM classifier by subject and training data type. The ANOVA test gives significant differences for both metrics (accuracy: $F = 216.92$, $p < 0.01$; Cohen's kappa index: $F = 1380.10$, $p < 0.01$). According to the results, when we trained the classifier with 2,000 averaged samples by class, classifier performance was significantly higher than training with the traditional approach with a significance level of 5%.

Table 2-1 discriminates the average of the results and pooled standard deviations obtained by each subject for accuracy and kappa. We performed a Student's t-test to yield the statistical significance of the difference between the methods. Results indicate that the difference is highly significant ($p < 0.01$) for most metrics and subjects for the proposed method.

2.3.2.2 Logistic regression

Figure 2-3 shows the average performance reached by employing the Logistic Regression classifier on each subject and type of training data. The ANOVA test gives significant differences for both metrics (accuracy: $F = 215.10$, $p < 0.01$; Cohen's kappa index: $F = 843.29$, $p < 0.01$). Results reveal that the classifier performance is higher with the proposed method for training.

Table 2-2 contrasts the averaged accuracy and kappa results and their standard deviations by each subject. The student's t-tests indicate that the statistical difference is highly significant ($p < 0.01$) for most subjects in the proposed method. The consistency

in the statistical analyses for Log Reg supports the improvement in the results with our approach. It also applies to the LSVM results.

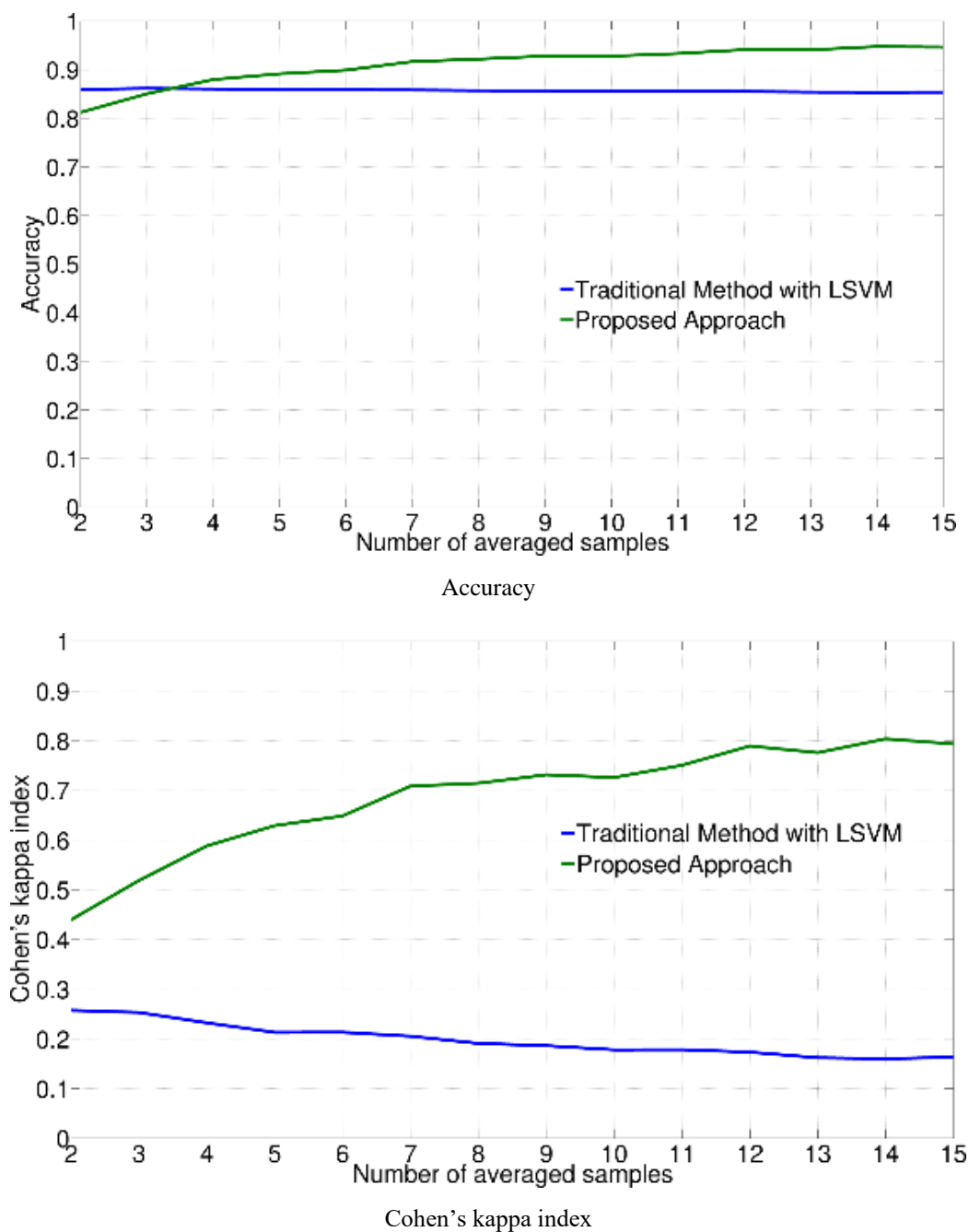


Figure 2-2. Averaged results of all subjects, for LSVM

Table 2-1. Averaged metrics by subject, for LSVM.

Subject	Traditional Approach		Proposed Method	
	Accuracy	kappa	Accuracy	kappa
1	0.95 ± 0.11	0.78 ± 0.33	$0.98 \pm 0.08^*$	$0.93 \pm 0.33^*$
2	0.88 ± 0.12	0.43 ± 0.27	$0.95 \pm 0.10^*$	$0.83 \pm 0.33^*$
3	0.84 ± 0.13	0.01 ± 0.07	$0.90 \pm 0.12^*$	$0.55 \pm 0.33^*$
4	0.84 ± 0.13	0.06 ± 0.18	$0.95 \pm 0.10^*$	$0.81 \pm 0.30^*$
5	0.88 ± 0.12	0.36 ± 0.26	$0.97 \pm 0.08^*$	$0.90 \pm 0.30^*$
6	0.83 ± 0.13	0.00 ± 0.00	0.84 ± 0.13	$0.52 \pm 0.27^*$
7	0.83 ± 0.13	0.00 ± 0.00	$0.87 \pm 0.12^*$	$0.48 \pm 0.27^*$
8	0.83 ± 0.13	0.00 ± 0.00	$0.87 \pm 0.12^*$	$0.59 \pm 0.28^*$
9	0.89 ± 0.12	0.43 ± 0.27	$0.94 \pm 0.10^*$	$0.81 \pm 0.29^*$
10	0.84 ± 0.12	0.06 ± 0.15	$0.94 \pm 0.10^*$	$0.79 \pm 0.30^*$
11	0.85 ± 0.13	0.18 ± 0.22	$0.94 \pm 0.10^*$	$0.80 \pm 0.29^*$
12	0.83 ± 0.13	0.00 ± 0.00	$0.90 \pm 0.12^*$	$0.53 \pm 0.28^*$
13	0.88 ± 0.12	0.44 ± 0.27	0.81 ± 0.13	$0.50 \pm 0.26^*$
14	0.83 ± 0.13	0.02 ± 0.27	$0.90 \pm 0.11^*$	$0.59 \pm 0.29^*$
Average	0.86 ± 0.12	0.20 ± 0.19	0.91 ± 0.11	0.69 ± 0.29

* The difference is highly significant, with a Student's t-test ($p < 0.01$). Number of samples: 372 for subject 1, 504 for the rest.

2.3.2.3 Stepwise and Bayesian LDA

Figure 2-4 displays the average performance obtained by using the SWLDA classifier on each subject and type of training data. The ANOVA test gives significant differences for both metrics (accuracy: $F = 28.15$, $p < 0.01$; Cohen's kappa index: $F = 20.86$, $p < 0.01$). When the Student's t-test contrasts the metrics, there is no statistical significance in most cases, as presented in the Table 2-3.

Identical results were obtained for BLDA, as illustrated in Figure 2-5. Although the ANOVA test gives significant differences between the traditional and the proposed methods (accuracy: $F = 13.21$, $p < 0.01$; Cohen's kappa index: $F = 6.21$, $p = 0.013$), the Student's t-tests do not reject the null hypothesis of equality of metrics, as shown in Table 2-4.

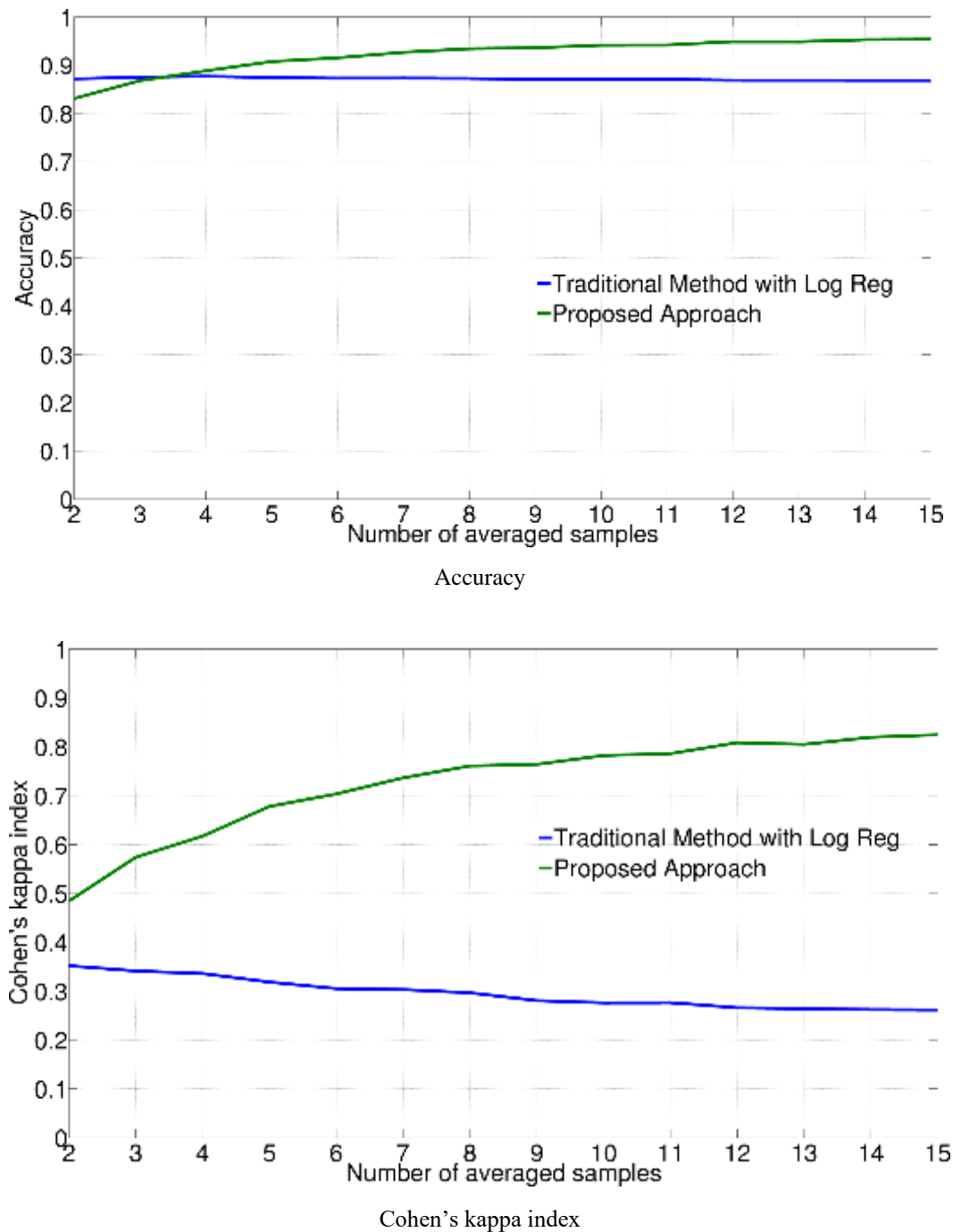


Figure 2-3. Averaged results of all subjects, for Log Reg.

Table 2-2. Averaged metrics by subject, for Log Reg.

Subject	Traditional Approach		Proposed Method	
	Accuracy	kappa	Accuracy	kappa
1	0.97 ± 0.09	0.88 ± 0.33	0.98 ± 0.09	0.92 ± 0.33
2	0.92 ± 0.11	0.65 ± 0.29	$0.96 \pm 0.09^*$	$0.86 \pm 0.30^*$
3	0.84 ± 0.13	0.03 ± 0.14	$0.93 \pm 0.11^*$	$0.68 \pm 0.29^*$
4	0.85 ± 0.13	0.15 ± 0.22	$0.95 \pm 0.10^*$	$0.84 \pm 0.30^*$
5	0.91 ± 0.11	0.56 ± 0.29	$0.97 \pm 0.08^*$	$0.91 \pm 0.30^*$
6	0.84 ± 0.13	0.05 ± 0.13	$0.86 \pm 0.12^*$	$0.59 \pm 0.27^*$
7	0.83 ± 0.13	0.00 ± 0.07	$0.88 \pm 0.12^*$	$0.51 \pm 0.28^*$
8	0.83 ± 0.13	0.00 ± 0.07	$0.88 \pm 0.12^*$	$0.62 \pm 0.28^*$
9	0.92 ± 0.11	0.64 ± 0.29	$0.95 \pm 0.09^*$	$0.84 \pm 0.29^*$
10	0.85 ± 0.13	0.15 ± 0.21	$0.95 \pm 0.09^*$	$0.83 \pm 0.30^*$
11	0.88 ± 0.12	0.41 ± 0.27	$0.95 \pm 0.09^*$	$0.82 \pm 0.30^*$
12	0.83 ± 0.13	0.00 ± 0.04	$0.90 \pm 0.11^*$	$0.55 \pm 0.29^*$
13	0.91 ± 0.11	0.60 ± 0.29	0.84 ± 0.13	0.57 ± 0.26
14	0.84 ± 0.13	0.03 ± 0.13	$0.91 \pm 0.11^*$	$0.61 \pm 0.29^*$
Average	0.87 ± 0.12	0.30 ± 0.22	0.92 ± 0.11	0.72 ± 0.29

* The difference is highly significant, with a Student's t-test ($p < 0.01$). Number of samples: 372 for subject 1, 504 for the rest.

2.3.3 Discussion

According to Figure 2-2 and Figure 2-3, results of LSVM and logistic regression are identical. When we trained the classifiers with the traditional approach, Cohen's kappa index reduced as the number of averaged samples increased in testing samples. The observed kappa reduction is due to the probability of classifying by chance p_e . The value of p_e increases as the increment in the number of averaged samples does. Meanwhile, the accuracy only had non-significant changes when the number of averaged samples by trial increased. Consequently, p_e was closer to the accuracy as the number of trials to get the averaged samples increased, so kappa decreased.

When we trained the classifiers with the proposed method, kappa improved significantly. It indicates that the classifiers learn features from P300 and non-P300 classes because the accuracy gets higher values than p_e when the number of averaged

samples by trial also increases. Consequently, accuracy, and thus kappa, will be higher, as seen in Figure 2-2 and Figure 2-3.

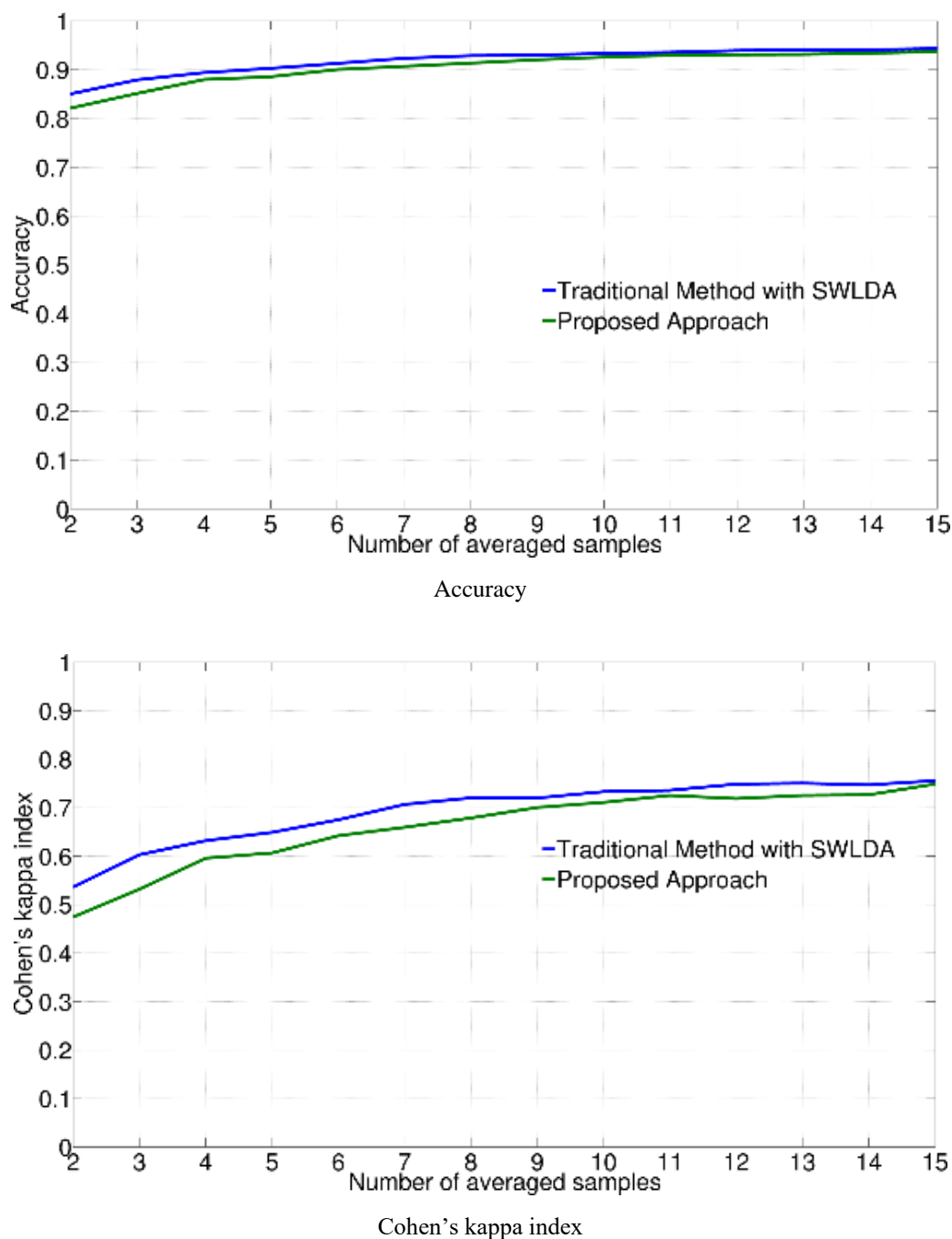


Figure 2-4. Averaged results of all subjects, for SWLDA

Table 2-3. Averaged metrics by subject, for SWLDA.

Subject	Traditional Approach		Proposed Method	
	Accuracy	kappa	Accuracy	kappa
1	0.98 ± 0.08	0.95 ± 0.33	0.98 ± 0.08	0.94 ± 0.33
2	0.96 ± 0.09	0.88 ± 0.30	0.95 ± 0.10	0.84 ± 0.29
3	0.86 ± 0.12	0.26 ± 0.25	0.87 ± 0.12	$0.36 \pm 0.26^*$
4	0.95 ± 0.10	0.83 ± 0.30	0.94 ± 0.10	0.79 ± 0.30
5	0.98 ± 0.07	0.94 ± 0.31	0.98 ± 0.08	0.92 ± 0.30
6	0.88 ± 0.12	0.63 ± 0.28	0.84 ± 0.13	0.53 ± 0.27
7	0.89 ± 0.12	0.54 ± 0.28	0.86 ± 0.12	0.48 ± 0.27
8	0.90 ± 0.11	0.68 ± 0.28	0.87 ± 0.12	0.60 ± 0.28
9	0.94 ± 0.10	0.82 ± 0.29	0.93 ± 0.10	0.80 ± 0.29
10	0.94 ± 0.10	0.78 ± 0.30	0.94 ± 0.10	0.78 ± 0.30
11	0.95 ± 0.10	0.82 ± 0.30	0.94 ± 0.10	0.79 ± 0.29
12	0.90 ± 0.12	0.50 ± 0.28	0.89 ± 0.12	0.48 ± 0.28
13	0.83 ± 0.13	0.54 ± 0.26	0.80 ± 0.13	0.48 ± 0.26
14	0.90 ± 0.12	0.57 ± 0.29	0.87 ± 0.12	0.47 ± 0.28
Average	0.92 ± 0.11	0.69 ± 0.29	0.90 ± 0.11	0.66 ± 0.29

* The difference is highly significant, with a Student's t-test ($p < 0.01$). Number of samples: 372 for subject 1, 504 for the rest.

The reasons are balanced data, similar statistical properties for training and test samples, and more statistical divergence by class due to the boost in sample size. It is necessary to remark that the trouble of using the traditional approach for training discriminative classifiers is due to the difference in statistical properties of training data and testing data. Again, the benefit of training with re-sampled and averaged samples is statistically significant.

Although stepwise and Bayesian LDA results were identical, they differ from the discriminative classifiers. ANOVA tests give significant differences in the methods, while the Student's t-test does not reject the statistical parity of the results, as presented Table 2-3 and Table 2-4. The disparity in the statistics is due to the origin of the standard deviation in each test. The Student's test employs a weighted pooling of the variances, whereas ANOVA uses the *mean square* of the error from data. For example, standard deviations lie between 0.07 and 0.13 for accuracy and between 0.25 and 0.33 for kappa.

Meanwhile, mean squares of error lie around 0.0006 for accuracy and 0.006 for kappa. It means that the differences of magnitude between standard deviations are around 183 for accuracy and 48 for Cohen's kappa index. So, with the same averaged metrics, both types of tests give different results: while ANOVA catches a gap between the groups of the design variable, the Student's t-test gives small values of the statistics.

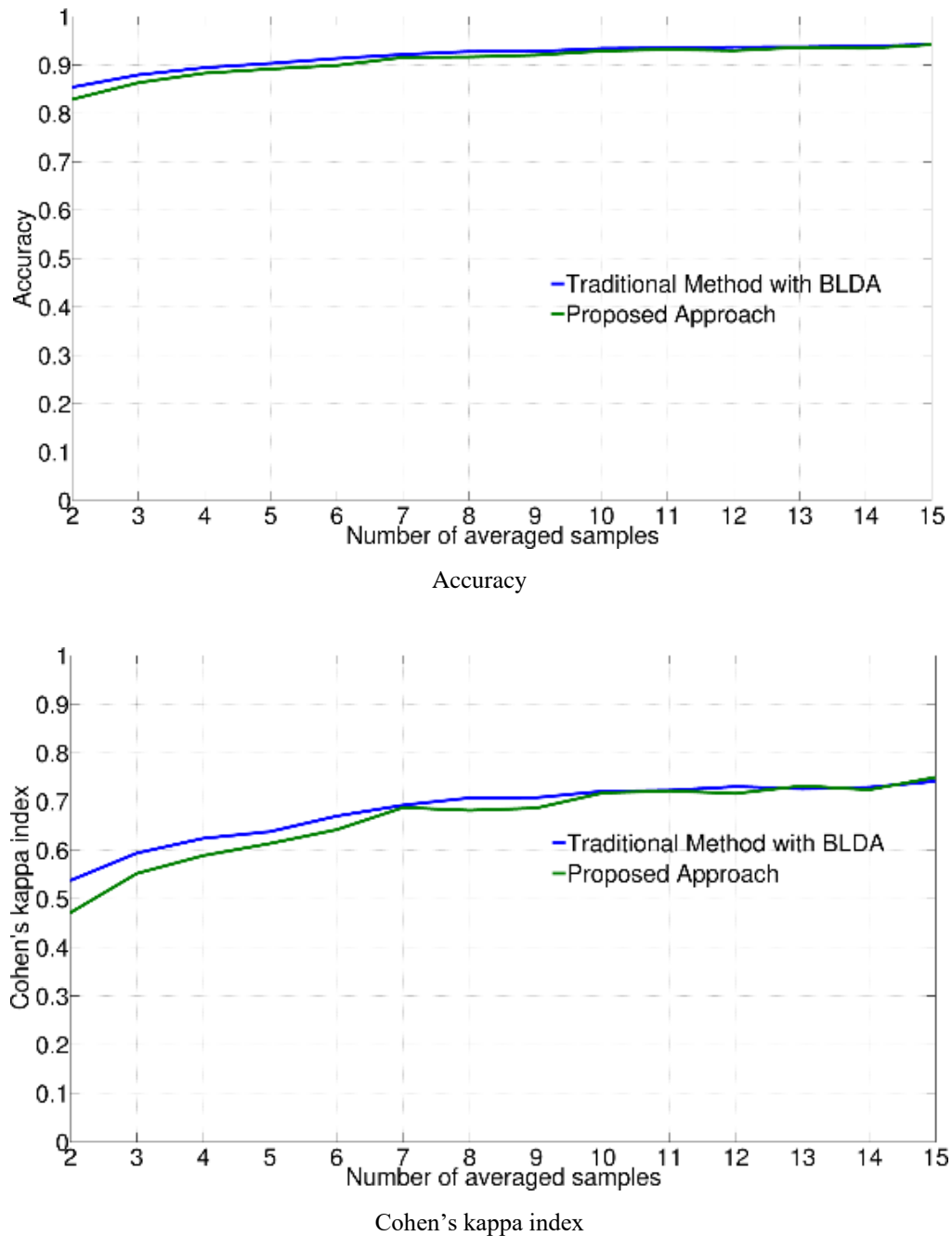


Figure 2-5. Averaged results of all subjects, for BLDA

Table 2-4. Averaged metrics by subject, for BLDA.

Subject	Traditional Approach		Proposed Method	
	Accuracy	kappa	Accuracy	kappa
1	0.99 ± 0.08	0.95 ± 0.33	0.98 ± 0.08	0.94 ± 0.33
2	0.96 ± 0.09	0.87 ± 0.30	0.95 ± 0.09	0.85 ± 0.30
3	0.86 ± 0.12	0.30 ± 0.25	0.86 ± 0.12	0.33 ± 0.26
4	0.95 ± 0.10	0.82 ± 0.30	0.95 ± 0.10	0.79 ± 0.30
5	0.98 ± 0.07	0.94 ± 0.31	0.98 ± 0.10	0.92 ± 0.30
6	0.86 ± 0.12	0.58 ± 0.27	0.84 ± 0.13	0.53 ± 0.27
7	0.88 ± 0.12	0.48 ± 0.28	0.87 ± 0.12	0.49 ± 0.28
8	0.90 ± 0.11	0.67 ± 0.28	0.88 ± 0.12	0.62 ± 0.28
9	0.95 ± 0.09	0.86 ± 0.30	0.94 ± 0.10	0.82 ± 0.29
10	0.95 ± 0.10	0.79 ± 0.30	0.94 ± 0.10	0.78 ± 0.30
11	0.95 ± 0.09	0.83 ± 0.30	0.94 ± 0.10	0.80 ± 0.29
12	0.89 ± 0.12	0.43 ± 0.27	0.89 ± 0.12	0.45 ± 0.27
13	0.85 ± 0.12	0.58 ± 0.27	0.82 ± 0.13	0.52 ± 0.26
14	0.89 ± 0.12	0.46 ± 0.28	0.88 ± 0.12	0.46 ± 0.28
Average	0.92 ± 0.29	0.68 ± 0.11	0.91 ± 0.29	0.66 ± 0.11

* The difference is highly significant, with a Student's t-test ($p < 0.01$). Number of samples: 372 for subject 1, 504 for the rest.

Another issue worth regarding is the nature of the LDA-based classifiers. They try to fit data to a set of Gaussian models, with a mean by class and a common covariance matrix [39]. When new data get in the classifier, the classifier compares new data with each model. Subsequently, a class is designated when the highest score or probability value is obtained from the corresponding model. This score or probability comes from the distance between the data and each model. In our study, both classifiers map data to a score value, according to a regression model before the Gaussian models. It means that the models are also scalar rather than multivariate, unlike discriminative classifiers, where the mapping to the class is direct [39]. Consequently, discriminative models are more affected by the statistical nature of the data, as the difference between generative and discriminative classifiers results reflects.

2.4 Conclusions

The study presents a bootstrapping method to solve some issues in the P300 speller. It generates a new training set by re-sampling with replacement from the original set, reaching two key goals.

First, the number of trials across classes is balanced. It avoids ignoring data, as suggested in other methods [53], [54], [138], [139], [141], while preventing a possible bias in the classification results. Second, the statistical properties of training data and test sets are equivalent. It is achieved when the number of averaged trials for each instance in training equals the number of averaged samples during testing.

Unbalanced classes and the difference in statistical properties are subjects present in the cutting-edge implementations of the P300 classification task. Results demonstrate that the proposed method improves the classification of P300 and non-P300 brain activity in linear discriminative classifiers significantly by dealing with the above issues.

3 A FIRST APPROXIMATION TO LINEAR CRF CLASSIFIERS FOR FINGER MOVEMENT CLASSIFICATION

3.1 Preliminaries

A Brain-Computer Interface (BCI) is a communication and control scheme that sends messages and commands to the external world without using nerves and muscles in the usual way [123]. A regular BCI system has a brain monitoring system, a signal pre-processing stage, a stage for extracting features of the pre-processed signal, and a classification stage where features are decoded into commands or messages. Some manners of BCI activation are evoked potentials, brain rhythms, and motor imagery, among others [19], [21]. Initially developed for electroencephalography (EEG) [63], [124], it was driven quickly to electrocorticography (ECoG) [64], [125] because of the advantages provided by the latter signals.

For classification, different methods have been proposed: Linear Discriminant Analysis (LDA) [153]–[155], Support Vector Machines (SVM) with different kernels [64]–[66], Hidden Markov Models (HMM) [65], [156], Template Matching [154], Logistic Regression (LR), Simultaneous Sparse Approximation and Conditional Random Fields (CRF) with long-range dependencies [157]. Only SVM, LR and CRF are discriminative classifiers, where the classifier maps data directly to classes [39]. The remaining classifiers are generative, where they learn the probabilistic models of data given the classes [158].

We present two based-CRF discriminative classifiers as methods to develop the task: a CRF classifier and a Latent Dynamic CRF (LDCRF) model. The proposed classifiers have the advantage of taking into account temporal dependencies without using generative models [69]. In both cases, the study employs down-sampled data to 20 Hz while other studies employed higher down-sampling rates [157]. Also, we use an additional measure of precision named *Cohen's kappa index* to discard any bias in the classifiers. We used three classifiers to compare the performance of the approach: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and a Support Vector Machine

with the linear kernel (LSVM). Results show that the performance of CRF-based models surpasses the other classifiers significantly.

3.2 Materials and methods

3.2.1 Experiment and dataset description

Dataset was recorded from nine patients. They underwent surgery for temporary placement of subdural electrodes due to intractable epilepsy. These data appeared initially in [159], has been collected by Dr. K. J. Miller *et al.* and are available online at <https://exhibits.stanford.edu/data/catalog/zk881ps0522> [160]. All of the patients participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12193). All patient data were anonymized according to the IRB protocol, under the HIPAA mandate.

The experiment task consisted in flexing a finger that a bedside monitor displays: thumb, index, middle, ring, or pinkie. Each subject moved the indicated finger for 2 seconds. The subject performed self-paced finger movements as a response, moving 2 – 5 times typically by trial. A 2 second rest period, indicated by a blank screen, pursued each trial. It makes the rest period an additional class. Each finger had 30 finger movement trials, given randomly.

A Neuroscan Synamps2 device performed ECoG recordings with a 1000 Hz sampling rate. ECoG 8x8 array electrodes were surgically placed on the sub-dural brain surface during the therapy for epilepsy. In addition to ECoG, a 5-degree-of-freedom data-glove sensor recorded the finger positions. Data-glove signals were initially sampled at 25 Hz and up-sampled at 1000 Hz to match the sampling rate of ECoG signals. We included only electrodes placed in the sensory and motor areas, where Figure 3-1 shows [161].

3.2.2 Data pre-processing

We previously detrended data and used a common average reference to reduce the effect of the reference electrode in all the recording electrodes. We standardized data with a mean of 0 and variance one before additional pre-processing. Two types of features are used in the study: High gamma (HG) band and low-frequency components (LFC) [161].

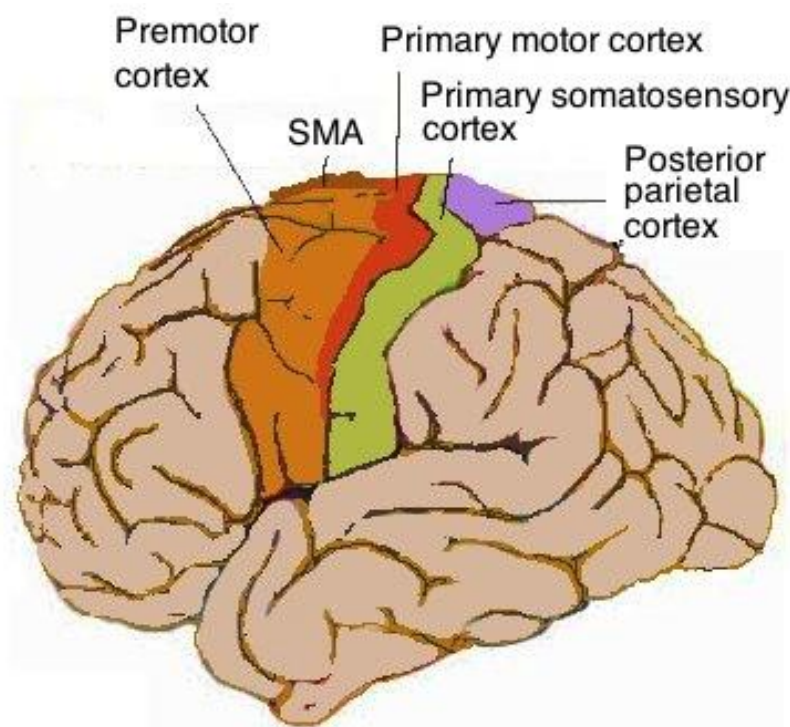


Figure 3-1. Location of primary sensorimotor cortex, indicated in red and green.

Source: https://commons.wikimedia.org/wiki/File:Human_motor_cortex.jpg

We filtered data between 70 and 170 Hz and passed them by a Hilbert Transform to generate HG features. The envelope is then low-pass filtered to 8 Hz using an 8th order Butterworth filter.

For LFC features, data were filtered between 0.5 and 8Hz using a 4th order Butterworth filter. The chosen bandwidth eliminates the trends of each channel and selects the frequencies of interest. Afterward, finger movements and ECoG data were down-sampled from 1 kHz to 20 Hz.

Both types of data will be used in the PGM models to analyze their performance, either HG, LFC, or both. We will use the features with the best performance to compare PGM models with other classifiers.

3.2.3 Classifiers to be compared

CRF and LDCRF are going to be compared with three classifiers: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Linear-Support Vector Machine (LSVM). Briefly, a discriminant classifier is a function that allocates an input

vector \mathbf{x} to one of K classes \mathcal{C}_k [39]. If each class has its covariance matrix Σ_k , the discriminant function is quadratic. However, if we suppose a shared covariance matrix Σ for all classes, the discriminant function becomes linear. On the other hand, the support vector machine (SVM) classifier builds its decision boundary by maximizing the perpendicular distance between the nearest training point and the boundary decision [39]. In the current study, we only take linear SVM into account.

3.2.4 Performance metrics

All performance measures described below are used rather than precision, recall, or area under the curve. It is due to the multiclass nature and the predominance of *rest* class in the data, making this an unbalanced and multiclass data problem. All measures were obtained by five rounds of three-fold cross-validation, being training data used to tune LSVM, CRF, and LDCRF model parameters by an intern four-fold cross-validation.

Briefly, accuracy is the rate between the number of correct predictions over the total number of predictions [68], [135], [93]. Cohen's kappa index relates the accuracy with the classification by chance probability as a rate between their subtraction by its maximum value [87], [135], [152]. Balanced accuracy is the average of the accuracy obtained by each class [93].

3.2.5 Statistical analysis

A one-way randomized blocks ANOVA tested the statistical significance of differences between HG and LFC data. Only CRF and LDCRF were used as classifiers because the goal is to choose the adequate type of data to compare the proposed models against other classifiers. If the ANOVA test rejected the null hypothesis of statistical equality of averages, a post hoc comparison was performed by a Tukey-Kramer test. The average value of each classifier and data type was compared with the overall average metric, rejecting the null hypothesis if the average by class is greater than the overall average.

Also, one-way randomized blocks ANOVA tested the statistical significance of differences between the metrics obtained by comparing CRF and LDCRF against LDA and LSVM. If the ANOVA test rejected the null hypothesis of statistical equality of averages, a post hoc comparison was performed by a Tukey-Kramer test.

3.3 Results and Discussion

Results showed here refer to the average performance obtained by each classifier, measured with the accuracy metric. All metrics were obtained from the testing data of each subject.

3.3.1 Type of Data: High Gamma Band vs. Low-Frequency Component

Table 3-1 displays the accuracies acquired by each subject. ANOVA test (F: 204.193; 2 degrees of freedom –d. f.-; $p < 0.001$) rejected the null hypothesis of average equality, so we performed the post hoc test. Their p-values are depicted in Table 3-1, showing that only data with the correlation between channels significantly outperforms the average accuracy.

To verify if the performance of classifiers with given data is adequate for unbalanced data, we also evaluated Cohen's kappa index of data, written in Table 3-2, as done with accuracy. Results of the ANOVA test (F: 263.240; 2 d. f.; $p < 0.001$) show again that the null hypothesis of average equality must be rejected. P-values from the post-hoc test confirm that solely HG data has the most significant performance values.

Table 3-1. Results of Accuracy, by subject and type of data. Overall performance: 0,526

Subject	CRF			LDCRF		
	High Gamma	LFC	HG+LFC	High Gamma	LFC	HG+LFC
bp	0,785	0,625	0,491	0,789	0,622	0,492
cc	0,746	0,474	0,325	0,749	0,469	0,316
ht	0,650	0,461	0,354	0,644	0,454	0,368
jc	0,761	0,625	0,466	0,768	0,619	0,451
jp	0,791	0,253	0,399	0,786	0,259	0,413
wc	0,747	0,226	0,372	0,693	0,125	0,438
wm	0,737	0,376	0,379	0,723	0,389	0,349
zt	0,715	0,489	0,523	0,712	0,486	0,510
mv	0,743	0,372	0,339	0,767	0,446	0,367
Average	0,742	0,433	0,405	0,737	0,430	0,411
Standard Deviation	0,041	0,142	0,071	0,048	0,159	0,067
P-value	< 0,001	0,980	0,994	< 0,001	0,983	0,992

Table 3-2. Results of Cohen's kappa index, by subject and type of data. Average performance: 0,394

Subject	CRF			LDCRF		
	High Gamma	LFC	HG+LFC	High Gamma	LFC	HG+LFC
bp	0,711	0,483	0,294	0,714	0,474	0,285
cc	0,693	0,365	0,183	0,697	0,359	0,173
ht	0,568	0,326	0,190	0,560	0,317	0,206
jc	0,689	0,511	0,290	0,699	0,504	0,269
jp	0,744	0,079	0,260	0,739	0,090	0,273
wc	0,700	0,096	0,243	0,637	-0,022	0,320
wm	0,682	0,239	0,247	0,665	0,258	0,212
zt	0,590	0,120	0,208	0,585	0,110	0,192
mv	0,691	0,244	0,205	0,719	0,333	0,238
Average	0,682	0,322	0,260	0,681	0,329	0,262
Standard Deviation	0,004	0,003	0,002	0,004	0,003	0,002
P-value	< 0,001	0,943	0,995	< 0,001	0,925	0,994

3.3.2 CRF and LDCRF against other classifiers

As before, one-way randomized blocks ANOVA tested the statistical significance of differences between the accuracies. If the ANOVA test rejects the null hypothesis of statistical equality of averages, a post hoc comparison is performed by a Tukey-Kramer test. The average value of each classifier is compared with the overall average accuracy, rejecting the null hypothesis if the average performance of a classifier is greater than the overall average.

Table 3-3 displays the accuracies obtained by each subject and their corresponding standard deviations. ANOVA test ($F: 608.121; 4 \text{ d. f.}; p < 0.001$) rejected the null hypothesis, so we performed the post hoc test. Their p-values are presented in Table 3-3, showing that based-PGM classifiers significantly surpass the average accuracy. However, Table 3-3 indicates that average values for CRF and LDCRF based classifiers have a difference of around 0,001 and standard deviations only have a difference of 0,002, so there is no statistical difference between both classifiers.

To prove if the performance of given classifiers is fair for unbalanced data, we also evaluated Cohen's kappa index data, reported in Table 3-4, as done with accuracy. Results

of the ANOVA test ($F: 1337.703$; 4 d. f.-; $p < 0.001$) reject the null hypothesis. P-values from the post-hoc test and written in Table 3-4 indicate that CRF and LDCRF classifiers have the most significant kappa values.

3.3.3 Discussion

Results from Table 3-1 and Table 3-2 show that HG features exceed the performance obtained by LFC or even their combination. It could seem to refuse other works where the performance of LFC features, either alone or combined with HG, is the same or higher than HG data, even with the same dataset [125], [161]. However, these investigations refer to a different duty that is decoding. It tries to predict the actual trajectory of a finger movement rather than what finger is moving.

When we selected HG features for evaluating the classifiers, CRF and LDCRF surpassed the performance of the other classifiers by evaluating the given metrics. Moreover, kappa values indicate that LDA, QDA, and LSVM have accuracy values near to the probability to classify correctly by chance. It is because they tend to assign the most probable class.

Table 3-3. Results of Accuracy, by subject and classifier. Overall performance: 0,433

Subject	Model				
	CRF	LDCRF	LDA	QDA	LSVM
bp	0,746	0,749	0,224	0,217	0,224
cc	0,785	0,789	0,226	0,254	0,433
ht	0,743	0,767	0,202	0,205	0,200
jc	0,791	0,786	0,169	0,164	0,189
jp	0,650	0,644	0,214	0,210	0,291
mv	0,747	0,693	0,136	0,160	0,119
wc	0,737	0,723	0,187	0,181	0,185
wm	0,715	0,712	0,237	0,226	0,497
zt	0,761	0,768	0,236	0,215	0,382
Average	0,742	0,737	0,204	0,204	0,280
Standard Deviation	0,041	0,048	0,034	0,031	0,129
P-value	< 0,001	< 0,001	1,000	1,000	0,998

Table 3-4. Results of Cohen's kappa index, by subject and classifier. Overall performance: 0,278.

Subject	Model				
	CRF	LDCRF	LDA	QDA	LSVM
bp	0,711	0,714	0,011	0,024	-0,002
cc	0,693	0,697	0,056	0,049	0,041
ht	0,568	0,560	0,020	0,015	-0,003
jc	0,689	0,699	0,047	0,037	-0,004
jp	0,744	0,739	0,015	0,023	-0,029
mv	0,700	0,637	-0,016	0,015	-0,031
wc	0,682	0,665	0,018	0,007	-0,026
wm	0,590	0,585	0,030	0,023	0,000
zt	0,691	0,719	0,039	0,040	0,030
Average	0,674	0,669	0,024	0,026	-0,003
Standard Deviation	0,057	0,062	0,022	0,014	0,025
P-value	< 0,001	< 0,001	1,000	1,000	1,000

To illustrate this issue, Table 3-5 displays the confusion matrices for CRF, LDA, and LSVM classifiers. The first row of each matrix refers to the assigned label by each classifier as Rest, which could be an indicator of the accuracy paradox because label Rest has more trials than the other ones. While CRF appoints most of the labels consistently, LDA and LSVM tend to assign the Rest class to most of the data, being more notorious in LSVM. It is a substantial indication of classification biasing due to the amount of data with the Rest label shown during the training stage.

When comparing our results with a previous study that employed a CRF model with long time dependencies [157], we noticed that the performance of our proposal is higher than reported by the earlier study. While it apprised accuracy values between 0.62 and 0.69, we got values between 0.65 and 0.79 for the CRF and LDCRF classifiers. Concerning Cohen's kappa index, values in the interval of 0.49 and 0.54 were reported in the previous study, while we reported values between 0.56 and 0.74 for CRF-based models. It exhibits a significant improvement, even with a higher number of subjects (3 subjects in the previous study and 9 in the current study) and a lower sampling rate (200 Hz and 20 Hz, respectively).

A significant improvement is observed when evaluating our results with other previous studies. For example, one of the most recent studies reports that its performance is up to 0.55 with features extracted from a continuous wavelets transform [154]. Another recent study that uses a HMM model with Gaussian mixtures and a bi-gram model [156] reports values up to 0.78, using additional movement insertions. However, this study only uses four classes (discarding ring finger movements) with three subjects and a generative classifier to classify the finger movement. In addition, the metric employed does not allow to perform further comparisons. Another previous study shows performances between 0.59 and 0.69 [66] with non-linear kernel SVM but classifiers employed only were evaluated by pairs of classes.

Moreover, a study from 2016 reported balanced accuracy values up to 0.77 [155] by employing four classes and 0.64 with five classes. The study used a two-stage classifier: a binary LDA for movement and no movement and a multiclass LDA for classifying the finger moved. Balanced accuracy values come from the multiclass LDA, and after excluding trials labeled as rest class. When we converted our metrics to balanced accuracies, as Table 3-6 exhibits, the average performance raised to 0.71 for both CRF-based classifiers. With the same data, LDA, QDA, and LSVM go down to 0.16 or less, depending on the subject.

The performance accomplished by CRF and LDCRF is due to the connectivity of data, as described in Equations (1-9) and (1-10). More specifically, the connectivity across time is taken into account by based-CRF classifiers while is ignored by the remaining classifiers used in our study. It compels that these classifiers tend to assign most of the labels whose number of samples is the highest or data have more co-occurrences in their statistical modeling, as LDA and QDA do [39]. However, when we contrasted the performance of CRF and LDCRF classifiers, there was no statistical difference between them when a paired t-Student post hoc test was performed (MSE: 0.013, T: 0.130, p: 0.899, d. f.: 8). It could indicate that there is no statistical difference among all hidden states assigned by the LDCRF classifier -or the classes themselves- or there are other relations in ECoG data further time. Finally, notice that most of the approaches proposed for finger classification from ECoG data are generative. It means that these classifiers use multiple

models for training and evaluation, while the proposed method employs only a CRF-based model directly for classification, even with a sampling rate as low as 20 Hz.

3.4 Conclusions

The proposed based linear-CRF classifiers have performances significantly higher than LDA, QDA, and LSVM classifiers in the HG band, from ECoG data. It is due that the latter classifiers try to fit their models to the class with the highest quantity of samples. In contrast, linear-CRF classifiers take into account temporal dynamics to detect attributes from other classes.

Table 3-5. Confusion Matrix across all subjects for CRF, LDA and LSVM classifiers. Grayscale indicates the degree of distribution of assigned labels.

		Confusion Matrix across all subjects - CRF						Confusion Matrix across all subjects - LDA						Confusion Matrix across all subjects – LSVM					
Assigned Label	Rest	0,838	0,081	0,044	0,109	0,072	0,060	0,353	0,258	0,253	0,277	0,274	0,268	0,814	0,598	0,589	0,651	0,598	0,619
	Thumb	0,034	0,853	0,024	0,017	0,021	0,007	0,158	0,228	0,196	0,167	0,170	0,186	0,055	0,114	0,122	0,098	0,099	0,097
	Index	0,018	0,025	0,772	0,092	0,015	0,033	0,117	0,152	0,140	0,142	0,152	0,156	0,049	0,148	0,112	0,110	0,131	0,117
	Middle	0,038	0,007	0,091	0,683	0,087	0,036	0,102	0,110	0,118	0,115	0,116	0,117	0,012	0,022	0,021	0,010	0,020	0,018
	Ring	0,029	0,019	0,027	0,059	0,537	0,221	0,140	0,122	0,150	0,151	0,141	0,140	0,038	0,070	0,095	0,074	0,088	0,094
	Pinky	0,030	0,015	0,041	0,040	0,261	0,642	0,116	0,121	0,131	0,133	0,141	0,127	0,027	0,038	0,048	0,045	0,057	0,048
	Not Found	0,013	0,000	0,001	0,000	0,007	0,000	0,014	0,008	0,011	0,015	0,006	0,007	0,005	0,009	0,013	0,013	0,008	0,006
		Rest	Thumb	Index	Middle	Ring	Pinky	Rest	Thumb	Index	Middle	Ring	Pinky	Rest	Thumb	Index	Middle	Ring	Pinky
	Actual Label						Actual Label						Actual Label						

Table 3-6. Results of balanced accuracy, by subject and classifier.

Subject	Model				
	CRF	LDCRF	LDA	QDA	LSMV
bp	0,737	0,734	0,166	0,172	0,166
cc	0,735	0,738	0,203	0,200	0,189
ht	0,602	0,593	0,180	0,175	0,165
jc	0,708	0,721	0,191	0,186	0,164
jp	0,769	0,772	0,178	0,185	0,150
mv	0,766	0,706	0,137	0,158	0,106
wc	0,731	0,716	0,182	0,173	0,150
wm	0,648	0,645	0,189	0,187	0,167
zt	0,732	0,757	0,197	0,198	0,190
Average	0,714	0,709	0,180	0,182	0,161

4 ANALYSIS OF INSTANTANEOUS BRAIN INTERACTIONS CONTRIBUTION TO A MOTOR IMAGERY CLASSIFICATION TASK

Using spatial interactions to enhance a two-class classification task

4.1 Preliminaries

A Brain-Computer Interface (BCI) is a communication and control scheme that sends messages and commands to the external world without using nerves and muscles in the usual way [123]. A regular BCI system has a brain monitoring system, a signal pre-processing stage, a stage for extracting features of the pre-processed signal, and a classification stage where features are decoded into commands or messages. Some manners of BCI activation are evoked potentials, brain rhythms, and motor imagery, among others [19], [21]. Initially developed for electroencephalography (EEG) [63], [124], it was driven quickly to electrocorticography (ECoG) [64], [125] because of the advantages provided by the latter signals.

Different studies have suggested that combining frequency-temporal features with blinding source separation techniques improves the performance of classifiers in motor imagery paradigms significantly. One of the most common of these techniques is the Common Spatial Patterns (CSP), that extracts mutual features from a mixture of two populations but maximizing the different proportion of the variances in each population [162]. In this way, a linear transformation (or *spatial filtering*) is performed, preserving the number of sources but missing any temporal or brain interactions between electrodes. To compensate the lack of information, a series of improvements have been introduced to CSP. One of them is Filter bank CSP (FBCSP). Here, the collected brain signal is passed by a set of filters with different frequency bands, and each filtered signal is processed by CSP [163]. Therefore, the contribution of a series of frequency bands is preserved while CSP extracts mutual features for each band. CSP and FBCSP can be applied over raw or pre-processed data, as filtered data extracted from wavelet packet decomposition [60], [61], recombined data from CSP [164], or frequency data extracted from CSP [165]. Also, some methods have

been developed to select the most discriminative features from CSP or FBCSP [60], [61], [165]–[167].

However, since CSP and FBCSP come from a linear transformation, the resultant operation only shows the projection of each electrode to the new space, i. e. its contribution to the total variance of the chosen population respect to the joint population [168]. Hence, the contribution of brain interactions between electrodes to the classification is, at best, implicit. The purpose of the present study is to analyze statistically the contribution of the interactions between electrodes, measured here as correlation or Jaccard distance, to the classification of two actions in a motor imagery paradigm: left hand or right hand movement. The analysis is performed in two classifier models: a static model, consisting in a linear discriminant analysis classifier (LDA), and a dynamic model, consisting in a linear conditional random fields (CRF) model where features only interact with hidden variables, rather than interacting with labels, named latent dynamic CRF (LDCRF). Also, the impact of using the sliding window technique (SWT) in the static and dynamic models is analyzed here.

4.2 Materials and methods

4.2.1 Experiment and dataset description

The analyzed dataset comes from the BCI Competition IV, dataset 2b [169]. It is available at <https://www.bbc.de/competition/iv/> [170]. Dataset was recorded from nine volunteer subjects. Each one was right-handed and had normal or corrected-to-normal vision. All subjects sat in an armchair and were watching a flat screen placed 1 m away at level eye. Five sessions were performed for each subject: first two without feedback and last three with feedback. Each session consists of several runs, preceded by 5 minutes of electrooculography (EOG) estimation at the beginning of a session, as follows.

Two first sessions used the following paradigm: a cue-based screening paradigm, consisting in two classes, namely left-hand and right-hand respectively. Each session consisted of six runs, and each run had ten randomized trials by class, for a total of 120 repetitions per session. Each trial started with a fixation cross and a warning tone by a few of seconds, followed by an arrow indicating either left or right side for 1.25 s.

Subsequently, the subject imagined the hand movement for 4 s. Next, there was a randomized pause for, at least, 1.5 s for avoiding adaptation.

Three later sessions made use of a smiley face for feedback, with four runs and 20 randomized trials by class and run, for a total of 160 repetitions per session. Each trial started with a gray smiley and a warning tone by a few of second, followed by a cue period of 3 s, where the smiley had to be displaced to the left or right. In this time interval, depending of the hand movement, the smiley became red when the subject was wrong and green when the movement was the correct. Moreover, the mouth of the smiley also changed to sad (corners of mouth downwards) or happy (corners of mouth upwards), with wrong or right movements, respectively. Next, there was a randomized pause between 1 and 2 seconds. We only employed data from feedback sessions.

Data were recorded from three EEG bipolar electrodes: C3, Cz and C4, with a frequency sample of 250 Hz as Figure 4-1 shows. Fz channel was used as EEG ground. Later, data was bandpass-filtered between 0.5 and 100 Hz followed by a 50 Hz notch filter. In addition, the EOG is available from three monopolar electrodes and similar amplification configuration. Additional details of the experiment are available in [171].

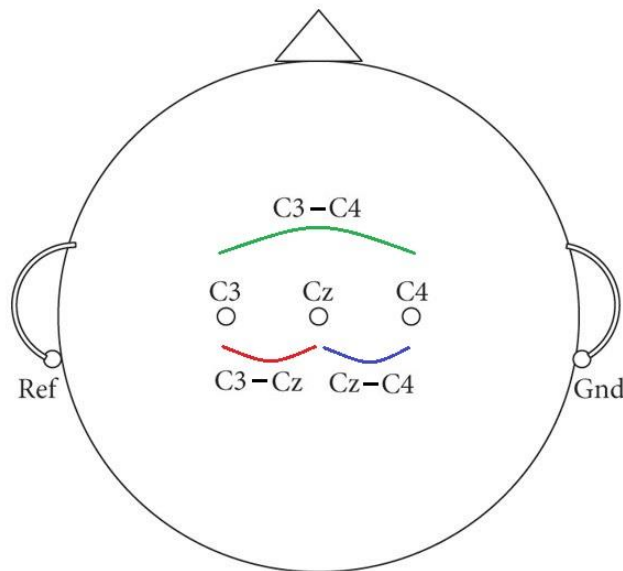


Figure 4-1. Electrodes location in the dataset BCI Competition IV, dataset 2b, and their interactions.

Source: adapted from [172].

4.2.2 Data pre-processing

We took each trial during the time interval that a subject was executing the imagery task, i. e. between 3 and 6 seconds after the beginning of the trial. Then, each trial was sub-segmented by sliding windows, varying the window size and the slide size along the experiments.

A single sub-segment can provide up to three types of features: alpha/mu (8 – 12 Hz) and beta (15 – 25 Hz) power band [173] by electrode, and a similarity (or distance) measure of interactions between a pair of electrodes signals (C3 – Cz, C4 – Cz, or C3 – C4, shown in Figure 4-1). Then, power band data was scaled such that the order of magnitude is 10^1 . Experiments used alpha and beta power band as main features and similarity (or distance) measures as additional features. All features were obtained using the sliding window technique.

4.2.3 Sliding window technique

Given a discrete time series arrangement $\mathbf{X} = \{x_1, x_2, \dots, x_{N-1}, x_N\}$ for N equally spaced time samples, an sliding window technique (SWT) is a set of instructions that is executed over a subset of k consecutive values of \mathbf{X} , whose initial point is given by x_i : $\mathbf{X}_{i,k} = \{x_i, x_{i+1}, \dots, x_{i+k-1}, x_{i+k}\}$. Once the set of instructions is executed, the position of the initial point is displaced by a distance Δi and the algorithm takes another k points to set the new subset $\mathbf{X}_{i+\Delta i,k}$. Instructions are executed in the given subset. The routine continues repeatedly until the value of \mathbf{X} corresponding to the final point x_N is reached. Size of the subset k and the displacement of initial point Δi could be fixed [174] or dynamically adaptive [175].

SWT is useful to get a simpler representation [175], [176] or finding dynamic patterns [177], [178] in a time series set. In this way, we propose to apply the SWT to find the alpha and beta power values and the similarity (or distance) measures, in order to reduce the memory consumption and time of execution [179].

4.2.4 Pearson correlation

Pearson correlation, defined for two zero-mean and real-valued random variables x, y , is the coefficient between the cross-correlation of the random variables $E[xy]$ and the product of the square root of their variances $\sigma_x \sigma_y$ [95], [96].

4.2.5 Jaccard distance

On the other hand, the Jaccard distance comes from its counterpart, the Jaccard index J . The latter is a similarity measure for two finite discrete sets A, B , defined as the coefficient between the size of the intersection $|A \cap B|$ and the size of its union $|A \cup B|$ [97]. It can be extended as the ratio between the measure of the intersection $\mu(A \cap B)$ and its union $\mu(A \cup B)$, with an arbitrary measure μ . Details about Jaccard distance are found in the section 1.5.2.

4.2.6 Performance metrics

A common measure of performance used for classification is the *accuracy*. Accuracy is defined as a metric of the closeness between measured or predicted values and their corresponding true values [68]. Accuracy varies from 0 to 1, where 1 gives a perfect classification. The study used five rounds of three-fold cross-validation, being training data used to tune LDA and LDCRF model parameters by an intern four-fold cross-validation.

4.2.7 Statistical analysis

A one-way randomized blocks ANOVA tested the statistical significance of differences between data with additional features or not. In the case that ANOVA test rejects the null hypothesis of statistical equality of averages, a post hoc comparison is performed using a Tukey-Kramer test. The average value of each classifier and type of data is compared against the overall average accuracy, rejecting the null hypothesis if the average by class is greater than the overall average.

Also, a linear multiple way randomized blocks ANOVA tested the statistical significance of differences of the data. Window size and slide size were the tested parameters, and subjects were implemented as randomized blocks in the model. In the case that ANOVA test rejects the null hypothesis of statistical equality of averages, a post hoc comparison is performed using a Tukey-Kramer test.

4.3 Results and Discussion

Results showed here refer to the average performance obtained by each classifier, measured with the accuracy metric. All metrics were obtained from the testing dataset of each subject.

4.3.1 Results with LDA model

Table 4-1 shows the accuracies obtained by subject, by comparing data with and without correlation features with the LDA classifier model. The whole window of 3s was used here, without modifying other parameters. According to the ANOVA test ($F: 34.922$; 1 degree of freedom –d. f.; $p < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. Their p-values are illustrated in Table 4-1, showing that only data with additional features surpasses significantly the average accuracy.

On the other hand, Table 4-2 shows the accuracies obtained by subject, by comparing data with and without Jaccard Distance features with the LDA classifier model. As before, a one-way randomized blocks ANOVA tested the statistical significance of differences between data with additional features or without. According to the ANOVA test ($F: 26.613$; 1 d. f.; $p < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. It indicates again that only data with additional features surpasses significantly the average accuracy.

The following step is to implement the sliding window algorithm in the data with power alpha and beta band, and additional features, in order to get more features. Hence, we used three sizes of sliding window (0.5, 1 and 2 s) against the whole 3 s window, and three slide sizes (0.125, 0.25 and 0.5 s) to compare the performance of the LDA classifier. Table 4-3 and Table 4-4 show the performance of LDA model by window size and slide size, by implementing the sliding window algorithm with correlation as additional features.

According to the ANOVA test ($F_{\text{window_size}}: 46.61$; 3 d. f.; $p_{\text{window_size}} < 0.001$; $F_{\text{slide_size}}: 144.60$; 2 d. f.; $p_{\text{slide_size}} < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. Results of Table 4-3 show that window sizes of 2 and 3 s have the most significant performance. Meanwhile, Table 4-4 show that slides of 0.25 and 0.5 have the greatest and most significant performances.

Table 4-1. Results of Accuracy in LDA model, with and without correlation. Overall performance: 0,74.

Subject	Presence of Correlation	
	YES	NO
1	0,7569	0,7144
2	0,5692	0,5267
3	0,6227	0,5802
4	0,9750	0,9325
5	0,7981	0,7556
6	0,7577	0,7152
7	0,7119	0,6694
8	0,9015	0,8590
9	0,7623	0,7198
Average	0,7617	0,7192
Standard Deviation	0,125	0,125
P-value	< 0,001	1,000

Table 4-2. Results of Accuracy in LDA, with and without Jaccard Distance Overall performance: 0,741.

Subject	Presence of Jaccard Distance	
	YES	NO
1	0,7432	0,7043
2	0,5645	0,5255
3	0,6041	0,5651
4	0,9763	0,9374
5	0,8209	0,7820
6	0,7638	0,7249
7	0,7072	0,6682
8	0,8978	0,8589
9	0,7674	0,7284
Average	0,7606	0,7216
Standard Deviation	0,13	0,13
P-value	< 0,001	1,000

Table 4-3. Results of Accuracy in LDA model with correlation as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,751.

Subject	Sliding Window Size (s)			
	0.5	1	2	3
1	0,667	0,685	0,698	0,699
2	0,534	0,553	0,566	0,567
3	0,620	0,638	0,651	0,653
4	0,906	0,925	0,938	0,939
5	0,792	0,810	0,823	0,824
6	0,713	0,731	0,744	0,745
7	0,772	0,790	0,803	0,805
8	0,848	0,866	0,879	0,880
9	0,721	0,739	0,752	0,753
Average	0,731	0,749	0,762	0,763
Standard Deviation	0,115	0,115	0,115	0,115
P-value	1,000	1,000	< 0,001	< 0,001

Table 4-4. Results of Accuracy in LDA model with correlation as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,751.

Subject	Slide Size (s)		
	0.125	0.25	0.5
1	0,661	0,694	0,705
2	0,529	0,562	0,573
3	0,615	0,648	0,659
4	0,901	0,934	0,945
5	0,787	0,819	0,831
6	0,708	0,741	0,752
7	0,767	0,800	0,811
8	0,843	0,876	0,887
9	0,716	0,749	0,760
Average	0,725	0,758	0,769
Standard Deviation	0,115	0,115	0,115
P-value	1,000	< 0,001	< 0,001

Respect to the Jaccard Distance as additional parameter, a similar procedure to the implemented for correlation was performed. According to the ANOVA test ($F_{\text{window_size}}$: 30.28; 3 d. f.; $p_{\text{window_size}} < 0.001$; $F_{\text{slide_size}}$: 137.73; 2 d. f.; $p_{\text{slide_size}} < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. Results of Table 4-5 show that window sizes of 2 and 3 s have the most significant performance. Meanwhile, Table 4-6 show that slides of 0.25 and 0.5 have the greatest and most significant performances.

4.3.2 Results with HCRF model

Since HCRF is a dynamic model, opposite to LDA model, implementation of sliding window algorithm is necessary to establish the corresponding timestamps of HCRF. To compare data with and without additional correlation features, we tested the model with two window sizes (0.5 and 2 s) and two slide sizes (0.125 and 0.5 s). Table 4-7 shows the accuracies obtained by subject, by comparing data with and without correlation features with the HCRF classifier model. According to the ANOVA test (F : 117.232; 1 degree of freedom –d. f.-; $p < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. Their p-values are illustrated in Table 4-7, showing that only data with additional features surpasses significantly the average accuracy.

On the other hand, Table 4-8 shows the accuracies obtained by subject, by comparing data with and without Jaccard Distance features with the HCRF classifier. As before, a one-way randomized blocks ANOVA tested the statistical significance of differences between data with additional features or without. According to the ANOVA test (F : 148.475; 1 d. f.; $p < 0.001$), the null hypothesis of averages equality must be rejected, so we performed the post hoc test. It indicates again that only data with additional features surpasses significantly the average accuracy.

The following step is to implement the sliding window algorithm in the data with power alpha and beta band, and additional features, in order to get more features. Hence, we used three sizes of sliding window (0.5, 1 and 2 s), and three slide sizes (0.125, 0.25 and 0.5 s) to compare the performance of the HCRF classifier. Table 4-9 and Table 4-10 show the performance of HCRF model by window size and slide size, by implementing the sliding window algorithm with correlation as additional features.

Table 4-5. Results of Accuracy in LDA model with Jaccard Distance as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,752.

Subject	Sliding Window Size (s)			
	0.5	1	2	3
1	0,681	0,691	0,705	0,703
2	0,539	0,549	0,563	0,561
3	0,613	0,622	0,637	0,634
4	0,911	0,920	0,934	0,932
5	0,835	0,844	0,859	0,856
6	0,728	0,737	0,751	0,749
7	0,752	0,762	0,776	0,774
8	0,860	0,870	0,884	0,882
9	0,725	0,734	0,748	0,746
Average	0,738	0,748	0,762	0,760
Standard Deviation	0,119	0,119	0,119	0,119
P-value	1,000	1,000	< 0,001	< 0,001

Table 4-6. Results of Accuracy in LDA model with Jaccard Distance as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,752.

Subject	Slide Size (s)		
	0.125	0.25	0.5
1	0,672	0,701	0,712
2	0,530	0,559	0,570
3	0,604	0,633	0,643
4	0,901	0,931	0,941
5	0,826	0,855	0,865
6	0,718	0,748	0,758
7	0,743	0,772	0,783
8	0,851	0,880	0,891
9	0,715	0,745	0,755
Average	0,729	0,758	0,769
Standard Deviation	0,119	0,119	0,119
P-value	1,000	< 0,001	< 0,001

Table 4-7. Results of Accuracy in HCRF model, with and without correlation. Overall performance: 0,757.

Subject	Presence of Correlation	
	YES	NO
1	0,7515	0,7121
2	0,5961	0,5567
3	0,5683	0,5288
4	0,9905	0,9511
5	0,7903	0,7509
6	0,7777	0,7383
7	0,7634	0,7239
8	0,9190	0,8796
9	0,8310	0,7916
Average	0,7764	0,7370
Standard Deviation	0,135	0,135
P-value	< 0,001	1,000

Table 4-8. Results of Accuracy in HCRF, with and without Jaccard Distance. Overall performance: 0,758.

Subject	Presence of Jaccard Distance	
	YES	NO
1	0,7528	0,7112
2	0,6000	0,5584
3	0,5707	0,5291
4	0,9919	0,9503
5	0,8086	0,7670
6	0,7883	0,7468
7	0,7498	0,7082
8	0,9192	0,8776
9	0,8319	0,7904
Average	0,7793	0,7377
Standard Deviation	0,135	0,135
P-value	< 0,001	1,000

Table 4-9. Results of Accuracy in HCRF model with correlation as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,777.

Subject	Sliding Window Size (s)		
	0.5	1	2
1	0,729	0,734	0,728
2	0,568	0,573	0,567
3	0,597	0,601	0,596
4	0,971	0,975	0,970
5	0,820	0,825	0,820
6	0,768	0,773	0,767
7	0,830	0,834	0,829
8	0,897	0,902	0,896
9	0,805	0,810	0,804
Average	0,776	0,781	0,775
Standard Deviation	0,130	0,130	0,130
P-value	1,000	< 0,001	1,000

Table 4-10. Results of Accuracy in HCRF model with correlation as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,777.

Subject	Slide Size (s)		
	0.125	0.25	0.5
1	0,732	0,730	0,728
2	0,572	0,569	0,568
3	0,600	0,598	0,596
4	0,974	0,972	0,970
5	0,824	0,822	0,820
6	0,771	0,769	0,767
7	0,833	0,831	0,829
8	0,900	0,898	0,896
9	0,808	0,806	0,804
Average	0,780	0,777	0,776
Standard Deviation	0,130	0,130	0,130

According to the ANOVA test ($F_{\text{window_size}}$: 4.20; 2 d. f.; $p_{\text{window_size}}$ = 0.016; $F_{\text{slide_size}}$: 1.95; 2 d. f.; $p_{\text{slide_size}}$ = 0.143), the null hypothesis of averages equality must be rejected only for the window size, so we performed the post hoc test. Results of Table 4-9 show that window size of 1 s has the most significant performance. Meanwhile, Table 4-10 shows that although the performance of the classifier is the highest when the size of slides is 0.125 s, the average result is slightly better than the average performance with whole data.

Respect to the Jaccard Distance as additional parameter, a similar procedure to the implemented for correlation was performed. According to the ANOVA test ($F_{\text{window_size}}$: 12.35; 3 d. f.; $p_{\text{window_size}}$ < 0.001; $F_{\text{slide_size}}$: 0.15; 2 d. f.; $p_{\text{slide_size}}$ = 0.86), the null hypothesis of averages equality must be rejected only for window size, so we performed the corresponding post hoc test. Results of Table 4-11 show that window sizes of 0.5 s has the most significant performance. Meanwhile, Table 4-12 shows that slides of 0.125 s have the greatest average performance, although it is not significant compared with the other slide sizes.

4.3.3 Discussion

Results from Table 4-1, Table 4-2, Table 4-7 and Table 4-8 suggest that adding either correlation or Jaccard distance to temporal features improves the performance of tested classifiers significantly, i. e. LDA and HCRF. It indicates that having available information of similarity or distance relations between channels gives additional knowledge about the classes that carry to a more accurate classification. However, it is the only common behavior of the models when brain interactions are added to the features. Also, it is important to remark that, although correlation and Jaccard distance are measures of interactions with distinct nature, we get an improvement of performance. It means that adding interactions between electrodes improves significantly the performance of each classifiers, regarding the nature of the interaction measure.

Table 4-11. Results of Accuracy in HCRF model with Jaccard Distance as additional feature, by varying the window size in the sliding window algorithm. Overall performance: 0,778.

Subject	Sliding Window Size (s)		
	0.5	1	2
1	0,733	0,731	0,723
2	0,576	0,573	0,565
3	0,606	0,603	0,595
4	0,975	0,972	0,964
5	0,853	0,850	0,842
6	0,780	0,778	0,770
7	0,803	0,801	0,793
8	0,902	0,900	0,892
9	0,809	0,806	0,798
Average	0,782	0,779	0,771
Standard Deviation	0,129	0,129	0,129
P-value	< 0,001	< 0,001	1,000

Table 4-12. Results of Accuracy in HCRF model with Jaccard Distance as additional feature, by varying the slide size in the sliding window algorithm. Overall performance: 0,778.

Subject	Slide Size (s)		
	0.125	0.25	0.5
1	0,730	0,729	0,728
2	0,572	0,571	0,571
3	0,602	0,601	0,601
4	0,971	0,970	0,970
5	0,849	0,848	0,848
6	0,777	0,776	0,776
7	0,800	0,799	0,798
8	0,899	0,898	0,898
9	0,805	0,804	0,804
Average	0,778	0,777	0,777
Standard Deviation	0,129	0,129	0,129

In the LDA model, results from Table 4-3 to Table 4-6 show that only a few additional information provided by the sliding window technique is enough to improve the classification. It is performed by adding brain interactions from 2 s sized windows and displacements of 0.5 s, or even with the whole trial with no shifts. It is due to the nature of the LDA model, where a dimensionality reduction of data to 1 dimension is necessary to perform the discrimination analysis [180]. When more dimensions is added to the features – by using smaller sizes of the displacement of the sliding window- the complexity of data is enhanced because of their dimensionality. Dimensionality reduction always implies loss and distortion of information, so preserving most of them in a tractable processing core is mandatory [181], [182]. One way is managing data with reduced dimensions before dimensionality reduction, which preserves information with less distortion and loss. It occurs when using brain interactions from 2 s or higher sized windows and displacements of 0.5 s.

On the other hand, HCRF model has a different behavior. Results from Table 4-9 to Table 4-12 show that the window size is the most relevant parameter to take into account to implement the SWT. Simultaneously, the change in the slide size had no or little effect in the classifying performance. This indicates that information from window size smaller than or equal to 1/3 trial size is more relevant than that from larger windows. It also leads to use the longest available window displacements without losing relevant information, reducing the features dimensionality and, hence, the computational load of the model.

4.4 Conclusions

Although it is proved that brain interactions do not contribute significantly to classification in a multiclass task by themselves [153], this study proved that their combination with temporal features provide significant information to improve the classification in a two class task, as motor imagery. Also, we showed that adding interactions between electrodes improves significantly the performance of each classifiers, regarding the nature of the interaction measure.

APPENDICES

APPENDIX 1. ADDITIONAL MODELS AND MULTIVARIATE BCI DATASETS TESTED

We also tested some additional models and features, with and without interactions between electrodes (correlation). However, their results did not have any significant differences. In this way, we present these results to show the behavior of some classifiers when we present a multiclass problem applied to BCIs.

Unless otherwise stated, we used the dataset 2a from BCI competition IV [183]. Here, dataset was recorded from nine volunteer subjects. Each one was right-handed and had normal or corrected-to-normal vision. All subjects sat in an armchair and were watching a flat screen placed 1 m away at level eye. Two sessions were performed for each subject. Each session consists of several runs, preceded by 5 minutes of electrooculography (EOG) estimation at the beginning of a session, as follows.

Both sessions used the following paradigm: a cue-based screening paradigm, consisting in four classes, namely left-hand, right-hand, foot or tongue respectively. Each session consisted of six runs, and each run had ten randomized trials by class, for a total of 288 repetitions per session. Each trial started with a fixation cross and a warning tone by a few of seconds, followed by an arrow indicating either left or right side for 1.25 s. Subsequently, the subject imagined the hand movement for 4 s. Next, there was a randomized pause for, at least, 1.5 s for avoiding adaptation [183].

Data were recorded from twenty-two EEG monopolar electrodes, with the left mastoid serving as reference and the right mastoid as ground, and a frequency sample of 250 Hz. Later, data was band-pass-filtered between 0.5 and 100 Hz followed by a 50 Hz notch filter. In addition, the EOG is available from three monopolar electrodes and similar amplification configuration. Additional details of the experiment are available in [183].

A-1 1. ADDITIONAL MODELS TESTED

A-1.1.1. Multi-view hidden conditional random fields (MVHCRF)

To expand the HCRF model, we could include a set of two or more attached linear hidden CRF models. This HCRF model is known as Multi-view hidden conditional random fields (MVHCRF) [184]. The main idea is to attach these models by their hidden states, either by linking those hidden states located in the same timestamp (linked MVHCRF) or coupling a hidden state in a specific timestamp and HCRF with the next timestamp in another HCRF (coupled MVHCRF), as Figure A-1-1 shows [184]. It implies that, unlike other CRF models, the MVHCRF model has other dependencies between hidden states, in addition to those established by the existing Markov chain in other linear CRF models. Note that both kind of MVHCRF models can be combined as a single linked-coupled MVHCRF model.

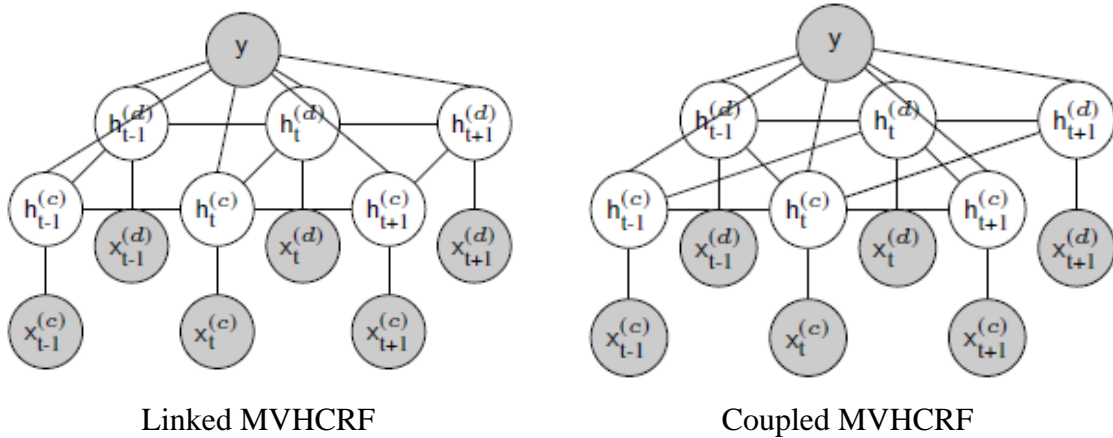


Figure A-1-1. A linked MVHCRF graphical model (left) and a coupled MVHCRF graphical model (right) with three timestamps and two HCRF models.

Source: [184].

A-1.1.2. Hidden-Unit CRF (HUCRF)

The Hidden-Unit CRF [86] could look similar to the LDCRF model, as Figure A-1-2 shows. However, HUCRF establishes the timestamps connection between outputs, as the linear CRF model does, rather than hidden states, as LDCRF does. In addition, its hidden states are conditionally independent given the features and labels [86] rather than establishing a set of hidden variables for each label, as LDCRF does [84].

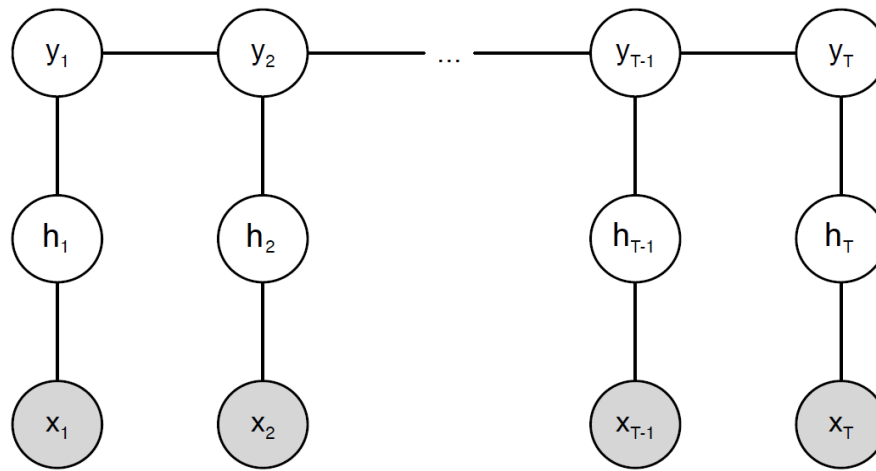


Figure A-1-2. A graphical model of HUCRF

Source: adapted from [86].

A-1 2. ADDITIONAL RESULTS

As mentioned before, we tested some additional models and features, with and without interactions between electrodes (correlation). However, their results did not have any significant differences. These results are presented below, indicating:

- Model (or models) used.
- Database and chapter whit their description (if applies).
- Methodology (including type of features and chapter whit their description).
- Accuracy values.

A-1.2.1. MVHCRF

- Model: Multi-view hidden conditional random fields
- Database: BCI competition IV 2B (chapter 4)
- Methodology
 - Detrending and Common Average Reference (CAR).
 - Trial interval used: 3 s (3rd to 6th second).
 - Types of features:
 - Alpha/mu (8 – 12 Hz) power band
 - Beta (15 – 25 Hz) power band
 - Alpha and beta power bands combined.
 - Sliding window technique (chapter 4):
 - Window size 2s
 - Delay size 500 ms.
 - 5 repetitions of 5-folds cross-validation.
 - Classifiers to be compared: LDA, HCRF, MVHCRF
- Results: in all of cases, MVHCRF gives the lower accuracy values, compared to LDA and HCRF (Table A-1-1).

Table A-1-1. Results of LDA, HCRF and MVHCRF implementation, by patient.

Patient	Alpha			Beta			Alpha and Beta		
	LDA	HCRF	MVHCRF	LDA	HCRF	MVHCRF	LDA	HCRF	MVHCRF
B01	0,677	0,670	0,598	0,664	0,670	0,645	0,705	0,670	0,520
B02	0,595	0,586	0,511	0,561	0,543	0,489	0,598	0,586	0,514
B03	0,550	0,468	0,498	0,518	0,475	0,495	0,557	0,468	0,489
B04	0,943	0,943	0,935	0,820	0,843	0,752	0,939	0,943	0,935
B05	0,622	0,567	0,511	0,772	0,796	0,726	0,674	0,567	0,565
B06	0,770	0,773	0,666	0,666	0,634	0,580	0,766	0,773	0,664
B07	0,730	0,723	0,730	0,632	0,627	0,595	0,727	0,723	0,709
B08	0,820	0,827	0,768	0,752	0,741	0,743	0,818	0,827	0,705
B09	0,745	0,734	0,650	0,773	0,789	0,668	0,770	0,734	0,673
Average	0,717	0,699	0,652	0,684	0,680	0,633	0,728	0,699	0,641

A-1.2.2. CRF

- Model: Conditional random fields
- Database: Finger flexion (Chapter 3)
- Methodology
 - Detrending and Common Average Reference (CAR).
 - Trial interval used: 3 s (3rd to 6th second).
 - Types of features:
 - High-gamma band (70 – 170 Hz) envelope.
 - Instantaneous interactions: correlation.
 - Down-sampling to 20 Hz.
 - Sliding window technique (chapter 4):
 - Window size 500 ms or 2s
 - Delay size 500 ms.
 - 3 repetitions of 5-folds cross-validation.
 - Classifiers to be compared: CRF
- Results: in all of cases, there are no significant differences by adding interactions between electrodes or not (Table A-1-2).

Table A-1-2. Results of CRF implementation, by patient.

Patient	No Correlation		Correlation	
	Window size 500	Window size 2000	Window size 500	Window size 2000
1	0,713	0,711	0,696	0,694
2	0,750	0,747	0,733	0,731
3	0,731	0,729	0,715	0,712
4	0,757	0,755	0,741	0,738
5	0,609	0,607	0,593	0,590
6	0,710	0,708	0,694	0,691
7	0,721	0,718	0,704	0,702
8	0,701	0,698	0,684	0,682
9	0,748	0,746	0,732	0,729
Average	0,715	0,713	0,699	0,697

A-1.2.3. HUCRF

- Model: Hidden-unit conditional random fields
- Database: Finger flexion (Chapter 3)
- Methodology
 - Detrending and Common Average Reference (CAR).
 - Trial interval used: 3 s (3rd to 6th second).
 - Types of features:
 - High-gamma band (70 – 170 Hz) envelope.
 - Instantaneous interactions: correlation.
 - Down-sampling to 20 Hz.
 - Sliding window technique (chapter 4):
 - Window size 500 ms or 2s
 - Delay size 500 ms.
 - 3 repetitions of 5-folds cross-validation.
 - Classifiers to be compared: HUCRF
- Results: in all of cases, HUCRF gives accuracy values lower than 0.25, compared to LDA and HCRF (Table A-1-3). Also, there are no significant differences by adding interactions between electrodes or not.

Table A-1-3. Results of HUCRF implementation, by patient.

Patient	No Correlation		Correlation	
	Window size 500 ms	Window size 2 s	Window size 500 ms	Window size 2 s
1	0,183	0,188	0,177	0,182
2	0,294	0,299	0,288	0,293
3	0,156	0,161	0,150	0,155
4	0,201	0,206	0,195	0,200
5	0,169	0,174	0,163	0,168
6	0,148	0,154	0,143	0,148
7	0,153	0,158	0,147	0,152
8	0,436	0,441	0,430	0,435
9	0,205	0,210	0,199	0,205
Average	0,216	0,221	0,210	0,215

A-1.2.4. CRF – Alpha and Beta power band

- Model: Conditional random fields
- Database: BCI competition IV 2A
- Methodology
 - Detrending and Common Average Reference (CAR).
 - Trial interval used: 3 s (3rd to 6th second).
 - Types of features:
 - Alpha/mu (8 – 12 Hz) power band
 - Beta (15 – 25 Hz) power band
 - Instantaneous interactions: correlation.
 - Features selection: Mutual Information (MI).
 - Sliding window technique (chapter 4):
 - Window size 1s
 - Delay size 500 ms.
 - 3 repetitions of 5-folds cross-validation.
- Results: in all of cases, CRF gives accuracy values lower than 0.3 (Table A-1-4).
There are no significant differences by adding interactions between electrodes or not.

Table A-1-4. Results of CRF implementation with Alpha and Beta power band features, by patient.

Patient	No Correlation					Correlation				
	Number of components					Number of components				
	3	6	9	12	15	3	6	9	12	15
1	0,327	0,342	0,337	0,330	0,322	0,333	0,347	0,342	0,335	0,328
2	0,231	0,246	0,241	0,234	0,226	0,236	0,251	0,246	0,239	0,231
3	0,291	0,305	0,300	0,293	0,285	0,296	0,310	0,306	0,299	0,291
4	0,265	0,280	0,275	0,268	0,260	0,271	0,285	0,280	0,273	0,265
5	0,234	0,249	0,244	0,237	0,229	0,240	0,254	0,249	0,242	0,234
6	0,230	0,245	0,240	0,233	0,225	0,236	0,250	0,245	0,238	0,231
7	0,282	0,297	0,292	0,285	0,277	0,287	0,302	0,297	0,290	0,282
8	0,285	0,300	0,295	0,288	0,280	0,291	0,305	0,300	0,293	0,285
9	0,280	0,295	0,290	0,283	0,275	0,286	0,300	0,295	0,288	0,281
Average	0,270	0,284	0,279	0,272	0,264	0,275	0,289	0,284	0,278	0,270

A-1.2.5. CRF – Filter bank Common Spatial Patterns (FBCSP)

- Model: Conditional random fields
- Database: BCI competition IV 2A
- Methodology
 - Detrending and Common Average Reference (CAR).
 - Trial interval used: 3 s (3rd to 6th second).
 - Types of features:
 - Filter bank CSP (chapter 4).
 - Instantaneous interactions: correlation.
 - Features selection: Mutual Information (MI).
 - Sliding window technique (chapter 4):
 - Window size 1s
 - Delay size 500 ms.
 - 3 repetitions of 5-folds cross-validation.
 - Classifiers to be compared: CRF

Results: in all of cases, CRF gives accuracy values lower than 0.55 (Table A-1-5 and Figure A-1-3). Also, there are no significant differences by adding interactions between electrodes or not. However, the number of components influences significantly in the results ($F: 11.44, 19 \text{ d.o.f.}, p\text{-value} < 0.001$), being 4 the number of components whose accuracies give the most significant values.

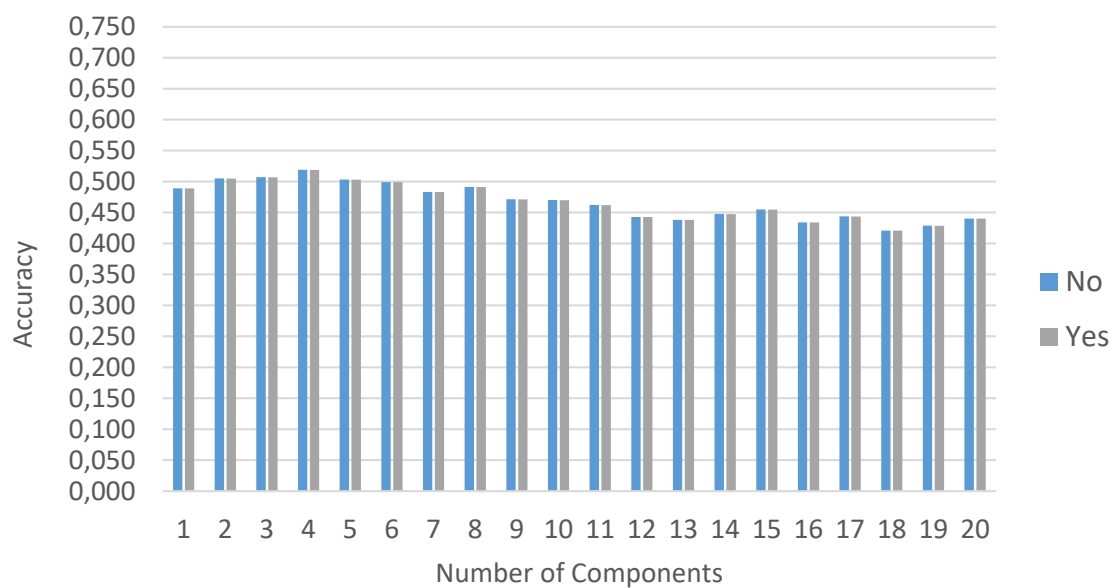


Figure A-1-3. Summarization of results of CRF implementation with CSP features.

Table A-1-5. Results of CRF implementation with CSP features, by patient.

Patient	Features with No correlation																			
	Number of components																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0,565	0,581	0,583	0,595	0,579	0,575	0,559	0,567	0,547	0,546	0,538	0,519	0,514	0,523	0,531	0,510	0,520	0,497	0,505	0,516
2	0,427	0,443	0,445	0,457	0,441	0,437	0,421	0,429	0,409	0,408	0,400	0,381	0,376	0,385	0,393	0,372	0,382	0,359	0,367	0,378
3	0,644	0,660	0,662	0,674	0,659	0,655	0,638	0,647	0,627	0,625	0,617	0,598	0,593	0,603	0,610	0,589	0,599	0,576	0,584	0,596
4	0,377	0,393	0,395	0,407	0,391	0,387	0,371	0,379	0,359	0,358	0,350	0,331	0,326	0,336	0,343	0,322	0,332	0,309	0,317	0,328
5	0,428	0,444	0,446	0,457	0,442	0,438	0,422	0,430	0,410	0,409	0,401	0,381	0,377	0,386	0,393	0,373	0,382	0,360	0,367	0,379
6	0,377	0,393	0,395	0,407	0,391	0,387	0,371	0,379	0,360	0,358	0,350	0,331	0,326	0,336	0,343	0,322	0,332	0,309	0,317	0,328
7	0,534	0,550	0,552	0,563	0,548	0,544	0,528	0,536	0,516	0,515	0,507	0,487	0,483	0,492	0,499	0,479	0,488	0,466	0,473	0,485
8	0,527	0,543	0,545	0,557	0,541	0,537	0,521	0,529	0,509	0,508	0,500	0,481	0,476	0,486	0,493	0,472	0,482	0,459	0,467	0,478
9	0,523	0,539	0,541	0,553	0,537	0,533	0,517	0,525	0,505	0,504	0,496	0,477	0,472	0,482	0,489	0,468	0,478	0,455	0,463	0,474
Average	0,489	0,505	0,507	0,519	0,503	0,499	0,483	0,491	0,471	0,470	0,462	0,443	0,438	0,448	0,455	0,434	0,444	0,421	0,429	0,440

Table A-1-5. (Cont.) Results of CRF implementation with CSP features, by patient.

Patient	Features with Correlation																			
	Number of components																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0,565	0,581	0,583	0,595	0,579	0,575	0,559	0,567	0,547	0,546	0,538	0,518	0,514	0,523	0,530	0,510	0,519	0,497	0,504	0,516
2	0,427	0,443	0,445	0,456	0,441	0,437	0,421	0,429	0,409	0,408	0,400	0,380	0,376	0,385	0,392	0,372	0,381	0,358	0,366	0,378
3	0,644	0,660	0,662	0,674	0,658	0,654	0,638	0,646	0,626	0,625	0,617	0,598	0,593	0,603	0,610	0,589	0,599	0,576	0,584	0,595
4	0,377	0,393	0,395	0,407	0,391	0,387	0,371	0,379	0,359	0,358	0,350	0,330	0,326	0,335	0,342	0,322	0,331	0,309	0,316	0,328
5	0,428	0,444	0,446	0,457	0,442	0,438	0,422	0,430	0,410	0,409	0,401	0,381	0,377	0,386	0,393	0,373	0,382	0,360	0,367	0,379
6	0,377	0,393	0,395	0,407	0,391	0,387	0,371	0,379	0,359	0,358	0,350	0,331	0,326	0,336	0,343	0,322	0,332	0,309	0,317	0,328
7	0,534	0,549	0,552	0,563	0,548	0,544	0,528	0,536	0,516	0,515	0,506	0,487	0,483	0,492	0,499	0,478	0,488	0,465	0,473	0,485
8	0,527	0,543	0,545	0,557	0,541	0,537	0,521	0,529	0,509	0,508	0,500	0,481	0,476	0,485	0,493	0,472	0,482	0,459	0,467	0,478
9	0,523	0,539	0,541	0,553	0,537	0,533	0,517	0,525	0,505	0,504	0,496	0,477	0,472	0,481	0,489	0,468	0,478	0,455	0,462	0,474
Average	0,489	0,505	0,507	0,519	0,503	0,499	0,483	0,491	0,471	0,470	0,462	0,443	0,438	0,447	0,455	0,434	0,444	0,421	0,429	0,440

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain Computer Interfaces for communication and control,” *Front. Neurosci.*, vol. 4, no. 113, pp. 767–791, 2002, doi: 10.3389/conf.fnins.2010.05.00007.
- [2] A. Kabbara, M. Khalil, W. El-Falou, H. Eid, and M. Hassan, “Functional brain connectivity as a new feature for P300 speller,” *PLoS One*, 2016, doi: 10.1371/journal.pone.0146282.
- [3] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, “Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic.,” *Comput. Intell. Neurosci.*, p. 8 pages, 2007, doi: 10.1155/2007/79642.
- [4] E. W. Sellers and E. Donchin, “A P300-based brain-computer interface: Initial tests by ALS patients,” *Clin. Neurophysiol.*, 2006, doi: 10.1016/j.clinph.2005.06.027.
- [5] H. Gevensleben *et al.*, “Is neurofeedback an efficacious treatment for ADHD? A randomised controlled clinical trial.,” *J. Child Psychol. Psychiatry.*, vol. 50, no. 7, pp. 780–9, 2009, doi: 10.1111/j.1469-7610.2008.02033.x.
- [6] A. Kübler and F. Nijboer, “Brain-Computer Interfaces for Communication and Motor Control — Perspectives on Clinical Applications,” *Toward. Brain-Computer Interfacing*, pp. 373–391, 2007.
- [7] C. Schnakers *et al.*, “Detecting consciousness in a total locked-in syndrome: An active event-related paradigm,” *Neurocase*, 2009, doi: 10.1080/13554790902724904.
- [8] R. Krepki, B. Blankertz, G. Curio, and K. R. Müller, “The Berlin Brain-Computer Interface (BBCI) - Towards a new communication channel for online control in gaming applications,” *Multimed. Tools Appl.*, vol. 33, no. 1, pp. 73–90, 2007, doi: 10.1007/s11042-006-0094-3.
- [9] G. Edlinger and C. Guger, “Social environments, mixed communication and goal-oriented control application using a brain-computer interface,” *Univers. Access Human-Computer Interact. Users Divers.*, pp. 545–554, 2011, doi: 10.1007/978-3-642-21663-3_59.
- [10] B. van de Laar, B. Reuderink, D. Plass-Oude Bos, and D. Heylen, “Evaluating User

- Experience of Actual and Imagined Movement in BCI Gaming,” *Int. J. Gaming Comput. Simulations*, vol. 2, no. 4, pp. 33–47, 2010, doi: 10.4018/jgcms.2010100103.
- [11] M. M. Jackson and R. Mappus, “Applications for Brain-Computer Interfaces,” *Brain-Computer Interfaces*, pp. 89–103, 2010, doi: 10.1007/978-1-84996-272-8_6.
- [12] A. Campbell *et al.*, “NeuroPhone: Brain-Mobile Phone Interface using a Wireless EEG Headset,” in *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds - MobiHeld '10*, 2010, pp. 3–8, doi: 10.1145/1851322.1851326.
- [13] Y.-T. Wang, Y. Wang, and T.-P. Jung, “A cell-phone-based brain-computer interface for communication in daily life,” *J. Neural Eng.*, vol. 8, p. 025018, 2011, doi: 10.1088/1741-2560/8/2/025018.
- [14] G. Edlinger, C. Holzner, C. Groenegress, C. Guger, M. Slater, and G. Technologies, “Goal-Oriented Control with Brain-Computer Interface,” *Methods*, pp. 732–740, 2009, doi: 10.1007/978-3-642-02812-0_83.
- [15] Y. Su *et al.*, “A hybrid brain-computer interface control strategy in a virtual environment,” *J. Zhejiang Univ. Sci. C*, vol. 12, no. 5, pp. 351–361, 2011, doi: 10.1631/jzus.C1000208.
- [16] C. Groenegress, C. Holzner, C. Guger, and M. Slater, “Effects of P300-Based BCI Use on Reported Presence in a Virtual Environment,” *Presence Teleoperators Virtual Environ.*, vol. 19, no. 1, pp. 1–11, 2010, doi: 10.1162/pres.19.1.1.
- [17] K. Förster, A. Biasiucci, R. Chavarriaga, J. Del R. Millán, D. Roggen, and G. Tröster, “On the use of brain decoded signals for online user adaptive gesture recognition systems,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6030 LNCS, pp. 427–444, 2010, doi: 10.1007/978-3-642-12654-3_25.
- [18] M. Fabiani, G. Gratton, K. D. Federmeier, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, “Event-related brain potentials: Methods, theory, and applications,” in *Handbook of psychophysiology (3rd ed.)*, 2007, pp. 85–119.
- [19] S. J. Luck, “An Introduction to Event-Related Potentials and Their Neural Origins,” *An Introd. to event-related potential Tech.*, 2005, doi: 10.1007/s10409-008-0217-3.
- [20] D. Plass-Oude Bos *et al.*, “Brain-Computer Interfacing and Games,” in *Brain-*

Computer Interfaces, 2010, pp. 149–178.

- [21] R. Scherer, A. Schloegl, F. Lee, H. Bischof, J. Janša, and G. Pfurtscheller, “The self-paced graz brain-computer interface: Methods and applications,” *Comput. Intell. Neurosci.*, 2007, doi: 10.1155/2007/79826.
- [22] H. Shibasaki and M. Hallett, “What is the Bereitschaftspotential?,” *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2341–2356, 2006.
- [23] J. Perelmouter, B. Kotchoubey, A. Kubler, E. Taub, and N. Birbaumer, “Language support program for thought-translation-devices,” *Automedica*, vol. 18, no. 1, pp. 67–84, 1999.
- [24] P. W. Ferrez, “Error-Related EEG Potentials in Brain-Computer Interfaces,” *Techniques*, vol. 3928, pp. 291–301, 2007, doi: 10.5075/epfl-thesis-3928.
- [25] P. Martinez, H. Bakardjian, and A. Cichocki, “Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm,” *Comput. Intell. Neurosci.*, vol. 2007, no. i, p. 94561, 2007, doi: 10.1155/2007/94561.
- [26] A. T. Herdman, O. Lins, P. Van Roon, D. R. Stapells, M. Scherg, and T. W. Picton, “Intracerebral sources of human auditory steady-state responses,” *Brain Topogr.*, vol. 15, no. 2, pp. 69–86, 2002, doi: 10.1023/A:1021470822922.
- [27] G. R. Müller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller, “Steady-state somatosensory evoked potentials: Suitable brain signals for brain-computer interfaces?,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 1, pp. 30–37, 2006, doi: 10.1109/TNSRE.2005.863842.
- [28] G. Pfurtscheller *et al.*, “The hybrid BCI,” *Front. Neurosci.*, vol. 4, p. 30, 2010, doi: 10.3389/fnpro.2010.00003.
- [29] R. Leeb, H. Sagha, R. Chavarriaga, and J. D. R. Millán, “A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities,” *J. Neural Eng.*, vol. 8, no. 2, p. 025011, 2011, doi: 10.1088/1741-2560/8/2/025011.
- [30] A. Kreilinger, V. Kaiser, C. Breitwieser, C. Neuper, and G. Müller-Putz, “Fusion of manual control and brain-computer interface using long term and short term quality measures,” *Front. Neurosci.*, 2011.
- [31] T. O. T. Zander, M. Gaertner, C. Kothe, and R. Vilimek, “Combining Eye Gaze Input

- With a Brain–Computer Interface for Touchless Human–Computer Interaction,” *Int. J. Hum. Comput. Interact.*, vol. 27, no. 1, pp. 38–51, 2010, doi: 10.1080/10447318.2011.535752.
- [32] A. Holm *et al.*, “Estimating brain load from the EEG,” *ScientificWorldJournal.*, vol. 9, pp. 639–651, 2009, doi: 10.1100/tsw.2009.83.
- [33] B. S. Oken, M. C. Salinsky, and S. M. Elsas, “Vigilance, alertness, or sustained attention: physiological basis and measurement,” *Clin. Neurophysiol.*, vol. 117, no. 9, pp. 1885–1901, 2006, doi: 10.1016/j.clinph.2006.01.017.
- [34] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” *Int. J. Hum. Comput. Stud.*, vol. 67, no. 8, pp. 607–627, 2009, doi: 10.1016/j.ijhcs.2009.03.005.
- [35] S. J. Johnston, S. G. Boehm, D. Healy, R. Goebel, and D. E. J. Linden, “Neurofeedback: A promising tool for the self-regulation of emotion networks,” *Neuroimage*, vol. 49, no. 1, pp. 1066–1072, 2010, doi: 10.1016/j.neuroimage.2009.07.056.
- [36] C. Zickler *et al.*, “A brain-computer interface as input channel for a standard assistive technology software,” *Clin. EEG Neurosci.*, vol. 42, no. 4, pp. 236–244, 2011, doi: 10.1177/155005941104200409.
- [37] D. J. McFarland, C. W. Anderson, K. R. Müller, A. Schlögl, and D. J. Krusienski, “BCI Meeting 2005 - Workshop on BCI signal processing: Feature extraction and translation,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 135–138, 2006, doi: 10.1109/TNSRE.2006.875637.
- [38] F. Lotte, M. Congedo, L. Anatole, F. Lotte, M. Congedo, and L. Anatole, “A review of classification algorithms for EEG-based brain – computer interfaces To cite this version : A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces,” 2007.
- [39] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2013.
- [40] S. M. M. Martens, J. M. Mooij, N. J. Hill, J. Farquhar, and B. Schölkopf, “A Graphical Model Framework for Decoding in the Visual ERP-Based BCI Speller,” *Neural Comput.*, vol. 23, no. 1, pp. 160–182, Jan. 2011, doi: 10.1162/NECO_a_00066.

- [41] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.
- [42] R. T. Mina, A. Atiya, M. I. Owis, and Y. M. Kadah, "Brain-Computer Interface Based on Classification of Statistical and Power Spectral Density Features," *Biomed. Eng. (NY)*, pp. 2–5, 2006.
- [43] J. F. Delgado Saa and M. Sotaquir, "Eeg signal classification using arpower spectral features and linear discriminant analysis," *Proc. Lat. Am. Caribb. Consort. Eng. Institutions, Arequipa-Per*, 2010.
- [44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer-Verlag New York, 2009.
- [45] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008, doi: 10.1016/j.jneumeth.2007.07.017.
- [46] R. R. Hocking, "A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, no. 1, p. 1, Mar. 1976, doi: 10.2307/2529336.
- [47] N. R. Draper and H. Smith, *Applied Regression Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014.
- [48] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *J. Neurosci. Methods*, 2008, doi: 10.1016/j.jneumeth.2007.03.005.
- [49] N. V. Manyakov, N. Chumerin, A. Combaz, and M. M. Van Hulle, "Comparison of Classification Methods for P300 Brain-Computer Interface on Disabled Subjects," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–12, 2011, doi: 10.1155/2011/519868.
- [50] D. J. C. MacKay, "Bayesian Interpolation," *Neural Comput.*, 1992, doi: 10.1162/neco.1992.4.3.415.
- [51] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 1, pp. 24–29, 2006, doi: 10.1109/TNSRE.2005.862695.
- [52] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *J. Neural Eng.*, vol. 6, no. 4, p. 046011, 2009, doi: 10.1088/1741-

2560/6/4/046011.

- [53] M. Kaper, P. Meinicke, U. Grossekhoefer, T. Lingner, and H. Ritter, “BCI Competition 2003—Data Set IIB : Support Vector Machines for the P300 Speller Paradigm,” *Ieee Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1073–1076, 2004, doi: 10.1109/TBME.2004.826698.
- [54] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado, “Ensemble of SVMs for improving brain computer interface P300 speller performances,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3696 LNCS, pp. 45–50, 2005, doi: 10.1007/11550822_8.
- [55] A. Rakotomamonjy and V. Guigue, “BCI competition III: Dataset II- ensemble of SVMs for BCI P300 speller,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, 2008, doi: 10.1109/TBME.2008.915728.
- [56] M. Bentlemsan, E. T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, and K. Ferroudji, “Random forest and filter bank common spatial patterns for EEG-based motor imagery classification,” *Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS*, vol. 2015-Sept, pp. 235–238, 2015, doi: 10.1109/ISMS.2014.46.
- [57] E. Dagdevir and M. Tokmakci, “Optimization of preprocessing stage in EEG based BCI systems in terms of accuracy and timing cost,” *Biomed. Signal Process. Control*, vol. 67, no. January, p. 102548, 2021, doi: 10.1016/j.bspc.2021.102548.
- [58] K. Ferroudji, B. Yahya-Zoubir, M. Bentlemsan, and E. T. Zemouri, “Features selection using differential evolution in motor-imagery based brain machine interface,” *ACM Int. Conf. Proceeding Ser.*, vol. 23-25-Nove, pp. 0–5, 2015, doi: 10.1145/2816839.2816844.
- [59] J. Luo, Z. Feng, and N. Lu, “Spatio-temporal discrepancy feature for classification of motor imageries,” *Biomed. Signal Process. Control*, vol. 47, pp. 137–144, 2019, doi: 10.1016/j.bspc.2018.07.003.
- [60] J. Luo, X. Gao, X. Zhu, B. Wang, N. Lu, and J. Wang, “Motor imagery EEG classification based on ensemble support vector learning,” *Comput. Methods Programs Biomed.*, vol. 193, 2020, doi: 10.1016/j.cmpb.2020.105464.
- [61] J. Luo, J. Wang, R. Xu, and K. Xu, “Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification,” *J. Neurosci. Methods*, vol.

- 323, no. May, pp. 98–107, 2019, doi: 10.1016/j.jneumeth.2019.05.011.
- [62] M. K. I. Molla, A. Al Shiam, M. R. Islam, T. Tanaka, T. Tanaka, and T. Tanaka, “Discriminative Feature Selection-Based Motor Imagery Classification Using EEG Signal,” *IEEE Access*, vol. 8, pp. 98255–98265, 2020, doi: 10.1109/ACCESS.2020.2996685.
 - [63] X. Liao, D. Yao, D. Wu, and C. Li, “Combining spatial filters for the classification of single-trial EEG in a finger movement task,” *IEEE Trans. Biomed. Eng.*, 2007, doi: 10.1109/TBME.2006.889206.
 - [64] P. Shenoy, K. J. Miller, J. G. Ojemann, and R. P. N. Rao, “Finger movement classification for an electrocorticographic BCI,” *Proc. 3rd Int. IEEE EMBS Conf. Neural Eng.*, pp. 192–195, 2007, doi: 10.1109/CNE.2007.369644.
 - [65] T. Wissel *et al.*, “Hidden Markov model and support vector machine based decoding of finger movements using electrocorticography,” *J. Neural Eng.*, 2013, doi: 10.1088/1741-2560/10/5/056020.
 - [66] K. Liao, R. Xiao, J. Gonzalez, and L. Ding, “Decoding Individual Finger Movements from One Hand Using Human EEG Signals,” *PLoS One*, vol. 9, no. 1, p. e85192, Jan. 2014, doi: 10.1371/journal.pone.0085192.
 - [67] R. Tomioka, K. Aihara, and K.-R. Müller, “Logistic regression for single trial EEG classification,” *Analysis*, vol. 19, pp. 1377–1384, 2007.
 - [68] P. D. Prasad, H. N. Halahalli, J. P. John, and K. K. Majumdar, “Single-Trial EEG Classification Using Logistic Regression Based on Ensemble Synchronization,” *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 3, pp. 1074–1080, May 2014, doi: 10.1109/JBHI.2013.2289741.
 - [69] C. Sutton and A. McCallum, “An introduction to conditional random fields,” *Found. Trends Mach. Learn.*, vol. 4, no. 4, pp. 267–373, 2011, doi: 10.1561/22000000013.
 - [70] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: The MIT Press, 2009.
 - [71] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, “Hidden Markov models for online classification of single trial EEG data,” *Pattern Recognit. Lett.*, vol. 22, no. 12, pp. 1299–1309, 2001, doi: 10.1016/S0167-8655(01)00075-7.
 - [72] C. Sutton and A. McCallum, “An Introduction to Conditional Random Fields,” *Found.*

- Trends Mach. Learn.*, vol. 4, no. 4, p. 90, 2010, doi: 10.1561/22000000013.
- [73] J. F. Delgado Saa, A. de Pestere, D. McFarland, and M. Çetin, “Word-level language modeling for P300 spellers based on discriminative graphical models,” *J. Neural Eng.*, vol. 12, no. 2, p. 026007, 2015, doi: 10.1088/1741-2560/12/2/026007.
 - [74] U. Orhan *et al.*, “Fusion with language models improves spelling accuracy for ERP-based brain computer interface spellers,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, vol. 100, no. 2, pp. 5774–5777, doi: 10.1109/IEMBS.2011.6091429.
 - [75] C. Ulas and M. Cetin, “The first Brain-Computer Interface utilizing a Turkish language model,” in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, Apr. 2013, pp. 1–4, doi: 10.1109/SIU.2013.6531174.
 - [76] A. Ö. Argunşah and M. Çetin, “AR-PCA-HMM approach for sensorimotor task classification in EEG-based brain-computer interfaces,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 113–116, 2010, doi: 10.1109/ICPR.2010.36.
 - [77] H. Il Suk and S. W. Lee, “Two-layer hidden Markov models for multi-class motor imagery classification,” *Proc. - Work. Brain Decod. Pattern Recognit. Challenges Neuroimaging, WBD 2010 - Conjunction with the International Conf. Pattern Recognition, ICPR 2010*, pp. 5–8, 2010, doi: 10.1109/WBD.2010.16.
 - [78] J. F. Delgado Saa and M. Cetin, “Modeling Differences in the Time-Frequency presentation of EEG Signals Through HMMs for classification of Imaginary Motor Tasks,” *Sabancı Univ.*, 2011.
 - [79] J. F. Delgado Saa and M. Çetin, “Hidden conditional random fields for classification of imaginary motor tasks from EEG data,” in *European Signal Processing Conference*, 2011, vol. 19, no. Eusipco, pp. 1377–1381.
 - [80] J. F. D. Saa and M. Çetin, “A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data,” *J. Neural Eng.*, vol. 9, no. 2, 2012, doi: 10.1088/1741-2560/9/2/026020.
 - [81] B. A. S. Hasan and J. Q. Gan, “Conditional random fields as classifiers for three-class motor-imagery brain–computer interfaces,” *J. Neural Eng.*, vol. 8, no. 2, p. 25013, 2011.
 - [82] J. F. Delgado Saa, A. De Pestere, and M. Cetin, “Asynchronous decoding of finger

- movements from ECoG signals using long-range dependencies conditional random fields,” *J. Neural Eng.*, vol. 13, no. 3, 2016, doi: 10.1088/1741-2560/13/3/036017.
- [83] S. Chiappa, N. Donckers, S. Bengio, and F. Vrins, “HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems,” *Eur. Symp. Artif. Neural Networks*, no. April, pp. 193–204, 2004, doi: 10.1017/CBO9781107415324.004.
- [84] L. P. Morency, A. Quattoni, and T. Darrell, “Latent-dynamic discriminative models for continuous gesture recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2007, doi: 10.1109/CVPR.2007.383299.
- [85] A. Quattoni, S. Wang, L. P. Morency, M. Collins, and T. Darrell, “Hidden conditional random fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1848–1853, 2007, doi: 10.1109/TPAMI.2007.1124.
- [86] L. Van Der Maaten, M. Welling, and L. K. Saul, “Hidden-Unit Conditional Random Fields,” in *JMLR Workshop and Conference Proceedings*, 2011, vol. 15, no. 1, pp. 479–488, [Online]. Available: <https://proceedings.mlr.press/v15/maaten11b.html>.
- [87] P. M. Clifford, “The international vocabulary of basic and general terms in metrology,” *Measurement*, vol. 3, no. 2, pp. 72–76, 1985, doi: 10.1016/0263-2241(85)90006-5.
- [88] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, “Characterization of four-class motor imagery EEG data for the BCI-competition 2005,” *J. Neural Eng.*, vol. 2, no. 4, pp. L14–L22, Dec. 2005, doi: 10.1088/1741-2560/2/4/L02.
- [89] X. Zhu, *Knowledge Discovery and Data Mining: Challenges and Realities: Challenges and Realities*. Information Science Reference, 2007.
- [90] M. J. Warrens, “Inequalities between multi-rater kappas,” *Adv. Data Anal. Classif.*, vol. 4, no. 4, pp. 271–286, 2010, doi: 10.1007/s11634-010-0073-4.
- [91] C. C. Burn and A. A. S. Weir, “Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers,” *Vet. J.*, vol. 188, no. 2, pp. 166–170, 2011, doi: 10.1016/j.tvjl.2010.04.021.
- [92] M. L. McHugh, “Lessons in biostatistics Interrater reliability : the kappa statistic,” *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, 2012, doi: 10.11613/BM.2012.031.
- [93] H. Carrillo, K. H. Brodersen, and J. A. Castellanos, “Probabilistic performance evaluation for multiclass classification using the posterior balanced accuracy,” *Adv.*

- Intell. Syst. Comput.*, vol. 252, pp. 347–361, 2014, doi: 10.1007/978-3-319-03413-3_25.
- [94] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, “The balanced accuracy and its posterior distribution,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 3121–3124, 2010, doi: 10.1109/ICPR.2010.764.
 - [95] J. Benesty, J. Chen, and Y. Huang, “On the importance of the pearson correlation coefficient in noise reduction,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 16, no. 4, pp. 757–765, 2008, doi: 10.1109/TASL.2008.919072.
 - [96] J. Benesty, J. Chen, Y. Huang, and I. Cohen, “Optimal filters in the time domain,” in *Springer Topics in Signal Processing*, vol. 2, 2009, pp. 1–18.
 - [97] S. Fletcher and M. Z. Isla, “Comparing sets of patterns with the Jaccard index,” *Australas. J. Inf. Syst.*, vol. 22, pp. 1–17, 2018, doi: 10.3127/ajis.v22i0.1538.
 - [98] S.-H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *City*, vol. 1, no. 2, p. 1, 2007.
 - [99] Scholar Google, “Brain Computer Interface - 2016 search,” 2016. .
 - [100] PubMed, “Brain Computer Interface - 2016 Search,” 2016. .
 - [101] SCOPUS, “Brain Computer Interface - 2016 Search,” 2016. .
 - [102] École Polytechnique Fédérale de Laussane, “The Blue Brain Project,” 2016. .
 - [103] US Department of Heath & Human Services and National Institutes of Health, “The Brain Initiative,” 2016. .
 - [104] Institut Català de Nanociència i Nanotecnologia, “BrainCom will develop a new generation of cortical implants for speech neural prostheses applications,” 2017. .
 - [105] Berlin Brain Computer Interface, “Berlin Brain Computer Interface,” 2014. .
 - [106] Graz University of Technology, “BNCI Horizon 2020. The Future of Brain/Neural Computer Interaction: Horizon 2020,” 2016. .
 - [107] Institute of Neural Engineering at Gratz University of Technology, “INE Home Page,” 2016. .
 - [108] The University of Sydney, “Penso and Brain Computer Interface,” 2015. .
 - [109] The University of British Columbia, “The Brain-Computer Interface Project,” 2016. .
 - [110] Aalborg University, “EEG/BCI Laboratory,” 2016. .
 - [111] TOBI: Tools for Brain Computer Interaction, “Welcome to TOBI,” 2016. .

- [112] Department of Biomedical Engineering and Computational Science at Aalto University, “COSY - Centre of Excellence in Computational Complex Systems Research,” 2016. .
- [113] Inria Rennes, “OpenViBE,” 2015. .
- [114] RIKEN Science Brain Institute, “RIKEN Science Brain Institute,” 2016. .
- [115] School of Electrical Engineering and Computer Sciences at NUST, “Neuro Informatics Lab,” 2016. .
- [116] Department of Human Physiology at Moscow State University, “Laboratory for Neurophysiology and Neuro-Computer Interfaces (NNCI),” 2016. .
- [117] École Polytechnique Fédérale de Laussane, “Multimedia Signal Processing Group,” 2016. .
- [118] Sabancı University - Computer Vision and Pattern Analysis Laboratory, “TUBITAK BCI: Development of Electroencephalography (EEG) Signal Analysis Techniques for Brain Computer Interface (BCI) Systems,” 2016. .
- [119] University of Twente, “HMI - Human Media Interaction,” 2016. .
- [120] Essex BCI Group, “Essex BCI-NE Lab,” 2015. .
- [121] New York State Department of Health - Wadsworth Center, “Neuroscience & Neurotechnology,” 2016. .
- [122] S. S. Fox and J. H. O’Brien, “Duplication of Evoked Potential Waveform by Curve of Probability of Firing of a Single Cell,” *Science* (80-.), vol. 147, no. 3660, pp. 888–890, Feb. 1965, doi: 10.1126/science.147.3660.888.
- [123] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*. 2002, doi: 10.1016/S1388-2457(02)00057-3.
- [124] Y. Li, X. Gao, H. Liu, and S. Gao, “Classification of single-trial electroencephalogram during finger movement,” *IEEE Trans. Biomed. Eng.*, 2004, doi: 10.1109/TBME.2004.826688.
- [125] J. Kubánek, K. J. Miller, J. G. Ojemann, J. R. Wolpaw, and G. Schalk, “Decoding flexion of individual fingers using electrocorticographic signals in humans,” *J. Neural Eng.*, 2009, doi: 10.1088/1741-2560/6/6/066001.
- [126] L. A. A. Farwell and E. Donchin, “Talking off the top of your head: toward a mental

- prosthesis utilizing event-related brain potentials,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, Dec. 1988, doi: 10.1016/0013-4694(88)90149-6.
- [127] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, “A Note on Brain Actuated Spelling with the Berlin Brain-Computer Interface,” *Int. Conf. Univers. Access Human-Computer Interact.*, vol. 2007, pp. 759–768, 2007, doi: 10.1007/978-3-540-73281-5_83.
- [128] M. S. Treder and B. Blankertz, “(C)overt attention and visual speller design in an ERP-based brain-computer interface,” *Behav. Brain Funct.*, vol. 6, p. 28, 2010, doi: 10.1186/1744-9081-6-28.
- [129] T. Kaufmann, S. M. Schulz, C. Grünzinger, and a Kübler, “Flashing characters with famous faces improves ERP-based brain–computer interface performance,” *J. Neural Eng.*, vol. 8, no. 5, p. 056016, 2011, doi: 10.1088/1741-2560/8/5/056016.
- [130] W. Speier, A. Deshpande, L. Cui, N. Chandravadia, D. Roberts, and N. Pouratian, “A comparison of stimulus types in online classification of the P300 speller using language models,” *PLoS One*, vol. 12, no. 4, p. e0175382, 2017, doi: 10.1371/journal.pone.0175382.
- [131] J. Polich, “Updating P300: An integrative theory of P3a and P3b,” *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007, doi: 10.1016/j.clinph.2007.04.019.
- [132] W. Speier, C. Arnold, J. Lu, A. Deshpande, and N. Pouratian, “Integrating Language Information With a Hidden Markov Model to Improve Communication Rate in the P300 Speller,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 678–684, May 2014, doi: 10.1109/TNSRE.2014.2300091.
- [133] R. K. Chaurasiya, N. D. Londhe, and S. Ghosh, “An efficient P300 speller system for Brain-Computer Interface,” in *2015 International Conference on Signal Processing, Computing and Control (ISPCC)*, Sep. 2015, pp. 57–62, doi: 10.1109/ISPCC.2015.7374998.
- [134] M. De Vos, M. Kroesen, R. Emkes, and S. Debener, “P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier,” *J. Neural Eng.*, vol. 11, no. 3, p. 036008, 2014, doi: 10.1088/1741-2560/11/3/036008.
- [135] 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 Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1057–1061, 2004, doi: 10.1109/TBME.2004.826702.
- [136] Yang Liu, Zongtan Zhou, Dewen Hu, and Guohua Dong, “T-weighted Approach for Neural Information Processing in P300 based Brain-Computer Interface,” in *2005 International Conference on Neural Networks and Brain*, 2005, vol. 3, pp. 1535–1539, doi: 10.1109/ICNNB.2005.1614924.
- [137] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “BCI2000: A General-Purpose Brain-Computer Interface (BCI) System,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004, doi: 10.1109/TBME.2004.827072.
- [138] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, “BCI competition 2003 - Data set IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1067–1072, 2004, doi: 10.1109/TBME.2004.826699.
- [139] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter, “Improving Transfer Rates in Brain Computer Interfacing: A Case Study,” in *Advances in Neural Information Processing Systems*, 2003, no. 15, pp. 1107–1114.
- [140] D. J. Krusienski *et al.*, “A comparison of classification techniques for the P300 Speller,” *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006, doi: 10.1088/1741-2560/3/4/007.
- [141] U. Hoffmann *et al.*, “A Boosting Approach to P300 Detection with Application to Brain-Computer Interfaces,” in *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.*, 2005, vol. 2005, pp. 97–100, doi: 10.1109/CNE.2005.1419562.
- [142] Xiaofeng Shi, Guoqiang Xu, Furao Shen, and Jinxi Zhao, “Solving the data imbalance problem of P300 detection via Random Under-Sampling Bagging SVMs,” in *2015 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2015, vol. 2015-Septe, pp. 1–5, doi: 10.1109/IJCNN.2015.7280834.
- [143] R. Tibon and D. A. Levy, “Striking a balance: Analyzing unbalanced event-related potential data,” *Front. Psychol.*, vol. 6, no. MAY, pp. 1–4, 2015, doi:

10.3389/fpsyg.2015.00555.

- [144] F. J. Valverde-Albacete and C. Peláez-Moreno, “100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox,” *PLoS One*, vol. 9, no. 1, 2014, doi: 10.1371/journal.pone.0084217.
- [145] J. Farquhar and N. J. Hill, “Interactions between pre-processing and classification methods for event-related-potential classification: Best-practice guidelines for brain-computer interfacing,” *Neuroinformatics*, vol. 11, no. 2, pp. 175–192, 2013, doi: 10.1007/s12021-012-9171-0.
- [146] A. Schögl, J. Kronegg, J. E. Huggins, and S. G. Mason, “Evaluation Criteria for BCI Research,” *Toward. Brain-computer Interfacing*, pp. 327–342, 2007.
- [147] A. E. Society, “American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature,” *J. Clin. Neurophysiol.*, vol. 8, no. November, pp. 200–202, 1991, doi: 10.1097/00004691-199104000-00007.
- [148] New York State Department of Health - Wadsworth Center and B. Blankertz, “BCI Competition III,” 2004. <https://www.bbci.de/competition/iii/>.
- [149] L. Batterink, C. M. Karns, and H. Neville, “Dissociable mechanisms supporting awareness: The P300 and gamma in a linguistic attentional blink task,” *Cereb. Cortex*, vol. 22, no. 12, pp. 2733–2744, 2012, doi: 10.1093/cercor/bhr346.
- [150] S. K. Thompson, *Sampling*, 3rd ed. Wiley, 2012.
- [151] T. M. Mitchell, *Machine Learning*. WCB / McGraw – Hill, 1997.
- [152] X. Zhu and I. Davidson, *Knowledge discovery and data mining: Challenges and realities*. 2007.
- [153] K. J. Miller, J. G. Ojemann, and J. M. Henderson, “Instantaneous interactions between brain sites can distinguish movement from rest but are relatively poor at resolving different movement types,” 2014, doi: 10.1109/EMBC.2014.6944797.
- [154] J. B. Salyers, Y. Dong, and Y. Gai, “Continuous Wavelet Transform for Decoding Finger Movements From Single-Channel EEG,” *IEEE Trans. Biomed. Eng.*, 2019, doi: 10.1109/TBME.2018.2876068.
- [155] G. Hotson *et al.*, “Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject,” *J. Neural Eng.*, 2016, doi: 10.1088/1741-2560/13/2/026017.

- [156] T. Pfeiffer, R. T. Knight, and G. Rose, “Hidden Markov model based continuous decoding of finger movements with prior knowledge incorporation using bi-gram models,” *Biomed. Phys. Eng. Express*, 2018, doi: 10.1088/2057-1976/aa99f3.
- [157] J. F. Delgado Saa, A. De Pesters, and M. Cetin, “Asynchronous decoding of finger movements from ECoG signals using long-range dependencies conditional random fields,” *J. Neural Eng.*, 2016, doi: 10.1088/1741-2560/13/3/036017.
- [158] S. M. M. Martens, J. M. Mooij, N. J. Hill, J. Farquhar, and B. Schölkopf, “A graphical model framework for decoding in the visual ERP-based BCI speller,” *Neural Computation*. 2011, doi: 10.1162/NECO_a_00066.
- [159] K. J. Miller *et al.*, “Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms,” *PLoS Comput. Biol.*, 2012, doi: 10.1371/journal.pcbi.1002655.
- [160] K. J. Miller, “A library of human electrocorticographic data and analyses,” *Stanford Digital Repository*, 2016. <https://exhibits.stanford.edu/data/catalog/zk881ps0522>.
- [161] J. Delgado Saa, A. Christen, S. Martin, B. N. Pasley, R. T. Knight, and A. L. Giraud, “Using Coherence-based spectro-spatial filters for stimulus features prediction from electro-corticographic recordings,” *Sci. Rep.*, 2020, doi: 10.1038/s41598-020-63303-1.
- [162] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, “Spatial patterns underlying population differences in the background EEG,” *Brain Topogr.*, vol. 2, no. 4, pp. 275–284, 1990, doi: 10.1007/BF01129656.
- [163] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 2390–2397, 2008, doi: 10.1109/IJCNN.2008.4634130.
- [164] M. Jalilpour Monesi and S. Hajipour Sardouie, “Extended common spatial and temporal pattern (ECSTP): A semi-blind approach to extract features in ERP detection,” *Pattern Recognit.*, vol. 95, pp. 128–135, 2019, doi: 10.1016/j.patcog.2019.05.039.
- [165] V. P. Oikonomou, S. Nikolopoulos, and I. Kompatsiaris, “Robust motor imagery classification using sparse representations and grouping structures,” *IEEE Access*, vol. 8, pp. 98572–98583, 2020, doi: 10.1109/ACCESS.2020.2997116.
- [166] S. Kumar, A. Sharma, and T. Tsunoda, “An improved discriminative filter bank

- selection approach for motor imagery EEG signal classification using mutual information,” *BMC Bioinformatics*, vol. 18, no. Suppl 16, 2017, doi: 10.1186/s12859-017-1964-6.
- [167] K. P. Thomas, C. T. Lau, A. P. Vinod, C. Guan, and K. K. Ang, “A New Discriminative Common Spatial Pattern Method for Motor Imagery Brain—Computer Interfaces,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, 2009, doi: 10.1109/TBME.2009.2026181.
- [168] J. C. Bezdek and N. R. Pal, “An index of topological preservation for feature extraction,” *Pattern Recognit.*, vol. 28, no. 3, pp. 381–391, Mar. 1995, doi: 10.1016/0031-3203(94)00111-X.
- [169] M. Tangermann *et al.*, “Review of the BCI competition IV,” *Front. Neurosci.*, vol. 6, no. JULY, pp. 1–31, 2012, doi: 10.3389/fnins.2012.00055.
- [170] European Network of Excellence PASCAL2 Institute for knowledge discovery, “BCI Competition IV,” 2008. <https://www.bbci.de/competition/iv/> (accessed Feb. 01, 2022).
- [171] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, “Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, 2007, doi: 10.1109/TNSRE.2007.906956.
- [172] M. Li, H. Xi, and X. Zhu, “An Incremental Version of L-MVU for the Feature Extraction of MI-EEG,” *Comput. Intell. Neurosci.*, vol. 2019, 2019, doi: 10.1155/2019/4317078.
- [173] G. Pfurtscheller and D. J. McFarland, “BCIs That Use Sensorimotor Rhythms,” in *Brain–Computer Interfaces Principles and Practice*, vol. 15, no. 1, Oxford University Press, 2012, pp. 228–240.
- [174] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, “Processing strategies for time-course data sets in functional mri of the human brain,” *Magn. Reson. Med.*, vol. 30, no. 2, pp. 161–173, 1993, doi: 10.1002/mrm.1910300204.
- [175] Y. BenYahmed, A. Abu Bakar, A. RazakHamdan, A. Ahmed, and S. M. S. Abdullah, “Adaptive sliding window algorithm for weather data segmentation,” *J. Theor. Appl. Inf. Technol.*, vol. 80, no. 2, pp. 322–333, 2015.

- [176] H. S. Hota, R. Handa, and A. K. Shrivastava, "Time Series Data Prediction Using Sliding Window Based RBF Neural Network," *Int. J. Comput. Intell. Res.*, vol. 13, no. 5, pp. 1145–1156, 2017, [Online]. Available: <http://www.ripublication.com>.
- [177] F. Mokhtari, M. I. Akhlaghi, S. L. Simpson, G. Wu, and P. J. Laurienti, "Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state," *Neuroimage*, vol. 189, no. November 2018, pp. 655–666, 2019, doi: 10.1016/j.neuroimage.2019.02.001.
- [178] V. M. Vergara, A. Abrol, and V. D. Calhoun, "An average sliding window correlation method for dynamic functional connectivity," *Hum. Brain Mapp.*, vol. 40, no. 7, pp. 2089–2103, 2019, doi: 10.1002/hbm.24509.
- [179] A. Villalba, J. L. Berral, and D. Carrera, "Constant-Time Sliding Window Framework with Reduced Memory Footprint and Efficient Bulk Evictions," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 3, pp. 486–500, 2019, doi: 10.1109/TPDS.2018.2868960.
- [180] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2006.
- [181] H. Zenil, N. A. Kiani, and J. Tegnér, "Quantifying loss of information in network-based dimensionality reduction techniques," *J. Complex Networks*, vol. 4, no. 3, pp. 342–362, 2016, doi: 10.1093/comnet/cnv025.
- [182] A. Gracia, S. González, V. Robles, and E. Menasalvas, "A methodology to compare Dimensionality Reduction algorithms in terms of loss of quality," *Inf. Sci. (Ny)*, vol. 270, pp. 1–27, 2014, doi: 10.1016/j.ins.2014.02.068.
- [183] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 957–964, 2007, doi: 10.1016/j.patrec.2007.01.002.
- [184] Y. Song, L. P. Morency, and R. Davis, "Multi-view latent variable discriminative models for action recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2120–2127, 2012, doi: 10.1109/CVPR.2012.6247918.