



Politecnico  
di Bari



SAPIENZA  
UNIVERSITÀ DI ROMA

ELECTRICAL AND INFORMATION ENGINEERING

Final Dissertation

---

# Accessible and Explainable AI for EEG Decoding in Brain-Computer Interfaces

---

by

Tommaso Colafiglio

*Supervisor*

Prof. Tommaso Di Noia

*Co-Supervisor*

Prof.ssa Angela Lombardi

*Coordinator of the PhD Program*

Prof. Giorgio Grisetti

---

Course XXXVIII, 01/11/2022 - 31/10/2025



Science is Music that thinks; Music is Science that breathes. Dedicated Anna, Vincenzo, and my beloved Mariagrazia, whose presence and deep sensitivity have inspired me in my continuous exploration of new and infinite frontiers.



# Abstract

The field of brain–computer interfaces (BCIs) is increasingly emerging as a highly relevant domain of research and experimentation, with growing implications for human well-being and daily activities. In this context, electroencephalography (EEG), a non-invasive measure of brain activity, is a crucial tool that enables both the investigation of neural mechanisms that remain largely unexplored and the development of interventions across multiple application areas, including motor rehabilitation, emotion regulation, and mental health.

In recent years, technological development has accelerated markedly, driven in particular by the rapid growth of artificial intelligence (AI) models and methodologies. The adoption of machine learning (ML) and deep learning (DL) techniques now enables the interpretation and decoding of highly complex patterns in EEG signals that are largely inaccessible through traditional analytical methods. At the same time, AI is increasingly recognized as a transformative force in healthcare, offering opportunities to improve diagnosis and clinical management and to develop innovative solutions in the BCI domain.

This doctoral thesis lies at the intersection of neuroscience and AI and is structured around two complementary research directions. First, it focuses on the development and integration of advanced EEG preprocessing techniques designed to operate in both offline analyses and real-time applications. Second, it investigates state-of-the-art ML and DL models for decoding and analyzing EEG patterns.

The contributions presented in this thesis include tools and frameworks that bridge the gap between AI research and its translation into real-world applications. By incorporating principles of explainable artificial intelligence (XAI), the proposed models are designed to be transparent and interpretable. Each contribution is presented in detail in its respective chapter, providing a comprehensive overview of the work conducted and its implications for artificial intelligence and computational neuroscience.



# Publications

A comprehensive list of my publications produced during the PhD is provided below. It should be noted that, in some instances, the order of authorship follows an alphabetical arrangement.

- [1] Carmelo Ardito, Ilaria Bortone, Tommaso Colafiglio, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Fedelucio Narducci, Rodolfo Sardone, and Paolo Sorino. “Brain computer interface: Deep learning approach to predict human emotion recognition.” In: *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2022, pp. 2689–2694.
- [2] Carmelo Ardito, Tommaso Colafiglio, Tommaso Di Noia, Angela Lombardi, Domenico Lofù, Fedelucio Narducci, and Paolo Sorino. “Towards a Neurofeedback Tool For Emotion Recognition Using Brain Computer Interface.” In: *WALS24 - The 3rd International Workshop on Web Applications for Life Sciences - In conjunction with the 24th International Conference on Web Engineering (ICWE 2024)* (2024).
- [3] Tommaso Colafiglio, Carmelo Ardito, Paolo Sorino, Domenico Lofù, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms.” In: *Cognitive Computation* (2024), pp. 1–24.
- [4] Tommaso Colafiglio, Tommaso Di Noia, Domenico Lofù, Angela Lombardi, Fedelucio Narducci, and Paolo Sorino. “Wearable Devices and Brain-Computer Interfaces for User Modelling (WeBIUM).” In: *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 2024, pp. 597–600.
- [5] Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “Exploring the Mental State Intersection by Brain-Computer Interfaces, Cellular Automata and Biofeedback.” In: *IEEE EUROCON 2023-20th International Conference on Smart Technologies*. IEEE. 2023, pp. 461–466.

- [6] Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. “Neural Musical Instruments through Brain-Computer Interface and Biofeedback.” In: *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 2025, pp. 489–494.
- [7] Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Angela Lombardi, Fedelucio Narducci, Fabrizio Festa, and Tommaso Di Noia. “EmoSynth Real Time Emotion-Driven Sound Texture Synthesis via Brain-Computer Interface.” In: *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 2024, pp. 616–621.
- [8] Tommaso Colafiglio, Angela Lombardi, Tommaso Di Noia, Maria Luigia Natalia De Bonis, Fedelucio Narducci, and Alice Mado Proverbio. “Machine learning classification of motivational states: Insights from EEG analysis of perception and imagery.” In: *Expert Systems with Applications* 275 (2025), p. 127076.
- [9] Tommaso Colafiglio, Angela Lombardi, Paolo Sorino, Elvira Brattico, Domenico Lofù, Danilo Danese, Eugenio Di Sciascio, Tommaso Di Noia, and Fedelucio Narducci. “NeuroSense: A Novel EEG Dataset Utilizing Low-Cost, Sparse Electrode Devices for Emotion Exploration.” In: *IEEE Access* (2024), pp. 1–1. DOI: 10.1109/ACCESS.2024.3487932.
- [10] Tommaso Colafiglio, Angela Lombardi, Paolo Sorino, Domenico Lofù, Danilo Danese, Fedelucio Narducci, Eugenio Di Sciascio, and Tommaso Di Noia. “Explainable Evaluation of Emotion Recognition with Low-Cost EEG: Feature Engineering and Interpretability Insights.” In: (2025).
- [11] Tommaso Colafiglio, Paolo Sorino, Domenico Lofu, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. “Combining Mental States Recognition and Machine Learning for Neurorehabilitation.” In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2023, pp. 3848–3853.
- [12] Tommaso Colafiglio, Paolo Sorino, Angela Lombardi, Domenico Lofù, T Di Noia, et al. “Predicting Human Emotions using EEG-based Brain computer Interface and Interpretable Machine Learning.” In: *CEUR WORKSHOP PROCEEDINGS*. Vol. 3486. CEUR-WS. 2023, pp. 200–205.
- [13] Domenico Lofù, Paolo Sorino, Tommaso Colafiglio, Caterina Bonfiglio, Rossella Donghia, Gianluigi Giannelli, Angela Lombardi, Tommaso Di Noia, Eugenio Di Sciascio, and Fedelucio Narducci. “MORIX: Machine learning-aided framework for lethality detection and MORTality inference with eXplainable artificial intelligence in MAFLD subjects.” In: *Computer Methods and Programs in Biomedicine Update* 7 (2025), p. 100176.

- [14] Domenico Lofù, Paolo Sorino, Tommaso Colafiglio, Caterina Bonfiglio, Fedelucio Narducci, Tommaso Di Noia, and Eugenio Di Sciascio. “MAFUS: a framework to predict mortality risk in MAFLD subjects.” In: *arXiv preprint arXiv:2301.06908* (2023).
- [15] Domenico Lofù, Paolo Sorino, Tommaso Colafiglio, Angela Lombardi, Tommaso Di Noia, and Fedelucio Narducci. “WeBIUM 2025: 2nd Workshop on Wearable Devices and Brain-Computer Interfaces for User Modelling.” In: *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 2025, pp. 467–470.
- [16] Angela Lombardi, Maria Luigia Natalia De Bonis, Giuseppe Fasano, Alessia Sportelli, Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Fedelucio Narducci, Eugenio Di Sciascio, Tommaso Di Noia, et al. “Time-to-event interpretable machine learning for multiple sclerosis worsening prediction: results from iDPP@ CLEF 2023.” In: *CLEF*. 2023.
- [17] Paolo Sorino, Giovanni Maria Biancofiore, Domenico Lofù, Tommaso Colafiglio, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. “ARIEL: Brain-Computer Interfaces meet Large Language Models for Emotional Support Conversation.” In: *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 2024, pp. 601–609.
- [18] Paolo Sorino, Gianluca Colonna, Domenico Lofù, Tommaso Colafiglio, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. “An Explainable Machine Learning Approach for Heartbeat Classification through Signal-based Features.” In: *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2024.
- [19] Paolo Sorino, Angela Lombardi, Domenico Lofù, Tommaso Colafiglio, Antonio Ferrara, Fedelucio Narducci, Eugenio Di Sciascio, and Tommaso Di Noia. “Detecting label noise in longitudinal Alzheimer’s data with explainable artificial intelligence.” In: *Brain Informatics* 12.1 (2025), p. 15.
- [20] Paolo Sorino, Vincenzo Paparella, Domenico Lofu, Tommaso Colafiglio, Eugenio Di Sciascio, Fedelucio Narducci, Rodolfo Sardone, and Tommaso Di Noia. “A Pareto-Optimality-based approach for selecting the best Machine Learning models in mild cognitive impairment prediction.” In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2023, pp. 3822–3827.



# Table of contents

<b>List of figures</b>	<b>xxi</b>
<b>List of tables</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.0.1 Thesis Statement . . . . .	1
1.0.2 Structure of the Thesis . . . . .	2
<b>2 Foundations of Electroencephalography</b>	<b>5</b>
2.1 Background . . . . .	5
2.1.1 Biophysical Basis, Electrode Montages . . . . .	5
2.1.2 Electrode positioning and standard montages . . . . .	6
2.1.3 Physiological EEG . . . . .	7
2.2 Preprocessing and Signal Quality . . . . .	8
2.2.1 Artifacts Correction . . . . .	8
2.2.2 Challenges in BCI Signal Preprocessing . . . . .	11
2.3 EEG Feature Extraction . . . . .	12
2.3.1 Univariate and Multivariate EEG Descriptors . . . . .	12
2.4 Brain Computer Interface Active Paradigms . . . . .	12
2.4.1 Motor Imagery Paradigms . . . . .	13
2.4.2 P300 Based Paradigms . . . . .	13
2.4.3 SSVEP Based Paradigms . . . . .	13
2.5 Brain Computer Interface Passive Paradigm . . . . .	14
2.5.1 Emotion Recognition . . . . .	14
<b>3 Artificial Intelligence for Brain Computer Interfaces</b>	<b>17</b>
3.1 Deep Learning Models for BCI . . . . .	20
3.1.1 Taxonomy of Deep Models in BCI . . . . .	21
3.1.2 Discriminative Models . . . . .	22

3.1.3	Unsupervised Models . . . . .	22
3.1.4	Generative Models . . . . .	23
3.1.5	Hybrid Architectures . . . . .	23
3.1.6	EEG Foundation Models . . . . .	23
3.1.7	Explainable BCI Decision Systems . . . . .	25
<b>4</b>	<b>EEG Based Emotion Recognition</b>	<b>27</b>
4.1	Introduction . . . . .	27
4.1.1	Problem Formulation and Aims . . . . .	27
4.2	NeuroSense: Low Cost sparse electrode Dataset . . . . .	28
4.2.1	Introduction . . . . .	28
4.2.2	Related Work . . . . .	29
4.2.3	Stimuli Selection . . . . .	30
4.2.4	Experimental Setup . . . . .	31
4.2.5	System Architecture . . . . .	33
4.2.6	Key Findings and Experiments . . . . .	37
4.2.7	Implications and Future Prospects . . . . .	40
4.3	Explainable Emotion Recognition with Low Cost EEG: Feature Engineering and Model Attributions . . . . .	40
4.3.1	Introduction . . . . .	40
4.3.2	Related Work . . . . .	41
4.3.3	Dataset, Protocol, and Preprocessing . . . . .	41
4.3.4	Feature Engineering and Classifiers . . . . .	42
4.3.5	Results and Discussion . . . . .	43
4.3.6	Conclusions and Future Work . . . . .	46
4.4	Deep Learning Applied in Human Emotion Recognition . . . . .	47
4.4.1	Introduction . . . . .	47
4.4.2	Related Work . . . . .	47
4.4.3	Materials and Methods . . . . .	48
4.4.4	BCI Devices . . . . .	48
4.4.5	Dataset . . . . .	49
4.4.6	Preprocessing . . . . .	49
4.4.7	Feature Set and Training Data . . . . .	50
4.4.8	Model . . . . .	50
4.4.9	System Performance and Real Time Inference . . . . .	50
4.4.10	Prototype Evaluation and Implications . . . . .	54

4.5	ARIEL: Brain Computer Interfaces meet Large Language Models for Emotional Support Conversation . . . . .	54
4.5.1	Introduction . . . . .	54
4.5.2	Related Work . . . . .	55
4.5.3	Brain Computer Interface . . . . .	55
4.5.4	Emotional Support Conversation with Large Language Models .	56
4.5.5	ARIEL Framework . . . . .	57
4.5.6	Emotion Recognizer . . . . .	58
4.5.7	Prompt Formatter . . . . .	59
4.5.8	Controlled Example . . . . .	59
4.5.9	Discussion and Future Directions . . . . .	60
<b>5</b>	<b>EEG for Motivation and Neurorehabilitation</b>	<b>61</b>
5.1	Machine Learning Classification of Motivational States: Insights from EEG Analysis of Perception and Imagery . . . . .	61
5.1.1	Introduction . . . . .	61
5.1.2	Related Work . . . . .	62
5.1.3	Problem Statement and Research Questions . . . . .	63
5.1.4	Dataset, Participants, and Protocol . . . . .	63
5.1.5	Preprocessing and ERP Construction . . . . .	65
5.1.6	Electrode Sub selection and Configurations . . . . .	65
5.1.7	Modeling Strategy . . . . .	66
5.1.8	Results . . . . .	67
5.1.9	Discussion . . . . .	70
5.1.10	Limitations . . . . .	70
5.1.11	Conclusion . . . . .	71
5.2	Combining Mental States Recognition and Machine Learning for Neurorehabilitation . . . . .	71
5.2.1	Introduction . . . . .	71
5.2.2	Background and Related Work . . . . .	72
5.2.3	Materials and Methods . . . . .	73
5.2.4	Data Acquisition Device . . . . .	73
5.2.5	Polyphonic Structure Generation . . . . .	74
5.2.6	Framework . . . . .	77
5.2.7	Machine Learning Engine . . . . .	77
5.2.8	Results and Discussion . . . . .	79
5.2.9	Conclusion and Future Work . . . . .	80

<b>6</b>	<b>BCI for Neuro Musical Interfaces</b>	<b>81</b>
6.1	NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms . . . . .	81
6.1.1	Introduction . . . . .	81
6.1.2	Related Work . . . . .	82
6.1.3	Example Scenario . . . . .	83
6.1.4	NeuralPMG Framework Architecture . . . . .	85
6.1.5	Polyphony Generation Process . . . . .	86
6.1.6	NeuralPMG Engine . . . . .	89
6.1.7	Evaluation Study . . . . .	90
6.1.8	Polyphony Assessment by Music Experts . . . . .	94
6.1.9	Discussion . . . . .	95
6.2	Neural Musical Instruments through Brain Computer Interface and Biofeedback . . . . .	95
6.2.1	Introduction . . . . .	95
6.2.2	Background and Related Work . . . . .	97
6.2.3	Materials and Methods . . . . .	97
6.2.4	System Architecture . . . . .	98
6.2.5	Muse Headset . . . . .	98
6.2.6	Virtual Musical Instrument . . . . .	98
6.2.7	Interpretable Model . . . . .	99
6.2.8	Experimental Settings . . . . .	99
6.2.9	EEG Preprocessing Pipeline . . . . .	100
6.2.10	Machine Learning LOSO Pipeline . . . . .	101
6.2.11	Results . . . . .	101
6.2.12	Discussion and Conclusion . . . . .	103
6.3	EmoSynth Real Time Emotion Driven Sound Texture Synthesis via Brain Computer Interface . . . . .	105
6.3.1	Introduction . . . . .	105
6.3.2	Background and Related Works . . . . .	106
6.3.3	Materials and Methods . . . . .	107
6.3.4	System Architecture . . . . .	107
6.3.5	MUSE . . . . .	107
6.3.6	Experimental Setting . . . . .	108
6.3.7	Emotion Dataset Description . . . . .	108
6.3.8	VAE Dataset Description . . . . .	108

6.3.9	EEG Preprocessing . . . . .	109
6.3.10	ML Emotion Recognition . . . . .	110
6.3.11	Variational Autoencoder . . . . .	111
6.3.12	VAE Training . . . . .	113
6.3.13	Limitations . . . . .	113
6.3.14	Conclusion and Discussion . . . . .	114
6.4	Exploring the Mental State Intersection by Brain Computer Interfaces, Cellular Automata and Biofeedback . . . . .	114
6.4.1	Introduction . . . . .	114
6.4.2	Related Work . . . . .	115
6.4.3	Materials and Methods . . . . .	116
6.4.4	System Architecture . . . . .	116
6.4.5	EEG signal preprocessing strategies . . . . .	117
6.4.6	EEG signal management and processing . . . . .	117
6.4.7	Cellular Automata . . . . .	119
6.4.8	Additive synthesis . . . . .	122
6.4.9	Data Visualization . . . . .	122
6.4.10	Discussion and further directions . . . . .	123
<b>7</b>	<b>Other research contributions</b>	<b>125</b>
7.1	MORIX: An ML–XAI Framework to Predict Mortality in MAFLD . . .	125
7.1.1	Overview . . . . .	125
7.1.2	Related work . . . . .	126
7.1.3	Proposed framework . . . . .	126
7.1.4	Proposed method . . . . .	127
7.1.5	Preprocessing and selection. . . . .	127
7.1.6	Models and evaluation. . . . .	127
7.1.7	Experimental results . . . . .	127
7.1.8	Evaluation metrics . . . . .	131
7.1.9	Best model performance . . . . .	131
7.1.10	Statistical analysis . . . . .	132
7.1.11	Explainability results . . . . .	133
7.1.12	MORIX web user interface . . . . .	135
7.1.13	Conclusion and further directions . . . . .	138
7.2	Machine learning in Mild Cognitive Impairment Prediction . . . . .	139
7.2.1	Introduction . . . . .	139
7.2.2	Background on MCI . . . . .	139

7.2.3	Study Aim and Contributions . . . . .	139
7.2.4	Pareto Optimality and Hypervolume . . . . .	139
7.2.5	Dataset Description . . . . .	140
7.2.6	Classification Models . . . . .	141
7.2.7	Pareto Optimality Based Selection . . . . .	141
7.2.8	Results . . . . .	142
7.3	ECG Based Machine Learning for Early Heart Failure Detection with Explainable AI . . . . .	144
7.3.1	Intoduction . . . . .	144
7.3.2	Related Work . . . . .	145
7.3.3	Data Source . . . . .	145
7.3.4	Methods . . . . .	145
7.3.5	Preprocessing Pipeline . . . . .	145
7.3.6	Beat Centered Subsampling . . . . .	146
7.3.7	Feature Engineering and Selection . . . . .	146
7.3.8	Classification Models and Tuning . . . . .	147
7.3.9	Explainability . . . . .	147
7.3.10	ECG Model Analysis . . . . .	148
7.3.11	Analysis Outcomes and Clinical Implications . . . . .	150
7.4	Time-to-event interpretable machine learning for multiple sclerosis worsening prediction: results from iDPP@ CLEF 2023 . . . . .	151
7.4.1	Introduction . . . . .	151
7.4.2	Related work . . . . .	152
7.4.3	Methodology . . . . .	152
7.4.4	Dataset . . . . .	153
7.4.5	Validation scheme . . . . .	154
7.4.6	Time to event learning . . . . .	154
7.4.7	Permutation feature importance . . . . .	155
7.4.8	Performance evaluation . . . . .	156
7.4.9	Experimental setup . . . . .	156
7.4.10	Results . . . . .	157
7.4.11	Conclusions and future work . . . . .	161

# List of figures

2.1	Synaptic integration and the analog digital transition. Excitatory (+) and inhibitory (−) presynaptic inputs generate EPSPs and IPSPs that sum over dendrites and soma. Integration at the axon initial segment implements a threshold nonlinearity: if exceeded, a digital action potential is generated. . . . .	7
2.2	From cortical dipoles to scalp potentials. (Left) Multiple synchronised cortical patches (conceptually equivalent to dipoles) project through layered tissues to scalp electrodes; (bottom) tissue stratification (cortex, cerebrospinal fluid, skull, scalp) imposes strong attenuation and spatial blurring; (right) Schematic distributions of EPSP/IPSP along apical dendrites producing net dipoles in the open field. [243] . . . . .	8
2.3	International 10–20 system and three standard clinical montages emphasizing different source projections. Panels (a)–(c) correspond to the sketches commonly reported in the clinical EEG atlas. [17, 41] . . . . .	9
2.4	Raw Electroencephalogram (EEG) from electrodes AF3, TP9, TP10, and AF4. The signal is affected by ocular, muscle, and slow drift artifacts.	10
2.5	Preprocessed EEG signals from the same electrodes after the full pre-processing pipeline. Noise and artifacts are significantly reduced. . . . .	10
4.1	Stimulus selection workflow: (1) affect tagged and curated pool; (2) one minute highlight detection (DEAP style); (3) video/audio feature extraction; (4) RVM based arousal/valence prediction and scoring; (5) final selection via online subjective annotation. . . . .	31
4.2	Self Assessment Manikins for valence, arousal, and dominance, and a 3 option liking scale. Ratings were entered on continuous 5 point visual analog scales. . . . .	33

4.3	Workflow for EEG based emotion recognition: acquisition, data creation, preprocessing, time sliding Region of Interest (ROI) estimation, Machine Learning (ML) classification, and statistical validation. . . . .	34
4.4	Horizontal bar plots are resulting from the STD analysis for each participant. . . . .	37
4.5	Self assessed <i>valence</i> (participants on the x axis; video IDs on the y axis).	38
4.6	Self assessed <i>arousal</i> (participants on the x axis; video IDs on the y axis).	38
4.7	Self assessed <i>dominance</i> (participants on the x axis; video IDs on the y axis). . . . .	38
4.8	Correlations between external labels and self assessments across videos.	39
4.9	Boxplots of pairwise SHAP value correlations across subjects for each emotional state. Each boxplot summarizes the distribution of Pearson correlation coefficients between SHAP vectors from different participants. Lower values reflect greater inter subject variability in feature importance.	44
4.10	Top 10 SHAP based features for each emotional state: (a) Excited, (b) Relaxed, (c) Sad, (d) Angry. Each panel shows the most relevant EEG features ranked by mean SHAP values across all subjects. . . . .	45
4.11	Spatial distribution of SHAP based feature importance across emotional states: (a) Excited, (b) Relaxed, (c) Sad, (d) Angry. Each panel highlights the most relevant EEG features and their associated electrode locations. . . . .	46
4.12	Emotiv Insight 5 electrodes (10–20 montage). . . . .	49
4.13	Valence error distribution. . . . .	51
4.14	Arousal error distribution. . . . .	51
4.15	Dominance error distribution. . . . .	52
4.16	Training loss across epochs. . . . .	52
4.17	User interface for real-time biofeedback. . . . .	53
4.18	System architecture overview. . . . .	53
4.19	ARIEL workflow. The Neuro Linguistic Interface (a) collects EEG and text; the Emotion Recognizer (b) infers an emotion label from EEG streams; the Prompt Formatter (c) merges the recognized label with the user’s message, constructing a conversation state prompt; the Large Language Model (LLM) (d) generates a supportive response. The session ends when a positive state is reached. . . . .	57

4.20	Example conversation with ARIEL. The dialogue begins in <i>Sad</i> , then gradually shifts to a positive affect as measured by the BCI pipeline (red balloons), with text exchanges (grey/blue) guided by dynamic prompts.	60
5.1	Schematic of the perception–imagery protocol (adapted from [270], CC BY).	64
5.2	14 channel Emotiv channels subset (10–20 system). Black circled electrodes indicate CMS/DRL reference sites.	66
5.3	18 channel variant with added midline Fz, Cz, Pz, Oz (hypothetical extension). Black circled electrodes indicate CMS/DRL.	66
5.4	Distributions of F1 scores across all pairwise classifiers under perception/imagery and 14/18 channel configurations.	68
5.5	F1 14 channels imagery	68
5.6	F1 18 channels imagery	69
5.7	System Architecture.	74
5.8	GUI: selecting octave/scale cardinality.	75
5.9	GUI: example submenu of pentatonic scales in one octave.	75
5.10	Raw EEG during <i>Focus</i> [a] and <i>Relaxation</i> [b] before filtering; ocular and muscle artifacts are evident [313].	76
5.11	Raw of EEG Signal during the Concentrating [a] and Relaxation [b] tasks, after the preprocessing pipeline.	76
5.12	Distribution of Ridge feature coefficients averaged across validation folds.	78
6.1	GUI of the NeuralPMG framework: three panels outline the main steps in creating the final polyphony, <i>Mental states training</i> , <i>Melodic pattern creation</i> , and <i>Polyphony generation</i> . Usage instructions appear in the top panel.	84
6.2	Overview of the NeuralPMG framework architecture.	85
6.3	Electrode placement under the 10–20 system.	86
6.4	Visualization of finger positions in the NeuralPMG GUI.	87
6.5	Finger positions mapped to the musical staff.	88
6.6	Infra/inter/ultrapositions on the augmented fourth axis C–F $\sharp$ .	88
6.7	Transposition of a pitch set by a constant (example: +300).	89
6.8	Maestro 1’s evaluation of the 19 polyphonies.	94
6.9	Maestro 2’s evaluation of the 19 polyphonies.	95
6.10	System architecture: EEG acquisition, Learning Engine (classification + focus index), audio/VST mapping	98

6.11	(a) Muse EEG headset; (b) 10–20 electrode layout. . . . .	98
6.12	Neural Guitar GUI: model indicators, focus index, and VST control. . .	99
6.13	Example Virtual Musical Instrument (VST) interface with controllable parameters. . . . .	99
6.14	EEG trial (4 s) after SPA denoising and preprocessing. . . . .	100
6.15	Confusion matrix for Subject 1 (LOSO test). . . . .	101
6.16	Confusion matrix for Subject 2 (LOSO test). . . . .	102
6.17	Confusion matrix for Subject 3 (LOSO test). . . . .	102
6.18	Confusion matrix for Subject 4 (LOSO test). . . . .	102
6.19	SHAP beeswarm plot for Subject 1: feature contributions. . . . .	103
6.20	SHAP beeswarm plot for Subject 2: feature contributions. . . . .	104
6.21	SHAP beeswarm plot for Subject 3: feature contributions. . . . .	104
6.22	SHAP beeswarm plot for Subject 4: feature contributions. . . . .	104
6.23	Russell’s circumplex model. . . . .	106
6.24	System architecture: acquisition, quality gate, emotion inference, condi- tional VAE synthesis, and performer control. . . . .	108
6.25	Mel spectrogram of an augmented sample. . . . .	109
6.26	Example epoch before preprocessing. . . . .	110
6.27	Example epoch after preprocessing. . . . .	111
6.28	System Architecture. . . . .	117
6.29	Cellular automaton with <code>rule_30</code> . . . . .	120
6.30	CA generated from Brain Computer Interface (BCI) features (length 20; 50 time steps). . . . .	120
6.31	Autoencoder architecture. . . . .	121
6.32	16 oscillators with partial amplitude sliders (dB) and global envelopes. . . . .	122
6.33	Autoencoder training: compressing the original Cellular Automata (CA). . . . .	122
6.34	Execution and control of the final sound texture. . . . .	123
7.1	MORIX classification framework architecture. . . . .	126
7.2	Confusion matrices on the test set. 0 = <i>Mortality (No)</i> , 1 = <i>Mortality (Yes)</i> . . . . .	130
7.3	ROC curves and AUC per model. . . . .	132
7.4	Bootstrap accuracy distributions per model. . . . .	133
7.5	Bootstrap accuracy differences: RF vs. other models. . . . .	133
7.6	Global SHAP analysis of anthropometric/biochemical predictors. . . . .	134
7.7	Cumulative explanatory contribution of selected features (RF). . . . .	134
7.9	Local SHAP analysis for a <i>Mortality (No)</i> prediction (RF). . . . .	135
7.8	Local SHAP analysis for a <i>Mortality (Yes)</i> prediction (RF). . . . .	135

7.10	MORIX web UI home: mortality risk prediction for MAFLD. . . . .	136
7.11	MORIX UI: <i>Mortality (Yes)</i> example with 97% probability. . . . .	137
7.12	MORIX UI: <i>Mortality (No)</i> example with 77% probability. . . . .	138
7.13	Per-model objective spaces (Accuracy vs. Recall for <i>Yes MCI</i> ). Black points lie on the model's Pareto frontier; red points are dominated. . .	142
7.14	Global objective space (Accuracy vs. Recall for <i>Yes MCI</i> ) across all models. Black marks: Pareto-optimal; red: dominated. Shapes encode model families. . . . .	143
7.15	Confusion matrices for the considered models. . . . .	148
7.16	Global Shapley Additive Explanations (SHAP) interpretation of Electrocardiogram (ECG) features (beeswarm). Horizontal position indicates feature effect toward NSR/ARR; color denotes feature value (red: high, blue: low). .	149
7.17	Local SHAP explanation for an NSR (negative) instance. . . . .	149
7.18	Local SHAP explanation for an ARR (positive) instance. . . . .	150
7.19	Workflow for Tasks 1a, 1b, 2a, 2b. . . . .	153
7.20	Subject overlap across feature types (all tasks). . . . .	154
7.21	Temporal AUCs (static variables, Dataset A). . . . .	158
7.22	Temporal AUCs (static variables, Dataset B). . . . .	158
7.23	Temporal AUCs (static+dynamic, Dataset A). . . . .	159
7.24	Temporal AUCs (static+dynamic, Dataset B). . . . .	159
7.25	Permutation importances (best Random Survival Forests (RSF), Dataset A). . . . .	160
7.26	Permutation importances (best RSF, Dataset B). . . . .	160



# List of tables

2.1	Physiological rhythms . . . . .	9
2.2	Summary of commonly used EEG features . . . . .	16
3.1	Taxonomy of deep learning models commonly used in BCI. . . . .	21
4.1	Time sliding accuracy by Russell quadrant. Values denote the average frequency (%) with which accuracy exceeds the threshold across participants. . . . .	39
4.2	Average LOSO test accuracy and standard deviation for stimulus vs. baseline classification. . . . .	40
4.3	Average performance of the ML models across the 4 emotions. . . . .	44
4.4	Global emotion assignment from V/A thresholds. . . . .	54
5.1	EEG acquisition summary. . . . .	65
5.2	Top accuracies by condition and configuration. Blank entries indicate pairs not among the top performers for that setting. . . . .	70
5.3	Training snapshot by EEG segment in sec. . . . .	78
5.4	Performance of the best models for different EEG segment lengths. Segment length is measured in seconds (sec). . . . .	79
5.5	Results of 5-fold cross-validation applied on Test set. MA = Mean Accuracy; STD = Standard Deviation. . . . .	79
6.1	Mental-state classification performance: LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayes, SVM = Support Vector Machine, FNN = Feedforward Neural Network. . . . .	90
6.2	Model accuracy per participant. LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayes, SVM = Support Vector Machine, FNN = Feedforward Neural Network. Grey cells indicate the selected model. . . . .	91

6.3	AttrakDiff means (AVG) and standard deviations (SD). . . . .	92
6.4	CSI dimensions: means (AVG) and standard deviations (SD). . . . .	92
6.5	NASA-TLX means (AVG) and standard deviations (SD). . . . .	92
6.6	UES dimensions: means (AVG) and standard deviations (SD). . . . .	93
6.7	Softmax normalized emotional scores. . . . .	93
6.8	Aspects of NeuralPMG appreciated by participants. . . . .	93
6.9	Criticisms aspects of neural PGM by participants. . . . .	94
6.10	Dataset size after preprocessing. . . . .	100
6.11	Best models per subject in LOSO. . . . .	101
6.12	BM: Best Model, BS: Best Scaler, BMP: Best Model Params. . . . .	101
6.13	Per-subject metrics in LOSO. . . . .	103
6.14	Mean performance across subjects. . . . .	103
6.15	Mean Model Accuracy. . . . .	111
6.16	EEG-VAE reconstruction error by emotion quadrant. . . . .	113
7.1	Hyperparameter list and corresponding values for the models. . . . .	129
7.2	Mortality prediction results (No/Yes denote classes). . . . .	129
7.3	Paired <i>t</i> -tests on bootstrapped accuracies (RF vs. others). . . . .	132
7.4	Hyperparameter list and values for the classification models. . . . .	141
7.5	Pareto-optimal configurations and performance. Best $\mathcal{H}\mathcal{V}$ values in bold. RF <sub>1</sub> : Entropy, depth=3, max-feat=log2, estimators=1; RF <sub>2</sub> : Entropy, depth=3, max-feat=sqrt, estimators=9; RF <sub>3</sub> : Entropy, depth=1, max-feat=sqrt, estimators=1; SVM <sub>1</sub> : $C=6$ , $\gamma=0.003$ , kernel=poly; XGBoost <sub>1</sub> : lr=0.1, depth=1, estimators=1; RF <sub>4</sub> : Entropy, depth=5, max-feat=log2, estimators=13. . . . .	143
7.6	Performance of top Random Forest candidates on the test set. . . . .	143
7.7	Hyperparameter grids for considered classifiers (seed=42). . . . .	147
7.8	Evaluation metrics across models. Best overall performance achieved by LGBM. . . . .	148
7.9	C-index averaged over repeated cross-validation. . . . .	157

# Chapter 1

## Introduction

Recent advances in the integration of neuroscience and Artificial Intelligence (AI) have significantly improved the field of Brain Computer Interface (BCI). Electroencephalogram (EEG), even when implemented with lightweight, low-density instrumentation, now facilitates the decoding of emotional and cognitive states, supports real-time neurofeedback, and enables interactive and creative systems. This dissertation addresses these developments by presenting pipelines and prototypes that emphasize robustness in real-world environments, Explainable Artificial Intelligence (XAI), and practical usability, validated through rigorous procedures such as Leave One Subject Out (LOSO) cross-validation and appropriate statistical evaluations.

The research is structured around three primary pillars:

1. emotion recognition from EEG using interpretable ML/Deep Learning (DL) models;
2. neurofeedback and assistive interaction, including the integration of BCI with large language models LLM;
3. *brain in the loop* approaches for musical generation and the control of virtual instruments.

### 1.0.1 Thesis Statement

This dissertation investigates how accessible and explainable artificial intelligence can improve EEG decoding in BCI, with a particular focus on low-cost and sparse-electrode devices and on deployment constraints in real-world settings. The core hypothesis is that robust EEG preprocessing—especially under continuous time–frequency analysis and streaming conditions—combined with interpretable machine learning and deep

learning models can enable reliable decoding of affective and cognitive states while supporting transparency, usability, and clinical translation. To this end, the thesis contributes (i) signal processing and preprocessing pipelines suitable for both offline and real-time operation, (ii) discriminative and hybrid AI models for EEG based state recognition, and (iii) explainable AI methodologies that provide feature and attribution level insights into model decisions. These contributions are validated across multiple application domains, emotion recognition, motivational state decoding for neurorehabilitation, and brain-in-the-loop creative systems for music interaction, and are complemented by additional healthcare-oriented studies that extend the same XAI driven perspective to other biomedical signals and clinical outcomes.

## 1.0.2 Structure of the Thesis

The thesis is organized as follows.

**Chapter 2** introduces the foundations of EEG, covering neurophysiological principles, electrode montages, signal quality issues, and core preprocessing challenges. It also reviews classical active BCI paradigms (motor imagery, P300, SSVEP) and introduces passive paradigms, with a focus on emotion related neurophysiological markers.

**Chapter 3** surveys artificial intelligence methods for BCI, including discriminative, generative, and hybrid architectures, and frames the role of explainable artificial intelligence in EEG decoding, with emphasis on interpretability, reliability, and evaluation.

**Chapter 4** presents the main research line on EEG based emotion recognition. It introduces the NeuroSense low-cost sparse-electrode dataset, develops interpretable emotion recognition pipelines based on feature engineering and model attributions, discusses deep learning approaches for affect decoding, and concludes with ARIEL, a framework that integrates BCI emotion inference with large language models for emotional support conversation.

**Chapter 5** focuses on EEG for motivation and neurorehabilitation. It reports experimental results on motivational state classification from perception and imagery protocols and presents a framework that combines mental state recognition and machine learning to support neurorehabilitation scenarios.

**Chapter 6** explores BCI for neuro musical interfaces, presenting brain-in-the-loop systems for generative and affective music, including NeuralPMG, neural musical instruments driven by BCI and biofeedback, EmoSynth for emotion-driven sound texture synthesis, and a creative framework combining BCI, cellular automata, and biofeedback.

**Chapter 7** collects additional research contributions in healthcare-oriented machine learning, including XAI supported mortality prediction in MAFLD (MORIX), Pareto-optimal model selection for mild cognitive impairment prediction, explainable ECG based classification for early heart failure detection, and time-to-event interpretable learning for multiple sclerosis worsening prediction.

**Chapter 8** concludes the dissertation by synthesizing findings across affective and cognitive EEG decoding, low-cost acquisition, explainable AI, and application-driven BCI design, outlining limitations and future research directions such as adaptive personalization, hybrid interpretable/deep models, and valid dataset studies.



# Chapter 2

## Foundations of Electroencephalography

### 2.1 Background

This chapter establishes the neurophysiological and methodological foundations for interpreting and applying the electroencephalogram EEG in clinical and BCI settings. We start with the biophysical origins of the EEG signal, then introduce electrode positioning under the international 10–20 system and the principal clinical montages, which highlight longitudinal bipolar, referential, and transverse configurations, emphasizing their consequences for source sensitivity and Signal to Noise Ratio (SNR). Then, physiological rhythms, their reactivity, common variants of the regular record, and practical guidelines for preprocessing and artifact mitigation are analysed. Finally, we connect these neurophysiological principles to decoding methodologies, including spectral and connectivity features, visual reading metrics, AI models for BCI, performance evaluation, realtime constraints, and elements of XAI that promote scientific validity, transparency, and responsible use.

#### 2.1.1 Biophysical Basis, Electrode Montages

The EEG represents the macroscopic summation of transmembrane postsynaptic currents generated by large, partially synchronized populations of cortical pyramidal neurons whose apical dendrites are oriented approximately perpendicular to the cortical surface. As illustrated in Fig. 2.1, individual synaptic events produce excitatory postsynaptic potentials (EPSPs) and inhibitory postsynaptic potentials (IPSPs), which summate across dendritic and somatic compartments. The resulting membrane poten-

tial is then compared with a threshold at the axon initial segment (hillock). When this threshold is exceeded, an action potential is triggered. Importantly, however, this spike does not directly contribute to the EEG recorded at the scalp, which predominantly reflects postsynaptic currents and their associated return. [41, 286]. The degree of phase alignment of EPSPs across adjacent cortical columns determines the resulting field configuration. When excitatory inputs are phase aligned, an open field configuration emerges, which can be approximated by an equivalent current dipole oriented along the apical dendrites. Conversely, when local return currents and heterogeneous source orientations cancel each other out, a closed field configuration arises, resulting in minimal expression on the scalp. Radially oriented dipoles produce focal scalp potentials centred below the corresponding electrodes, whereas tangential sources produce more widespread lateralised scalp topographies. Due to the layered and inhomogeneous nature of the cortex, cerebrospinal fluid, skull and scalp, the head acts as a strong spatial low-pass filter, attenuating signal amplitudes and smearing spatial patterns, substantially reducing activity originating from deep or highly focal sources [243, 286].

From a mathematical perspective, the EEG signal can be described as an overlap of sinusoidal components, expressed as

$$\text{EEG}(t) = \sum_{i=1}^N A_i \sin(2\pi f_i t + \phi_i),$$

where  $A_i$ ,  $f_i$ , and  $\phi_i$  represent the amplitude, frequency, and phase of the  $i$ -th oscillatory component, respectively.

## 2.1.2 Electrode positioning and standard montages

Electrodes are positioned according to the international 10–20/10-10 system, which ensures standardized coverage of the frontal, central, parietal, temporal, and occipital regions, with midline sites at Fpz, Fz, Cz, Pz, and Oz. Diagnostic interpretation depends not only on electrode location but also on the type of montage, that is, the referencing scheme applied to the channels. The most common clinical montages are: (i) referential (for example, linked ears A1/A2 or a mastoid or midline reference), (ii) longitudinal bipolar, and (iii) transverse bipolar. Montage selection influences sensitivity to radial versus tangential generators, the accuracy of spatial localization, and the SNR. It also substantially shapes the visual appearance of physiological and paraphysiological patterns [17, 286]. Figure 2.3 shows three different montages. Rereferencing to a common average approximates a reference at infinity and often

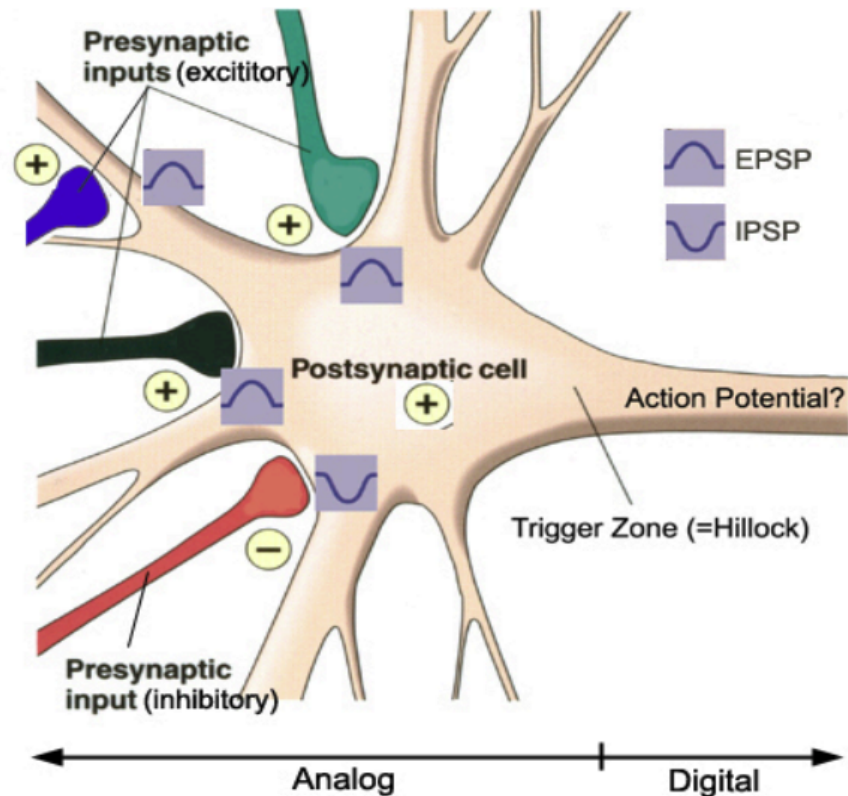


Fig. 2.1 Synaptic integration and the analog digital transition. Excitatory (+) and inhibitory (-) presynaptic inputs generate EPSPs and IPSPs that sum over dendrites and soma. Integration at the axon initial segment implements a threshold nonlinearity: if exceeded, a digital action potential is generated.

improves SNR by reducing reference site bias [336]. Spatial second derivative estimates surface Laplacian, enhance local contrast, and attenuate broadly distributed volume conducted activity, which is advantageous for focal generators ( $\mu$ -rhythm over C3/C4 in motor imagery) [134, 159, 286]. In low density BCI recordings, montage choice (referential vs. bipolar vs. Laplacian) strongly affects robustness and generalization across sessions and subjects [17, 202].

### 2.1.3 Physiological EEG

Visual interpretation of the EEG takes into account multiple factors, including age, state of alertness, dominant rhythm and its topography, reactivity, symmetry and synchrony, continuity, frequency, amplitude, and waveform morphology. The posterior dominant rhythm (alpha, 7–14 Hz) is typically maximal over occipital parietal regions during relaxed wakefulness with eyes closed. It attenuates or blocks with eye opening,

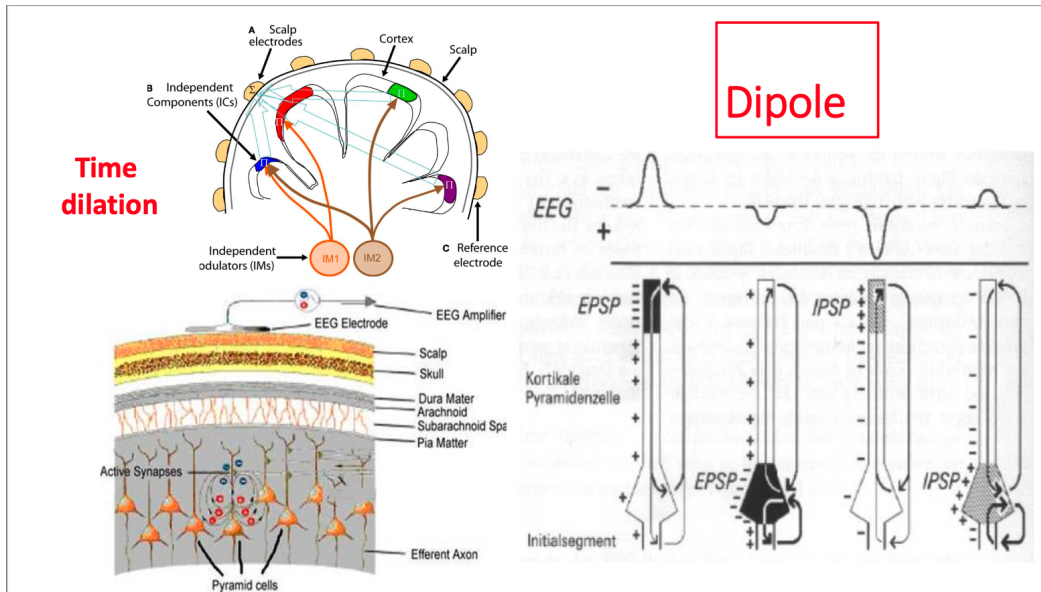


Fig. 2.2 From cortical dipoles to scalp potentials. (Left) Multiple synchronised cortical patches (conceptually equivalent to dipoles) project through layered tissues to scalp electrodes; (bottom) tissue stratification (cortex, cerebrospinal fluid, skull, scalp) imposes strong attenuation and spatial blurring; (right) Schematic distributions of EPSP/IPSP along apical dendrites producing net dipoles in the open field. [243]

attention, or mental effort (alpha blocking). Marked amplitude asymmetry (greater than 50%) may be clinically relevant, and diffuse slowing of the Posterior Dominant Rhythm (PDR) represents an early sign of global encephalopathy [41]. An overview of the principal physiological rhythms is provided in Table 2.1.

Fronto central beta activity (13 – 30  $Hz$ ) is usually of low amplitude (less than 20  $\mu V$ ) and is often enhanced by sedative, hypnotic or anxiolytic agents, as well as by skull defects. The central  $\mu$  rhythm (8 – 12  $Hz$ ) reflects activity of sensorimotor circuits and undergoes event related desynchronization during contralateral movement, motor intention, or somatosensory input [41]. Theta activity (4 – 7  $Hz$ ) is often seen over fronto central regions and can increase with drowsiness, cognitive load, or emotional processing. Delta activity (0.5 – 4  $Hz$ ) predominates in deep sleep and may appear focally or diffusely in pathological conditions when present during wakefulness [41].

## 2.2 Preprocessing and Signal Quality

### 2.2.1 Artifacts Correction

Scalp EEG is susceptible to ocular, muscular, cardiac, motion, and environmental artifacts. Standard preprocessing pipeline recommends bandpass filtering, typically

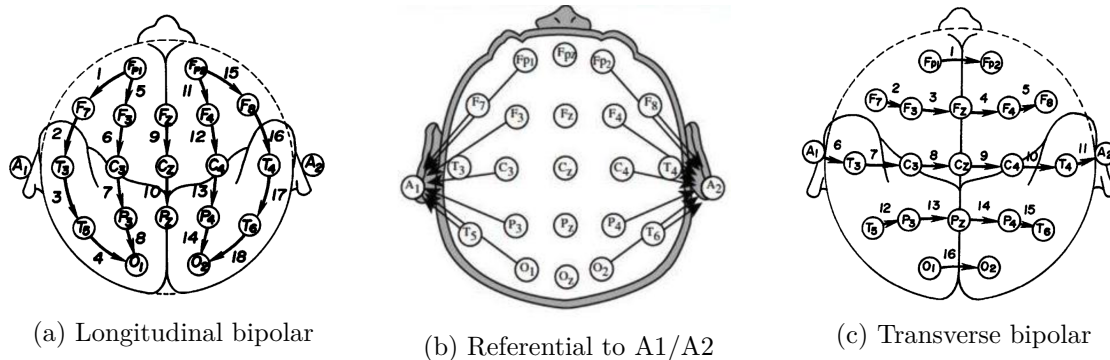


Fig. 2.3 International 10–20 system and three standard clinical montages emphasizing different source projections. Panels (a)–(c) correspond to the sketches commonly reported in the clinical EEG atlas. [17, 41]

Table 2.1 Physiological rhythms

Rhythm	Frequency (Hz)	Topography
Gamma	> 30	Fronto central; cognitive load
Beta	13–30	Fronto central; low amplitude
$\mu$	8–12	Central (C3/C4); ERD/ERS
Alpha	7–14	Posterior; blocks with eye opening
Theta	4–7	Fronto entral; cognitive load
Delta	0.5–4	Posterior; attenuates with eye opening

0.5 – 40  $Hz$  and 50/60  $Hz$  notch filtering [16, 18]. Rereferencing (standard average reference, CAR) and dedicated artifact handling methods such as Independent Component Analysis (ICA) and Signal space projection (SSP) are adopted to separate neural from non neural activity [85, 319]. For online applications, robust realtime techniques like Artifact sub space reconstruction (ASR) are efficient in suppressing transient, high variance contamination while preserving brain activity [234]. In preprocessing pipelines, it is essential to consider the order in which the individual noise reduction steps are applied. Sometimes it is advisable to first reduce ocular artefacts with ICA and then Rereference to the average.

Figure 2.4 presents the raw example EEG acquired from frontal and temporal parietal electrodes (AF3, AF4, TP9, TP10). Artifacts, including slow drifts, ocular activity, and high-amplitude fluctuations, contaminate signals. These distortions limit the quality of the signal and, if not corrected, compromise the reliability of subsequent feature extraction and classification steps in BCI applications.

In contrast, Figure 2.5 illustrates the same EEG channels after the complete preprocessing pipeline has been applied. Following notch filtering, bandpass filtering, and ASR, the resulting signal shows improved stability and clarity.

This comparison underscores the critical role of preprocessing in EEG based systems.

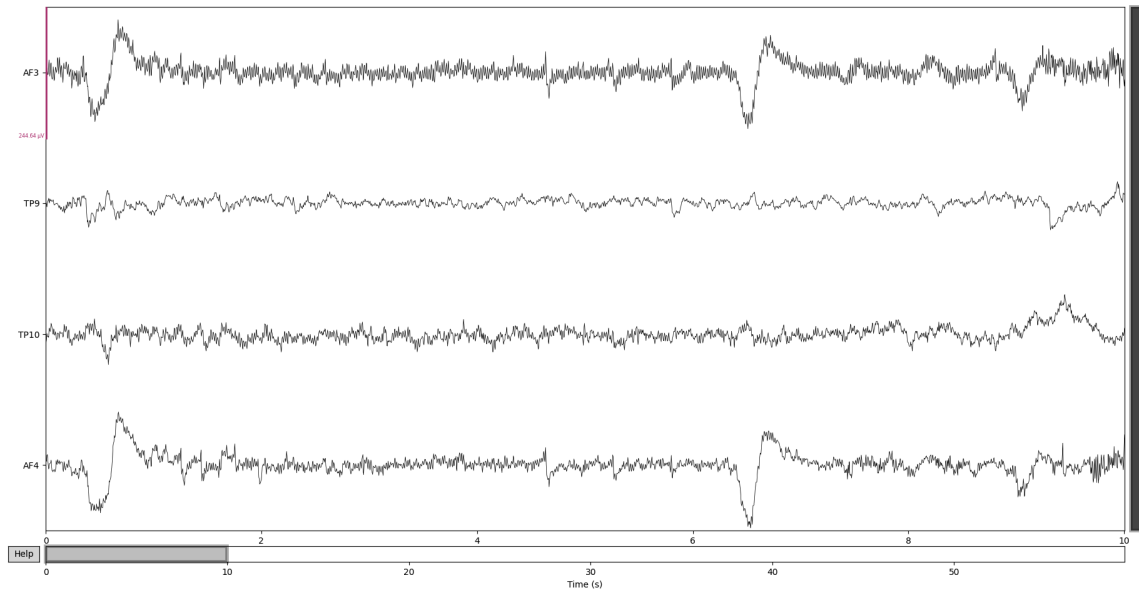


Fig. 2.4 Raw EEG from electrodes AF3, TP9, TP10, and AF4. The signal is affected by ocular, muscle, and slow drift artifacts.

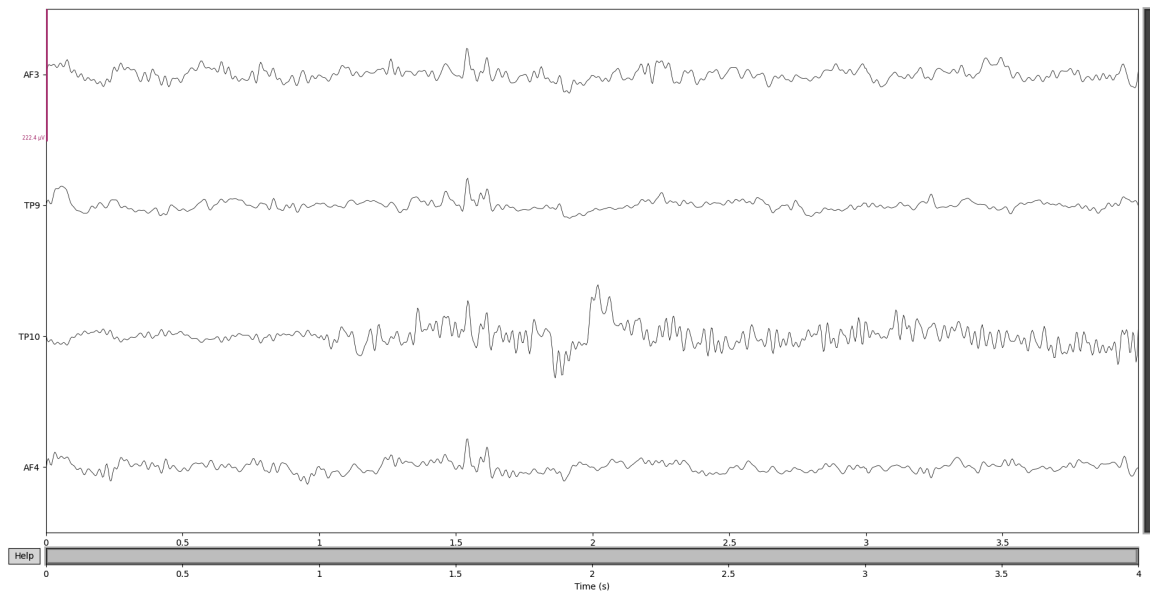


Fig. 2.5 Preprocessed EEG signals from the same electrodes after the full preprocessing pipeline. Noise and artifacts are significantly reduced.

## 2.2.2 Challenges in BCI Signal Preprocessing

In the clinical domain, EEG preprocessing pipelines typically follow well established standards, as recordings are usually acquired with medical grade devices equipped with approximately 32 or more electrodes. The reference tasks in this context are predominantly focused on the detection and diagnosis of neurological disorders, such as epilepsy or brain cancers, where high density electrode configurations provide spatial resolution sufficient to support robust artifact rejection and source separation techniques [3, 237].

In contrast, BCI applications pose a different set of challenges. The target tasks are not limited to pathological detection, but rather include the decoding of emotional states, the control of external devices such as wheelchairs or robotic arms, the classification of mental states, age prediction, or even the translation of imagined speech into text [131, 203, 326]. BCI systems are designed to be portable and easy to use, and are often based on low density EEG configurations, sometimes only 1–8 electrodes, located in specific regions of interest. This constraint significantly complicates preprocessing, as the reduced number of channels limits the applicability of advanced noise removal strategies such as ICA or SSP, which rely on multi channel data to isolate artifacts [86, 148].

In scenarios with only 4 channels, alternative solutions must be adopted. These typically include adaptive bandpass filtering, wavelet denoising, or localized artifact correction techniques, which aim to suppress noise without removing discriminative information required by machine learning classifiers [318]. Furthermore, epoch averaging represents another critical preprocessing step. In traditional paradigms, reliable averaging often requires a minimum of 30 epochs to effectively reduce artifacts and enhance the signal to noise ratio. However, this requirement severely limits realtime performance. In the BCI domain, machine learning approaches aim to reduce the number of required epochs to a maximum of 10, and in many cases as few as 5 epochs of approximately two seconds each, thereby balancing artifact reduction with system responsiveness [131, 203].

Finally, recent advances in deep learning have introduced new approaches to EEG denoising and artifact rejection. Variational Autoencoder (VAE) and other generative models have been successfully employed to reconstruct corrupted EEG segments, reducing noise while preserving task relevant signal dynamics [173, 334]. Such methods are becoming increasingly popular for low density BCI systems, where traditional artifact removal strategies are infeasible. They may play a pivotal role in enabling robust, realtime brain signal decoding.

## 2.3 EEG Feature Extraction

### 2.3.1 Univariate and Multivariate EEG Descriptors

The characterization of EEG signals typically relies on a heterogeneous set of features designed to capture both channel-specific properties and inter-channel relationships. As summarized in Table 2.2, these descriptors include statistical and amplitude measures (mean, variance, skewness, kurtosis), power and energy indices (RMS, band energy, log-bandpower), and time–frequency metrics derived from STFT or wavelet transforms (wavelet bandpower, wavelet packet energies, Hilbert-envelope power). Complementary feature families comprise complexity and entropy measures (Approximate/Sample/Permutation/Spectral Entropy, Lempel–Ziv complexity), fractal and temporal parameters (Hurst exponent, Higuchi/Katz fractal dimensions, detrended fluctuation analysis, zero crossings), and frequency-domain descriptors (PSD-based features, spectral slope/ $1/f$  exponent, spectral edge and peak frequency, Hjorth parameters). In more advanced application settings, the table also highlights covariance and spatial filtering approaches (CSP and its variants), as well as geometric representations within the Riemannian framework (covariance matrices on SPD manifolds, tangent-space mapping, distance-based descriptors, and MDM). Beyond univariate measures, bivariate and connectivity-related features are crucial for quantifying functional interactions (cross-correlation, coherence, PLV/PLI, mutual information, connectivity matrices, and graph-theoretic metrics), together with cross-frequency coupling measures (PAC, AAC, PPC) and nonlinear dynamical indicators (Lyapunov exponent, correlation dimension, and recurrence quantification analysis). Overall, this taxonomy underscores the multilevel nature of EEG signal description, enabling robust analyses for classification tasks, pattern recognition, and comparisons across experimental conditions.

## 2.4 Brain Computer Interface Active Paradigms

In the domain of BCI, three principal paradigms have been established for eliciting and characterizing neural responses. The following subsections provide a concise overview of each paradigm.

### **2.4.1 Motor Imagery Paradigms**

In Motor Imagery (MI) based BCI, the user engages in the imagination of a voluntary movement [233], which activates cortical regions responsible for the execution of actual movements [260]. The most widely employed motor imagery paradigm relies on sensorimotor rhythms [341]. Here, the imagined movement is defined as the kinesthetic imagination of body parts, such as hands, feet, or tongue, which induces specific modulations in brain activity. These modulations occur predominantly in the upper alpha (10–13 Hz) and beta (13–30 Hz) frequency bands and are reflected as Event Related Desynchronization (ERD)–Event Related Synchronization (ERS). They are most prominent at electrodes C3 and C4, located approximately above the sensorimotor cortex [258]. MI based BCI, have demonstrated practical utility, for example in controlling prosthetic limbs [13], or in neurorehabilitation contexts [259].

### **2.4.2 P300 Based Paradigms**

In addition to internally induced changes in brain activity, external visual stimuli are also frequently employed inattention positive event related potential (P300) and Steady State Visually Evoked Potentials (SSVEP) based paradigms [116]. P300 paradigm remains the most widely adopted EEG driven paradigm [2], originating from the pioneering work of Farwell and Donchin (1988), who introduced the P300 speller [107]. This paradigm exploits the P300 response, an attention dependent positive deflection that occurs between 250 and 500 ms after stimulus onset [263, 307]. The P300 typically peaks in the midline electrodes (Fz, Cz, Pz), increasing from the frontal to the parietal regions.

The most commonly used paradigm is the oddball paradigm, in which infrequent oddball stimuli are inserted into a sequence of frequent standard stimuli [107]. When the participant pays attention to the anomalous stimulus, a strong response is elicited. Applications include P300 communication systems [65, 238]

### **2.4.3 SSVEP Based Paradigms**

In visual SSVEP based BCI, the user attends to a blinking or flickering stimulus presented at a fixed frequency. This elicits a periodic brain response at the same frequency, often accompanied by its harmonics, primarily over parietal and occipital electrode sites [322]. The phenomenon was first described by Adrian and Matthews in 1934 [4, 5]. Reviews such as [242] have further characterized the SSVEP as a robust and reliable neural marker. When multiple flickering stimuli are simultaneously presented,

each at a distinct frequency, the corresponding SSVEP responses can be mapped to specific commands. This principle enables users to control external devices such as prostheses [235] or operate SSVEP spellers [338].

## 2.5 Brain Computer Interface Passive Paradigm

Passive BCI systems have emerged as a complementary branch of BCI research, shifting the focus from intentional command generation to the implicit monitoring of the user’s psychophysiological state. Rather than requiring explicit control, these paradigms continuously decode indicators of cognitive, affective, or perceptual processes from ongoing neural activity, to enable context-aware and adaptive human–machine interaction. Within this framework, emotion recognition is among the most extensively investigated applications, as affective states are tightly coupled to measurable neurophysiological modulations and provide actionable information for personalization, safety, and usability in interactive systems. A crucial methodological aspect in passive paradigms is the continuous time-frequency analysis of EEG activity, which aims to track non-stationary neural dynamics over time and across spectral bands. This requirement necessitates a robust, non-standard EEG preprocessing pipeline.

### 2.5.1 Emotion Recognition

Beyond active paradigms that rely on deliberate user intent, passive BCI paradigms aim to infer latent cognitive or affective states without requiring explicit user control. Within this category, Emotion recognition has attracted substantial interest, as affective processes modulate both cortical and autonomic activity and can be monitored to enable adaptive human–machine interaction. In EEG based emotion recognition, affect is commonly represented either in terms of discrete classes (happiness, fear, sadness) or along continuous dimensions such as valence and arousal, as formalized in the circumplex model of affect [279]. Experimental protocols typically elicit emotional responses using standardized stimuli, including images [178], music or audiovisual clips, and, more recently, immersive virtual reality scenarios, with self–assessment measures such as the Self Assessment Manikin (SAM) frequently used as ground–truth annotations [39]. From a neurophysiological perspective, emotion–related modulations are often investigated through spectral power and asymmetry indices, particularly in the alpha band, where frontal alpha asymmetry has been associated with approach withdrawal tendencies, as well as through band–specific changes in theta and beta activity and connectivity

measures across fronto–limbic networks [11, 81]. Methodologically, emotion recognition pipelines typically integrate artifact preprocessing and feature extraction with statistical learning or deep architectures, enabling the decoding of affective states from short temporal windows and supporting real time affect–aware applications. Practical use cases include adaptive interfaces, mental health monitoring, workload, stress regulation, and personalized neurofeedback, where emotion–aware systems can optimize interaction strategies by continuously estimating the user’s affective state from ongoing EEG activity [170].

Table 2.2 Summary of commonly used EEG features

<b>Category</b>	<b>Features</b>
Statistical	Mean, Variance, Standard deviation, Skewness, Kurtosis, Peak-to-peak amplitude, Median, Interquartile range
Signal Power/Energy	Root Mean Square (RMS), Signal energy, Band energy, Log-bandpower, Wavelet coefficient energy
Time–frequency	Short-Time Fourier Transform (STFT) power, Wavelet bandpower, Wavelet packet energies, Hilbert envelope power
Entropy/Complexity	Approximate Entropy, Sample Entropy, Permutation Entropy, SVD Entropy, Spectral Entropy, Lempel–Ziv complexity
Fractal/Temporal	Hurst exponent, Higuchi and Katz fractal dimensions, Detrended fluctuation analysis, Zero crossings
Frequency domain	Band-specific power, Relative power, Power spectral density (PSD) features, Spectral slope / $1/f$ exponent, Spectral edge frequency, Peak frequency, Hjorth parameters (activity, mobility, complexity)
Spatial filtering / covariance-based	Common Spatial Patterns (CSP), Filter Bank CSP (FBCSP), Regularized CSP
Riemannian / geometric	Covariance matrices on SPD manifolds, Riemannian distance to class means, Tangent space mapping coefficients, Log-Euclidean features, Minimum Distance to Mean (MDM) features
Bivariate (inter-channel)	Cross-correlation, Coherence (magnitude-squared), Phase Locking Value (PLV), Phase Lag Index (PLI) / weighted PLI, Mutual information, Nonlinear interdependence, Time-domain correlation
Connectivity / network-based	Functional connectivity matrices, Graph metrics (degree, clustering coefficient, characteristic path length, efficiency), Synchronizability indices
Cross-frequency coupling	Phase–amplitude coupling (PAC), Amplitude–amplitude coupling (AAC), Phase–phase coupling (PPC)
Nonlinear dynamics	Lyapunov exponent, Correlation dimension, Recurrence quantification analysis

# Chapter 3

## Artificial Intelligence for Brain Computer Interfaces

AI models decode brain activity via EEG to infer latent states and control external systems in real time. The central goal is to learn reliable mappings from neural signals to outputs, such as emotion labels, continuous affective values, or device commands. Core paradigms are ML, DL, and, when relevant, Natural Language Processing (NLP) for multimodal setups.

- **Machine Learning:** Algorithms learn from labeled or unlabeled EEG to classify emotions, mental states, or intentions, while adapting to inter- and intra-subject variability.
- **Deep Learning:** CNNs, RNNs, and Transformers learn temporal spectral features from raw or lightly preprocessed EEG, enabling end to end decoding.
- **NLP:** Natural language models integrate with neuro adaptive agents or condition generative systems (text to music) using brain derived affect and cognition.

**AI driven decisions.** Streaming EEG is preprocessed, features are extracted and the model outputs decisions with low latency for tasks such as emotion decoding, workload tracking, or sound synthesis control. These systems must be fast, stable, and robust. We write the online system as

$$\text{Real time BCI: } R(t) = f(\text{EEG}(t); \theta), \quad (3.1)$$

where  $R(t)$  is the output at time  $t$  (class probability, continuous arousal/valence value, or a control parameter for sound). The function  $f$  must be implemented to be efficient and resilient to noise.

## Benefits

- **Speed:** Low latency processing enables responsive closed loops (neurofeedback, real time music).
- **Accuracy/robustness:** Learned features capture subtle temporal spectral patterns under noise and non-stationarity.
- **Adaptivity:** Online learning, light calibration, and domain adaptation improve transfer across sessions and users.
- **Interpretability:** XAI exposes band channel contributions for insight and debugging.

## Challenges

- **Non stationarity:** EEG distributions shift across time and users.
- **Data limits:** Available datasets are small, noisy, and imbalanced.
- **Latency:** Realtime deployment on wearable devices requires efficiency.
- **Artifacts:** Eye, muscle, and motion noise are common outside controlled labs.
- **Ethics/UX:** Privacy, consent, and agency demand transparent, explainable design.

**Problem Statement** We consider BCI decoding as a supervised mapping from brain signals to activity outputs. We denote EEG features (time frequency values, spatial filter outputs, or compact embeddings learned from a network) by  $X$  and desired outputs (such as emotional labels, continuous arousal/valence) by  $Y$ . The decoding process consists of learning a function

$$f : X \rightarrow Y \tag{3.2}$$

parameterized by  $\theta$ . This function  $f$  is meant to approximate an unknown ground truth mapping  $f^*$  that links brain signals to their corresponding outputs.

To estimate the parameters  $\theta$ , we optimize them on a training dataset of paired examples  $\{(x_i, y_i)\}_{i=1}^N$ . The optimization objective is to minimize the average prediction error, as measured by a suitable loss function:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N L(f(x_i; \theta), y_i). \quad (3.3)$$

Once training is complete, the model can be applied to a new EEG data. Given a new input feature vector  $x$ , the trained model outputs a prediction  $\hat{y}$ :

$$\hat{y} = f(x; \theta). \quad (3.4)$$

This predicted output  $\hat{y}$  represents the system's estimate of the user's mental or emotional state, which can be used to guide applications downstream in a BCI system.

**Evaluation Metrics** We assess models with standard measures for classification and regression.

**Classification.** Let  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  denote the entries of the confusion matrix, where  $TP$  (true positives) is the number of positive samples correctly classified as positive,  $TN$  (true negatives) is the number of negative samples correctly classified as negative,  $FP$  (false positives) is the number of negative samples incorrectly classified as positive, and  $FN$  (false negatives) is the number of positive samples incorrectly classified as negative.

Accuracy is the fraction of correct predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.5)$$

Recall measures how many true positives are recovered:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3.6)$$

while Precision measures the fraction of predicted positives that are correct:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (3.7)$$

The F1 score balances Precision and Recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3.8)$$

Finally, the Area Under the ROC Curve (AUC) summarizes ranking performance across thresholds. Given a scoring function  $f(\cdot)$ , and letting  $X^+$  and  $X^-$  denote the sets of positive and negative samples, respectively, AUC can be expressed as the probability that a randomly chosen positive sample is assigned a higher score than a randomly chosen negative sample:

$$\text{AUC} = \frac{\sum_{x^- \in X^-} \sum_{x^+ \in X^+} \mathbf{1}(f(x^-) < f(x^+))}{|X^-| |X^+|}. \quad (3.9)$$

**Regression.** When the target is continuous, we evaluate prediction errors between the ground-truth values  $y_i$  and the corresponding model predictions  $\hat{y}_i$ , for  $i = 1, \dots, n$ , where  $n$  denotes the number of samples. Moreover,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  is the sample mean of the targets, and in the adjusted coefficient of determination  $R_{\text{adj}}^2$ ,  $k$  denotes the number of predictors (or, more generally, the number of fitted degrees of freedom, excluding the intercept). The metrics used are:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (3.10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (3.11)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3.12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.13)$$

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}, \quad (3.14)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (3.15)$$

### 3.1 Deep Learning Models for BCI

This section introduces the deep learning models most frequently employed in BCI research. Deep learning is a family of ML methods that organize non linear transforma-

Class	Input	Output	Core Function	Training
Discriminative	Data	Label	Feature learning + classification	Supervised
Representation	Data	Latent representation	Unsupervised feature learning	Unsupervised
Generative	Data ( $\pm$ label)	New samples / reconstructions	Data modeling & augmentation	(Mostly) unsup.
Hybrid	Data	Label / reconstruction	Combined objectives	Mixed

Table 3.1 Taxonomy of deep learning models commonly used in BCI.

tions into multiple layers, enabling hierarchical feature learning and decision making [180]. In contrast to shallow neural networks, deep neural networks comprise an input layer, two or more hidden layers, and an output layer, which allows the extraction of progressively more abstract features relevant to brain signal decoding.

Because BCI research often involves collaborators outside computer science (biomedical engineering and neuroscience), we emphasize core concepts, architectural choices, and training procedures alongside typical use cases.

### 3.1.1 Taxonomy of Deep Models in BCI

Deep learning approaches in BCI can be organized by their primary objective:

- **Discriminative models** learn a mapping from inputs to labels and are used for classification (imagined speech categories, motor imagery classes). They simultaneously perform representation learning and decision making. Canonical examples include the Multilayer Perceptron (MLP), recurrent neural networks Recurrent Neural Networks (RNN) [62, 135], and convolutional neural networks Convolutional Neural Network (CNN) [181].
- **Unsupervised models** learn compact and informative encodings of unlabelled inputs. Typical models include Autoencoder (AE), VAE and restricted Boltzmann machines (RBMs).[27, 132, 133].
- **Generative models** learn the joint structure of inputs to synthesize new samples or reconstruct signals; in BCI they are commonly used for data augmentation and reconstruction. Examples are VAE [165] and Generative Adversarial Network (GAN) [120].
- **Hybrid models** combine two or more of the above (AE $\rightarrow$ CNN, CNN+RNN, VAE $\rightarrow$ CNN), typically using representation or generative modules to bolster the data efficiency and robustness of discriminative classifiers.

### 3.1.2 Discriminative Models

Let  $(\mathbf{x}, \mathbf{y})$  denote an input label pair, with  $\mathbf{x} \in \mathbb{R}^N$  and  $\mathbf{y} \in \{0, 1\}^M$  (one hot). Discriminative models learn  $f_\theta : \mathbf{x} \mapsto \hat{\mathbf{y}}$  by minimizing a loss such as cross-entropy using stochastic gradient methods (e.g., SGD with momentum, Adam [164]) with regularization (dropout [303], weight decay, normalization [145]).

**Multi Layer Perceptron** The MLP is a fully connected feed forward network with two or more hidden layers. Layer activations  $\mathbf{h}^{(l)} = \sigma(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)})$  use nonlinearities (e.g., ReLU, *tanh*) to approximate complex decision boundaries. MLP serve as baselines for tabular or compact feature inputs (bandpower, Riemannian features) and as classifier heads atop learned representations.

**Recurrent Neural Networks** RNN capture temporal dependencies in sequential neural data. Long Short Term Memory (LSTM) cells maintain an internal state regulated by input, forget, and output gates, enabling long range memory and mitigating vanishing gradients [135]. Gated Recurrent Unit (GRU) cells streamline this with update and reset gates, often training faster with comparable accuracy on shorter sequences [62]. In BCI, RNN are well suited to raw or minimally processed EEG/Magnetoencephalography (MEG) time series and sequence to sequence tasks (continuous imagine speech decoding).

**Convolutional Neural Networks** CNN excel at learning spatially structured features and local invariances [181]. In BCI, they are adapted to (i) sensor space EEG (channels $\times$ time as a 2D image), often with temporal convolutions first (to learn band limited filters) followed, by spatial convolutions (to learn sensor topography patterns); and (ii) source space or imaging data ( Functional Near-Infrared Spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI)). Pooling layers reduce resolution and parameters; Global pooling or fully connected layers translate feature maps into class probabilities.

### 3.1.3 Unsupervised Models

These models learn latent codes  $\mathbf{z}$  that summarize the salient structure of  $\mathbf{x}$ , improving downstream classification and robustness.

**Autoencoders** An AE comprises an encoder  $\mathbf{z} = g_\phi(\mathbf{x})$  and decoder  $\hat{\mathbf{x}} = h_\psi(\mathbf{z})$ , trained to minimize a reconstruction loss  $\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})$ . Variants (denoising, sparse, contrac-

tive) promote noise robustness and compactness valuable for artifact EEG. Deep/s-tacked AE adds depth to the encoder and decoder, yielding more expressive representations.

**Restricted Boltzmann Machines and Deep Belief Networks** An RBM is a bipartite stochastic network (visible–hidden) that models data distributions and supports feature extraction via contrastive divergence [133]. DBNs stack AEs or RBMs with greedy layer wise pretraining, passing hidden representations upward to learn progressively abstract codes [27, 132].

### 3.1.4 Generative Models

**Variational Autoencoder** The VAE constrains the latent space to follow a chosen prior, enabling continuous sampling and interpolation; training optimizes a sum of reconstruction loss and a KL divergence term between the approximate posterior and the prior [165]. In BCI, VAE support data augmentation, denoising, and domain adaptation.

**Generative Adversarial Networks** GAN uses a generator and a discriminator to synthesise realistic signals. Conditional variants allow for class conditional EEG generation, enriching small training sets and improving generalisation. [120].

### 3.1.5 Hybrid Architectures

Hybrid systems integrate complementary inductive biases: (i) CNN+RNN chains temporal and spatial modeling for sequence labeling; (ii) AE→CNN/MLP uses unsupervised pretraining to initialize robust features before supervised classification; (iii) VAE/GAN→Classifier augments training data to mitigate overfitting and class imbalance.

### 3.1.6 EEG Foundation Models

Recent advances in self-supervised learning have fostered the emergence of EEG foundation models EEG-FMs, general-purpose encoders pre-trained on large-scale, heterogeneous, and predominantly unlabeled EEG corpora, with the goal of learning transferable representations that can be adapted to a broad range of tasks. A recent survey of first-generation EEG-FMs highlights both their rapid methodological convergence and the persistent gaps that currently hinder fair comparison and real-world

readiness [177]. In particular, many EEG-FMs share a common pretraining workflow in which multichannel recordings  $X \in \mathbb{R}^{T \times C}$  are segmented into temporal patches to reduce computational cost and facilitate long-range modeling, then mapped to latent embeddings via learnable projections and combined with time–frequency–position encodings. The resulting representations are typically processed by Transformer-style temporal encoders (capturing long-term dependencies) and spatial/channel encoders (capturing inter-channel dynamics) [177].

Across models, pretraining is most frequently implemented through masked self-supervision, but with notable variability in the reconstruction targets and objectives. This diversity, while reflecting active exploration, complicates direct performance attribution and motivates more systematic ablation of design choices, such as encoding strategies [177].

From a data-handling perspective, that most EEG-FMs apply relatively minimal preprocessing—typically filtering and resampling to standardize heterogeneous sources—while more aggressive artifact suppression, harmonization across acquisition sites, and even the reporting of normalization strategies are often inconsistent or absent, leaving open questions about the impact of offline handling on both pretraining and downstream performance [177]. In parallel, the lack of explicit interpretability and trustworthy modeling remains a key limitation for clinical adoption [177].

In addition to these general design principles, representative EEG-FM instantiations include BIOT, which adopts a biosignal Transformer to enable cross-dataset learning under heterogeneous channel configurations and missing electrodes [333]; LaBraM, which introduces a neural tokenizer and masked modeling over channel patches to pre-train a Transformer backbone on large multi-dataset EEG corpora [151]; and BrainGPT (EEGPT), which explores autoregressive pre-training via next-signal prediction and integrates electrode-aware modeling to better accommodate varying montages and high-density recordings [343]. These examples illustrate how current architectures typically revolve around Transformer-style encoders, while differing in tokenization strategies (continuous patches vs. learned discrete tokens), attention factorization across time and channels, and the specific self-supervised objective used during pre-training [177].

**Practical Considerations.** The choice of loss function, optimiser [164], regularisation, batch normalisation [145] substantially affects performance, as do data strategies (artefact rejection, subject splitting, cross session adaptation). Implementation, subject independence, and calibration efficiency remain central challenges; representation and generative modules often help bridge distribution shifts across sessions and users.

Explainability is essential for scientific validity, ethics, and user trust in BCI [98, 200]. It clarifies how EEG is transformed into outputs, allows comparison with known physiology, and supports informed consent in affective and creative applications. Explainable models highlight the channels, frequency bands, and time windows that drive decisions. They justify changes in control in interactive systems and support co-design with artists, clinicians, and end users. Greater transparency improves acceptance and helps systems adapt safely in closed loops [8].

### 3.1.7 Explainable BCI Decision Systems

Design choices for explainability include maps over time frequency representations, scores over spatial filters, and simple text or visual summaries for sequence models. These elements create an auditable trail for decisions and guide better stimulus and interface design. We also seek to understand why a model makes a decision. Frameworks such as Lime and Shap are beneficial in this regard.

*LIME* [272] explains a single prediction by fitting a simple local model  $\xi$  around the input  $x$  so that it mimics the complex model  $f$  near that point:

$$\xi(x) \approx f(x; \theta). \quad (3.16)$$

The weights of  $\xi$  indicate which features (for example, specific channels or frequency bands) were most influential.

*Global feature importance* estimates how much a feature helps overall. A simple method is permutation importance:

$$\text{Importance}(j) = \text{Error}_{\text{all features}} - \text{Error}_{\text{without feature } j}, \quad (3.17)$$

which compares model error with and without feature  $j$ .

*SHAP* [207] Shapley Additive Explanations, a post-hoc technique based on cooperative game theory. SHAP values quantify each feature’s contribution to a model’s prediction, enabling fine-grained attribution analyses. Given a feature  $x_i$ , its Shapley value  $\phi_i$  is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3.18)$$

where  $N$  is the set of all features and  $f(S)$  the model output when only features in  $S$  are considered.



# Chapter 4

## EEG Based Emotion Recognition

### 4.1 Introduction

The integration of AI with BCI is enabling new capabilities for affect sensing, neurofeedback, and creative Human Computer Interaction (HCI). This chapter advances AI powered BCI across three system axes: (i) realtime, explainable emotion recognition with low-latency decoding, (ii) neurofeedback and assistive interaction (including BCI and LLM for supportive dialogue), and (iii) brain in the loop music systems for expressive and potentially therapeutic control. Throughout, we emphasize sparse/low cost EEG, robustness outside the lab, and XAI consistent models.

#### 4.1.1 Problem Formulation and Aims

BCI decoding is formalized as an online mapping of EEG derived features to activity outputs (emotional labels, arousal/valence trajectories, or control parameters for musical systems), which are learned with supervised ML/DL under latency and stability constraints. Specifically, the objectives are:

1. Emotion recognition from EEG using machine and deep learning models, with post-hoc interpretability through XAI frameworks (LIME, SHAP), enabling physiological insight into feature contributions across channels and frequency bands.[67, 71, 74, 75].
2. Development of the NeuroSense dataset, a publicly available EEG resource acquired with a consumer-grade four electrode device, to democratize affective BCI research and benchmarking.[71].

3. Design of explainable and interactive neurofeedback systems that transform EEG derived affective states into feedback or adaptive behaviors in real time[21, 301].
4. Integration of BCI with LLM for supportive and affect aware dialogue systems, bridging neurophysiological sensing with natural language understanding.
5. Brain in the loop musical interfaces, such as EmoSynth and NeuralPMG, where decoded emotional or cognitive states modulate generative sound synthesis and polyphonic composition, combining neuroscience, AI, and computational creativity.[68, 72, 73].

From a methodological perspective, this work contributes pipelines for signal preprocessing, feature extraction, model training, and LOSO. It also develops hybrid architectures that combine manually created interpretable features with DL, demonstrating that even sparse EEG configurations can achieve competitive performance in emotion decoding and creative control tasks, provided that preprocessing and modelling are carefully optimised. Overall, this research places BCI at the intersection of neuroscience, machine learning, and digital creativity, proposing practical and interpretable frameworks for emotional decoding, neurofeedback, and expressive interaction. The contributions presented here extend the scope of EEG beyond controlled environments, opening the way for accessible, explainable, and human centred neurotechnologies applicable to health, education, and the arts.

## 4.2 NeuroSense: Low Cost sparse electrode Dataset

### 4.2.1 Introduction

This chapter presents NeuroSense, a dataset designed to enable emotion recognition using a low cost, sparse electrode EEG device. NeuroSense extends prior work on EEG based affective decoding by explicitly targeting accessibility and deployability. Whereas many influential studies demonstrate strong performance with high density, laboratory grade systems, such configurations can limit scalability and real world applicability. In contrast, NeuroSense adopts a minimal hardware configuration four electrodes on a consumer device to examine the feasibility of reliable affect decoding with portable instrumentation.

The dataset comprises EEG recordings from 30 participants and one set of music videos designed to evoke specific emotions. The acquisition was conducted in a controlled environment, combined with multidimensional self assessments (arousal,

valence, dominance). In addition to raw signals, NeuroSense includes a specific preprocessing workflow, statistical validations, and comparisons with external labelling schemes to articulate both the strengths and limitations of sparse EEG for affective computing.

Below, we explain the design choices, describe the stimulus selection protocol, illustrate the experimental setup in detail, and present the system architecture used to process, analyse, and model the data. We conclude with the main empirical results and implications for future research on realtime emotion recognition for HCI and mental health monitoring.

### 4.2.2 Related Work

Several open source datasets have been developed specifically for EEG based emotion recognition, each differing in electrode density, participant cohort, elicitation protocol, labeling scheme, and acquisition hardware.

**DEAP** [169] is a multimodal corpus designed for emotion analysis from physiological signals. 32 participants viewed 40 one minute music videos while EEG and peripheral signals were recorded with a Biosemi ActiveTwo system. After each clip, subjects provided self assessments on arousal, valence, dominance, and liking, enabling dimensional modeling alongside multimodal analysis.

**MAHNOB** [296] targets affective computing and HCI with rich multimodality: 30 participants (19–40 years) were recorded with 32 channel EEG, ECG, and EDA, synchronized with multi camera facial video. Participants watched 20 emotional clips eliciting disgust, amusement, fear, sadness, and joy; subjective reports and annotations accompany both physiological and behavioral data.

**SEED** [348] comprises 62 channel EEG from 15 subjects exposed to film excerpts that induce positive, negative, and neutral affect. Trials include a prompt, self assessment, and rest period to standardize temporal structure. The follow-up SEED-IV extends this design to four categories (happiness, sadness, fear, neutral) and augments EEG with eye movement features, supporting cross-subject and cross session research.

**MPED** [297] integrates four physiological modalities 62 channel EEG, galvanic skin response, respiration, and ECG from 23 participants, each completing a single session with 28 emotion eliciting videos. Labels cover discrete emotions (joy, funny, anger, fear, disgust, sadness, neutrality), facilitating studies on feature extraction and sequence modeling across modalities.

**DREAMER** [157] emphasizes mobile, affordable acquisition, pairing 14 channel EEG (Emotiv EPOC) with ECG while 23 participants viewed 18 audio visual clips. Post-stimulus self reports provide valence, arousal, and dominance ratings, enabling dimensional analysis under consumer grade hardware conditions.

### 4.2.3 Stimuli Selection

We adopted the DEAP style protocol for emotion elicitation due to its standardization, multidimensional ratings (arousal, valence, dominance) with audiovisual stimuli. The selection pipeline (Fig. 4.1) comprised five stages:

1. **Video Clip selection** We assembled 120 candidate music videos in two phases: 60 based on affective tags derived from an expanded set of Parrott’s emotion keywords [252], and 60 by manual curation. Using the Last.fm<sup>1</sup> catalogue, we queried songs tagged for emotions positioned across Russell’s circumplex quadrants [279].
2. **One Minute Highlight Detection.** We identified high emotion one minute segments using a Relevance Vector Machine (RVM) [295] trained to predict arousal/valence scores.
3. **Video/Audio Feature Extraction.** We extracted motion, color and rhythm related video features after encoding content in MPEG-1 [309]. Lighting keys (HSV) and color variance (CIE LUV) were computed frame wise [46, 209]; rhythm by average shot change rates [129]; and dynamics by motion vectors [273]. Audio channels were converted to mono MP3 (MPEG-1 Layer III) at 44.1 kHz, normalized, and summarized by 53 low level descriptors (MFCCs, energy, pitch) [122]. Pipelines are summarized in Panels B and C of Fig. 4.1.
4. **RVM Based Ranking.** The RVM [314] was trained on annotated shots from 21 films [295] to predict valence/arousal for each segment. For a segment  $i$ , we computed a composite highlight score

$$e_i = \sqrt{a_i^2 + v_i^2} \quad (4.1)$$

where  $a_i$  and  $v_i$  are centered arousal and valence predictions. High scoring segments were retained and, when necessary, refined by manual inspection to ensure musical salience and public recognizability.

---

<sup>1</sup>www.last.fm

5. **Online Subjective Annotation.** From the 120 curated segments, 40 test clips were selected via web based ratings (9-point scales for arousal, valence, dominance) by 14 volunteers. The order of presentation was randomized, with no clip repeated per rater. For each clip, we computed an intensity index  $\mu_x/\sigma_x$  (mean over standard deviation) and selected quadrant extreme exemplars to populate the final set (Panel E).

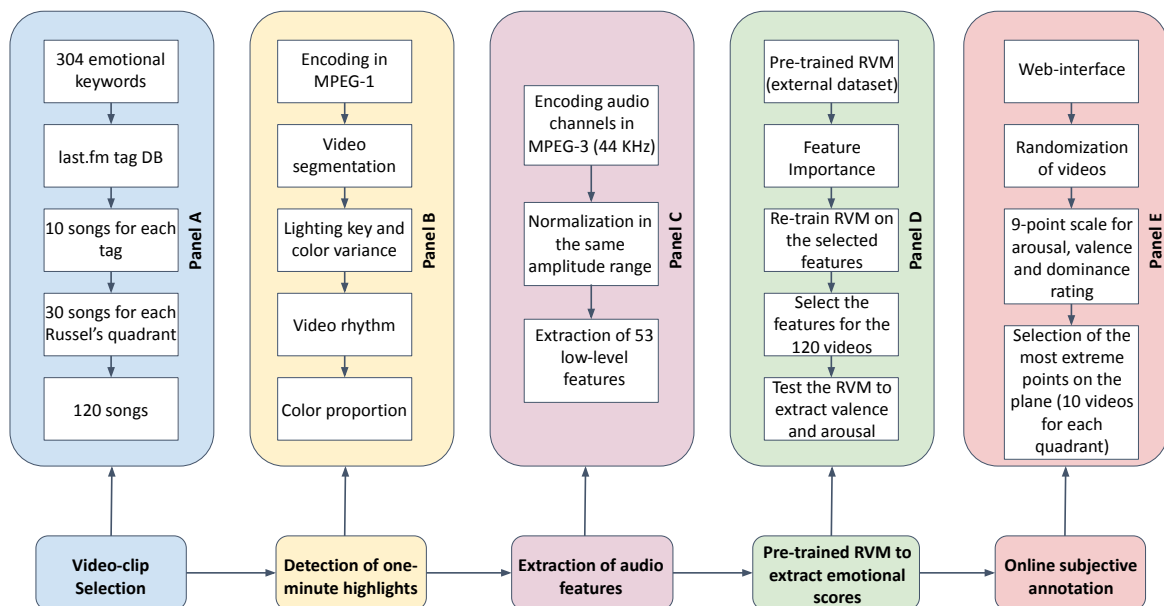


Fig. 4.1 Stimulus selection workflow: (1) affect tagged and curated pool; (2) one minute highlight detection (DEAP style); (3) video/audio feature extraction; (4) RVM based arousal/valence prediction and scoring; (5) final selection via online subjective annotation.

**Notes on Feature Engineering.** Lighting keys (HSV) and color variance (CIE LUV) capture luminance and chromatic dynamics [46, 209]; shot change rates quantify rhythm [129]; motion vectors summarize dynamic scenes [273]; and audio features (MFCCs) model timbre and spectral envelope [122]. These descriptors provide complementary predictors for elicitation strength and valence/arousal polarity.

#### 4.2.4 Experimental Setup

**Materials and Setting.** Experiments were conducted in a controlled laboratory at the Polytechnic University of Bari. EEG was recorded with a Muse 2<sup>2</sup> device with MacBook Pro (Retina, 15", Mid 2015). Stimuli were presented on a 17" display

<sup>2</sup><https://choosemuse.com/products/muse-2>

(1280 × 1024 px); videos were rendered at 800 × 600 px to reduce eye movements. Participants were seated ~1 m from the screen; audio was delivered via Yamaha HS8 monitors at a comfortable audio level.

Stimulus control was implemented in Max/MSP<sup>3</sup> with a GUI indexing video IDs. On each trial, a marker was sent over LabStreamingLayer (LSL)<sup>4</sup> and synchronized in LabRecorder<sup>5</sup> to align EEG and events.

**Participants and Ethics.** Thirty healthy adults (50% women; aged between 19 and 30 years, mean age 23.5) were recruited from the university. All provided informed consent, and no monetary compensation was offered. Ethical approval was obtained from the Local Ethics Committee of the University of Bari. After instruction on the protocol and scales, the EEG headset was fitted and signal quality verified [76].

**Protocol.** Each participant completed 40 one minute trials (the 40 selected videos). Trial structure:

- 2 s trial counter,
- 5 s resting baseline (fixation cross),
- 60 s music video,
- self ratings (arousal, valence, dominance, liking).

A brief pause followed trial 20 for re-checking signal quality and electrode placement.

**Self Assessment.** After each clip, participants completed Self Assessment Manikin (SAM) ratings [39] on 5-point scales for valence (unpleasant→pleasant), arousal (calm→excited), and dominance (submissive→dominant), plus a 3 point liking scale (dislike/neutral/like). Familiarity was rated post session on a 5 point scale (1 = never heard before, 5 = knew very well). Fig. 4.2 illustrates the SAM interfaces.

---

<sup>3</sup><https://cycling74.com/products/max>

<sup>4</sup><https://labstreaminglayer.org>

<sup>5</sup><https://github.com/labstreaminglayer/App-LabRecorder>

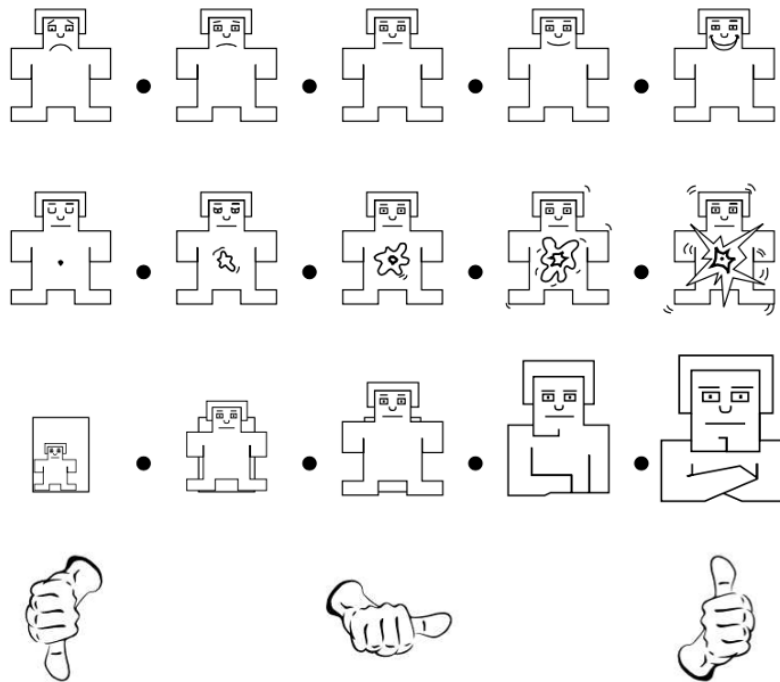


Fig. 4.2 Self Assessment Manikins for valence, arousal, and dominance, and a 3 option liking scale. Ratings were entered on continuous 5 point visual analog scales.

## 4.2.5 System Architecture

We developed an ML framework (Fig. 6.24) comprising six components:

- **Acquisition module:** streams EEG from Muse 2 with event markers.
- **Data creation:** structures per-participant recordings with precise baseline/s-stimulus timestamps.
- **Preprocessing:** MNE based<sup>6</sup> pipeline:
  1. epoch construction around baseline and stimulus intervals,
  2. FIR bandpass (1–45 Hz; Hamming window),
  3. ringing/artifact mitigation while preserving neural content,
  4. segmentation into 5 s sub-epochs for time-resolved analysis.
- **Time sliding estimation:** dynamic identification of ROI using sliding estimators to capture evolving discriminative structure.

<sup>6</sup><https://mne.tools/stable/index.html>

- **ML module:** feature extraction and classification of emotional states.
- **Statistical analyses:** (i) Pearson correlations between self ratings and external labels (arousal/valence) to assess label reliability; (ii) comparison of self assessment variability with the model’s decision probabilities to flag potentially poor participants.

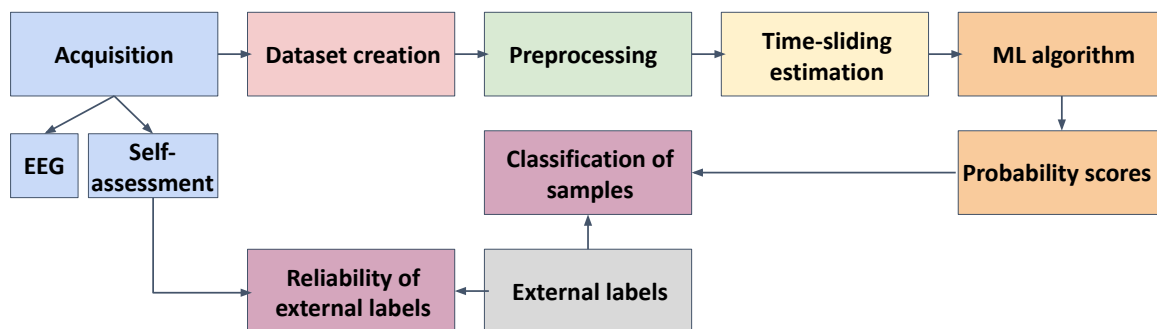


Fig. 4.3 Workflow for EEG based emotion recognition: acquisition, data creation, preprocessing, time sliding ROI estimation, ML classification, and statistical validation.

A high level pseudocode is provided in Algorithm 3. The dataset access page is available at: <https://sisinflab.poliba.it/neurosense-dataset-request/>

**Preprocessing.** To reduce sensitivity to artifacts in the EEG signal, we minimized user movements and applied a conservative denoising strategy tailored to the specific montage. Ocular artefacts were treated as outliers; segments exceeding thresholds derived from K-Nearest Neighbors (KNN) were interpolated. For temporal decoding, we used MNE’s SlidingEstimator to identify EEG segments of interest. 5 second epochs were selected for both the baseline and stimulus conditions. Emotions were discretized with reference to Russell’s circumplex model, which represents affective states along the dimensions of arousal and valence. The resulting bidimensional space was divided into 4 quadrants. The classification task was then structured as a series of binary stimulus versus baseline comparisons, one for each of the four emotional quadrants. In this way, for every combination of arousal and valence, a specific classifier was constructed to distinguish experimental conditions from the baseline.

The LOSO validation method was employed. Feature extraction was performed using MiniRocket [87] (with sktime framework [201]), which convolved each multivariate time series with random kernels. The emotional quadrant was then fitted and predicted by an Support Vector Machine (SVM) classifier.

---

**Algorithm 1:** Pseudocode for data processing and classification using the LOSO strategy

---

- 1: Initialize a data table containing user data, with each user having a unique subject ID
  - 2: **for** each unique subject ID **do**
  - 3: Create subject specific training dataset
  - 4: Create subject specific testing dataset
  - 5: **for** each file in the subject's training dataset **do**
  - 6: Extract trial epochs and corresponding labels from the training dataset
  - 7: Extract baseline epochs from the training dataset
  - 8: **end for**
  - 9: **for** each file in the subject's testing dataset **do**
  - 10: Extract trial epochs and corresponding labels from the testing dataset
  - 11: Extract baseline epochs from the testing dataset
  - 12: **end for**
  - 13: **end for**
  - 14: Initialize a processing pipeline with predefined configurations (feature extraction, normalization, classification)
  - 15: Define hyperparameter grids for optimization (number of features, maximum dilations per feature, regularization parameters)
  - 16: Initialize grid search for hyperparameter optimization based on the processing pipeline
  - 17: Train the model using the training data ( $X_{\text{train}}, y_{\text{train}}$ )
  - 18: Generate predictions and classification probabilities using the testing data ( $X_{\text{test}}, y_{\text{test}}$ )
  - 19: Store the results in a dictionary for further analysis
-

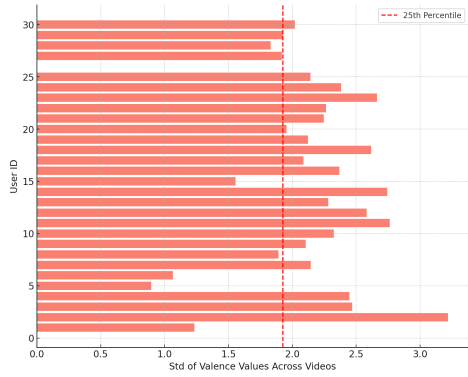
Hyperparameters were optimized in Python pipelines: MiniRocket kernel settings, feature normalization (MinMax, Standard, or Robust scaling [253]), and SVM regularization ( $C$ ). Within each LOSO fold, we performed 3 fold cross validation and randomized search over 50 configurations. All models were implemented with scikit-learn [253].

**Statistical Analysis and Participant Screening.** After training the ML model to predict the quadrant location of each video on a bivariate plane defined by arousal and valence scores, we exploited the decision probability scores to assess the average performance of each participant. The probability scores were averaged across the videos to obtain a single numeric performance score for each participant. Moreover, for each participant, we calculated the standard deviation (STD) of the self-assessment scores for arousal, dominance, and valence across the 40 videos. These STD values served as proxies for the credibility of the participants' engagement and understanding of the task, given the expected high variability across the videos. We compared these STD values to the decision probabilities output by the ML model. By setting the 25<sup>th</sup> percentile as the lower threshold, we identified participants with low variability in their scores and/or low classifier decision probabilities, flagging them as potentially non-credible or poorly performing participants.

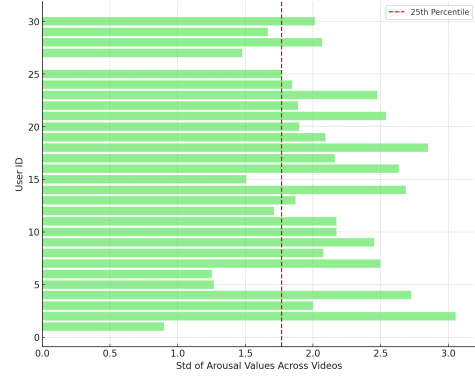
Figure 4.4 shows the horizontal bar plots of the STD values of the self-assessed valence scores (4.4a), the STD values of the self-assessed arousal scores (4.8b), and the average probability scores across the 40 videos (4.4c). The participant with user ID 26 was not considered in the analysis due to a lack of acceptable trials after the preprocessing steps. The STD analysis identified two participants (user ID 1 and 5) with consistently low variability across the two metrics (arousal and valence), thus flagging them as participants with untrustworthy labels. Interestingly, some participants exhibited low variability in arousal and valence (user ID 6 and 15) but maintained high mean decision probabilities, suggesting that their self-assessments did not align well with the classifier's performance. These cases likely indicate misunderstandings or inaccuracies in self-assessment, reinforcing the need for external labels.

Conversely, another group of participants showed high variability in their self-assessments but low classifier performance (user ID 10, 11, 20, 23), suggesting that the external labels may not accurately reflect their actual emotional states. This outcome highlights the complex interplay between subjective emotional experiences and objective measures. However, the overall use of external labels remains justified, as they provide a more stable reference point for the classifier, facilitating the identification of broader

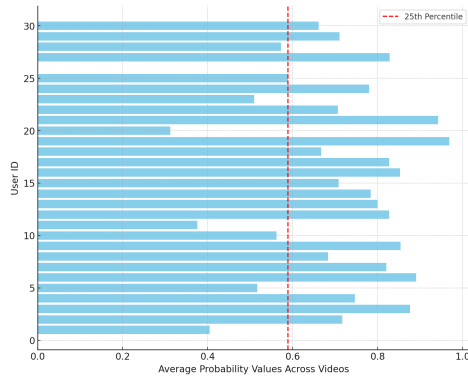
patterns in EEG data that the subjective nature of self-assessment scores might obscure. This approach ultimately supports the development of more generalizable and reliable emotion recognition models.



(a) STD of valence scores across videos.



(b) STD of arousal scores across videos.



(c) Average probability scores across videos.

Fig. 4.4 Horizontal bar plots are resulting from the STD analysis for each participant.

## 4.2.6 Key Findings and Experiments

**External vs. Self Reported Labels.** The heat maps of self rated valence, arousal, and dominance are shown in Figures 4.5–4.7. The correlations (Fig. 4.8) indicate a strong alignment for valence. At the same time, arousal and dominance show weaker and non significant associations, in line with the known difficulties of self assessing one’s own emotional states. Given the standardized nature of external labels and assessments made by users outside the context of the experiment, we adopted the DEAP valence/arousal quadrant estimates as primary targets.

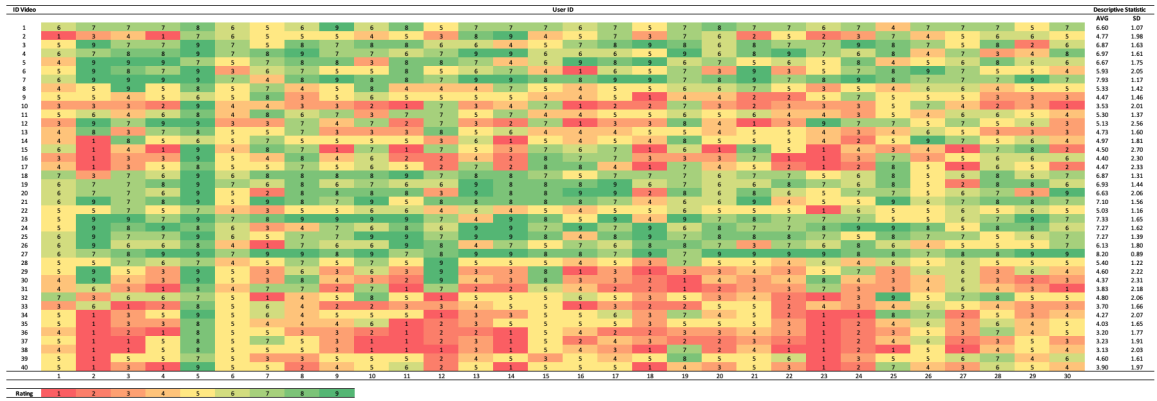


Fig. 4.5 Self assessed *valence* (participants on the x axis; video IDs on the y axis).

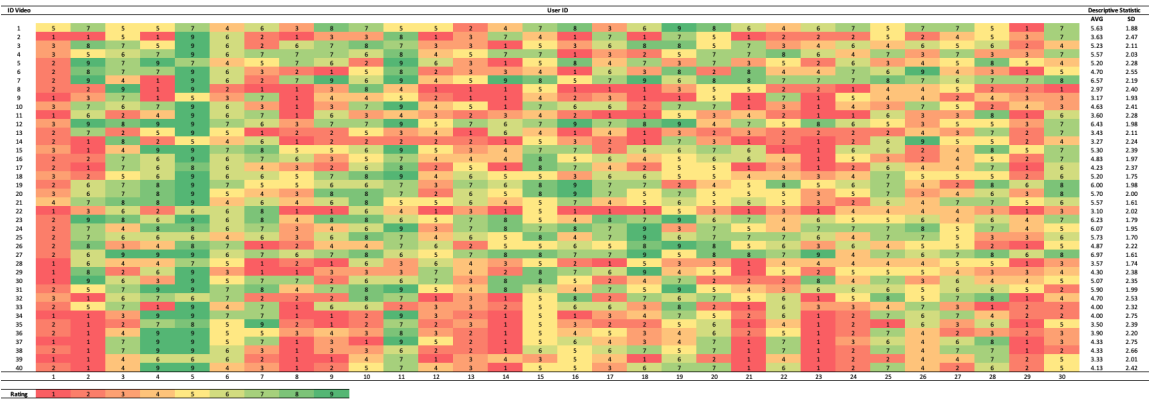


Fig. 4.6 Self assessed *arousal* (participants on the x axis; video IDs on the y axis).

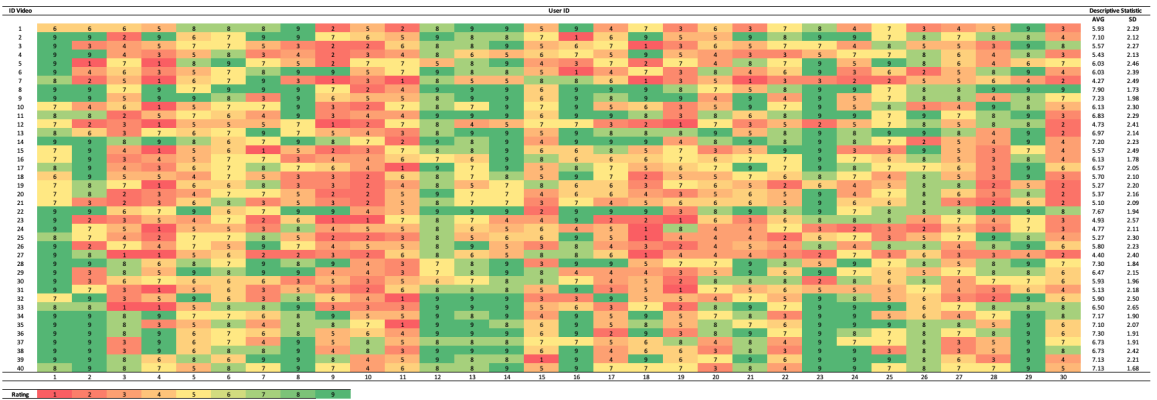


Fig. 4.7 Self assessed *dominance* (participants on the x axis; video IDs on the y axis).

**Time Sliding Region of Interest (ROI).** Table 4.1 shows the frequency (in %) with which accuracy exceeds a predefined threshold on 5s windows relative to each Russell quadrant. The 0–5s interval produces the most consistent performance,

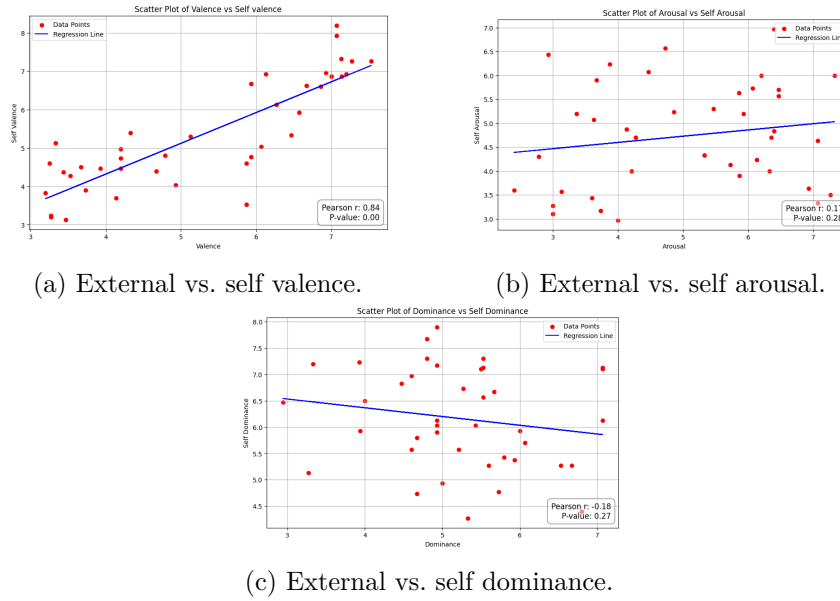


Fig. 4.8 Correlations between external labels and self assessments across videos.

suggesting that the segments of the epochs at the beginning of the audio/video stimulus contain discriminating and robust information for classifying the baseline relative to the emotional stimulus.

Table 4.1 Time sliding accuracy by Russell quadrant. Values denote the average frequency (%) with which accuracy exceeds the threshold across participants.

Time segment (s)	1° RQ	2° RQ	3° RQ	4° RQ
0–5	63.36	62.60	62.12	58.24
5–10	62.44	61.76	60.68	55.60
10–15	62.76	61.12	60.56	55.92
15–20	62.88	62.24	61.72	58.08
20–25	61.47	61.04	61.36	57.88

(a) Columns 2–5 correspond to the first through fourth Russell quadrants.

**Classification Performance.** Table 4.2 summarizes average LOSO test accuracy and standard deviation for binary classification (stimulus vs. baseline) across four model configurations. Most models achieve accuracy  $\geq 0.75$  with low variance, indicating reliable discrimination using sparse EEG.

Table 4.2 Average LOSO test accuracy and standard deviation for stimulus vs. baseline classification.

Model	Mean Accuracy	Std. Dev.
1	0.77	0.035
2	0.78	0.022
3	0.76	0.024
4	0.80	0.028

### 4.2.7 Implications and Future Prospects

NeuroSense shows that a consumer grade EEG device can support affect decoding with competitive accuracy when combined with specific and accurate preprocessing. ROI selection allows us to identify the section of the EEG that is richest in discriminating information. Convolution filters and the LOSO training strategy allow for a more in depth analysis of the model’s capacity in the generalisation step. NeuroSense expands participation in affective EEG research in real world scenarios and improves understanding of emotional processes decoded through the EEG signal. Future work will investigate continuous monitoring of emotional state and further analysis of models in terms of explainability.

## 4.3 Explainable Emotion Recognition with Low Cost EEG: Feature Engineering and Model Attributions

### 4.3.1 Introduction

This work investigates whether sparse, low cost electroencephalography EEG can support reliable explainable emotion recognition suitable for real world BCI scenarios. Building on the NeuroSense dataset acquired with a 4 electrode consumer headset (Muse 2), we develop pipeline that (i) performs transparent signal preprocessing, (ii) extracts features informed by neurophysiology (spectral, entropy, autoregressive descriptors), (iii) trains classical ML models under LOSO validation, and (iv) applies Post hoc explainability to quantify feature importance across subjects and affective states.

We address three research questions:

1. **RQ1:** How do engineered features influence decoding performance for EEG based emotion recognition under LOSO evaluation?
2. **RQ2:** Are the feature attributions consistent across subjects for the same target emotion?
3. **RQ3:** Do different emotions rely on shared EEG markers or on emotion specific descriptors?

### 4.3.2 Related Work

Research on EEG based emotion recognition spans classical multimodal pipelines, advanced domain adaptation, lightweight deep models, and graph-based representations. Kumar *et al.*[174] show that classifier choice should be signal specific: Random Forest (RF) excels on EEG/Galvanic Skin Response (GSR), whereas KNN/KNN/Logistic Regression (LR) perform better on ECG, underscoring modality dependent inductive biases. Yu *et al.*[340] advance cross subject/session robustness via Fine-grained Mutual Learning Adaptation Network (FMLAN), coupling multiple domain adaptation branches with mutual learning and a fine grained alignment module to improve generalization on SEED/SEED-IV. Gunda *et al.*[125] demonstrate that dual stream (time/frequency) architectures with attention can cut parameters by up to 98% while retaining high accuracy of 95.18% on SEED, supporting real time BCI use. Li *et al.*[190] exploit the intrinsic graph structure of multichannel EEG with Multiple Emotion-related Spatial Network Topology Patterns (MESNP), learning discriminative spatial topology patterns for stable, accurate emotion classification—even in online scenarios. Atkinson *et al.*[23] integrate feature selection minimum-Redundancy-Maximum-Relevance (mRMR) with kernel SVM Radial Basis Function (RBF) to boost multi class performance on DEAP while reducing computational cost.

### 4.3.3 Dataset, Protocol, and Preprocessing

We analyze the NeuroSense Emotion Recognition Dataset (30 participants) recorded with Muse 2 (four channels: AF7/AF8/TP9/TP10). Participants watched music video clips selected with a DEAP pipeline and provided self reports on arousal, valence, and dominance. Acquisition was synchronized via LabStreamingLayer; stimuli were presented in a controlled laboratory environment.

Preprocessing follows a conservative, reproducible procedure: epoching around baseline and stimulus intervals; FIR bandpass filtering ( $1 - 45$  Hz; Hamming window);

mitigation of non neural artifacts through ASR; segmentation into 5s sub epochs to enable time resolved analysis. A sliding estimator strategy identifies the most informative ROI related to emotion and subject, contrasting baseline vs. stimulus segments.

#### 4.3.4 Feature Engineering and Classifiers

We identify and extract interpretable clusters of well known features observed in EEG:

- **Spectral power** in  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  ranges (for channel);
- **Entropy/complexity** indices (approximate/sample entropy, fractal dimensions, Hurst exponent) to capture irregularity and nonlinearity;
- **Autoregressive coefficients** as compact time domain summaries.

We compare standard classifiers SVM, RF, MLP, KNN within LOSO training strategy. Hyperparameters (SVM kernel/regularization; RF depth/trees; MLP hidden sizes/learning rate; KNN neighbors/metrics) are optimized via randomized search with inner cross validation and appropriate scaling (MinMax, Standard, Robust). Performance is reported for the 4 quadrants of Russell’s circumplex (Excited, Relaxed, Sad, Angry) using Accuracy, Precision, Recall, and F1.

#### Inter subject similarity

To evaluate the consistency of feature attributions across individuals, standard on correlation matrices were computed on SHAP vectors for each emotion. These correlations quantified the similarity of feature attribution patterns among participants. The resulting distributions were summarized using boxplots to facilitate the assessment of inter subject variability.

#### Cross emotion comparison

To investigate whether EEG features were common across different emotions or specific to particular affective states, SHAP value distributions were compared across the 4 emotional quadrants. This comparison allowed us to assess whether the same characteristics played a similar role in multiple emotional states or whether, instead, they contributed in different ways, depending on the emotion.

### 4.3.5 Results and Discussion

The results indicate that SVM achieved the highest performance in three quadrants (Excited, Sad, Angry), whereas Random Forest RF yielded the best accuracy for the Relaxed condition. These differences suggest that linear separability in the engineered feature space is particularly relevant for high arousal emotions. At the same time, RF may better capture the variability present in low arousal states. Table 4.3 shows the average performance of the ML models across the four emotions.

When compared with the previous work introducing the NeuroSense dataset, which reported an average accuracy of 75% across emotional quadrants, the performance achieved in this study is slightly lower. The earlier work employed MiniRocket, an algorithm that applies random convolutional kernels followed by global max pooling, thereby generating a high dimensional feature space capable of capturing complex temporal dynamics in EEG signals. These features were then classified using SVM, benefiting from the rich representations produced by the MiniRocket transformation.

By contrast, the present study relied on handcrafted features, specifically spectral and entropy based descriptors derived directly from raw EEG signals. While these features provided interpretable physiological insights, they yielded marginally lower classification performance compared to MiniRocket. This comparison suggests that convolutional feature extraction approaches such as MiniRocket are particularly effective at exploiting hidden temporal dependencies in EEG. In contrast, engineered features may be more suitable when interpretability and explainability are prioritized.

To assess the consistency of SHAP based feature attributions across participants, Pearson correlation coefficients were computed between SHAP vectors for each pair of subjects. The distributions of these coefficients, summarized in the boxplots of Figure 4.9, highlight the degree of similarity in feature attribution across individuals. The results revealed generally low correlations for all four emotional states, with values clustered near zero, indicating substantial inter subject variability.

These findings suggest that the EEG features contributing to emotion classification differ markedly between individuals, regardless of the emotional quadrant. No emotion exhibited consistently higher agreement across participants. This inter subject variability may be attributed to differences in brain physiology, cognitive processing, or subjective emotional perception. Overall, these results highlight a critical limitation of subject independent EEG based emotion recognition and motivate the exploration of adaptive or personalized models capable of accommodating individual differences in feature attribution.

Table 4.3 Average performance of the ML models across the 4 emotions.

Model	Emotion	Precision	Recall	F1 score	Accuracy
KNeighbors	Excited	0.69	0.70	0.69	0.69
	Relaxed	0.68	0.68	0.67	0.68
	Sad	0.63	0.63	0.62	0.62
	Angry	0.67	0.67	0.67	0.67
RandomForest	Excited	0.70	0.70	0.70	0.70
	Relaxed	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>
	Sad	0.65	0.66	0.65	0.65
	Angry	0.68	0.68	0.68	0.68
MLP	Excited	0.66	0.66	0.66	0.66
	Relaxed	0.70	0.70	0.69	0.69
	Sad	0.64	0.64	0.64	0.64
	Angry	0.68	0.68	0.68	0.68
SVC	Excited	<b>0.72</b>	<b>0.72</b>	<b>0.72</b>	<b>0.72</b>
	Relaxed	0.68	0.69	0.69	0.68
	Sad	<b>0.69</b>	<b>0.69</b>	<b>0.69</b>	<b>0.69</b>
	Angry	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>	<b>0.70</b>

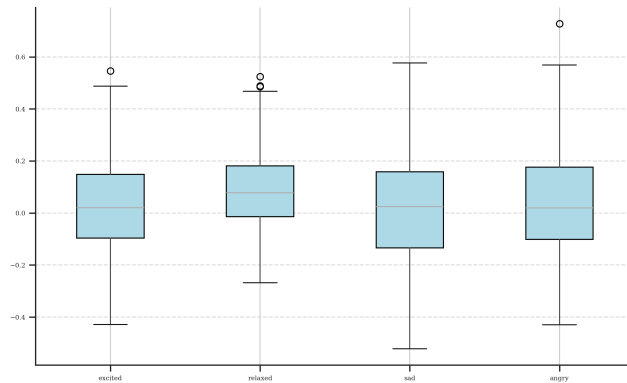


Fig. 4.9 Boxplots of pairwise SHAP value correlations across subjects for each emotional state. Each boxplot summarizes the distribution of Pearson correlation coefficients between SHAP vectors from different participants. Lower values reflect greater inter subject variability in feature importance.

To explore whether EEG features were shared across emotions or uniquely associated with specific affective states, we compared SHAP based feature rankings across the four quadrants. Figure 4.10 displays the ten most important features for each emotional state, averaged across participants.

The analysis revealed both shared and emotion-specific neural markers. Excited states were dominated by spectral characteristics of the theta and alpha bands at frontal electrodes, consistent with previous evidence linking these rhythms to attentional engagement and arousal. Relaxation states were primarily associated with beta

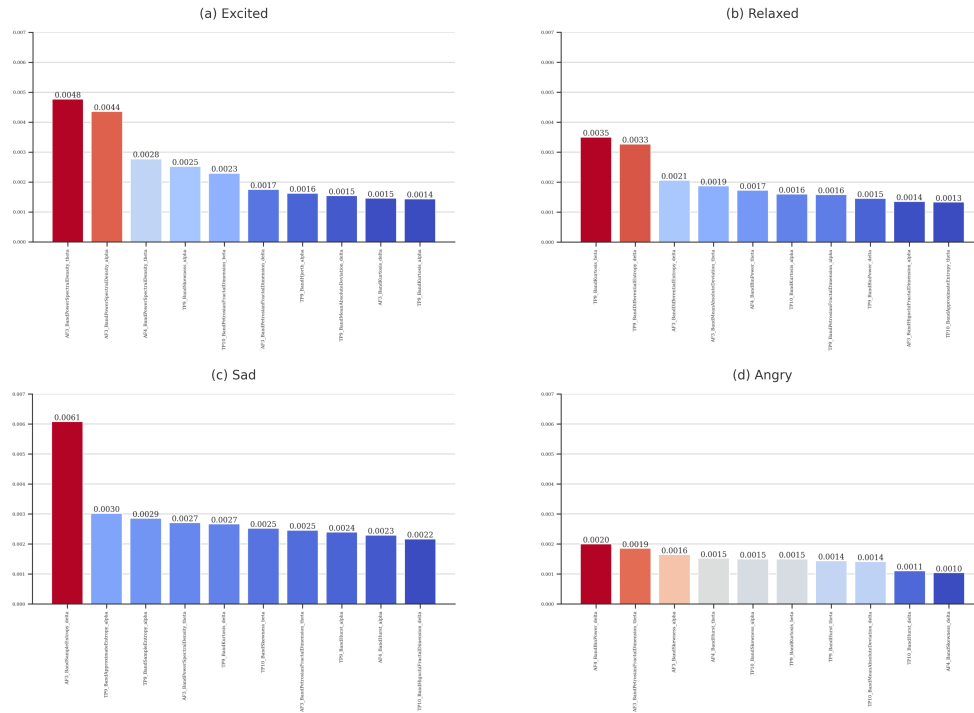


Fig. 4.10 Top 10 SHAP based features for each emotional state: (a) Excited, (b) Relaxed, (c) Sad, (d) Angry. Each panel shows the most relevant EEG features ranked by mean SHAP values across all subjects.

band kurtosis and delta band entropy at TP9, potentially reflecting reduced cortical excitability and inactivity. [237].

States of sadness were characterised by entropy based measures, particularly approximate entropy in the delta and alpha bands at the frontal and temporal electrodes. This is consistent with previous findings that emotional distress is associated with increased EEG complexity [78]. In contrast, anger was characterised by non linear and higher order features, such as delta power and fractal dimension, consistent with increased cortical activation and arousal [195].

From a spatial perspective (Figure 4.11), excited states showed a predominantly left lateralized pattern, supporting the frontal asymmetry hypothesis [12]. Relaxed and sad states were associated with TP9, pointing to stronger temporal parietal involvement [168]. Anger, in contrast, exhibited a more bilateral distribution, possibly reflecting motor-preparatory activity linked to action oriented affective responses [239].

In terms of frequency content, theta and alpha bands were most relevant for high arousal states (excited, angry), while delta and beta related entropy measures predominated in low arousal states (relaxed, sad). These findings underscore the need

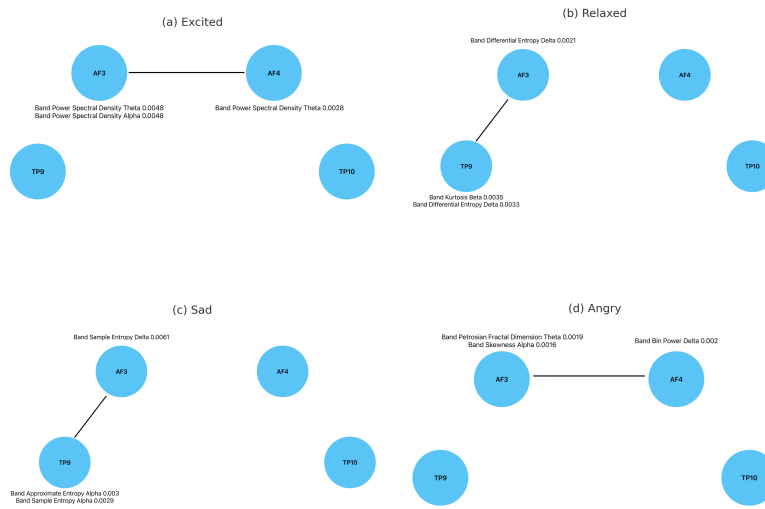


Fig. 4.11 Spatial distribution of SHAP based feature importance across emotional states: (a) Excited, (b) Relaxed, (c) Sad, (d) Angry. Each panel highlights the most relevant EEG features and their associated electrode locations.

to combine broadly generalizable spectral features with emotion specific non linear descriptors to effectively capture EEG correlates of affective processing.

### 4.3.6 Conclusions and Future Work

This study investigated interpretable EEG based emotion recognition using a low cost, sparse electrode device, combining engineered features with SHAP based explainability. Results demonstrated that spectral and entropy based features enable competitive performance in emotion classification, while also revealing inter subject variability in their importance.

Although some EEG features showed cross emotion generalizability, others were clearly emotion specific, reflecting distinct neural mechanisms. These findings underline the limitations of subject independent models and support the development of adaptive, personalized approaches for affective EEG systems.

Future research will focus on dynamic feature selection and online adaptation strategies to enhance robustness and subject specificity. While the present work was limited to a single dataset, future validation across multiple datasets will help assess generalizability across acquisition settings. Moreover, hybrid pipelines that

combine engineered and learned representations may capture complex EEG patterns more effectively without sacrificing interpretability. Finally, collaboration with affective neuroscience experts will be pursued to strengthen the neurophysiological interpretation of EEG markers identified in this study.

## 4.4 Deep Learning Applied in Human Emotion Recognition

### 4.4.1 Introduction

Growing interest in affective computing and BCI has accelerated progress in emotion recognition from electroencephalography EEG signals. In particular, estimating continuous affective dimensions, such as Valence, Arousal, and Dominance, offers practical value for neuromotor rehabilitation and psychological therapies. This chapter presents a prototype regression based emotion recognition system that provides real time EEG based biofeedback. The system continuously predicts affective scores and displays them interactively, positioning it as a practical tool for clinical studies and therapeutic settings.

### 4.4.2 Related Work

In this section, we briefly review the state of the art in terms of current ML or DL approaches used to solve the problem of emotion recognition using neurofeedback tools.

Ante *et al.* [315] propose transforming EEG features into topographic and holographic representations—Feature Maps Based on the Topographic Representation (TOPO-FM) and Feature Maps Based on the Holographic Representation (HOLO-FM), respectively. TOPO-FM visualizes feature values as topographic maps, whereas HOLO-FM employs computer generated holography to synthesize two dimensional feature maps from spatial point sets. Deep neural models are then trained on these maps, and the learned representations are fused into a single feature matrix to infer affective state. The authors highlight potential applications in assistive technologies, authentication/i-identification, driver fatigue detection, and neurogaming.

Najam Dar *et al.* [80] introduce a subject-independent deep architecture for emotion recognition that combines CNN and LSTM blocks. Specifically, a 2D CNN processes 14 channel EEG arrays, while ECG and GSR streams are modeled via a hybrid 1D CNN and LSTM pipeline. Evaluated on DREAMER [157] and AMIGOS [282] both acquired

with low cost, wearable sensors suited to real world settings the method reports improved performance on 4 class arousal recognition under a subject-independent protocol.

Galvão *et al.* [115] combine engineered features, band limited brain rhythms, and regression models to predict continuous valence and arousal. Among tested ML regressors, KNN and RF perform best; features in the beta and gamma ranges are most informative overall, and alpha band asymmetry is particularly useful for valence prediction. A KNN based model attains the strongest results across arousal, valence, and auxiliary classification tasks.

### 4.4.3 Materials and Methods

#### 4.4.4 BCI Devices

Multiple public datasets and reference devices exist for EEG based affect studies, including DEAP, DREAMER, and IDEA [154, 157, 169]. We use the DREAMER dataset [157], recorded with the Emotiv EPOC device with 14 channels; AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4; mastoid references M1/M2;  $f_s = 128$  Hz). Electrode placement follows the international 10–20 standard [230]

A primary design objective is to minimize the number of EEG channels while preserving the model’s predictive capability. To this end, we employ the Emotiv Insight device, which electrode from the AF3, AF4, T7, T8, and Pz electrodes at a sampling rate of  $f_s = 128$  Hz. For inter device compatibility, the number of electrodes is reduced to four (AF3, AF4, T7, and T8), excluding Pz. Our results demonstrate that a one dimensional convolutional neural network (1D-CNN) can achieve high classification performance even when limited to these four channels(Figure 4.12).

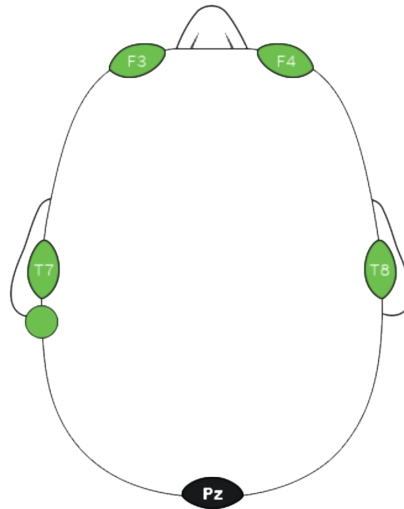


Fig. 4.12 Emotiv Insight 5 electrodes (10–20 montage).

#### 4.4.5 Dataset

DREAMER includes EEG recordings from 23 participants during audiovisual emotion induction using 18 video clips mapped to nine basic emotions: amusement, excitement, happiness, calm, anger, disgust, fear, sadness, and surprise. After each video clip, participants rated Valence, Arousal, and Dominance on a 5-point Likert scale using Sam Manikin’s questionnaire [39]. The raw EEG recordings are provided in MATLAB format. We follow the authors’ recommendations in selecting the final 60 seconds of each trial for labeling [157].

#### 4.4.6 Preprocessing

Robust real time performance requires automated artifact handling. Our pipeline used both for offline training and online inference proceeds as follows; its efficacy was checked by visual inspection of all sessions.

- **Trial selection & structuring:** Select all labeled trials (414 initial) and convert to MNE compatible objects with 10–20 metadata [123].
- **DC offset removal:** Manufacturer recommended correction for Emotiv raw streams.
- **Line noise & detrending:** 50 Hz notch; high-pass at 1 Hz; low-pass at 40 Hz.
- **Epoching:** Segment continuous data into 1 s sub-epochs for quality screening.

- **Noisy epoch pruning:** Remove sub epochs exceeding thresholds on amplitude, variance, and deviation from channel means.
- **ICA for ocular correction:** Run ICA; use AF3/AF4 as EOG surrogates to flag blink/saccade components [192].
- **Residual artifact repair:** Apply *Autoreject* [149] for EMG/slow drifts; detect/interpolate bad channels and segments with the PREP approach [31] via `pyprep`.
- **Reconstruction:** Reassemble denoised continuous signals.

#### 4.4.7 Feature Set and Training Data

For each continuous trial, 4-second epochs are extracted with overlapping every 128 samples. For each window and channel, we extract bandwidth power features, using NeuroKit2 [210]: Theta (4–8Hz), Alpha (8–13Hz), Beta1 (13–16Hz), Beta2 (16–20Hz), Beta3 (20–30Hz). We obtain 46,991 labeled epochs.

#### 4.4.8 Model

The dataset is split into training/validation/test using `scikit-learn` [254] (80/20; with 80 further split into 75/25). The features were scaled with `MinMaxScaler`. The regressor is a 1D CNN implemented in Keras/TensorFlow [1, 63]:

- **Conv stack:** 1D conv layers ( $k = 3$ ) with 128, 128, 64 filters; ReLU activations; BatchNorm after the first two; MaxPool1D after each conv.
- **Dense head:** Flatten  $\rightarrow$  fully connected layers (128  $\rightarrow$  128  $\rightarrow$  32) with tanh, tanh, ReLU; Dropout=0.2.
- **Output:** 3 linear units for Valence, Arousal, Dominance (multitask regression).
- **Optimization:** Adam [166]; loss is Mean Squared Error (MSE); early stopping with patience=10.

#### 4.4.9 System Performance and Real Time Inference

On the test set, the proposed model achieves excellent predictive accuracy, yielding a coefficient of determination of  $R^2 = 0.93$ , a mean absolute error of  $MAE = 0.08$ , and a

mean absolute percentage error of  $MAPE = 0.07$ . These results indicate a high degree of agreement between predicted and observed values, confirming the model's robustness and generalization capability.

Error distributions are shown in Figures 4.13–4.15; Loss curve is shown in Figure 4.16.

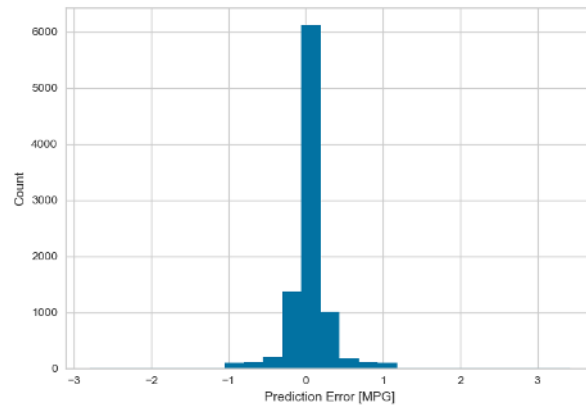


Fig. 4.13 Valence error distribution.

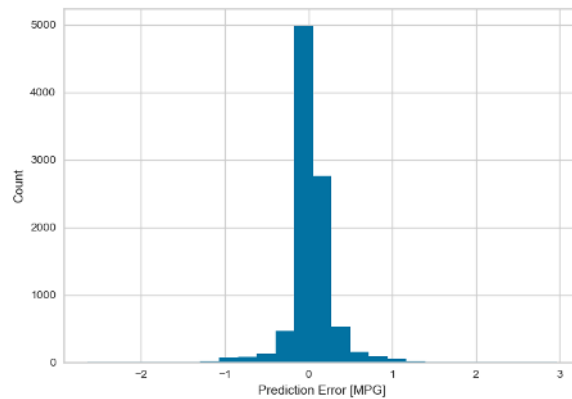


Fig. 4.14 Arousal error distribution.

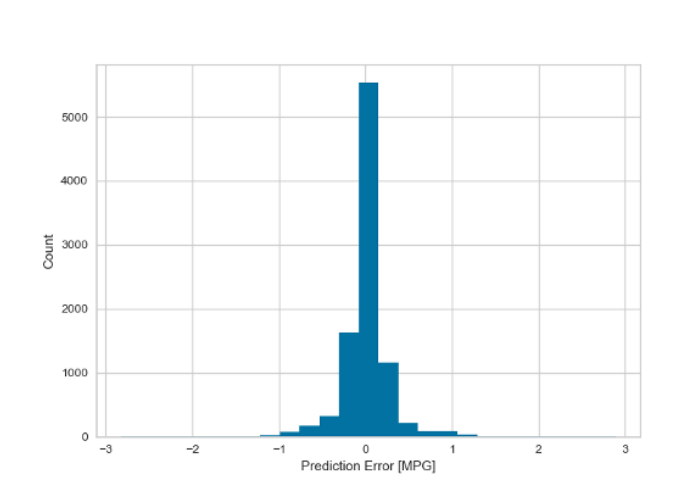


Fig. 4.15 Dominance error distribution.

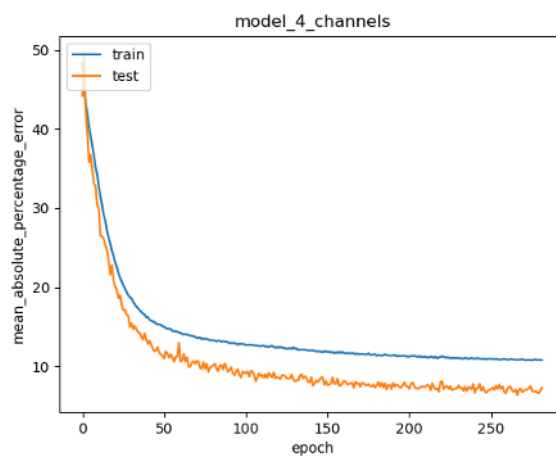


Fig. 4.16 Training loss across epochs.

A front end based on  $Max/MSP^7$  is used to visualize the continuous predictions (Figure 4.17 Figure 4.18 displays the entire system architecture.

<sup>7</sup><https://cycling74.com/products/max>

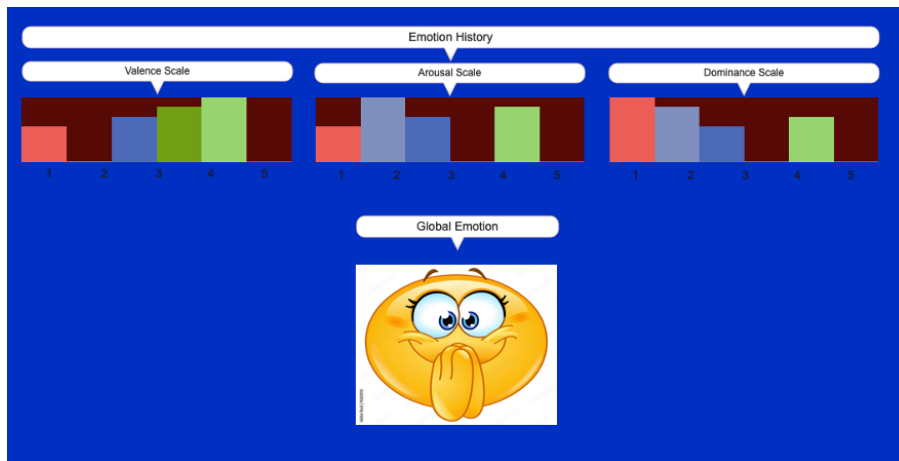


Fig. 4.17 User interface for real-time biofeedback.

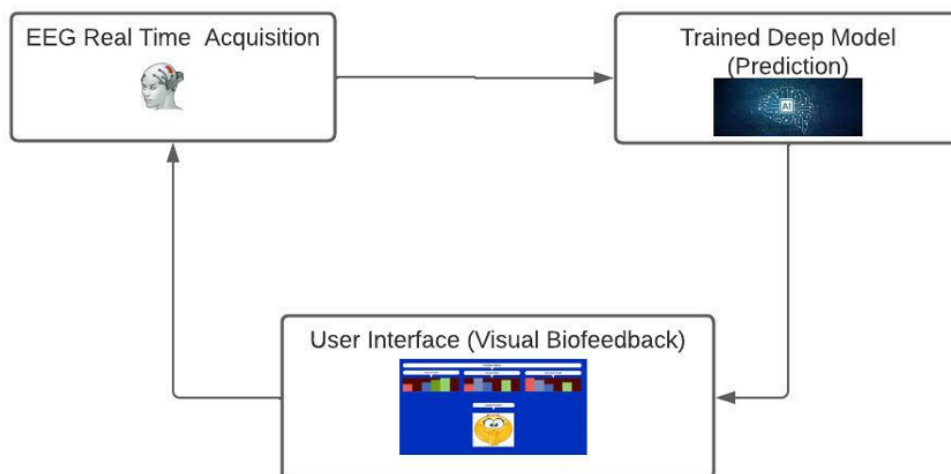


Fig. 4.18 System architecture overview.

The predicted Valence, Arousal, and Dominance values are further accumulated through a spatial coding block (`ml.spatial`<sup>8</sup>) in `max/msp` to calculate the most recurring emotions. The prevalent emotion is displayed through an explicit emoji. Table 4.4 shows the threshold values for valence and arousal to obtain the corresponding emotional quadrant related to the Russell model.

<sup>8</sup><https://www.benjamindaysmith.com/ml-machine-learning-toolkit-in-max>

Table 4.4 Global emotion assignment from V/A thresholds.

Valence	Arousal	Global Emotion
> 2.5	> 2.5	Happy
> 2.5	< 2.5	Relaxed
< 2.5	< 2.5	Sad
< 2.5	> 2.5	Angry

#### 4.4.10 Prototype Evaluation and Implications

We introduced a real time EEG regression based emotion recognition system that continuously estimates Valence, Arousal, and Dominance. The 1D CNN achieves high adherence ( $R^2=0.93$ ) on DREAMER using only four frontotemporal electrodes (AF3/AF4/T7/T8), offering a practical and minimally invasive hardware implementation. The reported metrics are based on a small cohort of users and require validation in intersubjective contexts, such as by applying the LOSO training strategy to a larger cohort. Future directions include adaptive artifact suppression, additional training strategies, and analysis of connectivity patterns between different brain areas. We also plan to evaluate our pre-trained models in clinical environments and real-world scenarios.

## 4.5 ARIEL: Brain Computer Interfaces meet Large Language Models for Emotional Support Conversation

### 4.5.1 Introduction

ARIEL merges BCI based emotion recognition with large scale linguistic models (LLMs) to dynamically perceive and respond to a user’s changing emotional state. Specifically, leveraging the LLaMA-2 model, ARIEL adapts its response strategy to real time affective inference, aiming to foster empathetic and contextualized conversation. A structured prompting scheme anchors the system’s dialogue logic so that responses remain consistent, supportive, and personalized based on the user’s emotional state.

The integration of BCI into affective computing highlights the promise of EEG for monitoring and interpreting emotions, with potential impact in neurorehabilitation and psychological support. Rapid progress in language models, many conversational systems still lack robust, grounded affect understanding. ARIEL addresses this gap

by uniting (i) BCI based emotion recognition and (ii) conversational AI to target the Emotional Support Conversation (ESConv), such as alleviating user distress through dialogue.

### Contributions.

1. **A unified framework** that couples EEG based emotion recognition with LLMs to provide real time emotionally supportive conversation.
2. **A detailed design and implementation** of ARIEL, including workflow, data handling, prompting strategy, and model components that sustain supportive interactions over time.
3. **A feasibility demonstration** via a controlled study illustrating end to end functionality and outlining pathways for in-vivo clinical applications.

### 4.5.2 Related Work

This work lies in the intersection between the field of LLM, employed as Conversational Agents (CAs), and BCI, which enables the acquisition of EEG signals and thus their processing to recognize human states like emotions. Although both domains are highly novel in the research community, reflected by the massive number of proposals shared as a preprint version, we dissect below only those works published after being peer reviewed into two separate subsections.

### 4.5.3 Brain Computer Interface

Atkinson *et al.*[23] proposed a feature based emotion recognition model for BCI systems using the DEAP dataset. The methodology integrated mutual information based feature selection with classifiers to improve in vivo validation across the Valence–Arousal dimensional space. The proposed support vector machine (SVM) achieved an accuracy of 73 for both valence and arousal recognition tasks.

Islam *et al.*[147] proposed a CNN based method using the Pearson correlation between EEG subband channels from the DEAP dataset. They tested binary and three level valence/arousal recognition, showing that using only the upper-triangular PCC reduced complexity without affecting accuracy. The model achieved 78.22 valence and 74.92 arousal accuracy.

Zheng *et al.*[349] transformed one dimensional EEG signals into two dimensional spatial representations and proposed an adaptive neural tree (ANT) model to improve

interpretability in emotion recognition. Their ANT architecture, optimized through a reinforcement learning exploration strategy, required no manual feature extraction. On the DEAP dataset, the model achieved accuracies of 99.12, 98.95, and 97.58 for binary, four-class, and eight-class tasks, respectively.

#### 4.5.4 Emotional Support Conversation with Large Language Models

Tu et al. [316] introduced MISC, a mixed strategy aware model that integrates COMET [38] for emotional support conversations. This framework formulates response generation as a probability distribution over a strategy codebook, guiding training to achieve coherent and strategically sound supportive dialogues across multiple evaluation metrics. In contrast, Peng et al. [256] investigated the use of the BlenderBot model within a framework for generating empathetic responses based on users' emotional cues. Their system adapts dynamically to conversational context and produces responses that promote positive emotional states.

Building on this direction, Liu et al. [193] extended the ESConv task to multimodal inputs (text, audio, and video) and proposed FEAT, a method that incorporates an emotion aware transformer and leverages fine grained emotional knowledge across modalities. FEAT achieved state of the art performance in generating supportive and emotionally aligned responses.

Despite these advances, such systems rely solely on textual interactions, limiting their generalizability to unseen emotional contexts. Moreover, the lack of reliable affective sensing constrains their capacity to detect complex emotional states. De Freitas et al. [83] underscored these limitations in companion AI and mental health applications, noting that current systems often fail to recognize distress or provide appropriate support, occasionally leading to negative user reactions.

Despite these advances, such systems rely solely on textual interactions, limiting their generalizability to unseen emotional contexts. Moreover, the lack of reliable affective sensing constrains their capacity to detect complex emotional states. De Freitas et al. [83] underscored these limitations in companion AI and mental health applications, noting that current systems often fail to recognize distress or to provide appropriate support, which can lead to negative user reactions.

This study proposes integrating EEG based emotional recognition via a BCI with a large language model fine tuned for emotional support dialogues. This approach improves the accuracy of emotion detection and enables a conversational agent to help

users mitigate distress and promote positive predictions, both emotional and aggregate, through dialogue in the emotion domain.

### 4.5.5 ARIEL Framework

The ESConv task can be decomposed into two subproblems: (i) detecting user emotions in real time and (ii) generating supportive linguistic responses. Although linguistic sentiment analysis models can identify affective cues in text [29, 327], they are less reliable for linguistic texts [88]. Furthermore, highly specialized sentiment analysis models risk overfitting to narrow topics and failing to generalize across human dialogue [7].

To address these limitations, ARIEL leverages (a) the EEG signal via BCI for emotional state recognition (b) and the conversational efficiency of modern LLM. Figure 7.1 presents the overall workflow consisting of four components: the neuro linguistic interface, emotion recognition, the prompt formatter, and the generative LLM.

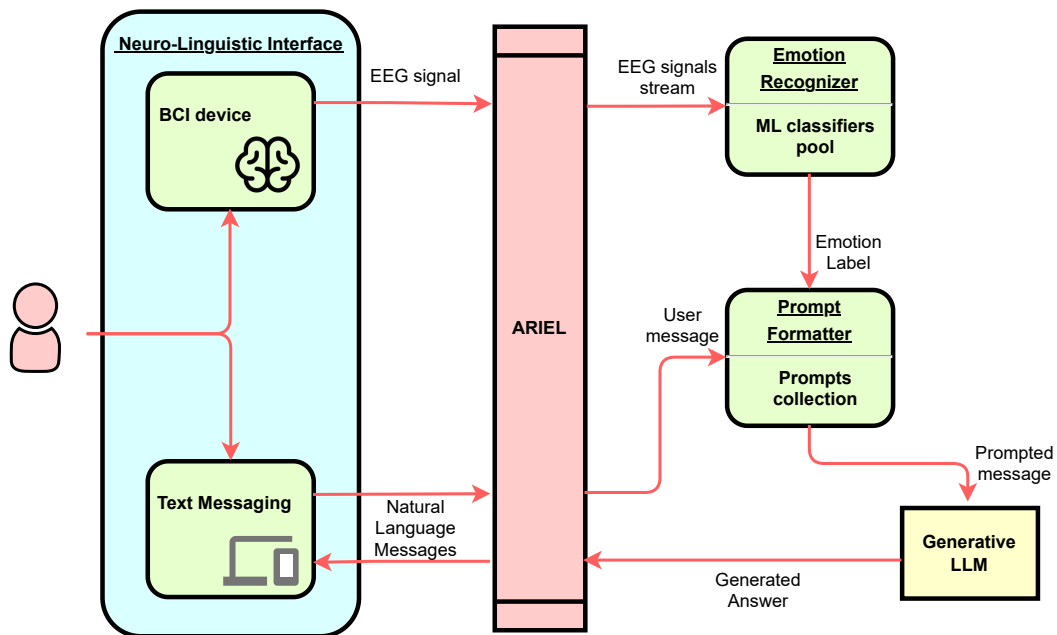


Fig. 4.19 ARIEL workflow. The Neuro Linguistic Interface (a) collects EEG and text; the Emotion Recognizer (b) infers an emotion label from EEG streams; the Prompt Formatter (c) merges the recognized label with the user’s message, constructing a conversation state prompt; the LLM (d) generates a supportive response. The session ends when a positive state is reached.

### Neuro Linguistic Interface

This bidirectional interface mediates all user interaction. It ingests two synchronized streams: (i) an EEG signal acquired via a non-invasive BCI and (ii) natural language

messages typed by the user. EEG conveys reliable affective information from cortical dynamics, whereas text provides topic and narrative content (and sometimes sentiment), enabling the system to both understand what the user says and sense how the user feels. The interface continuously buffers EEG from session start; chat messages are registered as submitted.

This bidirectional interface mediates all user interactions. It receives two synchronized streams: (i) an EEG signal acquired via a non-invasive BCI and (ii) natural language messages typed by the user. EEG transmits reliable affective information from cortical dynamics, while text provides narrative and thematic content, allowing the system to both understand what the user is saying and sense how they are feeling. The interface continuously stores EEG from the beginning of the session; chat messages are recorded as sent.

#### 4.5.6 Emotion Recognizer

The Emotion Recognizer estimates the user’s current affective state from streaming EEG. It receives time framed segments aligned to dialogue turns and outputs a discrete label (*Happy, Angry, Sad, Relaxed*) used downstream by the Prompt Formatter.

**Dataset description.** We followed the DEAP protocol [169] to collect training data at the Polytechnic University of Bari. Thirty participants provided informed consent . EEG was recorded with Muse 2<sup>9</sup> while subjects viewed 40 music videos. Data were anonymized, securely stored, and organized into 40 subject folders, each containing 40 trials.

**Preprocessing.** The preprocessing pipeline applies a 1-40 Hz bandpass filter; artifacts are detected using a KNN anomaly scorer trained on the filtered data. Evidence exceeding the threshold is removed using MEGKit’s artifact reduction algorithm. The data is structured in the MNE compatible format. Clean EEG signal epochs are considered for emotion prediction.

**Models and validation.** We compared LR, Extreme Gradient Boosting (XGB), RF, and SVM classifiers under a 5 fold GridSearchCV<sup>10</sup> with LOSO cross validation [175]. We trained 120 specialized models, 30 per emotion across four quadrants of Russell’s

---

<sup>9</sup><https://choosemuse.com/products/muse-2>

<sup>10</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

circumplex, then aggregated predictions via majority vote. SVM provided the best model accuracy for Happy, Angry, Sad, and Relaxed: 0.77, 0.78, 0.76, and 0.80, respectively.

### **4.5.7 Prompt Formatter**

The Prompt Formatter combines the recognized emotion label with the user's last message and conversation history to generate a concise, conditional prompt. This dynamic prompting scheme selects prompt templates based on the conversation state, starting with a role-play opening that instructs the LLM to act as a virtual assistant supporting users during emotional distress, explicitly emphasizing that affect is measured with BCI. The prompts evolve step by step to maintain consistency and goal alignment. [344].

### **4.5.8 Controlled Example**

To illustrate the framework's behavior, we report a controlled pilot study with a user (Figure 4.20). A user (Eric) begins a session wearing a BCI. The Emotion Recognizer predicts a feeling of Sadness. ARIEL opens with an empathic prompt and directs the conversation to personally meaningful topics (music). As the EEG derived emotion indicators progress from Sadness to Neutrality and then to Happiness, ARIEL adapts its prompts, maintaining engagement until a positive state is reached.

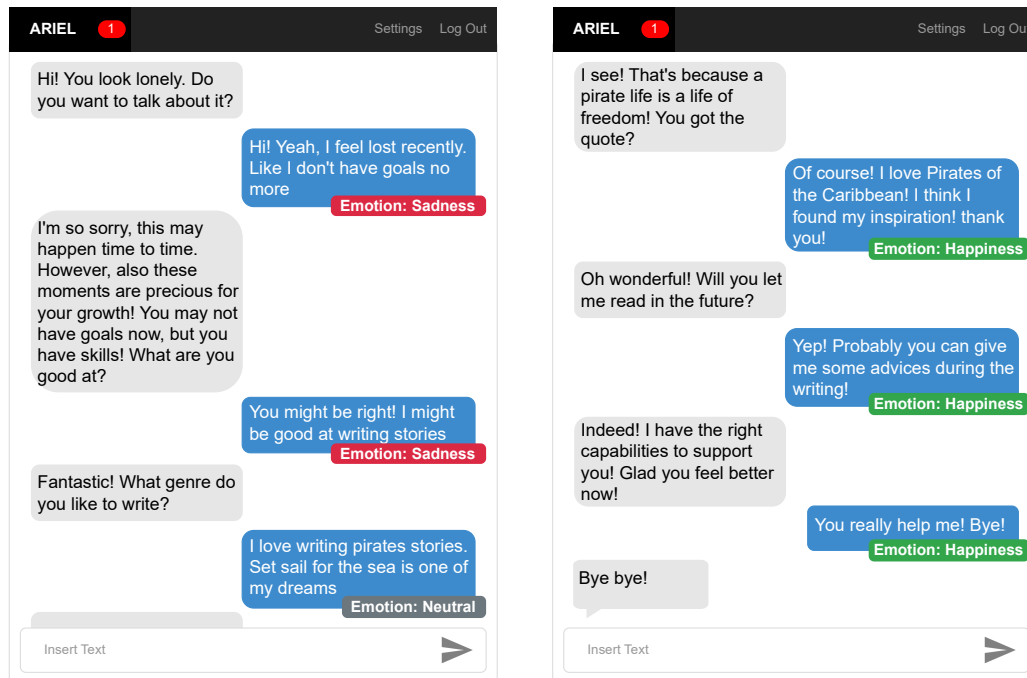


Fig. 4.20 Example conversation with ARIEL. The dialogue begins in *Sad*, then gradually shifts to a positive affect as measured by the BCI pipeline (red balloons), with text exchanges (grey/blue) guided by dynamic prompts.

#### 4.5.9 Discussion and Future Directions

ARIEL demonstrates how physiologically based affective perception can be combined with cutting-edge conversational models to address ESConv with responsive, emotionally aware dialogues. The proposed modular architecture extends to clinical and support settings (neurorehabilitation, counseling, triage), where objective affective signals can improve safety and personalization. Future work will include a user study on a much larger user population, extensive model customization, and formal outcome assessment (mental health, engagement, and work performance scales). We will also examine the ethical implications related to privacy, consent, transparency, and the sociotechnical dynamics of intimate human-machine interaction in personalized emotional care.

# Chapter 5

## EEG for Motivation and Neurorehabilitation

### 5.1 Machine Learning Classification of Motivational States: Insights from EEG Analysis of Perception and Imagery

#### 5.1.1 Introduction

This work examines the neural correlates of motivational states, desires, needs, and cravings under perception and imagery conditions using electroencephalography EEG and machine learning ML. While electrophysiological markers of motor and perceptual imagery are relatively well documented, the signatures of motivational imagery remain comparatively unexplored. We address this gap by classifying and comparing twelve motivational states recorded with medium density EEG (14 and 18 channels), and by quantifying the effect of electrode density on decoding performance. Across LOSO evaluations, perception consistently yields higher accuracy than imagery. Primary needs and somatosensory states exhibit stronger, clearer patterns during perception, with peak accuracy = 0.88 with 18 channels, whereas imagined states are more variable. Increasing electrode count from 14 to 18 provides only marginal gains and does not significantly change overall performance.

Understanding how the brain represents imagined content compared to perceived content is fundamental to cognitive neuroscience [9, 91, 92, 232, 302]. Numerous studies have described the electrophysiological markers of motor, perceptual, and mental images [99, 231, 310]. In contrast, a smaller number of studies have started to

analyse motivational images, particularly in the context of food or substance craving. Interest in these topics spans both basic science and applications to BCI, including potential communication pathways for patients with severe motor impairments or disorders of consciousness [43, 158, 184, 185, 216, 269, 320].

Event Related Potentials (ERP) studies have identified components modulated by craving and affect (P1, P2, P300, LPP, N200, N170, N400) and have shown that imagined stimuli can evoke measurable neurophysiological responses [47, 127, 277]. In parallel, ML methods have proved effective in extracting discriminative patterns from EEG for a range of affective and cognitive states [6, 58, 69, 70, 163, 280, 290, 312]. However, a systematic comparison of *motivational* states across perception and imagery, as well as an assessment of the incremental value of increased electrode density in this setting, is still lacking.

This study fills that gap by: (i) constructing a multiclass analysis across twelve motivational states under perception and imagery; (ii) formulating a comprehensive pairwise (binary) classification protocol to probe separability within and across conditions; (iii) evaluating the effect of 14 versus 18 EEG channels on performance. Our results indicate that perception outperforms imagination across the board, with somatosensory and primary need states being particularly robust during perception. Furthermore, modest increases in electrode density do not significantly alter overall accuracy distributions.

### 5.1.2 Related Work

BCI translate brain activity into control signals for external systems [118, 330], enabling assistive communication, neurofeedback, rehabilitation, and interactive media [6]. Classical EEG-BCI paradigms include motor imagery, slow cortical potentials, sensorimotor rhythms, and ERP based protocols (P300 spellers) [54, 139, 161, 176, 215]. ERP based frameworks have revealed systematic spatiotemporal patterns for perception across auditory/visual domains [269]. Subject specific morphology, multi channel variability, and limited data often complicate fully rule based systems, motivating data driven approaches [96, 184–186, 268].

Deep learning has shown promise for ERP classification in constrained settings [196, 292], though interpretability and data demands remain practical concerns. Classical linear methods (Linear Discriminant Analysis (LDA), SVM) and modern CNN methods have each achieved strong results in specific tasks [35, 53, 346]. Somatosensory and auditory/tactile imagery have also been explored for BCI [105, 337, 345]. Systematic BCI studies focused on motivational states, such as desire for food, warmth, or music, despite broader affective decoding successes in neuroimaging [94, 294].

### 5.1.3 Problem Statement and Research Questions

We investigate the EEG signatures of twelve motivational states under two conditions: perception (viewing pictograms) and imagery (generating the status internally at the appropriate time). From an ML perspective, this is a multiclass problem. To robustly probe separability and generalization across subjects, we cast it as a comprehensive set of pairwise binary tasks. With  $n = 12$  classes, the number of unordered pairs is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{with } n = 12, k = 2 \Rightarrow \binom{12}{2} = 66, \quad (5.1)$$

leading to 66 pairwise problems per condition and a total of 132 experiments (imagery and perception).

We address:

1. **RQ1 (Intra condition separability):** Within perception and within imagery, which state pairs are most separable?
2. **RQ2 (Inter condition comparison):** How do pairwise patterns differ between perception and imagery?
3. **RQ3 (Electrode density):** Does increasing the number of electrodes (14  $\rightarrow$  18) significantly affect performance?

### 5.1.4 Dataset, Participants, and Protocol

**Participants.** We recorded EEG from 20 right handed healthy students (8 males, 12 females; mean age 23.2, SE 1.7), normal/corrected vision, no neurological/psychiatric disorders, and no psychoactive medication. Procedures were conducted in accordance with institutional ethics (protocol RM-2020-242); all participants provided informed consent.

**Stimuli and Conditions.** Stimuli were drawn from a validated Motivational Pictionary comprising 60 color plates depicting 12 motivational states (5 variants/state) [270]. States spanned (i) primary homeostatic needs (hunger, thirst, sleep), (ii) somatosensory sensations (hot, cold, pain), (iii) affective states (happy, sad, fear), and (iv) secondary needs (move, play, listen to music). Each plate depicted a character with a pictorial cloud indicating the relevant need or desire; luminance was equated across micro/macro categories.

**Experimental Procedure.** A high density cap (10<sup>5</sup> system) was applied; impedances were kept < 5 k $\Omega$ . Participants sat in an anechoic chamber, staring at a central cross and minimizing movement. The experiment comprised 10 runs of 36 stimuli (total 360). Each stimulus was shown for 2000 ms, followed by a blank interval of 900  $\pm$  100 ms. A yellow frame (2000 ms) cued *imagery* of the preceding content; IT was 150  $\pm$  50 ms. Stimuli were 640  $\times$  480 px (6 $^{\circ}$ 47'16"  $\times$  4 $^{\circ}$ 46'37" visual angle) at 114cm viewing distance. Written instructions emphasized vivid, personal imagery; a short practice preceded recording. Figure 5.1 schematizes the procedure.

**Acquisition.** EEG was recorded from 128 sites (10<sup>5</sup> system) plus hEOG/vEOG using an EEProbe (ANT; Synamps amplifiers), referenced to averaged mastoids, at a sampling rate of 512Hz, with a bandpass filter of 0.16–70Hz. Data were stored via Cognitrace.

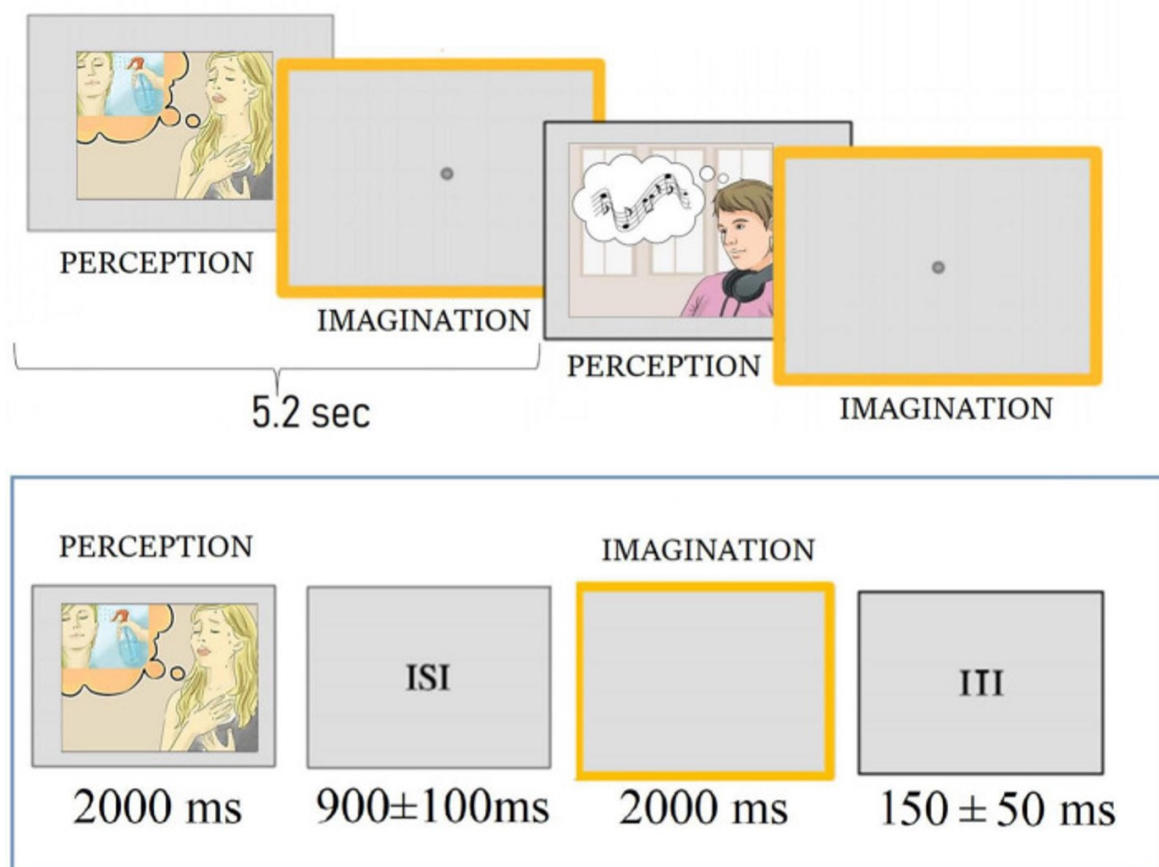


Fig. 5.1 Schematic of the perception-imagery protocol (adapted from [270], CC BY).

### 5.1.5 Preprocessing and ERP Construction

Trials with amplitudes exceeding  $\pm 50 \mu V$  were rejected before averaging. Epochs were extracted from  $-100$  to  $+1200$ ms relative to stimulus onset and baseline corrected using the  $[-100, 0]$  ms window [331]. Separate ERP were computed for perception (pictograms) and imagery (frame cue), time locked to the corresponding onsets. Additional acquisition metadata are summarized in Table 5.1.

Table 5.1 EEG acquisition summary.

<b>Participants</b>	20 right handed students (8M/12F), mean age 23.2 (SE 1.7)
<b>Montage</b>	128 electrodes (10–5 system) + hEOG/vEOG; averaged mastoids reference
<b>Sampling / Filter</b>	512Hz; 0.16–70Hz acquisition band-pass
<b>Artifact policy</b>	Automatic rejection at $\pm 50 \mu V$ prior to averaging
<b>Epochs / Baseline</b>	$[-100, 1200]$ ms; baseline $[-100, 0]$ ms
<b>Conditions</b>	Perception (pictograms) and imagery (frame cue)
<b>Stimuli</b>	12 states $\times$ 5 variants = 60 plates; each repeated 6 times (360 total)

### 5.1.6 Electrode Sub selection and Configurations

The initial data structure includes 126 EEG channels, encompassing the Mastoidei, Vertical, and Horizontal EOG channels. We aimed to reduce the number of channels required for the BCI domain. For this purpose, we considered the channel configuration of a well-known BCI, such as the Emotive headset. Figure 5.2 illustrates the electrode positions according to the 10-20 international system. Additionally, we explored the feasibility of a BCI with four extra electrodes positioned along the head’s centerline at Fz, Cz, Pz, and Oz. Figure 5.3 shows the placement of these additional electrodes. It should be noted that a BCI version with these four additional electrodes is not currently available on the market. Hence, two potential EEGcaps were evaluated with 14–and 18–channel configurations.

The latter is a hypothetical extension motivated by prior evidence that midline/fronto-central and centro-parietal sites are informative for imagery and memory processes [186, 206, 270]. For both configurations, epochs spanned  $[0, 1199]$ ms (baseline segment excluded). Data structures were prepared in MNE [123].

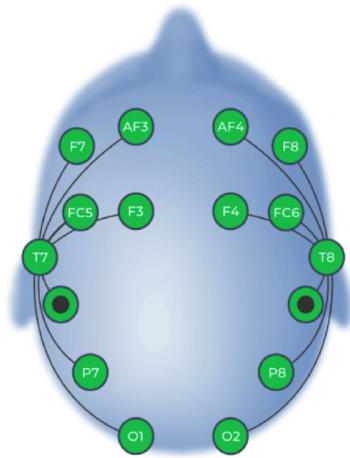


Fig. 5.2 14 channel Emotiv channels subset (10–20 system). Black circled electrodes indicate CMS/DRL reference sites.

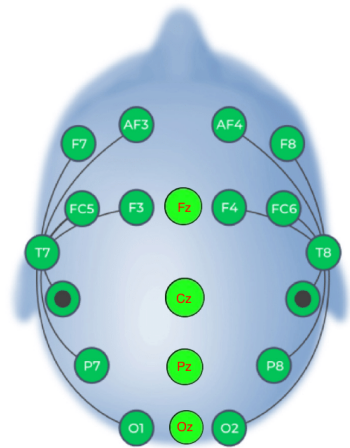


Fig. 5.3 18 channel variant with added midline Fz, Cz, Pz, Oz (hypothetical extension). Black circled electrodes indicate CMS/DRL.

### 5.1.7 Modeling Strategy

**Pairwise Design.** For each unordered pair among the 12 states (66 pairs), we trained binary classifiers separately for perception and imagery. This design probes fine grained separability within each condition and enables direct inter condition comparisons.

**Time Sliding Estimation.** Inspired by the time sliding fMRI/EEG strategy, we trained classifiers over successive samples within each epoch to identify ROI. This identifies early versus late contributions to separability within the range of [0, 1199]ms.

**Classifiers and Validation.** We evaluated SVM, RF, and a shallow feedforward Artificial Neural Network (ANN). Hyperparameters were tuned by grid search with inner cross validation and appropriate scaling (MinMax/Standard/Robust). To assess subject independent generalization, we used the LOSO method: for each hold out subject, models were trained on the remaining 19 subjects. For each pair, predictions were concatenated across all 20 LOSO folds to compute the final metric. For global method comparisons, one way ANOVA and Tukey’s HSD were used on F1 distributions across pairs and conditions.

**Performance Metrics.** Given TP, TN, FP, FN:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (5.2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (5.3)$$

$$\text{F1} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}. \quad (5.4)$$

## 5.1.8 Results

**Global Pattern.** Across all conditions and pairs, SVM provided the most reliable performance; RF and the Feed Forward Neural Network (FFNN) were competitive on specific pairs but did not outperform SVM in aggregate. In general perception consistently outperformed imagery.

**Condition and Density Effects.** Fig. 5.4 display probability density plots of pairwise F1. Clear separation between perception and imagery for both the 14- and 18 channel conditions is shown. Bonferroni corrected  $t$  tests confirmed perception  $>$  imagery for 14 channels ( $t = -6.06, p < 0.0001$ ) and 18 channels ( $t = -7.01, p < 0.0001$ ). The differences between 14 and 18 channels were not significant for either imagery ( $t = 0.27, p = 0.78$ ) or perception ( $t = -0.49, p = 0.61$ ).

**Best Pairs.** Figures 5.55.6 report F1 heatmaps. Table 5.2 summarizes top accuracies for selected pairs across categories and configurations. Perception achieved the highest values overall (music\_cold = 0.88 with 14 and 18 channels). Imagery peaks were lower (cold\_fear) = 0.82 at 14 ch; sadness\_sleep = 0.75 at 18 ch.

**Intra condition Highlights (RQ1).** In the context of imagination with a 14 channel device, cold\_fear stood out with an accuracy of 0.82 and happiness\_hunger with

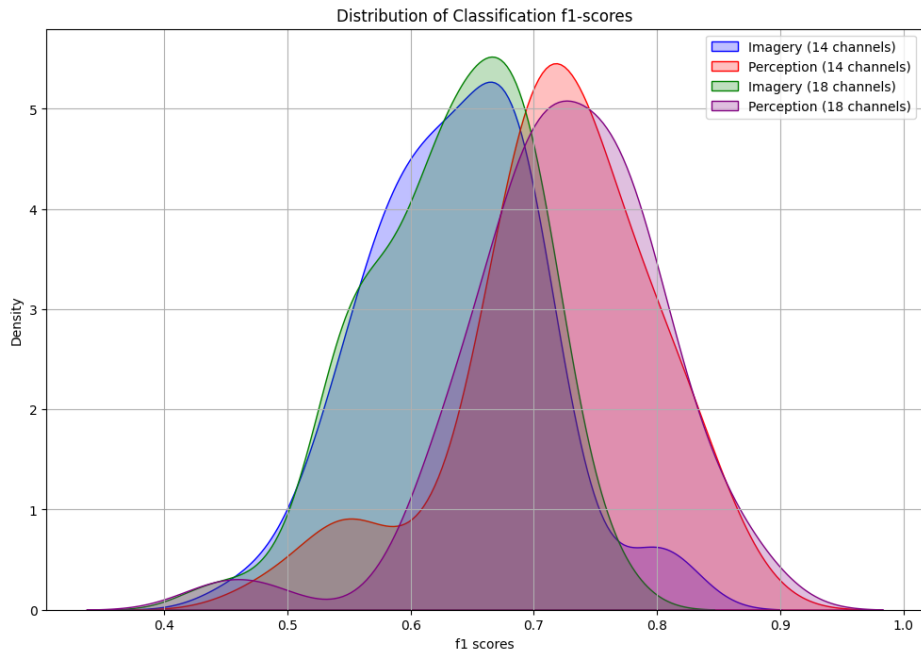


Fig. 5.4 Distributions of F1 scores across all pairwise classifiers under perception/imagery and 14/18 channel configurations.

	fear	happiness	hot	hunger	movement	music	pain	play	sadness	sleep	thirst	
	0.81	0.52	0.71	0.73	0.5	0.67	0.59	0.7	0.7	0.61	0.54	cold
		0.67	0.57	0.8	0.63	0.68	0.7	0.67	0.54	0.73	0.68	fear
			0.59	0.8	0.63	0.57	0.63	0.6	0.68	0.7	0.57	happiness
				0.63	0.67	0.46	0.6	0.63	0.7	0.67	0.52	hot
					0.63	0.6	0.67	0.6	0.63	0.67	0.71	hunger
						0.6	0.63	0.59	0.55	0.68	0.67	movement
							0.57	0.63	0.64	0.59	0.58	music
								0.67	0.68	0.64	0.59	pain
									0.74	0.68	0.59	play
										0.73	0.63	sadness
											0.55	sleep

Fig. 5.5 F1 14 channels imagery

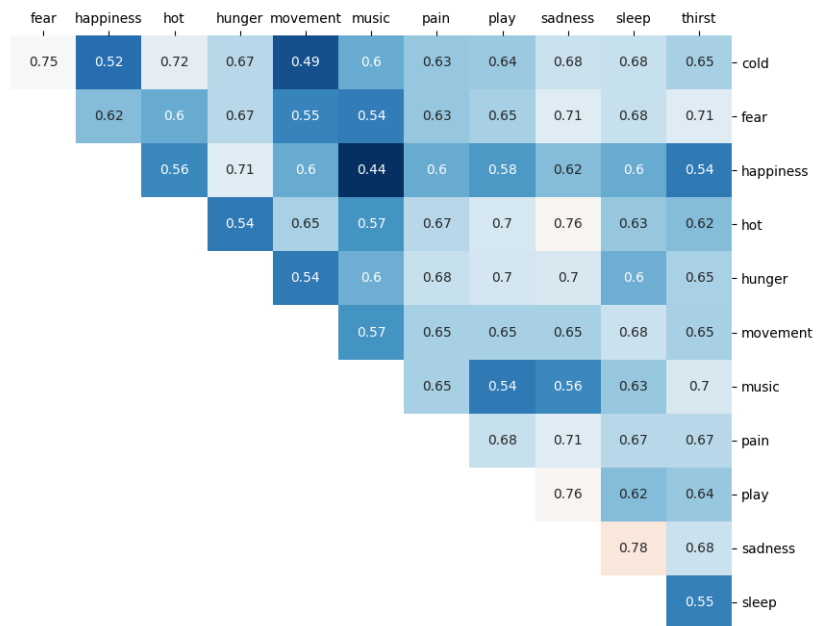


Fig. 5.6 F1 18 channels imagery

an accuracy of 0.80. Music hot proved difficult with an accuracy of 0.47. With 18 channels, sadness sleep and cold fear achieved an accuracy of 0.75. Within perception, several pairs exceeded 0.80 accuracy. Music / cold peaked at 0.88 for both 14 and 18 channels.

**Inter condition Contrasts (RQ2).** Perception > imagery across most categories, consistent with stronger, earlier sensory driven activations and with prior ERP literature showing larger components during perception than imagination [93, 113, 171, 213, 267, 270]. For example, thirst\_pain was 0.80 (perception, 14 ch) vs 0.65 (imagery, 14 ch); hot\_pain was 0.85 (perception, 18 ch) vs 0.68 (imagery, 18 ch).

**Electrode Density (RQ3).** Across all pairs, moving from 14 to 18 channels did not significantly alter the overall performance distributions for either condition. However, best pair summaries sometimes showed minor improvements with 18 channels. This suggests diminishing returns from modest density increases when using mid sparse montages and LOSO generalization.

Table 5.2 Top accuracies by condition and configuration. Blank entries indicate pairs not among the top performers for that setting.

Category Pair	State Pair	Perception (14)	Imagery (14)	Perception (18)	Imagery (18)
Primary-Primary	hunger_sleep	0.80			
Primary-Primary	thirst_sleep			0.82	
Primary-Somatosensory	thirst_pain	0.80	0.65		
Somatosensory-Somatosensory	hot_pain			0.85	0.68
Primary-Affective	hunger_sadness			0.72	
Primary-Affective	hunger_happiness		0.80	0.70	0.70
Primary-Affective	sleep_sadness				0.75
Primary-Secondary	sleep_movement		0.70		
Primary-Secondary	thirst_music				0.70
Somatosensory-Affective	cold_fear		0.82		0.75
Somatosensory-Affective	hot_happiness	0.80			
Somatosensory-Secondary	hot_music		0.47		
Somatosensory-Secondary	music_cold	0.88		0.88	
Affective-Secondary	happiness_movement			0.65	
Affective-Secondary	fear_play	0.82		0.80	
Affective-Secondary	sadness_play		0.75		
Affective-Secondary	sadness_movement	0.65			
Secondary-Secondary	movement_play				0.62

### 5.1.9 Discussion

Performance patterns align with neurophysiological plausibility. Perception benefits from direct stimulation of early sensory cortices, producing stronger, more reliable ERP; imagery relies more on higher order networks and memory systems, typically yielding lower amplitudes and higher variability [93, 113, 171]. The particularly robust decoding of somatosensory states (cold, pain) and positive affect (happiness) likely reflects generators closer to the scalp and well studied cortical circuits [270]. In contrast, hunger, thirst, sleep involve deep structures that contribute weakly to scalp potentials [244], which plausibly explains the lower separability for some primary need comparisons, especially during imagery.

Electrode density findings are practical: for medium sized BCI, a carefully chosen 14 channels montage may capture most discriminative content for these tasks. Adding a few midline sites (18 channels) provides only incremental benefits, consistent with the idea that temporal resolution, task design, and robust cross subject modelling can outweigh small spatial gains in sparse EEG.

### 5.1.10 Limitations

The binary pair design enables fine grained analysis but scales quadratically with the class count (132 experiments here). Future work should evaluate multiclass models and calibration light personalization. The sample size of 20 is typical for ERP studies but limits power for subtle effects. Larger cohorts would strengthen the generalizability

of LOSO. Finally, the 18 channel montage is a hypothetical extension; hardware with flexible midline coverage would facilitate confirmatory studies.

### **5.1.11 Conclusion**

This study provides a systematic, subject independent comparison of motivational perception and imagery using EEG and ML. Perception consistently shows higher separability than imagery. Somatosensory and certain affective states are particularly robust, and modest electrode density increases have a limited impact on overall performance distributions. These results contribute to the design of a practical, medium density BCI system for assessing motivational states and highlight the potential for improving image decoding through task design, personalization and modelling.

## **5.2 Combining Mental States Recognition and Machine Learning for Neurorehabilitation**

### **5.2.1 Introduction**

In recent years, significant progress has been made in automated control and monitoring applications. From this perspective, BCI technology provides a framework for interacting with computers and external devices by analyzing electroencephalographic EEG signals. The rapid development of AI has further encouraged researchers to employ Machine Learning ML techniques for the classification of various EEG related tasks within BCI systems.

Indeed, ML methods have been increasingly integrated into numerous research domains, yielding remarkable scientific advancements in recent years. ML approaches have become integral to rehabilitation and assistive technologies [179, 352], supporting tasks ranging from biofeedback monitoring [112, 289] to BCI based emotion and cognition analysis [20, 22].

Among emerging research directions, one of the most promising focuses on identifying variations in mental states and employing BCI based technologies for both motor and cognitive rehabilitation. However, there is still a lack of unified approaches in the literature addressing sound based biofeedback systems designed to detect emotional states using BCI. Sound driven rehabilitation frameworks have already shown significant effectiveness in neurorehabilitation contexts and hold strong potential for further enhancing biofeedback performance [276, 284].

The main contributions of this work can be summarized as follows:

- (i) We propose a novel biofeedback system based on BCI technology to detect two mental states of users: *Focus* and *Relaxation*;
- (ii) Depending on the detected state, the system generates either a *Focused* or *Relaxed* musical output;
- (iii) We demonstrate the applicability of the proposed system for potential integration into rehabilitation and assistive tasks, where the generated auditory biofeedback can support motor engagement.

The remainder of this paper is structured as follows. Section 6.3.2 reviews related work relevant to this study. Section 5.2.3 describes the materials and methods. Section 5.2.6 presents the proposed framework, while Section 7.4.10 reports the obtained results. Finally, Section 7.4.11 concludes the paper and outlines directions for future work.

## 5.2.2 Background and Related Work

Parivash *et al.* [251] proposed a system based on BCI, Virtual Reality (VR), and DL to detect three different neurostimulations during specific motor imagination tasks and generate the appropriate avatar virtual neurostimulations in the VR world. They analyzed the task of brain patient arousal to generate correct stimulations based on the instructed motor imagination task. The real-time system contains the following four units: (i) Preprocessing unit that employs a bandpass filter and discrete wavelet transform to prepare the sequenced EEG signals for feature extraction; (ii) a Feature selection unit that calculates the energy of the sequenced EEG signals from its decompositions and forms the normalized feature vectors for classification; (iii) MLP classification unit to detect three different types of neurostimulations and (iv) Logic unit to generate the correct virtual neurostimulations from the appropriate classifications and delete the incorrect classifications.

McMahon *et al.*[219] show a prototype development environment for game control. It is a low cost open source system based on BCI, VR, which uses real time processing of EEG ERD and ERS signals within the precentral gyrus (motor cortex), allowing the user to control a 3D object within a VR environment. The authors conducted a study on patients who needed neurorehabilitation. The feedback and results from the participants were enthusiastic, and the BCI, VR system worked as expected.

Karacsony et al.[156] present a real EEG via BCI system using a VR game as motivational feedback for stroke rehabilitation. The proposed game is based on targets that must be hit. If the subject successfully hits a target, it explodes. This gives feedback on a successfully imagined and virtually executed movement of hands or feet. A CNN with a unique process onset detection technique was used. This achieved better results than previous architectures on datasets from the offline PhysioNet database. Ten participants said the training was interesting, fun, and engaging. The system has achieved excellent results. A simultaneous approach was proposed for detecting trial onset. This optimized the classification for real-time performance. The experiment confirms the clinical potential of such a system by addressing the problem of low motivation in stroke neuro rehabilitation.

However, the current literature does not report the use of BCI and ML to generate music as a neurorehabilitation tool for patients with motor disorders. In summary, the preceding discussion demonstrates a significant gap in existing research. This work presents a state of the art advancement by introducing a novel approach that leverages ML and BCI to provide an acoustic neurorehabilitation tool.

### 5.2.3 Materials and Methods

We adopt a polyphony generation paradigm conditioned on Focus vs. Relaxation. Polyphony here denotes the coordinated superposition of four melodic lines.

The framework illustrated in Figure 6.2 comprises five main steps. Figure 6.2 outlines five stages. First, EEG is acquired via the Muse headset [24]<sup>1</sup>. Second, the data are preprocessed. Third/*Fourth*, the ML Engine estimates the mental state and instantiates the matched music generator and score visualizer. *Fifth*, a musician plays the generated score in real time, closing the biofeedback loop as the user listens and adapts (striving for higher rhythmicity to encourage movement).

### 5.2.4 Data Acquisition Device

We employ the Muse BCI, a dry electrode headset with four sensors placed according to the international 10–20 layout [136] at *AF3*, *AF4*, *TP9*, *TP10*, referenced to *FPz* (CMS/DRL), sampling at 256 Hz.

---

<sup>1</sup><https://choosemuse.com/muse-2/>

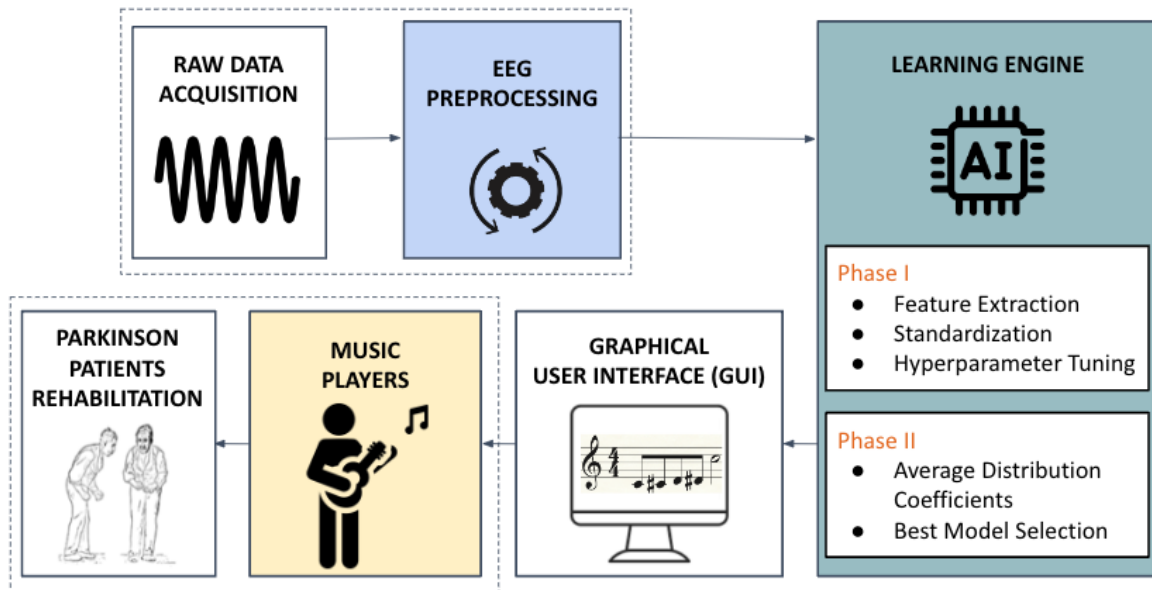


Fig. 5.7 System Architecture.

### 5.2.5 Polyphonic Structure Generation

The classifier output selects one of two rhythmic sets used to derive four voice-specific rhythmic profiles:

1. **Focus:** quarter ( $1/4$ ), eighth ( $1/8$ ), sixteenth ( $1/16$ ) notes and corresponding rests;
2. **Relaxation:** whole ( $4/4$ ), half ( $2/4$ ), quarter ( $1/4$ ) notes and corresponding rests.

Random permutations generate rhythmic strings per voice. Pitch material is constrained to user-selected scales. The framework enumerates all scales of size 3–11 over the 12-tone equal temperament [143, 153]. Figures 5.8–5.9 illustrate the GUI for selecting scale size and choosing (e.g.) among pentatonic options. All polyphony is realized over the selected scale.

#### Dataset Description

We use the public EEG brainwave mental state dataset [32], collected from four adults (two male, two female) across focus, neutral, and relaxation conditions.<sup>2</sup> Relaxation trials included soothing audio; focused trials involved a three-cup attention task. Due

<sup>2</sup>[https://github.com/jordan-bird/eeg-feature-generation/tree/master/dataset/original\\_data](https://github.com/jordan-bird/eeg-feature-generation/tree/master/dataset/original_data)

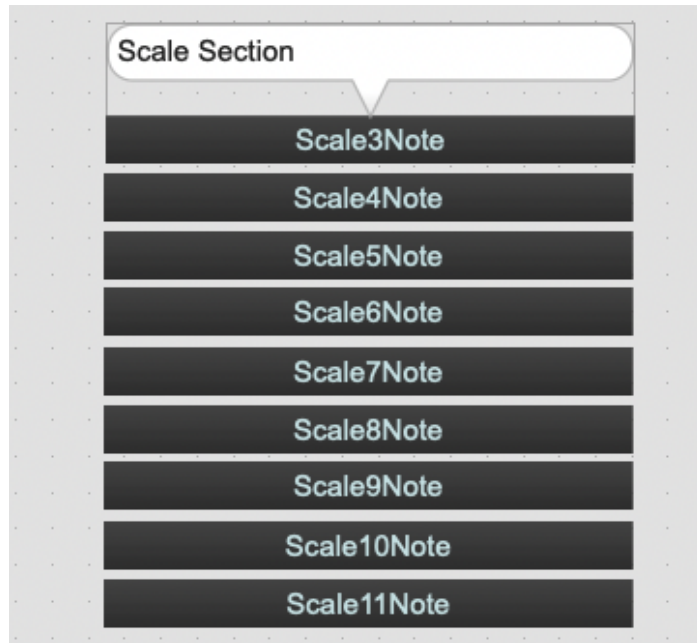


Fig. 5.8 GUI: selecting octave/scale cardinality.

Pentatonic Scales

1	1	2	3	4	5
2	1	9	10	11	12
3	1	8	10	11	12
4	1	8	9	11	12
5	1	8	9	10	12
6	1	8	9	10	11
7	1	7	10	11	12
8	1	7	9	11	12
9	1	7	9	10	12
10	1	7	9	10	11
11	1	7	8	11	12

play stop

noteout

Fig. 5.9 GUI: example submenu of pentatonic scales in one octave.

to noise in  $TP9/TP10$ , we limit analysis to  $AF3/AF4$ . Figures 5.10–5.11 show raw vs. filtered signals; artifacts include ocular (low frequency) and muscle (high frequency) components [313]. To mitigate these, we filter within the alpha band and proceed with  $AF3/AF4$  channels.

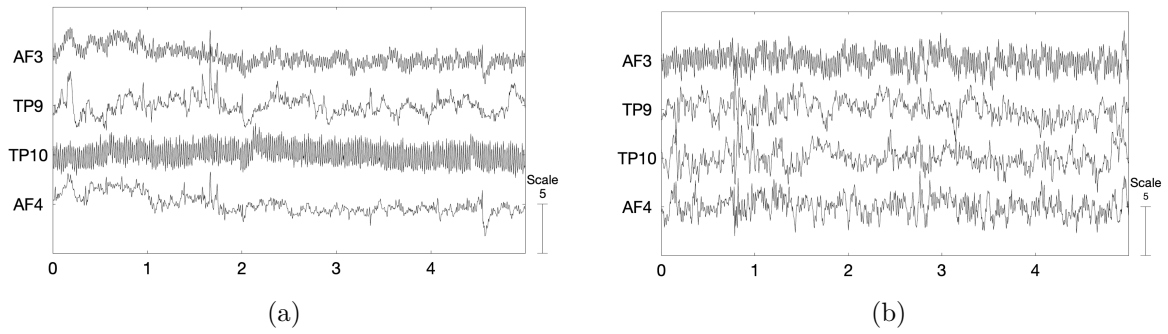


Fig. 5.10 Raw EEG during *Focus* [a] and *Relaxation* [b] before filtering; ocular and muscle artifacts are evident [313].

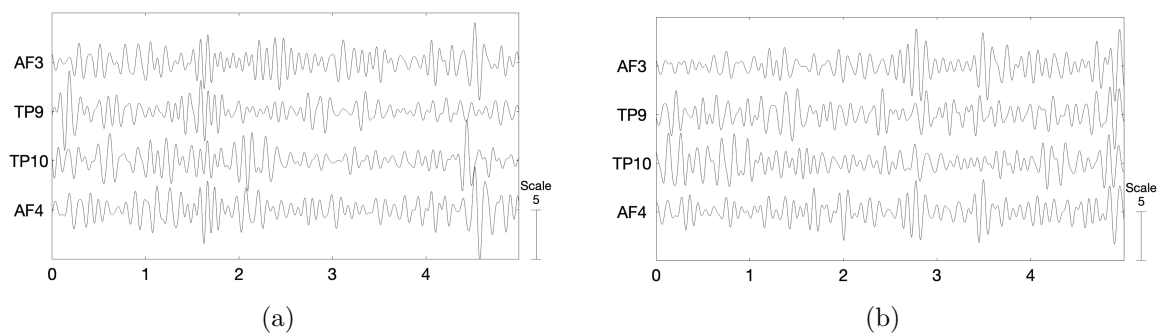


Fig. 5.11 Raw of EEG Signal during the Concentrating [a] and Relaxation [b] tasks, after the preprocessing pipeline.

### 5.2.6 Framework

The primary objective of the framework is to distinguish between two mental states: activation and relaxation, using an ML classifier. Based on the user's activation level, the system generates music in real time that corresponds to the identified mental state. The subsequent section details the preprocessing steps and the machine learning engine's operations.

### 5.2.7 Machine Learning Engine

The ML approach is structured into two distinct steps. In the first step, the MiniRocket algorithm [87] was applied using the implementation provided by the sktime library [201]. MiniRocket is a variant of the Rocket algorithm designed for feature extraction from time series data. Each time series is processed using random convolutional kernels followed by global max pooling.

Proportion pooling is applied to the convolutional output, resulting in two features per kernel. The extracted features encode information relevant to class membership, which can be modeled by a linear classifier. The initial process involves extracting features from each time series by applying convolutional filter kernels to multivariate EEG segments. These features are subsequently input to a Ridge classifier, selected as a linear method to address feature collinearity through L2 regularization and to perform feature ranking [150]. The dimensionality of the extracted features is determined by two hyperparameters: kernel size and kernel length. Hyperparameter optimization is conducted using Python pipelines, which include: (i) MiniRocket tuning (kernel size and length), (ii) RidgeClassifier tuning, and (iii) selection of normalization functions (MinMaxScaler, StandardScaler, RobustScaler). This process employs 5 fold cross validation and random search (50 iterations), implemented in Python using the Scikit-learn library [254]<sup>3</sup>.

The process of selecting the optimal EEG segment and tuning hyperparameters is repeated for different EEG signal lengths. Determining the appropriate signal length is critical for maximizing information extraction in real time classification. The maximum EEG epochs length is set at 5 seconds to ensure suitability for real time applications. Table 5.3 reports dataset shapes per window.

Accuracy, F1 score, precision, and recall have been computed to assess the models' performance. Accuracy is defined as follows:

---

<sup>3</sup><http://scikit-learn.org>

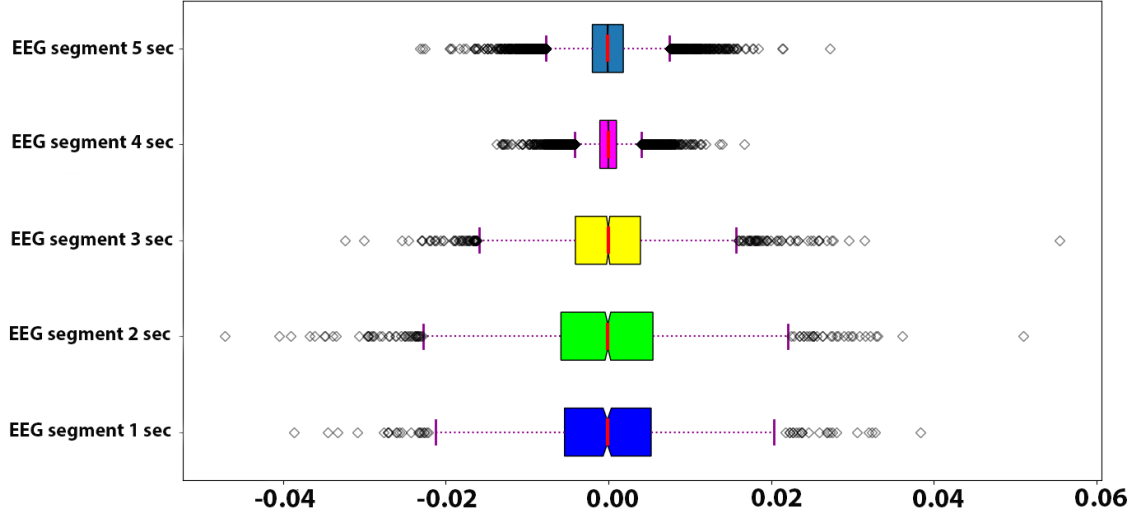


Fig. 5.12 Distribution of Ridge feature coefficients averaged across validation folds.

Table 5.3 Training snapshot by EEG segment in sec.

EEG Segment	Raw Trials	Channels	Samples
1	822	2	256
2	409	2	512
3	272	2	768
4	203	2	1024
5	162	2	1280

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \quad (5.5)$$

where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative.

Recall is defined as the proportion of correct positive classifications among all positive samples:

$$Recall = \frac{TP}{TP + FN} \quad (5.6)$$

Precision represents the ratio of correct positive classifications to the total positive classifications:

$$Precision = \frac{TP}{TP + FP} \quad (5.7)$$

Recall and accuracy are combined to form the F1 score:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5.8)$$

The main goal of the second phase is obtaining the average distribution of the coefficients of the features corresponding to the best model-RidgeClassifier. Figure 5.12 shows the feature coefficients averaged across k-fold validation rounds. Afterwards, the best predictive model and EEG segment are selected. We used a statistical criterion to reduce the set of 3,864 features based on the selection of the most significant features (lower 10th percentile and upper 90th percentile, given the symmetry of the model weights). The new set of reduced features is 774.

Consequently, the training dataset is redefined according to the new subset of extracted features. Finally, we trained three ML algorithms (SVM, RF and KNN) with hyperparameter tuning and k-fold strategy.

Table 5.4 Performance of the best models for different EEG segment lengths. Segment length is measured in seconds (sec).

EEG Segment Length (s)	Pipeline Step: MiniRocketMultivariate	Pipeline Step: Scaler	Pipeline Step: RidgeClassifier	Precision	Recall	F1-Score	Accuracy	Confusion Matrix
1	kernel max dilations: 352 num kernels: 1110	<i>RobustScaler</i>	$\alpha = 621.01$ max_iter = 891 random_state = 42 solver = auto $\alpha = 148.73$	class -1 = 0.73 class 1 = 0.77	class -1 = 0.70 class 1 = 0.79	class -1 = 0.71 class 1 = 0.78	0.75	51 22 19 73
2	kernel max dilations: 848 num kernels: 261	<i>StandardScaler</i>	max_iter = 769 random_state = 42 solver = auto $\alpha = 148.73$	class -1 = 0.78 class 1 = 0.78	class -1 = 0.69 class 1 = 0.85	class -1 = 0.74 class 1 = 0.81	0.78	25 12 7 39
3	kernel max dilations: 512 num kernels: 3910	<i>RobustScaler</i>	max_iter = 591 random_state = 42 solver = auto $\alpha = 0.78$	class -1 = 0.83 class 1 = 0.87	class -1 = 0.83 class 1 = 0.87	class -1 = 0.83 class 1 = 0.87	<b>0.85</b>	20 4 4 27
4	kernel max dilations: 384 num kernels: 12010	<i>MinMaxScaler</i>	max_iter = 16 random_state = 42 solver = auto $\alpha = 57.36$	class -1 = 0.75 class 1 = 0.76	class -1 = 0.67 class 1 = 0.83	class -1 = 0.71 class 1 = 0.79	0.76	12 6 4 19
5	kernel max dilations: 560 num kernels: 16910	<i>RobustScaler</i>	max_iter = 270 random_state = 42 solver = auto	class -1 = 0.91 class 1 = 0.77	class -1 = 0.67 class 1 = 0.94	class -1 = 0.77 class 1 = 0.85	0.82	10 5 1 17

## 5.2.8 Results and Discussion

Table 7.2 presents the best performing model for the 3 second EEG epochs. Figure 5.12 displays the coefficient distributions of the best models for each EEG segment length. Notably, models trained with 1, 2, and 3-second segments exhibit feature distributions with the greatest variance. These results indicate that implementing a percentile based feature selection criterion could enhance model performance.

Table 5.5 Results of 5-fold cross-validation applied on Test set. MA = Mean Accuracy; STD = Standard Deviation.

<i>Model</i>	<i>Fold 1</i>	<i>Fold 2</i>	<i>Fold 3</i>	<i>Fold 4</i>	<i>Fold 5</i>	<i>MA</i>	<i>STD</i>
<i>SVM</i>	<b>0.8727</b>	<b>0.9454</b>	<b>0.9444</b>	<b>0.9259</b>	<b>0.8148</b>	<b>0.9006</b>	<b>0.0504</b>
<i>Random Forest</i>	0.8363	0.9636	0.8148	0.9074	0.7592	0.8563	0.0716
<i>KNN</i>	0.8545	0.8545	0.7037	0.9444	0.8333	0.8381	0.0773

In contrast, the coefficient distributions for the 4 and 5 second segments are more concentrated near zero, suggesting that feature impact is reduced relative to models using 1, 2, and 3 second segments. Table 7.2 and Figure 5.12 illustrate that the optimal final model configuration is achieved by selecting 3 second EEG segments.

Three candidate models were identified as optimal: SVM, RF, and KNN. SVM constructs a geometric separation hyperplane using support vectors. RF determines the decision boundary based on feature informativeness. KNN forms centroids according to feature similarities and differences. Table 5.5 presents the results of the 5 fold cross-validation strategy following the feature selection process described previously. SVM, with a mean accuracy of 0.90 and a standard deviation of 0.050, demonstrates the highest performance for two mental state recognition tasks in a 5 fold cross-validation during both training and testing step. This approach advances the state of the art [32] by preprocessing EEG signals to remove noise and artifacts, and achieving high classification performance using only alpha range variations.

In summary, SVM outperforms the other models due to its practical decision boundary definition and suitability for the data distribution. Compared to RF and KNN, SVM also demonstrates greater consistency between training and testing performance, shows no evidence of overfitting, and achieves the lowest standard deviation during the test phase.

### 5.2.9 Conclusion and Future Work

We presented a portable BCI+ML biofeedback tool that discriminates *Focus* from *Relaxation* and maps the decision to polyphonic rhythmic structures suitable for neurorehabilitation. SVM classifier achieved an average test accuracy of 0.90 on 5-fold CV. Future work will (i) extend to multi-class affective states, (ii) personalize models with transfer learning, (iii) integrate continuous arousal/valence estimation, and (iv) evaluate clinical efficacy with standardized usability and workload instruments.

# Chapter 6

## BCI for Neuro Musical Interfaces

### 6.1 NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms

#### 6.1.1 Introduction

Following the developments discussed above, this section introduces NeuralPMG, a system that integrates BCI based mental state detection with polyphonic music generation to enhance creativity and user engagement in composition. NeuralPMG extends prior lines of work by combining ML and BCI to classify users' mental states and using these classifications to shape the compositional process in realtime. The framework brings together Slonimsky's theoretical model, AI driven pattern generation, and an interface that fuses EEG sensing with hand motion input, yielding a fluid, end to end composition experience.

Beyond recognizing emotion, NeuralPMG aims to apply these device in artistic and therapeutic contexts. Users' mental states are mapped to musical structures, providing a responsive composition environment that adapts to ongoing cognitive and affective changes. This marks a substantive step in our research agenda, showing how BCI can act both as conversation support systems and as interactive composition tools suitable for neurorehabilitation and creative education.

In what follows, we detail NeuralPMG's design: EEG based classification of mental states, hand tracking for melodic material, and realtime generation of polyphonic textures. The system illustrates practical, user centered applications of BCI and AI across healthcare, education, and the creative industries.

### 6.1.2 Related Work

This section reviews studies that employ EEG signals as direct drivers of music generation.

Folgeri *et al.* [111] introduce a low cost BCI approach that aims to let users deliberately trigger specific musical notes via brain activity within a multimodal conditioning paradigm (audio, gesture, and visual cues). Their study proceeds in two phases: (i) recording unconscious EEG from multiple participants and transforming the resulting brainwaves into sound to assess whether subject specific or common EEG characteristics can be identified, and (ii) training users to consciously reproduce single notes using commercial BCI devices and targeted reinforcement (listening, gesture association, and note-specific visual prompts). The resulting application links each note to a unique gesture and color coded visual label, guiding users to generate the intended note from their EEG signals.

Hamadicharef *et al.* [128] present *I<sup>2</sup>R NeuroComm*, a P300-based [325] BCI system for interactive music composition. The interface allows users to insert or delete notes and to play short melodies directly from brain responses. The core contribution lies in enabling end to end composition through a novel Graphical User Interface (GUI) that displays both textual and staff-based melodies and provides controls for note entry, deletion, and playback.

Pinegger *et al.* [261] propose Brain Composing, a P300-driven composition environment with a stepwise selection process and dynamic stopping. Signals are classified after each sequence using a shrinkage Linear Discriminant Analysis (sLDA) classifier; if the same element is detected three times consecutively, it is selected, otherwise the sequence restarts. In tests with five volunteers using a tap water-based electrode amplifier to improve usability three participants exceeded 77% accuracy and successfully copy composed a target melody. Reported accuracies range from 77.8% to 95.7%, with total task times between 15 and 22 minutes, and questionnaires indicating high perceived attractiveness and ease of use.

Overall, while prior work demonstrates that P300-based BCI interfaces can support symbolic music entry and simple composition [128, 261] and that multimodal conditioning can help users map EEG patterns to discrete notes [111], the literature remains sparse on systems that leverage modern ML techniques to compose music by exploiting BCI signals end to end as pursued in our work.

### 6.1.3 Example Scenario

This section presents a usage scenario for NeuralPMG. Scenarios are widely adopted in HCI to “bring requirements to life” [265], offering concrete narratives of people performing tasks within specific contexts [130].

Tom is a professional composer who produces soundtracks for films, commercials, music videos, audiovisual installations, and contemporary classical music projects. He frequently manages multiple commissions simultaneously, spanning different genres and timelines. His goal is to create complex works as both audio recordings and complete scores, suitable for performance with virtual instruments or live ensembles. To streamline production, Tom creates semi finished musical materials that he can refine at a later stage. An expert in Slonimsky’s theory, he regularly explores generated scales and melodic patterns, constructing polyphonies using contemporary techniques of note permutation, inversions, transpositions and rhythmic variations. He is also well versed in the rules of polyphony and counterpoint for tonal and post tonal writing.

Tom prefers not to relinquish creative control to a fully automated system. Instead, he adopts NeuralPMG, whose GUI is shown in Figure 6.1.

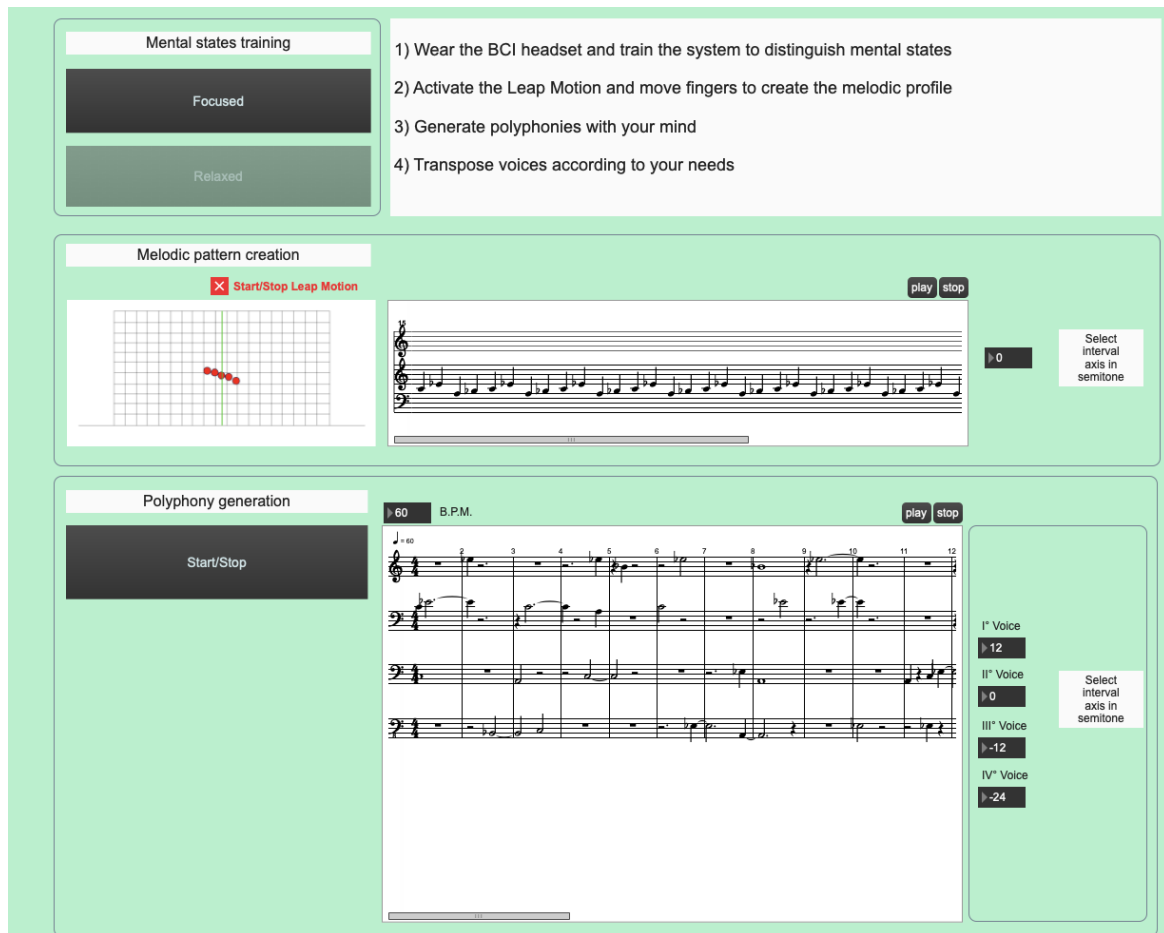


Fig. 6.1 GUI of the NeuralPMG framework: three panels outline the main steps in creating the final polyphony, *Mental states training*, *Melodic pattern creation*, and *Polyphony generation*. Usage instructions appear in the top panel.

As a first step, Tom puts on an Emotiv BCI headset and trains the system to recognise two mental states: Focused and Relaxed. He achieves this by selecting the “Focused” and then “Relaxed” buttons in the mental state training panel. Once training completes, Tom activates the Leap Motion to create a melodic pattern: moving his fingers over the sensor, the device captures finger coordinates. It renders the corresponding notes on a staff. Both the hand pose and the notated staff appear in the Melodic pattern creation panel. The pattern can be repeated or interrupted. Once you are satisfied, explore alternatives using the Interval Axes in Semitones control panel, transposing and readjusting the pattern on different interval axes. Then save the pattern as MIDI or XML. Using this seed, the system applies Slonimsky’s grammar to generate all related melodic variations.

Tom proceeds to Polyphony Generation. The trained classifier determines his mental state and drives polyphonic output. He knows that in the Focused state the

system uses quarter, eighth, and sixteenth note values, while in the Relaxed state it uses whole, half, and quarter notes. Depending on the desired character, he intentionally adopts one state or the other. He can also adjust interval sets, the performance tempo (beats per minute, *bpm*), and the interval of each voice. When satisfied, he exports the result as MIDI or XML.

Over time, Tom builds an archive of melodic and polyphonic materials to arrange and refine in a Digital Audio Workstation (DAW) or notation editor as needed.

### 6.1.4 NeuralPMG Framework Architecture

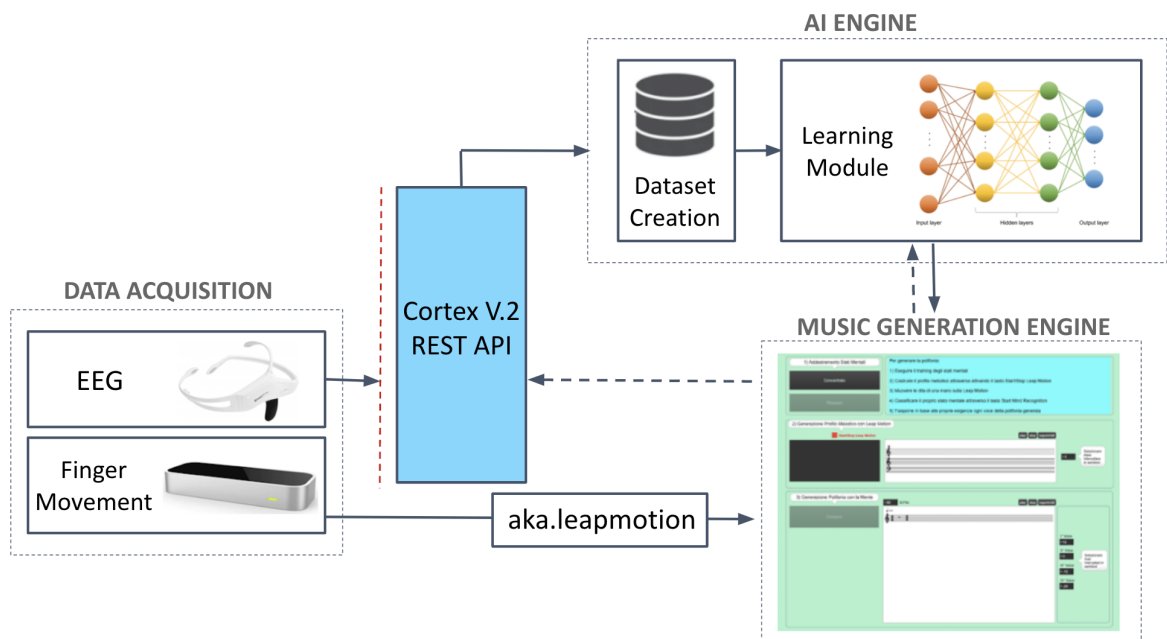


Fig. 6.2 Overview of the NeuralPMG framework architecture.

As shown in Figure 6.2, NeuralPMG comprises three macro components. The Data Acquisition module ingests signals from Leap Motion and the Emotiv headset. The AI Engine receives EEG power band values via the Cortex V.2 REST API and applies ML to first build a user training set and then classify the current mental state. The Music Generation Engine (i) derives the melodic pattern from Leap Motion data; (ii) generates a polyphony conditioned on the pattern and the AI Engine’s mental state prediction; and (iii) hosts the GUI for interaction and feedback. Dashed lines indicate control signals among modules.

The modules are orchestrated with Python Flask<sup>1</sup>, exposing an API REST over TCP/IP for client–server communication.

<sup>1</sup>Python Flask library: <https://flask.palletsprojects.com/en/2.0.x/>

## Data collection devices.

- **Brain Computer Interface.** The Emotiv Insight<sup>2</sup> is a five electrode passive headset that records EEG. Semi dry polymer electrodes typically operate without gels. Signal quality is monitored in EmotivPRO<sup>3</sup> (from Green “excellent” to Black “very poor”). Through Cortex V.2, the headset streams power spectra for *Delta* (0–4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Beta* (12–35 Hz), and *Gamma* (35–43 Hz). Electrodes are placed at *AF3*, *AF4*, *T7*, *T8*, and *Pz* under the 10–20 standard [136].
- **Leap Motion.** Leap Motion (LM)<sup>4</sup> tracks palm and finger motion in 3D. Streams are consumed via `aka.leapmotion`<sup>5</sup> in Max/MSP for immediate visualization in the GUI.

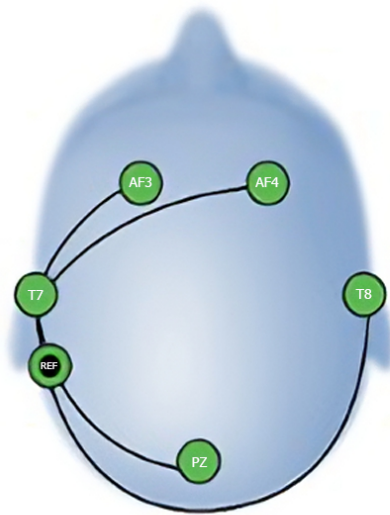


Fig. 6.3 Electrode placement under the 10–20 system.

### 6.1.5 Polyphony Generation Process

NeuralPMG’s generation pipeline has two stages:

---

<sup>2</sup><https://www.emotiv.com/insight/>

<sup>3</sup><https://www.emotiv.com/emotivpro/>

<sup>4</sup><https://www.ultraleap.com/product/leap-motion-controller/>

<sup>5</sup><https://github.com/akamatsu/aka.leapmotion>

- **Basic melodic pattern.** A reference interval axis (per Slonimsky) is selected. Infra/inter/ultra polations and their permutations are driven by finger motion (Section 6.1.5).
- **Polyphonic structure.** The seed pattern is projected to four voices. Rhythmic profiles for the voices are permuted from two sets: {quarter, eighth, sixteenth} and {whole, half, quarter}, each with rests. Dynamics per voice derive from scaled amplitudes of *Alpha*, *Beta*, *Theta*, *Gamma*, and *Delta* bands (MIDI 0–127). Selection of rhythmic set and dynamic trajectories is governed by the user’s current mental state.

**Mental State Generation** The system records band power values for *Alpha*, *Beta*, *Theta*, *Gamma*, and *Delta*. Consistent with the literature [42], Focused is associated with increased *Beta/Gamma* power; Relaxed correlates with elevated *Alpha*. Users first train the classifier to distinguish these states. To elicit Focused, the interface presents random numeric strings for subvocal reading. To elicit Relaxed, it plays natural soundscapes (lowing water, birdsong). The resulting samples form the training set for ML models and shallow feed forward networks.

**Melodic Pattern Generation** Users hold one hand roughly 30 cm above the Leap Motion. The system reads the *Y*-coordinates of five fingers (left or right hand). Three fingers determine a four-note cell: the first note is the reference; the thumb, middle, and little finger set the remaining notes.

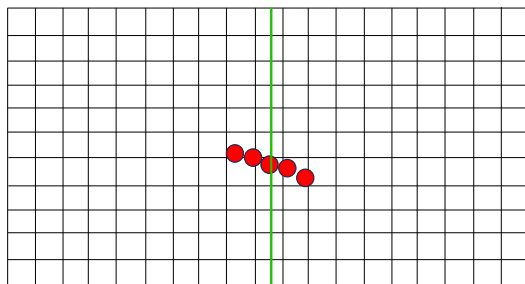


Fig. 6.4 Visualization of finger positions in the NeuralPMG GUI.

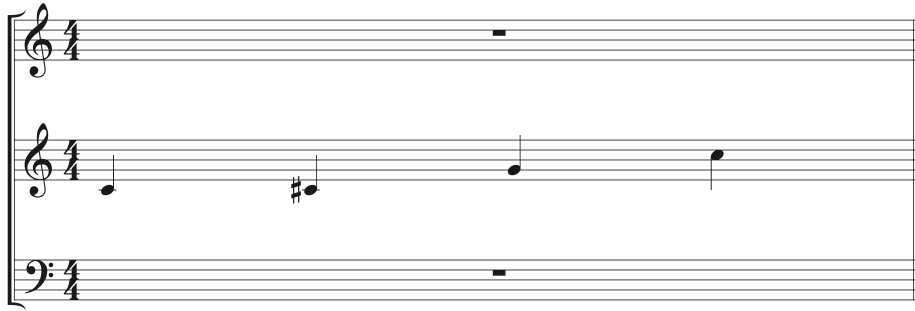


Fig. 6.5 Finger positions mapped to the musical staff.

The LM-derived pattern is scaled to two octaves via `bach.mod`<sup>6</sup>:

1. **Thresholding.** For inputs  $< 2400$  cents, output equals input; for inputs  $\geq 2400$  cents, output is input  $- 2400$ .
2. **Clef offset.** Add  $+6000$  cents for correct display in treble and bass clefs.
3. **Base cell & axis.** The four note base cell is then shifted by an axis constant: 0 cents for unison;  $+600$  cents for augmented fourth, etc.
4. **Notation.** The pattern is rendered in mensural notation [187] with a uniform quarter note rhythm.



can be transposed along any chosen interval axis by adding the corresponding constant to each pitch (Figure 6.7).

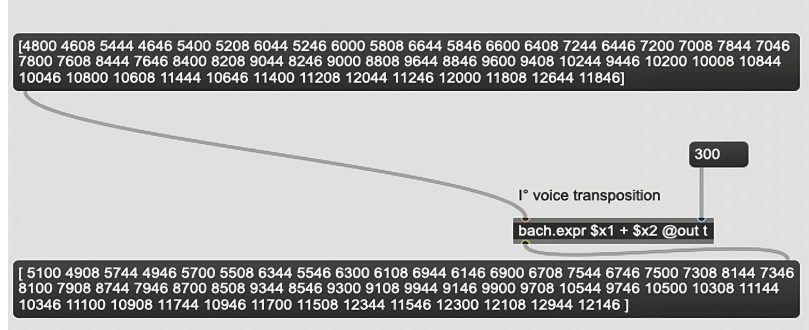


Fig. 6.7 Transposition of a pitch set by a constant (example: +300).

Dynamics are set from band amplitudes mapped to MIDI velocity (0–127) using Max/MSP’s scale with input range [0, 100] (Emotiv’s nominal maxima) and output [10, 127] to avoid silence. The piece can be auditioned at a user specified tempo (quarter =  $n$ BPM). Additional controls allow per voice transposition; output can be exported as MIDI or MusicXML.

### 6.1.6 NeuralPMG Engine

NeuralPMG records two mental states Focused/Relaxed to build a user dataset and selects the best ML model for that user. Leap Motion input supplies melodic seeds; the classifier’s output conditions polyphonic generation.

**Dataset Generation.** Communication between the BCI and Cortex API is handled via Flask<sup>7</sup>. The API streams absolute band powers ( $\mu V^2/Hz$ ) for each electrode. Each user’s dataset comprises 25 features (five bands  $\times$  five channels) across 100 acquisitions: the first 50 while reading number strings (Focused); the next 50 while listening to natural sounds (Relaxed).

**Model Selection.** Inter individual variability in EEG responses complicates classification. We therefore trained, for user, multiple models in parallel:

LDA [117], Decision Tree (DT) [55], Naive Bayes (NB) [274], SVM [335], KNN [306], and ANN [308] ( $1 \times 10$ ,  $1 \times 25$ , and  $2 \times 10$  hidden units). Using MATLAB’s Classification Learner<sup>8</sup>, we applied  $Z$  score normalization and 5 fold CV with randomized hyperpa-

<sup>7</sup><https://flask.palletsprojects.com/en/2.2.x/>

<sup>8</sup><https://it.mathworks.com/help/stats/classificationlearner-app.html>

parameter search. Table 6.1 reports pilot results (KNN under performed; FNN1, though weaker, remains implemented for completeness).

Table 6.1 Mental-state classification performance: LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayes, SVM = Support Vector Machine, FNN = Feedforward Neural Network.

User	<i>Accuracy Performance</i>							
	LD	DT	NB	SVM	KNN	FNN1	FNN2	FNN3
1	0.67	0.86	0.90	0.88	0.48	0.66	0.78	0.45
2	0.43	0.52	0.67	0.69	0.34	0.57	0.60	0.59
3	0.84	0.65	0.67	0.80	0.62	0.64	0.80	0.82
4	0.67	0.78	0.65	0.89	0.56	0.59	0.98	0.92
5	0.89	0.67	0.64	0.76	0.40	0.68	0.67	0.52

### 6.1.7 Evaluation Study

This section reports a user study evaluating interaction with NeuralPMG from multiple perspectives.

**Participants and Design** Participants were students from the Electronic Music laboratory at the E.R. Duni Conservatory of Matera, chosen for experience with algorithmic composition and sound manipulation software. Traditional composition students were not included due to limited emphasis on computer tools. All participants had moderate experience with computer aided composition; some were professional instrumentalists, and all played at least one instrument. None knew Slonimsky’s method. Demographics: age min = 19, max = 61 (avg = 30); Females = 10, Males = 9; 1<sup>st</sup> level degree = 11, 2<sup>nd</sup> level = 8.

**Procedure** Three HCI experts supported each individual session in the Conservatory’s “Rota” hall (moderate lighting, ~21°C). The study proceeded as follows: consent and setup; headset calibration (10~20 placement via EmotivPRO, with conductive gel as needed); mental state training (reading numbers for Focused, diaphragmatic breathing with nature sounds for Relaxed); melodic pattern creation with Leap Motion (start/stop, hand motion at ~10 cm, save/audition/transpose; ~10 minutes); and polyphony generation (press *Dial*, audition, adjust BPM/axes, export MIDI; ~20 minutes including calibration).

**Data Collection** We collected questionnaires (demographics; UX, creativity support, workload, engagement, emotion; self assessment; appreciation; comments) and observer notes; sessions were audio video recorded. Average completion time was  $\sim 12$  minutes.

**Experiments and Results** *Classification performance.* Table 6.2 reports participant accuracies for the selected model ( LDA, DT, NB, SVM, ANN). Selections were balanced across LDA, DT, ANN and SVM; NB was chosen twice;

Table 6.2 Model accuracy per participant. LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayes, SVM = Support Vector Machine, FNN = Feedforward Neural Network. Grey cells indicate the selected model.

Participant	<i>Accuracy Performance</i>					
	LD	DT	NB	SVM	FNN1	FNN2
1	0.67	0.78	0.85	0.66	0.51	0.43
2	0.54	0.65	0.45	0.69	0.52	0.58
3	0.84	0.76	0.67	0.82	0.64	0.47
4	0.34	0.55	0.41	0.86	0.78	0.94
5	0.89	0.78	0.68	0.72	0.81	0.84
6	0.90	0.87	0.77	0.82	0.67	0.78
7	0.55	0.52	0.41	0.48	0.50	0.51
8	0.34	0.45	0.44	0.49	0.31	0.30
9	0.80	0.87	0.90	0.97	0.87	0.88
10	0.45	0.80	0.51	0.78	0.80	0.97
11	0.30	0.73	0.45	0.56	0.70	0.75
12	0.67	0.56	0.78	0.80	0.78	0.90
13	0.45	0.91	0.34	0.78	0.68	0.87
14	0.30	0.40	0.46	0.44	0.23	0
15	0.56	0.39	0.40	0.76	0.60	0.50
16	0.70	0.80	0.64	0.79	0.73	0.77
17	0.43	0.44	0.65	0.32	0.70	0.98
18	0.67	0.91	0.80	0.78	0.79	0.46
19	0.31	0.20	0.30	0.43	0.38	0.40

**User eXperience** AttrakDiff results (Table 6.3) classify NeuralPMG as desirable, with strong hedonic qualities (HQ) and solid pragmatic qualities (PQ). HQ (1.55, 0.29) exceeds PQ (0.76, 0.35). Component scores: ATT  $7.50 \pm 0.63$ , HQ-I  $6.48 \pm 0.60$ , HQ-S  $7.25 \pm 0.73$ , PQ  $5.83 \pm 0.81$ .

Table 6.3 AttrakDiff means (AVG) and standard deviations (SD).

	<i>AVG Score</i>	<i>SD Score</i>
PQ	5.83	0.81
HQ-I	6.48	0.60
HQ-S	7.25	0.73
ATT	7.50	0.63

**Support for Creative Design** The Creativity Support Index (CSI) averaged 70/100 (SD = 14.86). Immersion scored highest (8.18), followed by Exploration (7.63), Expressiveness (7.32), and Effort/Reward (7.45). Collaboration was lowest (5.03), consistent with the system’s single user design (Table 6.4).

Table 6.4 CSI dimensions: means (AVG) and standard deviations (SD).

	<i>AVG Count</i>	<i>SD Count</i>	<i>AVG Score</i>	<i>SD Score</i>	<i>AVG weigh. Score</i>	<i>SD weigh. Score</i>
Exploration	4	1.08	7.63	1.67	30.12	1.80
Collaboration	3	1.18	5.03	2.83	16.14	3.35
Immersion	2	1.60	8.18	1.72	18.52	2.74
Expressiveness	1.42	1.07	7.32	1.93	10.40	2.07
Enjoyment	3	1.33	6.74	2.01	20.21	2.68
Effort/Reward	1	1.61	7.45	2.34	8.62	3.77

**Workload** NASA-TLX indicated high mental demand ( $7.42 \pm 7.68$ ) and low physical demand ( $0.63 \pm 2.75$ ). Performance was rated positively ( $8.11 \pm 3.02$ ), with low frustration ( $3.95 \pm 5.03$ ), though variability likely reflects headset calibration effort (Table 6.5).

Table 6.5 NASA-TLX means (AVG) and standard deviations (SD).

	<i>AVG Count</i>	<i>SD Count</i>	<i>AVG Score</i>	<i>SD Score</i>	<i>AVG weigh. Score</i>	<i>SD weigh. Score</i>
Mental Demand	1.21	0.92	4.63	2.83	7.42	7.68
Physical Demand	0.11	0.46	2.68	1.73	0.63	2.75
Temporal Demand	1.37	1.16	3.00	1.83	5.53	5.57
Performance	0.89	0.32	8.68	1.57	8.11	3.02
Effort	1.00	2.47	1.00	2.47	4.95	6.84
Frustration	0.74	2.35	0.74	2.35	3.95	5.03

**User Engagement** The UES short form averaged 4.13 (SD = 0.41). Dimension means (Table 6.6) show high Perceived Usability (PU), Aesthetic Appeal (AE), and Reward (RW), with moderate Focused Attention (FA).

Table 6.6 UES dimensions: means (AVG) and standard deviations (SD).

	<i>AVG Score</i>	<i>SD Score</i>
FA-S	3.51	0.56
PU-S	4.28	0.64
AE-S	4.28	0.71
RW-S	4.46	0.60

**Emotional Response** Emotions were grouped as {Excited/Delighted/Happy}, {Sleepy/Calm/Content}, {Neutral}, and {Sad/Depressed/Bored}. Table 6.7 displays a softmax on the category scores, producing a dominant positive class (0.99978), with negligible results on the other emotions.

Table 6.7 Softmax normalized emotional scores.

Excited, Delighted, Happy	0.99978
Sleepy, Calm, Content	0.00005
Neutral	0.00012
Sad, Depressed, Bored	0.00005

**Self assessment and Appreciation** Participants indicated that further pitch/dynamic refinement would be needed (AVG = 3.42/7). Preferred follow-up techniques were serial/atonal (AVG = 5.21) and aleatory (AVG = 5.42). Target domains skewed academic (AVG = 5.16) and multimedia (AVG = 5.84). Thematic analysis of open responses highlighted novelty, unconventional materials, and value of BCI/Leap Motion; critiques focused on headset comfort, GUI, and efficiency (see Tables 6.8 and 6.9).

Table 6.8 Aspects of NeuralPMG appreciated by participants.

<i>Theme</i>	<i>Frequency</i>
Overall system novelty	7
Use of BCI system	7
Production of unconventional polyphonies	3
Use of Leap Motion device	3
Production of unconventional melodies	3
Power of generative grammar	2

Table 6.9 Criticisms aspects of neural PGM by participants.

<i>Theme</i>	<i>Frequency</i>
Use of BCI system	5
Use of Leap Motion device	4
System graphical user interface	2
System appreciation	2
System efficiency	3

### 6.1.8 Polyphony Assessment by Music Experts

Two Music Maestro independently evaluated the 19 polyphonies using a six section questionnaire (aesthetics, harmony, elaboration, application scenarios, nine emotions, and open comments), each on 1~7 Likert scales where applicable. Results are summarized via heatmaps with means and standard deviations (Figures 6.8 and 6.9).

<i>Section</i>	<i>Specific aspect</i>	<i>Participant</i>																			<i>Descriptive statistics</i>			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	AVG	SD	AVG	SD
<i>Aesthetic</i>	Likeability	4	5	6	6	6	6	7	7	7	7	6	5	7	5	7	6	7	5	7	6.11	0.94		
	Elegance	6	6	6	6	6	6	7	7	7	7	5	7	7	7	6	7	5	7	6.42	0.69	6.32	0.78	
	Consonance/Dissonance	5	6	6	6	6	6	7	7	7	7	5	7	7	7	6	7	6	7	6.42	0.69			
<i>Innovativeness</i>	Harmonic Interest	5	5	6	6	6	6	2	3	2	2	2	2	2	2	2	2	2	2	3.21	1.75			
	Harmonic Innovativeness	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2.11	0.46	2.37	1.03	
	Innovativeness Vertical Intervallar Structures	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2.05	0.23			
	Innovativeness Horizontal Intervallar Structures	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2.11	0.32			
<i>Elaboration</i>	No Need to Adjust Parameters	2	5	6	5	5	5	6	6	6	6	6	6	6	7	6	7	6	7	5.74	1.10			
	Polyphony Thoroughness	2	5	5	5	5	5	5	4	5	4	4	5	5	5	4	6	4	5	4.63	0.83	4.18	1.84	
	Polyphony Needs No Further Modification	6	3	4	3	3	2	1	2	2	1	2	2	1	2	1	2	1	2	1	2.16	1.26		
<i>Usage</i>	Music Improvisation Boosting	1	2	3	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1.26	0.56		
	Music Composition Boosting	7	6	6	5	4	4	3	3	1	1	3	3	3	3	2	2	3	2	3	3.37	1.64	2.25	1.39
	Music Application Domain	1	1	1	3	2	1	3	2	2	2	2	1	3	2	3	2	3	3	3	2.11	0.81		
<i>Emotion</i>	Happiness	3	2	1	1	1	1	2	2	2	3	1	1	3	1	3	1	3	1	3	1.84	0.90		
	Sweetness	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.00		
	Anger	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.00		
	Sadness	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.00		
	Fear	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.00	1.61	0.96
	Negativity	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.00		
	Arousal	5	4	3	3	3	3	3	3	3	3	2	3	3	3	3	3	4	2	3	3.11	0.66		
	Positivity	3	2	1	1	1	1	1	1	2	1	1	2	1	3	1	1	1	1	2	1.42	0.69		
	Strength	4	2	3	3	4	3	3	3	3	3	4	3	3	3	3	3	3	3	3	3.11	0.46		

Fig. 6.8 Maestro 1's evaluation of the 19 polyphonies.

Section	Specific aspect	Participant																			Descriptive statistics			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	AVG	SD	AVG	SD
Aesthetic	Likeability	4	6	5	6	7	7	6	6	7	7	6	7	5	7	6	6	5	5	5	5.95	0.91		
	Elegance	4	7	5	6	7	6	5	6	7	6	5	7	5	6	6	6	4	5	5	5.68	0.95	5.95	0.91
	Consonance/Dissonance	5	7	6	7	7	6	6	7	7	7	6	7	6	7	6	6	4	6	5	6.21	0.85		
Innovativeness	Harmonic Interest	5	7	6	6	7	6	5	7	7	7	6	7	5	6	7	6	5	5	6	6.11	0.81		
	Harmonic Innovativeness	5	7	6	6	7	5	5	7	7	7	6	6	6	6	7	6	6	5	6	6.11	0.74	6.01	0.77
	Innovativeness Vertical Intervallar Structures	6	7	5	6	5	5	5	6	7	7	6	6	5	6	7	6	6	5	6	5.89	0.74		
Elaboration	Innovativeness Horizontal Intervallar Structures	6	7	5	7	5	5	5	7	7	7	6	6	5	6	7	6	6	5	5	5.95	0.85		
	No Need to Adjust Parameters	3	7	3	3	3	5	2	6	7	7	2	6	3	6	6	5	2	2	3	4.26	1.91		
	Polyphony Thoroughness	4	2	2	3	6	4	2	6	7	7	4	2	5	7	4	4	2	3	2	4.00	1.86	4.18	1.91
Usage	Polyphony Needs No Further Modification	3	1	2	4	7	6	2	6	6	7	4	6	5	1	5	5	1	4	6	4.26	2.05		
	Music Improvisation Boosting	5	7	6	6	7	5	2	4	7	7	5	6	5	7	6	5	2	4	5	5.32	1.53		
	Music Composition Boosting	6	7	7	6	7	5	3	7	7	7	5	7	6	7	7	6	3	5	6	6.00	1.29	5.88	1.36
Emotion	Music Application Domain	6	7	7	7	7	6	4	7	7	7	6	7	7	7	7	6	3	6	6	6.32	1.11		
	Happiness	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.21	0.92		
	Sweetness	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.05	0.23		
	Anger	1	1	1	1	1	1	1	2	5	4	2	2	1	4	7	1	2	1	3	2.16	1.71		
	Sadness	1	1	5	6	7	5	1	5	7	5	2	6	3	7	5	7	3	5	3	4.42	2.12		
	Fear	1	1	2	3	5	4	1	7	6	5	2	6	3	7	6	2	3	4	3	3.74	2.02	3.02	2.20
	Negativity	1	1	4	4	6	7	1	6	6	5	6	6	6	7	6	7	3	5	3	4.74	2.05		
	Arousal	5	5	2	2	1	1	6	4	6	7	5	5	3	5	7	1	5	5	7	4.32	2.06		
	Positivity	5	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1.26	0.93		
	Strength	6	7	5	2	3	3	2	2	6	7	2	5	3	6	7	4	3	5	4	4.32	1.83		

Fig. 6.9 Maestro 2's evaluation of the 19 polyphonies.

### 6.1.9 Discussion

Overall, the user experience (UX) was positive. Creativity scores highlighted exploration, immersion, and enjoyment; expressiveness was less relevant. The workload had a low physical impact, but variability in calibration caused frustration. Engagement was high, and emotional responses were predominantly positive. Participants found NeuralPMG effective for organizing complex presentation materials. However, modifications to the presentation/dynamics were anticipated; the results were useful starting points, particularly for academic and multimedia contexts in line with serial/atonal practice. Feedback suggested moderating the sensitivity of Leap Motion and improving the clarity of the GUI. To improve explainability [285, 293]), future versions could display decision indicators to boost user confidence.

## 6.2 Neural Musical Instruments through Brain Computer Interface and Biofeedback

### 6.2.1 Introduction

Electroencephalography EEG is a primary noninvasive technique for investigating the brain's electrical dynamics. Recording spontaneous neural activity from multiple scalp electrodes enables the analysis of signal modulations driven by external stimuli and internal processes. Supervised ML models can leverage these modulations to classify

user intent and mental states, thereby enabling closed loop, human in the loop systems. In this context, BCI systems with a small number of electrodes placed over ROI have emerged as practical enablers of real time interaction between people and external device.

AI methods have already shown impact in a wide range of fields how BCI interaction [22], information security [36], precision agriculture [225], Unmanned Aerial Vehicle (UAV) sensing [199], emotion recognition [69], biofeedback and self regulation [66], motor imagery [14], SSVEP based control and device operation [121].

The interaction between BCI and music is a rapidly emerging area of EEG research. However, the real time control of musical instruments with BCI remains largely unexplored. This research addresses this gap by introducing a comprehensive system for the neural control of instrument parameters during live performances. The system is based on the robust classification of mental states of concentration and relaxation. Furthermore, an explainability study accompanies this work to clarify which features and electrode positions drive the model’s decisions.

**Motivation.** Classical BCI metrics (Beta/Theta and Alpha/Delta) are often used for real time neurofeedback on activation and relaxation [214]. In practice, however, these two ratios alone may be insufficient to discriminate focus and relaxation across individuals consistently. Inter subject variability suggests that both (i) a richer feature set and (ii) spatial information from the electrode montage should be considered to improve robustness and interpretability.

**Research Questions.** We investigate:

- **RQ1:** To what extent do multiple band power features jointly contribute to activation detection across subjects?
- **RQ2:** Are there consistent feature location patterns that emerge across users?

To answer these questions, we apply SHAP [208] to quantify feature attributions and visualize model behavior at the subject level.

**Contributions.**

1. BCI driven performance framework that maps *focused* vs. *relaxed* mental states to real time control of virtual instrument parameters, enabling expressive, neural modulation of sound.

2. A subject agnostic pipeline (preprocessing, LOSO evaluation, model selection) that achieves high accuracy on public EEG data for focus/relax detection.
3. A comprehensive explainability analysis identifying feature electrode combinations that consistently drive performance and generalize across users.

**Roadmap.** Section 7.4.2 surveys background and related work. Section 6.2.3 details our architecture, BCI device, VST integration, and interpretability tools. Section 6.2.8 describes the experimental settings and preprocessing/ LOSO pipelines, while Section 7.4.10 reports results and SHAP analyses. Section 6.4.10 discusses implications and concludes.

## 6.2.2 Background and Related Work

While BCI have been explored for creative interaction, real time instrument control remains comparatively sparse in the literature. Early bioelectric controllers such as BioMuse [167] demonstrated that alpha delta dynamics correlate with relaxation, whereas beta theta trends may accompany increased activation. This inspired subsequent attempts to map physiological states to musical control parameters.

Miranda and Brouse’s BCMI Piano [222] formalized an EEG driven music generation pipeline: Hjorth parameters (Activity, Mobility, Complexity) modulated tempo and dynamics, while band specific features informed pitch organization; MIDI served as the inter device protocol. Earlier still, Alvin Lucier’s pioneering work [204, 304] harnessed alpha amplitude fluctuations, measured from frontal electrodes, to actuate percussion surfaces through loudspeakers, effectively turning brain rhythms into sound producing control signals.

More recent BCI [189] implementations combine EEG and accelerometers, connecting bandwidth powers to development environments for direct musical mapping. Building on these foundations, our system focuses on reliably decoding the binary mental state (concentration/relaxation) to enable continuous control of the parameters of a virtual instrument during performance, as well as identifying which features and electrodes contribute most to classification.

## 6.2.3 Materials and Methods

This section outlines the overall system 6.3.4, BCI hardware 6.2.5, the virtual instrument integration 6.2.6, and our interpretability method 6.2.7.

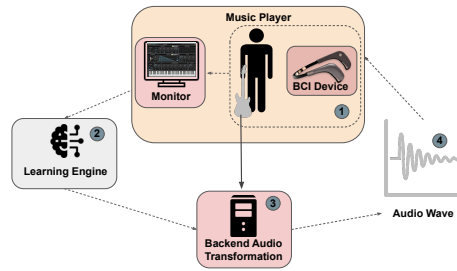


Fig. 6.10 System architecture: EEG acquisition, Learning Engine (classification + focus index), audio/VST mapping

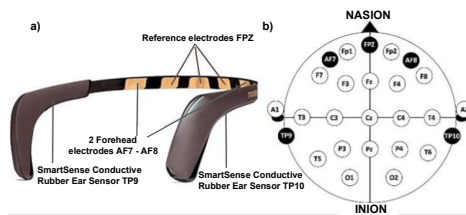


Fig. 6.11 (a) Muse EEG headset; (b) 10–20 electrode layout.

## 6.2.4 System Architecture

Figure 6.24 summarizes the four stages of the process. First, a musician wears a BCI and plays an instrument. Second, the learning engine pre-processes the EEG stream and generates a concentration/relaxation prediction from a pre-trained classifier. In parallel, a Beta ratio protocol [214] provides a continuous index of concentration intensity. Third, an audio back end maps the classifier output and concentration index to VST parameters for real time sound modeling. Finally, the musician detects these timbral changes and can guide the sound by modulating attention or relaxation.

## 6.2.5 Muse Headset

We employ the Muse2 6.11, a dry electrode BCI with four sensors placed according to the international 10–20 system. The device provides adequate signal quality without the need for conductive gels and is comfortable for extended sessions, facilitating live performance scenarios.

## 6.2.6 Virtual Musical Instrument

The popularity of musical instruments that interface directly with computers has caused an exponential increase in the use of VST plugins [311]. VST is a software interface that integrates software audio synthesizer and effect plugins with audio editors and recording systems. We implemented the system GUI in Max/MSP, exposing (i)

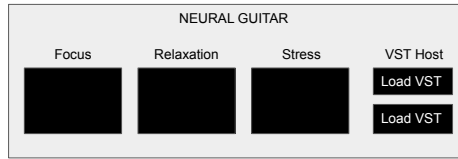


Fig. 6.12 Neural Guitar GUI: model indicators, focus index, and VST control.



Fig. 6.13 Example VST interface with controllable parameters.

model outputs (probabilities/indices) and (ii) on-screen controls to load and display the target VST. Figure 6.12 shows our Neural Guitar interface, while Figure 6.13 illustrates the exposed sound synthesis parameters.

### 6.2.7 Interpretable Model

We use SHAP [208] to quantify each feature’s contribution to the predicted class, enabling subject and cross subject analysis of the most influential band electrode combinations and supporting design decisions about electrode selection and mapping strategies.

### 6.2.8 Experimental Settings

The available dataset “EEG Brainwave Dataset: Mental State” [33] was employed in this study. It comprises EEG recordings from four participants (two female and two male) under three distinct mental states: concentration, neutrality, and relaxation. The relaxation condition was elicited through exposure to soothing auditory stimuli, whereas the neutral state involved no specific external stimulation. The concentration state was induced with a visual attention task based on the “three cups” paradigm.

Table 6.10 Dataset size after preprocessing.

$S_{ID}$	$Fcs$	$Rx$
1	240	698
2	346	418
3	256	610
4	80	506

(a)  $S_{ID}$ : Subject ID,  $Fcs$ : Focus,  $Rx$ : Relaxation.

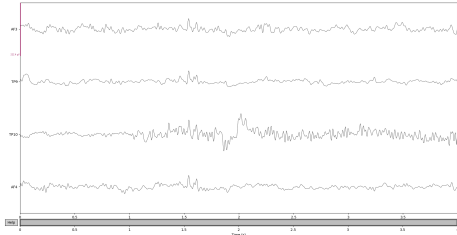


Fig. 6.14 EEG trial (4 s) after SPA denoising and preprocessing.

## 6.2.9 EEG Preprocessing Pipeline

We emphasize artifact mitigation and robust feature extraction. After visual inspection,  $TP9$  and  $TP10$  were excluded due to noise. Signals were segmented into sub epochs and segments exceeding  $150 \mu V$  were discarded. Autoreject [149] was used to estimate a global threshold and remove additional artifact epochs. The whole pipeline is described below.

1. Build a calibration buffer; divide into 32 sub epochs of 32 samples each.
2. Concatenate sub-epochs; bandpass filter 1–40  $Hz$ ; apply SPA denoising [248].
3. Extract Delta, Theta, Alpha and Beta band powers for  $AF3$  and  $AF4$  using NeuroKit2 [210].
4. Once 50 trials are buffered, compute the trial wise average band powers.
5. For each subject, assemble a dataframe with  $User\_ID$  and averaged band powers.
6. Eliminate the oldest sub epoch and add a new test from 32 samples; repeat until the end of the recording.

Figure 6.14 shows a representative 4 second trial after SPA denoising algorithm. The final dataset contains  $N_{focus} = 922$  and  $N_{relax} = 2232$  epochs. Table 6.10 details the number of trials for the subject.

Table 6.11 Best models per subject in LOSO.

S <sub>ID</sub>	BM	BS	BMP
1	Logistic Regression	Standard Scaler	C=2.3
2	Logistic Regression	Standard Scaler	C=0.1
3	SVM Classifier	MinMax Scaler	C=14
4	Random Forest	Standard Scaler	Max_depth=3, n_estimators=62

Table 6.12 BM: Best Model, BS: Best Scaler, BMP: Best Model Params.

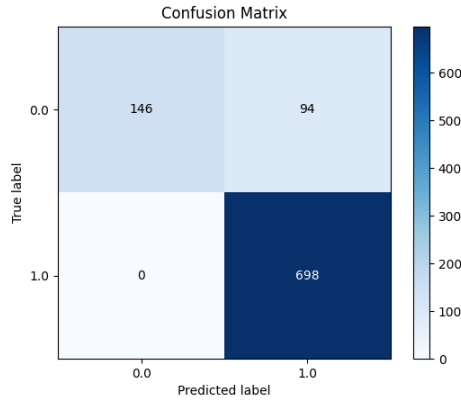


Fig. 6.15 Confusion matrix for Subject 1 (LOSO test).

### 6.2.10 Machine Learning LOSO Pipeline

We adopt LOSO cross validation, a recommended strategy for biometrics, coupled with grid search over hyperparameters. Models considered: (i) shallow Neural Network, (ii) AdaBoost, (iii) Logistic Regression, (iv) Random Forest, (v) SVM. Using scikit-learn [253]. Whole ML training strategy is described below.

- Assign unique IDs to the four subjects.
- For each LOSO split, train on  $N-1$  subjects on the train and test on remain subject.
- Grid search to select the best model.

### 6.2.11 Results

Table 6.11 summarizes the best model and hyperparameters for all subjects. Confusion matrices for all test subject are shown in Figures 6.15,6.16,6.17,6.18, and Table 6.13 reports metrics for the test subject. Mean performance across subjects is in Table 6.14.

To interpret and explain these results, Figures 6.19,6.20,6.21,6.22 provide SHAP beeswarm plots for the test sets. In all subjects, the power Alpha for AF3 consistently

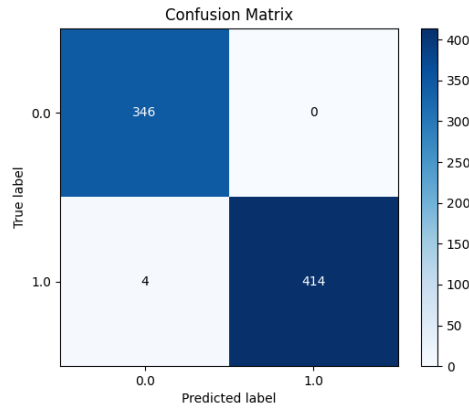


Fig. 6.16 Confusion matrix for Subject 2 (LOSO test).

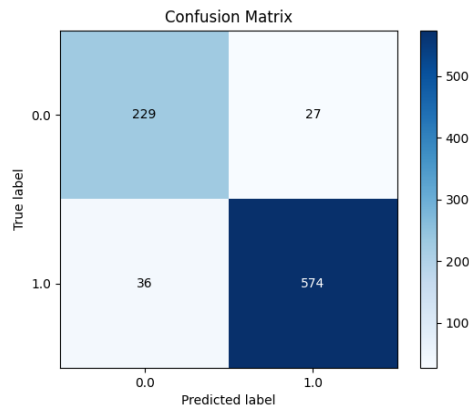


Fig. 6.17 Confusion matrix for Subject 3 (LOSO test).

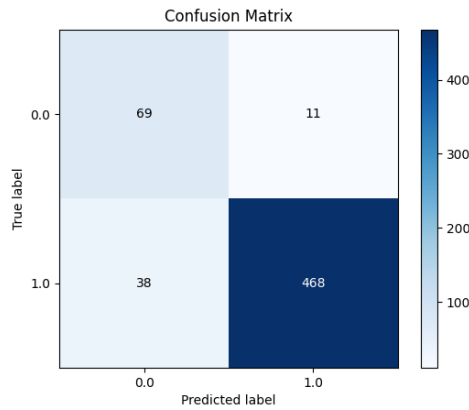


Fig. 6.18 Confusion matrix for Subject 4 (LOSO test).

shows a high positive contribution to correct classification. However, interactions between multiple bands in both AF3 and AF4 are prominent, highlighting the need to preserve spatial information and consider multiband feature sets in practical systems.

Table 6.13 Per-subject metrics in LOSO.

S <sub>ID</sub>	Class	Precision	Recall	F1	Accuracy
1	Focus	1.00	0.61	0.76	0.90
	Relaxation	0.88	1.00	0.94	
2	Focus	0.99	1.00	0.99	0.99
	Relaxation	1.00	0.99	1.00	
3	Focus	0.86	0.89	0.88	0.93
	Relaxation	0.96	0.94	0.95	
4	Focus	0.64	0.86	0.74	0.92
	Relaxation	0.98	0.92	0.95	

Table 6.14 Mean performance across subjects.

Class	MP	MR	MF1	MA
Focus	0.87	0.84	0.88	0.92
Relaxation	0.95	0.96	0.96	

(a) MP: Mean Precision; MR: Mean Recall; MF1: Mean F1; MA: Mean Accuracy.

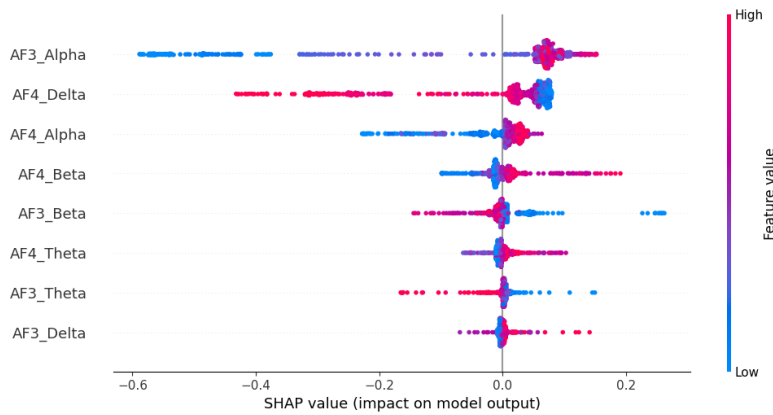


Fig. 6.19 SHAP beeswarm plot for Subject 1: feature contributions.

## 6.2.12 Discussion and Conclusion

We presented a BCI based prototype that classifies the state of concentration and relaxation in real time and maps these states onto VST parameters for expressive neural control of sound. The LOSO training strategy demonstrates robust generalization across users, while SHAP analysis reveals consistent and interpretable performance factors. From a design perspective, it seems advisable to use at least two frontal electrodes to preserve the spatial critical information for concentration detection. Future work will (i) extend the multi parametric control of dynamics, timbre, and

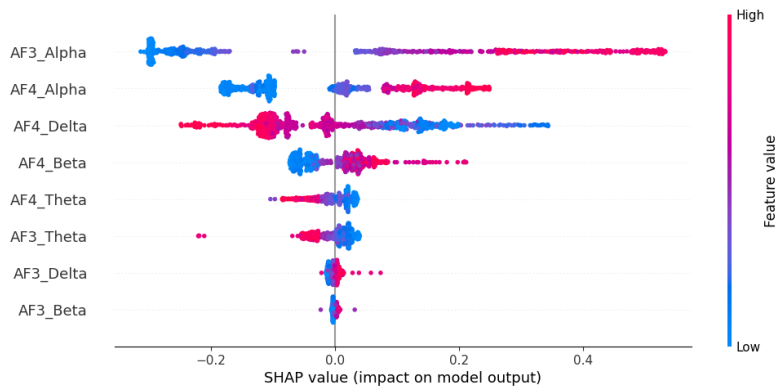


Fig. 6.20 SHAP beeswarm plot for Subject 2: feature contributions.

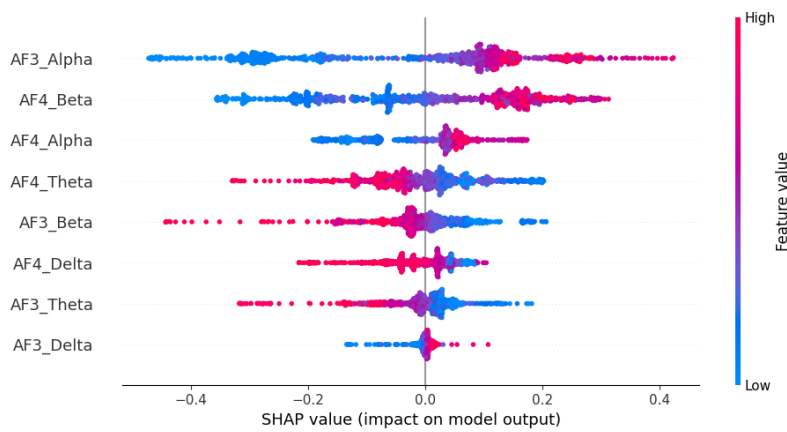


Fig. 6.21 SHAP beeswarm plot for Subject 3: feature contributions.

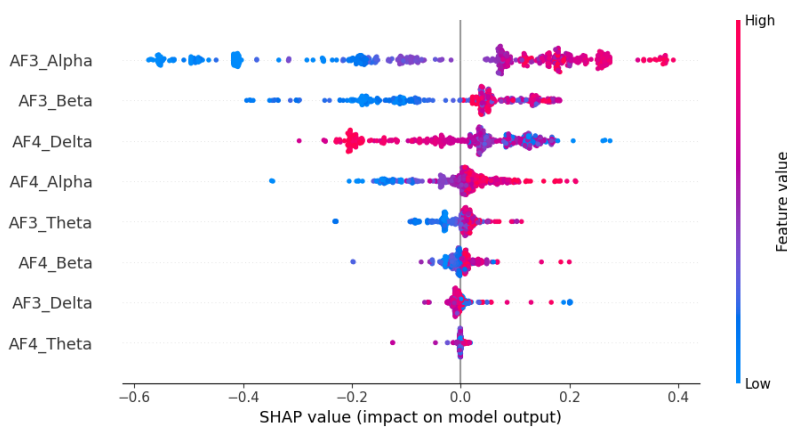


Fig. 6.22 SHAP beeswarm plot for Subject 4: feature contributions.

effects using multidimensional controls, (ii) explore more complex spatial features, and (iii) validate the system in live performance contexts with professional musicians.

## 6.3 EmoSynth Real Time Emotion Driven Sound Texture Synthesis via Brain Computer Interface

### 6.3.1 Introduction

Electroacoustic composition has made significant advances, from Luigi Russolo’s *In-tonarumori* to contemporary programmable ecosystems for digital sound synthesis (Max/MSP<sup>9</sup>, Pure Data<sup>10</sup>, SuperCollider<sup>11</sup>).

Modern algorithmic techniques and domain specific languages have enabled intricate procedures for sound generation and opened new conceptual frameworks for compositional practice. With the pervasive diffusion of portable synthesizers and software instruments, both traditional and contemporary musical repertoires now incorporate digital timbres, often beyond the tempered system. [227]

Recent DL based generative models further expand this scenario. Transformer architectures, widely used in natural language processing DL [278], have been adapted to produce music and sounds in different styles, often through interaction based on text prompts and high levels of autonomy [142, 155].

From a compositional standpoint, however, there is value in *conscious*, moment-to-moment control that reflects the creator’s embodied state. Here, BCIs offer an effective interface to steer computational processes in real time; they have demonstrated strong performance in time-critical settings such as healthcare [198, 300] and security [36, 199]. Complementarily, emotion recognition from EEG affords continuous monitoring of a composer’s affect, including responses to auditory stimuli [22, 66, 69].

From a compositional point of view, musicians need to have conscious control over their emotional and mental state. In this case, BCI offers an effective interface for recognizing emotions from EEG and cognitive states such as focus, relaxation, stress level, and engagement. This technology allows for continuous monitoring of a composer’s affectivity, including responses to auditory stimuli [22, 66, 69].

This work proposes a music composition support system capable of generating and conditioning sound textures based on the musician’s current emotional state. The framework offers two complementary operating modes and is designed for musicians who already have a reference timbral palette and wish to refine or vary it. To enable fluid tone morphology within a user defined corpus, we use VAE trained on datasets curated by the composer to meet specific expressive needs.

---

<sup>9</sup><https://cycling74.com/products/max>

<sup>10</sup><https://puredata.info/>

<sup>11</sup><https://supercollider.github.io/>

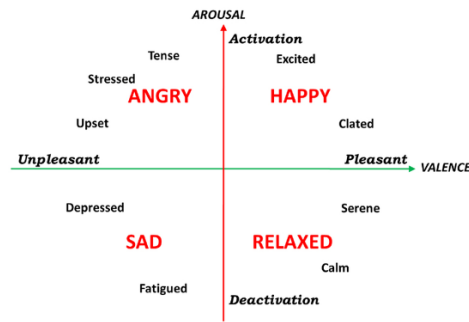


Fig. 6.23 Russell's circumplex model.

### 6.3.2 Background and Related Works

Two main models guide the scientific study of emotions: discrete [194] taxonomies and dimensional [279] schemes. Discrete models enumerate primary emotions as positive/negative. Dimensional models, on the other hand, place emotions on continuous spaces by crossing arousal, valence, and dominance. Russell's circumplex model is the reference template for the discretization of emotions on orthogonal axes of arousal and valence, which include every possible combination of valence and arousal. Figure 6.23 depicts the discretization over this plane.

In affective computing, EEG is a very attractive signal source for emotion recognition due to its direct neurophysiological origin [255]. Hu et al. [138] combined wavelet packet features, spectral power density, and entropy with a width learning classifier, achieving an accuracy of up to 81.5 when combining different entropy measures. Further advancements in EEG emotion recognition were made through the development of a Multi Task Joint Neural Network (MT-2DCNN-LSTM) model [332]. This model leverages both two dimensional convolutional neural network and LSTM networks to construct three dimensional feature matrix graphs and two dimensional sequence information from EEG signals. Validated on the DEAP dataset, the MT-2DCNN-LSTM model demonstrated high average accuracies of 97.29 and 97.72 for arousal and valence, respectively.

In recent developments for generative audio models, a hierarchical approach has been adopted to address the challenge of capturing temporal dependencies across varying timescales. Wavenet [245] utilised dilated CNN layers to selectively sample data points from prior layers, enhancing the model's ability to process longer time dependencies. Similarly, S. Mehri et al [220], introduced SampleRNN which incorporated layers of RNN to learn abstract representations on distinct temporal dimensions. Despite these advances, the models struggled to maintain coherent structures over long durations without external conditioning signals. More contemporary research has improved upon

this by segmenting the learning of local and broader structural patterns, training distinct models for different levels of abstraction. In this work [90] Dieleman *et al.* proposed using the latent space of a model trained on longer timescales to inform another model focused on shorter timescales. Currently, various transformer architectures such as Suno <sup>12</sup> and Udio <sup>13</sup> appear to be very effective tools for both music and sound texture generation.

### 6.3.3 Materials and Methods

#### 6.3.4 System Architecture

Figure 6.24 outlines a five stage loop:

1. **Acquisition start** The performer initiates the BCI acquisition.
2. **Signal quality gate.** The Learning Engine ingests a 4s EEG window, performs preprocessing and quality checks; if the check fails, the user is prompted to adjust the headset and restart.
3. **Emotion inference** A subsequent 4s segment is processed to predict the emotion class using pre trained models; frontal asymmetry indices are computed in parallel.
4. **Conditional synthesis** One of four pre trained VAE is activated according to the predicted class, generating a sound texture aligned with the inferred quadrant.
5. **Performance control** The musician listens and decides to loop the sample or attenuate it via a MIDI controller<sup>14</sup>; the loop continues as new EEG windows arrive.

#### 6.3.5 MUSE

We employ the Muse2 headset, a dry electrode BCI for EEG acquisition. It uses four sensors positioned under the international 10–20 system at *Fpz*, *AF7*, *AF8*, *TP9*, and *TP10*, sampled at 256 Hz. The adjustable band and dry contacts improve comfort and setup speed, which is crucial for live contexts.

---

<sup>12</sup><https://suno.com/>

<sup>13</sup><https://www.udio.com/>

<sup>14</sup><https://www.algameko.com/it/midi-controller/akai-professional>

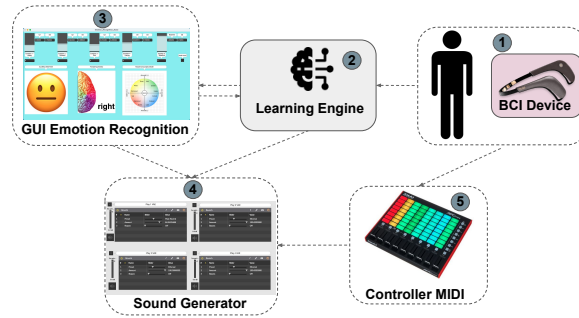


Fig. 6.24 System architecture: acquisition, quality gate, emotion inference, conditional VAE synthesis, and performer control.

### 6.3.6 Experimental Setting

The experiments were conducted using 5 dedicated datasets. A primary dataset consisting of EEG signals was employed for training models capable of classifying a user’s emotional state in real time. The other 4 datasets were utilized to train VAE to generate new sound textures. The datasets and specific procedures are detailed below.

### 6.3.7 Emotion Dataset Description

Following the DEAP protocol [169], we acquired EEG from 30 participants (age  $\sim 20$ –30) at Politecnico di Bari Sisinflab. After informed consent and headset fitting, each participant viewed 40 music videos while EEG was recorded. Anonymization and secure storage were enforced throughout. The dataset is organised as follows: (i) 30 subjects are collected and separated into different folders. (ii) Each folders contains 40 EEG trial organised into sessions folder.

### 6.3.8 VAE Dataset Description

We curated four tone corpora from pairs of seed sounds. For each seed, we applied augmentation: (i) additive Gaussian noise, (ii) time stretch, (iii) pitch shift, and (iv) slicing, yielding 1000 variants per seed and 8000 samples overall. We computed Mel spectrograms and converted them to dB using `librosa` [218]. Figure 6.25 illustrates a representative Mel spectrogram.

Four distinct datasets were generated from pairs of original audio recordings. To increase data diversity, a data augmentation pipeline was applied to each set of source sounds. This pipeline incorporated four techniques: (i) addition of Gaussian noise, (ii) time stretching, (iii) pitch shifting, and (iv) audio slicing. Through this process, 1000 augmented variants were produced per sound, yielding a total of 8000 audio

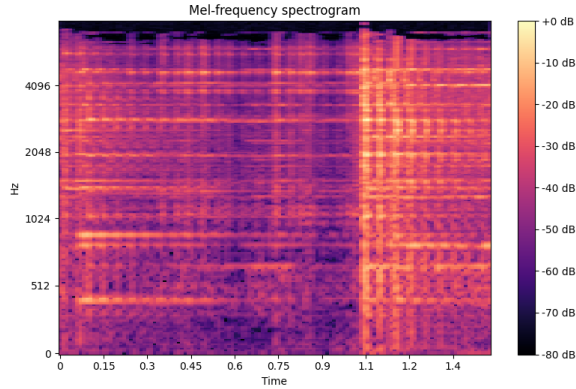


Fig. 6.25 Mel spectrogram of an augmented sample.

samples. Mel spectrograms were subsequently computed for each sample and converted to a decibel scale using the Python librosa library [218]. Figure 6.25 illustrates a Mel spectrogram generated from one of the augmented audio samples.

After feature extraction steps, normalization was conducted following the designated formula.

$$\text{MelSpectrogram} = \frac{S - \min(S)}{\max(S) - \min(S)} \quad (6.1)$$

where  $S$  is the matrix representing the spectrogram,  $\min(S)$  and  $\max(S)$  are its minimum and maximum values, respectively.

After normalization, to rescale an array  $A$  back to a specific range  $[\min, \max]$ , the following formula is used:

$$A = A \cdot (\max - \min) + \min \quad (6.2)$$

where  $\min$  and  $\max$  define the new range for the array.

Each mel spectrogram is saved along with its minimum and maximum values.

### 6.3.9 EEG Preprocessing

We duplicate epochs, band-pass filter (1–40 Hz), and apply transposition and amplitude scaling for efficient processing. Artifact detection uses a KNN anomaly score trained on filtered data; epochs exceeding a threshold are marked. We then remove artifacts with meegkit [25] (ringing-reduction), resize/reshape the data, and encapsulate in MNE EpochsArray. If no artifacts are detected, raw epochs are preserved. Figures 6.26, 6.27 shows the effect of the pipeline.

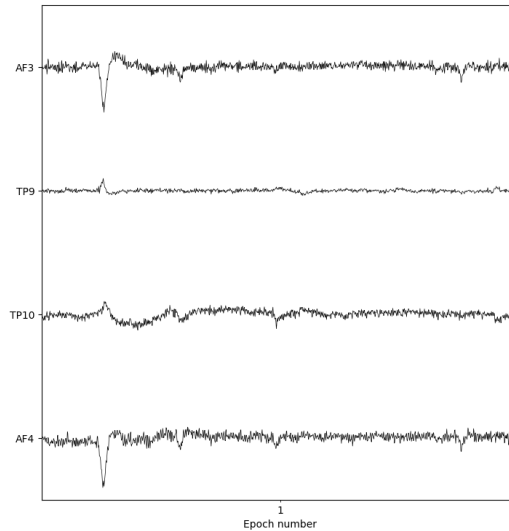


Fig. 6.26 Example epoch before preprocessing.

The preprocessing pipeline begins by duplicating the EEG data epochs and applying a bandpass filter to isolate frequencies between 1 and 40 Hz. Following filtering, the data undergo transposition and amplitude scaling to enhance computational efficiency. Artifact detection is then performed using a KNN classifier, trained on the filtered dataset to compute anomaly scores that identify noise contaminated epochs. Epochs with scores exceeding a predefined threshold are marked as containing significant artifacts. Artifact removal is carried out using the megkit framework [25], specifically employing the ringing artifact reduction method. After correction, the EEG data are resized, restructured, and encapsulated in an MNE EpochsArray object. When no artifacts are detected, the data remain unaltered. This rigorous preprocessing pipeline ensures the integrity of the EEG signals and improves the reliability of subsequent analyses. Figures 6.26 and 6.27 illustrate the effectiveness of the procedure in removing noise components from representative EEG trials.

### 6.3.10 ML Emotion Recognition

We evaluate Logistic Regression, SVM, Random Forest, and XGBoost under exhaustive grid search with LOSO cross validation. For each subject, models are trained on the remaining  $N-1$  subjects and tested on the held-out subject; this is repeated across all subjects. To operationalize Russell’s model, we train 30 binary classifiers per quadrant (vs. neutral), yielding 120 specialized models. Final decisions use majority voting across quadrant-specific models. Table 6.15 reports best model for quadrant with mean test

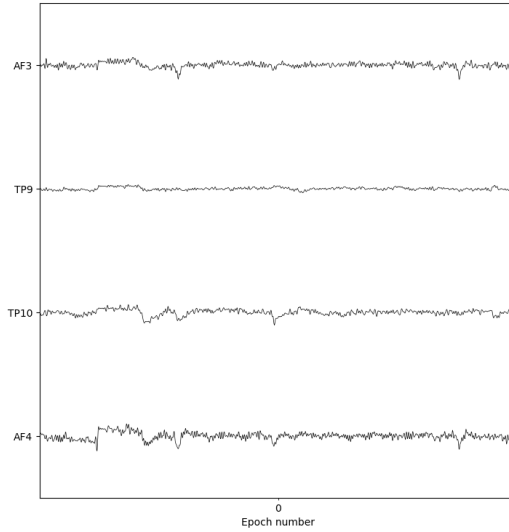


Fig. 6.27 Example epoch after preprocessing.

Table 6.15 Mean Model Accuracy.

Best Model	Russel's Quadrant	Mean Test Accuracy	Mean std score
SVM	1	0.77	0.035
SVM	2	0.78	0.022
SVM	3	0.76	0.024
SVM	4	0.80	0.028

accuracy and standard deviation. Several ML models, including LR, SVM, RF, XGB were evaluated to identify the most effective classifier for distinguishing emotional states. An exhaustive hyperparameter optimization was performed using a Grid Search combined with a LOSO cross validation scheme. This process is systematically repeated for every subject in the dataset. The Grid Search efficiently explored the hyperparameter space to determine the optimal configuration for each model, resulting in the development of 120 specialized classifiers. Specifically, for each quadrant of Russell's circumplex model, 30 models were trained to discriminate between that emotional quadrant and the neutral state. Final classification decisions were obtained through a majority voting strategy. Table 6.15 presents the mean accuracy scores for all quadrants of Russell's model, along with the corresponding best performing classifier.

### 6.3.11 Variational Autoencoder

To obtain a compact representation of the latent EEG space and enable sound generation, five VAE with identical architectures were trained. The models were implemented

using the Keras framework [124] and organized into three main components: (i) the encoder, (ii) the decoder, and (iii) the complete autoencoder. Convolutional layers were used for encoding, while transposed convolutional layers performed the decoding.

**Encoder:** The encoder receives input data of a predefined shape and processes it through a sequence of convolutional layers, each followed by ReLU activation and batch normalization. The layers are configured as follows:

- 512 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 256 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 128 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 64 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 32 filters,  $(3 \times 3)$  kernel, stride  $(2,1)$ , padding 'same'.

**Latent Bottleneck:** The flattened output of the final convolutional layer is passed through two dense layers that generate the mean  $\mu$  and log variance  $\log \sigma^2$ . These parameters are used to sample the latent vector  $z$  via the reparameterization trick:

$$z = \mu + e^{\frac{\log \sigma^2}{2}} \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (6.3)$$

This step defines the latent space representation that acts as the information bottleneck of the encoder.

**Decoder:** The decoder begins with a dense layer that reshapes the latent vector back into the bottleneck dimensions. The reshaped data are then passed through a series of transposed convolutional layers mirroring the encoder's structure in reverse:

- 32 filters,  $(3 \times 3)$  kernel, stride  $(2,1)$ , padding 'same';
- 64 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 128 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 256 filters,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same';
- 1 filter,  $(3 \times 3)$  kernel, stride  $(2,2)$ , padding 'same'.

A sigmoid activation function is applied at the output layer to reconstruct the input image.

Table 6.16 EEG-VAE reconstruction error by emotion quadrant.

Quadrant	MSE	MAE
1	0.020	0.118
2	0.010	0.103
3	0.015	0.100
4	0.020	0.115

**Full Autoencoder:** The encoder and decoder are combined into a whole model, with the encoder’s output feeding directly into the decoder.

**Loss Function:** The autoencoder employs a custom loss function that combines the MSE for reconstruction accuracy and the Kullback-Leibler divergence for regularization of the latent space distribution:

$$L = 10^6 \cdot \text{MSE}(y, \hat{y}) + \frac{1}{2} \sum \left( 1 + \log \sigma^2 - \mu^2 - e^{\log \sigma^2} \right) \quad (6.4)$$

where  $y$  denotes the target output and  $\hat{y}$  the reconstructed output.

The model is compiled with the Adam optimizer using a learning rate of 0.001. Training focuses on achieving high fidelity reconstruction while ensuring that the latent space remains smooth and well regularized.

### 6.3.12 VAE Training

The participants’ EEG trials, categorized according to Russell’s model, were combined with corresponding emotion labels to train separate VAEs for each emotional quadrant, yielding distinct latent space representations for each affective state. The performance metrics of these models are summarized in Table 6.16, reporting both the MSE and Mean Absolute Error (MAE) values for each quadrant.

Similarly, subsets of audio samples have been used to train VAEs aimed at generating new sound textures by exploring constructed latent spaces.

### 6.3.13 Limitations

The work presented demonstrates the functionality of the proposed system and its potential applications in computer assisted music composition. However, a dedicated user study is required to evaluate the system’s usability, cognitive load, and the effort and workload associated with user system interaction within the context of HCI. The design and execution of this evaluation will be the focus of future work.

### 6.3.14 Conclusion and Discussion

In this study, we presented a sound generation system driven by the musician’s emotional state through a BCI. Sound synthesis is achieved through the interaction of the models described in Sections 6.3.10, 6.3.11, and 6.3.12. The machine learning models classify the composer’s emotional state every 4 seconds. Once an emotional label is determined, the corresponding EEG VAE model is activated. This model processes the incoming EEG signal, and its encoder produces a compact latent representation. The resulting latent vector defines a trajectory within the latent space of the sound VAE decoder associated with the detected emotional state. Consequently, the VAE corresponding to the identified quadrant generates a matching sound texture, which is continuously looped for playback.

During performance, the musician can use a MIDI controller to audition all generated sounds or selectively trigger specific ones. Notably, the system is compatible with the MUSE headband, as the dataset used for training was collected through this device in a user study. Future work will focus on extending the framework with more advanced architectures such as Transformer based models to enable the generation of increasingly sonic textures.

## 6.4 Exploring the Mental State Intersection by Brain Computer Interfaces, Cellular Automata and Biofeedback

### 6.4.1 Introduction

AI refers to computational techniques that achieve intelligent behaviour with minimal human intervention. Its adoption spans autonomous driving [240], finance [324], agriculture and decision support [126], the Internet of Things (IoT) and security [36], clinical analysis [299] and healthcare [52, 198]. A particularly active frontier uses AI to process EEG signals acquired via BCI, targeting emotion recognition [22], biofeedback [19], motor imagery [56, 217], evoked potentials, and SSVEP based device control [350]. This growth is enabled, in part, by accessible EEG based BCI that deliver acceptable signal quality at low cost.

EEG records spontaneous electrical activity originating from ionic currents in neuronal populations. Multi-electrode recordings taken from the scalp capture ongoing dynamics that shift in response to internal or external stimuli. Beyond detection and

classification, an emerging research direction considers representation: how to convey mental states graphically and sonically in ways that are interpretable and actionable. To the best of our knowledge, simultaneous visual and auditory rendering of mental state from EEG has not been systematically addressed. We approach this challenge with CA.

CAs [64, 283] are discrete, rule based dynamical systems that generate complex spatiotemporal patterns from simple local interactions. Their appeal lies in the emergence of rich global behavior from minimal ingredients, making them suitable for mapping compact feature vectors into expressive visual structures.

**Contribution.** we present a prototype system based on AI, BCI, and CA for mental state description and representation. They are generated by CA and play, in biofeedback mode. The proposed system is a biofeedback enhancement tool that can be used in clinical studies as an application in psychological therapies and musical timbre generation.

**Roadmap.** Section 7.4.2 reviews work on mental state recognition with BCI. Section 6.4.3 details our materials and methods: signal acquisition and preprocessing, EEG to CA mapping, autoencoder modeling, and additive synthesis. Section 6.4.9 illustrates the user interface and visualization. Section 6.4.10 discusses findings and future directions.

## 6.4.2 Related Work

We briefly survey BCI applications to mental state inference using ML/DL.

Borisov *et al.* [37] introduced a method for assessing the functional and mental states of healthy individuals by integrating EEG data, collected using an Emotiv EPOC+ headset, with accelerometer readings. To measure heart rate variability Heart Rate Variability (HRV), the study employed a wireless EEG system featuring 14 channels, an accelerometer, and an electrocardiogram ECG recorder. EEG and ECG signals were recorded from 10 healthy participants under four distinct functional conditions: baseline (rest), TOVA test (mental workload), hyperventilation (physical load), and post test recovery. The experimental findings were validated through HRV analysis and psychological evaluation. The results revealed that suprasegmental control exerted a significant influence on the power of the very low frequency component of the HRV spectrum. Additionally, the study demonstrated a reduction in cortical control effects within the cerebral cortex under specific conditions.

Joadder *et al.* [152] proposed a system for identifying patients' mental states using supervised spatial filtering and four feature families (Katz Fractal Dimension, Sub-band

Energy, Log-Variance, Root Mean Square (RMS)), with LDA classification, achieving an average precision of 84.35% with Katz features.

Chatterjee *et al.* [57] introduced a Multiple Attractor Cellular Automata (MACA) based method to classify left/right motor imagery using bandpass filtering and features such as wavelet energy entropy, band power, and autoregressive coefficients; SVM and ensemble classifiers yielded accuracies up to 91.16% (10-fold CV).

While CA have been used in classification pipelines [57], we find no prior work that employs CA as a visual and sonic renderer of EEG inferred mental state. Our framework closes this gap by coupling CA visualization with autoencoder driven sonification.

### 6.4.3 Materials and Methods

We describe the overall architecture (Section 6.4.4), EEG preprocessing (Section 6.4.5), feature handling and discretization (Section 6.4.6), CA construction (Section 6.4.7), and additive-synthesis mapping in Max/MSP (Section 6.4.8).

### 6.4.4 System Architecture

The following section outlines the main stages of data and signal processing. The user's EEG signal is first acquired using a 4-electrode BCI, specifically the MUSE 2 headset<sup>15</sup>. The EEG signal is recorded over a 5 second window selected as the minimum duration required to ensure reliable preprocessing of EEG data. The acquired signal is then preprocessed to remove artifacts and intrinsic noise components. From the cleaned data, the amplitudes of five canonical frequency bands are extracted: *Delta* (1–4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Sigma* (12–16 Hz), and *Beta* (16–30 Hz).

Each of these five bands is subsequently discretized into two classes, 0 and 1, using the `scikit-learn`<sup>16</sup> library. The resulting binary features are concatenated into a single vector of 20 elements. This vector serves as the initial condition for generating an elementary CA governed by rule 30 over 50 temporal generations. The choice of the rule is left to the user.

Once the automaton development is obtained, an Autoencoder Neural Network (shown in Figure 6.31) is overfitted to obtain the compressed representation of the zero and one matrix constituting the CA. Subsequently, the 16 features from the neural network control both the amplitudes and the multiple frequencies of the 16 sinusoidal

---

<sup>15</sup><https://choosemuse.com/muse-2/>

<sup>16</sup><https://scikit-learn.org/stable/>

oscillators. All this is aimed at the timbre generation of a sound built with the additive synthesis in Max/Msp<sup>17</sup>. Figure 6.28 shows the architecture of the system.

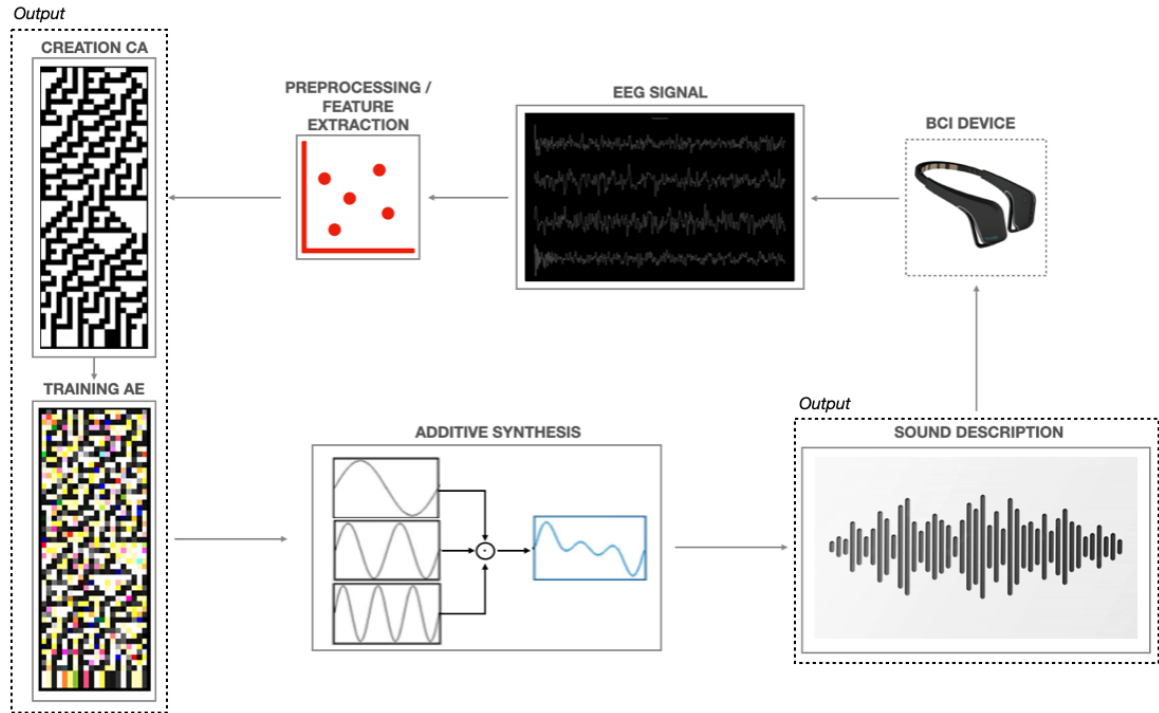


Fig. 6.28 System Architecture.

### 6.4.5 EEG signal preprocessing strategies

Real time utility requires robust artifact handling. Algorithm 2 summarizes the workflow.

### 6.4.6 EEG signal management and processing

First, the acquired EEG signal is bandpass filtered within the 1–30 Hz range and segmented into 1-second epochs. This segmentation facilitates the removal of epochs containing samples that exceed a defined voltage threshold of 0.000005 volts, effectively discarding those affected by ocular or motion related artifacts. Once contaminated epochs are excluded, the SNR can be further enhanced [28].

To achieve this, the XDAWN algorithm [275], implemented in the MNE framework, is applied. XDAWN is a spatial filtering technique that computes a matrix of coefficients

<sup>17</sup><https://cycling74.com/products/max/>

---

**Algorithm 2:** EEG workflow for visual and sonic state rendering.

---

**Functions:**

- *Clean*: real time denoising of raw EEG.
- *FeaturesExtraction*: compute band features from cleaned EEG.
- *CellularAutomata*: generate 1D CA from discretized features.
- *AutoEncoder*: learn compressed representation of CA.
- *AdditiveSynthesizer*: render sound texture from latent code.

**Input:**

- $EEG_{signal}$ : 5 s raw EEG;
- $EEG_{cleaned}$ : denoised EEG;
- $EEG_{features}$ : discretized band features (0–1);
- $EEG_{features\_compressed}$ : autoencoder latent code.

**Result:**

- $GMS$ : graphical mental state representation;
- $SMS$ : sound texture mental state representation.

**while true do**

```
     $EEG_{cleaned} \leftarrow Clean(EEG_{signal})$   
     $EEG_{features} \leftarrow FeaturesExtraction(EEG_{cleaned})$   
     $GMS \leftarrow CellularAutomata(EEG_{features})$   
     $EEG_{features\_compressed} \leftarrow AutoEncoder(GMS)$   
     $SMS \leftarrow AdditiveSynthesizer(EEG_{features\_compressed})$ 
```

**end**

---

optimized for separating signal and noise components. The filter matrix is derived from the covariance matrices of the signal ( $C_s$ ) and noise ( $C_n$ ) as follows:

$$W = C_s(C_s + C_n)^{-1} \quad (6.5)$$

where  $W$  represents the spatial filter coefficients. Applying this matrix to the raw data suppresses unwanted noise, yielding a cleaner EEG signal.

From the filtered and denoised signal, five canonical EEG frequency bands are extracted: *Delta* (1–4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Sigma* (12–16 Hz), and *Beta* (16–30 Hz).

The resulting band power values are then discretized<sup>18</sup> into binary values (0 and 1). These binary outputs are subsequently used as the *initial condition* for generating a Cellular Automaton (CA) governed by `rule_30` (Definition 1).

### 6.4.7 Cellular Automata

CA is a grid of cells updated synchronously by local rules [328, 329]. In one dimension elementary CA, a cell’s next state depends on itself and its two neighbors. With `rule_30`, iterating from a simple seed produces the well known triangular structure (Fig. 6.29).

**Definition 1.** *rule\_30 maps each 3 bit neighborhood  $(l, c, r) \in \{0, 1\}^3$  to a next state bit as follows: 000 → 0, 001 → 1, 010 → 1, 011 → 1, 100 → 1, 101 → 0, 110 → 0, 111 → 0.*

In our proposed system, the creation of the Initial Vector (IV) is its core. According to `rule_30`, the creation of the IV conditions the generation of the cellular automaton. Therefore, we start with an array of 20 features (4 electrodes for 5 bands) with a disposition related to the features’ discretization. We apply an iteration of 50 timesteps to optimize the execution of the process, but nothing prevents exploring longer iterations. The results are shown in Figure 6.30.

To obtain a compact control representation, we train an Autoencoder (Fig. 6.31) to compress the  $50 \times 20$  binary image (flattened) into a 16 element latent vector suitable for real time synthesis control.

---

<sup>18</sup>Using the `KBinsDiscretizer` method from scikit-learn: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.KBinsDiscretizer.html>

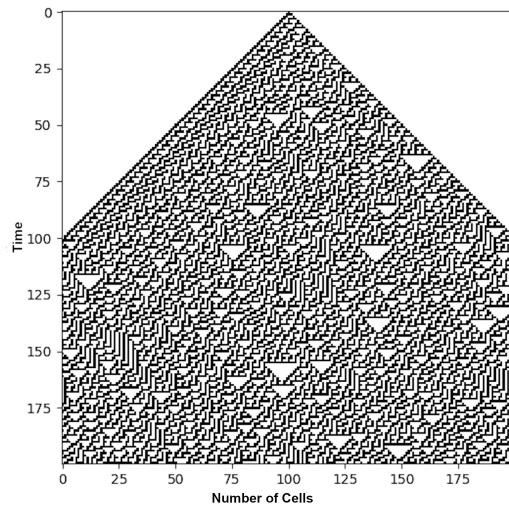


Fig. 6.29 Cellular automaton with rule\_30.

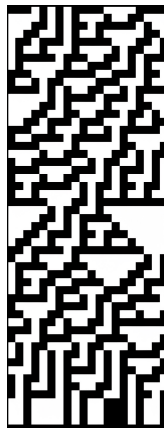


Fig. 6.30 CA generated from BCI features (length 20; 50 time steps).

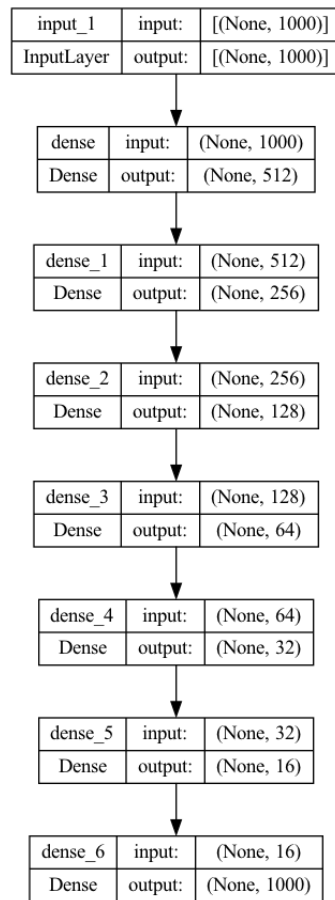


Fig. 6.31 Autoencoder architecture.

## 6.4.8 Additive synthesis

Max/MSP provides a visual environment for rapid audio prototyping. We implement an additive synthesizer comprising 16 sinusoidal oscillators whose partial frequencies and amplitudes are modulated by the 16-D latent vector (Fig. 6.32). Tone depends on the number and type of partials, their frequency relationships, and the amplitude envelope. A random modulator (0–6, at 10 ms) perturbs a user selected fundamental in the Basic Tones section (Fig. 6.34); the resulting MIDI value is converted to frequency via `mtof`. Amplitudes are scaled between  $-69$  dB and  $-3$  dB using `scale`, with latent values normalized via `sklearn.preprocessing.normalize`.

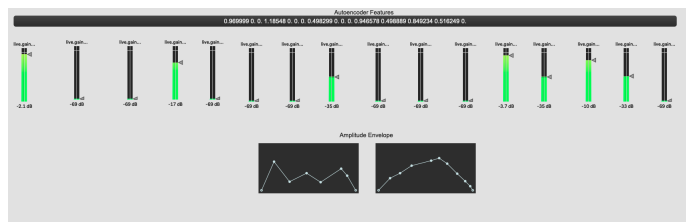


Fig. 6.32 16 oscillators with partial amplitude sliders (dB) and global envelopes.

## 6.4.9 Data Visualization

Figures 6.33 and 6.34 show the User Interface (UI). Fig. 6.33 illustrates training, where the autoencoder learns a compressed code that reconstructs the CA. Fig. 6.34 displays the performance state: perfect reconstruction of the CA, selection of a fundamental pitch, and dry/wet controls for the rendered sound texture.

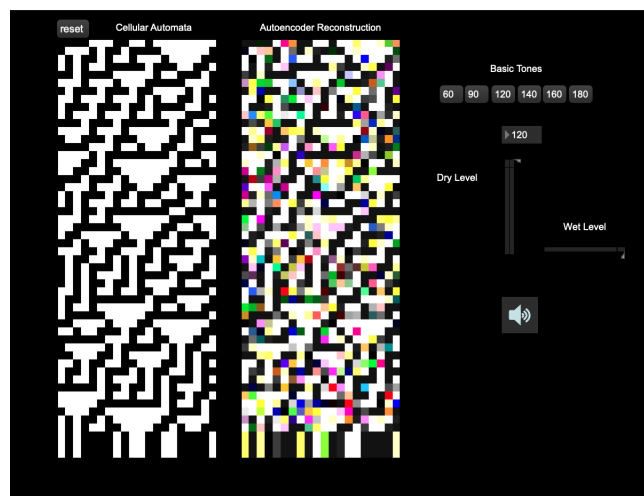


Fig. 6.33 Autoencoder training: compressing the original CA.

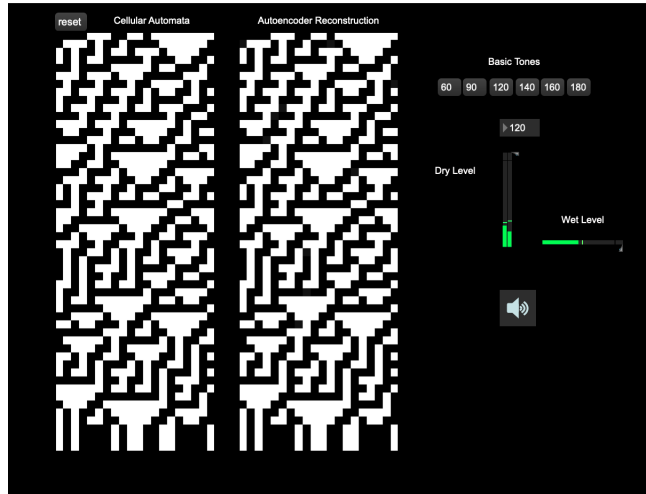


Fig. 6.34 Execution and control of the final sound texture.

#### 6.4.10 Discussion and further directions

In this paper, we propose an approach demonstrating how BCI and AI can be combined to develop a biofeedback tool that integrates both graphical and auditory representations. By employing CA within specific time intervals, users can perceive: (i) variations (or stability) in their mental states, and (ii) the corresponding visual patterns generated by the automaton alongside their sonic representations. In our study, a 5 second EEG epoch is defined as the temporal unit representing the user's mental state in real time. However, further investigation is needed to assess the adequacy of this temporal window for accurately capturing cognitive dynamics. Our experimental results demonstrate the feasibility of simultaneously representing real time mental states through two complementary modalities: a graphical visualization and a personalized sound texture. Moreover, the use of CA enables the exploitation of the matrices derived from automaton evolution as a potential method for feature augmentation and signal characterization. Future work will focus on a deeper analysis to enhance the system's ability to characterize and classify mental states more precisely, thereby improving the robustness and interpretability of the proposed framework.



# Chapter 7

## Other research contributions

### 7.1 MORIX: An ML–XAI Framework to Predict Mortality in MAFLD

#### 7.1.1 Overview

Metabolic dysfunction associated fatty liver disease (MAFLD) updates fatty liver diagnosis independently of alcohol intake and viral hepatitis [101, 102]. Given the substantial burden associated with NAFLD/MAFLD including cardiovascular, metabolic, renal, and all cause mortality risks [108, 188] and European prevalence around 23.7% (near 30% in Southern Italy) [77, 224, 247, 266, 339], accurate risk stratification is clinically relevant. While MAFLD is linked to CVD and malignancies, its independent association with overall mortality remains debated [287]. In parallel, ML is increasingly explored as a cost-effective complement to imaging and specialized testing [45, 48–50, 182, 291, 300], including NAFLD detection [298, 299]. Motivated by evidence that lifestyle interventions can improve NAFLD [79], we aim to improve all cause mortality prediction to support proactive counseling. (We retain MAFLD/NAFLD terminology for consistency with source cohorts [82].)

#### Contributions

- **Integrated ML–XAI:** End-to-end pipeline (RFE–RF + multi-model benchmarking + SHAP) for MAFLD mortality risk.
- **Parsimony:** A compact feature subset enhances generalizability and interpretability.

- **Transparency:** Global and local SHAP explanations support clinician trust and decision-making.
- **Deployment:** A web interface provides real time, patient level risk and explanations.

### 7.1.2 Related work

Studies have characterized MAFLD versus NAFLD in terms of CKD risk and metabolic imbalances [191, 305], and compared non invasive indices for mortality [84]. Population analyses reported differential mortality across MAFLD/NAFLD strata [141, 162, 236]. ML has been applied to MAFLD related CVD mortality [95]. Unlike prior work, MORIX delivers an end to end, explainable pipeline for all-cause mortality using readily accessible, routinely collected variables.

### 7.1.3 Proposed framework

Figure 7.1 summarizes the architecture. The Data Preparer cleans, normalizes, rebalances, and selects features; the Predictor trains/tunes multiple ML models and outputs hold out predictions; the Explainer provides SHAP based global/local attributions; the User Interface visualizes risks and explanations for clinical use.

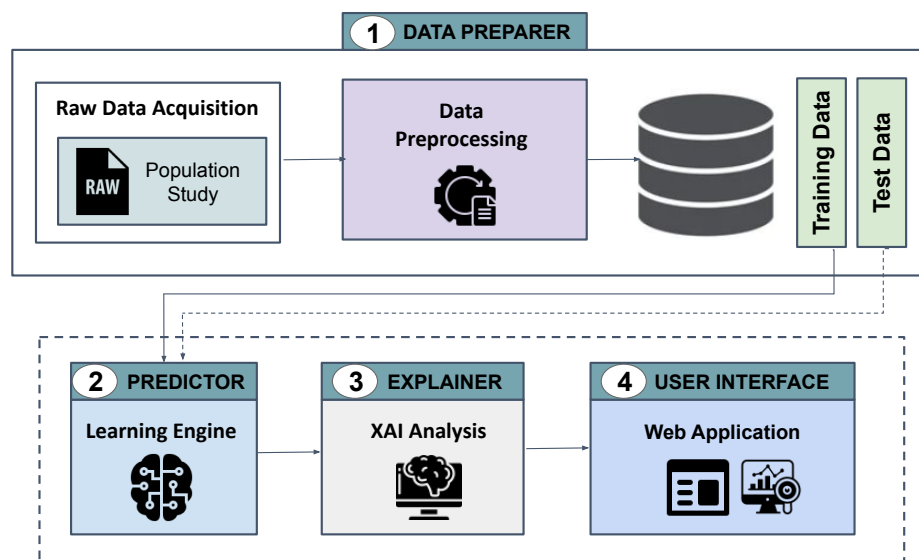


Fig. 7.1 MORIX classification framework architecture.

### 7.1.4 Proposed method

We analyze 1,675 MAFLD subjects  $> 30$  years (543 women, 1,131 men) from MICOL/NUTRIHEP [223], recruited from January 2005 to December 31, 2020, with standardized assessments and regulatory approvals (MICOL: DDG-CE-589/2004, -782/2013; NUTRIHEP: DDG-CE-502/2005, -792/2014) in accordance with the Helsinki declaration [246, 288]. Hepatic steatosis was assessed by ultrasonography (Hitachi H21 Vision, 3.5 MHz). We consider 30 biochemical, anthropometric, and sociodemographic features; the analytic dataset includes 1,674 rows (target included).

### 7.1.5 Preprocessing and selection.

Using Python/Pandas,<sup>1</sup> we removed missing data (final  $n = 1,459$ ) and addressed class imbalance (No: 1,374; Yes: 301) via random undersampling to  $N = 528$  with imbalanced-learn.<sup>2</sup> Features were standardized with RobustScaler<sup>3</sup>. RFE with RF was embedded in a grid search pipeline with 5-fold CV<sup>4</sup> to avoid double dipping. Six features were retained: Weight, Cholesterol, Blood Glucose, Alkaline Phosphatase, LDL, Age.

### 7.1.6 Models and evaluation.

We compared MLP [26], RF [249], SVM [241], XGBoost [61], and LGBM [106] using scikit-learn v1.4.2,<sup>5</sup> tuning via 5-fold grid search. Metrics include Accuracy, ROC AUC, Precision, Recall, and F1 (Sec. 7.1.8). The formal pipeline is summarized in Algorithm 3.

### 7.1.7 Experimental results

An 80/20 split with `random_state=42` was used.<sup>6</sup> All models underwent 5-fold grid search optimizing Accuracy (Table 7.7).

---

<sup>1</sup><https://pandas.pydata.org/>

<sup>2</sup><https://imbalanced-learn.org/stable/index.html>

<sup>3</sup><https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>

<sup>4</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

<sup>5</sup><http://scikit-learn.org>

<sup>6</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

---

**Algorithm 3:** Algorithm of MORIX framework

---

1: **Functions:**2: *splitting\_dataset*( $D$ ): Function to split the dataset  $D$ .3: *clean\_data*( $D$ ): Function to remove records with missing values from  $D$ .4: *balance\_datasets*( $D$ ): Function to equalize the number of records for each class in  $D$ .5: *train\_model*( $D$ ): Function to train a model on the dataset  $D$ .6: *predict\_outcomes*( $model, D$ ): Function to predict outcomes using the model on the dataset  $D$ .7: *compute\_shap\_values*( $model, instance$ ): Function to calculate SHAP values for the instance using the model.8: **Inputs:**9:  $D$ : Original dataset.10: **Outputs:**11: *predictions*: Predicted outcomes for the test set.12: *shap\_values*: SHAP values for the test instances.13: **procedure** SETUP PHASE14:  $D_{train}, D_{test} \leftarrow \textit{splitting\_dataset}(D)$ 15:  $D_{train} \leftarrow \textit{clean\_data}(D_{train})$ 16:  $D_{test} \leftarrow \textit{clean\_data}(D_{test})$ 17:  $D_{train} \leftarrow \textit{balance\_datasets}(D_{train})$ 18:  $D_{test} \leftarrow \textit{balance\_datasets}(D_{test})$ 19:  $model \leftarrow \textit{train\_model}(D_{train})$ 20:  $predictions \leftarrow \textit{predict\_outcomes}(model, D_{test})$ 21: **procedure** SUMMARIZATION PHASE22: **For** each instance in  $D_{test}$ 23:      $shap\_value \leftarrow \textit{compute\_shap\_values}(model, instance)$ 24:     Store instance, prediction, and *shap\_value*25: **End For**26: **procedure** FINALIZATION PHASE27:     Make the summarized data available for further use or analysis.

---

Table 7.1 Hyperparameter list and corresponding values for the models.

Algorithm	Hyperparameter	Values
Multilayer Perceptron	random_state	{42}
	hidden_layer_sizes	{(50,), (100,), (50,50), (100,150), (100,100)}
	activation	{relu, tanh, logistic}
	solver	{sgd, adam}
	learning_rate	{constant, adaptive}
Random Forest	random_state	{42}
	n_estimators	{50, 100, 200}
	max_depth	{10, 20, 30, None}
	min_samples_split	{2, 5, 10}
	min_samples_leaf	{1, 2, 4}
Support Vector Machine	random_state	{42}
	C	{0.1, 1, 10, 100}
	gamma	{0.1, 0.01, 0.001}
	kernel	{linear, rbf, poly}
XGBoost	random_state	{42}
	max_depth	{3, 4, 5}
	learning_rate	{0.1, 0.01, 0.001}
	n_estimators	{100, 200, 300}
LightGBM	random_state	{42}
	learning_rate	{0.1, 0.01, 0.001}
	num_leaves	{31, 50, 100}
	max_depth	{-1, 10, 20, 30, 300}

RF exhibited the most balanced performance and the fewest Yes class errors (Table 7.2, Fig. 7.2).

Table 7.2 Mortality prediction results (No/Yes denote classes).

		Classifier				
		RF	XGB	MLP	LGBM	SVM
Precision	No	0.83	0.81	0.78	0.78	0.78
	Yes	0.83	0.84	0.77	0.84	0.82
Recall	No	0.82	0.85	0.76	0.85	0.82
	Yes	0.84	0.80	0.79	0.76	0.78
F1 score	No	0.83	0.83	0.77	0.81	0.80
	Yes	0.83	0.82	0.78	0.80	0.79
Accuracy		0.83	0.82	0.77	0.81	0.80
AUC		0.88	0.88	0.88	0.88	0.86

Output Class	0	60 37.7%	19 11.9%	75.9% 24.1%
	1	17 10.7%	63 39.6%	78.8% 21.3%
		77.9% 22.1%	76.8% 23.2%	77.4% 22.6%
		0	1	Target Class

(a) MLP

Output Class	0	65 40.9%	14 8.8%	82.3% 17.7%
	1	13 8.2%	67 42.1%	83.8% 16.2%
		83.3% 16.7%	82.7% 17.3%	83.0% 17.0%
		0	1	Target Class

(b) RF

Output Class	0	65 40.9%	14 8.8%	82.3% 17.7%
	1	18 11.3%	62 39.0%	77.5% 22.5%
		78.3% 21.7%	81.6% 18.4%	79.9% 20.1%
		0	1	Target Class

(c) SVM

Output Class	0	67 42.1%	12 7.5%	84.8% 15.2%
	1	16 10.1%	64 40.3%	80.0% 20.0%
		80.7% 19.3%	84.2% 15.8%	82.4% 17.6%
		0	1	Target Class

(d) XGB

Output Class	0	67 42.1%	12 7.5%	84.8% 15.2%
	1	19 11.9%	61 38.4%	76.2% 23.8%
		77.9% 22.1%	83.6% 16.4%	80.5% 19.5%
		0	1	Target Class

(e) LGBM

Fig. 7.2 Confusion matrices on the test set. 0 = Mortality (No), 1 = Mortality (Yes).

### 7.1.8 Evaluation metrics

Let  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  denote counts. We report:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (7.1)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (7.2)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (7.3)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (7.4)$$

AUC is the probability that a randomly chosen positive is ranked above a negative [44]:

$$\text{AUC} = \frac{\sum_{x^- \in X^-} \sum_{x^+ \in X^+} \mathbb{1}\{f(x^-) < f(x^+)\}}{|X^-| |X^+|}.$$

### 7.1.9 Best model performance

RF achieved the strongest class balance (Fig. 7.2b, Table 7.2), with top Recall (0.84) and F1 (0.83) for *Yes*, and top Precision/F1 (0.83/0.83) for *No*. It yielded the fewest *Yes*-class errors and the highest Accuracy (0.83). XGB provided the best *No*-class Recall (0.85) and *Yes*-class Precision (0.84). SVM had the highest AUC (0.89) but less favorable class-wise balance. We therefore select RF, prioritizing minimization of *Yes* misclassification. Best RF hyperparameters: `max_depth=10`, `min_samples_leaf=4`, `min_samples_split=10`, `n_estimators=100`, `random_state=42`. ROC curves are shown in Fig. 7.3.

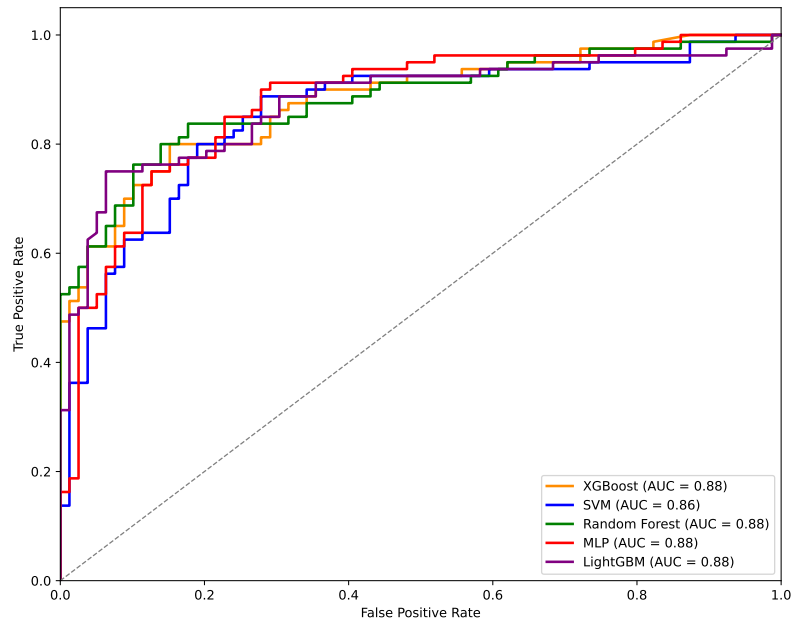


Fig. 7.3 ROC curves and AUC per model.

### 7.1.10 Statistical analysis

Bootstrap (1,000 replicates) estimated accuracy distributions [100]. Paired  $t$ -tests against RF (Table 7.3) indicate significantly higher RF accuracy (all  $p < 0.01$ ). Distributional summaries are in Figs. 7.4–7.5.

Table 7.3 Paired  $t$ -tests on bootstrapped accuracies (RF vs. others).

Model	T-statistic	P-value	Mean Difference (95% CI)
XGBoost	2.62	$< 0.01$	0.004 (−0.082, 0.082)
MLP	20.03	$< 0.01$	0.028 (−0.057, 0.113)
LightGBM	17.20	$< 0.01$	0.023 (−0.063, 0.101)
SVM	19.94	$< 0.01$	0.028 (−0.057, 0.113)

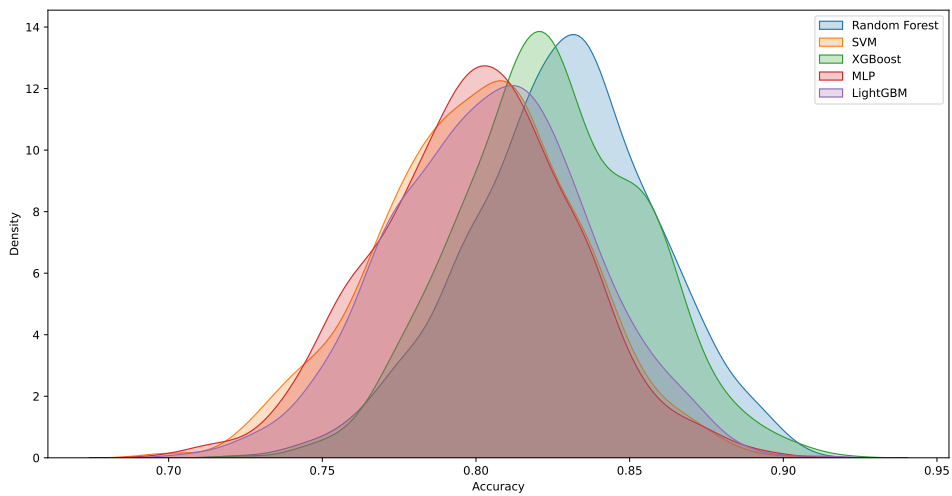


Fig. 7.4 Bootstrap accuracy distributions per model.

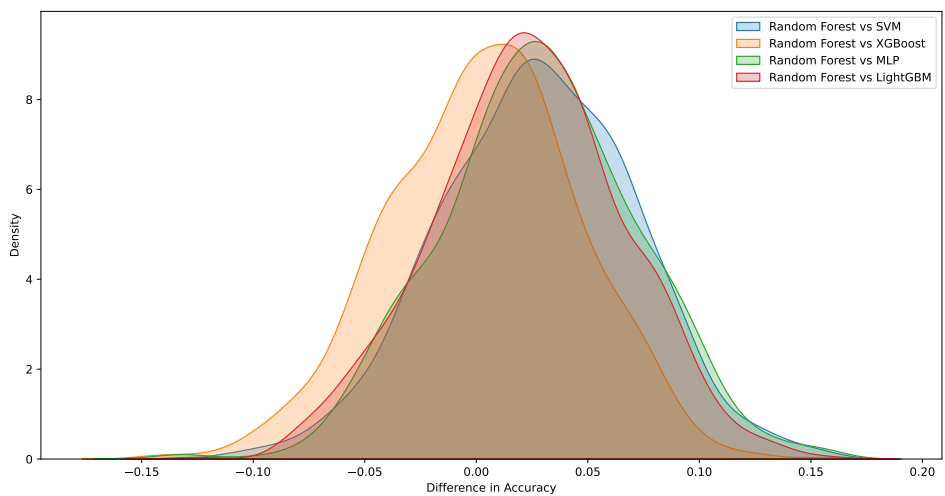


Fig. 7.5 Bootstrap accuracy differences: RF vs. other models.

### 7.1.11 Explainability results

SHAP analyses of the selected RF clarify decision logic. Global importance and directionality (Fig. 7.6) show Age as the dominant predictor, with Weight contributing least among selected features. Cumulative contribution (Fig. 7.7) indicates Age explains  $\sim 60\%$ , and adding Blood Glucose surpasses 70%; LDL, Alkaline Phosphatase,

Cholesterol, and Weight add incremental value. Local waterfalls illustrate case level reasoning for *Yes/No* outcomes (Figs. 7.8, 7.9).

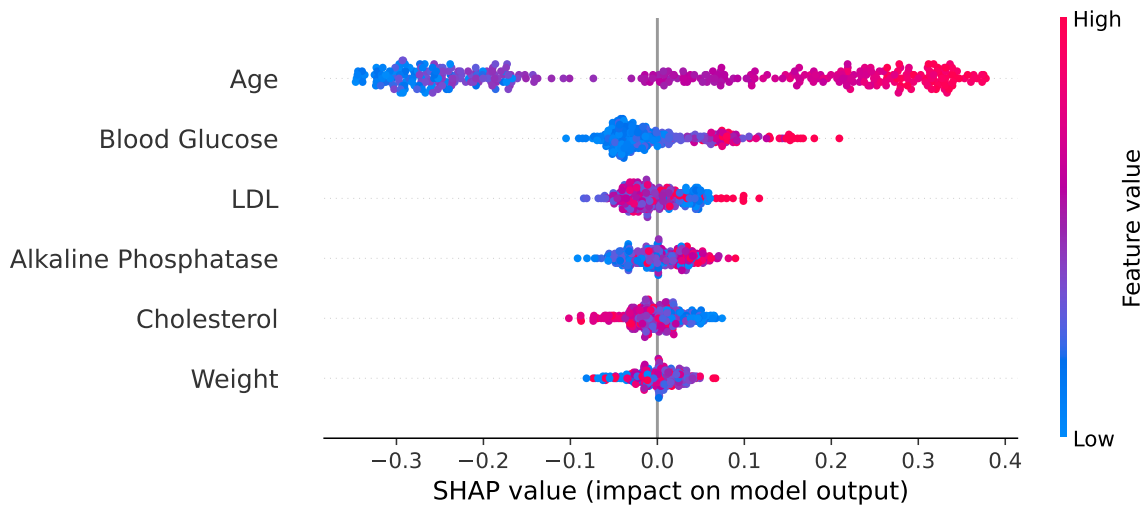


Fig. 7.6 Global SHAP analysis of anthropometric/biochemical predictors.

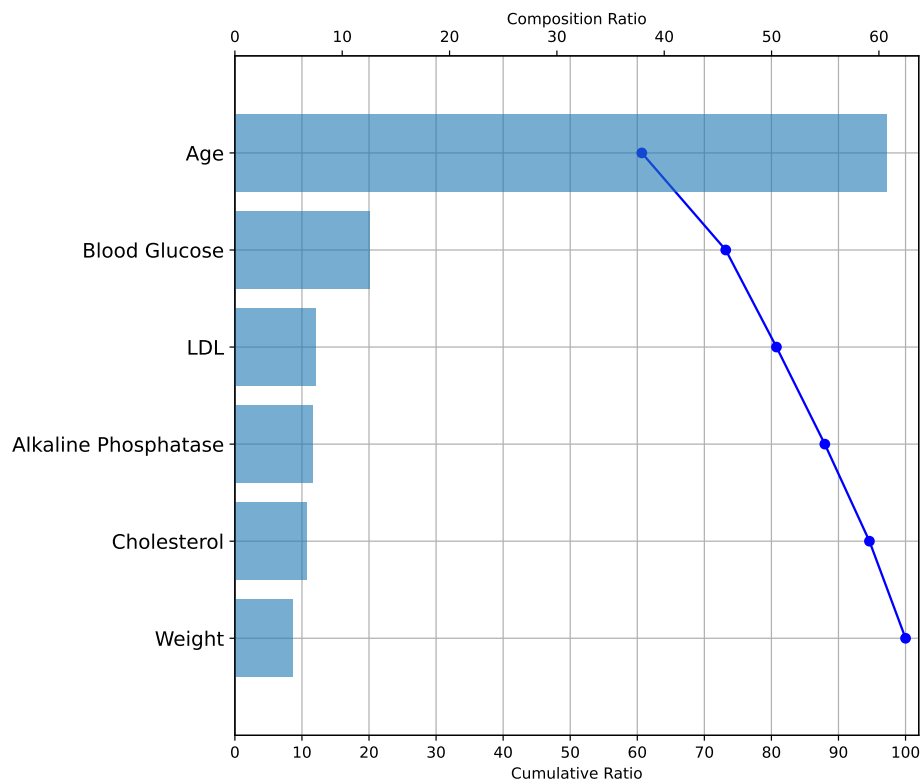


Fig. 7.7 Cumulative explanatory contribution of selected features (RF).

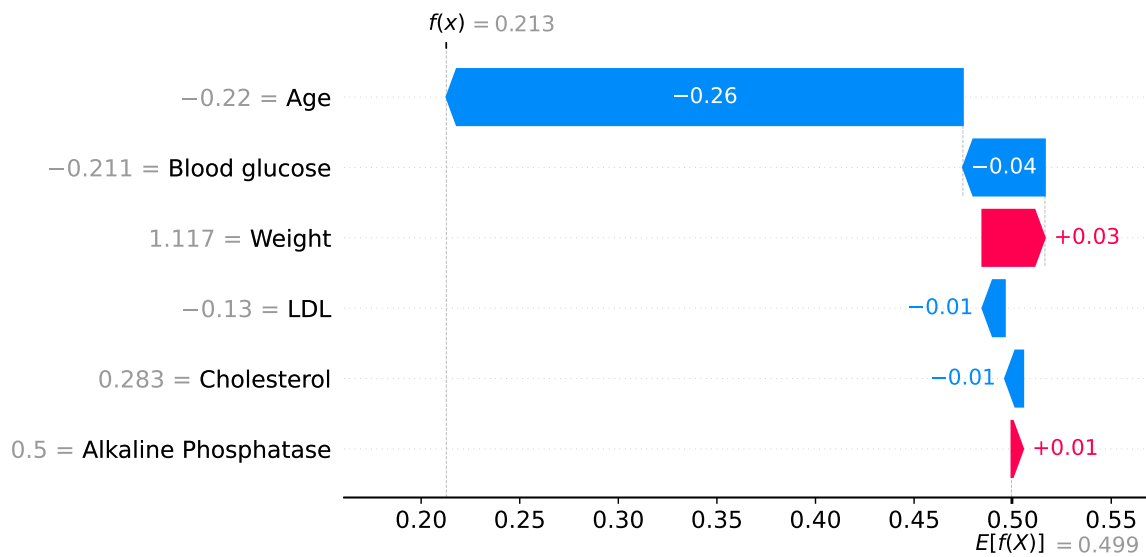


Fig. 7.9 Local SHAP analysis for a *Mortality (No)* prediction (RF).

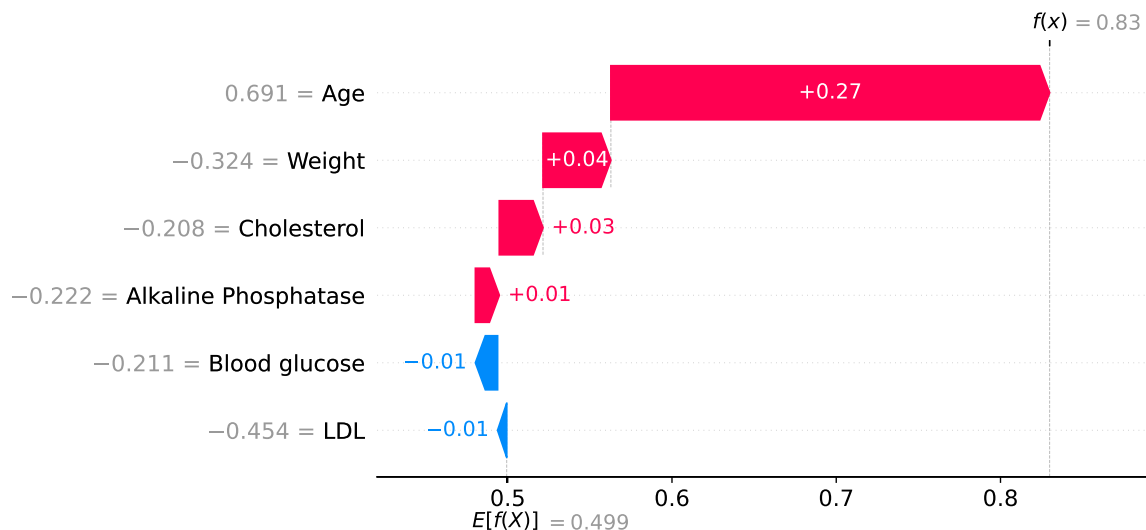


Fig. 7.8 Local SHAP analysis for a *Mortality (Yes)* prediction (RF).

### 7.1.12 MORIX web user interface

The trained RF is serialized with `pickle`<sup>7</sup> and served via `Flask`,<sup>8</sup> with `joblib` for model loading.<sup>9</sup> The front end (HTML/CSS [97]) collects inputs and renders predicted risk and SHAP waterfalls (Figs. 7.10–7.12).

<sup>7</sup><https://docs.python.org/3/library/pickle.html>

<sup>8</sup><https://flask.palletsprojects.com/en/3.0.x/>

<sup>9</sup><https://joblib.readthedocs.io/en/stable/>

## Mortality Prediction in MAFLD Subjects

Weight (kg)

Cholesterol (mg/dL)

Blood Glucose (mg/dL)

Alkaline Phosphatase (U/L)

LDL (mg/dL)

Age (years)

Predict

Fig. 7.10 MORIX web UI home: mortality risk prediction for MAFLD.

## Prediction Result

Prediction: Risk Yes

Probability: 97%

### SHAP Waterfall Plot



Fig. 7.11 MORIX UI: *Mortality (Yes)* example with 97% probability.

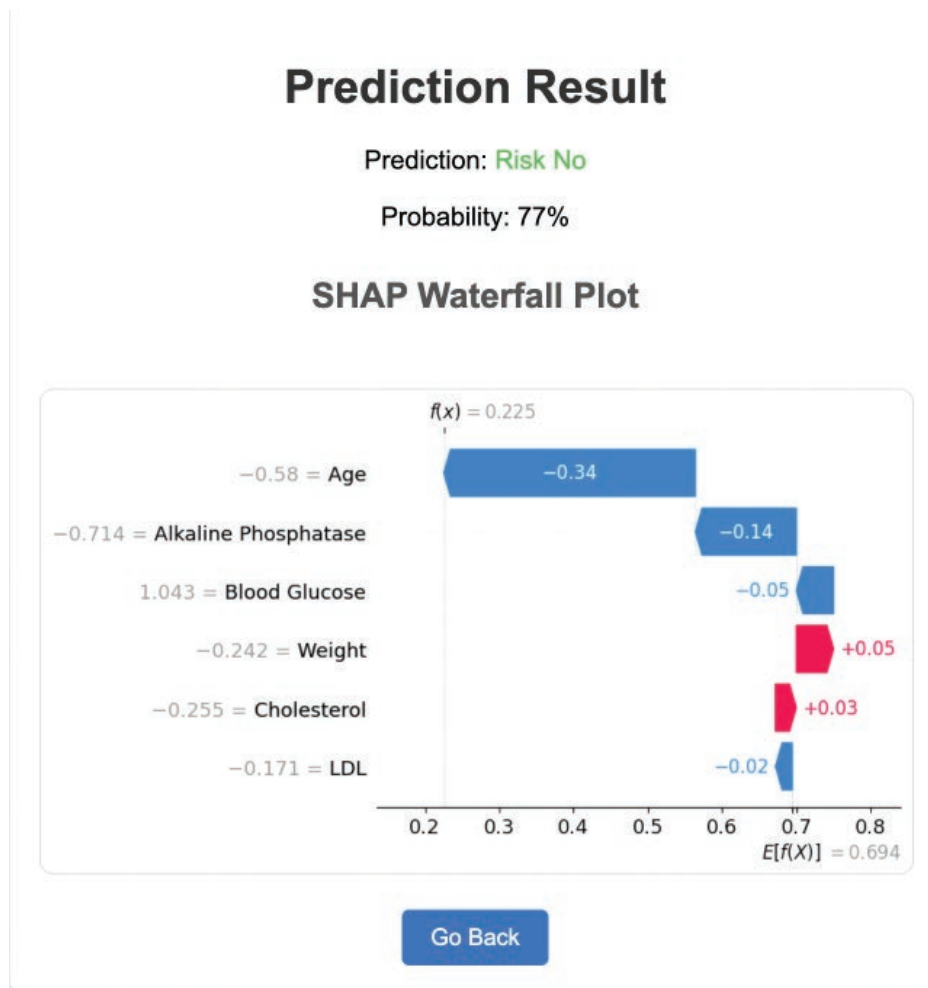


Fig. 7.12 MORIX UI: *Mortality (No)* example with 77% probability.

### 7.1.13 Conclusion and further directions

MORIX provides an end to end, explainable pipeline for mortality prediction in MAFLD using routinely collected variables. Among five classifiers, RF offered the best balance across classes, with Precision/Recall/F1 (0.83,0.82,0.83) for *NO* label and (0.83,0.84,0.83), Accuracy 0.83, and AUC 0.88 for *Yes* labels. Table 7.2 displays all results. SHAP confirmed Age as the dominant driver, with additional contributions from Blood Glucose, LDL, Alkaline Phosphatase, Cholesterol, and Weight. By exposing the rationale of predictions, MORIX supports individualized counseling aligned with prior evidence [79]. Limitations include initial class imbalance and the scarcity of directly comparable AI benchmarks for MAFLD mortality. Future work will investigate deep learning models, multi center data sharing, and strengthened app scalability, validation, and security.

## 7.2 Machine learning in Mild Cognitive Impairment Prediction

### 7.2.1 Introduction

This study investigates the prediction of neurodegenerative outcomes, with a focus on mild cognitive impairment (MCI). Advanced machine learning (ML) techniques are implemented and systematically compared using large-scale healthcare data from the Salus in Apulia cohort to identify early predictive patterns of neurodegeneration. The aim is to improve early diagnostic frameworks by identifying robust markers from routinely collected epidemiological variables.

### 7.2.2 Background on MCI

Initially introduced by Petersen *et al.* in the late 1990s [257], Mild Cognitive Impairment (MCI) denotes an age independent decline in cognitive abilities insufficient to meet dementia criteria. It may affect memory, language, attention, and other domains and is a known risk factor for dementia. The Alzheimer’s Association estimates that approximately 15–20% of people aged 65 and older have MCI<sup>10</sup>.

### 7.2.3 Study Aim and Contributions

We propose a Pareto optimality based approach to select the best predictive model for MCI. Using the *Salus in Apulia* dataset [51], we train multiple classifiers on easily accessible clinical and epidemiological features. Our contributions are:

- A predictive pipeline for MCI using readily available healthcare data;
- A Pareto based model selection procedure that balances overall accuracy with sensitivity (Recall) to *Yes MCI* cases, minimizing clinically costly misclassifications.

### 7.2.4 Pareto Optimality and Hypervolume

We frame model selection as a multi objective optimization problem (MOOP) [212]:

---

<sup>10</sup>[https://www.alz.org/alzheimers-dementia/what-is-dementia/related\\_conditions/mild-cognitive-impairment](https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment)

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{f}(\mathbf{x}) = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})\} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{7.5}$$

Here,  $\mathbf{x} \in \mathbb{R}^n$  collects  $n$  *decision variables*,  $\mathcal{X} \subseteq \mathbb{R}^n$  is the feasible set defined by constraints, and  $\mathbf{f}(\cdot)$  maps solutions to  $k \geq 2$  scalar objective functions  $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$  (objective space  $\mathbb{R}^k$ ).

In MOOPs, there is typically no single global optimum; instead, the *Pareto dominance* relation is used [205]. A solution  $\mathbf{x}^*$  dominates  $\mathbf{x}$  (denoted  $\mathbf{x}^* \prec \mathbf{x}$ ) if  $f_i(\mathbf{x}^*) \leq f_i(\mathbf{x})$  for all  $i \in \{1, \dots, k\}$  and there exists at least one strict inequality. A solution is *Pareto optimal* if it is not dominated by any other feasible solution. The set of non-dominated solutions forms the *Pareto optimal set*, whose image in objective space is the *Pareto frontier* [250].

To select a single representative solution from the Pareto frontier, we use the *hypervolume* indicator [351]. Given a Pareto-optimal point  $\mathbf{x}^* \in \mathbb{R}^k$ , a reference point  $\mathbf{r} \in \mathbb{R}^k$ , and Lebesgue measure  $\lambda$ , the hypervolume is  $\mathcal{HV} = \lambda(\{\mathbf{x} \in \mathbb{R}^k \mid \mathbf{x}^* \prec \mathbf{x} \prec \mathbf{r}\})$ . Intuitively, it measures the dominated volume between  $\mathbf{x}^*$  and  $\mathbf{r}$ ; larger  $\mathcal{HV}$  indicates a better trade-off across objectives.

## 7.2.5 Dataset Description

We analyzed participants from the *Salus in Apulia Study*. The dataset comprises  $N = 1,929$  observations with 70 continuous/categorical features spanning biochemical, anthropometric, and cognitive assessments obtained via interviews, blood analyses, and clinical visits. After removing irrelevant categorical features, except sex, we retained 57 predictors, including the target.

Data were parsed with `pandas`<sup>11</sup>. To manage class imbalance, we applied Random-Under-Sampler from `imblearn` framework [183], yielding  $N = 818$  (414 *No MCI*, 414 *Yes MCI*). We split the data into training and test sets (80/20). Continuous variables were standardized with `StandardScaler`<sup>12</sup> to improve learning stability [114]. We performed univariate feature selection with `SelectKBest` (`scikit-learn`) using a 5-fold `GridSearchCV`<sup>13</sup> and a Random Forest classifier [249] to choose  $k \in \{1, \dots, 57\}$ . The best  $k$  was 9, selecting: Pure Tone Audiometry (PTA)(0.5–2kHz) left/right,

<sup>11</sup><https://pandas.pydata.org/>

<sup>12</sup><https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

<sup>13</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

Total Cholesterol, HDL Cholesterol, Mini Mental State Examination (MMSE), Frontal Assesment Battery (FAB) raw score, Vitamin D, mean PTA(0.5–2kHz), and mean PTA(8kHz). Although *sex*/gender were not selected, we retained them for clinical relevance [342].

## 7.2.6 Classification Models

We compared the following classifiers: MLP [26], RF [249], SVM [241], and XGB [61]. Models were implemented in Python using `scikit-learn` [254] and tuned via 5 fold grid search to maximize Area Under the Receiver Operating Characteristic Curve (AUC). The hyperparameter grids are listed in Table 7.4.

Table 7.4 Hyperparameter list and values for the classification models.

Algorithm	Hyperparameter	Values
<b>Multilayer Perceptron</b>	seed	{42}
	hidden_layer_sizes	{(50,50,50), (50,100,50), (100,,)}
	activation	{tanh, relu}
	solver	{sgd, adam}
	alpha	{np.arange(0.1, 1, 0.2)}
	learning_rate	{constant, adaptive}
<b>Random Forest</b>	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	max_features	{sqrt, log2}
	max_depth	{np.arange(1, 30, 1)}
	criterion	{gini, entropy}
<b>Support Vector Machines</b>	seed	{42}
	kernel	{rbf, poly}
	gamma	{0.1, 0.001, 0.3, 0.003, 0.5, 0.05}
<b>eXtreme Gradient Boosting</b>	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	learning_rate	{np.arange(0.1, 1, 0.2)}
	scale_pos_weight	{1.000}

## 7.2.7 Pareto Optimality Based Selection

Model configurations define points in a two objective space: overall Accuracy (x-axis) and Recall on Yes MCI (y-axis). For each model, we compute the Pareto frontier (black points in Fig. 7.13; red points are dominated). We then merge frontiers across models

to obtain a global Pareto frontier (Fig. 7.14) and use the hypervolume indicator to select the single best solution.

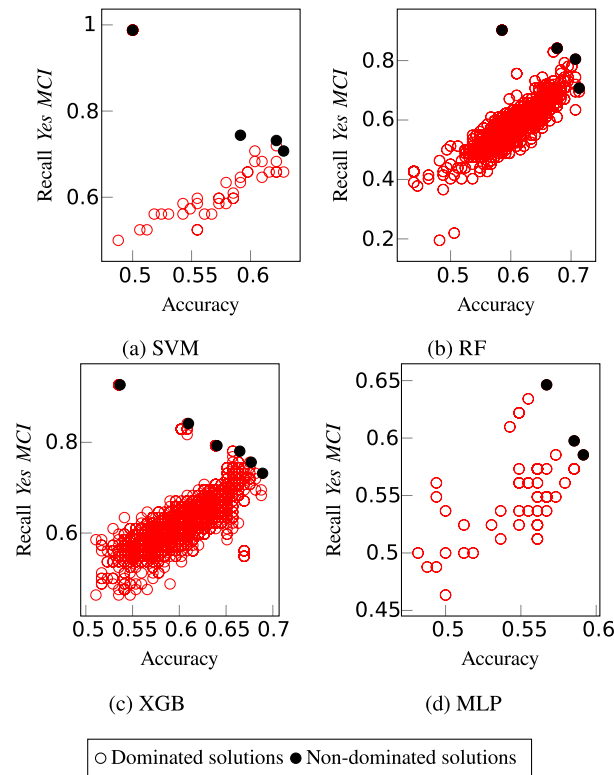


Fig. 7.13 Per-model objective spaces (Accuracy vs. Recall for *Yes MCI*). Black points lie on the model’s Pareto frontier; red points are dominated.

## 7.2.8 Results

Table 7.5 reports the hypervolume  $\mathcal{HV}$  for all solutions lying on the global Pareto frontier in Fig. 7.14. Six configurations are Pareto optimal. No MLP configuration lies on the frontier; four of the six solutions are Random Forest variants. Two RF settings achieve the highest  $\mathcal{HV}$  and are therefore shortlisted. Comparing their precision/recall/F1 by class (Table 7.6), we select RF<sub>2</sub> as the overall best via multi-criteria evaluation: it balances accuracy (0.71) and strong performance on *Yes MCI* (Precision 0.67, Recall 0.80, F1 0.73), with improved metrics over RF<sub>1</sub> except for a marginally lower recall on *Yes MCI*.

Table 7.5 Pareto-optimal configurations and performance. Best  $\mathcal{HV}$  values in bold. RF<sub>1</sub>: Entropy, depth=3, max-feat=log2, estimators=1; RF<sub>2</sub>: Entropy, depth=3, max-feat=sqrt, estimators=9; RF<sub>3</sub>: Entropy, depth=1, max-feat=sqrt, estimators=1; SVM<sub>1</sub>:  $C=6$ ,  $\gamma=0.003$ , kernel=poly; XGBoost<sub>1</sub>: lr=0.1, depth=1, estimators=1; RF<sub>4</sub>: Entropy, depth=5, max-feat=log2, estimators=13.

Model	Accuracy	Recall <i>Yes MCI</i>	$\mathcal{HV}$
RF <sub>1</sub>	0.68	0.84	<b>0.57</b>
RF <sub>2</sub>	0.71	0.80	<b>0.57</b>
RF <sub>3</sub>	0.58	0.90	0.53
SVM <sub>1</sub>	0.50	0.99	0.49
XGBoost <sub>1</sub>	0.54	0.93	0.50
RF <sub>4</sub>	0.71	0.71	0.50

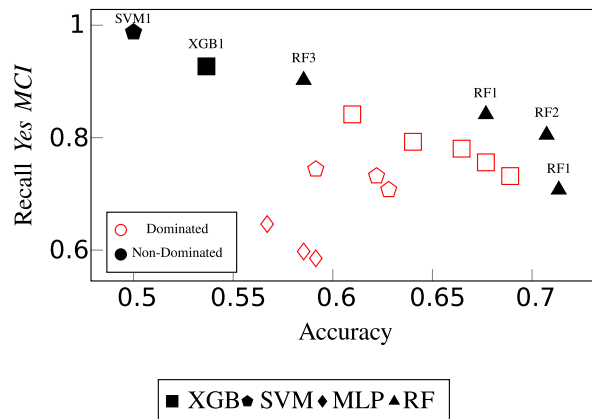


Fig. 7.14 Global objective space (Accuracy vs. Recall for *Yes MCI*) across all models. Black marks: Pareto-optimal; red: dominated. Shapes encode model families.

Table 7.6 Performance of top Random Forest candidates on the test set.

Model	Class	Precision	Recall	F1-score	Accuracy
RF <sub>1</sub>	No MCI	0.76	0.51	0.61	0.68
	Yes MCI	0.63	0.84	0.72	
RF <sub>2</sub>	No MCI	0.76	0.61	0.67	0.71
	Yes MCI	0.67	0.80	0.73	

Our Pareto based selection framework enables principled, multi criteria model choice for MCI prediction. Rather than relying solely on accuracy, we explicitly

prioritize sensitivity to *Yes MCI*, which is critical for screening and early intervention. The selected model (RF<sub>2</sub>) attains Accuracy = 0.71 and, for *Yes MCI*, Precision = 0.67, Recall = 0.80, and F1 = 0.73. These results compare appreciatively with prior work [40, 119], though external validation on larger cohorts remains necessary. In practice, this strategy supports epidemiological screening by leveraging a compact set of accessible features, potentially shortening the time to assessment and easing the burden on specialized diagnostic pathways. The Pareto/hypervolume framework is general and can be reused to tailor model selection to other high risk clinical targets where recall is paramount.

## 7.3 ECG Based Machine Learning for Early Heart Failure Detection with Explainable AI

### 7.3.1 Introduction

This work investigates the use of AI and ML in detecting Cardiovascular Diseases (CVD), focusing specifically on Heart Failure (HF). Heart failure is a primary contributor to morbidity and mortality and poses substantial challenges to conventional diagnostic approaches that depend on patient history and physical examination. This research assesses ECG based methods that extract temporal and morphological features from signals and compares multiple ML classifiers for heart failure screening. The study introduces an Light Gradient Boosting Machine (LGBM) model to distinguish Normal Sinus Rhythm (NSR) from Arrhythmia (ARR) with high accuracy. Additionally, XAI techniques are integrated to improve the clinical interpretability of model predictions.

With the expansion of big data and wearable sensing, ML pipelines have growing potential for scalable, real time HF detection. Our objectives are twofold: (i) build a robust ECG feature-based classifier; (ii) employ post hoc XAI to elucidate model decisions for clinicians. Using the MIT-BIH PhysioNet database, we evaluate the following research questions (RQs):

- **Research Question (RQ)1:** Can newly extracted ECG features train a model to accurately predict HF?
- **RQ2:** Which classifier trained on these features is most effective for HF prediction?
- **RQ3:** Can subject level predictions from the best model be explained to support diagnosis?

All RQs target real time HF prediction using routinely derivable features.

### 7.3.2 Related Work

Merdjanovska *et al.* [221] train a CNN for three class arrhythmia classification across four annotated ECG datasets, addressing class imbalance with SMOTE and enforcing patient level splits to prevent leakage. They report substantially lower PPV/TPR/F1 than accuracy especially in cross database tests likely due to heterogeneous leads and dataset shifts. Sannino *et al.* [281] use the MIT–BIH Arrhythmia dataset with denoising, peak detection, segmentation, and RR-interval features, applying subsampling and subject-wise splits before CNN training; they achieve  $> 98\%$  accuracy, sensitivity, and specificity. Ali *et al.* [10] adopt a data mining workflow on a Kaggle heart disease dataset (biological, non ECG features), evaluating KNN, DT, RF, LR, AdaBoostM1, and MLP; RF attains 100% across accuracy, sensitivity, and specificity. Malakoutiica [211] compares NBG, RF, LR, LDA, and a Dummy Classifier with 10 fold cross-validation; NBG performs best with 96% accuracy and 97% precision.

### 7.3.3 Data Source

Experiments leverage MIT-BIH PhysioNet signals<sup>14</sup>, a reference repository for arrhythmia and normal rhythm:

- **MIT-BIH Arrhythmia (ARR)** [229]: 48 ambulatory 2 channel ECG (360Hz) from 47 subjects (25 men, 22 women; 23–89 years), 30 minutes per recording, with expert rhythm/beat annotations and four pacemaker records.
- **MIT-BIH Normal Sinus Rhythm** [140]: 18 long-term ECG (128Hz) from healthy adults (5 men, 13 women; 20–50 years),  $\sim 24$  hours each, with rhythm/beat annotations.

Beat-wise classification excludes paced beats for consistency.

### 7.3.4 Methods

### 7.3.5 Preprocessing Pipeline

Signals were parsed with `pandas`<sup>15</sup> and curated to retain entries with raw signals and annotations only. For computational tractability, NSR recordings were segmented

<sup>14</sup><https://physionet.org/content/mitdb/1.0.0/>

<sup>15</sup><https://pandas.pydata.org/>

into 1 hour windows. Denoising, peak detection (P, Q, R, S, T), and delineation were performed with NeuroKit2 [210]. Subjects were split 74%/26% into train/test with no overlap. Continuous features were scaled with `RobustScaler`<sup>16</sup> [271].

### 7.3.6 Beat Centered Subsampling

Beat annotations centered at R peaks defined the beat segments. NeuroKit2-assisted segmentation produced pandas DataFrames with labels aligned to R-peak annotations. Regular beats were taken from NSR; abnormal beats from ARR. Non-beat annotations were removed.

### 7.3.7 Feature Engineering and Selection

From segmented beats, we computed 26 candidate features:

- Segment durations (P, PQ, QRS, QR, QT, RS, ST, T, PT; in ms via sample distance/sampling rate);
- Angular features at key junctions ( $\angle P_{\text{on}}PQ$ ,  $\angle R_{\text{off}}QR$ ,  $\angle QRS$ ,  $\angle RST$ ,  $\angle STT_{\text{off}}$ );
- Segment slopes (PQ, QR, RS, ST);
- Peak voltages (P, Q, R in mV);
- Ratios (QR/QS, RS/QS) to quantify QRS substructure.

After removing missing data, we ran recursive feature elimination Recursive Feature Elimination (RFE) with 5 fold `GridSearchCV`<sup>17</sup> across XGB [60], RF [30], and LGBM [160]. The final 13 feature set included: PQ, QRS, RS, ST durations;  $\angle QRS$ ,  $\angle STT_{\text{off}}$ ; PQ, QR, RS, ST slopes; and P, Q, R peak voltages. Class balance was enforced by `RandomUnderSampler` with `imblearn` framework [183], yielding 8259 beats per class NSR/ ARR.

---

<sup>16</sup><https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>

<sup>17</sup>[https://scikitlearn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

### 7.3.8 Classification Models and Tuning

We compared RF, XGB, and XGB implemented in Python via Scikit-learn [254]<sup>18</sup>, XGB<sup>19</sup>, and XGB<sup>20</sup>. Hyperparameters were tuned with 5 fold GridSearchCV to maximize precision, recall, and F1. Search spaces are summarized in Table 7.7.

Table 7.7 Hyperparameter grids for considered classifiers (seed=42).

Algorithm	Hyperparameter	Values
Random Forest	n_estimators	{np.arange(50, 201, 50)}
	max_features	{np.arange(1, x <sub>train</sub> .shape[1])}
	max_depth	[4, 6, 8]
	seed	{42}
XGBoost	n_estimators	{np.arange(50, 201, 50)}
	learning_rate	[0.001, 0.01, 0.1, 1]
	max_depth	[4, 6, 8]
	max_features	{np.arange(1, x <sub>train</sub> .shape[1])}
	seed	{42}
LightGBM	n_estimators	{np.arange(50, 201, 50)}
	learning_rate	[0.001, 0.01, 0.1, 1]
	max_depth	[4, 6, 8]
	max_features	{np.arange(1, x <sub>train</sub> .shape[1])}
	seed	{42}

### 7.3.9 Explainability

We employed SHAP [208] to explain predictions on the independent test set. Rooted in Shapley values from cooperative game theory, SHAP quantifies each feature’s marginal contribution for every instance, enabling both global (feature importance, directionality) and local interpretability.

<sup>18</sup><http://scikit-learn.org>

<sup>19</sup><https://xgboost.readthedocs.io/en/stable/>

<sup>20</sup><https://lightgbm.readthedocs.io/en/stable/>

### 7.3.10 ECG Model Analysis

Table 7.8 reports evaluation metrics for all models; Figure 7.15 shows confusion matrices. The best configuration was LGBM with `learning_rate=1`, `max_depth=4`, `max_features=1`, `n_estimators=200`. Compared to RF and XGB, LGBM achieved the fewest misclassifications across both classes, as shown in Figure 7.15, with overall Accuracy = 0.99. For Normal Sinus, it attained Precision = 1.00, Recall = 0.99, F1 = 0.99; for Arrhythmia, Precision = **0.99**, Recall = **1.00**, F1 = 0.99.

Figure 7.16 displays global feature importance that indicates broadly similar contributions across features, with comparatively marginal roles for  $\angle QRS$ , PQ segment,  $\angle STT_{off}$ , P, and Q peak values. Figures 7.17–7.18 show local explanations that reveal instance-specific drivers: in a negativ NSR case,  $R$  peak, RS and QRS segments, and PQ slope dominate, while in a positive ARR case,  $QRS/RS$  segments and  $ST$  slope contribute more strongly to the decision.

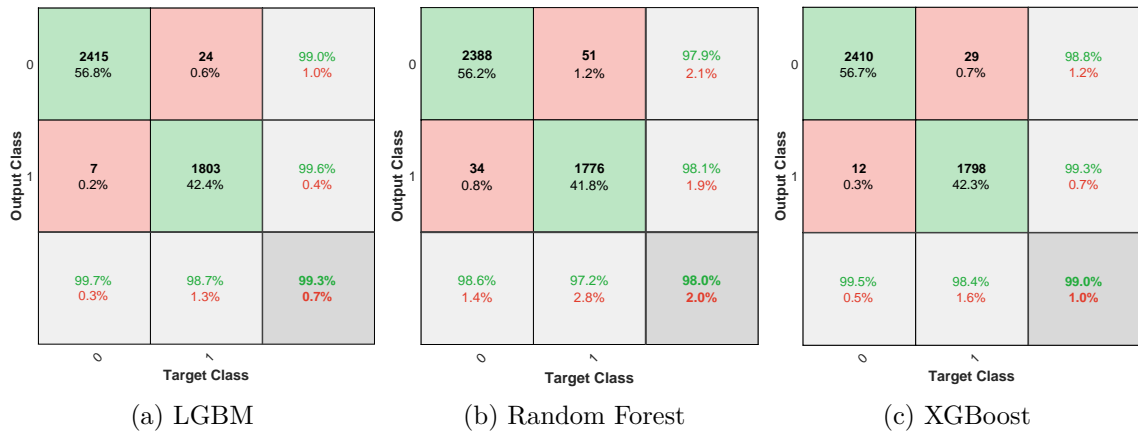


Fig. 7.15 Confusion matrices for the considered models.

Model	Condition	Precision	Recall	F1-score	Accuracy
LGBM	Normal sinus	1.00	0.99	0.99	0.99
	Arrhythmia	<b>0.99</b>	<b>1.00</b>	0.99	
XGBoost	Normal sinus	1.00	0.99	0.99	0.99
	Arrhythmia	0.98	0.99	0.99	
Random Forest	Normal sinus	0.99	0.98	0.98	0.98
	Arrhythmia	0.97	0.98	0.98	

Table 7.8 Evaluation metrics across models. Best overall performance achieved by LGBM.

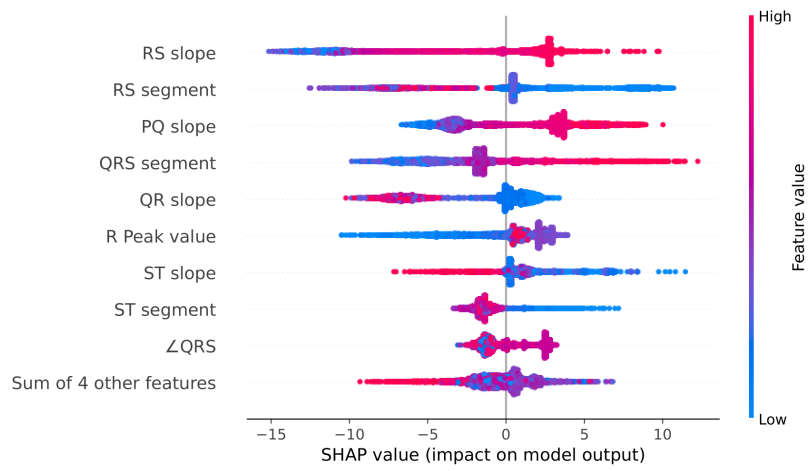


Fig. 7.16 Global SHAP interpretation of ECG features (beeswarm). Horizontal position indicates feature effect toward NSR/ARR; color denotes feature value (red: high, blue: low).

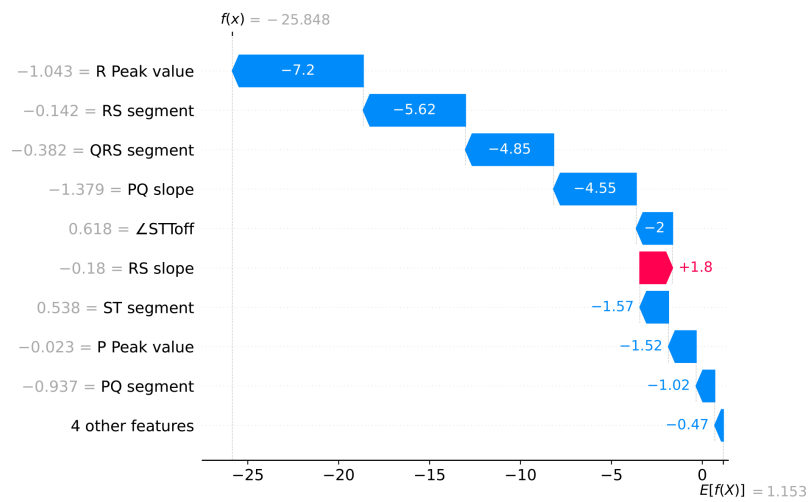


Fig. 7.17 Local SHAP explanation for an NSR (negative) instance.

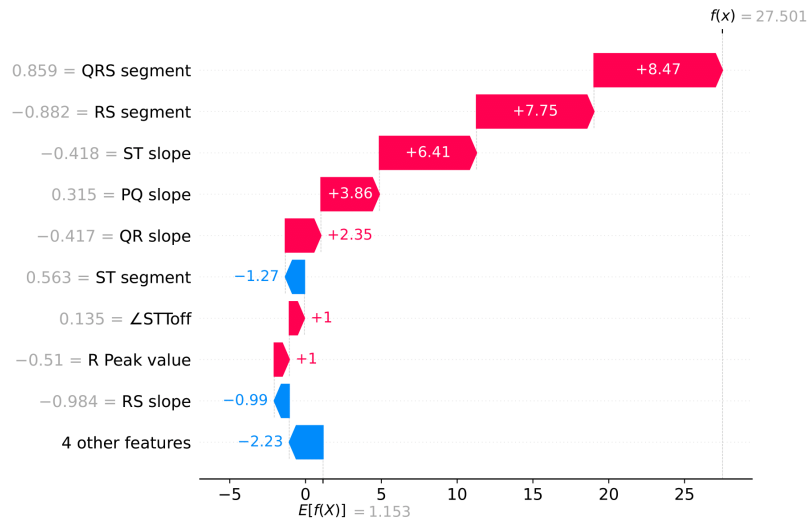


Fig. 7.18 Local SHAP explanation for an ARR (positive) instance.

### 7.3.11 Analysis Outcomes and Clinical Implications

We present an end to end pipeline combining feature extraction/selection with XAI to identify the optimal ML model for real time HF detection. By emphasizing discriminative ECG features and post hoc interpretability, the framework supports both model selection and clinical trust.

**RQ1.** The engineered ECG features successfully trained accurate models for HF screening, indicating their suitability for real time pipelines.

**RQ2.** LGBM achieved the best overall performance (Accuracy = 0.99), with top precision/recall/F1 in both NSR and ARR classes, outperforming RF and marginally improving upon XGB in arrhythmia precision/recall.

**RQ3.** SHAP based XAI yielded clinically meaningful explanations at both global and subject levels, clarifying feature contributions and enhancing diagnostic support.

While the current experiments used a limited benchmark, subsequent research will expand the datasets to encompass a broader range of cardiac conditions and explore the application of DL and anomaly detection methods. The proposed pipeline is designed to streamline epidemiological screening and clinical triage by employing a compact, interpretable feature set derived from raw electrocardiogram data. This approach may reduce delays and resource requirements in subsequent diagnostic testing.

Carefully engineered electrocardiogram features, combined with LGBM and explainable artificial intelligence techniques, achieve state of the art performance in HF and ARR classification. This combination also provides actionable interpretability, facilitating integration into clinical decision making workflows.

## 7.4 Time-to-event interpretable machine learning for multiple sclerosis worsening prediction: results from iDPP@ CLEF 2023

### 7.4.1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease of the Central nervous system (CNS) marked by demyelination, inflammation, and neurodegeneration, yielding heterogeneous symptoms and complicating diagnosis. Advances in computational diagnostics have improved accuracy and efficiency. Among clinical tools, magnetic resonance imaging Magnetic Resonance Imaging (MRI) often enhanced with gadolinium contrast detects active lesions and tracks spatiotemporal lesion dynamics for monitoring and treatment assessment [109, 197]. cerebrospinal fluid (CSF) analysis (oligoclonal bands, IgG, myelin basic protein) supplies evidence of CNS inflammation, while electrophysiological tests (VEP, BAEP, SSEP) reveal demyelination and neural conduction abnormalities [89].

ML and AI increasingly support MS diagnosis, prognosis, and therapy selection by processing clinical and imaging data to identify lesions, quantify disease burden, and forecast progression [226]. These approaches also aid in predicting disease course and treatment response by identifying prognostic patterns and biomarkers from large-scale cohorts, thereby enabling personalized care [262].

Time to event survival ML models predict the time until outcomes such as progression or death [146]. Compared with traditional parametric methods, which assume specific survival distributions, ML survival models better accommodate high dimensional predictors, nonlinearities, interactions, censoring, and time-varying covariates, often achieving superior performance in progression tasks [323].

In this work, using the iDPP@CLEF 2023 MS dataset with  $\sim 1,800$  patients; demographics and clinical variables. We evaluate multiple time to event ML approaches for predicting MS progression. Our objective is an interpretable framework that maximizes task performance while illuminating potential biases and the impact of clinical predictors. The manuscript proceeds as follows: Section 7.4.2 reviews related work; Section 7.4.3 details the methodology; Section 7.4.9 outlines the experimental setup; Section 7.4.10 reports results; Section 7.4.11 concludes and sketches future directions.

## 7.4.2 Related work

Several studies have used ML to predict MS worsening and stratify risk. Zhao *et al.* [347] modelled worsening (increase in EDSS  $\geq 1.5$  within 5 years) using clinical and MRI features. Fiorini *et al.* [110] discriminated MS courses using clinical data. Montolio *et al.* [228] combined Retinal nerve fiber layer (RNFL) thickness with clinical variables for diagnosis and disability course prediction. While many works classify progression states, continuous time to event approaches remain comparatively underexplored for MS course prediction.

## 7.4.3 Methodology

We addressed four challenge tasks:

- **Task 1a** (survival): risk of worsening defined by Expanded Disability Status Scale (EDSS)  $\geq 3$  at least twice within one year.
- **Task 1b** (survival): risk of worsening defined by the first recorded value, per clinical protocol.
- **Task 2a** (probabilistic): cumulative worsening probabilities over windows [0–2], [0–4], [0–6], [0–8], [0–10] years for Task 1a subjects.
- **Task 2b** (probabilistic): same windows for Task 1b subjects.

Figure 7.19 summarizes the pipeline.

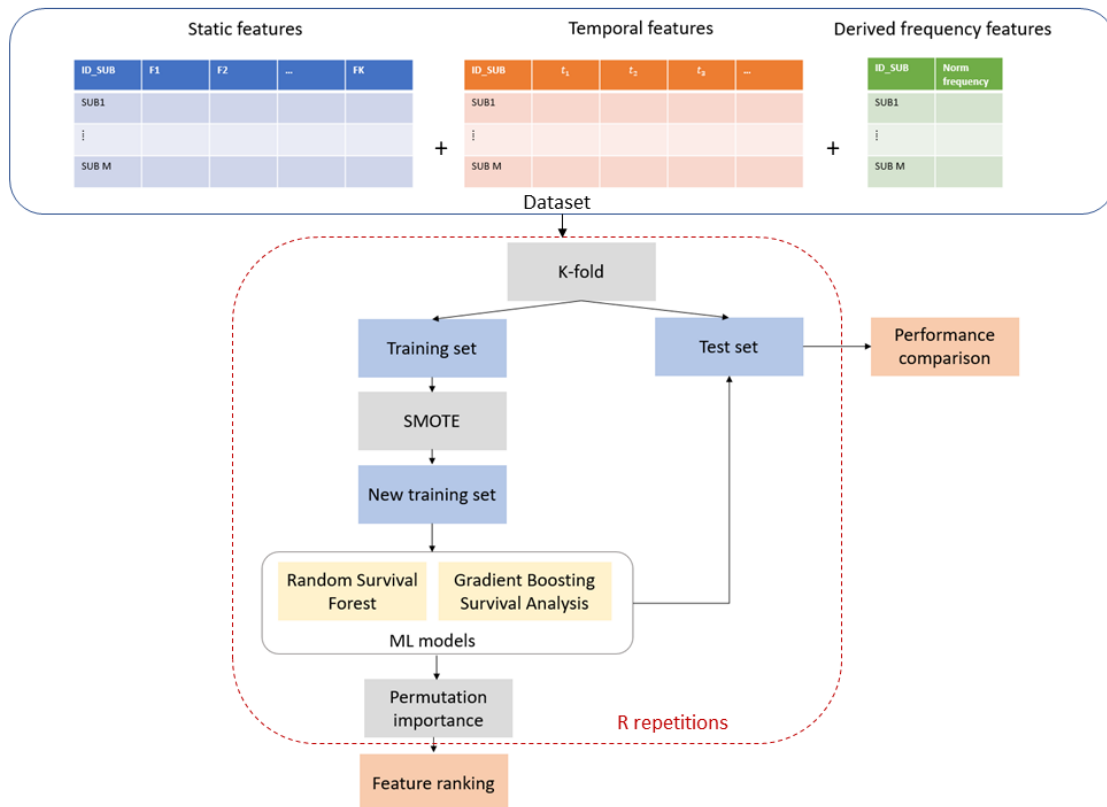


Fig. 7.19 Workflow for Tasks 1a, 1b, 2a, 2b.

#### 7.4.4 Dataset

Challenge data include: (i) static descriptors (age, sex, onset features, centre), (ii) relapse dates, (iii) evoked potentials, (iv) MRI regions/lesions, (v) MS course, (vi) EDSS dates and scores, and (vii) survival outcomes (worsening occurrence and time). See [103, 104] for details.

Figure 7.20 shows the quantified subject overlaps across feature groups, where we retained static features, relapse information, and MRI metadata. MRI lesion features were excluded due to excessive missingness.

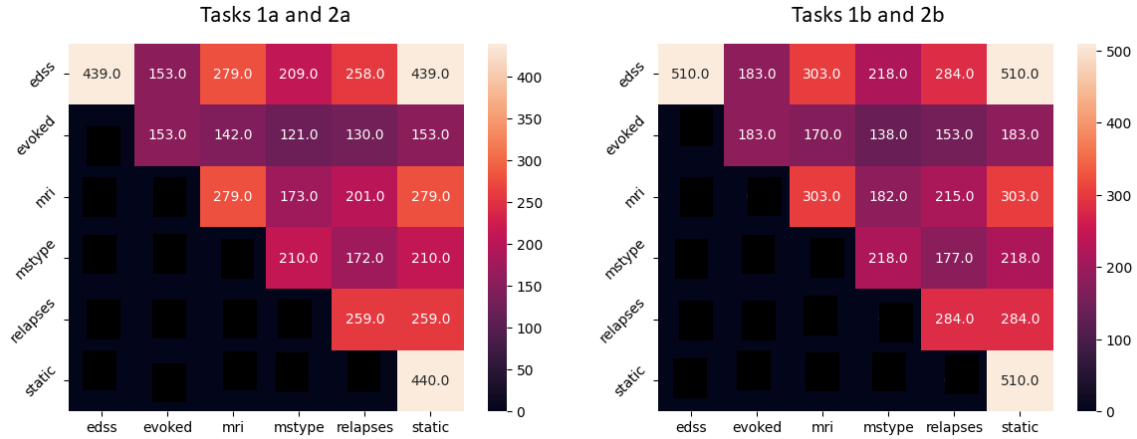


Fig. 7.20 Subject overlap across feature types (all tasks).

We adopted a landmarking strategy [137], sampling the `delta_edss_time0` axis into 16 time points per patient; missing EDSS at a landmark was imputed as 0.

For patient  $i$ , we defined the normalized relapse variation:

$$\text{variation\_relapses}_i = \frac{n\_relapses}{(t_{\text{end}} - t_1)}, \quad (7.6)$$

where  $n\_relapses$  is the count between first ( $t_1$ ) and last ( $t_{\text{end}}$ ) recorded times.

Missing values were imputed with the median. The centre variable was excluded to avoid leakage of site effects.

## 7.4.5 Validation scheme

We used repeated  $k$ -fold cross-validation [172, 321] to improve performance stability and hyperparameter tuning. Within each training fold, we addressed event/censoring imbalance with SMOTE oversampling of the minority class [59].

## 7.4.6 Time to event learning

Each observation comprises features  $X = (x_1, \dots, x_n)$ , event time  $t$  (if observed), censoring time  $c > 0$ , and indicator  $\delta \in \{0, 1\}$ . The observed time is

$$y = \min(t, c) = \begin{cases} t, & \delta = 1, \\ c, & \delta = 0. \end{cases} \quad (7.7)$$

Key functions:

- **Survival**  $S(t) = \Pr[T \geq t]$ ,
- **Hazard**  $h(t) = \lim_{dt \rightarrow 0} \frac{\Pr[t \leq T < t + dt | T \geq t]}{dt}$  with cumulative hazard  $\int_0^t h(u) du$ ,
- **Risk score** (time-discrete) for sample  $x$ :  $r(x) = \sum_{j=1}^J H(t_j, x)$ .

We considered:

- **RSF**: ensembles of survival trees trained on bootstrap samples with random feature selection at splits; terminal nodes estimate survival functions, and ensemble risk is aggregated over trees.
- **Gradient Boosting (GB)** for survival: a sequence of weak learners fit to residuals under a survival loss, combined into a strong predictor via weighted aggregation.

#### 7.4.7 Permutation feature importance

To evaluate feature importance for both ML models, we computed permutation feature importance [15], which involves the following steps:

- **Model Training**: A machine learning model is trained using the dataset.
- **Baseline Performance Calculation**: The model's performance metric is computed using cross-validation. This metric serves as the baseline for subsequent comparisons.
- **Feature Permutation**: The values of the feature of interest are randomly shuffled while all other features remain unchanged. This process disrupts the original relationship between the feature and the target variable.
- **Model Performance Evaluation**: The trained model is applied to the permuted dataset, and the performance metric is recalculated. This metric reflects model performance with the disrupted feature-target relationship.
- **Feature Importance Computation**: Feature importance is quantified by measuring the decrease in model performance resulting from feature permutation. A larger decrease indicates greater importance. Feature importance may be expressed as a difference, ratio, or percentage change between baseline and post-permutation performance.

In this study, we employed this technique to determine which features most significantly influence model performance and to elucidate the relationships between the features and the target variable.

#### 7.4.8 Performance evaluation

The following metrics were utilized to evaluate model performance during the training phase:

- the concordance index (C-index), i.e., a generalization of the area under the ROC curve (AUC) that can take into account censored data. It estimates the probability that the order of the predictions of a pair of comparable patients is consistent with the individual risk scores [317]. It can be computed as:

$$C - index = \frac{\sum_{i,j} 1_{T_j < T_i} \cdot 1_{r_j > r_i} \cdot \delta_j}{\sum_{i,j} 1_{T_j < T_i} \cdot \delta_j}, \quad (7.8)$$

where  $r_i$  is the risk score for patient  $i$ ,  $1_{T_j < T_i} = 1$  if  $T_j < T_i$  else 0;  $1_{r_j > r_i} = 1$  if  $r_j > r_i$  else 0.

- The cumulative dynamic AUC. The receiver operating characteristic (ROC) curve and the area under the ROC curve AUC can be extended to survival data by defining sensitivity (true positive rate) and specificity (true negative rate) as time-dependent measures [144]. Cumulative cases are all individuals who experienced an event prior to or at time  $t$  ( $t_i \leq t$ ), whereas dynamic controls are those with  $t_i > t$ . The associated cumulative dynamic AUC quantifies how well a model distinguishes subjects who fail by a given time ( $t_i \leq t$ ) from those who fail after this time ( $t_i > t$ ).

#### 7.4.9 Experimental setup

The experimental setup was organized to address the following research questions:

- RQ1 What is the impact of the static features on the model performance for the two datasets A and B?
- RQ2 What is the additional contribution of the dynamic features?
- RQ3 Is there a significant ranking of all the features for both definitions of MS worsening?

- RQ4 Does the imbalance of the types of events (censored and non-censored) affect the performance of the models?

Accordingly, 16 models were trained: two RSF models (with and without SMOTE) and two GB models (with and without SMOTE) for each dataset. This was performed for both the partial dataset, which included only static features, and the total dataset, which included both static and dynamic features. All experiments were conducted on Google Colab [34]. The machine learning models were implemented using the Python package scikit-survival-0.20.0 [264]. All submissions were completed in collaboration with the "SisInfLab-AIBio" team.

### 7.4.10 Results

#### RQ1: static features only

Table 7.9 C-index averaged over repeated cross-validation.

Model	Static A	Complete A	Static B	Complete B
RSF	0.49	0.65	0.50	0.53
RSF + SMOTE	0.47	0.58	0.48	0.51
GB	0.53	0.55	0.55	0.53
GB + SMOTE	0.53	0.53	0.56	0.46

For static-only inputs, GB slightly outperforms RSF in both A and B (Figures 7.21, 7.22), but performance remains near chance, indicating static data alone are insufficient.

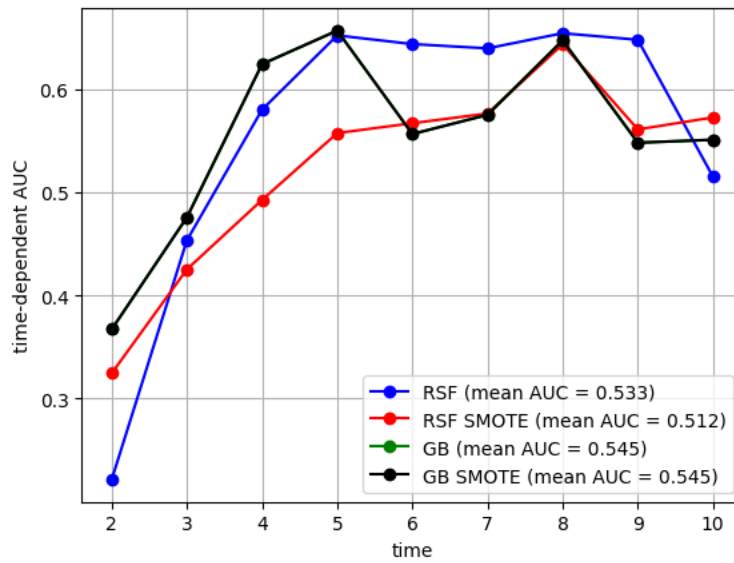


Fig. 7.21 Temporal AUCs (static variables, Dataset A).

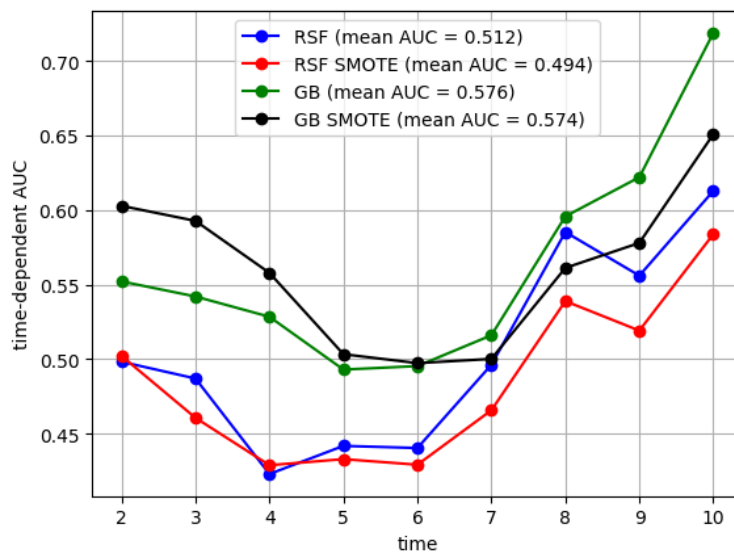


Fig. 7.22 Temporal AUCs (static variables, Dataset B).

### RQ2: adding dynamic features

With dynamic EDSS trajectories and relapse frequency, RSF outperforms GB on Dataset A and clearly improves over static-only models (Figure 7.23, Table 7.9). For Dataset B, only RSF shows a modest gain over GB trained on static features (Figure 7.24).

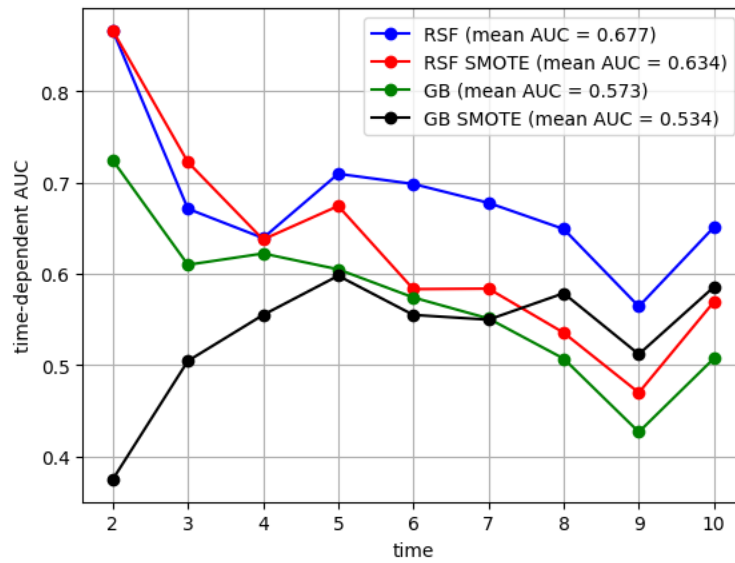


Fig. 7.23 Temporal AUCs (static+dynamic, Dataset A).

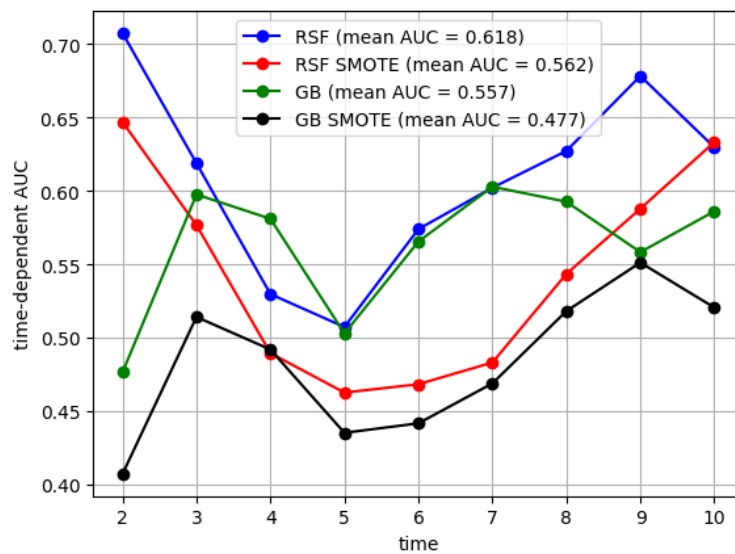


Fig. 7.24 Temporal AUCs (static+dynamic, Dataset B).

### RQ3: feature ranking

For both datasets, the best models were RSF without SMOTE. Permutation importance (Figures 7.25, 7.26) shows, for Dataset A, multiple EDSS landmarks rank highest, alongside disease-onset variables and variation\_relapses. For Dataset B (with a different worsening definition), fewer EDSS landmarks dominate, but disease-onset and variation\_relapses remain salient.

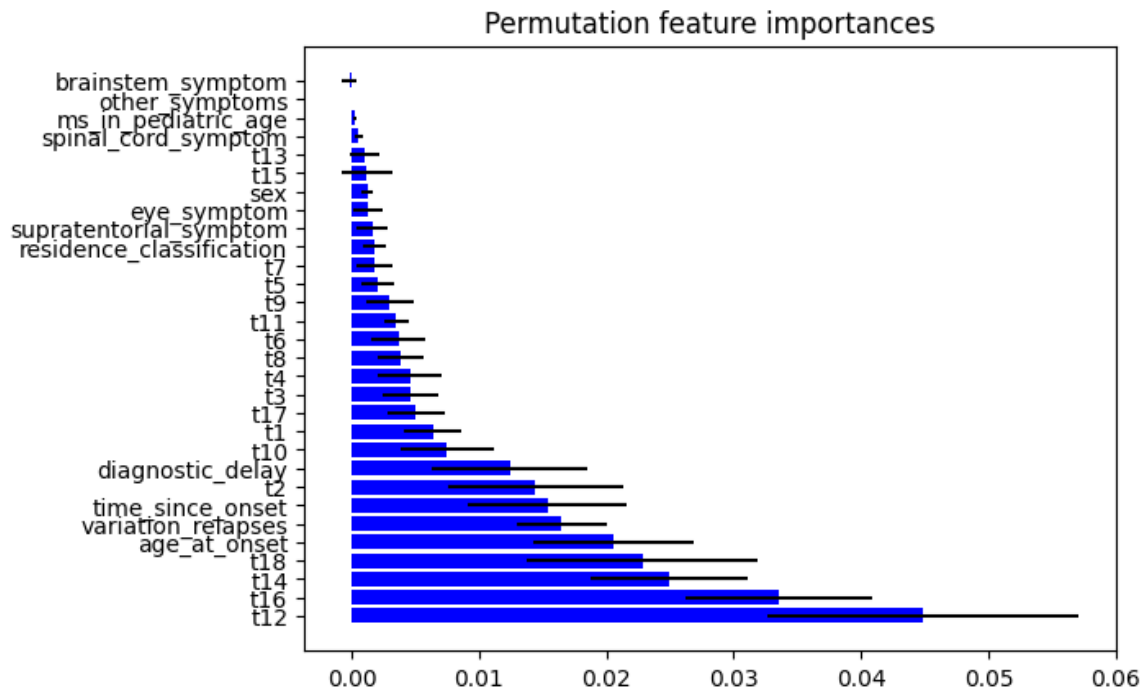


Fig. 7.25 Permutation importances (best RSF, Dataset A).

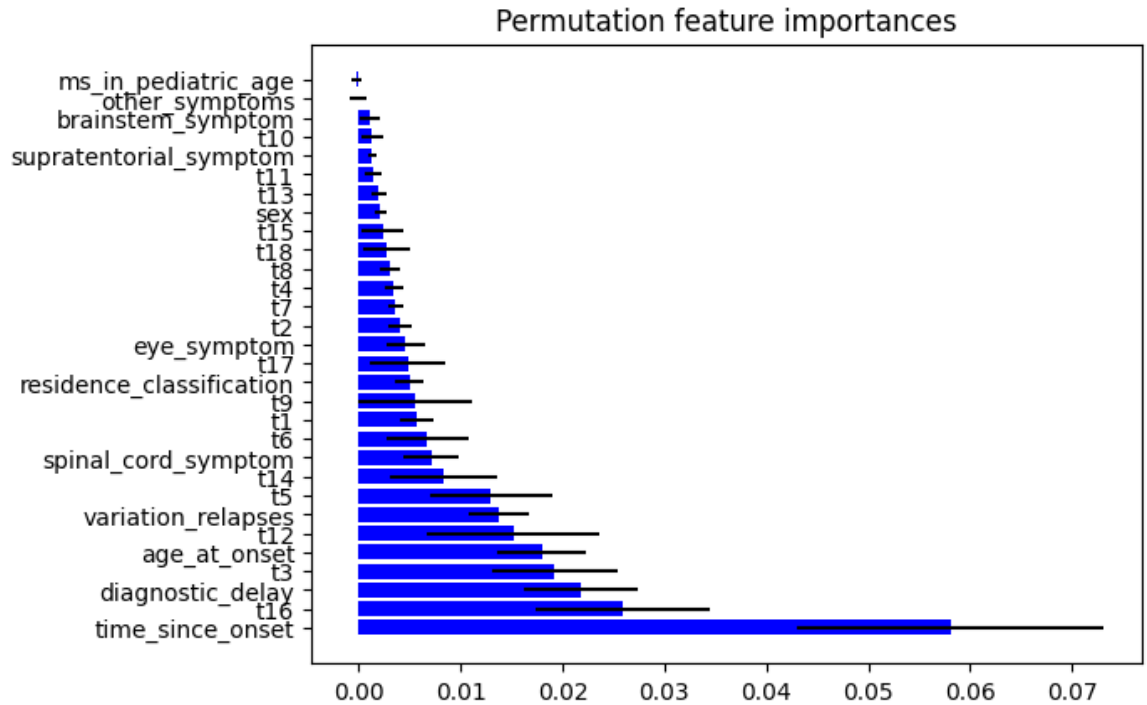


Fig. 7.26 Permutation importances (best RSF, Dataset B).

#### **RQ4: effect of class imbalance**

SMOTE did not improve performance for either learner or dataset (Table 7.9, Figures 7.23, 7.24), suggesting it is ill-suited to censored/not-censored imbalance in these survival settings.

#### **7.4.11 Conclusions and future work**

We proposed a modular, interpretable survival analysis framework for MS worsening on the iDPP@CLEF 2023 data. Dynamic EDSS information and a normalized relapse-frequency feature substantially improve performance; RSF models consistently yielded the best results. Worsening definitions, more complex and condition-dependent (Tasks 1b/2b), led to noticeably lower performance than simpler definitions (Tasks 1a/2a).

We limited our study to two learners with a landmarking approach. Future work will evaluate deep survival models capable of handling variable-length sequences end-to-end. We also excluded the centre variable to avoid site bias; harmonization methods could enable its safe inclusion. More broadly, integrating multi-site harmonization and richer temporal modelling may further improve both discrimination and calibration while preserving interpretability.



# Chapter 8

## Conclusion and Future Works

This thesis presents a comprehensive investigation of EEG decoding methods and systems in the context of BCI. The research is organized around three main lines: emotion recognition using low-cost devices with sparse electrode montages, interactive and creative brain-in-the-loop applications, and the transferability of the proposed methodological framework to clinical and healthcare domains supported by XAI. Across all studies, rigorously designed pipelines—validated through subject-independent protocols such as LOSO and complemented by explainability analyses—demonstrate robust decoding capabilities in settings consistent with lightweight acquisition and real-time integration, supporting both foundational research and application-oriented deployments.

NeuroSense introduces a low-cost dataset together with a validated LOSO pipeline, revealing the particular informativeness of the 0–5s post-stimulus interval. The study further contrasts engineered feature sets with MiniRocket representations and leverages SHAP based interpretation to reveal substantial inter-subject variability and partially emotion-specific explanatory patterns. The Motivational States study compares perception and imagination, showing that perception-based discrimination is consistently more effective, while modest increases in channel density provide only marginal benefits. In parallel, NeuralPMG, Neural Musical Instruments, EmoSynth, and ARIEL demonstrate how the deliberate modulation of mental and emotional states can be operationalized for polyphonic music generation, VST control, sound texture synthesis, and neurolinguistic interaction via BCI and LLM. Finally, Chapter 7 extends the same methodological perspective to healthcare-oriented problems, including MORIX for MAFLD risk modeling, Pareto-optimal model selection for MCI prediction, and XAI-supported ECG classification, thereby illustrating portability and clinical relevance.

From a scientific standpoint, three overarching findings emerge. First, affect-related decoding with low-cost EEG is feasible provided that the processing pipeline is sufficiently rigorous. Using the NeuroSense dataset and a workflow comprising acquisition, MNE-based preprocessing, segmentation into 5s epochs, time-sliding evaluation, and subject-independent classification (MiniRocket and SVM with hyperparameter optimization), the initial 0–5s stimulus window consistently concentrates discriminative information for separating baseline from stimulus across Russell’s four valence/arousal quadrants, suggesting an early temporal dynamic that remains informative even under sparse montages.

Second, inter-subject variability imposes concrete constraints on subject-independent system design. The comparison between engineered spectral/entropy and nonlinear descriptors and MiniRocket representations highlights a systematic trade-off between interpretability and performance: feature-based approaches facilitate neurophysiological interpretation with a modest cost in accuracy, whereas learned representations capture latent temporal dependencies more effectively. In both cases, SHAP attributions exhibit low cross-subject agreement and partially emotion-specific patterns, such as frontal  $\theta/\alpha$  contributions under high arousal and entropy-based or nonlinear indices associated with relaxed or negative states. These results underscore the limitations of fully subject-independent models and motivate personalization and adaptation mechanisms.

Third, perception yields greater separability of motivational states than imagination, while electrode density is not the primary determinant of performance. In the study Machine Learning Classification of Motivational States: Insights from EEG Analysis of Perception and Imagery, twelve motivational states were recorded with medium-density EEG using 14 and 18 channels; perception-based conditions consistently produced more discriminable patterns than imagery-based conditions. By contrast, increasing the montage from 14 to 18 channels resulted in only marginal gains, indicating that task design, stimulus timing, and the temporal organization of the protocol may exert a stronger influence on decoding accuracy than modest increases in channel count under practical acquisition constraints.

Three cross-cutting limitations were also identified. (i) Subject-independent generalization remains sensitive to inter-individual variability, and the analyses confirm that predictive mechanisms differ substantially across participants. (ii) Portions of the evidence are conditioned on the specific datasets and devices considered, motivating broader validation across multiple datasets and acquisition systems, including cross-session and cross-device scenarios. (iii) The interpretability–accuracy trade-off remains open: interpretable feature sets provide neuro-consistent explanations, yet in

some tasks they may be outperformed by convolutional or representation-learning approaches. These observations motivate immediate future directions, including adaptive personalization through lightweight calibration, hybrid models that combine interpretable features with learned temporal representations to balance transparency and performance, multi-dataset evaluations to assess external validity, and controlled clinical studies to quantify effectiveness, acceptability, and user benefit in rehabilitation, emotional support, and educational/creative applications.

Overall, this thesis demonstrates that accessible and explainable BCI systems can (i) decode affective and motivational states using rigorous evaluation protocols, even on lightweight, low-cost hardware; (ii) enable brain-in-the-loop interactive applications, including creative systems and brain-controlled musical instruments; and (iii) transfer ML–XAI best practices to clinically and healthcare-relevant settings through decision-support pipelines and interpretable interfaces. Taken together, the resulting constellation of datasets, preprocessing pipelines, models, and demonstrators advances BCI beyond laboratory prototypes toward more human-centered, interpretable, and deployment-ready neurotechnologies spanning health, education, and the arts.

Finally, the healthcare-oriented contributions substantiate the methodological portability of the proposed approach: the same focus on robust pipelines and XAI-supported interpretation is extended to clinical prediction and decision-support settings, including mortality risk modeling in MAFLD (MORIX), Pareto-optimal model selection for MCI screening, explainable ECG-based classification for heart failure detection, and interpretable time-to-event modeling for multiple sclerosis worsening prediction. Taken together, these studies frame explainability not as an optional add-on, but as a practical prerequisite for integrating machine learning into clinical workflows, where transparency, calibration, and actionable insights are essential.

Future work will address subject variability and external validity. This includes personalization and adaptation mechanisms with light calibration, wider multi-session and cross-device validation to assess generalization under realistic acquisition variability, and hybrid modeling strategies that balance interpretability and accuracy by combining neurophysiological features with learned temporal representations. In parallel, the transition from windowed decoding to continuous monitoring in real scenarios, along with the conduct of controlled user/clinical studies, will be essential for quantifying the effectiveness, acceptability, and implementation readiness of explainable BCI pipelines and healthcare decision support.



# Bibliography

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. <https://www.tensorflow.org/>. Software available from tensorflow.org. 2015.
- [2] Reza Abiri, Soheil Borhani, Eric W. Sellers, Yongmin Jiang, and Xiaopeng Zhao. “A comprehensive review of EEG-based brain–computer interface paradigms.” In: *Journal of Neural Engineering* 16.1 (2019), p. 011001. DOI: 10.1088/1741-2552/aaf12e.
- [3] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. “Automated EEG-based screening of depression using deep learning techniques.” In: *Knowledge-Based Systems* 107 (2016), pp. 1–7.
- [4] E.D. Adrian and B.H.C. Matthews. “The Berger rhythm: Potential changes from the occipital lobes in man.” In: *Brain* 57.4 (1934), pp. 355–385. DOI: 10.1093/brain/57.4.355.
- [5] E.D. Adrian and B.H.C. Matthews. “The interpretation of potential waves in the cortex.” In: *Journal of Physiology* 81.4 (1934), pp. 440–471. DOI: 10.1113/jphysiol.1934.sp003147.
- [6] Aman Aggarwal, Nidhi Chugh, and Rishabh Gupta. “EEG-based brain–computer interfaces: A comprehensive review of machine learning techniques.” In: *ACM Computing Surveys* 55.3 (2022). Match to your cited version, pp. 1–35. DOI: 10.1145/3519021.
- [7] Ali Ahmadvand et al. “Would You Like to Talk About Something Else? On Topic Breadth and Robustness in Conversational Agents.” In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*. Discusses overfitting and topic generalization in CAs. 2020.
- [8] Ahmed S. Albahri, Ameer A. Zaidan, B. B. Zaidan, Osamah S. Albahri, Mohanad M. Salih, et al. “A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion.” In: *Information Fusion* 90 (2023), pp. 316–352. DOI: 10.1016/j.inffus.2023.03.008.
- [9] K. Alho et al. “Brain basis of visual imagery: Recent advances.” In: *Nature Reviews Neuroscience* 24.x (2023). Verify, pp. xx–xx. DOI: 10.1038/s41583-023-xxxx-x.
- [10] Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M Bui, Julian MW Quinn, and Mohammad Ali Moni. “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison.” In: *Computers in Biology and Medicine* 136 (2021), p. 104672.

- [11] John JB Allen, James A Coan, and Maria Nazarian. “Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion.” In: *Biological psychology* 67.1-2 (2004), pp. 183–218.
- [12] John JB Allen, Philipp M Keune, Michael Schönenberg, and Robin Nusslock. *Frontal EEG alpha asymmetry and emotion: From neural underpinnings and methodological considerations to psychopathology and social cognition*. 2018.
- [13] Luz Maria Alonso-Valerdi, Rosa A. Salido-Ruiz, and Ricardo A. Ramírez-Mendoza. “Motor imagery-based brain–computer interfaces: An emerging technology to rehabilitate motor deficits.” In: *Neuropsychologia* 79 (2015), pp. 354–363. DOI: 10.1016/j.neuropsychologia.2015.02.001.
- [14] H. Altaheri et al. “Deep Learning for EEG-Based Motor Imagery: A Review.” In: *Sensors* 23.— (2023). Confirm volume/issue/pages.
- [15] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. “Permutation importance: a corrected feature importance measure.” In: *Bioinformatics* 26.10 (2010), pp. 1340–1347.
- [16] American Clinical Neurophysiology Society. *Guideline 1: Minimum Technical Requirements for Performing Clinical Electroencephalography*. ACNS Guidelines. Accessed YYYY-MM-DD. 2016.
- [17] American Clinical Neurophysiology Society. *Guideline 3: A Proposal for Standard Montages to Be Used in Clinical EEG*. ACNS Guidelines. Accessed YYYY-MM-DD. 2016.
- [18] American Clinical Neurophysiology Society. *Guideline 4: Recording Clinical EEG on Digital Media*. ACNS Guidelines. Accessed YYYY-MM-DD. 2016.
- [19] Carmelo Ardito et al. “Biofeedback Interfaces Driven by EEG: Design and Evaluation.” In: *Proc. of \**. Update venue/DOI. 2021.
- [20] Carmelo Ardito et al. “Brain–Computer Interface Applications in Rehabilitation: A Review.” In: *Proc. of \**. Please complete. 2021.
- [21] Carmelo Ardito, Tommaso Colafiglio, Domenico Lofù, Elena Longo, Fedelucio Narducci, and Eugenio Di Sciascio. “Towards a Neurofeedback Tool for Emotion Recognition Using Brain–Computer Interface.” In: *Web Engineering Workshops (ICWE 2024 Workshops, WALIS 2024)*. Vol. 2188. Communications in Computer and Information Science. Cham: Springer, 2025, pp. 292–299. DOI: 10.1007/978-3-031-75110-3\_8.
- [22] Carmelo Ardito, Tommaso Colafiglio, Paolo Sorino, Domenico Lofù, Tommaso Di Noia, and Eugenio Di Sciascio. “Brain–Computer Interfaces and AI: Opportunities for Interactive Systems.” In: — (2022). Please update venue details if available.
- [23] John Atkinson and Daniel Campos. “Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers.” In: *Expert Systems with Applications* 47 (2016), pp. 35–41.
- [24] N. Author et al. “On the Signal Quality of the MUSE EEG Headband.” In: *Proc. IEEE EMBC*. Placeholder key used in text; please replace with the exact Muse reference. 2021.

- [25] Nicolas Barascud et al. *meegkit: MEG/EEG Preprocessing Toolbox in Python*. GitHub repository. <https://github.com/nbara/python-meegkit>. 2021.
- [26] William G Baxt. “Application of artificial neural networks to clinical medicine.” In: *The lancet* 346.8983 (1995), pp. 1135–1138.
- [27] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. “Greedy layer-wise training of deep networks.” In: *NIPS*. 2007.
- [28] Anna M. Beres. “Time is of the Essence: A Review of ERP Components Sensitive to Time and Timing in Cognitive Neuroscience.” In: *Frontiers in Human Neuroscience* 11 (2017), p. 303. DOI: 10.3389/fnhum.2017.00303.
- [29] Stefano Biancofiore and Colleagues. “Aspect-Based Sentiment Analysis: A Review of Methods and Applications.” In: *Information Processing & Management* (2022). Overview of aspect extraction and sentiment classification.
- [30] Gérard Biau and Erwan Scornet. “A random forest guided tour.” In: *Test* 25 (2016), pp. 197–227.
- [31] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyle M. Su, and Kay A. Robbins. “The PREP Pipeline: Standardized Preprocessing for Large-Scale EEG Analysis.” In: *Frontiers in Neuroinformatics* 9 (2015), p. 16. DOI: 10.3389/fninf.2015.00016.
- [32] Jordan J. Bird et al. “EEG-Based Mental State Recognition: Dataset and Baselines.” In: *IEEE Access* (2019). Key matches text; please confirm exact title/DOI.
- [33] Simon Bird. *EEG Brainwave Dataset: Mental State*. Dataset. Available online; please add persistent identifier if available. 2018.
- [34] Ekaba Bisong and Ekaba Bisong. “Google colab.” In: *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners* (2019), pp. 59–64.
- [35] Benjamin Blankertz, Klaus-Robert Müller, et al. “The BCI Competition 2003/2005 — SVM-based single-trial EEG classification.” In: *Lecture Notes in Computer Science*. Vol. 3696. Check title/pages to match this chapter. 2005, pp. xx–xx. DOI: 10.1007/11550822\_8.
- [36] Pietro Boccadoro et al. “Watermarking and AI Methods for Information Security.” In: — (2022). Please update with exact venue/DOI.
- [37] V. Borisov et al. “Mobile System for Assessing Functional and Mental States Using EEG and ECG.” In: *Biomed Eng* (2017). Update volume/issue/pages.
- [38] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. “COMET: Commonsense Transformers for Automatic Knowledge Graph Construction.” In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, pp. 4762–4779.
- [39] Margaret M. Bradley and Peter J. Lang. “Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential.” In: *Journal of Behavior Therapy and Experimental Psychiatry* 25.1 (1994), pp. 49–59. DOI: 10.1016/0005-7916(94)90063-9.

- [40] Brankica Bratić, Vladimir Kurbalija, Mirjana Ivanović, Iztok Oder, and Zoran Bosnić. “Machine learning for predicting cognitive diseases: methods, data sources and risk factors.” In: *Journal of medical systems* 42 (2018), pp. 1–15.
- [41] Jeffrey W. Britton, Lorraine C. Frey, Jennifer L. Hopp, Erik K. St. Louis, et al. “The Normal EEG.” In: *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. Ed. by Erik K. St. Louis and Lorraine C. Frey. Chicago: American Epilepsy Society, 2016.
- [42] Joseph D Bronzino and Donald R Peterson. “Principles of electroencephalography.” In: *Biomedical Engineering Fundamentals*. CRC press, 2006, pp. 445–456.
- [43] S. Cai et al. “Motor imagery BCI for rehabilitation: A review of clinical studies.” In: *Journal of Neural Engineering* 19.x (2022). Verify, p. xxx. DOI: 10.1088/1741-2552/xxxxxx.
- [44] Toon Calders and Szymon Jaroszewicz. “Efficient AUC optimization for classification.” In: *European conference on principles of data mining and knowledge discovery*. Springer. 2007, pp. 42–53.
- [45] Francesco Calivà, Nikan K Namiri, Maureen Dubreuil, Valentina Pedoia, Eugene Ozhinsky, and Sharmila Majumdar. “Studying osteoarthritis with artificial intelligence applied to magnetic resonance imaging.” In: *Nature Reviews Rheumatology* 18.2 (2022), pp. 112–121.
- [46] K Cantrell, MM Erenas, I de Orbe-Payá, and LF Capitán-Vallvey. “Use of the hue parameter of the hue, saturation, value color space as a quantitative analytical parameter for bitonal optical sensors.” In: *Analytical chemistry* 82.2 (2010), pp. 531–542.
- [47] Kelsey A. Carbine, Michael V. Larson, Chad D. Stevens, and T. L. et al. “The utility of EEG/ERP components in food-craving research: A systematic review and meta-analysis.” In: *Psychophysiology* 55.10 (2018). Verify full author list, e13046. DOI: 10.1111/psyp.13046.
- [48] Gabriella Casalino, Giovanna Castellano, Arianna Consiglio, Nicoletta Nuzziello, and Gennaro Vessio. “MicroRNA expression classification for pediatric multiple sclerosis identification.” In: *Journal of Ambient Intelligence and Humanized Computing* (2021), pp. 1–10.
- [49] Gabriella Casalino, Giovanna Castellano, and Gianluca Zaza. “Evaluating the robustness of a contact-less mHealth solution for personal and remote monitoring of blood oxygen saturation.” In: *Journal of Ambient Intelligence and Humanized Computing* (2022), pp. 1–10.
- [50] Fabio Castellana, Simona Aresta, Paolo Sorino, Ilaria Bortone, Domenico Lofù, Fedelucio Narducci, Tommaso Di Noia, Eugenio Di Sciascio, and Rodolfo Sardone. “An Artificial Neural Network Model to Assess Nutritional Factors Associated with Frailty in the Aging Population from Southern Italy.” In: *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2022, pp. 3228–3233.

- [51] Fabio Castellana, Luisa Lampignano, Ilaria Bortone, Roberta Zupo, Madia Lozupone, Chiara Griseta, Antonio Daniele, Giovanni De Pergola, Gianluigi Giannelli, Rodolfo Sardone, et al. “Physical frailty, multimorbidity, and all-cause mortality in an older population from Southern Italy: results from the Salus in Apulia Study.” In: *Journal of the American Medical Directors Association* 22.3 (2021), pp. 598–605.
- [52] M. Castellana et al. “Artificial Intelligence in Healthcare: Trends and Challenges.” In: *Journal of Medical Systems* (2022). Please update details.
- [53] Hubert Cecotti and Axel Grützmaier (if any). “Convolutional Neural Networks for P300 detection with application to BCI.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19.3 (2011). Verify authors/pages, pp. 341–352. DOI: 10.1109/TNSRE.2011.2108667.
- [54] Chia-Hung Chang et al. “EEG frequency analysis for brain–computer interface: a review.” In: *Sensors* 20.10 (2020). Verify, p. xxxx. DOI: 10.3390/s201xxxx.
- [55] Bnar Charbuty and Adnan Abdulazeez. “Classification Based on Decision Tree Algorithm for Machine Learning.” In: *Journal of Applied Science & Technology Trends* 2.1 (2021). Check and confirm venue/page details if needed, pp. 20–28.
- [56] R. Chatterjee et al. “A Novel MI-BCI Classification Framework Using EEG Features.” In: *Proc. IEEE EMBS*. Update pages/DOI. 2019.
- [57] R. Chatterjee et al. “Cellular Automata-based Classification of Motor Imagery from EEG.” In: *IEEE Access* (2020). Update volume/issue/pages.
- [58] I. Chattopadhyay et al. “Motivation-related EEG decoding with machine learning.” In: *IEEE Access* 9 (2021). Verify, pp. xxxx–xxxx. DOI: 10.1109/ACCESS.xxxxxxx.
- [59] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. “SMOTE: synthetic minority over-sampling technique.” In: *Journal of artificial intelligence research* 16 (2002), pp. 321–357.
- [60] Tianqi Chen and Carlos Guestrin. “Xgboost: A scalable tree boosting system.” In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.
- [61] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. “Xgboost: extreme gradient boosting.” In: *R package version 0.4-2* 1.4 (2015), pp. 1–4.
- [62] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. “Learning phrase representations using RNN encoder–decoder for statistical machine translation.” In: *EMNLP*. 2014, pp. 1724–1734.
- [63] François Chollet et al. *Keras*. <https://keras.io>. Accessed: 2025-09-05. 2015.
- [64] Bastien Chopard and Michel Droz. *Cellular Automata Modeling of Physical Systems*. Cambridge University Press, 1998.

- [65] P. Cipresso, L. Carelli, F. Solca, D. Meazzi, P. Meriggi, B. Poletti, G. Riva, N. Ticozzi, and V. Silani. “The use of P300-based brain–computer interface in amyotrophic lateral sclerosis: From augmentative and alternative communication to cognitive assessment.” In: *Brain and Cognition* 79.1 (2012), pp. 79–83. DOI: 10.1016/j.bandc.2012.02.005.
- [66] Tommaso Colafoglio et al. “Exploring Biofeedback with EEG and AI.” In: — (2023). Please update venue/DOI.
- [67] Tommaso Colafoglio, Carmelo Ardito, Domenico Lofù, Paolo Sorino, Fedelucio Narducci, Tommaso Di Noia, and Eugenio Di Sciascio. “Brain–Computer Interface: Deep Learning Approach to Predict Human Emotion Recognition.” In: *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2022. DOI: 10.1109/SMC53654.2022.9945554.
- [68] Tommaso Colafoglio, Carmelo Ardito, Paolo Sorino, Domenico Lofù, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms.” In: *Cognitive Computation* 16.5 (2024), pp. 2779–2802. DOI: 10.1007/s12559-024-10280-6.
- [69] Tommaso Colafoglio and coauthors. “Predicting Human Emotions using EEG-based Brain–Computer Interface and Interpretable Machine Learning.” In: *Ital-IA 2023 (CINI National Conference on AI)*. Extended abstract/paper; add pages/DOI if available. Italy, 2023.
- [70] Tommaso Colafoglio and coauthors. “EmoSynth: Real-Time Emotion-Driven Sound Texture Synthesis via Brain–Computer Interface.” In: *UMAP 2024 WeBIUM Workshop (ACM Companion)* (2024). Add DOI or proceedings details.
- [71] Tommaso Colafoglio, Domenico Lofù, Paolo Sorino, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “Machine learning classification of motivational states: Insights from EEG analysis of perception and imagery.” In: *Expert Systems with Applications* 275 (2025), p. 127076. DOI: 10.1016/j.eswa.2025.127076.
- [72] Tommaso Colafoglio, Paolo Sorino, Domenico Lofù, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “EmoSynth: Real-Time Emotion-Driven Sound Texture Synthesis via Brain–Computer Interface.” In: *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’24 Adjunct), WeBIUM Workshop*. New York, NY, USA: ACM, 2024. DOI: 10.1145/3631700.3665196.
- [73] Tommaso Colafoglio, Paolo Sorino, Domenico Lofù, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. “Neural Musical Instruments through Brain–Computer Interface and Biofeedback.” In: *Proceedings of the 2nd Workshop on Wearable Devices and Brain–Computer Interfaces for User Modelling (WeBIUM ’25), co-located with UMAP 2025*. New York, NY, USA: ACM, 2025. DOI: 10.1145/3708319.3733644.

- [74] Tommaso Colafiglio, Paolo Sorino, Domenico Lofù, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. “Combining Mental States Recognition and Machine Learning for Neurorehabilitation.” In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2023, pp. 3848–3853. DOI: 10.1109/SMC53992.2023.10394347.
- [75] Tommaso Colafiglio, Paolo Sorino, Angela Lombardi, Domenico Lofù, and Tommaso Di Noia. “Predicting Human Emotions using EEG-based Brain–Computer Interface and Interpretable Machine Learning.” In: *Proceedings of Ital-IA 2023: 3rd National CINI Conference on Artificial Intelligence*. Vol. 4017. CEUR Workshop Proceedings. CEUR-WS.org, 2023, pp. 200–205.
- [76] Alexandre Cormier et al. *MuseLSL: Stream Muse EEG with LabStreamingLayer*. <https://github.com/alexandre-oll/muse-lsl>. Accessed: YYYY-MM-DD. 2017.
- [77] Raffaele Cozzolongo, Alberto Ruben Osella, Silvana Elba, Josè Petruzzi, Giampiero Buongiorno, Vito Giannuzzi, Giuseppe Leone, Caterina Bonfiglio, Elsa Lanzilotta, Onofrio Giuseppe Manghisi, et al. “Epidemiology of HCV infection in the general population: a survey in a southern Italian town.” In: *Official journal of the American College of Gastroenterology ACG* 104.11 (2009), pp. 2740–2746.
- [78] Milena Čukić, Miodrag Stokić, Slavoljub Radenković, Miloš Ljubisavljević, Slobodan Simić, and Danka Savić. “Nonlinear analysis of EEG complexity in episode and remission phase of recurrent depression.” In: *International journal of methods in psychiatric research* 29.2 (2020), e1816.
- [79] Ritanna Curci, Antonella Bianco, Isabella Franco, Angelo Campanella, Antonella Mirizzi, Caterina Bonfiglio, Paolo Sorino, Fabio Fucilli, Giuseppe Di Giovanni, Nicola Giampaolo, et al. “The Effect of Low Glycemic Index Mediterranean Diet and Combined Exercise Program on Metabolic-Associated Fatty Liver Disease: A Joint Modeling Approach.” In: *Journal of Clinical Medicine* 11.15 (2022), p. 4339.
- [80] Muhammad Najam Dar, Muhammad Usman Akram, Sajid Gul Khawaja, and Amit N Pujari. “CNN and LSTM-based emotion charting using physiological signals.” In: *Sensors* 20.16 (2020), p. 4551.
- [81] Richard J Davidson. “The functional neuroanatomy of affective style.” In: *Cognitive neuroscience of emotion* (2000), pp. 371–388.
- [82] Arka De, Naveen Bhagat, Manu Mehta, Sunil Taneja, and Ajay Duseja. “Metabolic dysfunction-associated steatotic liver disease (MASLD) definition is better than MAFLD criteria for lean patients with NAFLD.” In: *Journal of hepatology* 80.2 (2024), e61–e62.
- [83] Julian De Freitas, Ahmet Kaan Uğuralp, Zeliha Oğuz-Uğuralp, and Stefano Puntoni. “Chatbots and mental health: insights into the safety of generative AI.” In: *Journal of Consumer Psychology* (2022).
- [84] Marie Decraecker, Dan Dutartre, Jean-Baptiste Hiriart, Marie Irles-Depé, Faiza Chermak, Juliette Foucher, and Victor de Lédinghen. “Long-term prognosis of patients with metabolic (dysfunction)-associated fatty liver disease by non-invasive methods.” In: *Alimentary Pharmacology & Therapeutics* 55.5 (2022), pp. 580–592.

- [85] Arnaud Delorme and Scott Makeig. “EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis.” In: *Journal of Neuroscience Methods* 134.1 (2004), pp. 9–21. DOI: 10.1016/j.jneumeth.2003.10.009.
- [86] Arnaud Delorme, Scott Makeig, and Terrence Sejnowski. “Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis.” In: *NeuroImage* 34.4 (2007), pp. 1443–1449.
- [87] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. “MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification.” In: *arXiv preprint arXiv:2012.08791* (2021).
- [88] Liang Deng et al. “A Survey of Sentiment Analysis in the Era of Deep Learning and Pretrained Language Models.” In: *IEEE Transactions on Affective Computing* (2021). Covers limitations of linear sentiment assumptions.
- [89] Giovanni Di Maggio, Roberto Santangelo, Simone Guerrieri, Mariangela Bianco, Laura Ferrari, Stefania Medaglini, Mariaemma Rodegher, Bruno Colombo, Lucia Moiola, Raffaella Chieffo, et al. “Optical coherence tomography and visual evoked potentials: which is more sensitive in multiple sclerosis?” In: *Multiple Sclerosis Journal* 20.10 (2014), pp. 1342–1347.
- [90] Sander Dieleman and other. “The Challenge of Realistic Music Modeling.” In: *NeurIPS Workshop on Machine Learning for Creativity*. 2018.
- [91] Nadine Dijkstra et al. “Subjective vividness and neural similarity between imagery and perception.” In: *NeuroImage* 274 (2023). Confirm volume/page, p. 120133. DOI: 10.1016/j.neuroimage.2023.120133.
- [92] Nadine Dijkstra et al. “Nuancing the overlap between imagery and perception.” In: *Trends in Cognitive Sciences* 28.x (2024). Verify, pp. xx–xx. DOI: 10.1016/j.tics.2024.xxxxxx.
- [93] Nadine Dijkstra, Sara E. H. van Gerven, and Laura D. Bosch? “Vividness of visual imagery depends on the neural overlap with perception in visual areas.” In: *Journal of Neuroscience* 37.5 (2017). Verify author list, pp. 1367–1373. DOI: 10.1523/JNEUROSCI.3022-16.2017.
- [94] Owen M. Doyle, R. Küblbeck, and R. I. “Single-trial affective state classification with SVM from physiological signals.” In: *IEEE EMBC*. Verify metadata. 2006, pp. xxxx–xxxx.
- [95] Karolina Drożdż, Katarzyna Nabrdalik, Hanna Kwiendacz, Mirela Hendel, Anna Olejarz, Andrzej Tomasik, Wojciech Bartman, Jakub Nalepa, Janusz Gumprecht, and Gregory YH Lip. “Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach.” In: *Cardiovascular Diabetology* 21.1 (2022), p. 240.
- [96] Y. Du et al. “Individuals’ differences in ERP-based BCI: Deep learning and transfer methods.” In: *IEEE Access* 10 (2022). Verify, pp. xxxx–xxxx. DOI: 10.1109/ACCESS.xxxxxx.
- [97] Jon Duckett. *HTML & CSS: design and build websites*. Vol. 15. Wiley Indianapolis, IN, USA: 2011.

- [98] Ritesh Dwivedi, Dimpal Dave, Hardik Naik, Shashank Singhal, Ritesh Omer, Pratik Patel, et al. “Explainable AI (XAI): Core Ideas, Techniques, and Solutions.” In: *ACM Computing Surveys* 55 (2023). DOI: 10.1145/3561048.
- [99] Amine Ectiouei et al. “Classification of motor imagery from EEG: Recent methods and challenges.” In: *Biomedical Signal Processing and Control* 89 (2024). Confirm, pp. 105–118. DOI: 10.1016/j.bspc.2024.xxxxxx.
- [100] Bradley Efron and Robert Tibshirani. “The bootstrap method for assessing statistical accuracy.” In: *Behaviormetrika* 12.17 (1985), pp. 1–35.
- [101] Mohammed Eslam, Philip N Newsome, Shiv K Sarin, Quentin M Anstee, Giovanni Targher, Manuel Romero-Gomez, Shira Zelber-Sagi, Vincent Wai-Sun Wong, Jean-François Dufour, Jörn M Schattenberg, et al. “A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement.” In: *Journal of hepatology* 73.1 (2020), pp. 202–209.
- [102] Mohammed Eslam, Arun J Sanyal, Jacob George, Arun Sanyal, Brent Neuschwander-Tetri, Claudio Tiribelli, David E Kleiner, Elizabeth Brunt, Elisabetta Bugianesi, Hannele Yki-Järvinen, et al. “MAFLD: a consensus-driven proposed nomenclature for metabolic associated fatty liver disease.” In: *Gastroenterology* 158.7 (2020), pp. 1999–2014.
- [103] G. Faggioli et al. “Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2023.” In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023)*. Ed. by A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, A. Aliannejadi, M. Vlachos, G. Faggioli, and N. Ferro. Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, 2023.
- [104] G. Faggioli et al. “Overview of iDPP@CLEF 2023: The Intelligent Disease Progression Prediction Challenge.” In: *CLEF 2023 Working Notes*. Ed. by M. Aliannejadi, G. Faggioli, N. Ferro, and M. Vlachos. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073., 2023.
- [105] J. Fan et al. “Recognizing five emotional states with k-NN on EEG.” In: *Biomedical Signal Processing and Control* xx.xx (2017). Complete details, pp. xx–xx. DOI: 10.xxxx/xxxxx.
- [106] Junliang Fan, Xin Ma, Lifeng Wu, Fucang Zhang, Xiang Yu, and Wenzhi Zeng. “Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data.” In: *Agricultural water management* 225 (2019), p. 105758.
- [107] Lawrence A. Farwell and Emanuel Donchin. “Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials.” In: *Electroencephalography and Clinical Neurophysiology* 70.6 (1988), pp. 510–523. DOI: 10.1016/0013-4694(88)90149-6.
- [108] Yousef Fazel, Aaron B Koenig, Mehmet Sayiner, Zachary D Goodman, and Zobair M Younossi. “Epidemiology and natural history of non-alcoholic fatty liver disease.” In: *Metabolism* 65.8 (2016), pp. 1017–1025.

- [109] Massimo Filippi, Maria A Rocca, Olga Ciccarelli, Nicola De Stefano, Nikos Evangelou, Ludwig Kappos, Alex Rovira, Jaume Sastre-Garriga, Mar Tintorè, Jette L Frederiksen, et al. “MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines.” In: *The Lancet Neurology* 15.3 (2016), pp. 292–303.
- [110] Samuele Fiorini, Alessandro Verri, Andrea Tacchino, Michela Ponzio, Giampaolo Bricchetto, and Annalisa Barla. “A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes.” In: *2015 37th Annual International Conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2015, pp. 4443–4446.
- [111] Raffaella Folgieri and Matteo Zichella. “A BCI-based application in music: Conscious playing of single notes by brainwaves.” In: *Computers in Entertainment* 10.1 (2012), pp. 1–10.
- [112] Simon Fortin et al. “Affective Biofeedback in Neurorehabilitation: Evidence and Design.” In: *IEEE Transactions on Affective Computing* (2019). Please complete.
- [113] Julian Fulford et al. “The neural correlates of visual imagery vividness.” In: *Cortex* 105 (2018), pp. 135–144. DOI: 10.1016/j.cortex.2018.04.013.
- [114] Thippa Reddy Gadekallu, Neelu Khare, Sweta Bhattacharya, Saurabh Singh, Praveen Kumar Reddy Maddikunta, and Gautam Srivastava. “Deep neural networks to predict diabetic retinopathy.” In: *Journal of Ambient Intelligence and Humanized Computing* (2020), pp. 1–14.
- [115] Filipe Galvão, Soraia M Alarcão, and Manuel J Fonseca. “Predicting Exact Valence and Arousal Values from EEG.” In: *Sensors* 21.10 (2021), p. 3414.
- [116] S. Gao, Y. Wang, X. Gao, and B. Hong. “Visual and auditory brain–computer interfaces.” In: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), pp. 1436–1447. DOI: 10.1109/TBME.2014.2300164.
- [117] X. Gao. “Linear Discriminant Analysis: Overview and recent advances.” In: — (2005). Placeholder entry; update with the exact source you used.
- [118] Felix Gembler, Patrick Stawicki, and Ivan Volosyak. “How to report BCI performance?” In: *Graz Brain-Computer Interface Conference*. Check proceedings details/DOI. 2017.
- [119] Jesus J Gomar, Maria T Bobes-Bascaran, Concepcion Conejero-Goldberg, Peter Davies, Terry E Goldberg, Alzheimer’s Disease Neuroimaging Initiative, et al. “Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer’s disease neuroimaging initiative.” In: *Archives of general psychiatry* 68.9 (2011), pp. 961–969.
- [120] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets.” In: *NIPS*. 2014, pp. 2672–2680.
- [121] S. Gordleeva et al. “Real-Time EEG-Based Brain–Computer Interfaces: A Review.” In: — (2020). Please update venue/DOI.

- [122] Bahman Gorjian, Abdolmajid Hayati, and Parisa Pourkhoni. “Using Praat software in teaching prosodic features to EFL learners.” In: *Procedia-Social and Behavioral Sciences* 84 (2013), pp. 34–40.
- [123] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S. Hämäläinen. “MEG and EEG data analysis with MNE-Python.” In: *Frontiers in Neuroscience* 7 (2013), p. 267. DOI: 10.3389/fnins.2013.00267.
- [124] Antonio Gulli and Sujit Pal. *Deep Learning with Keras*. Packt Publishing, 2017.
- [125] Naresh Kumar Gunda, Mohammed I Khalaf, Shaleen Bhatnagar, Aadam Quraishi, Leeladhar Gudala, Ashok Kumar Pamidi Venkata, Faisal Yousef Alghayadh, Shtwai Alsubai, and Vaibhav Bhatnagar. “Lightweight attention mechanisms for EEG emotion recognition for brain computer interface.” In: *Journal of Neuroscience Methods* 410 (2024), p. 110223.
- [126] V. K. Gunjan et al. “Prediction and Decision Support in Agriculture Using AI.” In: *Proc. Intl. Conf. on AI in Agriculture*. Please update venue and pages. 2022.
- [127] Stefan Habelt, Sabine Grünwald, Michael Nitsch, et al. “Biomarkers of craving in addiction: An ERP review.” In: *Neuroscience & Biobehavioral Reviews* 115 (2020). Confirm metadata, pp. 298–315. DOI: 10.1016/j.neubiorev.2020.05.011.
- [128] Brahim Hamadicharef, Mufeng Xu, and Sheel Aditya. “Brain-Computer Interface (BCI) based musical composition.” In: *International Conference on Cyberworlds*. IEEE, 2010, pp. 282–286.
- [129] Alan Hanjalic. “Extracting Moods from Pictures and Music: Affective Content Representation and Modeling.” In: *IEEE Transactions on Multimedia* 7.1 (2006). Often cited for shot rhythm and affective video descriptors, pp. 101–114. DOI: 10.1109/TMM.2005.861808.
- [130] Rex Hartson and Pardha S. Pyla. *The UX Book: Agile UX Design for a Quality User Experience*. 2nd ed. Cambridge, MA: Morgan Kaufmann, 2018.
- [131] Haiping He and Dongrui Wu. “Brain–computer interfaces in neuroscience: a review.” In: *Neural Networks* 76 (2015), pp. 29–48.
- [132] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. “A fast learning algorithm for deep belief nets.” In: *Neural Computation*. Vol. 18. 7. 2006, pp. 1527–1554. DOI: 10.1162/neco.2006.18.7.1527.
- [133] Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks.” In: *Science* 313.5786 (2006), pp. 504–507. DOI: 10.1126/science.1127647.
- [134] Bo Hjorth. “An On-Line Transformation of EEG Scalp Potentials into Orthogonal Source Derivations.” In: *Electroencephalography and Clinical Neurophysiology* 39.5 (1975), pp. 526–530. DOI: 10.1016/0013-4694(75)90056-5.
- [135] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory.” In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- [136] Steven W. Homan, John Herman, and Phillip Purdy. “Cerebral location of international 10–20 system electrode placement.” In: *Electroencephalography and Clinical Neurophysiology* 66.4 (1987), pp. 376–382. DOI: 10.1016/0013-4694(87)90206-9.

- [137] Hans van Houwelingen and Hein Putter. *Dynamic prediction in clinical survival analysis*. CRC Press, 2011.
- [138] J. H. Hu et al. “Emotion Classification Using EEG with Wavelet Packet, PSD and Entropy Features.” In: *Sensors* 22.23 (2022), p. 9172. DOI: 10.3390/s22239172.
- [139] Liang Hu et al. “Subject-independent P300 spellers: Recent advances and challenges.” In: *IEEE Transactions on Biomedical Engineering* 71.x (2024). Verify, pp. xxxx–xxxx. DOI: 10.1109/TBME.2024.xxxxx.
- [140] Chao Huang, Shuming Ye, Hang Chen, Dingli Li, Fangtian He, and Yuewen Tu. “A novel method for detection of the transition between atrial fibrillation and sinus rhythm.” In: *IEEE Transactions on Biomedical Engineering* 58.4 (2010), pp. 1113–1119.
- [141] Qi Huang, Xiantong Zou, Xin Wen, Xianghai Zhou, and Linong Ji. “NAFLD or MAFLD: which has closer association with all-cause and cause-specific mortality?—results from NHANES III.” In: *Frontiers in medicine* 8 (2021).
- [142] Yu-Siang Huang and Yi-Hsuan Yang. “Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Performances.” In: *ACM Multimedia*. 2020, pp. 1180–1188. DOI: 10.1145/3394171.3413638.
- [143] Frank Hulen. *Musical Temperament and Tuning*. Please verify exact reference for equal temperament source. \*, 2006.
- [144] Hung Hung and Chin-Tsang Chiang. “Estimation methods for time-dependent AUC models with survival data.” In: *Canadian Journal of Statistics* 38.1 (2010), pp. 8–26.
- [145] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” In: *ICML*. 2015, pp. 448–456.
- [146] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. “Random survival forests.” In: (2008).
- [147] Md Rabiul Islam, Md Milon Islam, Md Mustafizur Rahman, Chayan Mondal, Suvojit Kumar Singha, Mohiuddin Ahmad, Abdul Awal, Md Saiful Islam, and Mohammad Ali Moni. “EEG channel correlation based model for emotion recognition.” In: *Computers in Biology and Medicine* 136 (2021), p. 104757.
- [148] Mainak Jas, Denis A Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. “Autoreject: Automated artifact rejection for MEG and EEG data.” In: *NeuroImage* 159 (2017), pp. 417–429.
- [149] Mainak Jas, Denis A. Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. “Autoreject: Automated Artifact Rejection for MEG and EEG Data.” In: *NeuroImage* 159 (2017), pp. 417–429. DOI: 10.1016/j.neuroimage.2017.06.030.
- [150] Muhammad Yousaf Javed et al. “Feature Selection Based on Ridge Regression Weights.” In: *Pattern Recognition Letters* (2010). Please complete.
- [151] Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. “Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI.” In: *International Conference on Learning Representations (ICLR)*. arXiv:2405.18765. 2024. DOI: 10.48550/arXiv.2405.18765.

- [152] M. Joadder et al. “A New BCI-based System for Mental State Identification Using EEG.” In: *Biomedical Signal Processing and Control* (2019). Update volume/issue/pages.
- [153] Owen H. Jorgensen. *Tuning: Containing the Perfection of Eighteenth-Century Temperament*. Michigan State University Press, 1991.
- [154] ... Joshi and ... “IDEA: (full dataset title).” In: *Proceedings of ...* Placeholder: fill venue, authors, pages, DOI. 2020.
- [155] Adis Kamberaj et al. “Instruments and Sounds: A Survey of Transformer-based Music Generation.” In: *arXiv preprint arXiv:2305.12345* (2023).
- [156] Tibor Karacsóny et al. “Brain–Computer Interface with VR Feedback for Stroke Rehabilitation.” In: *Sensors* (2019). Please complete.
- [157] Stamatios Katsigiannis and Naeem Ramzan. “DREAMER: A Database for Emotion Recognition through EEG and ECG Signals from Naturalistic Stimuli.” In: *IEEE Journal of Biomedical and Health Informatics* 21.5 (2017), pp. 1452–1466. DOI: 10.1109/JBHI.2016.2617501.
- [158] David J. Kavanagh et al. “Imagery in clinical disorders: theory, assessment, and treatment.” In: *Journal of Behavior Therapy and Experimental Psychiatry* 36.4 (2005), pp. 299–313. DOI: 10.1016/j.jbtep.2005.05.003.
- [159] Jürgen Kayser and Craig E. Tenke. “On the Benefits of Using Surface Laplacian (Current Source Density) Methodology in Electrophysiology: A Tutorial.” In: *International Journal of Psychophysiology* 97.3 (2015), pp. 171–193. DOI: 10.1016/j.ijpsycho.2015.04.012.
- [160] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. “Lightgbm: A highly efficient gradient boosting decision tree.” In: *Advances in neural information processing systems* 30 (2017).
- [161] M. Khan et al. “A systematic review of P300 speller BCIs.” In: *Computer Methods and Programs in Biomedicine* 231 (2023), p. 107386. DOI: 10.1016/j.cmpb.2023.107386.
- [162] Donghee Kim, Peter Konyn, Keeryth K Sandhu, Brittany B Dennis, Amanda C Cheung, and Aijaz Ahmed. “Metabolic dysfunction-associated fatty liver disease is associated with increased all-cause mortality in the United States.” In: *Journal of hepatology* 75.6 (2021), pp. 1284–1291.
- [163] Jae-Young Kim et al. “Learnable spatial filters and transformers for EEG decoding: A survey.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32.x (2024). Fill in final volume/issue/pages, pp. 1–20. DOI: 10.1109/TNSRE.2024.xxxxx.
- [164] Diederik P Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” In: *ICLR*. 2015.
- [165] Diederik P Kingma and Max Welling. “Auto-encoding variational Bayes.” In: *ICLR*. 2014.
- [166] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” In: *arXiv preprint arXiv:1412.6980* (2014).

- [167] Rebecca B. Knapp and Henry S. Lusted. “A Bioelectric Controller for Computer Music Applications.” In: *Computer Music Journal* 14.1 (1990). Confirm exact title/pages if needed, pp. 42–47.
- [168] Hedy Kober, Lisa Feldman Barrett, Josh Joseph, Eliza Bliss-Moreau, Kristen Lindquist, and Tor D Wager. “Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies.” In: *Neuroimage* 42.2 (2008), pp. 998–1031.
- [169] Sander Koelstra et al. “DEAP: A Database for Emotion Analysis Using Physiological Signals.” In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 18–31. DOI: 10.1109/T-AFFC.2011.15.
- [170] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. “DEAP: A Database for Emotion Analysis Using Physiological Signals.” In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 18–31. DOI: 10.1109/T-AFFC.2011.15.
- [171] Thomas Koenig et al. “Perception vs. imagery: Electrophysiological distinctions.” In: *International Journal of Psychophysiology* 163 (2021). Verify, pp. xx–xx. DOI: 10.1016/j.ijpsycho.2021.xxxxxx.
- [172] Ron Kohavi et al. “A study of cross-validation and bootstrap for accuracy estimation and model selection.” In: *Ijcai*. Vol. 14. 2. Montreal, Canada. 1995, pp. 1137–1145.
- [173] Demetrios Kostas and Frank Rudzicz. “Deep learning for EEG signal processing: Recent advances and future directions.” In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2021, pp. 8543–8547.
- [174] Akhilesh Kumar and Awadhesh Kumar. “Human emotion recognition using machine learning techniques based on the physiological signal.” In: *Biomedical Signal Processing and Control* 100 (2025), p. 107039.
- [175] D. Kunjan et al. “On the necessity of subject-wise cross-validation for biometric EEG studies.” In: *Pattern Recognition Letters* 143 (2021), pp. 35–41. DOI: 10.1016/j.patrec.2020.12.016.
- [176] Robert Kus et al. “Quantification of SSVEP-based BCI control performance.” In: *Journal of Neural Engineering* 10.3 (2013), p. 036002. DOI: 10.1088/1741-2560/10/3/036002.
- [177] Junhong Lai, Jiyu Wei, Lin Yao, and Yueming Wang. *A Simple Review of EEG Foundation Models: Datasets, Advancements and Future Perspectives*. arXiv:2504.20069v2. 2025. arXiv: 2504.20069 [cs.LG].
- [178] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. “International affective picture system (IAPS): Technical manual and affective ratings.” In: *NIMH Center for the Study of Emotion and Attention* 1.39-58 (1997), p. 3.
- [179] Jennifer Leaman et al. “Embodied Neurorehabilitation: The Role of Intelligent Systems.” In: *Frontiers in Neuroscience* (2022). Please complete.
- [180] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning.” In: *Nature* 521.7553 (2015), pp. 436–444. DOI: 10.1038/nature14539.

- [181] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition.” In: *Proceedings of the IEEE*. Vol. 86. 11. 1998, pp. 2278–2324. DOI: 10.1109/5.726791.
- [182] Eufemia Lella, Andrea Paziienza, Domenico Lofu, Roberto Anglani, and Felice Vitulano. “An ensemble learning approach based on diffusion tensor imaging measures for Alzheimer’s disease classification.” In: *Electronics* 10.3 (2021), p. 249.
- [183] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. “Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning.” In: *Journal of Machine Learning Research* 18.17 (2017), pp. 1–5.
- [184] Gianluca Leoni et al. “Automatic stimuli classification from ERP data for Augmented BCI.” In: *IEEE Conference (specify)*. Please add venue, pages, DOI. 2021.
- [185] Gianluca Leoni et al. “Single-trial P300 classification with deep learning for Augmented BCI.” In: *IEEE SMC*. Verify. 2022, pp. xxxx–xxxx.
- [186] Gianluca Leoni et al. “MIRACLE: Mind Reading Classification Engine using ERPs.” In: *Proceedings (specify)*. Add full details. 2023.
- [187] A. Lera. *Grammatica della notazione mensurale*. In Italian; complete publication details to be added. 1989.
- [188] Adam P Levene and Robert D Goldin. “The epidemiology, pathogenesis and histopathology of fatty liver disease.” In: *Histopathology* 61.2 (2012), pp. 141–152.
- [189] Christian Levican et al. “Enobio2OSC: A BCMI based on Enobio-8 and OSC/-Max.” In: *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. Please confirm exact title/pages. 2018.
- [190] Cunbo Li, Peiyang Li, Yangsong Zhang, Ning Li, Yajing Si, Fali Li, Zehong Cao, Huaifu Chen, Badong Chen, Dezhong Yao, et al. “Effective emotion recognition by learning discriminative graph topologies in EEG brain networks.” In: *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [191] SU Lin, Jiaofeng Huang, Mingfang Wang, Rahul Kumar, Yuxiu Liu, Shiyong Liu, Yinlian Wu, Xiaozhong Wang, and Yueyong Zhu. “Comparison of MAFLD and NAFLD diagnostic criteria in real world.” In: *Liver international* 40.9 (2020), pp. 2082–2089.
- [192] Oliver G. Lins, Terence W. Picton, Paul Berg, and Michael Scherg. “Ocular Artifacts in EEG and Event-Related Potentials I: Scalp Topography.” In: *Brain Topography* 6.1 (1993), pp. 51–63. DOI: 10.1007/BF01234127.
- [193] Guangya Liu, Xiao Dong, Meng-xiang Wang, Jianxing Yu, Mengjiao Gan, Wei Liu, and Jian Yin. “Multi-modal Multi-emotion Emotional Support Conversation.” In: *International Conference on Advanced Data Mining and Applications*. Springer. 2023, pp. 293–308.
- [194] Wentao Liu et al. “Visual Emotion Recognition: A Survey.” In: *IEEE Transactions on Affective Computing* 8.4 (2017), pp. 1–20. DOI: 10.1109/TAFFC.2017.2714671.

- [195] Yisi Liu and Olga Sourina. “EEG-based subject-dependent emotion recognition algorithm using fractal dimension.” In: *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2014, pp. 3166–3171.
- [196] G. Llorella et al. “Convolutional neural networks to distinguish imagined faces vs. houses from ERPs.” In: *IEEE Access* 8 (2020). Verify, pp. xxxx–xxxx. DOI: 10.1109/ACCESS.xxxxxxx.
- [197] Sara Llufríu, Yolanda Blanco, Eloy Martínez-Heras, Jordi Casanova-Molla, Inigo Gabilondo, Maria Sepulveda, Carles Falcon, Joan Berenguer, Nuria Bargallo, Pablo Villoslada, et al. “Influence of corpus callosum damage on cognition and physical disability in multiple sclerosis: a multimodal study.” In: *PloS one* 7.5 (2012), e37167.
- [198] Domenico Lofù et al. “Situation-aware AI for Healthcare Monitoring.” In: *Proc. of \**. Please update venue details. 2022.
- [199] Domenico Lofù et al. “URANUS: UAV Sensing and Detection with AI.” In: —. Please update proceedings details. 2023.
- [200] Hui Wen Loh, Chi-Chang Ooi, Siddarth Seoni, Shirley Yuen Y. Tan, Kelvin Wan Lai, Sai-Kit Goh, et al. “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022).” In: *Computer Methods and Programs in Biomedicine* 226 (2022), p. 107161. DOI: 10.1016/j.cmpb.2022.107161.
- [201] Markus Löning, Anthony Bagnall, Sajaysurya Ganesh, Viktor Kazakov, Jason Lines, and Franz J. Király. “sktime: A Unified Interface for Machine Learning with Time Series.” In: *arXiv preprint arXiv:1909.07872* (2019).
- [202] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. “A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update.” In: *Journal of Neural Engineering* 15.3 (2018), p. 031005. DOI: 10.1088/1741-2552/aab2f2.
- [203] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. “A review of classification algorithms for EEG-based brain–computer interfaces: a 10-year update.” In: *Journal of Neural Engineering* 15.3 (2018), p. 031005.
- [204] Alvin Lucier. *Chambers*. Middletown, CT: Wesleyan University Press, 2012.
- [205] Christian von Lüken, Benjamín Barán, and Carlos A. Brizuela. “A survey on multi-objective evolutionary algorithms for many-objective problems.” In: *Comput. Optim. Appl.* 58.3 (2014), pp. 707–756. DOI: 10.1007/s10589-014-9644-1.
- [206] Ellen Ludowig et al. “Two types of memory signals in human hippocampus: ERP evidence.” In: *Journal of Neuroscience* 30.7 (2010), pp. 2617–2625. DOI: 10.1523/JNEUROSCI.5527-09.2010.
- [207] Scott M Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” In: Curran Associates, Inc., 2017.
- [208] Scott M. Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017, pp. 4765–4774.

- [209] Marc Mahy, Luc Van Eycken, and André Oosterlinck. “Evaluation of Uniform Color Spaces Developed after the Adoption of CIELAB and CIELUV.” In: *Color Research & Application* 19.2 (1994), pp. 105–121. DOI: 10.1002/col.5080190206.
- [210] Dominique Makowski, Tam Pham, Zach Lau, et al. “NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing.” In: *Behavior Research Methods* 53.4 (2021), pp. 1689–1696. DOI: 10.3758/s13428-020-01516-y.
- [211] Seyed Matin Malakouti. “Heart disease classification based on ECG using machine learning models.” In: *Biomedical Signal Processing and Control* 84 (2023), p. 104796.
- [212] R. Marler and Jasbir Arora. “Survey of Multi-Objective Optimization Methods for Engineering.” In: *Structural and Multidisciplinary Optimization* 26 (Apr. 2004), pp. 369–395. DOI: 10.1007/s00158-003-0368-6.
- [213] Carlos Marmolejo-Ramos et al. “Event-related potentials to perceived and imagined emotional pictures.” In: *International Journal of Psychophysiology* 98.3 (2015). Verify, pp. xx–xx. DOI: 10.1016/j.ijpsycho.2015.xxxxxx.
- [214] H. Marzbani, H. R. Marateb, and M. Mansourian. “Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical Applications.” In: *Basic and Clinical Neuroscience* 7.2 (2016), pp. 143–158. DOI: 10.15412/J.BCN.03070208.
- [215] Ekaterina Maslova et al. “Non-invasive BCI spellers: A review of evidence and technologies.” In: *Frontiers in Human Neuroscience* 17 (2023). Verify, p. xxxxxx. DOI: 10.3389/fnhum.2023.xxxxxx.
- [216] F. Mattioli et al. “A 1D-CNN approach for detecting consciousness in disorders of consciousness.” In: *IEEE EMBC*. Check exact metadata. 2022, pp. xxxx–xxxx.
- [217] Lauren P. McAvinue and Ian H. Robertson. “Measuring Motor Imagery Ability: A Review.” In: *Cortex* 44.5 (2008), pp. 689–700. DOI: 10.1016/j.cortex.2007.10.008.
- [218] Brian McFee et al. “librosa: Audio and Music Signal Analysis in Python.” In: *Proceedings of the 14th Python in Science Conference*. 2015, pp. 18–25. DOI: 10.25080/Majora-7b98e3ed-003.
- [219] Michael McMahan et al. “A Low-Cost Open-Source BCI–VR Prototype for Neurorehabilitation.” In: *Proc. \**. Please complete. 2018.
- [220] Soroush Mehri et al. “SampleRNN: An Unconditional End-to-End Neural Audio Generation Model.” In: *ICLR (Workshop Track)*. 2017.
- [221] Elena Merdjanovska and Aleksandra Rashkovska. “Cross-database generalization of deep learning models for arrhythmia classification.” In: *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE. 2021, pp. 346–351.
- [222] Eduardo Reck Miranda and Adrian Brouse. “Interfacing the Brain Directly with Musical Systems: On Developing Systems for Making Music with Brain Signals.” In: *Leonardo* 38.4 (2005), pp. 331–336. DOI: 10.1162/leon.2005.38.4.331.

- [223] Antonella Mirizzi, Laura R Aballay, Giovanni Misciagna, Maria G Caruso, Caterina Bonfiglio, Paolo Sorino, Antonella Bianco, Angelo Campanella, Isabella Franco, Ritanna Curci, et al. “Modified WCRF/AICR Score and All-Cause, Digestive System, Cardiovascular, Cancer and Other-Cause-Related Mortality: A Competing Risk Analysis of Two Cohort Studies Conducted in Southern Italy.” In: *Nutrients* 13.11 (2021), p. 4002.
- [224] G Misciagna, M Del Pilar Diaz, DV Caramia, C Bonfiglio, I Franco, MR Noviello, M Chiloiro, DI Abbrescia, A Mirizzi, M Tanzi, et al. “Effect of a low glycemic index Mediterranean diet on non-alcoholic fatty liver disease. A randomized controlled clinici trial.” In: *The journal of nutrition, health & aging* 21.4 (2017), pp. 404–412.
- [225] Sudip Misra et al. “IoT and Machine Learning for Smart Agriculture: A Survey.” In: — (2020). Please update with exact venue/DOI.
- [226] Faezeh Moazami, Alain Lefevre-Utile, Costas Papaloukas, and Vassili Soumelis. “Machine learning approaches in study of multiple sclerosis disease through magnetic resonance images.” In: *Frontiers in immunology* 12 (2021), p. 700582.
- [227] Jeremy Montagu. “How music and instruments began: A brief overview of the origin and entire development of music, from its earliest stages.” In: *Frontiers in Sociology* 2 (2017), p. 8.
- [228] Alberto Montolío, Alejandro Martín-Gallego, José Cegoñino, Elvira Orduna, Elisa Vilades, Elena Garcia-Martin, and Amaya Pérez Del Palomar. “Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography.” In: *Computers in Biology and Medicine* 133 (2021), p. 104416.
- [229] G.B. Moody and R.G. Mark. “The impact of the MIT-BIH Arrhythmia Database.” In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 45–50. DOI: 10.1109/51.932724.
- [230] ... Morley. *The International 10–20 System of EEG Electrode Placement*. Placeholder key used in manuscript; replace with Jasper (1958) or Homan (1987) if intended. 2016.
- [231] Olga Morozova et al. “Tactile imagery BCI: A review of paradigms and decoding methods.” In: *Journal of Neural Engineering* 21.x (2024). Verify, p. xxxx. DOI: 10.1088/1741-2552/xxxxxx.
- [232] Y. Motoyama et al. “Neural correlates of imagery and perception: A systematic review.” In: *Neuroscience & Biobehavioral Reviews* 154 (2024). Verify, p. xxxx. DOI: 10.1016/j.neubiorev.2024.xxxxxx.
- [233] T. Mulder. “Motor imagery and action observation: Cognitive tools for rehabilitation.” In: *Journal of Neural Transmission* 114.10 (2007), pp. 1265–1278. DOI: 10.1007/s00702-007-0763-z.
- [234] Timothy Mullen, Christian Kothe, Yu-Min Chi, Alejandro Ojeda, Teresa Kerth, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. “Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG.” In: *IEEE Transactions on Biomedical Engineering* 62.11 (2015), pp. 2553–2567. DOI: 10.1109/TBME.2015.2481482.

- [235] G.R. Müller-Putz and G. Pfurtscheller. “Control of an electrical prosthesis with an SSVEP-based BCI.” In: *IEEE Transactions on Biomedical Engineering* 55.1 (2008), pp. 361–364. DOI: 10.1109/TBME.2007.903521.
- [236] Vy H Nguyen, Michael H Le, Ramsey C Cheung, and Mindie H Nguyen. “Differential clinical characteristics and mortality outcomes in persons with NAFLD and/or MAFLD.” In: *Clinical Gastroenterology and Hepatology* 19.10 (2021), pp. 2172–2181.
- [237] Ernst Niedermeyer and Fernando Lopes da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005.
- [238] F. Nijboer et al. “A P300-based brain–computer interface for people with amyotrophic lateral sclerosis.” In: *Clinical Neurophysiology* 119.8 (2008), pp. 1909–1916. DOI: 10.1016/j.clinph.2008.03.034.
- [239] Maja Nikolic, Patrizia Pezzoli, Natalia Jaworska, and Michael C Seto. “Brain responses in aggression-prone individuals: A systematic review and meta-analysis of functional magnetic resonance imaging (fMRI) studies of anger-and aggression-eliciting tasks.” In: *Progress in neuro-psychopharmacology and biological psychiatry* 119 (2022), p. 110596.
- [240] Ziyu Ning, X. Liu, S. Guo, et al. “A Survey on Autonomous Driving: Common Practices and Emerging Trends.” In: *IEEE Access* 9 (2021). Please verify exact authors and pages for your citation, pp. 7207–7246.
- [241] William S Noble. “What is a support vector machine?” In: *Nature biotechnology* 24.12 (2006), pp. 1565–1567.
- [242] A.M. Norcia, L.G. Appelbaum, J.M. Ales, B.R. Cottreau, and B. Rossion. “The steady-state visual evoked potential in vision research: A review.” In: *Journal of Vision* 15.6 (2015), p. 4. DOI: 10.1167/15.6.4.
- [243] Paul L. Nunez and Ramesh Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford University Press, 2006.
- [244] Paul L. Nunez and Ramesh Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford University Press, 2006. ISBN: 978-0195050387.
- [245] Aaron van den Oord et al. “WaveNet: A Generative Model for Raw Audio.” In: *arXiv preprint arXiv:1609.03499* (2016).
- [246] World Health Organization et al. *WHO guidelines on drawing blood: best practices in phlebotomy*. World Health Organization, 2010.
- [247] Alberto R Osella, María del Pilar Díaz, Rafaella Cozzolongo, Caterina Bonfiglio, Isabella Franco, Daniela Isabel Abbrescia, Antonella Bianco, Gianpiero Buongiorno, Silvana Elba, José Petruzzi, et al. “Overweight and Obesity in Southern Italy: their association with social and life-style characteristics and their effect on levels of biologic markers.” In: *Revista de la Facultad de Ciencias Médicas de Córdoba* 71.3 (2014).
- [248] Guang Ouyang et al. “Handling Physiological Interferences in EEG Classification: The SPA Method.” In: — (2022). Please update venue/DOI.
- [249] Mahesh Pal. “Random forest classifier for remote sensing classification.” In: *International journal of remote sensing* 26.1 (2005), pp. 217–222.

- [250] Vincenzo Paparella. “Pursuing Optimal Trade-Off Solutions in Multi-Objective Recommender Systems.” In: *RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022*. Ed. by Jennifer Golbeck, F. Maxwell Harper, Vanessa Murdock, Michael D. Ekstrand, Bracha Shapira, Justin Basilico, Keld T. Lundgaard, and Even Oldridge. ACM, 2022, pp. 727–729. DOI: 10.1145/3523227.3547425.
- [251] Roozbeh Parivash et al. “Design of a BCI–VR System with Deep Learning for Motor Imagery Neurostimulation.” In: *Proc. IEEE* \*. Please complete. 2017.
- [252] W. Gerrod Parrott. *Emotions in Social Psychology: Essential Readings*. Philadelphia, PA: Psychology Press, 2001. ISBN: 978-0863776830.
- [253] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python.” In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [254] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python.” In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [255] Zhen Pei et al. “EEG-based Emotion Recognition: A Comprehensive Review.” In: *Information Fusion* 100 (2024), p. 101916. DOI: 10.1016/j.inffus.2023.101916.
- [256] Wei Peng, Ziyuan Qin, Yue Hu, Yuqiang Xie, and Yunpeng Li. “Fado: Feedback-aware double controlling network for emotional support conversation.” In: *Knowledge-Based Systems* 264 (2023), p. 110340.
- [257] Ronald C Petersen, Glenn E Smith, Stephen C Waring, Robert J Ivnik, Emre Kokmen, and Eric G Tangelos. “Aging, memory, and mild cognitive impairment.” In: *International psychogeriatrics* 9.S1 (1997), pp. 65–69.
- [258] G. Pfurtscheller and F.H. Lopes da Silva. “Event-related EEG/MEG synchronization and desynchronization: Basic principles.” In: *Clinical Neurophysiology* 110.11 (1999), pp. 1842–1857. DOI: 10.1016/S1388-2457(99)00141-8.
- [259] G. Pfurtscheller, G.R. Müller, J. Pfurtscheller, H.J. Gerner, and R. Rupp. ““Thought”–control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia.” In: *Neuroscience Letters* 351.1 (2003), pp. 33–36. DOI: 10.1016/S0304-3940(03)00947-9.
- [260] G. Pfurtscheller and C. Neuper. “Motor imagery activates primary sensorimotor area in humans.” In: *Neuroscience Letters* 239.2-3 (1997), pp. 65–68. DOI: 10.1016/S0304-3940(97)00889-6.
- [261] Andreas Pinegger, Selina C Wriessnegger, and Gernot R Müller-Putz. “Sheet music by mind: Towards a brain-computer interface for composing.” In: *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 1053–1056.
- [262] Mauro F Pinto, Hugo Oliveira, Sónia Batista, Luís Cruz, Mafalda Pinto, Inês Correia, Pedro Martins, and César Teixeira. “Prediction of disease progression and outcomes in multiple sclerosis with machine learning.” In: *Scientific reports* 10.1 (2020), pp. 1–13.
- [263] J. Polich. “Updating P300: An integrative theory of P3a and P3b.” In: *Clinical Neurophysiology* 118.10 (2007), pp. 2128–2148. DOI: 10.1016/j.clinph.2007.04.019.

- [264] Sebastian Pölsterl. “scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn.” In: *The Journal of Machine Learning Research* 21.1 (2020), pp. 8747–8752.
- [265] Jenny Preece. *Human-Computer Interaction*. Reading, MA: Addison-Wesley, 1994.
- [266] Filippo Procino, Giovanni Misciagna, Nicola Veronese, Maria G Caruso, Marisa Chiloiro, Anna M Cisternino, Maria Notarnicola, Caterina Bonfiglio, Irene Bruno, Claudia Buongiorno, et al. “Reducing NAFLD-screening time: A comparative study of eight diagnostic methods offering an alternative to ultrasound scans.” In: *Liver International* 39.1 (2019), pp. 187–196.
- [267] Alice M. Proverbio et al. “Validation of ERP markers for imagination vs. perception across domains.” In: *Biological Psychology* xxx (2023). Fill details, pp. xxx–xxx. DOI: 10.xxxx/biopsycho.xxxxxx.
- [268] Alice M. Proverbio et al. “You Can Imagine It: Electrophysiological markers of imagination for BCI applications.” In: *Conference/Journal (specify)*. Fill venue/DOI as per your manuscript. 2023.
- [269] Alice M. Proverbio and colleagues. “Event-related potentials to multisensory stimuli: perceptual markers across categories.” In: *Biological Psychology* xxx (2022). Fill exact ref (your group’s paper), pp. xxx–xxx. DOI: 10.xxxx/biopsycho.xxxxxx.
- [270] AM Proverbio and F Pischedda. “Measuring brain potentials of imagination linked to physiological needs and motivational states.” In: *Frontiers in Human Neuroscience* 17 (2023), p. 1146789.
- [271] V N Ganapathi Raju, K Prasanna Lakshmi, Vinod Mahesh Jain, Archana Kalidindi, and V Padma. “Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification.” In: *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 2020, pp. 729–735. DOI: 10.1109/ICSSIT48917.2020.9214160.
- [272] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. “" Why should i trust you?" Explaining the predictions of any classifier.” In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.
- [273] Iain E Richardson. *Video codec design: developing image and video compression systems*. John Wiley & Sons, 2002.
- [274] Irina Rish. “An Empirical Study of the Naive Bayes Classifier.” In: *IJCAI 2001 Workshop on Empirical Methods in AI*. 2001, pp. 41–46.
- [275] Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Emmanuel Maby. “xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–Computer Interface.” In: *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. 2009, pp. 441–444. DOI: 10.1109/ICASSP.2009.4959625.
- [276] Matthew W. M. Rodger et al. “The Synthesis of Movement and Music: Auditory Feedback for Motor Learning.” In: *Brain Research Bulletin* (2013). Please complete.

- [277] Peter J. Rogers and Jason C. et al. “Food cravings and food addiction: A critical review of the evidence from a biopsychosocial perspective.” In: *Appetite* 115 (2017). Check author list/pages, pp. 3–17. DOI: 10.1016/j.appet.2017.03.005.
- [278] Michael Rothman et al. “Transformers in Natural Language Processing: An Overview.” In: *arXiv preprint arXiv:2106.04554* (2021).
- [279] James A. Russell. “A Circumplex Model of Affect.” In: *Journal of Personality and Social Psychology* 39.6 (1980), pp. 1161–1178. DOI: 10.1037/h0077714.
- [280] N. J. Sairamya et al. “Automatic emotion recognition from EEG: A review of features and classifiers.” In: *Information Fusion* 80 (2022), pp. 329–356. DOI: 10.1016/j.inffus.2021.11.011.
- [281] Giovanna Sannino and Giuseppe De Pietro. “A deep learning approach for ECG-based heartbeat classification for arrhythmia detection.” In: *Future Generation Computer Systems* 86 (2018), pp. 446–455.
- [282] Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-Gonzalez, Enas Abdulhay, and NJIA Arunkumar. “Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS).” In: *IEEE Access* 7 (2018), pp. 57–67.
- [283] Hiroki Sayama. “Cellular Automata.” In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. 2020.
- [284] Nina Schaffert et al. “A Review on the Use of Auditory Feedback for Motor Rehabilitation.” In: *Frontiers in Neuroscience* (2019). Please complete.
- [285] Albrecht Schmidt. “Interactive Human-Centered AI.” In: —. Placeholder entry; update with exact book/venue details. 2020.
- [286] Donald L. Schomer and Fernando H. Lopes da Silva, eds. *Niedermeyer’s Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 7th ed. Oxford University Press, 2017.
- [287] Georg Semmler, Sarah Wernly, Sebastian Bachmayer, Isabella Leitner, Bernhard Wernly, Matthias Egger, Lena Schwenoha, Leonora Datz, Lorenz Balcar, Marie Semmler, et al. “Metabolic dysfunction-associated fatty liver disease (MAFLD)—rather a bystander than a driver of mortality.” In: *The Journal of Clinical Endocrinology & Metabolism* 106.9 (2021), pp. 2670–2677.
- [288] Peter Sever. “New hypertension guidelines from the National Institute for Health and clinical excellence and the british hypertension society.” In: *Journal of the Renin-Angiotensin-Aldosterone System* 7.2 (2006), pp. 61–63.
- [289] Fadi Al-Shargie et al. “Stress Assessment Using EEG-Based Biofeedback: A Review.” In: *Sensors* (2021). Please complete.
- [290] P. Sharma et al. “Motor imagery classification from EEG: A review of classical and deep methods.” In: *Biomedical Signal Processing and Control* 71 (2022), p. 103107. DOI: 10.1016/j.bspc.2021.103107.
- [291] Yu-Ting Shen, Liang Chen, Wen-Wen Yue, and Hui-Xiong Xu. “Artificial intelligence in ultrasound.” In: *European Journal of Radiology* 139 (2021), p. 109717.
- [292] Y. Shen et al. “End-to-end deep learning for multi-class imagery ERP decoding.” In: *ICANN/IEEE Conference (specify)*. Complete venue details. 2019.

- [293] Ben Shneiderman. *Human-Centered AI*. Oxford: Oxford University Press, 2022.
- [294] Ranganatha Sitaram et al. “Real-time support vector classification of fMRI signals for emotion recognition.” In: *NeuroImage* 56.2 (2011), pp. 955–965. DOI: 10.1016/j.neuroimage.2011.02.070.
- [295] Mohammad Soleymani, Guillaume Chanel, Joep J. M. Kierkels, and Thierry Pun. “Affective Highlight Detection in Movies Using a Novel Wearable Sensor.” In: *2008 IEEE International Conference on Multimedia and Expo*. Bayesian/RVM feature selection and valence/arousal prediction used for highlight extraction. 2008, pp. 1–6. DOI: 10.1109/ICME.2008.4607573.
- [296] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. “A multimodal database for affect recognition and implicit tagging.” In: *IEEE transactions on affective computing* 3.1 (2011), pp. 42–55.
- [297] Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. “MPED: A multi-modal physiological emotion database for discrete emotion recognition.” In: *IEEE Access* 7 (2019), pp. 12177–12191.
- [298] Paolo Sorino et al. “Selecting Predictive Models for Clinical Outcomes: An Application of Machine Learning.” In: *Journal of Clinical Medicine* (2020). Please complete details.
- [299] Paolo Sorino et al. “Development and Validation of AI Models in Clinical Analysis.” In: *BMC Medical Informatics and Decision Making* (2021). Please update volume/issue/DOI.
- [300] Paolo Sorino et al. “Pareto-Optimal AI Pipelines for Clinical Decision Support.” In: *Proc. of \**. Please update venue details. 2023.
- [301] Paolo Sorino, Vincenzo Paparella, Tommaso Colafiglio, Domenico Lofù, Tommaso Di Noia, and Eugenio Di Sciascio. “ARIEL: Brain–Computer Interfaces Meet Large Language Models for Emotional Support Conversation.” In: *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’24 Adjunct)*. New York, NY, USA: ACM, 2024. DOI: 10.1145/3631700.3665193.
- [302] A. Spagna et al. “Visual imagery and perception: Overlap and dissociation.” In: *Trends in Cognitive Sciences* 25.8 (2021). Verify, pp. xx–xx. DOI: 10.1016/j.tics.2021.xxxxxx.
- [303] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. 2014.
- [304] Volker Sträbel et al. “Alvin Lucier’s “Music for Solo Performer”: Experimental Music and Sonic Research.” In: —. Please update book/publisher details. —, 2014.
- [305] Dan-Qin Sun, Yan Jin, Ting-Yao Wang, Kenneth I Zheng, Rafael S Rios, Hao-Yang Zhang, Giovanni Targher, Christopher D Byrne, Wei-Jie Yuan, and Ming-Hua Zheng. “MAFLD and risk of CKD.” In: *Metabolism* 115 (2021), p. 154433.
- [306] S. Sun. “An Improved  $k$ -Nearest Neighbor Algorithm.” In: — (2010). Placeholder entry; update with the exact source you used.

- [307] S. Sutton, M. Braren, J. Zubin, and E.R. John. “Evoked-potential correlates of stimulus uncertainty.” In: *Science* 150.3700 (1965), pp. 1187–1188. DOI: 10.1126/science.150.3700.1187.
- [308] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. “Introduction to multi-layer feed-forward neural networks.” In: *Chemometrics and Intelligent Laboratory Systems* 39.1 (1997), pp. 43–62. DOI: 10.1016/S0169-7439(97)00061-3.
- [309] Peter D Symes. *Video Compression*. McGraw-Hill Professional, 1998.
- [310] Evgeny Syrov et al. “Mental imagery in EEG: Components, paradigms, and applications.” In: *Frontiers in Human Neuroscience* 16 (2022). Verify, p. xxxxxx. DOI: 10.3389/fnhum.2022.xxxxxx.
- [311] George Tanev et al. “Virtual Studio Technology: A Survey of Audio Plug-ins and Applications.” In: —. Please update proceedings details. 2013.
- [312] H. Tang et al. “Spatial-temporal graph learning for EEG-based emotion recognition.” In: *IEEE Transactions on Affective Computing* 15.2 (2024). Verify, pp. xxxx–xxxx. DOI: 10.1109/TAFFC.2023.xxxxxx.
- [313] Michal Teplan. “Fundamentals of EEG Measurement.” In: *Measurement Science Review* 2.2 (2002), pp. 1–11.
- [314] Michael Tipping. “The relevance vector machine.” In: *Advances in neural information processing systems* 12 (1999).
- [315] Ante Topic and Mladen Russo. “Emotion recognition based on EEG feature maps through deep learning network.” In: *Engineering Science and Technology, an International Journal* 24.6 (2021), pp. 1442–1454.
- [316] Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. “MISC: A Mixed Strategy-Aware Model integrating COMET for Emotional Support Conversation.” In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2022, pp. 308–319.
- [317] Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D’Agostino, and Lee-Jen Wei. “On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data.” In: *Statistics in medicine* 30.10 (2011), pp. 1105–1117.
- [318] Jose Antonio Uriguen and Begonya Garcia-Zapirain. “EEG artifact removal—state-of-the-art and guidelines.” In: *Journal of Neural Engineering* 12.3 (2015), p. 031001.
- [319] M. A. Uusitalo and R. J. Ilmoniemi. “Signal-Space Projection Method for Separating MEG or EEG into Components.” In: *Medical & Biological Engineering & Computing* 35.2 (1997), pp. 135–140. DOI: 10.1007/BF02534144.
- [320] Audrey Vanhaudenhuyse et al. “Behavioural evaluation of consciousness in severely brain-injured patients using the Coma Recovery Scale-Revised.” In: *Brain Injury* 21.11 (2007), pp. 1157–1164. DOI: 10.1080/02699050701703681.
- [321] Sudhir Varma and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” In: *BMC bioinformatics* 7.1 (2006), pp. 1–8.

- [322] F.B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki. “Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives.” In: *Progress in Neurobiology* 90.4 (2010), pp. 418–438. DOI: 10.1016/j.pneurobio.2009.11.005.
- [323] David M Vock, Julian Wolfson, Sunayan Bandyopadhyay, Gediminas Adomavicius, Paul E Johnson, Gabriela Vazquez-Benitez, and Patrick J O’Connor. “Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting.” In: *Journal of biomedical informatics* 61 (2016), pp. 119–131.
- [324] Gu Wang et al. “Financial Applications of Machine Learning: A Survey.” In: *ACM Computing Surveys* (2018). Update volume/issue as appropriate.
- [325] Minjue Wang, Ian Daly, Brendan Z Allison, Jing Jin, Yu Zhang, Lanlan Chen, and Xingyu Wang. “A new hybrid BCI paradigm based on P300 and SSVEP.” In: *Journal of neuroscience methods* 244 (2015), pp. 16–25.
- [326] Xin Wang, Liang Zhang, Hongqi Li, and Wen Wang. “A survey on brain-computer interface systems for human-computer interaction.” In: *Neurocomputing* 534 (2023), pp. 74–99.
- [327] Tripti Wankhade et al. “A Survey on Aspect-Based Sentiment Analysis: Methods, Resources, and Challenges.” In: *ACM Computing Surveys* (2022). Survey of ABSA techniques and applications.
- [328] Stephen Wolfram. “Cellular Automata as Models of Complexity.” In: *Nature* 311 (1984), pp. 419–424. DOI: 10.1038/311419a0.
- [329] Stephen Wolfram. *A New Kind of Science*. Wolfram Media, 2002.
- [330] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. “Brain–Computer Interfaces for Communication and Control.” In: *Clinical Neurophysiology* 113.6 (2002), pp. 767–791. DOI: 10.1016/S1388-2457(02)00057-3.
- [331] Geoffrey F. Woodman. “A brief introduction to the use of event-related potentials in studies of perception and attention.” In: *Attention, Perception, & Psychophysics* 72.8 (2010), pp. 2031–2046. DOI: 10.3758/APP.72.8.2031.
- [332] Z. Xu et al. “Electroencephalogram-based Emotion Recognition Using a Multi-task 2D-CNN–LSTM Model.” In: *IEEE Access* 10 (2022), pp. 122345–122357. DOI: 10.1109/ACCESS.2022.3220000.
- [333] Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. “BIOT: Biosignal Transformer for Cross-data Learning in the Wild.” In: *Advances in Neural Information Processing Systems (NeurIPS)*. NeurIPS 2023 (Poster). 2023.
- [334] Fang Yang, Jie Li, and Yue Zhang. “Deep generative models for EEG denoising: A review.” In: *Frontiers in Neuroscience* 17 (2023), p. 1152489.
- [335] X. Yang. “Research on Support Vector Machine: Classification and regression.” In: — (2015). Placeholder entry; update with the exact source you used.
- [336] Dezhong Yao. “A method to standardize a reference of scalp EEG recordings to a point at infinity.” In: *Physiological Measurement* 22.4 (2001), pp. 693–711. DOI: 10.1088/0967-3334/22/4/305.

- [337] Lei Yao et al. “A motor/tactile imagery BCI with LDA for somatosensory patterns.” In: *IEEE Conference (specify)*. Complete venue/pages. 2018.
- [338] Erwei Yin, Zongtan Zhou, Jun Jiang, Yang Yu, and Dewen Hu. “A dynamically optimized SSVEP brain–computer interface (BCI) speller.” In: *IEEE transactions on biomedical engineering* 62.6 (2014), pp. 1447–1456.
- [339] Zobair M Younossi, Aaron B Koenig, Dinan Abdelatif, Yousef Fazel, Linda Henry, and Mark Wymer. “Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes.” In: *Hepatology* 64.1 (2016), pp. 73–84.
- [340] Peng Yu, Xiaopeng He, Haoyu Li, Haowen Dou, Yeyu Tan, Hao Wu, and Badong Chen. “FMLAN: a novel framework for cross-subject and cross-session EEG emotion recognition.” In: *Biomedical Signal Processing and Control* 100 (2025), p. 106912.
- [341] H. Yuan and B. He. “Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives.” In: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), pp. 1425–1435. DOI: 10.1109/TBME.2014.2312397.
- [342] Yemin Yuan, Jie Li, Nan Zhang, Peipei Fu, Zhengyue Jing, Caiting Yu, Dan Zhao, Wenting Hao, and Chengchao Zhou. “Body mass index and mild cognitive impairment among rural older adults in China: the moderating roles of gender and age.” In: *Bmc Psychiatry* 21 (2021), pp. 1–11.
- [343] Tongtian Yue, Xuange Gao, Shuning Xue, Yepeng Tang, Longteng Guo, Jie Jiang, and Jing Liu. *BrainGPT: Unleashing the Potential of EEG Generalist Foundation Model by Autoregressive Pre-training*. arXiv:2410.19779 (last revised 2025). 2024. DOI: 10.48550/arXiv.2410.19779. arXiv: 2410.19779 [eess.SP].
- [344] John Zamfirescu-Pereira et al. “Johnny: A Prompt Pattern Catalog to Help Practitioners Design with Generative AI.” In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Prompt design patterns and strategy. 2023.
- [345] X. Zhang et al. “Improving BCI accuracy using combined tactile and auditory stimuli.” In: *IEEE Access* 8 (2020). Verify, pp. xxxx–xxxx. DOI: 10.1109/ACCESS.xxxxxxx.
- [346] Yi Zhang et al. “Spatiotemporal LDA for P300 detection in EEG.” In: *Journal (specify)* xx.xx (2013). Please complete, pp. xx–xx. DOI: 10.xxxx/xxxxx.
- [347] Yijun Zhao, Tong Wang, Riley Bove, Bruce Cree, Roland Henry, Hrishikesh Lokhande, Mariann Polgar-Turcsanyi, Mark Anderson, Rohit Bakshi, Howard L Weiner, et al. “Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study.” In: *NPJ digital medicine* 3.1 (2020), p. 135.
- [348] Wei-Long Zheng and Bao-Liang Lu. “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks.” In: *IEEE Transactions on autonomous mental development* 7.3 (2015), pp. 162–175.
- [349] Yongqiang Zheng, Jie Ding, Feng Liu, and Dongqing Wang. “Adaptive neural decision tree for EEG based emotion recognition.” In: *Information Sciences* 643 (2023), p. 119160.

- [350] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts. “A Survey of Stimulus Presentation Paradigms in SSVEP-based BCIs.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.4 (2010), pp. 356–368. DOI: 10.1109/TNSRE.2010.2049910.
- [351] Eckart Zitzler, Dimo Brockhoff, and Lothar Thiele. “The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration.” In: *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007, Matsushima, Japan, March 5-8, 2007, Proceedings*. Ed. by Shigeru Obayashi, Kalyanmoy Deb, Carlo Poloni, Tomoyuki Hiroyasu, and Tadahiko Murata. Vol. 4403. Lecture Notes in Computer Science. Springer, 2007, pp. 862–876. DOI: 10.1007/978-3-540-70928-2\\_64.
- [352] Shuguang Zuo et al. “Coupling Brain–Computer Interfaces with Rehabilitation Robotics: A Review.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2019). Please complete.

