Leveraging artificial intelligence for enhanced and human-centered healthcare solutions

Department of Electrical and Information Engineering

Electrical and Information Engineering Ph.D. Program

SSD: ING-INF/05 - Information Processing Systems

Final Dissertation

Leveraging Artificial Intelligence for Enhanced and Human-centered Healthcare Solutions

---

by

Paolo Sorino

Supervisor

Prof. Tommaso Di Noia

Co-supervisor

Prof. Fedelucio Narducci

Dr. Rodolfo Sardone

Coordinator of the Ph.D. Program

Prof. Mario Carpentieri

Course XXXVII, 01/11/2021 - 31/10/2024

Department of Electrical and Information Engineering

Electrical and Information Engineering Ph.D. Program

SSD: ING-INF/05 - Information Processing Systems

Final Dissertation

Leveraging Artificial Intelligence for Enhanced and Human-centered Healthcare Solutions

---

by

Paolo Sorino

Referees

Prof. Pierangelo Veltri

Prof. Pietro Pinoli

Supervisor

Prof. Tommaso Di Noia

Co-supervisor

Prof. Fedelucio Narducci

Dr. Rodolfo Sardone

Coordinator of the Ph.D. Program

Prof. Mario Carpentieri

Course XXXVII, 01/11/2021 - 31/10/2024

Dedicated to those I love, near, far, and beyond this world.

Abstract

Artificial Intelligence (AI) is increasingly recognized as a transformative force in healthcare, offering unprecedented opportunities to enhance disease diagnosis, management, and prevention. This PhD thesis is rooted in two fundamental research areas: the application of AI to health and epidemiological data for the purposes of disease prevention and monitoring, and the utilization of AI techniques for the analysis of bioelectrical signals to support clinical decision-making.

The first research area delves into the sophisticated analysis of extensive health and epidemiological datasets using cutting-edge machine learning (ML) methodologies. The objective is to uncover significant patterns that can inform and improve the prevention and management of chronic diseases. By identifying these patterns, the research enables the creation of personalized intervention strategies tailored to individual patient profiles, while also optimizing disease management on a broader, population-wide scale. This approach not only contributes to the advancement of public health but also sets the stage for more proactive healthcare practices.

The second research focus of this thesis explores the development and application of advanced ML and deep learning (DL) models for the interpretation of bioelectrical signals, such as electroencephalograms (EEG), electrocardiograms (ECG), and electromyograms (EMG). It is important to point out that non-invasive technologies such as brain-computer interfaces (BCIs) were used for the analysis of EEG signals. The AI-driven models developed in this PhD thesis aim to enhance the accuracy and reliability of medical diagnostics, facilitating more precise and personalized clinical decisions. The integration of these models into clinical workflows has the potential to revolutionize patient care by providing healthcare professionals with powerful tools for diagnosis and treatment planning.

The practical outcomes of this research are profound, offering novel tools and frameworks that bridge the gap between AI innovation and clinical application. By incorporating explainable Artificial Intelligence (XAI) principles, the models developed in this thesis are designed to be transparent and interpretable, ensuring that healthcare

professionals can trust and effectively use these advanced technologies in their daily practice.

In summary, this PhD thesis makes significant contributions to the intersection of AI and medicine, addressing key challenges in the interpretation of health and epidemiological data as well as the analysis of bioelectrical signals. The findings presented here lay a robust foundation for future advancements in personalized medicine and public health, ultimately aiming to improve patient outcomes and the overall efficacy of healthcare systems.

All contributions made in this thesis are detailed in the respective chapters, providing a comprehensive overview of the research conducted and its impact on the field of AI in healthcare.

Publications

Some ideas and figures have appeared previously in other publications. A complete list of my publications during the PhD is available in the following. Note that, in some cases, the author list follows the alphabetical order.

[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). **To appear**.

[3] Simona Aresta, Ilaria Bortone, Francesco Bottiglione, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Mariapia Musci, Fedelucio Narducci, Andrea Pazienza, Rodolfo Sardone, et al. "Combining biomechanical features and machine learning approaches to identify fencers' levels for training support." In: *Applied Sciences* 12.23 (2022), p. 12350.

[4] 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)*. IEEE. 2022, pp. 3228–3233.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[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, 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.

[11] 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.

[12] 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.

[13] Danilo Danese, Tommaso Di Noia, Angela Lombardi, Domenico Lofù, Fatemeh Nazary, Rodolfo Sardone, and Paolo Sorino. "Integrating eXplainable AI in Healthcare: A Web Application Framework for Advancing the One Health Paradigm." 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). **To appear**.

[14] Domenico Lofù, Pietro Di Gennaro, Paolo Sorino, Tommaso Di Noia, and Eugenio Di Sciascio. "Cpu-side comparison for key agreement between tree parity machines and standard cryptographic primitives." In: *2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT)*. IEEE. 2022, pp. 1–6.

[15] Domenico Lofù, Paolo Sorino, Tommaso Colafiglio, Alberto Rubén Osella, Caterina Bonfiglio, 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: ***submitted to Computer Methods and Programs in Biomedicine Update*** (2024).

[16] Domenico Lofù, Paolo Sorino, Tommaso Di Noia, and Eugenio Di Sciascio. "Towards a Federated Intrusion Detection System based on Neuromorphic Computing." In: *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE. 2024, pp. 1–5.

[17] 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.

[18] Paolo Sorino. "Blockchain and AI to Build an Alzheimer's Risk Calculator." In: *International Conference on Web Engineering*. Springer. 2022, pp. 432–436.

[19] 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.

[20] 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)*. **To appear**. 2024.

[21] 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.

[22] Rossella Tatoli, Luisa Lampignano, Ilaria Bortone, Rossella Donghia, Fabio Castellana, Roberta Zupo, Sarah Tirelli, Sara De Nucci, Annamaria Sila, Annalidia Natuzzi, et al. "Dietary patterns associated with diabetes in an older population from Southern Italy using an unsupervised learning approach." In: *Sensors* 22.6 (2022), p. 2193.

[23] Paola Zinno, Francesco Maria Calabrese, Emily Schifano, Paolo Sorino, Raffaella Di Cagno, Marco Gobbetti, Eugenio Parente, Maria De Angelis, and Chiara Devirgiliis. "FDF-DB: A Database of Traditional Fermented Dairy Foods and Their Associated Microbiota." In: *Nutrients* 14.21 (2022), p. 4581.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Thesis Statement

The thesis is divided into thematic chapters, each covering the basics, analyses, and proposals of documents related to a specific theme. It should be noted that the order of the papers in the chapters doesn't necessarily follow chronological order. The goal was to provide a complete and cohesive narrative of the entire thesis, from using Artificial intelligence (AI) models to support diagnosis, to acquiring and analyzing bioelectric signals, to propose concrete support tools in the medical domain.

To begin with, **Chapter 2** introduces the foundational concepts and methodologies for utilizing AI to accelerate and refine diagnostic practices. This chapter is essential for understanding how AI can enhance effective patient care and timely treatment decisions. Simultaneously, **Chapter 3** delves into the application of AI for analyzing bioelectric signals such as Electroencephalograms (EEGs), ECGs, and, Electromyography (EMG) showcasing the potential of these technologies to significantly enhance clinical decision-making processes.

The narrative progresses with **Chapter 4**, where the focus shifts to the formalization, theoretical and empirical analysis, and the development of new AI-based diagnostic systems. This chapter discusses how AI models not only improve the accuracy of medical diagnostics but also the efficiency of healthcare operations, making AI integration a pivotal improvement in healthcare systems.

Following this, **Chapter 5** provides a comprehensive examination of the integration of AI with electroencephalographic technologies to advance neurological diagnostics and therapeutic interventions. This discussion extends the potential of AI beyond traditional diagnostics to include therapeutic applications, highlighting the critical role of AI in advancing medical technology.

While, **Chapter 6**, discusses the broader implications and applications of AI in processing bioelectrical signals, extending beyond traditional healthcare settings to include aspects like performance enhancement in sports and rehabilitation.

The thesis concludes with **Chapter 7**, synthesizes the insights and advancements discussed throughout the thesis. It highlights significant contributions and outlines potential future research pathways that could stem from this work, emphasizing the transformative impact of AI in healthcare.

Each chapter, while independent in focus, contributes to a cohesive narrative about the transformative role of AI in healthcare, underlining the necessity of an interdisciplinary approach to harness the full potential of AI in clinical and medical research settings.

1.2 Research Contributions

The current section aims to offer a synthetic but comprehensive overview of the research contributions from this thesis, as organized into thematic chapters. For each of them, we briefly summarize the content, report on the related publications, and give complete details about the role of the Ph.D. candidate, **Paolo Sorino**, in such publications. Note that, in all the papers cited in this section, and as already indicated in the thesis preamble, Paolo Sorino contributed as **main author** or is the **corresponding author**.

1.2.1 Healthcare Assisted by Artificial Intelligence-based Techniques

Contributions

This chapter provides a detailed exploration of the application of AI-based techniques in healthcare, demonstrating their versatility and transformative potential in various settings.

It initially discusses how AI can analyze and understand complex dietary and nutritional data—illustrating methods for identifying dietary patterns and assessing their impact on health conditions such as diabetes and frailty in specific populations. Techniques such as unsupervised learning and neural networks are applied to extract meaningful information from large unstructured datasets, enabling personalized nutrition planning and intervention. Then, we move from diet analysis to the broader predictive applications of AI in health care. Developments in AI tools that improve predictive accuracy for various health conditions, including cognitive disorders and

chronic diseases such as multiple sclerosis, are outlined. Discussion includes innovative data security and reliability techniques in creating decision support systems, as well as advanced machine learning approaches. Finally, the critical role of explainability in AI applications in healthcare is highlighted. Solutions are detailed that not only perform complex analyses, but also provide understandable and usable information to ensure that healthcare providers and patients can make informed decisions based on AI-generated data and predictions. Throughout the chapter, the narrative is structured to highlight the empirical research supporting these applications, discussing both the technological advancements and the theoretical underpinnings that make AI a valuable tool in modern healthcare. The discussion is designed to be comprehensive, covering technical details while also considering practical implications, thus providing a holistic view of the current and potential uses of AI in improving healthcare outcomes.

Publications

The chapter covers the topics presented and explored in “Dietary Patterns Associated with Diabetes in an Older Population from Southern Italy, Using an Unsupervised Learning Approach” [283] published in the international journal on the science and technology of sensors (Sensors).; “An Artificial Neural Network Model to Assess Nutritional Factors Associated with Frailty in the Aging Population from Southern Italy“ [44], presented at IEEE International Conference on Systems, Man and Cybernetics (SMC2022); “Blockchain and AI to Build an Alzheimer’s Risk Calculator” [270], presented at 22nd International Conference, (ICWE 2022); "A Pareto-Optimality-Based Approach for Selecting the Best Machine Learning Models in Mild Cognitive Impairment Prediction" [273], presented at IEEE International Conference on Systems, Man and Cybernetics (SMC2023); "Time-to-Event Interpretable Machine Learning for Multiple Sclerosis Worsening Prediction: Results from iDPP@CLEF 2023" [172], presented at Conference and Labs of the Evaluation Forum; "Integrating eXplainable AI in Healthcare: A Web Application Framework for Advancing the One Health Paradig" [69], presented at the $3rd$ International Workshop on Web Applications for Life Sciences In conjunction with the $24th$ International Conference on Web Engineering (ICWE 2024)**(To appear)**.

Ph.D. Candidate’s Role

Paolo Sorino is the corresponding author of the papers [270, 273]

1.2.2 Enhancing Devices for Bioelectrical Signal Analysis with Artificial Intelligence

Contributions

The chapter deals a progressive exploration of how AI can be utilized with bioelectrical signal analysis across various healthcare and auxiliary applications, beginning with foundational concepts and advancing to more complex integrations. The narrative begins with an investigation into the use of deep learning techniques for emotion recognition through Brain-Computer Interfacess (BCIs). This segment focuses on developing systems that can interpret human emotions by analyzing brain signals, setting the stage for real-time applications that could significantly impact mental health therapies and enhance user interaction with technology. The discussion then transitions into the development of neurofeedback tools. These tools are designed to provide users with real-time insights into their emotional states, enabling them to actively manage and adjust their emotions through guided feedback from the system. This area is of growing interest for its potential in psychiatric treatments and improving overall emotional well-being. Further exploring the intersection of BCIs and emotional recognition, the chapter elaborates on innovative applications that integrate BCI with advanced AI technologies, like large language models. These applications are designed to facilitate emotional support conversations, creating responsive and empathetic interactions that can offer support in mental health contexts. The chapter also covers creating an emotion-driven sound synthesis system, which utilizes bioelectrical signals to control sound environments. This technology demonstrates potential uses in therapeutic settings, where tailored soundscapes could aid in relaxation and stress reduction, as well as in interactive media and artistic performances. Extending the application of BCI beyond emotional recognition, the narrative delves into their use in recognizing mental states for neurorehabilitation. This section highlights how integrating BCI with machine learning can enhance therapeutic strategies, particularly for patients recovering from neurological impairments, by providing customized treatments based on the users's current mental state. Exploration continues by looking at combining mental state recognition with other innovative technologies like cellular automata and biofeedback. This multifaceted approach aims to create more sophisticated systems that can adapt to and interact with users' mental states in various medical and recreational contexts. The focus then shifts to applying explainable AI in classifying heartbeat signals, an essential aspect of medical diagnostics. This approach ensures that the AIs decision-making process is transparent and interpretable, fostering trust and

understanding among medical professionals and patients. As the chapter progresses, it broadens the scope to general applications of AI in bioelectrical signal analysis beyond healthcare. It examines how biomechanical and bioelectrical data can be utilized in sports science to enhance training programs for athletes, such as fencers, by identifying skill levels and providing targeted training feedback. Additionally, the narrative explores creative applications, discussing systems that generate music based on neural networks and bioelectrical inputs. These systems highlight the potential for BCI to transform artistic creation, allowing artists to generate music or control musical instruments directly through brain activity. To conclude, we propose a dataset collected using low-electrode devices for emotion recognition. This dataset takes advantage of advances in AI technologies to analyse EEG signals with minimal hardware, with the aim of facilitating the development of a robust tool to understand and interpret human emotions efficiently.

The chapter emphasizes the broad and versatile implications of integrating AI with bioelectrical signal analysis. This timeline not only reflects the current state of research and application but also charts a course for future innovations that could transform multiple aspects of human life, from healthcare and therapy to sports and the arts.

Publications

The chapter covers the topics presented and explored in "Brain Computer Interface: Deep Learning Approach to Predict Human Emotion Recognition" [12], presented at IEEE International Conference on Systems, Man and Cybernetics (SMC2022); "Predicting Human Emotions using EEG-based Brain Computer Interface and Interpretable Machine Learning" [65], presented at the 3rd National CINI Conference on Artificial Intelligence Ital-IA2023; "Combining Mental States Recognition and Machine Learning for Neurorehabilitation" [65], IEEE International Conference on Systems, Man and Cybernetics (SMC2023); "Exploring the Mental State Intersection by Brain-Computer Interfaces, Cellular Automata, and Biofeedback" [62], presented at IEEE EUROCON 2023 - 20th International Conference on Smart Technologies; "Combining Biomechanical Features and Machine Learning Approaches to Identify Fencers Levels for Training Support" [15], published in the International Journal of Applied Sciences; "NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms" [60], published in Cognitive Computation Journal; "Towards a Neurofeedback Tool For Emotion Recognition Using Brain Computer Interface" [13], presented atthe $3rd$ International Workshop on Web Applications for Life Sciences In conjunction with the $24th$ International Conference on Web Engineering

(ICWE 2024)**(To appear)**. "ARIEL: Brain-Computer Interfaces Meet Large Language Models for Emotional Support Conversation" [271], presented at ACM WeBIUM24 - 1st Workshop on Wearable Devices and Brain-Computer Interfaces for User Modelling - In conjunction with the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2024); "EmoSynth Real Time Emotion-Driven Sound Texture Synthesis via Brain-Computer Interface" [63], presented at ACM WeBIUM24 - 1st Workshop on Wearable Devices and Brain-Computer Interfaces for User Modelling - In conjunction with the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2024). "An Explainable Machine Learning Approach for Heartbeat Classification Through Signal-Based Features" [272], presented at the IEEE International Conference on Systems, Man and Cybernetics (SMC2024) **(To appear)**. "NeuroSense: A Novel EEG Dataset Utilizing Low-Cost, Sparse Electrode Devices for Emotion Exploration" [64], published in IEEE Access Journal.

Ph.D. Candidate's Role

Paolo Sorino is the corresponding author of the papers [12, 15, 65, 271]

1.3 Bibliographical Notes

For the sake of completeness, in this last section, we report on the publications which have been listed in the Publications section in the preamble of this thesis, but have not been cited among the research contributions. Specifically, we categorize them into two groups, namely: (i) research contributions regarding the main topics of this thesis that cannot properly be considered as paper publications (e.g., workshops and tutorials); (ii) other publications whose topics are similar/related to those of this thesis and where, *in same cases*, the Ph.D. candidate (Paolo Sorino) has contributed.

Workshops and Tutorials

- "WeBIUM 1st Workshop on Wearable Devices and Brain-Computer Interfaces for User Modelling" [61].
- "Invited speaker to the Workshop: "The potential of enabling technologies in tailoringand adapting neuromotor and cognitive rehabilitation in children", IEEE BHI BSN 2022"

Different Approaches in Cybersecurity

- "CPU-side comparison for Key Agreement between Tree Parity Machines and standard Cryptographic Primitives" [169] accepted at the 12th International Conference on Dependable Systems, Services and Technologies (DESSERT).
- "Towards a Federated Intrusion Detection System based on Neuromorphic Computing" [170] Accepted at the 9th International Conference on Smart and Sustainable Technologies (SpliTech).

Contributions in Clinical Papers

- "FDF-DB: A Database of Traditional Fermented Dairy Foods and Their Associated Microbiota" [312] Accepted at international journal of human nutrition (Nutrients)
- "The Effect of Low Glycemic Index Mediterranean Diet and Combined Exercise Program on Metabolic-Associated Fatty Liver Disease: A Joint Modeling Approach" [67] Accepted at international journal of clinical medicine

Chapter 2

Background on intelligent diagnostics for fast screening

The integration of Big Data and AI techniques in the development of intelligent diagnostic tools represents a pivotal evolution in the healthcare sector. These technologies are central to creating advanced decision support systems that assist physicians in making more accurate and faster diagnostic decisions. This background section will focus on how big data and AI have come together to revolutionise diagnostics, emphasising their role in supporting rapid screening and decision-making processes [111, 179, 182].

Big Data in Healthcare

Big data in healthcare refers to the extensive volumes of data generated from numerous sources such as Electronic Health Recordss (EHRs), genomic sequencing, wearables, imaging, and patient-reported data [71]. The sheer volume, velocity, and variety of this data surpass traditional data management tools and require advanced analytical approaches to derive meaningful insights.

In particular, the role of Big Data in this scenario are related to:

- **Comprehensive Insights:** Big data allows for the aggregation of information from multiple sources, providing a holistic view of a patient's health status.
- **Population Health Management:** Analysis of large datasets can reveal trends, predict outbreaks, and identify risk factors on a population level, which is invaluable for public health planning and intervention.
- **Personalized Medicine:** Big data enables the analysis of patient-specific factors such as genetics and lifestyle, which can inform personalized treatment plans.

Artificial Intelligence Techniques

AI in healthcare uses algorithms and software to approximate human cognition in the analysis of complex medical data. The primary goal of AI applications in healthcare is to analyze relationships between treatment techniques and patient outcomes. AI techniques commonly used include Machine Learning (ML) [181], Deep Learning (DL) [263], and Natural Language Processing (NLP) [168].

Specifically, it is possible to analyse individually the main different AI techniques commonly used in healthcare:

- **Machine Learning:** ML algorithms learn from data and can improve their accuracy over time without being explicitly programmed. In diagnostics, ML can be used to identify patterns in data that are indicative of specific health conditions.
- **Deep Learning:** A subset of ML, DL uses neural networks with many layers (hence "deep") to analyze various details of the data. It is particularly effective in image recognition, which is crucial for diagnosing diseases from imaging scans like X-rays, MRIs, OCT-Scans, etc.
- **Natural Language Processing:** NLP helps interpret and categorize text data in patient records, which is essential for extracting meaningful patient information from unstructured data.
- **Reinforcement Learning:** Reinforcement Learning (RL) focuses on training agents to make a sequence of decisions by rewarding them for correct actions and penalizing incorrect ones. In healthcare, RL has been applied to optimize treatment protocols, manage chronic diseases, and personalize therapies.
- **Computer Vision:** Computer Vision (CV) enables the analysis of visual data, such as medical images, to assist in detecting abnormalities, segmenting regions of interest, and tracking changes over time. Applications include tumor detection, retinal analysis, and monitoring patient activity.

Artificial Intelligence-driven Decision Support Systems

AI-driven Decision support systems (DSS) in healthcare provide clinicians with incredibly powerful tools for fast screening and diagnosis [14, 82, 111, 126]. These systems analyze the data provided by big data technologies and offer prediction based on patterns that may not be apparent to human observers. The benefits and challenges of these tools for physicians are different and listed below

Benefits to Physicians:

- **Speed and Efficiency:** AI can process and analyze data much faster than humans, reducing the time from screening to diagnosis.
- **Accuracy and Precision:** AI systems reduce human error and increase the precision of diagnoses, particularly in complex cases where multiple conditions may exist.
- **Predictive Capabilities:** AI can predict disease progression and outcomes, aiding physicians in choosing the best course of treatment.
- **Interpretability:** eXplainable Artificial Intelligence (XAI) systems can provide physicians with feature contributions during the prediction, assisting them in understanding the decision-making process.

Challenges in Healthcare:

- **Lack of Trust:** Physicians and patients may be reluctant to rely on AI systems due to concerns about reliability, transparency, and accountability.
- **Data Imbalance:** Medical datasets often suffer from imbalances, with rare conditions underrepresented, leading to biased predictions.
- **Data Privacy and Security:** Strict regulations like GDPR and HIPAA make accessing and sharing healthcare data challenging, hindering AI training and deployment.
- **Integration with Clinical Workflows:** Adapting AI systems to existing workflows can be difficult, requiring significant adjustments and training for healthcare professionals.
- **Ethical Concerns:** Issues such as bias in algorithms, fairness, and equitable access to AI-driven solutions remain key challenges.

The convergence of big data and AI techniques is transforming the landscape of diagnostic medicine by enabling the development of sophisticated decision support systems. These tools not only enhance the efficiency of diagnostic processes but also support physicians in making informed, data-driven decisions quickly. As these technologies continue to evolve, they hold the promise of significantly improving patient outcomes and optimizing healthcare delivery. The content of this chapter is inspired by [275] and other papers we will cite in the following.

2.1 Preliminaries

In healthcare, AI models often function as mappings from input features to predictive outputs. Let X represent the input features (e.g., patient data), and let Y be the predictions (e.g., diagnosis probabilities). The AI model can be described by a function:

$$f : X \to Y$$

where f is parameterized by θ and trained to approximate the true function f^* that maps inputs to outputs.

The model is trained by optimizing the parameters θ to minimize a loss function L, which quantifies the difference between predicted outputs $\hat{y}$ and actual outputs y over a dataset D:

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

where N is the number of training samples, and (x_i, y_i) are the feature vector and label for the i-th sample.

The trained model uses its learned parameters θ for predicting outputs on new, unseen data:

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

Evaluation Metrics

The metrics used to evaluate classifications models are the following.

The first parameter considered to evaluate the performance of machine learning model was the accuracy, defined as [28]:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of prediction}}$$

More specifically, the accuracy of a model is calculated with the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative. The Recall metric measures the ratio of correct positive classifications among

the total number of positive samples:

$$Recall = \frac{TP}{TP + FN}$$

The Precision measures the ratio of correct positive classifications among the total positive classifications:

$$Precision = \frac{TP}{TP + FP}$$

The F1 score is the harmonic mean between recall and accuracy:

$$F1 = 2 \cdot \frac{\text{Precision } \cdot \text{ Recall}}{\text{Precision } + \text{ Recall}}$$

The F1 score combines precision and recall into a single metric. The Area Under the Receiver Operating Characteristic Curve (AUC) is a metric that measures the capability of a classifier to separate the positive class from the negative one. It is formulated as follows:

$$AUC = \frac{\sum_{x^- \epsilon X^-} \sum_{x+ \epsilon X+} \left(1\left(f\left(x^-\right) < f\left(x^+\right)\right)\right)}{X^- + X^+}$$

where $1(\cdot) = 1$ if $f\left(x^-\right) < f\left(x^+\right)$else $1(\cdot) = 0$ and, $X^+$is the set of positive samples, $X^-$is the set of negative samples, $f(\cdot)$ is the result of model prediction, and $1(\cdot)$ an indicator function [40].

The metrics described above are essential for evaluating the performance of classification models, where the objective is to correctly distinguish between different classes. However, for regression models, where the objective is to predict a continuous value, different metrics are needed to measure the accuracy of the predictions against actual values. Below, we describe the most commonly used metrics for evaluating regression models:

- **Mean Squared Error (MSE)**

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

- **Root Mean Squared Error (RMSE)**

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

- **Mean Absolute Error (MAE)**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **R-Squared (R^2)**

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

  where $\bar{y}$ is the mean of the observed data.

- **Adjusted R-Squared (R^2_{adj})**

$$R^2_{\text{adj}} = 1 - \left(\frac{1 - R^2}{n - k - 1}\right)(n - 1)$$

  where n is the number of observations and k is the number of predictors.

- **Mean Absolute Percentage Error (MAPE)**

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

In the context of healthcare, it is also crucial to understand the model's predictions for acceptance and ethical reasons. For this reason, it is important to define various XAI approaches most commonly used.

LIME (Local Interpretable Model-agnostic Explanations)

Local Interpretable Model-Agnostic Explanations (LIME) is a technique that explains the predictions of any classifier in an interpretable and faithful manner, by approximating it locally with an interpretable model[233]. The predictions are explained as follows:

$$\xi(x) \approx f(x;\theta)$$

where $\xi(x)$ is a simple model, such as a linear regression, trained to approximate the predictions of the complex model f around the vicinity of a given instance x. The local model is generated by perturbing x and obtaining the predictions of f on these new samples, then weighting these samples according to their proximity to x.

Feature Importance

Feature importance is a global explanation method where the importance of each input feature is evaluated in terms of how much they contribute to the predictive model's accuracy [177, 242]. This can be quantitatively measured in several ways, such as the decrease in model performance when a feature's information is shuffled (permutation importance), or the increase in prediction error when a feature is removed:

$$\text{Importance}(j) = \text{Error}_{\text{model with all features}} - \text{Error}_{\text{model without feature } j}$$

where Importance(j) measures the increase in prediction error caused by omitting feature j from the model. This metric provides a sense of how crucial a feature is for the model's performance.

SHAP (SHapley Additive exPlanations)

Shapley Additive exPlanations (SHAP) values explain the prediction of an instance by computing the contribution of each feature to the prediction [178].

The prediction can be expressed as:

$$\hat{y} = \phi_0 + \sum_{j=1}^{M} \phi_j$$

where M is the number of features, and ϕ_j represents the contribution of the j-th feature to the prediction. These contributions are computed using Shapley values from cooperative game theory, defined as:

$$\phi_j = \sum_{S \subseteq M \setminus \{j\}} \frac{|S|!(M - |S| - 1)!}{M!} \left(v(S \cup \{j\}) - v(S)\right)$$

Here, $v(S)$ is the value function that outputs the prediction of the model using the subset of features S.

The mathematical formalism of AI in healthcare, especially with the integration of SHAP for XAI, bridges the gap between technical accuracy and practical transparency, enhancing trust and reliability in AI-driven diagnostics.

2.1.1 Artificial Intelligence-Driven Diagnostics: Tools and Technologies

The application of AI in healthcare diagnostics incorporates a variety of sophisticated tools and technologies, each designed to enhance the accuracy and efficiency of medical diagnostics [112, 145, 157]. These AI systems are built on the foundations discussed in the Preliminaries, where models trained on extensive datasets help predict and diagnose diseases with remarkable precision.

The core technologies in AI-driven diagnostics include:

- **Machine Learning and Deep Learning Models:** Utilized for pattern recognition and predictive analytics, these models can interpret complex datasets and extract clinically relevant insights that assist in decision-making [39, 149].
- **Cloud Computing:** Facilitates the storage and processing of large volumes of healthcare data, providing the necessary computational power to support sophisticated AI algorithms and real-time data analytics [8, 108].
- **IoT Devices:** Wearables and other connected devices collect continuous health data from patients, which are then processed using AI to monitor conditions and detect anomalies in real-time [225, 289, 291].
- **Data Analytics Platforms:** These platforms integrate data from various sources and use AI to provide comprehensive analytics, which is crucial for diagnosing complex cases and planning treatment strategies [2, 174, 282].

The implementation of these tools and technologies in AI-driven diagnostics not only speeds up the process of medical evaluation but also enhances the granularity and accuracy of the diagnostics. This is crucial in all areas of healthcare, particularly in emergency care, where rapid response and accurate diagnostics are paramount. By seamlessly connecting data collection, model application, and result interpretation, AI-driven tools and technologies lay the groundwork for more informed and timely medical decisions.

2.1.2 Real-Time Data Processing and Its Importance in Emergency Care

Real-time data processing in healthcare involves the continuous input, processing, and analysis of patient data to provide immediate insights that are crucial in emergency

care settings [55, 216]. This capability is particularly important in scenarios where time-sensitive decisions can significantly impact patient outcomes.

$$\text{Real-time AI System:} \quad R(t) = f(\text{data}(t); \theta)$$

Where $R(t)$ represents the real-time output or decision at time t, and $\text{data}(t)$ is the live streaming patient data at time t. The function f, parameterized by θ, is designed to operate under constraints of low latency and high reliability, often using lightweight models that are optimized for speed without sacrificing accuracy.

The implementation of real-time data processing systems in emergency care can be transformative by:

- **Accelerating Diagnosis:** Real-time systems can quickly analyze critical data, such as vital signs or ECG readings, to identify life-threatening conditions like myocardial infarctions or strokes. By speeding up the diagnosis, treatment can be initiated more promptly, which is often crucial for patient survival [5, 200, 281].
- **Monitoring and Alerts:** Continuous monitoring of patient vitals and automated alert systems help in early detection of patient deterioration, allowing healthcare staff to intervene before the condition worsens. AI-driven systems can predict severe events before they occur, enabling proactive care [137, 304].
- **Resource Optimization:** In emergency departments, where resources are often limited and time is critical, real-time AI systems can aid in prioritizing patient care based on severity and predicted medical needs. This ensures that the available medical staff and equipment are used as efficiently as possible [90, 124].

These systems rely on advanced data acquisition devices, high-speed computing platforms, and robust AI models that are capable of functioning accurately and efficiently in high-pressure environments. The integration of these technologies in emergency care represents a critical advancement in medical diagnostics and patient management.

2.2 Artificial Intelligence as a Driving Tool in Decision-Making Processes

AI usually applies to computational technologies that emulate mechanisms assisted by human intelligence, such as reasoning, deep learning, adaptation, and engagement [280].

Some devices can play a role that typically involves human interpretation and decision making [210]. These techniques have an interdisciplinary approach and can be widely applied to medicine and health. The literature recognizes several application areas of AI in medicine, however, two of the main ones can be defined as follows: (i) Clinical Decision Support Systems (CDSS), (ii) Role of AI in Enhancing Diagnostic Accuracy. Below, we present each of these macro areas of application outlining their main application concepts, along with some of the most popular state-of-the-art techniques.

2.2.1 Clinical Decision Support Systems (CDSS)

Clinical Decision Support Systems (CDSS) [262] are crucial in healthcare, crafted to assist clinicians in making well-informed decisions through robust, evidence-based support. Traditional CDSS harness guidelines and protocols derived from extensive clinical research and expert consensus, enabling healthcare providers to diagnose accurately and manage patient care effectively. The integration of CDSS into healthcare routines aims to improve patient outcomes, reduce errors, and maintain a consistently high standard of care across various medical disciplines. Furthermore, the infusion of AI into CDSS transforms these systems from static repositories of knowledge into dynamic tools capable of learning and adapting. AI enhances CDSS by enabling the analysis of large datasets beyond human capability, providing real-time insights, and offering predictive analytics that anticipate patient risks and outcomes before they manifest, significantly boosting the efficacy and responsiveness of medical interventions [210, 258, 262].

Overview of CDSS in Healthcare

Nowadays, hospitals have access to vast amounts of data. The application of AI techniques has significantly increased interest in CDSS. These systems are increasingly employed across various medical domains, including medical imaging, epidemiology, and triage operations in emergency departments, providing essential decision support to physicians [84, 276]. CDSS can be categorized into two main types based on the AI methodologies they employ: knowledge-based AI and data-driven AI.

Knowledge-based AI models human knowledge in computational terms, typically involving a top-down methodology where human experts describe the concepts and knowledge used to solve problems in a specific domain. Techniques in this category include rule-based systems, ontologies, semantic networks, and case-based reasoning. These systems provide explicit reasoning paths, making the decision process transparent

and explainable, but often face scalability issues due to the extensive effort required for knowledge acquisition and formalization [196].

Data-driven AI, on the other hand, starts with large datasets and uses statistical and machine learning methods to identify patterns and make predictions. Operating in a bottom-up manner, data-driven AI abstracts patterns from data to emulate human decision-making processes. Common techniques include neural networks, support vector machines (SVMs), decision trees, and ensemble learning methods. Data-driven AI excels in handling large volumes of data, such as electronic health records (EHRs), and can automatically build models from data with high accuracy, though these systems often lack transparency and explainability [51][262].

A promising trend in CDSS is the integration of both knowledge-based and data-driven approaches, creating hybrid systems that leverage the strengths of both methodologies. Hybrid systems can use formalized domain knowledge to structure the model and enhance interpretability, while also utilizing large datasets to improve predictive accuracy and handle complex patterns. Examples include probabilistic graphical models like Bayesian networks and case-based reasoning systems that integrate structured knowledge with data-driven methods for retrieval and adaptation [196].

Key challenges for CDSS include ensuring transparency and explainability, integrating diverse data sources for effective interoperability, and addressing scalability issues. The field is evolving towards the adoption of hybrid approaches that combine the robustness of knowledge-based AI with the flexibility and accuracy of data-driven AI, aiming to provide robust, scalable, and explainable decision support systems. Future research will likely focus on enhancing transparency and ensuring seamless integration of diverse data sources to support comprehensive clinical decision-making [220, 262].

Integration of Artificial Intelligence into CDSS: Enhancing Accuracy and Speed

The integration of AI into CDSS has notably enhanced both the accuracy and speed of medical diagnostics and decision-making processes. AI-driven CDSS leverage complex algorithms to process and analyze vast amounts of healthcare data rapidly, leading to quicker and more accurate patient assessments. This acceleration is critical in environments where time is of the essence, such as in emergency departments or during acute care scenarios.

Furthermore, AI enhances the precision of CDSS by employing advanced data analytics techniques, such as machine learning and deep learning, which can identify patterns and anomalies that may be invisible to the human eye. For instance, in

medical imaging, AI algorithms can detect subtle changes in imaging data that signify early stages of diseases such as cancer or neurological disorders, often before clinical symptoms are apparent [120, 300].

AI's capability to continuously learn from new data also means that CDSS can become more refined and accurate over time, adapting to new medical findings and evolving disease patterns without the need for explicit reprogramming. This adaptability ensures that CDSS remain at the forefront of medical technology, providing clinicians with the most current and relevant decision-support tools.

The use of AI in CDSS also addresses previous limitations related to the scalability of knowledge-based systems. By automating the data analysis process, AI allows CDSS to handle larger datasets efficiently, thereby supporting broader applications across various medical specialties and increasing the system's overall utility and impact [84, 294].

In summary, the integration of AI into CDSS significantly enhances the systems' ability to deliver rapid, accurate, and high-quality medical decision support, transforming how healthcare providers diagnose and manage patient care, and underscoring the indispensable role of AI in modern healthcare systems [210, 258]

2.2.2 Role of Artificial Intelligence in Enhancing Diagnostic Accuracy

AI significantly enhances diagnostic precision across various medical fields. By integrating deep learning algorithms, AI systems can analyze medical images, genetic data, and biomarkers with higher accuracy and granularity than traditional methods. This capability allows for the detection of minute abnormalities that often elude human observers, leading to earlier and more precise diagnoses. For example, AI applications in dermatology have outperformed seasoned dermatologists in identifying skin cancers from images, showcasing AI's superior pattern recognition capabilities [79, 147].

How Artificial Intelligence Contributes to Finer Diagnostic Precision. AI's strength in handling complex and voluminous data sets comes into play notably in the interpretation of complex medical data. In genomics, AI helps in mapping out genetic mutations and predicting their implications on diseases, thus facilitating personalized medicine. Similarly, in neurology, AI tools analyze brain imaging scans to detect early signs of disorders such as Alzheimer's or Parkinson's disease, often before clinical symptoms manifest. These applications underscore AI's pivotal role in leveraging complex datasets to provide actionable insights, thereby supporting clinicians in making

informed decisions [16, 244, 303].

Artificial Intelligence in the Interpretation of Complex Medical Data. The integration of AI into diagnostic workflows has revolutionized the speed and efficiency of patient care delivery. In emergency settings, AI-driven tools quickly assess patient data to prioritize care based on severity and immediate needs. For instance, AI systems in triage analyze symptoms and vital signs to recommend immediate interventions or further testing. This not only speeds up the diagnostic process but also ensures that critical resources are allocated efficiently, enhancing overall healthcare delivery. Moreover, AI integration helps in continuous monitoring of hospitalized patients, detecting potential complications before they become critical [206, 229].

2.3 Explainable Artificial Intelligence for Understanding Model Diagnoses

Introduction to XAI and its Necessity in Healthcare. XAI pertains to systems and methodologies developed within the artificial intelligence framework that are structured to be interpretable by humans [80]. This capability is critical in healthcare, a field where decisions have profound implications on patient outcomes. XAI demystifies the AI decision-making process, making it transparent and understandable for practitioners [171, 201]. This transparency is not just a technical requirement but a clinical necessity, ensuring that AI-driven decisions align with ethical medical practices and are grounded in understandable, logical reasoning [96]. As AI systems become more prevalent in diagnostic and treatment pathways, their explainability will be essential for clinicians to trust and effectively integrate these technologies into patient care.

Benefits of XAI in Patient and Practitioner Trust. The integration of XAI in healthcare significantly fosters trust, a cornerstone in patient-practitioner relationships. Explainable models contribute to this by providing insights into their analytical processes, justifying their recommendations in a scientifically sound manner [154]. For practitioners, this means being able to verify AI-generated advice against their clinical knowledge and ethical considerations, enhancing confidence in utilizing AI tools as part of treatment plans. For patients, understanding how decisions about their health are made can improve satisfaction and compliance with treatment, essential components of successful healthcare outcomes [4].

2.3.1 Explainable Decision Support Systems

Designing Explainable Artificial Intelligence for Transparency in Clinical Decisions. In the realm of clinical decision-making, transparency is not merely beneficial; it is imperative. AI systems designed with explainability at their core help in elucidating complex decision-making processes, thus providing clinicians with a clear understanding of how conclusions were reached [254]. This level of transparency is crucial for compliance with medical regulations and for maintaining an audit trail that justifies decision-making processes in patient care. It also plays a pivotal role in interdisciplinary collaborations where understanding the rationale behind AI decisions can enhance teamwork and integrated care strategies [211].

Case Examples where Explainable Artificial Intelligence has Transformed Decision-making. Several instances demonstrate how XAI has significantly improved clinical decision-making. For instance, in radiology, XAI tools assist radiologists by highlighting features in imaging scans that contribute to diagnoses, thus acting as a second, highly reliable opinion that aids in early disease detection. Another notable example is in oncology, where XAI provides rationales for recommending specific chemotherapy regimens based on patient-specific data, historical treatment outcomes, and ongoing research. These case examples not only illustrate the practical benefits of XAI but also showcase its potential to significantly improve diagnostic accuracy and patient outcomes [4, 296].

Future Trends in Explainable Systems within Medical Diagnostics. Looking forward, the potential of XAI in healthcare is vast and varied. As medical data becomes increasingly complex and voluminous, XAI will be crucial in managing this complexity and ensuring that AI-driven diagnostics remain both high in quality and understandable. Future developments may see XAI facilitating real-time diagnostics in emergency scenarios, integrating multimodal data for comprehensive patient assessments, and enhancing personalization in treatment planning. These advancements will not only refine diagnostic processes but also make them more patient-centric [123, 247, 298].

2.3.2 Implementation Challenges and Solutions

Identifying Common Hurdles in Deploying explainable artificial intelligence in Healthcare. The deployment of XAI in healthcare environments encounters numerous challenges. These range from technical issues, such as data integration and system complexity, to human factors, including resistance from healthcare staff and a steep learning curve associated with new technologies. Moreover, ensuring the reliability

and robustness of XAI systems in diverse clinical settings remains a significant hurdle [248].

Strategies for Overcoming these Challenges. To address these challenges, a multi-faceted approach is necessary. This includes rigorous training programs for healthcare staff to increase familiarity with AI technologies, enhancements in user interface design for XAI systems to improve usability, and robust data governance frameworks that ensure the integrity and security of patient data. Additionally, ongoing collaboration between AI developers, healthcare professionals, and regulatory bodies is crucial to align system capabilities with clinical needs and compliance requirements.

Ethical Considerations and Managing Patient Data. The ethical dimensions of deploying AI and XAI in healthcare are complex and critical. They encompass concerns about patient privacy, consent processes for the use of personal medical data, and the potential biases inherent in AI algorithms. Managing these ethical considerations requires transparent practices, rigorous system testing to uncover and mitigate biases, and clear communication with patients about how their data is used and protected. These measures are essential to maintain trust and uphold the ethical standards expected in healthcare [3, 208, 267].

Chapter 3

Advanced Bioelectric Signal Analysis in Healthcare

The healthcare domain has undergone substantial advancements driven by the integration of modern technologies into various diagnostic and therapeutic fields. Among these technologies, the analysis of bioelectric signals has emerged as one of the most vital areas, playing a critical role in medical diagnostics, patient monitoring, and treatment. Bioelectric signals, including EEG [26], EMG [59], and ECG [100], provide rich information about the physiological activities of the brain, muscles, and heart, respectively. These signals form the backbone of numerous clinical applications, allowing for the assessment of neurological, muscular, and cardiac functions.

This chapter delves into the foundational concepts of bioelectric signals, offering an overview of how these signals are generated and utilized in clinical practice. It further explores the technological advancements that have significantly improved the acquisition and processing of bioelectric signals, emphasizing the emerging role of AI in transforming bioelectric signal analysis. By integrating ML, DL, and XAI techniques, the field has seen significant improvements in refining signal processing workflows, automating feature extraction, and enhancing diagnostic accuracies.

The deployment of AI-based methods has enabled more precise and personalized assessments, making bioelectric signals an indispensable tool in modern healthcare. Through these advancements, healthcare providers can make informed decisions that ultimately lead to better patient outcomes and more efficient healthcare systems. Moreover, examines the importance of in bioelectric signal analysis.

3.1 Foundational Concepts of Bioelectric Signals

Bioelectric signals are electrical manifestations of the physiological activities of cells and tissues within the body, primarily generated by nerve and muscle cells. These signals, such as EEG, EMG, and ECG, play a vital role in healthcare for diagnosing, monitoring, and treating various medical conditions. By providing a real-time window into the functioning of critical systems like the brain, muscles, and heart, bioelectric signals enable healthcare professionals to assess organ function, detect abnormalities, and guide therapeutic interventions.

The importance of bioelectric signals lies in their non-invasive nature and ability to offer early detection of physiological changes. For instance, EEG is essential for identifying seizures, EMG is fundamental for diagnosing neuromuscular disorders, and ECG is the gold standard for monitoring heart health and detecting arrhythmias. Advancements in technology have further extended their use from routine check-ups to critical care settings, enabling continuous monitoring and timely intervention.

Bioelectric signals also significantly impact patient management and rehabilitation by assessing treatment effectiveness and tailoring personalized rehabilitation programs. The integration of these signals into wearable technologies facilitates long-term monitoring in everyday environments, thus supporting personalized healthcare and leading to improved patient outcomes.

Given the complexity and variability of bioelectric signals, accurate interpretation requires sophisticated tools, including preprocessing techniques and advanced analytical models. These methodologies enhance signal quality, remove artifacts, and extract meaningful features, transforming raw data into actionable clinical insights that support decision-making and optimize patient care.

3.1.1 Types and Roles of Bioelectric Signals

Bioelectric signals are essential in monitoring the physiological state of different organs, helping detect abnormalities, guiding interventions, and evaluating treatment efficacy. Here, we discuss the different types of bioelectric signals, each playing a specific role in healthcare.

Electroencephalogram (EEG)

The EEG is a non-invasive method to measure the electrical activity of the brain. It captures spontaneous electrical signals generated by the synchronized activity of large

populations of neurons in the cortex. EEG is represented mathematically as:

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

where A_i, f_i, and ϕ_i denote the amplitude, frequency, and phase of the i-th component of the signal, reflecting the rhythmic activity of cortical neurons. EEG is extensively used to diagnose epilepsy, sleep disorders, and brain function in coma patients, as well as to monitor cognitive activity in neurological conditions. The ability to record the brain's activity with high temporal resolution makes EEG an essential tool in neuroscience research.

Electromyogram (EMG)

EMG records the electrical activity produced by skeletal muscles. It is critical for understanding the activation and coordination of muscles during movement, and it is particularly useful for diagnosing neuromuscular diseases, monitoring rehabilitation progress, and controlling prosthetic devices. EMG signals are modeled as a stochastic process, given the random nature of muscle fiber activation:

$$EMG(t) = \sum_{n} a_n \delta(t - t_n)$$

where a_n and t_n represent the amplitude and firing time of the n-th motor unit action potential, respectively. By analyzing EMG data, clinicians can assess muscle health and function, detect myopathies, and evaluate muscle fatigue.

Electrocardiogram (ECG)

The ECG is a widely used technique to measure the electrical activity of the heart. It provides crucial information regarding cardiac rhythm, heart rate variability, and possible abnormalities like arrhythmias, ischemia, or hypertrophy. The ECG signal can be approximated by a series of Gaussian functions representing different components of the heart's electrical cycle:

$$ECG(t) = \sum_{j} A_j e^{-\frac{(t-\mu_j)^2}{2\sigma_j^2}}$$

where A_j, μ_j, and σ_j correspond to the amplitude, position, and width of the j-th wave component, representing the P wave, QRS complex, and T wave. The non-invasive

nature and diagnostic capabilities of ECG make it a cornerstone of cardiovascular assessment, with applications in both acute and chronic disease monitoring.

3.1.2 Technological Advancements in Signal Acquisition

The acquisition of bioelectric signals has evolved significantly over recent years due to improvements in sensor technologies and data acquisition systems. These advancements have enabled more accurate, reliable, and user-friendly devices for clinical and research purposes.

Cutting-edge Technologies for Signal Capture

Advances in sensor technology and instrumentation have led to higher fidelity in bioelectric signal recording. For EEG, high-density electrode arrays, often with 64 or more channels, have improved spatial resolution, allowing more detailed monitoring of cortical activity. Advances in BCI technologies have further pushed the frontier of EEG acquisition, enabling communication and control for individuals with severe motor impairments.

Wearable devices have also played a transformative role in signal acquisition. Portable ECG monitors, equipped with wireless data transfer capabilities, provide continuous, real-time data collection, which is crucial for long-term monitoring of cardiac health. These devices offer a convenient and non-invasive way for patients to track their heart health while maintaining their daily routines, thus improving compliance and overall health outcomes.

The development of dry electrodes has also significantly enhanced patient comfort. Traditional wet electrodes require a conductive gel, which can cause skin irritation and requires regular reapplication. Dry electrodes eliminate these issues, making it easier to obtain high-quality recordings in ambulatory settings without compromising signal integrity.

Data Processing and Analysis Techniques

The processing and analysis of bioelectric signals employ advanced mathematical and computational techniques to enhance the quality and interpretability of physiological data. This process involves a series of steps to convert raw bioelectric signals into actionable information for clinicians.

Preprocessing. The preprocessing stage is critical for ensuring the accuracy of bioelectric signal analysis. Signal acquisition is often contaminated by noise and

artifacts, such as muscle activity, power line interference, and eye blinks. Techniques like bandpass filtering, notch filtering, and Independent Component Analysis (ICA) are employed to remove unwanted artifacts while preserving the integrity of the signals. Proper preprocessing is essential, as it lays the foundation for accurate feature extraction and classification.

Feature Extraction and Signal Segmentation. Once noise and artifacts are removed, the next step is to extract meaningful features from the bioelectric signals. Feature extraction involves identifying significant components, such as frequency bands in EEG (e.g., delta, theta, alpha, beta, gamma), the time-domain features of ECG waves, or motor unit action potentials in EMG. Extracted features provide insights into the physiological state of the organ of interest. Signal segmentation techniques, which divide signals into meaningful epochs, are also employed to focus on specific events or activities within the data.

Advanced Analytical Techniques. Advanced analytical techniques, including wavelet transforms, Fourier transforms, and statistical analysis, help in understanding the underlying physiological patterns. These techniques allow researchers and clinicians to interpret complex bioelectric signals by converting them into a more comprehensible form, such as time-frequency maps or statistical distributions. Machine learning models can then be applied to classify or predict specific outcomes based on these features.

3.2 Artificial Intelligence in Bioelectric Signal Analysis

AI has revolutionized the analysis of bioelectric signals, pushing the boundaries far beyond what traditional signal processing methods could achieve. The integration of AI techniques, including ML and DL, has enhanced the precision, efficiency, and scope of bioelectric signal analysis, particularly for EEG, EMG, and ECG signals.

By leveraging AI algorithms, bioelectric signal analysis has evolved to achieve more sophisticated levels of interpretation, enabling a deeper understanding of the physiological processes behind these signals. The capabilities of AI to automatically extract complex patterns and features from large datasets have proven invaluable in managing the variability and intricacy of bioelectric data. This not only allows for more accurate diagnostics but also provides new opportunities for predictive modeling and personalized treatment approaches.

The transformative impact of AI on bioelectric signal analysis has led to significant advancements in clinical and research settings. These advancements have enabled real-

time monitoring of patient health, improved detection and classification of abnormalities, and supported the development of novel applications such as BCIs and advanced rehabilitation tools. Furthermore, the use of DL models has been particularly impactful due to their ability to learn directly from raw data, thereby eliminating the need for extensive feature engineering and reducing human bias in the analytical process.

3.2.1 Artificial Intelligence-enhanced Electroencephalography Analysis

The use of AI in EEG analysis has significantly advanced the understanding of the human brain and its various functional states. Machine learning, particularly DL, has enabled the identification of complex patterns in EEG data that were previously difficult to decipher using conventional methods.

Deep Learning for EEG Analysis. Deep learning methods, such as Convolutional Neural Networkss (CNNs) and Recurrent Neural Networkss (RNNs) [94, 240, 309], have been particularly useful for analyzing EEG signals due to their ability to handle large datasets and their robustness in feature extraction. CNNs are primarily used for spatial feature extraction, identifying distinct patterns in brain activity associated with different tasks or emotional states. On the other hand, RNNs and Long short-term memorys (LSTMs) [295] are highly effective in handling time-series data, making them ideal for capturing temporal dependencies in EEG signals for different tasks [31, 78, 102, 261].

Applications include emotion recognition, seizure detection, and sleep stage classification. For instance, LSTMs have been used in real-time emotion recognition systems to identify changes in emotional states, which can be applied in areas such as mental health monitoring and adaptive human-computer interaction [155, 295].

Real-time Monitoring and Brain-Computer Interfaces. AI-powered EEG systems are also at the forefront of BCI technology. These systems allow for direct communication between the brain and external devices, empowering individuals with physical disabilities to control prosthetic limbs or computer systems using their brain activity alone. AI algorithms are critical for accurately decoding the neural commands, thus making BCIs reliable and practical for real-world applications.

3.2.2 Artificial Intelligence Applications in Electromyography

AI techniques have also been employed extensively in the analysis of EMG signals [93], aiding both clinical and non-clinical applications, such as rehabilitation, prosthetics,

and human-machine interfaces [228, 232, 301].

Movement Intention Recognition. One of the key applications of AI in EMG analysis is movement intention recognition, which is critical for prosthetic limb control. Machine learning algorithms, such as Support Vector Machines (SVMs) and CNNs, are used to classify muscle signals associated with specific movements, thereby allowing amputees to control prosthetic devices more naturally [158, 292]. The accuracy of these models has significantly improved with the inclusion of DL, which can learn abstract features directly from raw signals, thus reducing the need for manual feature engineering.

Muscle Fatigue Monitoring and Rehabilitation. AI models are also being utilized for muscle fatigue detection, an important application in sports science and rehabilitation [70, 195, 212]. Fatigue analysis provides insights into muscle health and helps in designing personalized training and rehabilitation protocols. RNN are well-suited for capturing the temporal characteristics of muscle activation, which is key in detecting and monitoring the progression of muscle fatigue during exercise or rehabilitation sessions [93].

Human-Machine Interaction. Beyond prosthetics, AI-enhanced EMG systems are used in human-machine interaction scenarios, such as controlling robotic exoskeletons for rehabilitation purposes [23, 77, 104]. These systems help restore motor function in individuals who have lost it due to injury or neurological conditions. AI algorithms are used to interpret the user's muscle activation signals, translating them into commands that control the exoskeleton's movements, thus assisting in physical therapy and mobility.

3.2.3 Artificial Intelligence in Electrocardiogram Analysis

The integration of AI into ECG analysis has fundamentally altered the landscape of cardiovascular diagnostics, providing more accurate and real-time assessments of cardiac health.

Diagnostic Applications of Deep Learning. Deep learning algorithms, particularly CNNs, have shown significant potential in diagnosing various cardiac conditions by analyzing ECG data [19, 56, 191]. These models are capable of automatically extracting features related to arrhythmias, myocardial infarctions, and other cardiovascular disorders without the need for manual intervention [161]. The automation provided by AI is crucial in clinical settings, where rapid and accurate diagnosis can make a significant difference in patient outcomes.

Real-time Monitoring Systems. AI has enabled the development of real-time

ECG monitoring systems that are capable of continuously analyzing ECG signals and alerting healthcare providers to any potential abnormalities [224, 243, 251]. These systems are particularly valuable for patients with chronic cardiovascular conditions, as they allow for timely intervention and reduce the risk of severe cardiac events [257].

Personalized Cardiovascular Risk Assessment. Another important application of AI in ECG analysis is personalized risk assessment. By analyzing ECG data in combination with other patient-specific information, AI models can provide a comprehensive assessment of cardiovascular risk [135, 148]. Techniques such as ensemble learning, which combines multiple machine learning algorithms, have been used to improve the accuracy of cardiovascular risk prediction models, helping clinicians make better-informed decisions regarding preventive measures and treatment strategies [6].

3.3 Explainable Artificial Intelligence in Bioelectric Signals

XAI plays a crucial role in increasing transparency and building trust in bioelectric signal analysis. In healthcare, it is essential for clinicians to understand and trust the decision-making processes of AI models, especially when they are used in diagnosis and treatment planning.

Enhancing Transparency and Trust with XAI. XAI techniques help demystify the "black-box" nature of many AI models by providing insights into how decisions are made. Techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) [256], SHAP (Shapley Additive Explanations), and LIME (Local Interpretable Model-agnostic Explanations) are widely used to highlight important features in bioelectric signal data and to visually explain model predictions [198, 227]. These explanations can reveal which aspects of a signal contributed most to a particular diagnosis, providing clinicians with greater confidence in the system's recommendations. In the context of EEG, XAI can be used to determine which regions of the brain were most influential in detecting abnormal neural activity, while in ECG analysis, it can help in pinpointing specific anomalies in heart rhythms that indicate cardiac risk. This not only improves the interpretability of AI models but also supports better clinical decision-making.

XAI in Clinical Practice: Case Studies and Applications. Several case studies demonstrate the practical use of XAI in clinical settings, particularly in neurorehabilitation and cardiovascular disease management [165, 222, 227]. For example, Class Activation Mapping (CAM) has been used to highlight which regions of an ECG signal are indicative of certain conditions, making it easier for cardiologists to validate the AI

model's output [134, 198]. Similarly, Random Forest models have been used for EMG signal classification, with feature importance techniques providing valuable insights into the features that are most predictive of specific motor activities [85, 98, 227]. These applications of XAI not only improve the transparency of AI models but also foster greater acceptance among clinical experts, as they make it easier for practitioners to understand and trust the models' predictions. This is particularly important in fields like mental health, where patient outcomes depend heavily on the accuracy and transparency of diagnostic tools.

3.3.1 Prospects and Challenges of XAI in Bioelectric Signal Analysis

The future of XAI in bioelectric signal analysis is promising, with the potential to transform how clinicians interact with AI-based diagnostic tools [106, 133, 184]. Emerging techniques, such as causal inference and domain adaptation, are expected to further enhance the interpretability and robustness of AI models in healthcare [227].

However, significant challenges remain. Ensuring the generalizability of XAI methods across different datasets and clinical scenarios is critical for their widespread adoption. Bioelectric signals are often subject to variability based on individual differences, such as age, sex, and comorbid conditions, making it challenging to create one-size-fits-all explainable models. Additionally, incorporating clinician feedback during the development of XAI systems can improve their usability and relevance, leading to better patient care and outcomes [128].

Ethical considerations also play a role in the adoption of XAI. Issues such as bias, fairness, and accountability must be addressed to ensure responsible and equitable use of AI technologies in healthcare [32, 198]. XAI provides an opportunity to mitigate these concerns by making the decision-making process more transparent and understandable for all stakeholders involved.

In conclusion, XAI represents a transformative approach in the analysis of bioelectric signals, with the potential to significantly enhance diagnostic and therapeutic processes in healthcare. By making AI systems more transparent and trustworthy, XAI can contribute to more personalized, effective, and ethical healthcare solutions.

Chapter 4

Healthcare Assisted by Artificial Intelligence-based Techniques

4.1 Research Objectives and Contributions

The increasing complexity of modern health challenges, particularly among aging populations and individuals at risk of neurodegenerative diseases, calls for advanced analytical tools to enhance precision in diagnosis and care. AI has emerged as a transformative technology in addressing these challenges, offering new ways to process and interpret complex epidemiological data. This chapter explores the application of machine learning techniques in the health and epidemiology domains, focusing on identifying patterns in dietary behaviors, assessing factors linked to physical frailty, and predicting cognitive decline.

4.1.1 Research Objectives

Chapter 4 is dedicated to harnessing the capabilities of AI to address complex challenges in the domains of health and epidemiology. The chapter sets forth several key objectives aimed at leveraging the power of both supervised and unsupervised machine learning techniques. Firstly, it seeks to identify dietary patterns within an elderly diabetic population and to analyze nutritional factors that are associated with frailty. This involves a detailed examination of large datasets to discern patterns that may not be immediately apparent to human analysts.

Additionally, the chapter aims to demonstrate the efficacy of AI in enhancing healthcare diagnostics. This includes efforts to reduce delays in diagnostics, lower the costs associated with healthcare delivery, and facilitate the development of personalized

treatment plans tailored to the specific needs of individual patients. These objectives underscore the potential of AI to transform traditional approaches to healthcare and improve both efficiency and effectiveness.

Another significant objective of this chapter is to utilize advanced machine learning methods to probe epidemiological data with a particular focus on neurodegenerative outcomes, such as Mild Cognitive Impairment (MCI). This involves sophisticated modeling techniques that can predict disease progression and potentially offer insights into preventive measures.

Finally, the chapter evaluates the integration of AI technologies into clinical settings, assessing their impact on the decision-making processes and the overall delivery of healthcare services. This includes a critical analysis of how AI tools are currently being adopted within healthcare infrastructures and their effects on the workflows and outcomes of medical practices.

By addressing these objectives, Chapter 4 not only highlights the practical applications of AI in healthcare but also explores the broader implications of these technologies on the field of epidemiology and public health.

4.1.2 State of the Art Limitations

While AI offers transformative potential for healthcare, it faces several challenges that may limit its efficacy in current methodologies. These include issues related to the quality, privacy, and representativeness of data, which are significant hurdles affecting the reliability of AI predictions. The opaque nature of many AI models, particularly those involving deep learning, complicates their acceptance and use in clinical practice due to a lack of transparency and interpretability. Furthermore, technical and cultural barriers complicate the integration of AI systems into existing healthcare infrastructures, requiring considerable adaptation from current practices. Additionally, AI deployment must navigate a complex landscape of ethical considerations and regulatory requirements to ensure patient safety and ethical compliance.

4.1.3 Chapter Contributions

Building upon the identified objectives and the outlined limitations, this chapter contributes to the field of healthcare through the application of AI. We begin by utilizing machine learning to analyze dietary data from elderly diabetics, providing new insights into the correlations between diet and health. This analysis helps in

identifying dietary patterns that may influence the onset and progression of various health conditions.

Furthermore, the relationship between nutritional factors and physical frailty is examined, which suggests potential pathways for dietary interventions aimed at reducing frailty risks among the elderly. This is particularly significant as it opens up avenues for preventive healthcare measures that can significantly improve quality of life.

In addition, we focus on developing predictive models for MCI, aiming to enhance our understanding and anticipation of neurodegenerative diseases. These models are instrumental in proposing preventive health strategies by identifying early indicators of cognitive decline, thereby allowing for timely intervention.

These contributions are designed not only to address the challenges posed by the integration of AI in healthcare but also to leverage AI's capabilities to improve healthcare outcomes and decision-making processes significantly. By advancing AI applications in healthcare, we aim to demonstrate its value in enhancing diagnostic accuracy, optimizing treatment efficacy, and ultimately, improving patient care.

4.2 Background

The following section provides an overview of the existing research on the application of AI in two key areas relevant to this thesis: nutritional pattern analysis and cognitive impairment detection. These topics form the foundational background upon which the subsequent work presented in this chapter is built.

First, we review recent advances in the use of ML techniques for analyzing nutritional patterns, with a focus on dietary habits and their health implications. This includes the identification of distinct dietary patterns and the potential for personalized nutrition interventions.

Next, we explore AI-based methods for detecting cognitive impairment, specifically focusing on Alzheimer's disease (AD) and MCI. This includes various machine learning and deep learning approaches that have been applied to neuroimaging and electrophysiological data to enhance early detection and prediction capabilities.

Artificial Intelligence in Nutritional Patterns

In recent years, there has been a growing interest in applying ML techniques to nutrition research to manage the complexity and high dimensionality of nutritional data.

Kirk et al. [151] provide a comprehensive review of the applications of ML in nutrition, particularly in precision nutrition and metabolomics. They emphasize the importance of ML for understanding the multifaceted interactions between nutrition and health, identifying key areas such as obesity and cardiovascular diseases where ML has shown significant promise.

Silva *et al.* [264] used clustering algorithms and ML to analyze dietary patterns in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Their study identified two primary patterns: a "Western" pattern characterized by high consumption of red meat and processed foods, and a "Prudent" pattern with higher intakes of fruits, vegetables, and whole grains. Algorithms like K-means, SVM, and random forest demonstrated moderate accuracy ($69\% - 72\%$) in predicting these patterns based on socio-demographic and clinical data.

In a recent systematic review, Armand *et al.* [285] explored the applications of AI in nutrition, highlighting the potential of ML and DL in areas such as personalized nutrition, food recognition, and dietary monitoring. The review emphasizes how these techniques can detect complex patterns within nutritional datasets and provide predictive models for disease prevention and monitoring, thereby optimizing dietary recommendations.

AI-based Methods in Cognitive Impairment Detection

AD, a type of dementia primarily affecting the elderly, is prevalent in society. Detecting AD in its early stages is challenging for medical professionals, with no entirely reliable biomarker available. Unfortunately, AD remains incurable, and clinical trials for AD drugs have largely been unsuccessful. Many DL and supervised classification algorithms have been developed for AD detection, but these methods are still imperfect and unable to achieve high accuracy.

MCI, a progressive neurological condition often preceding AD, leads to cognitive decline in older individuals. Early detection is crucial to implementing treatment approaches that can improve quality of life for patients with MCI, although no current remedies can halt or reverse its progression.

This section discusses studies on the use of Machine Learning and DL techniques for analyzing MCI. Ansart *et al.* [10] highlighted the importance of following guidelines for using ML as a clinical decision-making tool, focusing on predicting the progression of individuals with MCI. The review of 234 experiments from 111 articles raised doubts about the effectiveness of Magnetic Resonance Imaging (MRI) alone for predicting the

advancement of MCI towards dementia, with studies using cognitive variables or FDG PET reporting better results.

Grueso *et al.* [110] conducted a comprehensive evaluation of 452 studies that applied ML to neuroimaging data to assess whether MCI patients were likely to develop AD. The findings showed that combining MRI and PET led to higher classification accuracy compared to using only one imaging technique. SVM was the most commonly used algorithm, achieving an average accuracy of 75.4%, while CNN performed slightly better with an accuracy of 78.5%.

Alvi *et al.* [9] proposed a DL approach involving a Gated Recurrent Unit (GRU) model for detecting MCI using EEG data. Their model outperformed other classifiers, including LSTM, SVM, and K-Nearest Neighbors (KNN), with high accuracy when distinguishing between MCI participants and Healthy Control (HC) subjects, validated using 5-fold cross-validation.

Forouzannezhad *et al.* [87] worked on distinguishing between Normal Controls (CN) and early MCI (EMCI) groups to facilitate early diagnosis. They proposed a ML approach combining MRI, PET, and neuropsychological test scores, achieving an accuracy of 84.0% in differentiating CN from EMCI. Their method demonstrated potential for multiclass classification, including CN, late MCI, and AD.

In summary, these studies illustrate the current advances and challenges in using ML and DL techniques for MCI detection and AD prediction, emphasizing the need for improved methods to enhance diagnostic accuracy.

These related works highlight both the opportunities and challenges in the application of AI to health-related data, providing a comprehensive background that underpins the novel approaches discussed later in this chapter.

4.3 Exploring Dietary and Nutritional Patterns in Diabetic Population

With regard to work related to the diabetes mellitus (DM) population it is important to define that DM is a group of metabolic diseases characterized by an increase in blood glucose concentrations (hyperglycemia). There are two major subtypes of DM: type 1 and type 2. Type 2 diabetes (T2DM) is the most common type of DM (around 90% of people with diabetes have T2DM) [83]. It is mainly linked to insulin resistance (IR) and relatively poor insulin secretion.

Diabetes has become a major public health concern worldwide due to its growing epidemic prevalence. According to the data of the International Diabetes Federation

(IDF), diabetes affected 463 million people between the ages of 20 and 79 worldwide in 2019, which will grow to an estimated 700 million by 2045 [83]. Disease prevalence has doubled in Italy in the last 30 years (now 5.7–6.2%, or one in every six people over 65 years old) [131].

Given these alarming trends, my doctoral research has concentrated on analyzing the dietary differences between subjects with and without diabetes among non-institutionalized older adults from Southern Italy using an unsupervised machine learning approach in the identification of dietary patterns based on principal component analysis (PCA).

4.3.1 Methods

Salus in Apulia study population

Study participants were residents of Castellana Grotte, Bari, Southern Italy, and the sampling framework is based on the health registry office list on 31 December 2014. This included 19675 people, 4021 of whom were aged 65 years or older. However, only 2192 eligible individuals responded to the enrollment. They were enrolled in the "Salus in Apulia Study", a public health initiative financed by the Italian Ministry of Health and the Regional Government of Apulia and carried out by the IRCCS research hospital Saverio De Bellis. Previous prospective MICOL studies [18], which began in 1981, included these same potential research subjects. From 2014 to 2018, all eligible subjects, starting with MICOL participants, were invited to take part in the study [47]. All participants signed informed consent acknowledgements after receiving full information about their medical data to be studied. The study was conducted in accordance with the Helsinki Declaration of 1975. Every examination and informed consent form was approved by the Institutional Review Board of the National Institute of Gastroenterology and Research Hospital. All study information is stored in electronic databases that are protected according to Italian privacy laws.

Dietary Assessment and Clinical Evaluation

Diet and eating habits were assessed with a validated food frequency questionnaire (FFQ) used in previous studies [159, 249]. FFQ referred only to the frequency of intake and did not consider differences in portion sizes. In detail FFQ is a structured tool that provides a checklist of foods and beverages along with a section for respondents to indicate how frequently each item was consumed during a specified time period. Semi-quantitative FFQs go a step further by including questions about portion sizes,

typically using standardized portion sizes or offering multiple portion size options for selection.

This questionnaire investigated dietary habits over the previous year and inquired about the consumption of 85 food items. The food items are expressed in grams per day as the unit of measurement. These values are subsequently converted into kilocalories in accordance with the CREA (Council for Agricultural Research and Economics) [203] guidelines to standardize dietary intake assessment. Afterwards the original food items were further summarized in 28 food groups [47]. Supplementary Table 1 in the apppendix shows the concordance of single foods in the questionnaire and the food grouping used in the analyses [314].

The self-administered questionnaire was checked for completeness during an interview conducted by a physician at the study centre. The questionnaire also included questions about lifestyle aspects such as educational level, physical activity, and smoking habits. Additionally, at the interview, anthropometric data on waist circumference (cm), weight (kg), and height (cm) were obtained. Weight and height were measured with the mechanical scale SECA 700 and stadiometer SECA 220 (Seca GmbH and Co., Hamburg, Germany), and the body mass index (BMI) was then derived and calculated as the ratio of weight (kg) to height squared (m^2). The waist circumference was assessed with respect to the National Cholesterol Education Program: Adult Treatment Panel III (NCEP: ATP III) criteria. Diabetes mellitus was categorized as fasting blood glucose (FBG) $\geq$ 126 mg/dL.

Statistical Analysis

Patients' characteristics were reported as mean $\pm$ standard deviation (M $\pm$ SD) for continuous variables and as frequencies and percentages (%) for categorical variables. To test associations with diabetes-related diseases between groups, the Chi-square test was used for categorical variables, and the Wilcoxon rank-sum (Mann–Whitney) test was used for continuous variables.

To further reduce the number of the 28 food groups, a dimensionality reduction algorithm based on unsupervised learning was used, namely principal component analysis (PCA).

The PCA algorithm finds linear combinations of raw features (also known as projection) such that they retain as much variation in the data as possible, summarized in as few new variables (components) as possible. The vectors (loadings) describing these linear transformations produce a new set of features called scores (eigenvalues), which are uncorrelated with each other. The principal components returned by statistical

software are often ranked in descending order by their corresponding eigenvalues, which simply comprise the amount of variance in the original data explained by each principal component. The PCs with the largest eigenvalues account for most of the variation in the data. We applied PCA to the 28 food intakes in the groups of diabetics and in the control groups of non-diabetics. We considered only the most predominant PC (higher eigenvalues) in both groups, describing the loadings for each food in that PC. Due to the nature of this method, the observed food group contribution to the PCA-derived habitual dietary patterns tended to be higher for large meals with a low consistency of consumption and high interindividual variation. We chose PCA as the basis of our analysis of dietary patterns because it is the exploratory method most frequently adopted [255].

All analyses were performed using StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX, USA: StataCorp LLC.

4.3.2 Health Profiles and Dietary Intake Evaluation

The sample analyzed in the present study included 1399 subjects drawn from the "Salus in Apulia Study" population, with an average age of 73.43 ± 6.30 years. These subjects were the only ones who had completed the questionnaire and had all the necessary health data, for the analysis. The male gender was slightly predominant, accounting for 53.6% ($p = 0.02$). The population was fairly well balanced for the education level, which averaged 6.79 ± 3.79 years of schooling, as well as for mean BMI, 28.98 ± 4.26, and waist circumference, 103.48 ± 10.25 cm. The population was generally overweight had a greater abdominal circumference value than recommended. The study population was subdivided into two groups based on the presence or absence of diabetic disease: i.e., the "Not Diabetic (121)" and the "Diabetic (187)" groups.

Table 4.1 shows differences between the two groups regarding socio-demographic, lifestyle, and biochemical parameters.

Table 4.1 Sociodemographic and clinical variables in patients with and without diabetic disease. * As mean and standard deviation for continuous and percentage (%) for categorical variables. ψ Wilcoxon rank-sum test (Mann–Whitney). ** Chi-square test.

Parameters	**No (n = 1212)**	**Yes (n = 187)**	**p-value**$^{\psi}$
Gender (%)			0.02**
M	634 (52.31)	115 (61.50)	
F	578 (47.69)	72 (38.50)	
Age (yrs)	73.24 ± 6.26	74.66 ± 6.39	0.003
Education (yrs)	7.07 ± 3.80	6.52 ± 3.78	0.05
BMI (kg/m^2)	28.90 ± 4.34	29.07 ± 4.18	0.60
Normal weight (BMI $\leq$ 24.90)	221 (18.54)	28 (15.05)	
Overweight (BMI 25.0–29.90)	548 (45.97)	93 (50.00)	
Obese (BMI $\geq$ 30)	423 (35.49)	65 (34.95)	
Waist (cm)	102.92 ± 10.42	104.05 ± 10.08	0.24
Biomarkers			
Glucose (mg/dL)	98.11 ± 11.33	160.63 ± 44.98	<0.0001
Cholesterol (mg/dL)	185.89 ± 36.87	167.47 ± 36.61	<0.0001
HDL (mg/dL)	49.41 ± 13.03	42.95 ± 10.63	<0.0001
LDL (mg/dL)	115.32 ± 31.14	97.78 ± 30.54	<0.0001
Triglycerides (mg/dL)	101.84 ± 54.25	133.58 ± 78.68	<0.0001
Systolic Blood Pressure (mmHg)	132.76 ± 14.30	136.90 ± 14.76	0.0006
Diastolic Blood Pressure (mmHg)	78.48 ± 7.72	77.46 ± 8.19	0.04
IL-6 (pg/mL)	3.85 ± 6.73	4.39 ± 6.48	0.0001
TNF-α (µg/mL)	2.76 ± 3.87	3.16 ± 2.98	0.01

Table 4.2 shows the characteristics of the two groups in terms of food group consumption. Diabetic subjects consumed more potatoes (14.01 ± 31.18 g vs. 13.31 ± 16.38 g), more ready-to-eat dishes (34.45 ± 94.18 g vs. 33.24 ± 34.83 g), fewer eggs (7.40 ± 8.64 g vs. 8.33 ± 9.12 g), fewer nuts (5.49 ± 16.04 g vs. 7.56 ± 15.72 g), and fewer sugary foods and beverages (sweets: 16.52 ± 22.87 g vs. 23.74 ± 35.81 g; sugary foods: 7.22 ± 15.96 g vs. 11.08 ± 21.85 g; juices: 4.80 ± 21.26 g vs. 6.96 ± 20.64 g) than non-diabetic (ND) subjects.

Table 4.2 Food groups consumption in patients with and without diabetic disease. * Data are presented as mean and standard deviation for continuous and percentage (%) for categorical variables. ¥ Food groups were calculated by quantity of daily consumption. ψ Wilcoxon rank-sum test (Mann–Whitney).

Parameters	**No (n = 1212)**	**Yes (n = 187)**	**p-value**$^{\psi}$
Food Groups ¥			
Dairy	104.19 ± 111.15	109.38 ± 99.20	0.41
Low-Fat Dairy	101.84 ± 108.35	98.18 ± 107.52	0.49
Eggs	8.33 ± 9.12	7.40 ± 8.64	0.02
White Meat	26.32 ± 32.52	28.19 ± 59.34	0.82
Red Meat	22.62 ± 23.62	25.99 ± 39.21	0.17
Processed Meat	15.11 ± 15.45	17.57 ± 40.64	0.50
Fish	25.20 ± 23.95	33.64 ± 100.18	0.39
Seafood/Shellfish	9.45 ± 13.75	14.84 ± 64.34	0.31
Leafy Vegetables	59.02 ± 60.42	65.59 ± 93.65	0.94
Fruiting Vegetables	93.39 ± 78.56	107.85 ± 105.38	0.08
Root Vegetables	11.81 ± 26.78	14.17 ± 33.44	0.17
Other Vegetables	80.28 ± 77.02	93.80 ± 106.76	0.28
Legumes	37.78 ± 27.66	41.27 ± 46.99	0.99
Potatoes	13.31 ± 16.38	14.01 ± 31.18	0.002
Fruits	620.23 ± 537.58	598.35 ± 485.11	0.89
Nuts	7.56 ± 15.72	5.49 ± 16.04	<0.0001
Grains	157.59 ± 108.42	145.80 ± 99.22	0.29
Sweets	23.74 ± 35.81	16.52 ± 22.87	<0.0001
Sugary Foods	11.08 ± 21.85	7.22 ± 15.96	<0.0001
Juices	6.96 ± 20.64	4.80 ± 21.26	0.002
High Calorie Drinks	7.31 ± 42.37	16.85 ± 95.24	0.53
Ready-to-Eat Dishes	33.24 ± 34.83	34.45 ± 94.18	0.01
Coffee	46.41 ± 29.97	50.32 ± 28.72	0.06
Wine	121.98 ± 162.88	124.38 ± 169.39	0.85
Beer	19.54 ± 73.26	19.56 ± 69.59	0.85
Spirits	1.54 ± 5.48	1.31 ± 5.31	0.62
Water	653.61 ± 297.74	705.75 ± 312.98	0.03

Principal Component Analysis (PCA)

PCA was used to evaluate dietary differences between the "Diabetic" and "Not Diabetic" groups. Figure 4.1 shows that the most significant PCA in the "Diabetic" group was dominated in terms of loading scores by foods of plant origin. The food pattern in this group not only reflected a high-frequency intake of dairy products, eggs, vegetables

and greens, nuts, and olive oil but also sweets and sugary foods. This pattern is named the "Vegetarian Pattern".

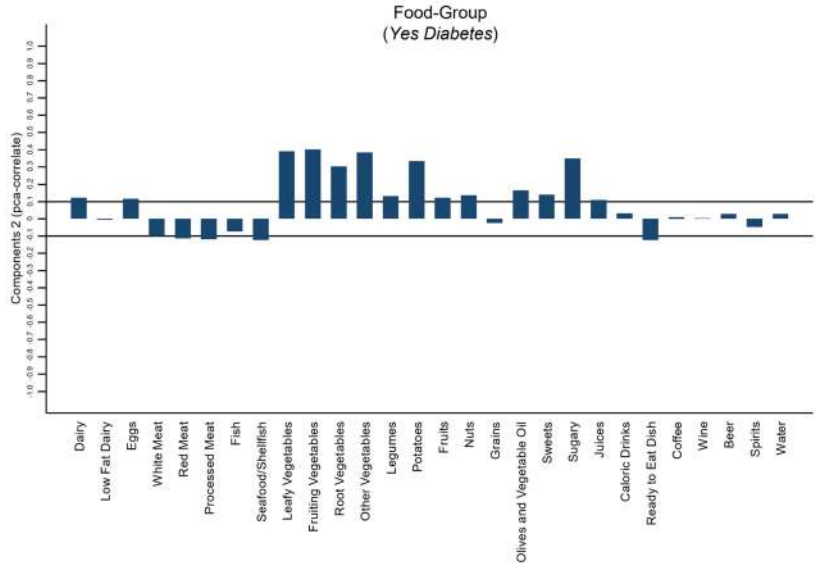


Fig. 4.1 Principal component analysis (PCA) used to identify a dietary pattern of "Diabetic" subjects.

Figure 4.2 shows the food pattern of the "Not Diabetic" group. It was characterized by a high-frequency intake of red and processed meat, seafood, high-calorie drinks, ready-to-eat dishes, wine, beer, and spirits.

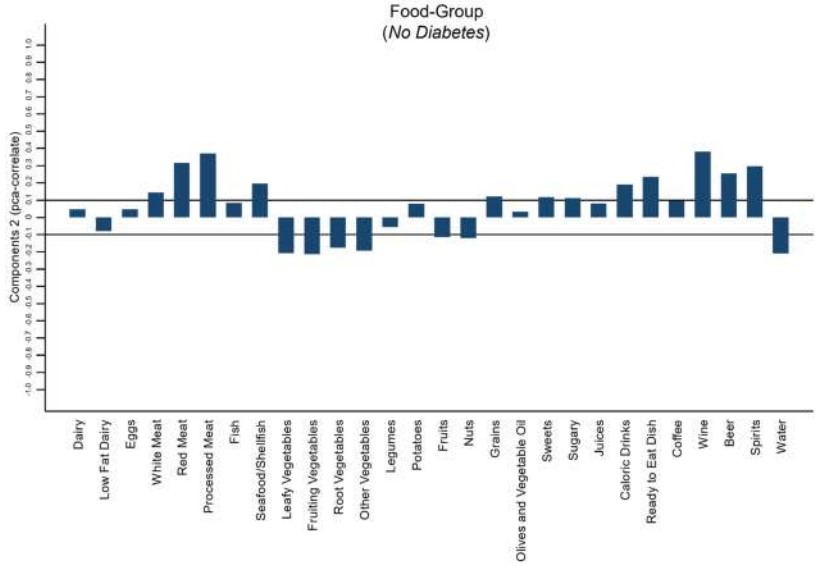


Fig. 4.2 Principal component analysis (PCA) used to identify a dietary pattern of "Not Diabetic" subjects.

4.3.3 Discussion of Key Outcomes

This cross-sectional study conducted in a sample of 1399 middle-aged Italians from Castellana Grotte (Puglia, Italy) examined dietary differences between subjects with and without diabetes, identifying a distinct dietary pattern among diabetics. Diabetic individuals exhibited a dietary pattern richer in fruits, vegetables, and nuts, and lower in red and processed meats compared to non-diabetic subjects. The diabetic group had an average age of 75, slightly older than the non-diabetic group (73 years), supporting the association between age and diabetes prevalence [132].

Biochemically, diabetic subjects showed higher levels of glucose, triglycerides, blood pressure, IL-6, and TNF-α, markers often linked to metabolic syndrome [81, 143]. Diabetes commonly coexists with obesity and hypertension, with obesity being characterized by proinflammatory markers that may contribute to metabolic disorders [50, 217].

Our study explored dietary habits more holistically by analyzing food groups rather than isolated nutrients. While both groups met the recommended portions of fruits and vegetables, the diabetic group still consumed sweets and sugary foods at higher levels than non-diabetics, despite guidelines advising limited intake of added sugars [290].

In line with the Italian Society of Diabetology (SID) recommendations, a diet for diabetes management should emphasize fiber-rich fruits, vegetables, legumes, and whole grains, which help control blood glucose and triglycerides [140, 250]. Regular consumption of five portions of fruits and vegetables per day and four portions of legumes per week aligns with these goals, with studies indicating that 40g or more of daily fiber intake can significantly reduce blood glucose levels [48].

Despite national data showing insufficient fruit and vegetable consumption among the elderly [245], participants in this study largely met Mediterranean diet guidelines, perhaps reflecting their strong ties to regional dietary traditions. The Mediterranean Diet, which places fruits and vegetables at its core while limiting red and processed meats, has been associated with numerous health benefits, including better glycemic control and reduced cardiovascular risk [68, 88].

Strengths and Limitations

The present study evaluated dietary differences between subjects with and without diabetes drawn from non-institutionalized older adults from Southern Italy, using an unsupervised machine learning algorithm. The main strength of our study is that no study has yet analyzed these aspects in similar populations using this novel approach.

Another strength is the description of a dietary pattern characteristic of diabetic subjects.

However, some limitations must be considered. One of the main limitations of the study is the use of food frequencies instead of calculating quantitative daily intake. This type of measurement could increase the bias that is usually associated with a retrospective dietary assessment over a period of one year when compared to the actual intake. Nevertheless, despite the reported limitation of this assessment method, FFQs remain the dietary assessment method most used to study dietary patterns and population eating habits [11, 68, 88, 129, 142, 180, 245, 290]. Another important limitation is the nature of the study, which was cross-sectional and does not allow a clear directionality of an association to be discerned. Moreover, conclusions in this study should be considered a descriptive comparison of the dietary pattern between two groups, without quantitative statistical inferences. Another limitation concerns the age of the diabetic subjects in our sample, which is reduced compared to the age range of the typical diabetic population.

4.4 Neural Network Assessment of Frailty-Related Nutrition in Aging Italians

The ongoing research described previously has naturally progressed to investigating nutritional factors associated with frailty among the elderly population in Southern Italy.

Building on the findings discussed earlier, this study introduces a machine learning approach using a Neural Network model. The model was developed to classify an aging cohort from the Salus in Apulia study population described above, based on their physical frailty condition. It aims to identify key dietary factors and readily available clinical data that best encapsulate the characteristics of the frailty phenotype [103].

4.4.1 Methods

Clinical and Laboratory Examination

Education was categorized according to the major divisions of the Italian national education system, i.e., lowest level: < 6 years reflecting primary school education; middle level: $6-8$ years reflecting lower secondary school education; highest level > 8 years reflecting upper secondary school/high school education and university education. A blood sample was collected in the morning after overnight fasting to measure the levels of fasting blood glucose (FBG), glycated hemoglobin (HbA1c), total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides using standard automated enzymatic colorimetric methods (AutoMate 2550, Beckman Coulter, Brea, CA, US) under strict quality control. LDL cholesterol was calculated using the Friedewald equation [89]. Plasma glucose was determined using the glucose oxidase method (Sclavus, Siena, Italy). Blood cell count was determined by a Coulter Hematology analyzer (Beckman Coulter, Brea, CA). The clinical evaluation included extemporaneous ambulatory systolic blood pressure (SBP) and diastolic blood pressure (DBP), determined in a sitting position after at least a 10-min rest, at least three different times, using the OMRON M6 automatic blood pressure monitor. Serum high-sensitivity C-reactive protein (CRP) was assayed using a latex particle-enhanced immunoturbidimetric assay (Kamiya Biomedical Company, Seattle, WA) (reference range: 0–5.5 mg/L; interassay coefficient of variation: 4.5%). Serum interleukin (IL)-6 and tumor necrosis factor-alpha (TNF-α) were assayed using the quantitative sandwich enzyme technique of ELISA (QuantiKine High Sensitivity Kit, R&D Systems, Minneapolis, MN, and QuantiGlo immunoassay from R&D Systems,

Minneapolis, MN). The interassay coefficient of variations was 11.7% for IL-6 and 13.0% for TNF-α. Inflammatory marker assays performed were analyzed at the same laboratory following strict quality control procedures.

Food Questionnaire Assessment

A food frequency questionnaire (FFQ) previously validated on our population [283, 315] was used following a semi-quantitative approach based on foods consumed by the Mediterranean population. For these foods, the frequency of predefined portion intake, in the past year, was probed in the questionnaire using a 9-category scale. Each portion was assigned a weight, and then the intake amount expressed as the average intake in grams per day was calculated. The FFQ was structured into 11 sections that partly reflect the sequence of foods throughout the day and include foods with similar characteristics: cereals, meat, fish, milk and dairy products, vegetables, legumes, fruits, miscellaneous foods, water and alcoholic beverages, olive oil and other edible fats, coffee/sugar, and salt. In a subsequent step, the FFQ was validated against the dietary records, and the results were reviewed to make necessary changes to the questionnaire [109]. The initial FFQ consisted of 85 food items and included questions regarding fat consumption. To improve efficiency and reduce redundancy, the questionnaire was refined to 77 food items, allowing for more streamlined analyzes and facilitating statistical modeling as well as comparisons with other studies. The food items are expressed in grams per day, and the recorded values are then converted into kilocalories according to the guidelines provided by the CREA (Council for Agricultural Research and Economics) [203], ensuring consistency in the assessment of dietary intake.

Physical Frailty Assessment

Assessment of physical frailty status was performed using the CHS operational criteria slightly modified for the present study, i.e., the positivity of three or more of the following components weight loss, exhaustion, low physical activity levels, weakness, and sluggishness, as detailed elsewhere [45]. The whole sample was categorized into two groups based on the number of physical frailty components. Subjects who met three or more criteria were included in the frail group, otherwise non-frail (or robust).

Statistical Analysis

The entire sample was divided into two groups according to physical frailty conditions. Normal distributions of quantitative variables were tested using the Kolmogorov-Smirnov test. Therefore, data were reported as mean ± standard deviation (M ±

SD) for continuous measures and frequency and percentages (%) for all categorical variables. A statistical approach based on the null hypothesis significance test (NHST) was not used to focus the reader's attention on practical differences between groups in terms of effect size. Differences in the prevalence of exposure groups, i.e., physical frailty and other categorical variables, and their 95% CIs were calculated and used to assess significant differences in the magnitude of the association, i.e., effect size (ES). Using Wilcoxon's effect size, differences between continuous variables were calculated by confidence intervals around them following a non parametric approach [109]. Differences between categorical variables were assessed using differences in prevalence and confidence intervals around them. Dietary records were normalized to a daily assumption of 2000 kcal. A Lasso regression was performed on Physical Frailty status, as the dependent variable, to reduce the number of predictors in the Neural Network model (NN). A minimum value of lambda was chosen to optimize the Lasso regression. There were 20 remaining predictors associated with the dependent variable in the Lasso regression and selected for inclusion in the NN as input. Statistical analysis were performed using RStudio software, Version 1.2.5042 (RStudio, Inc.,Boston, MA, USA).

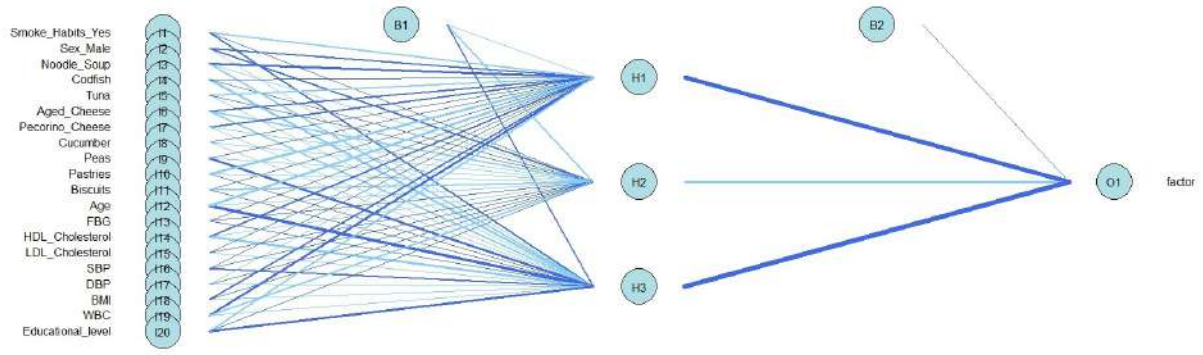


Fig. 4.3 Neural Network Architecture. Input units: from I1 to I20; Hidden units: from H1 to H3; Output unit: O1. B1 and B2 are, respectively, hidden neuron bias and output neuron bias.

Machine Learning Approach

The Machine Learning algorithm considered was the NN, which is a reinforcement learning algorithm. The 20 predictors obtain from Lasso regression, were used to predict the frailty condition. The split percentages for the training and test were, respectively, 70% and 30%. All data were normalized using Min-Max normalization.The Caret

package was used to build and train the NN model. The hyperparameters tuning was performed, using 5-fold cross validation on the training set, on:

- hidden layer size: from 3 to 9;
- decay values: $0.001, 0.01, 0.1$.

The logistic activation function was used for each neuron. The best selection of hyperparameters was obtained using the best accuracy, actually the hidden layer size selected was 3 and the decay value was 0.1. Therefore the NN architecture was composed of one input layer, one hidden layer and one output layer. The input layer size was 20, the hidden layer size was 3 and the output layer size was 1. Fig. 5.1 shows the architecture of the NN. Fig. 5.2 shows the variation of the accuracy value according to the combination of the hidden layer size and decay values.

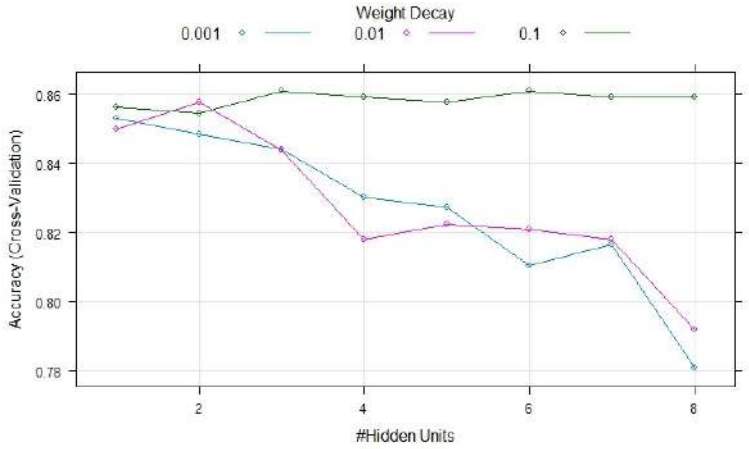


Fig. 4.4 Accuracy plots of Training dataset.

The absolute importance was evaluated using Garson's algorithm [95]. This algorithm decomposes the connection weights of a neural network to estimate the relative contribution of each input feature to the network's output. It calculates importance scores by distributing the connection weights of hidden neurons proportionally across input features, providing an interpretable measure of feature relevance.

Olden's algorithm [207] evaluated the relative importance of the input variables in the neural network (NN). This algorithm assesses importance by directly analyzing the connection weights of the NN, accounting for the magnitude and direction (positive or negative) of each weight. Unlike Garson's algorithm, Olden's method preserves the sign of the weights, providing a more accurate and interpretable measure of the contribution of each input variable to the model's predictions.

Accuracy metrics were evaluated for both datasets to assess model performance and overfitting. The Machine Learning algorithm was performed using RStudio software, Version 1.2.5042 (RStudio, Inc.,Boston, MA, USA).

4.4.2 Frailty Status Analysis and Predictive Modeling

The examined population (N = 1,929) appeared gender-balanced (53.20% male, n = 493). The mean age was 74.03 ± 6.36 years. Table 4.3 summarizes the main differences among the population variables according to physical frailty status. Physical frailty was prevalent in 14% of the entire sample (N=130), compared with 86% of those who were not frail (N= 796). Older age (ES: 0.14, 95%CI 0.07 to 0.19), female gender (ES: 12.72, 95%CI 3.55 to 21.89), and high BMI (ES: 0.08, 95%CI 0.01 to 0.14) were associated with physical frailty, as high diastolic (ES: 0.07, 95%CI 0.01 to 0.13) and systolic (ES: 0.08, 95%CI 0.02 to 0.15) blood pressure. Higher educational level and smoking habit prevailed in non-frail individuals compared with the counterpart (ES: 5.98, 95%CI −9.20 to −2.77). As for the metabolic biomarkers, the lipid profile showed significantly lower total, HDL, and LDL cholesterol values in frail subjects (ES: 0.08, 95%CI 0.01 to 0.13, ES:0.07, 95%CI 0.01 to 0.13, and ES: 0.08, 95%CI 0.02 to 0.14). RBC levels were lower in the frail than the counterpart (ES: 0.06, 95%CI 0.01 to 0.13). Conversely, serum levels of interleukin-6 and CRP were higher in the frail group (ES: 0.07, 95%CI 0.03 to 0.12, and ES: 0.06, 95%CI 0.02 to 0.11). Table 4.4 shows the daily calorie intake (%) per food of the whole sample according to physical frailty condition. Consumption of spaghetti soup, pecorino cheese, fennel, and chocolate was found to be significantly higher in the frail group than non-frail (ES: 0.08, 95%CI 0.01 to 0.14, ES: 0.08 95%CI 0.02 to 0.14, ES: 0.06 95%CI 0.01 to 0.13, and ES: 0.08 95%CI 0.01 to 0.14, respectively). Ham consumption was statistically lower in the frail group than its counterpart (ES: 0.02 95%CI: 0.01 to 0.14). There were 20 features selected using the Lasso regression model on physical frailty condition (see Table 4.5).

Table 4.3 Description of the whole sample socio-demographic and biochemical variables according to Physical Frailty status. N: 926
Legend: BMI: Body Mass Index, DBP: Diastolic Blood Pressure, SBP: Systolic Blood Pressure, FBG: Fasting Blood Glucose, HbA1c: Glycated Hemoglobin, GGT: Gamma Glutamyl Transferase, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, RBC: Red Blood Cells, WBC: White Blood Cells, TNF-alpha: Tumor Necrosis Factor, CRP: C-Reactive Protein, PA score: InChianti Physical Activity score, PF score: Global Physical Frailty score.
All data are shown as mean ± sd, median (min to max) for continuous variables and as n (%) for proportions.
*Wilcoxon's effect size.

	Without Physical Frailty		**With Physical Frailty**		**Effect Size***	
	Mean ± SD	**Median (Min to Max)**	**Mean ± SD**	**Median (Min to Max)**	**Value**	**95% CI**
Proportions (%)	796 (86.00)		130 (14.00)			
Age (years)	73.69 ± 6.32	73 (65 to 92)	76.1 ± 6.25	76 (65 to 95)	**0.14**	**0.07 to 0.19**
Sex						
Male	438 (55.00)		55 (42.30)		12.72	3.55 to 21.89
Female	358 (45.00)		75 (57.70)		12.72	3.55 to 21.89
Educational Level (years)	6.92 ± 3.81	5 (0 to 18)	5.94 ± 4.15	5 (0 to 17)	**0.13**	**0.07 to 0.21**
Smoke Habits (yes)	66 (8.30)		3 (2.30)		**-5.98**	**-9.20 to -2.77**
BMI (kg/m^2)	28.58 ± 4.89	28.13 (14.09 to 48.47)	29.73 ± 5.35	29.55 (18.61 to 48.77)	**0.08**	**0.01 to 0.14**
DBP (mmHg)	78.36 ± 7.9	80 (40 to 110)	77.35 ± 8.59	80 (50 to 100)	0.04	−0.02 to 0.09
SBP (mmHg)	133.51 ± 14.71	130 (80 to 180)	137.19 ± 15.01	140 (100 to 200)	**0.08**	**0.02 to 0.15**
FBG (mg/dL)	105.45 ± 29.49	99 (62 to 365)	110.78 ± 38.1	101 (54 to 435)	**0.07**	**0.01 to 0.13**
HbA1c (mmol/mol)	40.64 ± 10.69	39 (18 to 128)	42.09 ± 11.2	40 (24 to 101)	0.06	−0.01 to 0.12
GGT (U/L)	35.39 ± 37.24	19 (6 to 158)	39.64 ± 40.87	22.5 (7 to 158)	0.05	−0.02 to 0.09
AST (U/L)	32.25 ± 28.77	23 (2 to 197)	35.8 ± 33.51	24 (1.2 to 189)	0.06	−0.01 to 0.11
ALT (U/L)	26.16 ± 22.71	19 (6 to 221)	26.02 ± 16.52	20 (8 to 81)	0.02	−0.03 to 0.06
Total Cholesterol (mg/dL)	184.01 ± 37.28	183.5 (76 to 386)	176.82 ± 40.21	178.5 (94 to 302)	**0.08**	**0.01 to 0.13**
LDL Cholesterol (mg/dL)	113.33 ± 31.18	114 (30 to 220)	107.41 ± 33.88	108.5 (37 to 209)	**0.07**	**0.01 to 0.13**
HDL Cholesterol (mg/dL)	49.07 ± 13.21	47 (23 to 117)	46.01 ± 11.54	44 (23 to 89)	**0.08**	**0.02 to 0.14**
Triglycerides (mg/dL)	103.86 ± 57.07	89 (21 to 498)	113.43 ± 65.31	93 (30 to 443)	0.06	−0.01 to 0.12
RBC (10^6 cells/mm^3)	4.82 ± 1.38	4.76 (3.34 to 40.8)	4.68 ± 0.56	4.7 (3.17 to 6.78)	**0.06**	**0.01 to 0.13**
WBC (10^3 cells/mm^3)	6.05 ± 1.71	5.8 (1.7 to 20.9)	6.34 ± 2.3	5.93 (3.29 to 24.36)	0.04	−0.02 to 0.08
Platelets (10^3 cells/mm^3)	218.62 ± 55.37	212.5 (65 to 520)	219.22 ± 70.5	211.5 (75 to 605)	0.02	−0.05 to 0.04
Interleukin 6 (pg/ml)	4.1 ± 7.36	1.94 (0.06 to 64.94)	4.26 ± 5.96	2.28 (0.36 to 41.9)	**0.07**	**0.03 to 0.12**
TNF-alpha (pg/ml)	2.99 ± 4.56	2.2 (0.03 to 46.48)	2.89 ± 2.17	2.27 (0.2 to 13.83)	0.05	−0.01 to 0.11
CRP (mg/dL)	0.58 ± 0.86	0.31 (0.1 to 10.96)	0.65 ± 1.12	0.36 (0.1 to 10.96)	**0.06**	**0.02 to 0.11**
Step test (Impaired)	30 (3.80)		71 (54.60)		**50.85**	**42.19 to 59.51**
Weight loss	32 (4.00)		21 (16.20)		**12.13**	**5.66 to 18.61**
Sit to stand test (Impaired)	161 (20.20)		120 (92.30)		**72.08**	**66.72 to 77.45**
Gait test	114 (14.30)		121 (93.10)		**78.76**	**73.76 to 83.75**
PA score (Impaired)	127 (16.00)		107 (82.30)		**66.35**	**59.32 to 73.39**
GPF score (Impaired)	0.58 ± 0.72	0 (0 to 2)	3.38 ± 0.56	3 (3 to 5)	**0.64**	**0.61 to 0.69**

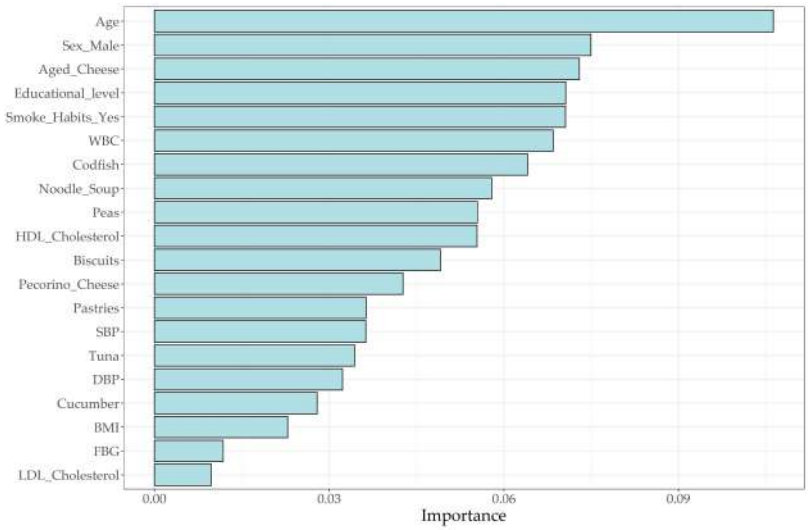
Table 4.4 Description of daily calorie intake by food category according to Physical Frailty status. N: 926. All data are shown as percentage of calories normalized to 2000 kcal/day. *Wilcoxon's effect size.

	Without Physical Frailty		**With Physical Frailty**		
	Mean ± SD	**Median (Min to Max)**	**Mean ± SD**	**Median (Min to Max)**	**Effect Size***
Proportions (%)	796 (86.00)		130 (14.00)		
Veal Meat	5.87 ± 4.78	4.28 (0 to 32.1)	5.34 ± 5.19	4.28 (0 to 44.94)	**0.03 (-0.02 to 0.08)**
Chicken Meat	6.38 ± 5.8	4.5 (1.15 to 47.25)	7.02 ± 7.92	4.5 (1.15 to 70.9)	**0.02 (-0.04 to 0.05)**
Rabbit Meat	3.68 ± 4.00	2.65 (1.36 to 39.83)	3.59 ± 3.4	2.01 (1.36 to 26.55)	**0.02 (-0.05 to 0.04)**
Lamb Meat	3.73 ± 8.91	2.38 (0 to 100.17)	3.29 ± 3.16	2.38 (1.59 to 23.85)	**0.02 (-0.03 to 0.06)**
Raw Ham	3.17 ± 4.28	2.35 (0.59 to 49.35)	2.84 ± 4.46	1.18 (0.59 to 41.12)	**0.02 (0.01 to 0.14)**
Cooked Ham	4.41 ± 5.35	2.88 (0.72 to 50.4)	4.67 ± 6.1	2.88 (0.72 to 40.32)	**0.03 (-0.03 to 0.08)**
Noodle Soup	11.99 ± 15.12	5.94 (0 to 124.77)	15.77 ± 17.92	11.88 (0 to 83.18)	**0.08 (0.01 to 0.14)**
Bread	76.11 ± 38.78	75.04 (0 to 168.84)	73.85 ± 38.07	75.04 (2.01 to 168.84)	**0.02 (-0.05 to 0.04)**
Flat Bread	15.41 ± 21.49	12.5 (0 to 175)	14.76 ± 19.13	12.5 (0 to 175)	**0.01 (-0.06 to 0.02)**
Citrus	21.21 ± 13.9	22.5 (0 to 47.25)	20.14 ± 13.61	22.5 (0.56 to 47.25)	**0.03 (-0.03 to 0.06)**
Apricot	11.69 ± 11.31	6.3 (0 to 39.69)	11.02 ± 10.43	6.3 (0.32 to 39.69)	**0.02 (-0.04 to 0.04)**
Watermelon	14.57 ± 12.53	10.42 (0 to 43.74)	14.96 ± 13.39	10.42 (0.42 to 43.74)	**0.01 (-0.07 to 0.01)**
Melon	6.19 ± 7.28	2.94 (0 to 30.88)	7.17 ± 8.38	2.94 (0.3 to 30.88)	**0.02 (-0.05 to 0.04)**
Apples/Pears	17.28 ± 11.99	16.5 (0 to 46.2)	17.74 ± 12.63	16.5 (0.55 to 46.2)	**0.03 (-0.06 to 0.02)**
Kiwi	6.55 ± 9.56	1.8 (0 to 50.4)	7.17 ± 9.86	2.4 (0.6 to 50.4)	**0.02 (-0.03 to 0.06)**
Banana	21.59 ± 15.16	20.9 (0 to 63.84)	21.24 ± 16.26	18.62 (0 to 63.84)	**0.02 (-0.05 to 0.04)**
Dried Fruits	10.3 ± 15.5	5.26 (0 to 110.56)	9.04 ± 13.23	2.81 (0 to 73.71)	**0.04 (-0.02 to 0.08)**
Canned Fruit	3.39 ± 6.77	0.74 (0 to 61.95)	3.52 ± 8.71	0.74 (0 to 61.95)	**0.01 (-0.06 to 0.02)**
Pecorino Cheese	12.16 ± 17.31	5.05 (1.62 to 148.47)	15.18 ± 18.47	6.06 (1.62 to 70.7)	**0.08 (0.02 to 0.14)**
Fennel	2.99 ± 3.3	1.88 (0 to 15.75)	2.43 ± 2.92	1.12 (0 to 11.81)	**0.06 (0.01 to 0.13)**

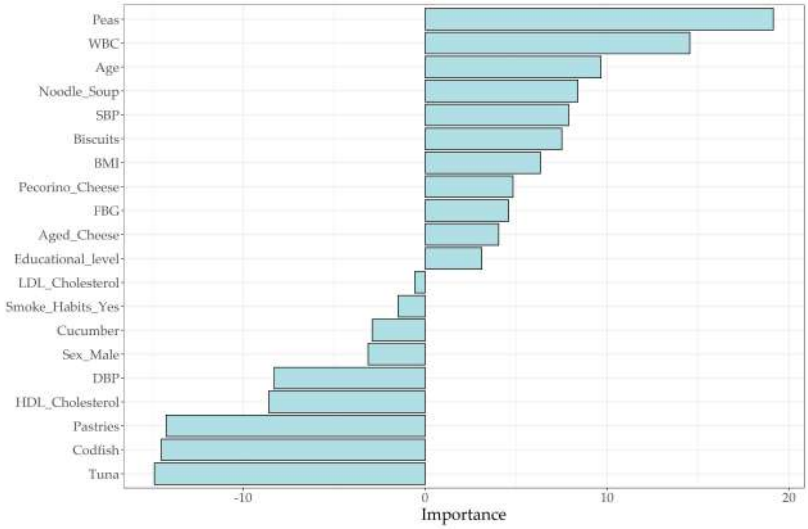
Table 4.5 Lasso regression model on Physical Frailty as dependent variable and selected variables regressors with minimum lambda value.
Nutritional data are normalized to 2000 *kcal/day and expressed as percentage.*

	Minimum Lambda: 0.01521157
	LASSO Coefficients
Noodle soup	0.01
Codfish	−0.01
Tuna	−0.01
Aged Cheese	0.01
Pecorino Cheese	0.01
Cucumber	−0.02
Peas	0.01
Pastries	−0.01
Biscuits	0.01
Age (years)	0.03
Smoke Habits (yes)	−0.19
Sex (Male)	−0.29
FBG (mg/dl)	0.01
HDL Cholesterol (mg/dl)	−0.01
LDL Cholesterol (mg/dl)	−0.01
SBP (mmHg)	0.01
DBP (mmHg)	−0.01
BMI (Kg/m^2)	0.01
WBC (10^3 cells/mm^3)	0.01
Educational level (years)	−0.01

The most important feature to identify physical frailty condition, according to Garson's algorithm, was age followed by gender and consumption of aged cheese (Fig. 5.3). On the other hand, according to Olden's algorithm, pea consumption was the most important characteristic used by the NN, while tuna consumption was the most negative (Fig. 5.4).



(a) Garson's variables importance plot.



(b) Olden's variables importance plot.

Fig. 4.5 Comparison of Garson's and Olden's variables importance plots.

The trained neural network demonstrated a robust accuracy of 86.49% on the training dataset and maintained a closely comparable accuracy of 85.77% on the test dataset, indicating a stable model performance with no significant evidence of overfitting.

4.4.3 Insights and Implications

In this study, a NN to support frailty diagnosis in the aging population has been developed on a model made up of easily available variables and nutritional data. In particular, in this work, we trained and tested a NN to identify frailty subjects

that can be applied in epidemiological study or screening. The aim was to identify frailty subjects quicker instead of using classical tests that are more complicated to perform, as they require information challenging to find. Previous studies used other features, more difficult available, to predict frailty outcome with good performances. However, our study is the first that uses easily available parameters to build a NN to support frailty diagnosis in an older population. Our previous results show that our NN has good performances on the train and test dataset and a good accuracy, and there is no presence of overfitting. The NN makes few mistakes and it predicts, with high specificity, healthy subjects. This allows us to conduct preliminary screening, ensuring more specific analyses only when necessary. Moreover, the most important variables were compared to identify those that have a major contribution to the outcome. As we expected, age, gender, foods rich in fats, and smoking habits are the most important features to predict the frailty condition according to Garson's method, whereas the direction of each variable importance was assessed with Olden's method. As underlined in a previous study [46], the population assessed is characterized by higher consumption of legumes and unrefined cereals; therefore, the association found between peas (legumes in general) and physical frailty confirms this pattern of consumption. Legumes' organoleptic characteristics and palatability are suitable for consumption in frail and advanced age subjects, threfore preferably consumed. In conclusion, the NN is a good support for frailty diagnosis and it can be applied in an epidemiological study or screening.

4.5 Machine learning in Mild Cognitive Impairment Prediction

Building on the exploration of machine learning techniques applied to epidemiological data, this work now shifts focus to predicting neurodegenerative disease outcomes, particularly MCI. This section presents advanced machine learning methods applied to model and predict MCI by examining key epidemiological factors that may signal the early stages of neurodegeneration. Through this approach, we aim to identify predictive patterns and markers within large-scale healthcare data, contributing to the field of neurodegenerative disease prevention and enhancing the accuracy of early diagnostic frameworks.

It is important to define the concept of MCI, initially introduced by Petersen *et al.* in the late 1990s, as part of the cognitive aging discipline [223]. MCI describes a condition where an individual experiences an age-independent decline in cognitive

abilities that is not severe enough to meet the criteria for dementia. MCI can impact memory, language, attention, and other cognitive functions, and it is considered a risk factor for dementia. According to the Alzheimer's Association, approximately $15-20\%$ of people aged 65 and older are affected by MCI[1].

In this work, we propose a Pareto-optimality-based approach to identify the best ML model. Specifically, we test our approach by training multiple ML models on data from the *Salus in Apulia study population* [45] to predict MCI using easily accessible clinical dataset features that best describe this outcome.

To summarize, the contributions of this work are:

- Developing a predictive approach based on readily available healthcare data for MCI prediction;
- Implementing a Pareto-based model selection approach to identify the optimal model with high accuracy while minimizing misclassifications in MCI predictions;

4.5.1 Pareto Optimality and Hypervolume

In this Section, we explain the concept of Pareto Optimality. Its definition is widely exploited in Multi-Objective Optimization Problems. Formally, a Multi-Objective Optimization Problem (MOOP) is defined as [188]:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \boldsymbol{f}(\mathbf{x}) = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})\} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{4.1}$$

The vector $\mathbf{x} \in \mathbb{R}^n$ is formed by n independent *decison variables*. The set $\mathcal{X} \subseteq \mathbb{R}^n$, generally known as *feasible set*, is defined by a set of equality and inequality constraints. The vector of functions $\boldsymbol{f(\cdot)}$ is composed by k scalar *objective functions* $f_i : \mathbb{R}^n \to \mathbb{R}^k$ with $i = 1, \ldots, k$; $k \geq 2$. The space $\mathbb{R}^k$ is known as *objective function space*.

In a MOOP, since typically there is no single global solution, it is usually adopted the concept of *Pareto optimality* which leverages on the *Pareto dominance* relation [176]. A vector $\mathbf{x}^\star$ Pareto-dominates vector $\mathbf{x}$, denoted by $\mathbf{x}^\star \prec \mathbf{x}$, if and only if: $f_i(\mathbf{x}^\star) \leq f_i(\mathbf{x}) \forall i \in \{1, \ldots, k\}$ and $\exists i \in \{1, \ldots, k\} \,|\, f_i(\mathbf{x}^\star) < f_i(\mathbf{x})$. Hence, a solution $\mathbf{x}^\star \in \mathcal{X}$ is Pareto optimal if there does not exist another solution $x \in \mathcal{X}$ such that $\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}^\star)$.

Then, solving the problem in Equation (4.1) means to find the solutions $\mathbf{x} \in \mathcal{X}$ such that their images $\mathbf{f}(\mathbf{x})$ are not Pareto-dominated by any other vector in the feasible set.

[1] https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment

The group of non Pareto-dominated solutions in the feasible set is called *Pareto optimal set*, whose image in the objective function space is known as *Pareto frontier* [213]. In essence, the Pareto frontier represents the solutions that accomodate the trade-offs between the different objective functions. However, sometimes there is a need to select a single solution from the Pareto frontier that best fits the considered task. In this regard, the *hypervolume* indicator helps to select a single solution from the Pareto frontier [313]. Given a Pareto optimal solution $\mathbf{x}^\star \in \mathbb{R}^k$, a reference point $\mathbf{r} \in \mathbb{R}^k$, and the Lebesgue measure λ, the hypervolume of $\mathbf{x}^\star$ with respect to $\mathbf{r}$ is $\mathcal{HV} = \lambda(\{\mathbf{x} \in \mathbb{R}^k \,|\, \mathbf{x}^\star \prec \mathbf{x} \prec \mathbf{r}\})$. The $\mathcal{HV}$ value is the volume of the hypercube determined by the solution $\mathbf{x}^\star$ and the reference point $\mathbf{r}$. Therefore, the solution with the highest hypervolume can be considered as the best solution among the Pareto optimal set.

4.5.2 Methods

4.5.3 Dataset Description

The population under study is derived from the *Salus in Apulia Study*. Subjects included in the analysis were characterized by 70 features, consisting of continuous or categorical variables related to biochemical, anthropometric, and cognitive tests, acquired through interviews, blood analyses, and clinical visits. The dataset is represented as tabular data with $1,929$ observations and 70 features, including the target variable. To obtain a ready-to-use dataset, all categorical non-relevance features were removed from the analysis except *sex*. A new dataset was obtained containing 57 features.

Preprocessing

We parsed the dataset using `pandas`[2]. Initially, the dataset was subsampled using the RandomUnderSampler technique from the `imblearn` library [162] to balance classes by randomly sampling observations, resulting in 818 observations (414 for *No MCI* and 414 for *Yes MCI*). The dataset was then split into training and test sets with an 80/20 ratio. Continuous values were standardized using the StandardScaler[3] method, widely used in ML to reduce variance [91]. Feature selection was performed using `SelectKBest` from `scikit-learn`, with the optimal value of k determined through a `GridSearchCV` (5-fold)[4] with a Random Forest Classifier [209], testing k values from 1 to 57. The best k was 9, selecting Pure tone audiometry (PTA)$(0.5-2\text{kHz})$ left and

[2]https://pandas.pydata.org/

[3]https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

[4]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

right, Total Cholesterol, HDL Cholesterol, Mini Mental State Examination (MMSE), Frontal Assesment Battery (Frontal Assesment Battery) raw score, Vitamin D, the mean PTA($0.5-2$kHz), and mean PTA(8kHz). Although gender and sex were not selected, they were included as medically relevant features [305].

Classification Models

In order to determine the best classifier to predict the MCI status, we compared the following models: Multilayer Perceptron (MLP) [22]; Random Forest (RF) [209]; SVM [202]; Extreme Gradient Boosting (XGBoost) [53]; The models are developed with Python, using Scikit-learn library[5] [219] and tuned using a grid search exploration strategy with a 5-fold cross-validation. The models maximize the Area Under the ROC Curve (AUC) value in order to obtain the best predictive power in binary classification. For each model under study, the list of explored hyperparameter values is reported in Table 4.6.

Table 4.6 Hyperparameter list and values for the classification models reported in this work.

Algorithm	**Hyperparameter**	**Values**
Multilayer Perceptron	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}
Random Forest	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	max_features	{sqrt, log2}
	max_depth	{np.arange(1, 30, 1)}
	criterion	{gini, entropy}
Support Vector Machines	seed	{42}
	kernel	{rbf, poly}
	gamma	{0.1, 0.001, 0.3, 0.003, 0.5, 0.05}
eXtreme Gradient Boosting	seed	{42}
	n_estimators	{np.arange(1, 50, 2)}
	learning_rate	{np.arange(0.1, 1, 0.2)}
	scale_pos_weight	{1.000}

[5] http://scikit-learn.org

Pareto Optimality-based Selection

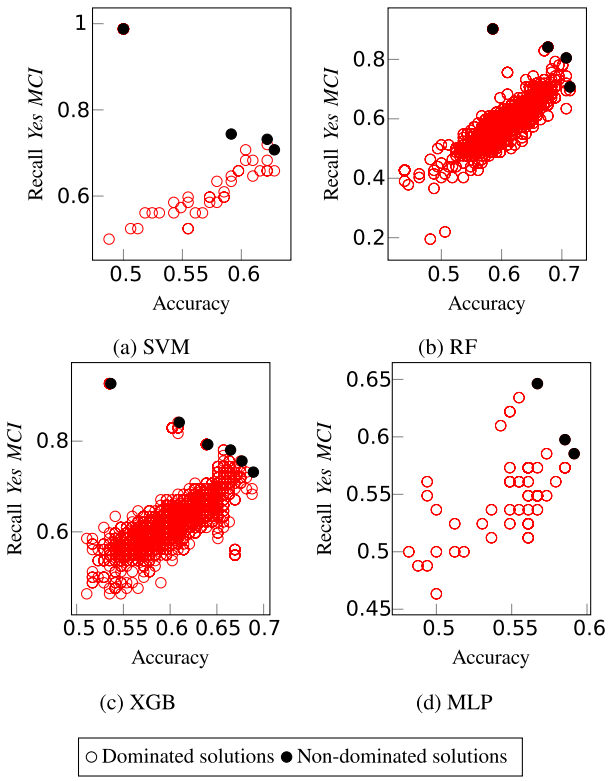


Fig. 4.6 Models' hyperparameter configurations in the objective function space Accuracy/Recall on YES MCI class. For each model, the black dots are on the Pareto frontier, while the red dots are dominated solutions.

In this Section, we describe the approach adopted to select the best classifier for our task. Oftentimes, traditional approaches consist in selecting the best models relying only on the accuracy of the candidate models. However, in MCI systems, it is of greater importance to classify MCI subjects rather than healthy subjects correctly. As a consequence, it is not sufficient to choose the model having the best overall accuracy. Indeed, the selected model should maximize the number of correctly classified MCI subjects (i.e., the Recall on *yes MCI* class), even paying some trade-off with the overall accuracy of the system. In other words, we should perform a *multi-criteria* selection of the best model. To this end, we exploit the concept of Pareto Optimality to identify the models that best accommodate the trade-off between overall accuracy and Recall on *yes MCI* class.

Once all the models' configurations described above are trained, they represent different solutions in the objectives function space having accuracy and Recall on *yes MCI* class on the x-axis and y-axis, respectively. For each model, we identify the Pareto-optimal solutions lying on the Pareto frontier. To clarify this aspect, we plot the objective function spaces and the solutions on them for each model in Figure 4.6. Then, we take into account the Pareto frontier of each model, and we find the Pareto-optimal points among all the models (Figure 4.7). Finally, we exploit the widely-used Hypervolume indicator to select one–best solution from the final Pareto frontier.

4.5.4 Pareto-optimal Models for Enhanced MCI Prediction

In this section, we detail the results obtained from our experimental approach. Specifically, in Table 4.7, we report the results obtained by computing the Hypervolume indicator for each model lying on the Pareto frontier in Figure 4.7. We observe that six models are Pareto-optimal. On the one hand, we notice that no configuration of MLP lies on the Pareto frontier. On the other hand, four out of six Pareto-optimal solutions are different configurations of Random Forest. Specifically, two Random Forest configurations achieve the highest $\mathcal{HV}$ value. Both of them are characterized by the Shannon entropy as information gain and 3 as the maximum depth of the tree. Hence, we have two candidates as the best model to perform the MCI prediction. To determine the best one, we also analyze the performance of such models in terms of Precision, Recall, and F1-score for each class. From Table 4.8, we claim that RF_2 is the best model since it achieves better performance than RF_1 with respect to all the considered metrics except for Recall on *Yes MCI* class. To summarize, RF_2 performs better according to a *multi-criteria* evaluation. In fact, the performance obtained by RF_2 is the most balanced as it achieves an accuracy of 71%, a high value of recall (0.80), the best value of precision, in the prediction of *Yes MCI* (0.67), and the best value of F1-score in the prediction of *Yes MCI* (0.73).

Table 4.7 Pareto-optimal models' configurations and their performance. The best values for $\mathcal{HV}$ indicator are in bold. The models are characterized by the following hyperparameters configurations. RF_1: Entropy as *Criterion*, *Max depth*=3, *Max Features*=log2, *Number of estimators*=1; RF_2: Entropy as *Criterion*, *Max depth*=3, *Max Features*=sqrt, *Number of estimators*=9; RF_3: Entropy as *Criterion*, *Max depth*=1, *Max Features*=sqrt, *Number of estimators*=1; SVM_1 C=6, γ =0.003, poly as *Kernel*; $XGBoost_1$: *Learning Rate*=0.1, *Max depth*=1, *Number of estimators*=1; RF_4: Entropy as *Criterion*, *Max depth*=5, *Max Features*=log2, *Number of estimators*=13.

Model	**Accuracy**	**Recall *Yes MCI***	**HV**
RF_1	0.68	0.84	**0.57**
RF_2	0.71	0.80	**0.57**
RF_3	0.58	0.90	0.53
SVM_1	0.5	0.99	0.49
$XGBoost_1$	0.54	0.93	0.50
RF_4	0.71	0.71	0.50

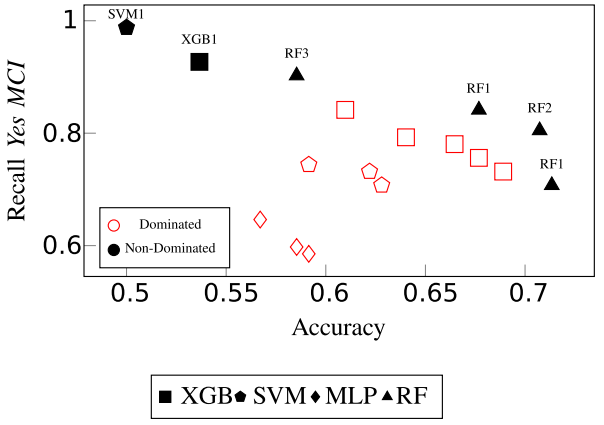


Fig. 4.7 Overall models' hyperparameter configurations in the objective function space Accuracy/Recall on YES MCI class. The black marks refer to Pareto-optimal solutions. The red marks represent dominated solutions. The marks' shapes indicate a particular class of model.

Table 4.8 Results for the MCI prediction of the Best Classifier.

Model	Class	Precision	Recall	F1-score	Accuracy
RF_1	No MCI	0.76	0.51	0.61	0.68
	Yes MCI	0.63	0.84	0.72	
RF_2	No MCI	0.76	0.61	0.67	0.71
	Yes MCI	0.67	0.80	0.73	

4.5.5 Results Interpretation and Impact

In this study, we propose a Pareto-based approach to identifying the best model for predicting MCI. Specifically, we trained and tested various machine-learning algorithms in order to choose the best one. From our perspective, the Pareto-based approach is relevant because it allows to select the best model according to multiple objectives. In this way, the identification of the best model does not solely depend on the overall accuracy of the model but also on other criteria tailored to the considered task. Practically, in MCI systems, we are able to select a model with both acceptable overall accuracy and effectiveness in classifying MCI subjects, which is crucial in medical tasks. Hence, this approach could be used to select the best machine learning algorithms for predicting high-risk classes that can be used in epidemiological studies or screening. The aim is to provide a smaller set of easy-to-find features in health datasets to predict MCI, thus reducing time and waiting lists for more extensive instrumental examinations.

The Random Forest considered achieved an Accuracy = 0.71, Precision = 0.76, Recall = 0.61, F1-Score = 0.67 respectively in the class *No MCI* and Precision = 0.67, Recall = 0.80, F1-Score = 0.73 respectively in the class *Yes MCI*. These results appear to be good in the current state of the art [35, 101]. However, the metrics achieved refer to a small population sample. Possible implementations could be aimed at validating this system on a larger test sample. The model selected by the Pareto-based approach was able to well discriminate subjects at risk of MCI.

4.6 Summary

To summarise the contribution of the reported work, our studies, based on the Salus in Apulia population, highlighted distinct dietary patterns and their clinical associations among diabetic and non-diabetic subjects, as well as the role of ML in identifying

complex features related to physical frailty. Among diabetic subjects, the diet was characterized by a high intake of fruits, vegetables, dairy products, eggs, greens, nuts, and olive oil, with limited consumption of red and processed meats. This dietary pattern aligns with a more "Vegetarian" style, likely influenced by nutritional counseling interventions, underscoring the importance of nutritional education as an effective tool in both primary and secondary prevention for diabetic patients. In contrast, non-diabetic individuals showed a higher consumption of red and processed meats, ready-to-eat meals, and high-calorie beverages, indicating a need for greater nutritional guidance within the healthy population as well.

Based on these findings, our focus expanded to using ML techniques on the Salus in Apulia data to predict MCI in an elderly population. We developed a Pareto-based approach for model selection, allowing us to identify the most balanced model by optimizing multiple objectives beyond simple accuracy—a key requirement in complex medical tasks. This approach enabled the identification of an optimal model with both acceptable overall accuracy and effectiveness in classifying MCI subjects, demonstrating a practical method for supporting high-risk class predictions in epidemiological studies and screenings. Further, leveraging ML to analyze dietary, clinical, and cognitive data enabled us to detect associations that traditional methods might miss, particularly in identifying features associated with both physical frailty and MCI. This approach not only validated classic associations in dietary and health studies but also highlighted the potential of AI to uncover early indicators of neurodegeneration, broadening the preventive scope to encompass both physical and cognitive decline. By incorporating ML in the study of MCI, we were able to model subtle patterns and risk factors that contribute to cognitive impairment, creating a basis for more proactive screening in at-risk populations.

These findings underscore the interdisciplinary nature of physical frailty and cognitive decline, suggesting that ML offers a powerful tool for early intervention across these domains. Future studies should focus on optimizing model characteristics, refining training parameters, and testing on larger datasets to enhance prediction accuracy, thus paving the way for more tailored dietary, clinical, and cognitive guidelines across populations. With continued refinement, these ML-based frameworks have the potential to support improved health outcomes by identifying critical risk factors in both physical and cognitive health.

Chapter 5

Enhancing Devices for EEG Signal Analysis with AI

5.1 Research Objectives and Contributions

The integration of AI with BCIs technologies represents a significant frontier in neuroscience, offering transformative potential for both diagnostic and therapeutic applications in healthcare. This chapter explores how AI-powered BCIs can advance emotion recognition, personalized neurorehabilitation, and therapeutic tools.

5.1.1 Research Objectives

Chapter 5 explores the advanced application of AI in neuroscience, focusing on the integration of BCIs and machine learning techniques to enhance diagnostic and therapeutic tools in healthcare. The primary objectives of this chapter are to develop real-time emotion recognition models powered by AI for improved human-machine interactions, create innovative therapeutic solutions that combine BCIs with natural language models for emotional support, apply adaptive algorithms for personalized neurorehabilitation, and advance research in affective computing through the NeuroSense dataset. The dataset, leveraging low-cost, sparse electrode BCIs, facilitates emotion exploration and promotes accessible and extensive research.

5.1.2 Addressing Current Literature Gaps

Despite advancements in BCI-based emotion recognition and neurorehabilitation, several gaps remain.

1. Cost-effective EEG Setups for Emotion Recognition: Conventional approaches rely on expensive, dense EEG systems, limiting practicality. Our framework demonstrates that accurate emotion detection can be achieved with low-cost, sparse EEG setups, utilizing refined feature extraction and robust models, thereby broadening accessibility.

2. Real-world BCI Applications: BCI technologies are often restricted to controlled environments, with insufficient exploration of real-world adaptability. We address this by developing adaptive models that function effectively beyond laboratory conditions, incorporating real-time emotion recognition and the integration of BCIs with Large Language Models (LLMs) for practical emotional support.

3. Therapeutic Integration of BCI: Traditional BCI research is limited to neurofeedback or basic communication. Our work bridges this gap by merging BCIs with AI frameworks for mental health, offering solutions like EEG-driven emotion recognition systems coupled with conversational agents for personalized support.

4. Music Generation for Therapy: BCI and ML applications in music have focused on artistic expression rather than therapy. Our system translates EEG signals for music composition into a therapeutic setting, offering novel interventions that improve emotional and psychological well-being.

5.1.3 Chapter Contributions

Chapter 5 makes several impactful contributions to the field. We have improved the precision of emotion recognition systems using deep learning models integrated with BCIs, enhancing user interaction through adaptive and responsive feedback. The ARIEL framework exemplifies the convergence of BCIs and LLMs, providing effective emotional support in therapeutic contexts and showcasing a new frontier for mental health technology.

The NeuralPMG system introduces a unique approach by using neural patterns to generate personalized music, facilitating novel therapeutic methods for neurorehabilitation. This contribution highlights the integration of emotional and physical state data in creating immersive and effective therapy experiences.

Finally, the NeuroSense dataset fills a crucial gap by offering a low-cost, sparse electrode solution for BCI research, democratizing access to emotion recognition technologies and encouraging broader participation and innovation in the field.

These advancements collectively underscore the transformative potential of AI and BCIs in healthcare, aiming to improve both clinical practices and research accessibility.

5.2 Background

To provide a clearer understanding of the following studies, it is essential to establish a comprehensive background and review the state-of-the-art relevant to our research themes.

Emotion Recognition Using AI Techniques

Automatic emotion recognition is a critical field that serves a myriad of applications such as automotive industry safety features, business decision-making supports, and robotic empathy interactions. The most precise methods for recognizing emotions typically involve processing images and videos through convolutional neural networks. Studies by Rakshit *et al.* [231], Dhall et al. [75], Zhang *et al.* [308], and Wang *et al.* [297] have explored various aspects of facial expression recognition critical to these applications. With the advent of improved computational capabilities and network designs by Szegedy *et al.* [279], He *et al.* [119], and Li *et al.* [164], the accuracy of these systems has significantly improved.

Furthermore, Hazourli *et al.* [118] highlight the transition of emotion recognition research from laboratory settings to real-world environments. While facial expressions can sometimes be masked, involuntary biological signals captured by biosensors provide a consistent basis for detecting emotional states, as explored by Menard *et al.* [193] and in the International Affective Picture System by Lang *et al.* [160] and Bradley *et al.* [33].

Brain-Computer Interface (BCI)

BCI technology has made significant advances in integrating EEG data for emotional recognition. The work of Atkinson *et al.* [17], Islam *et al.* [130], and Zheng *et al.* [311] demonstrates how novel feature-based models and sophisticated classifiers can greatly enhance the accuracy of emotion classification systems. These advancements underline the potential of BCIs to interpret cognitive states effectively, paving the way for applications that range from adaptive interfaces to mental health monitoring.

Emotional Support Conversation with Large Language Models (LLMs)

The integration of LLMs for emotional support has seen innovative applications, as demonstrated in the work of Tu *et al.* [287], who designed the MISC model, and Peng et al. [221], who explored how BlenderBot can be adapted for emotional support. Liu *et al.* [167] furthered this by integrating multi-modal information into their FEAT model for emotional support conversations. These developments, however, highlight the limitations of training data scope and the challenge of adapting to the variability of real-world interactions, a critical area for future research as noted by De *et al.* [72].

BCI and ML for Neurorehabilitation

The fusion of BCI, Virtual Reality (VR), and ML has been transformative in neurorehabilitation. Technologies that interpret and respond to neurophysiological signals during rehabilitation tasks, as detailed by Parivash *et al.* [214], McMahon *et al.* [192], and Karacsony *et al.* [141], have shown promising results in enhancing patient engagement and therapeutic outcomes.

BCI in Music Generation

The creative intersection of BCI and music has been explored by pioneers such as Rosenboom [238], Lucier [175], and others like Knapp *et al.* [152] and Brouse *et al.* [38]. These efforts have led to innovative systems that transform EEG signals into musical outputs, enhancing both artistic expression and scientific exploration.

EEG Emotion Recognition Databases

The development of EEG-based emotion recognition has been greatly aided by databases such as DEAP [153], MAHNOB [266], SEED [310], MPED [268], and DREAMER [144]. These resources provide a rich foundation for validating and improving emotion recognition technologies.

Table 5.1 Summary of published EEG databases for emotion recognition/classification.

Dataset	EEG Electrodes	Participants	Stimuli
DEAP [153]	32	32	Music videos
MAHNOB [266]	32	30	Movie clips
SEED [310]	62	15	Film clips
MPED [268]	62	23	Emotion videos
DREAMER [144]	14	23	Audio-visual clips

The summary of insights from these studies provides a robust background for the understanding and advancement of the technologies discussed in this chapter.

5.3 Deep Learning Applied in Human Emotion Recognition

The growing interest in affective computing and brain-computer interfaces (BCI) has fostered advancements in emotion recognition through EEG signal analysis. In particular, recognizing and quantifying emotional states such as Valence, Arousal, and Dominance holds significant potential for applications in neuromotor rehabilitation and

psychological therapies. This work introduces a prototype regression-based emotion recognition system capable of detecting a user's emotional state in real-time through EEG-based biofeedback. A practical focus of this system is to provide a real-time, interactive feedback loop, making it a promising tool for clinical trials and other therapeutic settings such as neuromotor rehabilitation or psychological therapies.

5.3.1 Material and Methods

BCI Device Description

In the domain of emotion recognition related to electroencephalographic signals, several studies based on multiple datasets and reference devices exist in the literature. We remember some of them, such as DEAP, Dreamer, and IDEA [139, 144, 153]. In this study, we focused on the Dreamer dataset [144] because it was made with a device that is easy to find and easy to use, such as the Emotive Epoc with 14 electrodes[1]. The device used has the following electrodes: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, plus two mastoid references, M1 and M2. In Fig. 5.1 we can observe the arrangement of the electrodes according to the standard configuration of the system 10-20 [199].

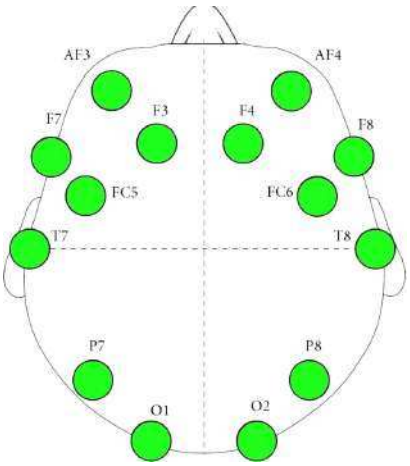


Fig. 5.1 Emotive Epochs 14 Channels.

The sampling frequency of the device is 128Hz in output. One goal of our study is to further reduce the channels of the EEG signal acquisition device while continuing to obtain good predictive values. According to this goal, we used the Emotive - Insight 5[2] electrode device with its electrodes AF3, AF4, T7, T8, Pz with an output rate

[1] www.emotiv.com/epoc-x/

[2] https://www.emotiv.com/product/emotiv-insight-5-channel-mobile-brainwear/

always at 128 Hz. The two devices share the same technical characteristics but have a different number of electrodes. Specifically, we only considered the electrodes related to the emotion, only AF3, AF4, T7, and T8 because the Pz is not shared. However, with only 4 electrodes, we found the possibility of obtaining good prediction results with our 1D Convolution Deep Learning model. In Fig. 5.2 are the electrodes of the Emotive Insight 5 device are shown according to the standard configuration 10-20.

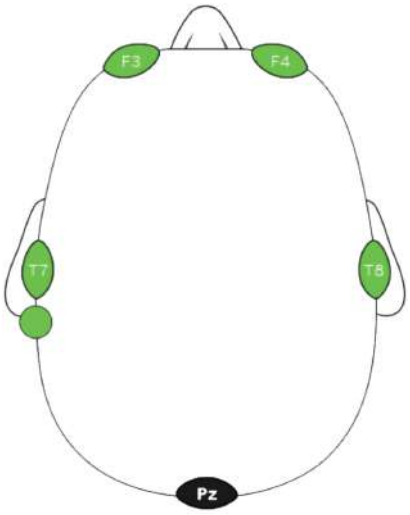


Fig. 5.2 Emotive Insight 5 electrodes.

Dataset Description

The Dreamer dataset is composed of EEG signals from 23 users during emotional induction. The emotional stimulation protocol was carried out using audio-video clips. 18 movie clips cataloged in nine basic emotions such as amusement, excitement, happiness, calmness, anger, disgust, fear, sadness, and surprise are used. Each user had to watch all the video clips of variable length between 65 and 393s. The self-assessment was given on a 5-point Likert scale for Valence, Arousal, and Dominance whenever the video clip ended. In order to perform this task, the participants fill out the Self-Assessment Manikin questionnaire at the end of each experiment [34]. In the dataset both the recordings without emotional elicitation (baseline) and the recordings during the emotion induction are collected. Finally, the authors of the study released the dataset in the Matlab format[3].

Preprocessing

One of the big problems with EEG signals is the strong presence of artifacts (noise) or faulty EEG channels that can seriously compromise the analysis of the data. An

[3]https://it.mathworks.com/products/matlab.html

important part of our study is the development of an automated preprocessing step in order to create an easy-to-use routine for capturing EEG signals in real-time. The preprocessing flow is crucial as our prototype aims to provide a real-time detection system of the user's emotional spectrum. The same preprocessing system is also used for preprocessing the Dreamer dataset for deep learning model training. We verified the efficiency of the automated preprocessing technique by visually inspecting all EEG trials.

Preprocessing Pipeline

- All the trials with related labels were selected, obtaining a total of 414 samples. Each trial is composed of signals coming from the EEG channels of the Emotive device and the corresponding values of Valence, Arousal, and Dominance.
- Once acquired all the trials, the first preprocessing step was to remove the Direct Current offset (DC offset) present in the raw data from the Emotive device. It has been used the script suggested by the helmet manufacturer[4].
- After the removal of the DC Offset, we move on making the data format compatible with the MNE [5] framework for adapting the skull electrode position according to the standard arrangement 10-20. In this way, the cap location was built in line with the electrodes of the device used in the study.
- The last 60 seconds are taken for each trial according to the study submitted directly by the authors of the dataset [144].
- The noise of commercial electric current are removed using a notch filter calibrated on the cutting frequency equal to 50hz.
- The whole trial was normalized in the 1-40Hz frequency range.
- From the continuous EEG signal, we have created epochs of length equal to 1 second.
- Before proceeding with the removal of ocular artifacts [166] through the Independent Component Analysis (ICA), all noisy epochs are removed controlling the deviation from the mean of the values for each channel, the amplitude of the epochs, the variance, and the distance from the mean.

[4]https://emotiv.gitbook.io/emotivpro-v3/notes-on-the-data/code-examples
[5]https://mne.tools/stable/index.html

- After that, ICA was applied for the identification of the components of the EEG signal. The AF3, AF4 reference electrodes are used as false EOG electrodes for the automatic recognition of components related to electro-oculographic artifacts (EOG). This operation allows us to automatically identify the artifact peaks related to the eye movements present in the signal.

- After removing EOG artifacts, a correction of the artifacts related to electromyographic patterns (EMG), ocular sacs, and any other artifact drifts is made through the Autoreject framework[6] with a particular setting of the hyperparameters. Specifically, a very high decimation threshold is applied.

- As a further step, the detection and interpolation of defective channels and epochs are done through the pyprep framework[7]. Epochs that exceed a certain noise threshold are, however, removed and not interpolated.

- As the last step, the continuous EEG signal is reconstructed by joining all the various epochs thus preprocessed. We add that some trials are removed if they do not pass one of these preprocessing phases, so after operation, the actual trials are 247.

Training Dataset

Once you get all the free trials from the artifacts, the dataset is built as follows:

1. Split into 4-second epochs of each continuous EEG trial.

2. All epochs are in overlap every 1280 samples. The new generated epochs have the same label as the originals.

3. For each epoch in overlap, five features are extracted related to the amplitudes of the bands Theta, Alpha, Beta1, Beta2, Beta3 with the framework neurokit2[8]. The Theta band is considered in the range 4-8Hz. The Alpha band in the range 8-13Hz, the Beta 1 in the 13-16 Hz, the Beta2 from 16-20Hz, and the Beta3 between 20-30Hz. In this modality, the dataset of 46,991 epochs with the relative labels of reference is created.

Model Description

[6]https://autoreject.github.io/stable/index.html
[7]https://pypi.org/project/pyprep/0.2.1/
[8]https://neurokit2.readthedocs.io/en/latest/

All the obtained features were initially split into train, validation, and test with the sklearn train_test_split library in proportion 80% for training and the remaining 20% for testing. The training dataset was then split into 75% for train and 25% for validation. After this operation, the normalization was performed with the MinMax scaler of sklearn [218]. The model used to make regression predictions is a 1D convolutional neural network (CNN) because it is useful in order to predict vectors of features at one size. The reference frameworks for the model are Keras [58] and Tensorflow [189]. The model consists of three convolutional layers, of which two to 128 neurons and a last to 64 neurons. A BatchNormalization was performed at the end of the first two layers of filters. Each layer was then condensed with the MaxPooling-1D in order to extract the most relevant correlation of engineered features. The kernel size is kept at 3, and the activation functions are Relu for convolutional layers. At the end of the convolutional layers, a Flatten operation is performed to create the input arrays for the next neural network. The neural network useful in order to predict regression values is a Fully Connected Layer composed of four layers, one of which is 128 neurons input, a second hidden at 128 neurons, and a third hidden layer at 32 neurons. The activation functions are relatively Tanh for the first two layers and Relu for the 32-neuron layer. Then a Dropout operation of 0.2 was performed in order to regularize learning to avoid overfitting. Finally, there is the last three neurons' output layer with linear output function useful for the purpose of the regression task. The three classes we want to predict are Valence, Arousal, and Dominance. During the learning, the mean_absolute_error was monitored as a loss function, and a callback was set to stop learning if the loss did not improve after ten iterations. (Patience = 10). The optimizer chosen is the Adam algorithm [150]. We have to schematize the structure of the network as in Fig. 5.3.

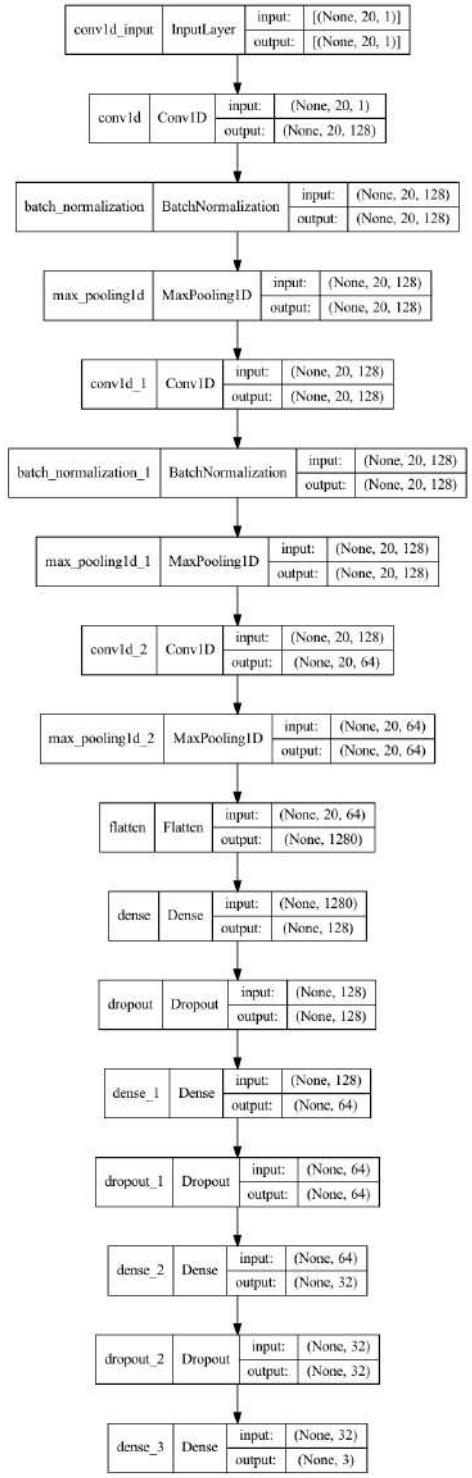


Fig. 5.3 Architecture of CNN-1D.

The main aim of our study is to provide an emotion recognition system that can provide real-time feedback on the user's emotional condition. In order to achieve this goal, the minimum length in terms of seconds was sought in relation to the greater level

of accuracy of the R2 metric. In practice, the minimum time that maintains the levels of accuracy above 0.9% of R2 was sought, achieving periods not less than 4 seconds. In the same way, the optimal overlap coefficient was chosen to maintain the value of the metric R2 not less than 0.9%. This optimization of the hyperparameters was carried out experimentally to directly find the best solution that could avoid considering the eras of the EEG signal not too long but neither too short. Assuming to use epochs of 1-second length, or 128 samples, is not representative of an emotional state. Increasing the length of the epoch, the value R2 increases, but consequently, it creates a problem relative to the time of scan of the signal EEG during the acquisition in real-time.

5.3.2 System Performance and Real-Time Emotion Recognition

The model thus trained achieves the following levels of predictive accuracy: R2 = 0.93 Mean Absolute Error = 0.08 Mean Absolute Percentage Error = 0.07 All metrics are calculated with sklearn.metrics.

in Figure5.4, 5.5, 5.6 is detailed the error distribution for each considered emotional metrics:

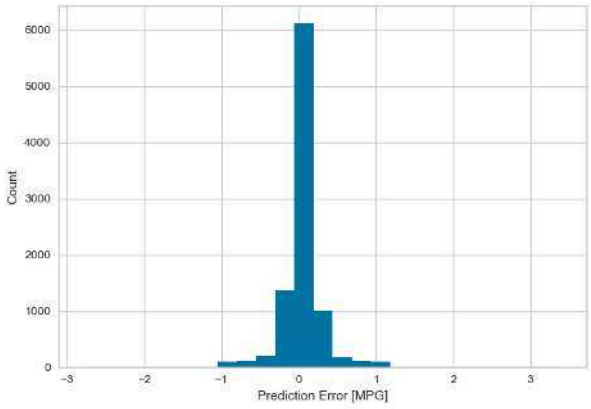


Fig. 5.4 Valence Error Distribution.

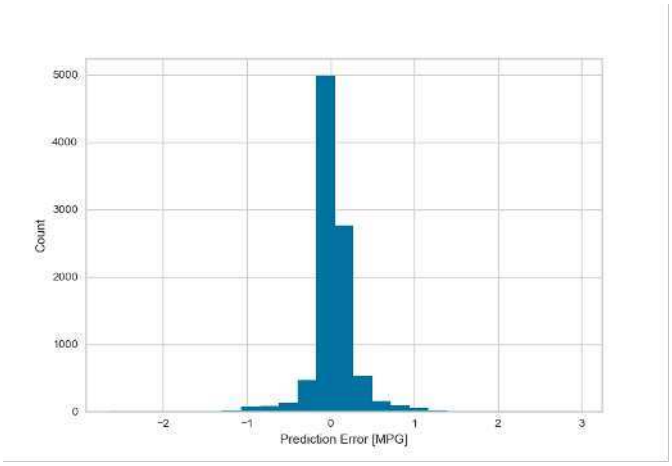


Fig. 5.5 Arousal Error Distribution.

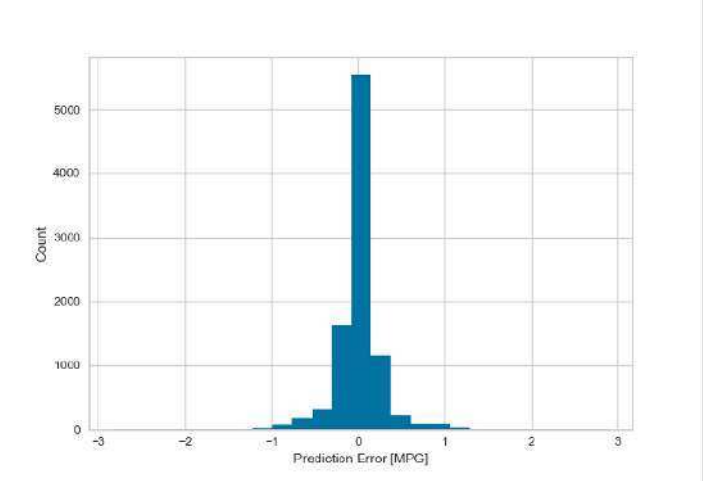


Fig. 5.6 Dominance Error Distribution.

Fig. 5.7 shows the course of the function of loss during the learning phase.

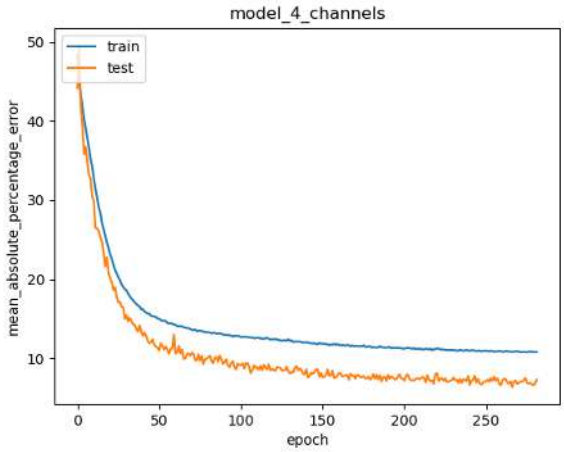


Fig. 5.7 Loss – Mean Absolute Percentage Error.

Our emotion recognition system aims to recognize them in real-time, so it must provide continuous visual feedback of how the user's emotion varies in the domain of time. Fig. 5.8 shows the user interface

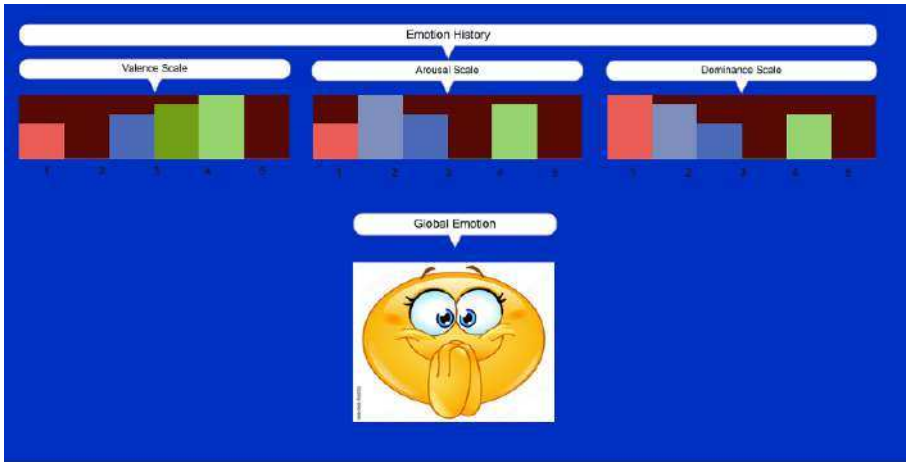


Fig. 5.8 User Interface.

The whole Front End is built in Max Msp[9], a software development environment mainly oriented to the development of applications in the domain of music but which offers great potential for the development of any stand-alone application. Fig.5.9. shows the full system architecture

[9]https://cycling74.com/products/max

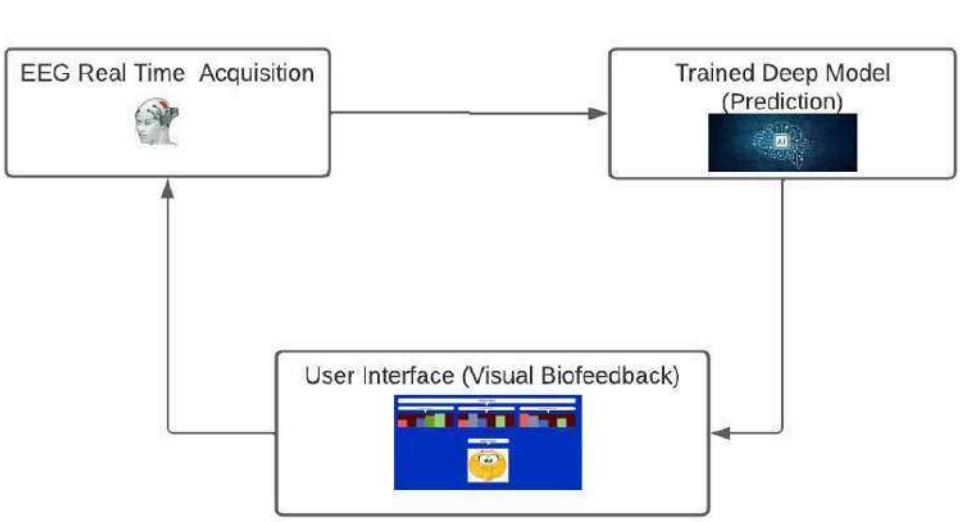


Fig. 5.9 System Architecture.

In Fig. 5.9 the user can continuously monitor their emotional state in order to be able to search, in real-time, to change it in the desired polarity. The steps necessary for the operation of the system are as follows: The user wears the device with the five electrodes. The EEG signal is sampled and preprocessed in real-time to make it conform to the subsequent analysis.

- The prediction of the values of Valence, Arousal, and Dominance on the EEG epochs that have passed the preprocessing flow, so we will get different prediction values in relation to the epochs analyzed. Remember that each epoch is 4 seconds.
- The prediction results are sent to the user interface, which in the back end performs the distribution of the received valence, arousal, and Dominance values. The distribution of these values is performed through a Spatial Model Encoding through the object ml.spatial[10] of Max MSP. The spatial encoder is a form of neural network that takes integer tokens and creates a vector that encodes the input sequence. The ml.spatial represents a form of "short-term memory" network where each vector element indicates the recursiveness of the various tokens. In this way, as you receive the input prediction data, you can see which classes are most present and which tend to decrease. The global emotion is instead calculated as the average of the various tokens in the time domain. For the assignment of the four emotional states, we refer to the circumflex model of Russel, which describes emotions as interpolation of the values of Valence and Arousal. Fig. 5.10 shows a schematization of the Russel model [42].

[10]https://cycling74.com/articles

Fig. 5.10 The schematization of the Russel's Circumplex Emotion Model

In Table 5.2, we observe the threshold for the allocation of the global emotional state.

Table 5.2 Global Emotion Parameter.

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

5.3.3 Prototype Evaluation and Implications

In this study, we developed a prototype of emotional bio-feedback based on EEG signals that can recognize human emotions. In particular, we have trained a deep neural network to classify valence, arousal, and dominance levels. From our point of view, classifying human emotions dynamically through the CNN-1D and the spatial encoder ml. spatial allows us to observe them better over time. From this perspective, we realized the possibility of observing the emotional history of a subject in real-time. This aspect makes our research original. The prototype could be used in epidemiological studies or screening. The purpose is to provide a real-time multimedia helpful stimulus as a feedback system for both the user and the clinician. The Convolutional Neural Network reached an R2 of 0.93, Mean Absolute Error = 0.08, and Mean Absolute Percentage Error = 0.07. These results are very good according to the current state of research. However, the high metrics reached refer to a small population sample. Possible implementations could be aimed at validating this prototype on a more

extensive test sample. In conclusion, these prototypes could be a good support tool in the clinical field and large-population studies.

5.4 ARIEL: Emotional Support Conversations with Brain-Computer Interfaces and Language Models

Building upon our previous work in user emotion recognition through EEG signals, which demonstrated the feasibility of detecting and classifying emotional states in real-time, this study extends those findings to explore interactive applications utilizing emotional feedback. The recent advances in Brain-Computer Interfaces (BCIs)—specifically devices that interpret electroencephalographic (EEG) signals—provide a minimally invasive and cost-effective method for monitoring user emotions with accuracy and reliability. These developments have paved the way for **ARIEL**—an **emotionAl suppoRt bcI dEvices and Llm-based conversational agent** designed to engage users in emotionally supportive conversations by harnessing real-time emotion data.

ARIEL integrates BCI-driven emotion recognition with Large Language Models (LLMs), enabling it to dynamically perceive and respond to users' emotional states, thus creating an interactive, emotionally aware experience. Leveraging the LLaMA 2 chat model, ARIEL adapts its conversational approach in response to real-time emotion analysis, fostering a supportive environment that actively assists the user's emotional well-being. Through a carefully structured hard-prompt strategy, the system ensures coherent, empathetic dialogues tailored to each user's emotional evolution, making the interaction both responsive and personalized.

The integration of BCIs in affective computing underscores the potential for EEG signals to monitor and interpret emotional states, especially given the growing need for emotionally responsive technology in fields like neurorehabilitation and psychological support. Despite substantial advancements in AI and large language models, many automated conversation systems still lack the nuanced emotional understanding required for meaningful support. ARIEL seeks to address this gap by uniting BCI-based emotion recognition and conversational AI to achieve the goal of the Emotional Support Conversation (ESConv) task—helping users alleviate emotional distress through dialogue.

Our contributions can be summarized as follows:

1. **A novel framework** combining EEG-based emotion recognition with LLMs, providing a system capable of real-time emotional support in conversational settings.

2. **Detailed design and implementation** of the ARIEL framework, describing the workflow and components necessary for sustained emotional support.

3. **Efficacy demonstration** through a controlled in-vitro experiment, showing ARIEL's functionality and highlighting its potential for in-vivo clinical applications.

Looking forward, we aim to conduct in-vivo experiments to further evaluate ARIEL's functionality, examining its potential applications within clinical and support-based environments. This innovative combination of affective computing and language modeling marks a significant step toward developing emotionally aware conversational agents with promising applications in mental health and emotional support contexts.

5.4.1 ARIEL Framework

The task of emotional support conversation can be generally divided into two subtasks: i) the modelling of users' psychological factors to detect their emotional state (i.e., emotion recognition task - ER) and ii) the design of a conversational mode able to have natural language dialogues (i.e., conversational agent - CA) with the final aim to generate supportive responses. Although multiple Sentiment Analysis techniques allow today for detailed identification of emotional aspects hidden within a text [24, 299], which helps to infer the emotional state of the person who wrote it, they are effective only when dealing with linear sentiment [74], failing the emotion recognition task otherwise. Moreover, the development of CAs strictly dedicated to generating supportive dialogues suffers from the overfitting problem. Specifically, they may be unskilled in talking about topics outside the addressed task [1], typical of human interactions, which would facilitate user emotional support.

To solve such issues, we design the ARIEL framework by taking advantage of the recent developments in the AC field through BCI devices and the unprecedented linguistic and cognitive properties characterizing the most recent LLM systems in having conversations with users. Figure 5.11 illustrates ARIEL's architecture and its workflow modelled to address the Emotional Support Conversation (ESConv) task, which mainly relies on four components: the Neuro-Linguistic Interface, the Emotion

Recognizer, the Prompt Formatter, and the LLM.

Neuro-linguistic Interface

The Neuro-Linguistic Interface identifies the principal component with which users can interact with the ARIEL framework. It allows information gathering from actions undertaken by users who decide to interact with our system, looking for supportive conversations. Such data is mandatory for ARIEL to accomplish the ESConv task, which exploits the Neuro-Linguistic Interface to return, in turn, the computed responses to users. In a few words, this component is essential to make people communicate with the framework and vice versa.

The input on which our framework works is two-folded, spanning over two different information channels: the EEG signal and natural language. The former is acquired through a BCI device, which measures the brain's electrical activities in a non-invasive way. This signal conveys more emotional information from humans than other kinds (e.g., linguistic data, facial expressions). The brain's more superficial areas are closely connected to emotions, with a higher concentration of electrical activity when feelings or emotional states are experienced. Therefore, it is a valuable source of information to detect users' emotional condition and act accordingly. The latter, instead, brings to the system what users want to communicate explicitly in textual form. Such messages may also include users' sentiments through descriptions or writing styles, although they can hardly be considered reliable. Nevertheless, textual messages are handled by the homonym sub-component, which allows the system to know the topics on which users prefer to talk, besides making possible dialogues.

The system consistently collects EEG signals starting when the user initiates interaction. In contrast, it exclusively registers chat messages upon the user's successful transmission through the Neuro-Linguistic Interface.

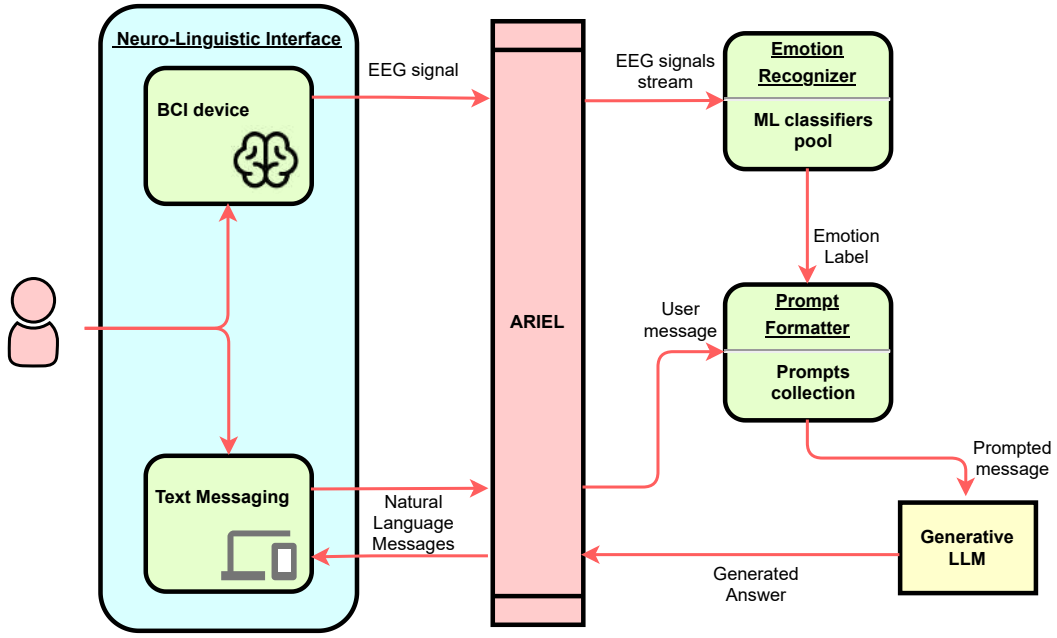


Fig. 5.11 The figure schematizes the overall workflow followed by the ARIEL framework. On the left side, the user interacts with ARIEL through the Neuro-Linguistic Interface (a) component, which contemporarily enables Text Messaging and measuring EEG signals through a BCI device. On the right side, the Emotion Recognizer (b) receives a stream of EEG signals on which several ML classifiers are asked to infer the emotion label that describes those signals. Such a label feeds the Prompt Formatter (c) with the user message. The (c) component wraps up this information within the most suitable prompt based on the conversation state. The LLM (d) is queried with the selected prompt, generating a supportive response delivered to the user for continuing the dialogue. The interaction ends when the user reaches a positive emotional state and leaves the conversation.

Emotion Recognizer

The Emotion Recognizer is a component that analyzes the EEG signals measured via BCI to infer the actual emotional state of the user. Specifically, it receives a stream of EEG signals, which is adequately framed based on the dialogue's timing users have with the system, from the Neuro-Linguistic Interface. These signals are processed and feed machine learning models to infer a label indicating the user's emotional state, which is essential to have a clear picture of the user's current state and act accordingly. To achive this goal, we rely on a novel approach granting high trustworthiness regarding the emotion recognized, which we describe below in tripartite paragraphs related to its founding elements.

Dataset Description

In this study, the Deap [153] dataset protocol approach was employed to collect data to train machine learning models based on the emotion state classification. The acquisition

process was conducted at the *Polytechnic University of Bari*, taking into account a student population with an average age between 20 and 30 years. Participants were informed about the procedure and provided written consent for the collection of EEG data before the experiment. Participants' EEG signal was acquired using the Muse 2 EEG device[11] to perform an accurate data acquisition. During the experiment, a series of 40 music videos was shown to each participant, and simultaneously the EEG data were continuously recorded. All collected EEG data have been anonymized to ensure participants' privacy, and stored securely and accessible only to the researcher team. The final dataset is organized as follows: (i) 30 subjects collected and divided into several folders. (ii) Each folder contains 40 EEG trials organized into session subfolders.

Preprocessing

The preprocessing pipeline is outlined below: EEG epochs duplication and application of a bandpass filter to isolate frequencies between 1 and 40 Hz was carried out as the first step. After filtering, the data are transposed and amplitude scaled to simplify computation. A K-nearest neighbors (K-NN) classifier, was trained on the filtered dataset to calculate anomaly scores in order to identify EEG data artifacts if a predefined threshold is exceeded. All artifacts detected by the threshold-based approach are subsequently removed by exploiting the megkit [21] framework using the ringing artifact reduction method. Cleaned EEG data are then resized, restructured, and encapsulated in an MNE[107] EpochsArray object, ensuring data consistency since it maintains names and channel types defined according to a standardized electrode fitting. Otherwise, if the EEG Signals are artifact-free, they are preserved in their original shape. This accurate preprocessing procedure is essential to preserve the integrity of the EEG data, increasing the reliability of the results.

Machine Learning Models

In order to find the best model able to classify emotional states, different ML models such as Logistic Regression (LR), eXtreme Gradient Boosting (XGB), RF and, SVM are compared. In order to identify the best model, a hyperparameter tuning using GridSearchCV (5-fold)[12] was applied, with a Leave-One-Subject-Out (LOSO) cross-validation strategy[156]. This approach has led to the development of 120 specialized models. Specifically 30 ML models for each of 4 emotions (Happy, Angry, Sad, Relaxed) able to predict by majority vote the outcome resulting from the quadrants of Russell's

[11] https://choosemuse.com/products/muse-2

[12] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

complex circumference. After comparison of the various ML models, the best model in predicting Happy, Angry, Sad and, Relaxed emotions was SVM achieving an accuracy of 0.77, 0.78, 0.76, 0.80 respectively for the considered emotions.

Prompt Formatter

ARIEL introduces also the Prompt Formatter component, a crucial element within our conversational system that orchestrates the initiation and progression of dialogues with a Large Language Model (LLM). The Prompt Formatter serves as the gateway through which the user's emotion label, found by the Emotion Recognizer, and her sent message are amalgamated into coherent prompts, setting the stage for meaningful interactions. A prompt is essentially a piece of natural language text given to the model by the user, serving as the initial input or question. This input initiates the model's response process and guides its direction and scope, acting as a catalyst for the output's relevance and utility [307].

At the heart of the Prompt Formatter lies a dynamic prompting strategy designed to optimize dialogue generation in handling the ESConv task. The strategy hinges on the selection and adaptation of prompts tailored to different stages of the conversation. Initially, the Prompt Formatter employs a role-play prompt to kickstart the interaction, providing the role the LLM has to play towards conversations as a context. In this case, we assign to the LLM the role of a virtual assistant specialized in supporting users' mental distress to evolve their emotional status to a positive condition. Furthermore, we also let the LLM know that the user's sentiment is measured with BCI sensors, favouring the LLM in considering possible mismatches between what the user says about her emotional state and what he actually feels. Below we report the adopted prompt for enabling the LLM to initiate the conversation.

Conversation start role-play prompt

You are a Virtual Assistant designed to converse with users supporting their emotional distress and transforming it into a positive feeling. You receive as input also information about users emotional state measured through an high reliable emotion recognition model. Just have a chit-chat conversation until the user feels happy or relaxed and wants to leave the conversation. The emotion recognition model measures the user feels {emotion label}.
The user sent {user message}.

As the dialogue unfolds, the prompt evolves in response to the ongoing discourse, ensuring relevance and coherence while leveraging the LLM's contextual understanding. Specifically, the prompt evolution is determined step-by-step through the conversation by selecting a prompt from a collection designed to take into account possible conversational states. Moreover, all the prompts include growing dialogue histories

to avoid information loss. This adaptive prompting mechanism not only sustains the coherence of the conversation but also facilitates the LLM's comprehension and response generation, leading to richer interactive exchanges.

Generative LLM

Finally, the generative LLM component identifies the core of the ARIEL framework. Based on the open weight LLaMA 2 chat model, this component exploits the language processing capabilities of the LLMs instructed to behave in a conversational manner. It serves as the virtual interlocutor, capable of understanding user inputs, processing contextual information, and generating coherent responses in natural language.

Large Language Models (LLMs) are based on advanced neural networks designed for natural language processing tasks. At their core, LLMs are probabilistic models designed to estimate the probability distribution of sequences of linguistic units such as words or sentences [239]. Upon receiving prompts from the Prompt Formatter component, the generative LLM builds responses based on the provided context and instructions. These responses are crafted to emulate human-like conversational behavior, incorporating linguistic nuances, contextually relevant information, and appropriate tone. Moreover, since it is coupled with prompting strategies described within the Prompt Formatter component, LLMs seamlessly integrate into conversational frameworks, effortlessly generating responses that align with the desired conversational goals.

5.4.2 ARIEL @ Work

With the growing sense of isolation and detachment people feel in today's scenario, compounded by the pervasive influence of digital communication platforms, ARIEL eavesdrops actively on the users' emotional problems and promulgates through conversations engaging discussions and chit-chat interactions to alleviate their mental distress. The ARIEL's goal is to let people talk about preferred and positive topics, opening up to confidence and problem expressions to relieve the experienced emotive pressure.

Below, we describe an example of a user interaction with the ARIEL system consisting of an AI-based emotion recognition framework and an LLM-based chatbot. Specifically, we have designed a running example immersed in one of the possible cases for which this tool is helpful, besides explaining how the ARIEL framework works in detail. Furthermore, Figure 5.12 contains an additional conversation example to ensure a better understanding of the functioning of the proposed framework. Such a controlled experiment resulted from a real session trial with the ARIEL system.

Eric is a young professional immersed in a stressful work environment. Particularly, in a very stressful period studded with relevant workload and imminent deadlines, at the end of his working day, Eric has limited time to engage in any social activity to relieve his accumulated tension. Therefore, to get relaxed Eric sits at his computer and wears a BCI device. With a habitual gesture, he starts the system called *ARIEL*, which immediately starts recording the EEG signals of his brain and using an AI-based Emotion Recognition algorithm to predict his emotional state in real-time. Indeed, this system represents an innovative interface between Eric's mind and the digital world, enabling a proactive understanding and reaction to his emotions. Based on the emotions predicted by ARIEL, a chatbot embedded in the system triggers a dialogue with Eric, designed to drive his emotional state towards a positive direction.

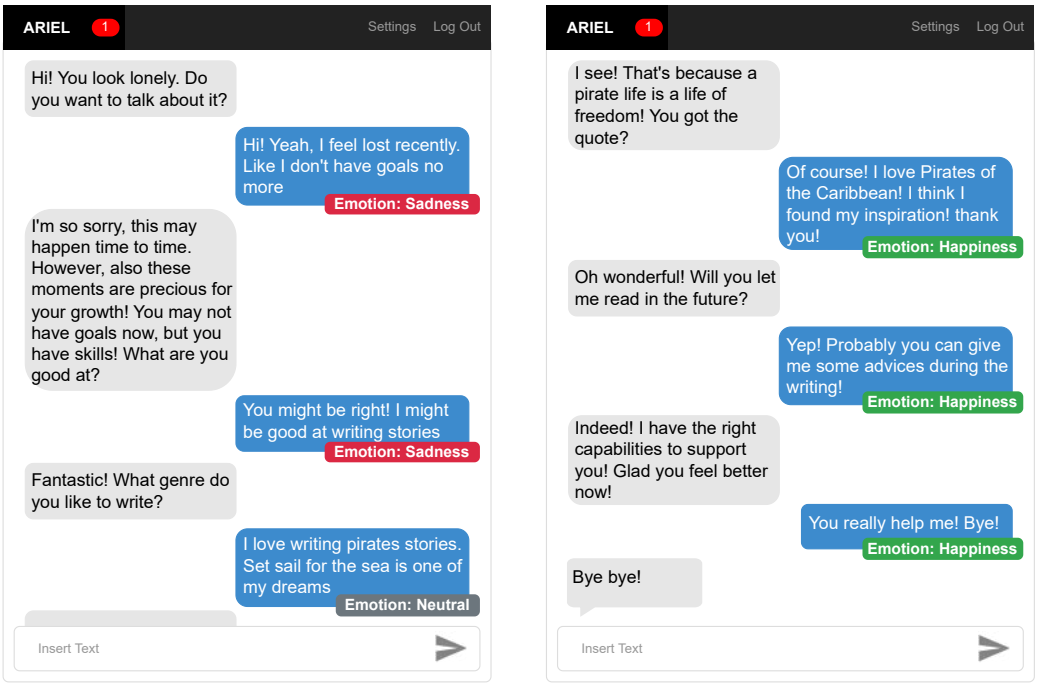


Fig. 5.12 The figure highlights a simplified conversation held by ARIEL with the user Eric. The first screenshot from the left captures the start of the conversation, where ARIEL engages Eric in talking about his emotional distress (grey balloon). Thus, Eric joins the dialogue with an answer describing what he feels (blue balloon) together with the emotion recognised by the Emotion Recognizer (cf. Figure 5.11) through the BCI device (red balloon). The two actors of the framework have a chit-chat conversation that brings the user's emotional state to evolve from sadness into happiness.

In this case, the ARIEL's emotion recognizer predicts that Eric is experiencing sadness. Hence, ARIEL's chatbot welcomes Eric and initiates working on the emotion felt by the user by encouraging them to have a light conversation with the message: "*Good evening Eric! I sense a certain sadness from you. What happened?*". Eric, who needs to talk to someone to relieve himself from daily worries, texts the system: "*Hi Ariel! Yes, my work is quite tiring these days, and deadlines certainly don't help me relax.*". Since ARIEL does not acknowledge a shift in Eric's emotional state, it tries to cheer up him by replying: "*I see! But do you know that after great effort there is always great satisfaction? You're working hard, you deserve it!*". Eric feels partly consoled by the message, but the great tiredness does not allow him to change his sadness to a positive feeling. Accordingly, Eric messages ARIEL: "*I know! However, I feel trapped in this routine. Tonight I would have liked to go and listen to a Pink Floyd cover band, but obviously, I can't.*". Although the overall emotional state did not have sensitive variations from the initial state, the BCI device measures a slight positive change in Eric's EEG signal that the emotion recognizer promptly reports to the ARIEL chatbot. Consequently, ARIEL continues the conversation by resuming the topic discussed in the previous message, which allowed the recording of this little positive perturbation (e.g., Pink Floyd). It answers Eric: "*I'm sorry! However, even if it is not the same, you can reproduce their songs to relax! I think you would love listening to "The Great Gig in the Sky"! Is one of my favourites! The vocal performance is exceptional!*". At this moment, Eric feels engaged by the conversation since the mentioned song is also one of his favourites. He replies: "*It is also one of my favourite songs! Clare was commendable and conveys many positive feelings!*". The emotion recognizer now detects a neutral state of the user from the initial sadness, which encourages ARIEL to continue talking about music. Through messages of encouragement, positive talk or joyful activities, the system will try to transform Eric's sadness into a more positive feeling, such as happiness. The conversational session will terminate whenever Eric reaches a positive emotional state and greets ARIEL, declaring the end of the dialogue.

ARIEL implements a dynamic approach for handling the ESConv task thanks to the Prompt Formatter. Depending on emotions predicted by the emotion recognizer, this component selects the most suitable prompt that guides ARIEL's LLM in conversing with users and accomplishing the emotional supportive mission. In this example, Eric has a bad day at work the following day, having had some discussions with colleagues. When interacting with ARIEL, the system predicts that Eric is experiencing anger. In this case, the chatbot will take a different approach. Therefore, ARIEL texts Eric: "*Hi Eric! Did you have a bad day? I feel your anger.*" Eric will then explain to ARIEL

his unfortunate day, and the emotion recognizer will detect his negative feelings. As a consequence, ARIEL engages in a conversation to manage nervousness issues, replying to Eric: "*What a pity! But first thing first, take a deep breath. Every problem has a solution. Let's find it together.*" During the dialogue, ARIEL will offer Eric further stress management techniques, such as calm reflection on his feelings, despite positively facing the emerging work issue. The aim is to help Eric calm down and regain his inner peace, gradually shifting his emotionality towards more neutral and ultimately positive states. Also, the whole conversation will be guided wisely thanks to the continuous monitoring of the user's EEG signals through the BCI device.

Whenever Eric starts a conversation with ARIEL while feeling happy or any other positive emotion, the emotion recognizer will detect that Eric is already in a positive emotional state, therefore the chatbot will keep the conversation light and joyful, without attempting to change his mood further. Instead, it might propose fun activities or encourage Eric to share his positive experiences, thus helping to consolidate his happy emotional state.

5.4.3 ARIEL Principles and Future Pathways

In this work, we presented a novel framework to handle the emotional support conversation (ESConv) task. ARIEL, an emotionAl suppoRt bcI dEvices and Llm-based system, embraces an innovative strategy in engaging users through conversations to alleviate their mental distress via the adoption of brain-computer interface (BCI) devices for reliably detecting their feelings. Our proposal is the first to take advantage of the recent discoveries in the fields of Affective Computing and Language Modelling, mixing BCI-based emotion recognition models and a large language model (LLM)-based conversational agent (CA) to effectively address the ESConv task. Indeed, ARIEL is composed of an emotion recognizer to detect the users' emotional states via BCI, an LLM to generate natural language messages given a text as the input, and a prompt formatter to enclose emotional labels and users' messages within different and dynamically selected prompts to guide the LLM in generating dialogues. One of the most interesting aspects of ARIEL is the high reliability in detecting users' emotions thanks to the electroencephalographic (EEG) signals acquired by the BCI device, showing effective functioning during our running example.

In future developments of this work, through a detailed user study, the impacts of this technology on mental health, emotional well-being and work performance will be analysed. Furthermore, the ethical and social implications of an increasingly intimate

interaction between man and machine in the field of personalised emotional assistance will be assessed.

5.5 Machine Learning for Neurorehabilitation via Mental State Recognition

The progression of our studies on the application of Machine Learning (ML) techniques to electroencephalographic (EEG) signals has led to a shift in focus towards neurorehabilitation and the potential of biofeedback systems to enhance therapeutic outcomes. Initially, our work concentrated on classifying users' mental states, such as *Focus* and *Relaxation*, through various ML models. This early phase established a solid foundation by achieving significant accuracy rates, specifically attaining 0.90 accuracy on the Test Set with the Support Vector Machine (SVM) classifier. However, recognizing the therapeutic potential of biofeedback-driven interactions, we have extended our investigation to explore how real-time emotional state detection via EEG signals could support neurorehabilitation processes.

This extension of our research is realized through the development of a prototype system that dynamically generates music based on detected mental states, categorized into *Focused* or *Relaxed* outputs. This biofeedback tool, leveraging BCI capabilities, allows the system to automatically adapt music generation to the user's detected mental state. Such adaptability could be highly effective in clinical neurorehabilitation scenarios, where tailored auditory stimuli play a vital role in enhancing motor recovery, emotional regulation, and cognitive engagement. Sound-based interventions have been shown to stimulate neuroplasticity and improve therapeutic outcomes in various clinical domains [237, 252]. By integrating these principles into a responsive system, we provide a novel approach to real-time rehabilitation support, driven by a biofeedback mechanism that encourages patient engagement and movement through personalized auditory cues.

The contributions of this work are threefold: (i) development of an ML-based biofeedback system using EEG signals to detect users' mental states (Focus and Relaxation); (ii) introduction of a real-time music generation component that adjusts according to the user's current state, thereby promoting focused or relaxed mental states as needed; (iii) demonstration of the system's applicability in neurorehabilitation, underscoring its potential to enhance patient outcomes through auditory biofeedback.

5.5.1 Methods

The system employs an approach based on new paradigms for polyphony generation by exploiting the user's Focus and Relaxation values. More specifically Polyphony generation refers to the creation of music by the coherent superimposition of four melodic lines.

The framework shown in Figure 5.21, is structured of 5 main steps. The *First Step* is based on the data acquisition by the BCI-Muse EEG Device [97][13]. Afterwards, the EEG signal is pre-processed during the *Second Step*. The *Third* and *Fourth Steps* are devoted to (i) the generation of the user's mental state classification values by the Machine Learning Engine and (ii) the generation of music based on the user's mental state with the visualization of the score. During the *Fifth Step*, a musician plays the generated music in real-time. Finally, the user hears and reacts to the generated music by trying to maintain the required concentration level. The system manages a biofeedback loop between the user-generated musical texture and the his mental state. A likely example scenario could be: Tom is a guy who needs to stimulate his motor skills. A specialized operator asks him to move in time with the generated music. The task is to generate high rhythmic music. To achieve this, Tom must strive to concentrate. The original music generated in real-time is performed by a music player. If the music is high rhythmically Tom will be induced to move more and enhance his movement. On the other hand, if the music is low rhythmically Tom will be induced to move slowly.

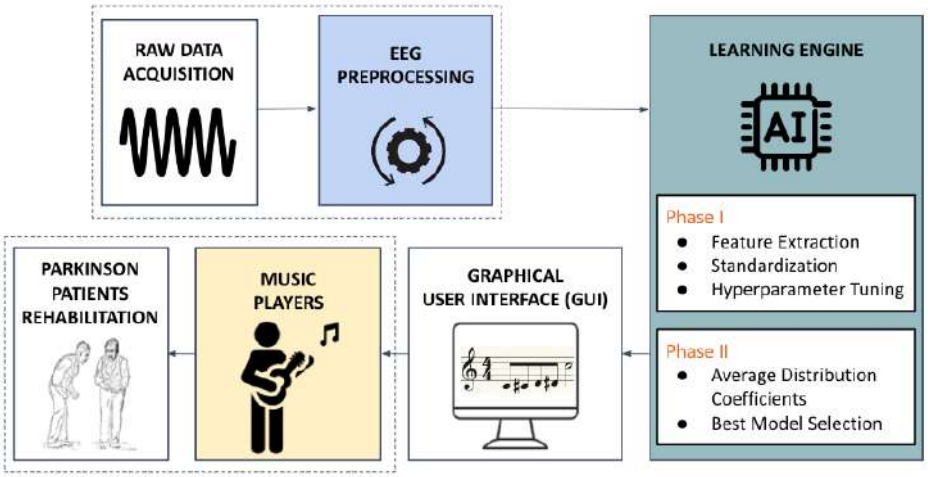


Fig. 5.13 NeuralPMG System Architecture.

Data Acquisition Device

[13]https://choosemuse.com/muse-2/

In this section, we describe the device *BCI-Muse*, used in user EEG signal acquisition. Muse device is a passive helmet with four electrodes capable of measuring electrical voltages on the head's surface (EEG). The silver electrodes are dry-conductive (conductive solutions are not required) and are placed according to the official $10-20$ system [122] in positions $AF3$, $AF4$, $TP9$, $TP10$ with reference on $FPz(CMS/DRL)$. The sampling frequency (fs) is 256 Hz. Figure 5.22 shows an example of electrode representation with relative positions.

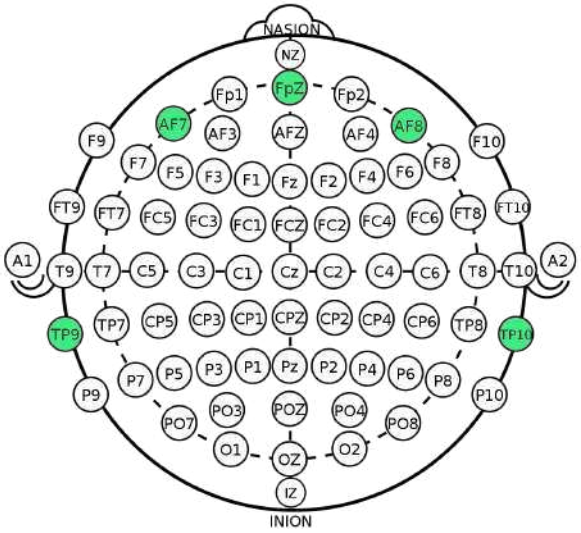


Fig. 5.14 Representation of Electrode placement according to the 10-20 system.

Polyphonic Structure Generation

The generation process starts by analyzing the result of the ML model prediction. Based on the classification result, a set of corresponding rhythmic values is chosen and then used to generate the rhythmic profiles for the four melodic lines. Rhythmic value samples are divided into the following two groups:

1. `Focus:` represented by the rhythmic values 1/4, 1/8, 1/16 and their corresponding rest values;

2. `Relaxation:` represented by the rhythmic values of 4/4, 2/4, 1/4 and their corresponding rest values.

Strings of the rhythmic pattern are created with random permutations between values belonging to the settings chosen by the classifier (0 for Focus, 1 for Relaxation). The melodic profile is generated through several steps based on the set of notes. The framework is able to compute all possible musical scales in an octave from 3 to 11 notes. An octave interval is defined as the distance between two musical notes, one of

which has twice the frequency of the other. The word musical scale means a subset of notes inside all 12 Equal temperament [127, 138]. To generate polyphony, the first steps are related to selecting the basic scale. Figure 5.15 shows the Graphical User Interface (GUI) of notes number choice for scale. All polyphony will be generated on the notes of the selected scale.

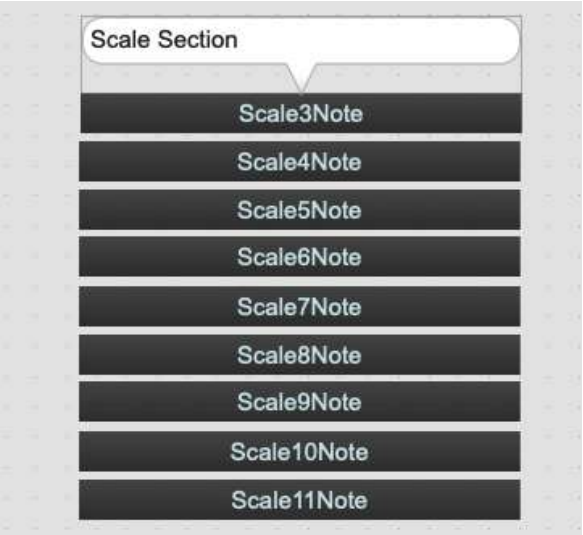
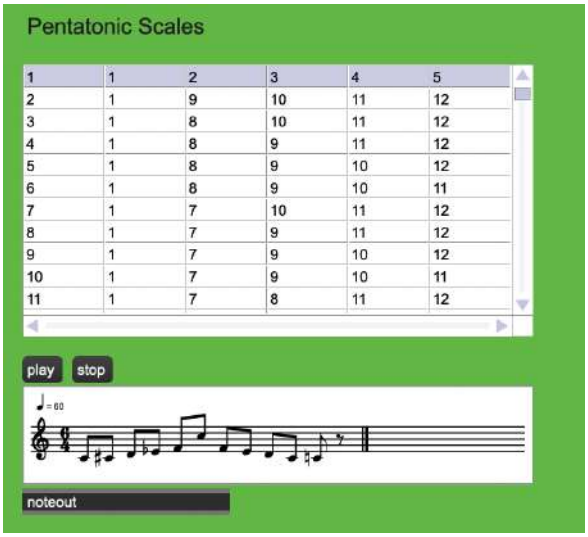


Fig. 5.15 GUI - Menu representation, allowing the user to select octaves.



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

Fig. 5.16 GUI - Example of submenu for all possible pentatonic scales in one Octave.

5.5.2 Dataset Description

In our system, the *'EEG brainwave mental state'* dataset [27] is used in order to apply the Muse device. The raw data are available in open source format[14].

The authors collected the EEG signals, from four users (two males and two females), related to the following mental states: (i) *focus* (ii) *neutral* and (iii) *relaxation*.

During the relaxation task, relaxing sounds were played, and users were asked to relax. Moreover, in the attention condition, users played a task that required them to pay attention: a ball was hidden under one of three cups. The goal of the test was to identify in which cup the ball was hidden. Figure 5.17 shows the user's EEG signal during concentration and relaxation tasks, respectively. Channels $TP9$ and $TP10$ were excluded because are noisy in whole dataset. Using only a few channels, it was not possible to perform EEG/Electrooculogram (EOG) artifact removal applying regression methods [76], ICA [293], or other standard techniques.

Figure 5.17 shows several problems related to noisy channels and the presence of ocular artefacts [284] related to the raw data. In particular, they show: (i) low-frequency artefacts (ocular) and (ii) high-frequency artefacts (muscle tension, etc.).

Therefore, each signal was filtered considering only the 7-14 Hz as wavelength. This technique excludes low-frequency artifacts (drift, head movements, EOG blink artifacts, etc.), as well as high-frequency artifacts related to noise environments, and muscle artifacts. Accordingly, the selected channels used in the analysis are the $AF3$ and $AF4$. Figure 5.18 shows the signal in the concentration and relaxation task after the filtering step.

[14]https://github.com/jordan-bird/eeg-feature-generation/tree/master/dataset/original_data

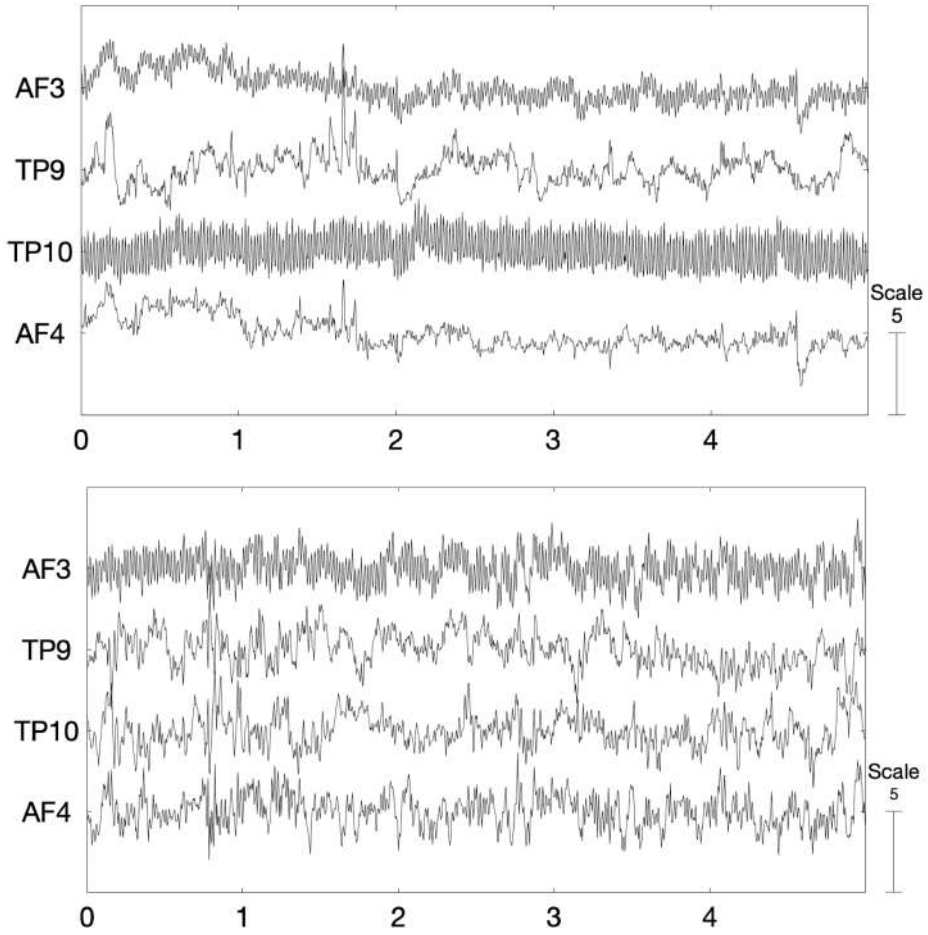


Fig. 5.17 Representation of EEG Raw Signal during the Concentrating (left) and Relaxation (right) tasks, before the filter operation.

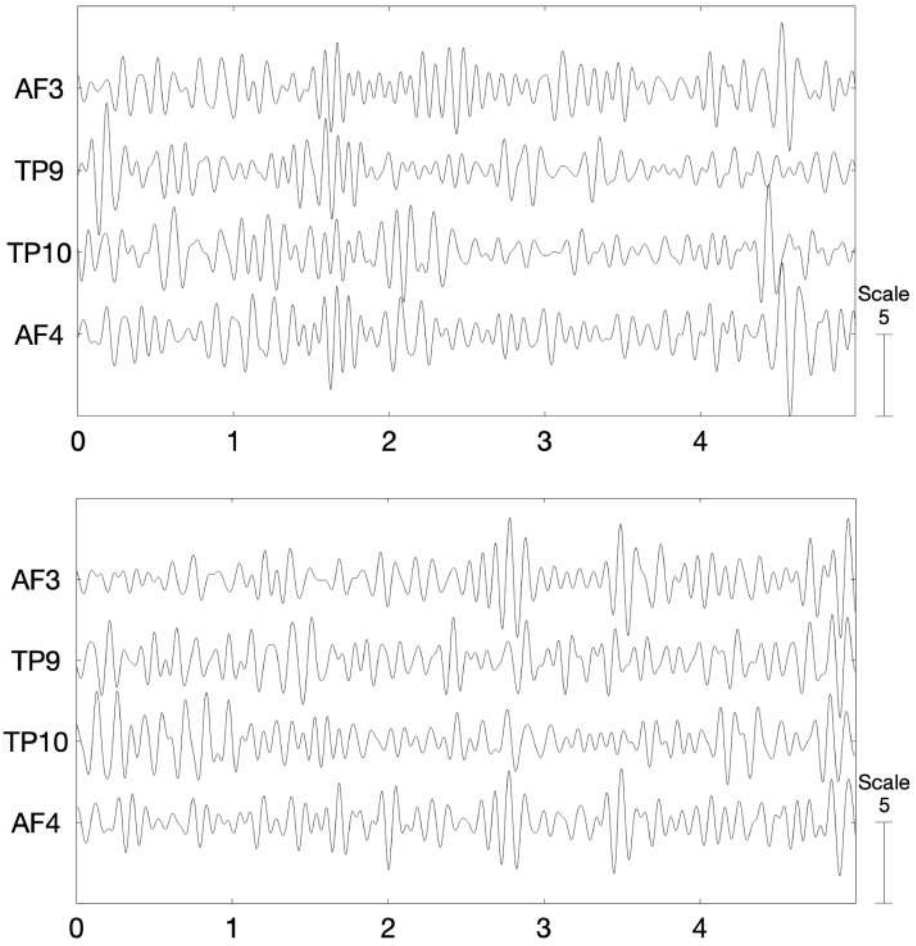


Fig. 5.18 Representation of EEG Raw Signal during the Concentrating (left) and Relaxation (right) tasks, after the filter operation.

5.5.3 Framework

The primary objective of the framework is to identify two mental states - activation and relaxation - through a ML classifier. After classifying the user's activation level, the system generates music in real-time based on the current mental state. The preprocessing and Machine Learning Engine operations are described in the following section.
Machine learning engine The ML approach is structured into two steps. In the first step, the MiniRocket [73] algorithm was applied. We adopted the implementation of the `sktime` [173] library. Briefly, Minirocket is a variant of Rocket, i.e., an algorithm for feature extraction from time series. Each time series is convolved by using random convolutional kernels, and global max pooling. The proportion of positive values pooling is applied to the convolutional output to produce two features per kernel. The extracted features contain information related to series class membership, which a

linear classifier can model. Hence, the first process involves feature extraction from each time series using a convolutional filter kernel. It extracts features from the multivariate EEG segment. These features are then input to the Ridge classifier, which we have selected as a linear method for the twofold goal of handling feature collinearity with L2 regularization and performing feature ranking [136]. The length of the features employs two hyperparameters: `Kernel size` and `Kernel length`. The hyperparameter phase of the whole process is carried out into Python Pipelines: (i) MiniRocket tuning (Kernel size and length), (ii) RidgeClassifier tuning (alpha values), and (iii) Normalization function selection (MinMaxScalar, StandardScaler, RobustScaler) in 5-fold cross-validation using Random Search (50 iterations) developed with Python, applying Scikit-learn library [219][15].

The whole process of choosing the best EEG segment and hyperparameter tuning is iterated with different lengths of the EEG signal. Determining the correct signal length is essential to obtain the highest amount of information from the signal in a real-time classification setting. From our proposed system, the maximum EEG signal length should not exceed 5 seconds to avoid it being too long and then not in real-time.

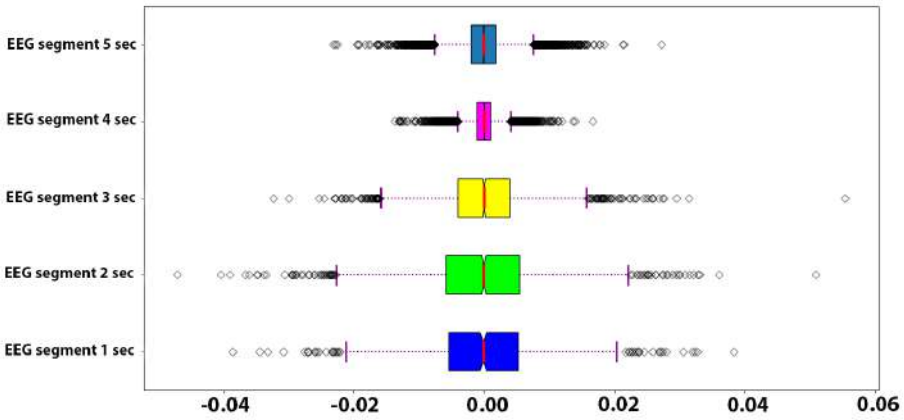


Fig. 5.19 Distribution of features coefficients averaged across the k validation rounds.

Table 5.3 shows the shapes of the data set in relation to the length of the EEG segments.

[15]http://scikit-learn.org

Table 5.3 Snapshot of the training dataset. EEG Segment is measured in seconds (sec).

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

Accuracy, F1 score, precision, and recall have been computed to assess the performance of the models.

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 5.19 shows the feature coefficients averaged across k-fold validation rounds. Afterwards, the best predictive model with the best EEG segment is selected. We used a statistical criterion to reduce the set of 3,864 features based on the selection of the most important 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, Random Forest, and KNN) with hyperparameter tuning and k-fold strategy.

Table 5.4 Results of the best models related to the EEG segment length. EEG segment is measured in seconds (sec).

EEG Segment Length	***Pipeline Steps: MiniRocketMultivariate***	***Pipeline Steps: Scaler***	***Pipeline Steps: RidgeClassifier***	***Precision***	***Recall***	***F1 Score***	***Accuracy***	***Confusion Matrix***
1	kernel max dilatations: 352 num kernels:1110	*RobustScaler*	alpha = 621.01 max_iter = 891 random_state = 42 solver =auto	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 dilatutions: 848 num kernels: 261	*StandardScale*	alpha = 148.73 max_iter = 769 random_state = 42 solver =auto	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 dilatations: 512 num kernels: 3910	*RobustScaler*	alpha = 148.73 max_iter = 591 random_state = 42 solver =auto	class -1 = 0.83 class 1 = 0.87	class -1 = 0.83 class 1 = 0.87	class -1 = 0.83 class 1 = 0.87	**0.85**	20 4 4 27
4	kernel max dilututions: 384 num kernels: 12010	*MinMaxScaler*	alpha = 0.78 max_iter = 16 random_state = 42 solver =auto	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 dilututions: 560 num kernels:16910	*RobustScaler*	alpha = 57.36 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

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

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

5.5.4 Mental State Recognition Insights and Analysis

In this section, we present the best model able to recognize two mental states related to Focus and Relaxation.

Table 5.4 highlights the best model in terms of performance for the 3-second EEG segment. Figure 5.19 shows the coefficients distributions of the best models for the different EEG segments length. It is remarkable that models trained with segments of length 1, 2, and 3 seconds show feature distributions with the highest variance. In fact, it suggests applying a feature selection criterion by exploiting percentiles. On the other hand, the distributions of the coefficient for the 4 and 5 seconds are more concentrated around zero. This phenomenon shows that the impact of the features is much more negligible than the models obtained with 1, 2, and 3-second segments. Table 5.4 and Figure 5.19, shows the best final model configuration by selecting 3-second EEG segments.

Afterwards, we identified three different candidate models as the best model. SVM builds a geometric separation hyperplane with support vectors, Random forest builds the decision boundary through the informative content of the features and the KNN algorithm creates centroids according to the similarities/differences between the features. Table 5.5 shows the results of the $5-fold$ cross-validation strategy after feature selection carried out in the previous step. SVM with a mean accuracy of 0.90 and STD of 0.050 is the best model because it achieves the best mean accuracy for two mental-state recognition in a $5-fold$ cross-validation strategy during the training and test phase.

In our approach, we used signal preprocessed; in contrast, the authors [27] used only downsampling data preprocessing, achieving 87.16% accuracy.

To summarize SVM is better than the others. Indeed it is able to define the decision boundary and it is more suitable for the distribution of characteristics. In fact, there is a better relationship between performance on the train and testing compared to Random Forest and KNN models. In addition, SVM shows no overfitting problem and achieved the best Standard Deviation (STD) during the test phase.

5.6 NeuralPMG System

Following the developments discussed above, this section introduces *NeuralPMG* a system that leverages BCI-based emotion detection with polyphonic music generation to support creativity and engagement in music composition. Here, NeuralPMG serves as a continuation of the themes addressed in the prior works: it integrates ML and BCI technologies to classify user mental states, using these insights to influence and enhance the music creation process in real-time. This framework incorporates Slonimsky's theoretical model, AI-driven pattern generation, and an interface combining EEG and hand movement detection to create a seamless composition experience.

With NeuralPMG, the aim is to go beyond pure emotion recognition to apply these techniques in artistic and therapeutic contexts. The system translates users' mental states into musical structures, offering a dynamic platform for musical composition that adapts to the user's cognitive and emotional states. This approach reflects a significant step forward in our research, highlighting how BCIs can be used both as conversation support systems (as in the case of ARIEL) and as interactive music composition tools that can be employed in neurorehabilitation and creative education environments.

In the following sections, we detail the NeuralPMG system's design, including its integration of EEG-based emotion classification, hand-tracking input for melodic structure, and the real-time generation of polyphonic compositions. This work marks a further evolution in leveraging BCI technology and AI for practical, user-centered applications across healthcare, education, and creative industries.

5.6.1 Example Scenario

This section proposes a scenario depicting a key usage situation of NeuralPMG. Scenarios are commonly used in Human-Computer Interaction for "bringing requirements to life" [226] since they provide, in a concrete narrative style, a description of specific people performing work activities in a specific work situation within a specific work context [116].

Tom is an expert music composer, who creates soundtracks for movies, advertisements, music videos, audio-visual setups, as well as more traditional 20th-century classical music. He frequently manages multiple composition projects in parallel and in different genres and styles, each taking from a week to several months to complete. He aspires to provide more complex compositions in the form of an audio and musical score, suitable for performance with virtual instruments or real music groups.

Tom needs to streamline the production process. He aims to obtain semifinished pieces that he can later refine, modify, or complete as desired. He is familiar with Nicolas Slonimsky's theory, which enables him to experiment with various generations of scales and melodic patterns. He can construct polyphonies using sound organisation grammar drawn from contemporary composition methods, such as the permutation of notes, inversions, transpositions of musical phrases, variations in rhythmic values, etc. Furthermore, Tom is well acquainted with the rules of polyphonic and counterpoint organisation, which he uses to create tonal and post-tonal musical compositions.

Tom does not wish to rely solely on an automated music generation system, even if it can be programmed with specific parameters, as he wants to retain control over his own artistic decisions in each composition. Thus, Tom decides to use the NeuralPMG, whose GUI is shown in Figure 5.20.

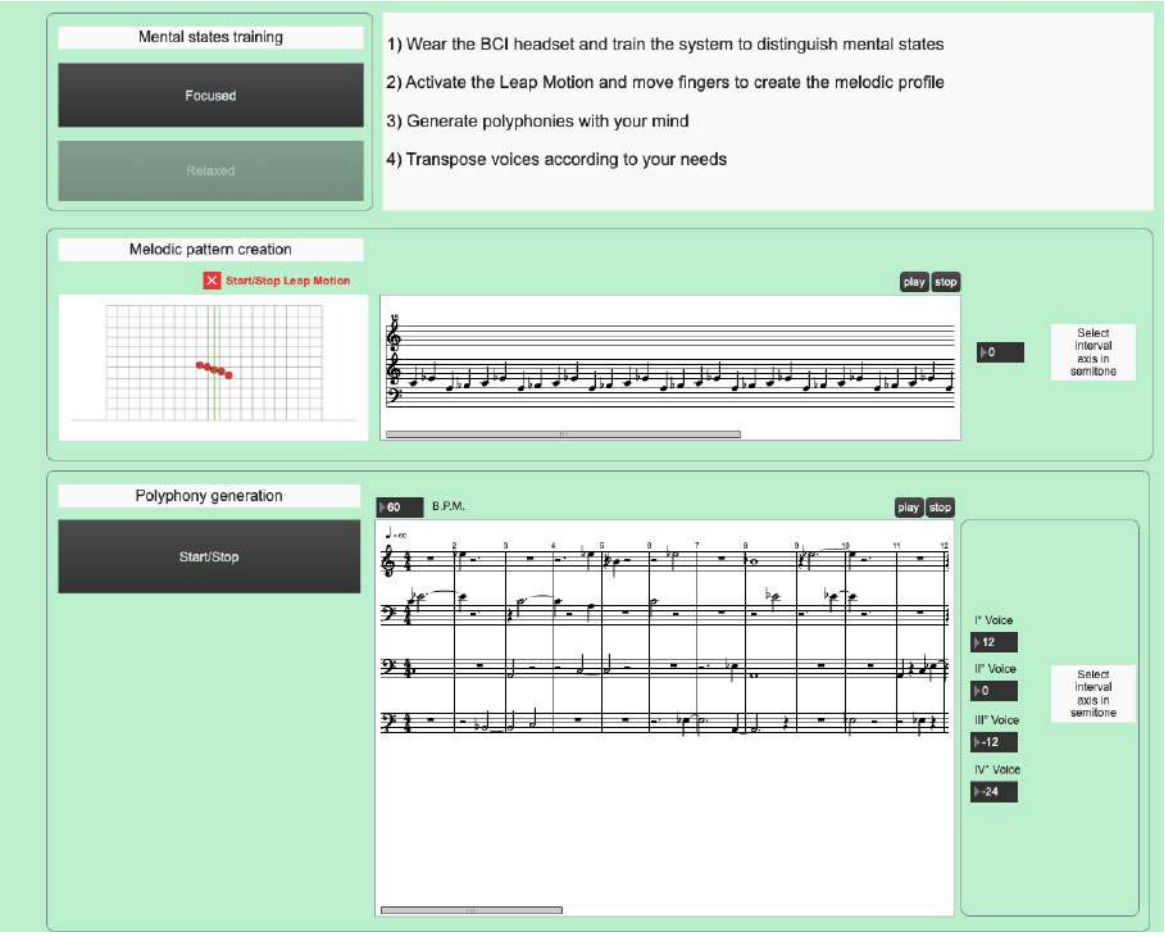


Fig. 5.20 GUI of the NeuralPMG framework: three panels outline the main steps in the process of creating the final polyphony, i.e., *Mental states training*, *Melodic pattern creation*, *Polyphony generation*. Instructions for using the system can be found in the panel at the top.

As first step, Tom wears a Emotiv headset, i.e. the BCI, and trains the system to recognize his mental state, so that it can determine if he is focused or relaxed. To accomplish this task, he selects the "Focused" and then the "Relaxed" buttons in the "Mental states training" panel of the system interface. Once the system has been trained on Tom's mental states, he can carry on using it. Tom activates the Leap Motion device to produce a melodic pattern: he moves his fingers on the device that detects the coordinates of Tom's fingers and translates them into musical notes that are displayed on the musical staff. The finger position and the musical staff are both displayed in the "Melodic pattern creation" panel, on the left and central side, respectively. The melodic pattern can be played and stopped. Once he has found a satisfactory melodic pattern, he can change and explore different version of melodic pattern through the box "selecting Interval Axes in Semitones". With this button he can transpose and refit the melodic pattern on different Interval axis. When Tom thinks he has found the right melodic pattern, saves it as a MIDI or XML file. Using this melodic pattern as a seed, the system exploits the Slonimsky's grammar to generate all the melodic patterns related to that seed. Tom then goes on with "Polyphony Generation". The trained classifier identifies Tom's mental state and generates a polyphonic composition. Tom knows that when he is focused, the polyphony has rhythmic values of quarter note, eighth note and sixteenth note, while when he is relaxed, the rhythmic values are whole note, half note and quarter note. Thus, depending on the polyphony he wants to generate, Tom tries to be focused or relaxed. Tom can modify the generation intervals, the "performance tempo" expressed in bits per minute (*bpm*) and the interval of each polyphonic voice according to what he feels for each composition. When he finds the generated polyphony useful, he saves and exports it in a MIDI or XML file format.

Later, having an archive of polyphonic and melodic compositions at his disposal, he can edit, arrange, and manipulate them as desired using either a Digital Audio Work Station (DAW) or a music notation software.

5.6.2 NeuralPMG Framework Architecture

In this section, we describe the NeuralPMG architecture.

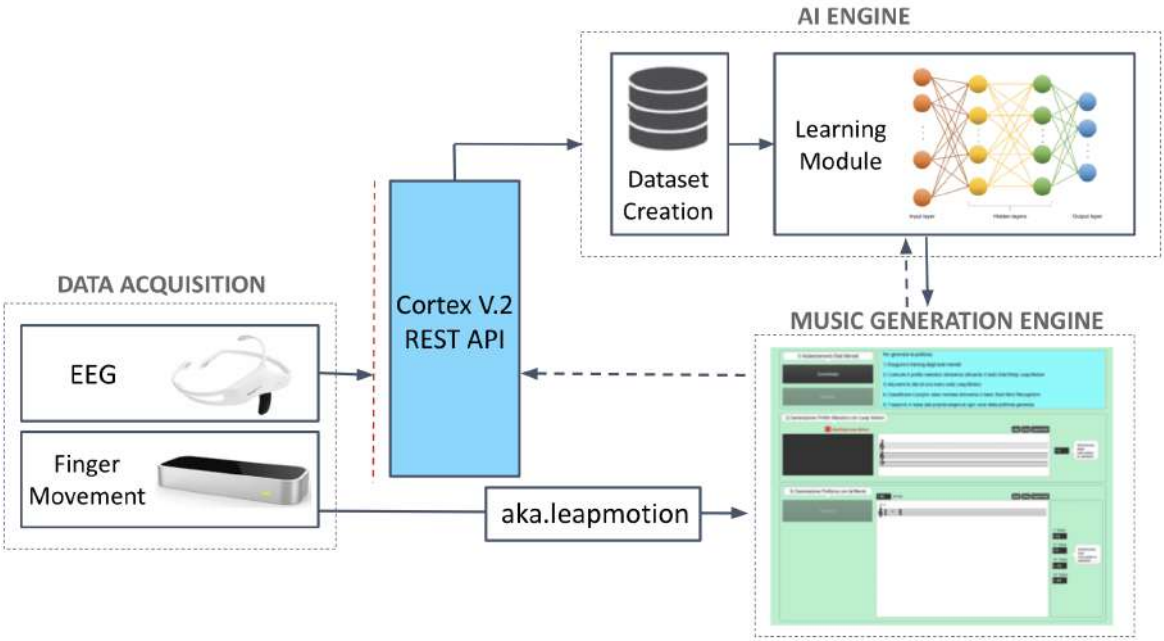


Fig. 5.21 Overview of the NeuralPMG framework architecture main components.

According to Figure 5.21, NeuralPMG consists of three macro-components. The *Data Acquisition* module is in charge of data acquisition from the two devices, i.e. the Leap Motion and the Emotiv headset. The *AI Engine* module receives the EEG power band values from the CortexV.2 REST API and elaborates them with ML techniques to first create a user's mental state training dataset and later classify the current mental state. The *Music Generation Engine* module is responsible for: i) creating the melodic pattern based on the data acquired by the Leap Motion; ii) creating polyphony based on the melodic pattern and the mental stated prediction value provided by the AI Engine; iii) displaying a GUI with the widgets for user interaction and system output. The dashed lines between the architecture modules represent control signal.

The processes of the modules are controlled using a Python Flask library[16], which allows us, through the API-Rest technology, to manage a communication protocol based on TCP-IP capable of handling server-client calls with related data exchange on a specific IP address.

Two devices are used for acquiring user data:

- **Brain Computer Interface.** The Emotive Insight device[17] is a 5-electrode passive headset capable of detecting electrical voltages on the head surface (i.e. the EEG). The electrodes are Semi-Dry Polymers, as they have a conductive

[16]Python Flask library: https://flask.palletsprojects.com/en/2.0.x/
[17]https://www.emotiv.com/insight/

rubber coating. Generally they can be used without gels or conductive solutions. The EEG signal quality is verified by the proprietary Emotive PRO[18] software interface, which provides a signal quality check by means of coloured indicators, from "Green = excellent quality" to "Black = very poor quality". As shown in Figure 5.21, the EEG headset interface communicates, with the Engine module by an API Gateway, using the native CortexV.2 REST API. Once the signal calibration phase has passed, data is received by the Cortex API SDK. Electrodes are placed according to the official $10-20$ system in positions $AF3$, $AF4$, $T7$, $T8$, and Pz [122]. Figure 5.22 shows a schematic representation of the electrodes with their positioning on the user's head.

The Cortex $V.2$ API by Emotive enables interfacing the device with various development environments. For each electrode, the API provides a series of numerical values corresponding to the power spectrum in the different bands of interest: $Delta$(0-4Hz), $Theta$(4-8Hz), $Alpha$(8-12Hz), $Beta$(12-35Hz) and $Gamma$(35-43Hz).

- **Leap Motion.** Leap Motion (LM)[19] is a powerful tool for recognizing the movement of different hand parts. Palm and fingers data of one or both hands in the spatial coordinates on the x,y, and z axes are detected by an infrared camera. The coordinates data streaming is received directly by `aka.leapmotion` object[20] in Max/Msp that visualizes them in a widget of the user interface.

[18]https://www.emotiv.com/emotivpro/
[19]https://www.ultraleap.com/product/leap-motion-controller/
[20]https://github.com/akamatsu/aka.leapmotion

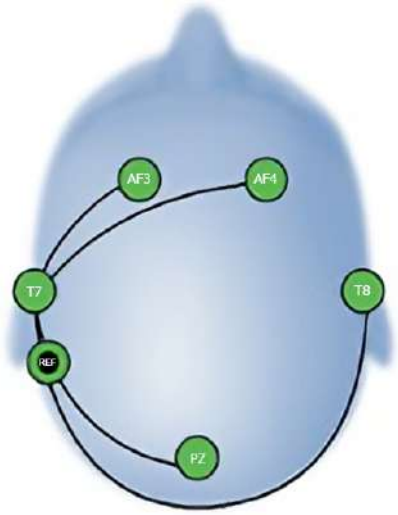


Fig. 5.22 Electrode placement according to the 10-20 system.

5.6.3 Polyphony Generation Process

The polyphony generation process is divided into two main parts, namely:

- **Generation of the basic melodic pattern:** It is generated by selecting a reference interval axis as proposed in Slonimsky's theory. The infra-inter-extrapolation process along with related permutations, is determined by the movement of one hand's fingers (refer to Section 5.6.3 for details).
- **Generation of Polyphonic structure:** The system employs the previously generated melodic pattern to create a four-part polyphony. This involves overlaying four rhythmic profiles, derived from permutations of rhythmic figures found in two distinct reference sets, onto the four voices. The first set comprises rhythmic values of quarter notes, eighth notes, and sixteenth notes, along with their corresponding rest values. The second set includes rhythmic values equivalent to whole notes, half notes, and quarter notes, also accompanied by their respective rest values. These divided sets are then utilized to form sequences of rhythmic figures. Concerning dynamic progression in each polyphonic voice, amplitude variations corresponding to *Alpha*, *Beta*, *Theta*, *Gamma*, and *Delta* brainwaves are scaled within the 0 - 127 range, in accordance with MIDI protocol standards. The selection of rhythmic value sets and dynamic progression is subject to the user's discretion, controlled through mental states assumed during system usage.

Mental State Generation

The system records power data from *Alpha*, *Beta*, *Theta*, *Gamma*, and *Delta* brainwave bands. This data is essential for defining mental states. As per literature [37], the *Focused* mental state correlates with increased power in Beta and Gamma bands, whereas the *Relaxed* state is associated with heightened Alpha band amplitudes. Initially, users must train the system to distinguish between these two mental states.

To elicit a Focused mental state, users engage in mental exercises that simulate an elevation in Beta and Gamma brainwave activities. The system sequentially displays numerical strings (e.g., [41-21-92-10-8-37-45-75-61-29-61-95-79- ...]) for the user to observe and read. Conversely, to induce a Relaxed state, the system plays natural sounds such as flowing water and bird chirps. The data collected during the focused and relaxed mental state form a training dataset for various ML algorithms and simple Feed Forward Neural Networks with up to three hidden layers.

Melodic Pattern Generation

Melodic pattern generation relies on the positioning of select fingers of one hand approximately 30 cm from a Leap Motion device. The Y-axis coordinates of the five fingers (left or right hand) are recognized. Figure 5.23 shows the position of the fingers at 45°angle relative to a horizontal reference plane. Three fingers are considered in building the melodic pattern: the first note is the reference tone; the second, third, and fourth notes are defined by thumb, middle, and little finger respectively. Figure 5.24 shows the same fingers translated onto the musical staff.

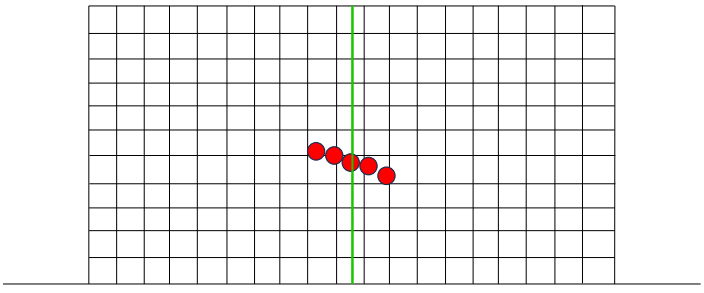


Fig. 5.23 Visualization of the user's left hand five fingers position in the NeuralPMG GUI.

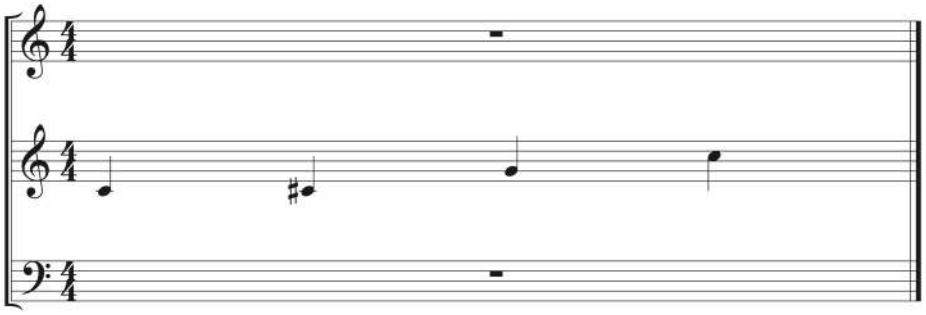


Fig. 5.24 Finger transposition on the musical staff.

The melodic pattern generated through LM is scaled across two octaves using the `bach.mod`[21] library. This scaling process involves specific handling of input values relative to a defined threshold:

1. **Threshold Processing**: The system employs a threshold of 2400 *cents*. For input values falling below this threshold, the output value directly corresponds to the input value. Conversely, if the input value surpasses 2400 *cents*, the output is recalculated as *inputvalue*-2400 cents.

2. **Constant Addition for Note Display**: To accurately represent the notes in both violin and bass clefs, a constant of +6000 *cents* is added to the output value. This adjustment ensures proper display and interpretation of the notes within the respective clef notations.

3. **Formation of the Base Melodic Cell**: The base four-note melodic cell is derived initially. Subsequent notes are generated by adding constants that correspond to the selected interval axis. This process is pivotal in determining the melodic structure and is dependent on the axis choice:

   - **Unison Interval Axis**: If a unison interval axis is chosen, a value of 0 *cents* is added to the four base notes.
   - **Augmented Fourth Interval Axis**: For an augmented fourth interval axis, each note of the base cell is incremented by +6000 *cents*.

4. **Display and Notation**: The resulting pattern is then visualized in a software object that provides a transcription in mensural notation (as referenced in [163]). This display includes a consistent rhythmic profile, standardizing all notes to quarter notes.

[21] https://www.bachproject.net/

Figure 5.25 illustrates the generated melodic pattern following these procedures.

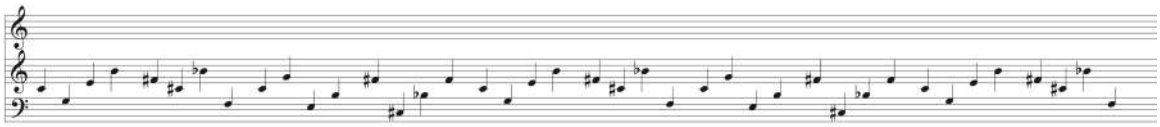


Fig. 5.25 Melodic pattern of infra-inter-ultrapolation on augmented fourth C-F# interval axis.

Polyphony Generation

The polyphony generation process starts with the analysis of the user's mental state, utilizing the Emotiv headset to capture EEG signals. These signals are transformed from the time domain to the frequency domain in the range 0.5-43 Hz. The system samples 50 time series at approximately 1-second intervals, each consisting of 25 features. This data, pertaining to *Alpha*, *Beta*, *Theta*, *Gamma*, and *Delta* bandwidth values, is then fed into a classification algorithm. The algorithm outputs a prediction value corresponding to a mental state class, which informs the rhythmic value selection in the Max/MSP software for generating the rhythmic profiles of the four voices. The rhythmic values are divided into the following two categories:

1. *Focused* state: consists of rhythmic figures 1/4, 1/8, 1/16 and their corresponding rest values;

2. *Relaxed* state: consists of rhythmic values of 4/4, 2/4, 1/4 and their corresponding rest values.

Random permutations of these rhythmic values generate strings of 49 rhythmic figures, assigning one rhythmic value to each note. These strings are then overlaid on the pitches of the basic patterns distributed across the four voices in the score, visualized in the bach.score object. This visualization facilitates observing the complete polyphonic score as shown in Figure 5.25.

Subsequently, the score undergoes further refinement by transposing each voice along a selected interval axis, as per Slonimsky's theory. This involves adding a numerical constant to each note and voice, corresponding to the chosen interval axis. Figure 5.26 elucidates this transposition operation, showcasing the initial set of pitches in the top box and the altered set in the bottom box after adding a numerical constant of 300.

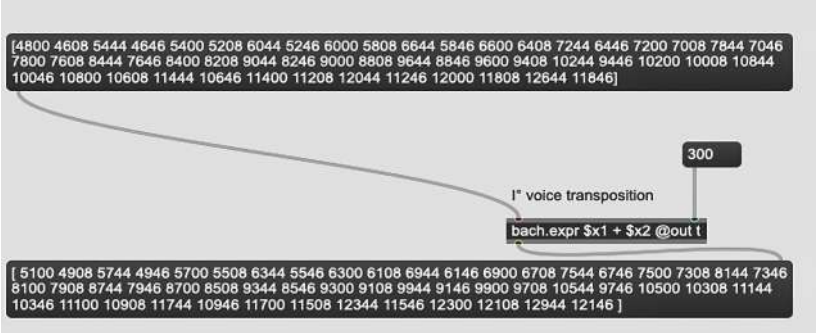


Fig. 5.26 Transposition operation of the pitch set.

The rectangle labeled 'I°voice transposition' in Figure 5.20 performs the summation operation between the top height list plus the numeric constant. The input height list in the `bach.expr` method is represented by the variable x_1, while the constant is represented by the variable x_2. In the lower panel of Figure 5.26, the output list on which the summing operation was performed can be observed.

Additionally, the amplitude values from the relevant EEG bands are utilized to set dynamic values for each note. These dynamics are translated into MIDI velocity values, ranging from 0 to 127. The Max/MSP's native scales object is calibrated to scale these values appropriately, with input ranges from 0 to 100 (reflecting the maximum bandwidths provided by the Emotiv Insight) and output ranges from 10 to 127. This ensures that notes are never assigned a velocity of 0, which would equate to silence.

Upon completion of polyphony generation, the piece can be audibly rendered. The tempo is adjustable via beats per minute ($B.P.M.$), allowing for the input of the desired metronomic speed. Additionally, input boxes are provided for further modification of the score, enabling transposition of each polyphonic voice to preferred intervals. The finalised composition can be exported in MIDI or XML formats.

NeuralPMG Engine

This section outlines the functioning of the NeuralPMG Engine. Two mental states are recorded (Focused/Relaxed) and a dataset is created. For each user, NeuralPMG identifies the best ML algorithm able to recognize the mental state. Using LM allows the user's finger movement to be recognized and generate the melodic pattern.

Different ML models are used and compared in the training phase. The best model is then used for predicting the user's mental state. The system then generates polyphony from the melodic pattern according to the user's detected mental state.

Dataset Generation. The acquisition of the mental state is managed by specific routines. There is TCP-IP call to Flask[22] enabling communication between the BCI device and the Cortex API[23]. Once the connection has been established, the API starts streaming power band data for each band and electrode. They are absolute values whose unit is expressed in uV^2/Hz. Each user has a dataset consisting of 25 features based on 5 power bands for 5 channels, with 100 total acquisitions: the first 50 items collected while the user was stimulated to the Focused state by reading sequences of numbers, the next 50 while listening to natural environment sounds to induce the Relaxed state. As an illustrative example, the Supplementary Table 2 available in the Appendix to this article presents the mental state dataset acquired for a single user.

Classification Model Evaluation. Determining the user's mental state is influenced by several factors, as it is well-known in the EEG analysis domain. First, values for delta, theta, alpha, beta, and gamma frequencies can vary among different users. Furthermore, users have different reactions on internal or external stimuli, thus having different abilities in relaxing or concentrating without being influenced by the surrounding environment. To mitigate these problems, we selected six ML models and tested their performances in classifying mental state by involving 5 users. The models are trained in parallel for each user; this approach is possible because the training dataset is not large and therefore it does require reasonable computational resources and time to execute. The following models were considered:

- Linear Discriminant [92];
- Decision Tree [49];
- Naive Bayesian [236];
- Support Vector Machine [302];
- K-nearest neighbors [274];
- Feedforward Neural Network [277] (1 hidden layer with 10-neuron);
- Feedforward Neural Network [277] (1 hidden layer with 25-neuron);
- Feedforward Neural Network [277] (2 hidden layers with respectively 10 neurons per layer);

[22]https://flask.palletsprojects.com/en/2.2.x/
[23]https://emotiv.gitbook.io/cortex-api/

To perform the training phases of the models, we used the Matlab toolbox - Classification Learner[24]. It performs the following steps: (i) Z-score standardization and (ii) Training phase with k-fold = 5. The toolbox returns the model with the best accuracy among all models considering also the best performance of the model on the 5 folds obtained in the training phase. For each model, a random search is performed to select the best hyperparameters.

Table 5.6 shows test results. KNN showed extremely low accuracy values; therefore, it was not considered in the system implementation. The Feed Forward neural network 1 (FNN1), despite its weak performance, was implemented in the system.

Table 5.6 Mental state classification performances of the considered models: LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayesian, SVM = Support Vector Machine, FNN = Feedforward Neural Network.

	Accuracy Performance							
User	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

5.6.4 Evaluation Study

In this section, we describe the user study that aimed at evaluating the interaction with the NeuralPMG from various perspective.

Participants and Design

The study on NeuralPMG specifically involved participants from the *Electronic Music laboratory* of the *E.R. Duni Conservatory of Matera*. This targeted selection was based on two key prerequisites: expertise in algorithmic composition and proficiency in both algorithmic composition and sound manipulation software. This focus on electronic music students was intentional. Traditional music composers, typically from classical backgrounds, were not considered suitable for this study. Their training often emphasizes manual composition methods, including hand-writing music, and frequently

[24]https://it.mathworks.com/help/stats/classificationlearner-app.html

avoids the use of digital tools like music notation software. In contrast, our study required familiarity with computer systems integral to algorithmic composition.

All participants in the study possessed a moderate level of experience in algorithmic and computer-aided music composition. While some were professional instrumentalists, all participants were skilled in at least one musical instrument, aligning with the Conservatory's educational standards. Notably, none of the participants were previously acquainted with Slonimsky's method.

The participant group was diverse in terms of age and gender: age min = 19, age max = 61, age avg = 30, Females = 10, Males = 9, 1^{st} academic degree level = 11 and, 2^{nd} academic degree level = 8.

Procedure

The experimental procedure involved three Human-Computer Interaction (HCI) experts: two observers, who recorded task execution times and challenges, and one facilitator, who managed participant-system interactions. Conducted individually for each participant in the "Rota" hall of the Conservatory, the study was performed under moderate lighting and comfortable room conditions (21°C). The procedure consisted of six stages:

1. **Initial Setup**: Each participant signed a consent form for photo documentation and research use before proceeding to the study station, where equipment was prearranged.

2. **Calibration Phase**: Participants wore an Emotive Insight headset, adjusted per the 10-20 standard electrode placement system. Impedance was checked using EmotivePro software, and conductive gel was applied as necessary to ensure signal quality. Participants were advised to minimize head and facial movements to maintain signal integrity.

3. **Mental State Training**: Machine Learning algorithms trained the system to distinguish between 'Focused' and 'Relaxed' mental states. For 'Focus,' participants read random number sequences on-screen; for 'Relaxed,' they practiced diaphragmatic breathing while listening to nature sounds. Accurate classification in this stage was crucial for generating subsequent polyphonic elements.

4. **Melodic Pattern Generation**: After calibration, participants created a melodic pattern using the Leap Motion device:

    - Activating the Leap Motion start/stop button.

- Moving their hand and fingers over the sensor (at approximately 10 cm) to compose a pattern.
- Saving the pattern by holding the hand steady and pressing the start/stop button.
- Modifying and playing the pattern as desired.

This phase took approximately 10 minutes per participant.

5. **Polyphony Generation**: Participants created a polyphony based on the melodic pattern with the following steps:

   - Pressing `Dial` to generate the polyphony.
   - Adjusting speed (BPM) and interval axes as needed.
   - Saving the final composition by pressing `export midi`.

   The polyphony generation phase, including headset calibration, required an average of 20 minutes per participant.

Data Collection

Both quantitative and qualitative data were collected through questionnaires completed by participants and observer notes on significant behaviors or comments. All interactions were audio-video recorded.

Initially, participants filled out a demographic questionnaire assessing their background in IT, electronic music composition, and familiarity with Slonimsky's method. Details are reported in Section 5.6.4.

The second questionnaire, divided into 8 sections, evaluated various aspects of the interaction with NeuralPMG. Each of the first five sections used established questionnaires, while the remaining three sections contained custom questions to explore specific aspects:

1. **User Experience** using the *AttrakDiff* questionnaire [117]: A 28-item, seven-step bipolar adjective scale (e.g., "confusing - clear"). This tool assesses:

   - *Pragmatic Quality (PQ)*: Usability and goal achievement.
   - *Hedonic Quality - Stimulation (HQ-S)*: Novelty and engagement.
   - *Hedonic Quality - Identity (HQ-I)*: User identification with the system.
   - *Attractiveness (ATT)*: Overall quality perception.

2. **Creativity Support** using the *Creativity Support Index (CSI)* questionnaire [57]: A 12-item survey measuring six creativity support dimensions: *Exploration, Expressiveness, Immersion, Enjoyment, Effort*, and *Collaboration*.

3. **Workload** using the *NASA-TLX* questionnaire [115]: A six-item survey rating perceived workload on *Mental Demand, Physical Demand, Temporal Demand, Performance, Effort*, and *Frustration*, combining scores into a NASA-TLX workload index.

4. **User Engagement** using the *User Engagement Scale (UES) short-form* [204]: A 12-item survey assessing engagement through four dimensions: *Focused Attention (FA), Perceived Usability (PU), Aesthetic Appeal (AE)*, and *Reward (RW)*.

5. **Emotional Response** using the *Self-Assessment Manikin (SAM)* questionnaire [34]: Two image-based questions measuring emotional response in terms of *pleasure, arousal*, and *dominance* on a 1-9 Likert scale.

6. **Self-Assessment of Generated Polyphony**: Eight custom questions on a 1-7 scale, covering:

   (a) Need for modifications in pitch and dynamics (2 questions).

   (b) Most suitable compositional techniques for further processing (3 questions).

   (c) Application domains best fitting the generated polyphony (3 questions).

7. **NeuralPMG Appreciation**: Two open-ended questions on the most and least appreciated aspects of NeuralPMG.

8. **Comments and Suggestions**: One open-ended question inviting feedback and suggestions.

The average time for participants to complete the questionnaire was approximately 12 minutes.

Experiments

In the following subsections data collected during the experimental study are analysed and results are provided.

Classification Performances of NeuralPMG Models

Table 5.7 shows the Accuracy obtained during the experiment by each model for each of the 19 participants. It emerges that the different models for classifying the participants' mental state were selected homogeneously by the engine (LD, DT, FNN2 = 4; SVM = 5), except for NB which was used for only two participants. FNN1, which in the pilot test had weak performance, has never been the best classification model and was used for no participants.

Table 5.7 Model accuracy for each of the 19 participants. LD = Linear Discriminant, DT = Decision Tree, NB = Naive Bayesian, SVM = Support Vector Machine, FNN = Feedforward Neural Network. The grey background indicates that model has been selected for that participant.

	Accuracy Performance					
Participant	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	0.4	0.76	0.6	0.5
16	0.70	0.80	0.64	0.79	0.73	0.77
17	0.43	0.44	0.65	0.32	0.7	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 (UX)

An overview of the AttrakDiff results is presented in Table 5.8, which summarizes the hedonic (HQ) and pragmatic (PQ) qualities of the system according to their respective confidence rectangles. In general, a larger rectangle indicates greater uncertainty about the system's classification. NeuralPMG displays high HQ and PQ values, categorizing it as a *desirable* product with promising User Experience (UX).

The HQ value (1.55, 0.29) is higher than PQ (0.76, 0.35), indicating a stronger hedonic appeal. The partial scores for the system are as follows: Attractiveness (ATT) (7.50, 0.63), Hedonic Quality - Identity (HQ-I) (6.48, 0.60), Hedonic Quality - Stimulation (HQ-S) (7.24, 0.73), and Pragmatic Quality (PQ) (5.82, 0.80).

Table 5.8 Mean (AVG score) and Standard deviation (SD score) for each category of the AttrakDiff questionnaire.

	AVG Score	*SD Score*
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

Using the Creative Support Index (CSI) questionnaire, participants' perceptions of NeuralPMG's creativity support were measured. The system achieved an average CSI score of 70/100, indicating strong support for creative design (CSI=70.00, STD=14.86). As shown in Table 5.9, the highest average score was for Immersion (8.18), followed by Expressiveness (7.32), Exploration (7.63), and the Effort/Reward trade-off (RWE 7.45). Collaboration scored lowest (5.03), as there were no collaborative functions in the system.

Table 5.9 Mean (AVG score) and Standard deviation (SD score) for CSI questionnaire dimensions.

	AVG Count	*SD Count*	*AVG Score*	*SD Score*	*AVG weigh. Score*	*SD weigh. Score*
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 tradeoff	1	1.61	7.45	2.34	8.62	3.77

Workload

Workload was assessed using the NASA-TLX questionnaire, with scores shown in Table 5.10. Mental Demand had a high weighted mean (7.42, SD=7.68), indicating significant cognitive workload. Physical Demand was low, reflecting minimal physical effort. Performance scored high (8.11, SD=3.02), suggesting positive feedback on outcomes. Frustration levels were low (3.95, SD=5.03), although variability was noted, likely due to calibration challenges with the headset.

Table 5.10 Mean (AVG) and Standard deviation (SD) for NASA-TLX questionnaire dimensions.

	AVG Count	*SD Count*	*AVG Score*	*SD Score*	*AVG weigh. Score*	*SD weigh. Score*
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	1.83	5.53	5.57
Performance	0.89	0.32	8.68	1.57	8.11	3.02
Effort	1	2.47	1	2.47	4.95	6.84
Frustration	0.74	2.35	0.74	2.35	3.95	5.03

User Engagement

The User Engagement Scale (UES) short form measured engagement, yielding an average score of 4.13 (SD=0.41). Table 5.11 shows average scores for each dimension, with Perceived Usability (PU) achieving a satisfactory score of 4.28 (SD=0.64). The Aesthetic Appeal (AE) and Reward (RW) dimensions were also highly rated, highlighting the system's visual appeal and user satisfaction.

Table 5.11 Mean (AVG Score) and Standard Deviation (SD Score) for UES Questionnaire Dimensions.

	AVG Score	*SD Score*
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

Participants' emotional responses to the generated polyphony were categorized into four main emotions: Excited/Delighted/Happy, Sleepy/Calm/Content, Neutral, and

Sad/Depressed/Bored. A softmax function was applied to determine the dominant emotion, yielding a high positive score (0.99978) for Excited/Delighted/Happy, with minimal responses in other categories (see Table 5.12).

Table 5.12 Emotional Score from Softmax Activation Function.

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

Self-assessment of Generated Polyphony

Participants provided feedback on the generated polyphony, suggesting that significant modifications were needed in pitch and dynamics (AVG=3.42/7). The preferred compositional techniques for further processing were serial/atonal (AVG=5.21) and aleatory (AVG=5.42). Most participants felt the polyphony was best suited for Multimedia (AVG=5.84) and Academic (AVG=5.16) domains.

NeuralPMG Appreciation and Suggestions

Open-ended responses from participants were analyzed thematically, revealing six positive themes, including "System Novelty," "Production of Unconventional Polyphonies," and "Use of BCI and Leap Motion Devices." Critical themes included "System GUI" and "System Efficiency." Tables 5.13 and 5.14 show theme frequency.

Table 5.13 Aspects of NeuralPMG Appreciated by Participants.

Theme	*Frequency*
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 5.14 Aspects of NeuralPMG Criticized by Participants.

Theme	*Frequency*
Use of BCI system	5
Use of Leap Motion device	4
System graphical user interface	2
System appreciation	2
System efficiency	3

Comments and Suggestions

Seven participants provided additional comments, all positive. Participants expressed appreciation for the system (4 comments), a desire to use it in their work (2 comments), and one suggested integrating other biometric inputs, such as heartbeat.

5.6.5 Polyphony Assessment by Music Experts

Two professors of the E.R. Duni Conservatory of Matera, both holding the title of Maestro, and with extensive experience in music composition were involved. They individually listened to the 19 polyphonies generated by the 19 participants in the previous study and, for each one, answered a questionnaire structured in 6 sections:

1. three questions focusing on the evaluation of the aesthetic aspects of the polyphony;
2. four questions addressing the technical evaluation of the harmonic aspects;
3. three questions analysing the polyphony level of elaboration;
4. three questions dealing with the evaluation of possible usage scenarios of the polyphony;
5. nine questions assessing each polyphony with respect to its ability to arouse 9 different emotions: happiness, tenderness, happiness, anger, sadness, fear, negativity, activity, positivity, tension.
6. one open question to provide further comments and opinions.

The first 5 sections were based on a Likert scale with values from 1 to 7, representing respectively the degree of agreement ranging from "not at all" to "very much".

Results of the music experts' assessment have been summarised by heatmaps, complemented with average values and standard deviation for every question and

section. Figure 5.27 and 5.28 refer to the polyphony evaluation provided by Maestro 1 and 2, respectively.

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	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	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	7	5	7	7	7	6	7	6	7	6.42	0.69		
Innovativeness	Harmonic Interest	5	5	6	6	6	6	2	3	2	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	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	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	2.11	0.32		
Elaboration	No Need to Adjust Parameters	2	5	6	5	5	5	6	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	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		
Usage	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		
Emotion	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	1	2	1	1	2	1	3	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		
	Rating	7	6	5	4	3	2	1																

Fig. 5.27 Maestro 1's evaluation of the polyphonies produced by the 19 participants.

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		
	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		
Elaboration	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
	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		
Usage	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
	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		
Emotion	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	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		
	Rating	7	6	5	4	3	2	1																

Fig. 5.28 Maestro 2's evaluation of the polyphonies produced by the 19 participants.

5.6.6 NeuralPMG Findings

In this section, we discuss the findings from the NeuralPMG evaluation, covering dimensions such as *User Experience (UX)*, *Creativity*, *Workload*, *Engagement*, *Self-assessment of Generated Polyphony*, *Application Domains*, and *Participant Feedback*.

The assessment by the two Maestros also provides critical insights, highlighting the system's strengths and potential areas for improvement, as well as its broader impact on music composition and user interaction.

User Experience (UX): The AttrakDiff results indicate that participants generally found NeuralPMG attractive. However, there is room for improvement in its *Pragmatic Quality*, as noted in user feedback.

Creativity: According to the Creativity Support Index (CSI), the exploratory component was highly rated, reflecting the system's alignment with the participants' professional interests. While *immersiveness* and *enjoyment* were prominent, *expressiveness* was less significant, which aligns with the participants' inclination to use tools for creative support rather than as direct expressions of artistry. Overall, the system fostered focused creativity among participants, enabling deep engagement in the composition task.

Workload: Low *physical effort* was reported, though some participants experienced *frustration* related to BCI headset calibration challenges.

Engagement: Engagement scores from the User Engagement Scale (UES) and emotional responses indicated positive interaction with NeuralPMG, affirming its role in creating an immersive and enjoyable experience.

Self-assessment of Generated Polyphony: Participants recognized NeuralPMG's effectiveness in assisting with complex note organization. Although refinement in pitch and dynamics was desired, the generated polyphonies were seen as a beneficial starting point for further development, especially for academic or multimedia applications. This aligns with serial and atonal compositional principles, emphasizing diverse organizational techniques.

Application Domains: Participants found the polyphony suitable for academic and multimedia applications, such as in film scoring, where it can convey dynamic emotions and narratives through sound textures and harmonic layers.

Participant Feedback: Insights from open-ended responses underscored the system's effectiveness in mental state decoding for polyphony generation, though improvements in BCI device comfort and the graphical user interface (GUI) were suggested. Enhancing device wearability and GUI customization in future versions could further improve user experience.

In terms of *Slonimsky's grammar* for harmonic and melodic exploration, feedback highlighted its potential in generating unconventional melodies and polyphonies, reinforcing the system's role in discovering innovative compositional materials.

Leap Motion Device Challenges: Some participants reported challenges in using the Leap Motion device for melody crafting. Future adjustments to tracking sensitivity could improve stability in melodic pattern generation, enhancing user control over the musical output.

Transparency and Explainability: Some participants expressed difficulty in understanding how BCI-generated signals influence polyphony generation, a common issue in Human-Computer Artificial Intelligence (HCAI) systems [253, 260]. Addressing this, we propose enhanced transparency in data flow management, with the addition of visual or auditory cues reflecting backend processes. Such modifications align with HCAI principles for reliable, safe, and trustworthy AI systems [259].

Maestros' Evaluation: The two Maestros provided nuanced insights into the system's aesthetic and innovative potential. Both agreed on the consonance/dissonance balance, indicating coherent harmonic structures in the generated polyphonies. Divergence in their views on *innovativeness* reflects differing interpretations of innovation, with Maestro 1 favoring harmonic coherence and Maestro 2 valuing the unconventional structures enabled by Slonimsky's theory.

In the *elaboration* domain, Maestro 1 viewed polyphonies as harmonically integrated textures, while Maestro 2 adopted a contrapuntal perspective, seeing them as foundations for further permutation. Emotional responses to the generated pieces varied, affirming the polyphonies' versatility across applications such as film scoring and pure composition, highlighting the breadth and interpretative depth of the material created.

In summary, the NeuralPMG evaluation underscores the system's role in supporting composers by facilitating the exploration of polyphonic structures. Future developments should focus on enhanced explainability, interface improvements, and real-world application validation. The findings reflect NeuralPMG's potential to foster creative innovation in the intersection of BCI, ML, and music composition, offering a promising tool for professional musicians and composers.

5.7 NeuroSense: Low-Cost sparse electrode Dataset

The final work detailed in this chapter presents the NeuroSense project, an innovative dataset for emotion recognition utilizing a low-cost, sparse-electrode EEG device. NeuroSense builds upon previous studies on emotion recognition, where advancements in EEG-based emotion classification revealed the viability of detecting and categorizing emotional states. While these studies provided significant insights, they predominantly

relied on complex and high-density EEG setups, which can limit accessibility and general applicability in practical environments.

Given the state of the art previously discussed, where the majority of emotion recognition studies employ numerous electrodes to capture brain activity across various regions, we chose to take a different path. The NeuroSense dataset is designed to explore the feasibility of emotion recognition using a minimalistic setup, employing only four electrodes. This approach addresses the growing interest in creating scalable, accessible, and cost-effective solutions for real-time emotion recognition, especially for applications in human-computer interaction, mental health monitoring, and other areas requiring portable systems.

The NeuroSense dataset comprises EEG signals collected from 30 participants who experienced a series of music videos in a controlled neurofeedback environment. Through a standardized assessment protocol, participants rated their emotional responses along key affective dimensions, such as arousal, valence, and dominance. This dataset not only provides a unique resource for ML-based emotion recognition with sparse EEG devices but also offers an extensive preprocessing pipeline, statistical validations, and comparisons to other widely used datasets to highlight both the advantages and limitations of using a minimal electrode setup for affective computing.

In this section, we will discuss the motivation, methodology, and design of the NeuroSense dataset, as well as the potential implications of using low-cost, sparse electrode setups in the future of emotion recognition research.

5.7.1 Stimuli Selection

We adopted the DEAP protocol for eliciting emotional responses, chosen for its standardization and reliability in emotion analysis via physiological signals. This protocol offers multidimensional evaluations, including arousal, valence, and dominance, aligning closely with our research objectives and enhancing ecological validity through the use of audiovisual stimuli to evoke genuine emotional responses.

The stimuli selection process consists of several key steps, as illustrated in Figure 5.29:

1. **Video Clip Selection**: A total of 120 videos were selected in two stages—60 videos through affective tags derived from an expanded list of emotional keywords based on Parrott's work [215], and 60 videos curated manually. Using the

Last.fm [25] database, we identified songs tagged with specific emotions, representing each of Russell's emotional quadrants [241].

2. **Detection of One-Minute Highlights**: Using the DEAP protocol and a Relevance Vector Machine (RVM) model [265], we identified high-emotion one-minute segments within the selected videos based on predicted arousal and valence scores.

3. **Video and Audio Feature Extraction**: Relevant video features, such as motion vectors, lighting, color variance, rhythm, and dynamic object movements, were extracted to enhance emotional content. The video content was encoded in MPEG-1 format [278] to facilitate these extractions. Lighting keys in the HSV color space and color variance in the CIE LUV color space were computed for each frame [43, 183]. Further, rhythm and emotional impact were evaluated by average shot change rates [114] and dynamic scenes quantified by motion vectors [234].

4. **Relevance Vector Machine Application**: The RVM model [265] was trained on annotated features from an initial set of 21 movies to predict valence and arousal scores. Selected segments with high emotional scores were manually refined to ensure relevance.

5. **Online Subjective Annotation**: The final selection of 40 test video clips was achieved via online subjective ratings from volunteers, assessing arousal, valence, and dominance. Each video was selected for its high-intensity emotional elicitation.

[25] www.last.fm

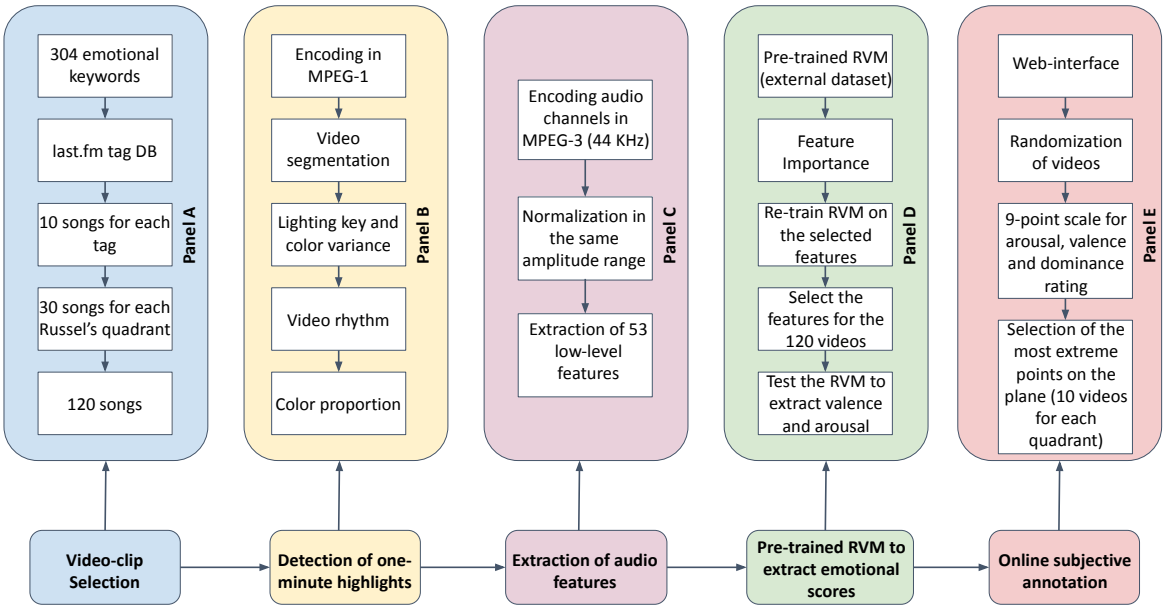


Fig. 5.29 Overview of the stimuli selection process, comprising five steps: (1) selection through affective tags and manual curation; (2) detection of one-minute highlights using DEAP; (3) video and audio feature extraction; (4) RVM-based valence/arousal prediction; and (5) final selection via online annotation.

Video and Audio Feature Extraction

The extracted video features included shot segmentation, lighting keys, color variance, rhythm, and dynamic scene analysis. Lighting keys and color variance were determined in the HSV and CIE LUV color spaces, respectively [43, 183]. Rhythmic impact was assessed by shot change rates [114], and dynamic object movements by motion vectors [234]. Audio channels were converted to mono MPEG-3 format at 44.1 kHz, normalized in amplitude, and analyzed for 53 low-level audio features (e.g., MFCCs, energy, pitch) [105]. The summarized pipelines are shown in Panels B and C of Figure 5.29.

Relevance Vector Machine

To predict the valence and arousal scores for each video segment among the selected 120 videos, the Relevance Vector Machine (RVM) [286] is trained using all shots from 21 annotated movies within the dataset presented in [265]. Additionally, the RVM identifies the importance of features. It is thus employed in the DEAP protocol to select a subset of features from the entire set of extracted features. The final prediction pipeline is delineated below:

- the music videos are segmented into one-minute intervals with a 55-second overlap.

- Content features are extracted from each segment.
- Prediction scores for arousal and valence are obtained from the trained RVM.
- A final score is computed for each segment using the equation: $e_i = sqrt(a_i^2 + v_i^2)$, where a_i and v_i represent centered arousal and valence scores.

The graphical representation shown in Panel D of Figure 5.29 summarizes the steps involved in the pipeline. Segments with higher emotional highlight scores are then selected. A manual override of the affective highlight detection is performed for some clips. This manual intervention is applied to segments deemed characteristic of the song, recognized by the public and expected to elicit emotional responses. Following this iterative process, a collection of 120 one-minute videos featuring high emotional content is obtained.

Online Subjective Annotation

This section describes the process for selecting the final 40 test video clips. The selection methodology involves a web-based subjective emotion assessment wherein participants viewed music videos and provided ratings on a 9-point scale for valence, arousal, and dominance. Key aspects of this process include:

1. participants used a web interface for rating videos.
2. They were allowed to watch as many videos as they wanted and end the rating process at any time.
3. The order of video clips was randomized.
4. Participants did not see the same video twice.
5. All 120 videos received ratings from 14 volunteers.

For each video, a score representing the intensity of elicited emotion was computed based on the ratings provided by volunteers. This score was calculated by dividing the mean rating (μ_x) by the standard deviation (σ_x), yielding $\frac{\mu_x}{\sigma_x}$. Subsequently, for each quadrant where the videos were positioned, the videos situated at the extreme corners of the quadrant were selected. Through this iterative process, 40 videos were ultimately selected. We defined the corresponding labels of these videos as external labels. Panel E of Figure 5.29 illustrates the main steps.

5.7.2 Experimental Setup

Materials and Setup The experiments were conducted in laboratory settings with controlled illumination in the Department of Electrical and Information Engineering premises at the Polytechnic University of Bari. EEG signals were recorded using a Muse 2 device [26], equipped with four electrodes, connected to a dedicated recording computer, a MacBook Pro (Retina, 15-inch, Mid 2015).

Stimuli presentation was facilitated through a dedicated PC monitor (HP), and the software for stimuli presentation was developed using Max/MSP [27]. This system featured a graphical user interface (GUI) with corresponding numerical identifiers linked to the videos to be played. Upon the operator pressing the button associated with a specific video, a marker was transmitted to a server using the LabStreamingLayer (LSL) protocol [28]. Labrecorder software [29] was employed to record EEG signals and synchronize markers.

The music videos were displayed on a 17-inch screen with a 1280×1024 pixel resolution. To minimize eye movements, the videos were presented at a reduced resolution of 800×600 pixels, filling approximately two-thirds of the screen. Participants were seated approximately one meter away from the screen. Yamaha-HS 8 speakers were used for audio playback, with the volume set relatively loud. However, participants were consulted beforehand regarding their comfort level with the volume, and adjustments were made accordingly.

The device used for the acquisition is the previously described Muse 2 EEG device.

Experimental Protocol

A total of 30 healthy participants (50% female), aged between 19 and 30 years (mean age 23.5), were recruited for the experiment. Participants for this study were recruited through university-wide advertisements and social media platforms to ensure a diverse and representative sample. All participants provided informed consent before participating, and no monetary compensation was offered. Ethical permission was obtained from the Local Ethical Committee of the University of Bari. The participants received comprehensive information regarding the experimental protocol, including detailed explanations of the various self-assessment scales. Once the protocol was thoroughly explained, the EEG device was positioned, and signal quality was meticulously assessed using the MuseLSL library for EEG signal window [20].

[26] https://choosemuse.com/products/muse-2
[27] https://cycling74.com/products/max
[28] https://labstreaminglayer.org
[29] https://github.com/labstreaminglayer/App-LabRecorder

The experiment began with the presentation of 40 videos across 40 separate trials, each lasting 1 minute. Each trial followed a standardized sequence of events comprising:

- a 2-second display indicating the current trial number to inform participants of their progress.
- A 5-second recording of baseline activity, represented by a fixation cross.
- Presentation of a 1-minute music video.
- Subsequent rating of the participants' arousal, valence, liking, and dominance levels.

Upon completion of 20 trials, participants were given a brief break. During this intermission, the facilitator assessed signal quality and electrode placement to ensure accuracy before instructing participants to proceed with the second part of the test.

Participant Self-assessment

Participants were tasked with assessing their levels of arousal, valence, liking, and dominance after each trial. To facilitate this process, Self-Assessment Manikins (SAMs) [34] were employed, providing visual representations of the scales. For example, the liking scale featured symbols of thumbs-down and thumbs-up, positioned at the center of the screen with numbers 1 to 9 displayed below them. Figure 5.30 illustrates an example of the various scales utilized in the questionnaire. Participants indicated their self-assessment levels horizontally, moving the mouse beneath the numbers and clicking on their chosen level. They were informed of the flexibility to click anywhere below or between the numbers, effectively creating a continuous scale for self-assessment.

The valence scale ranged from unhappy or sad to happy or joyful, allowing participants to rate the emotional tone of their experience. The arousal scale spanned from calm or bored to stimulated or excited, enabling participants to gauge their level of stimulation or excitement. The dominance scale ranged from submissive (indicating a lack of control) to dominant (indicating a sense of control or empowerment).

A fourth scale was also included to assess participants' personal liking for the video. It was essential to distinguish this scale from the valence scale, as it measured preferences rather than emotional responses. For instance, participants could like videos that elicited feelings of sadness or anger.

Following the experiment, participants were requested to rate their familiarity with each song on a scale from 1 ("Had never heard it before the experiment") to 5 ("Knew the song very well").

The questionnaire structure is outlined below:

- User ID (i.e., the numerical code representing each participant).
- Video ID (i.e., the numerical identifier associated with each of the 40 videos).
- Valence score (Likert scale 1-9).
- Arousal score (Likert scale 1-9).
- Dominance score (Likert scale 1-9).
- Liking (3 options: liked, neutral, dislike).
- Familiarity with the song heard (Likert scale 1-5, ranging from "Never heard of it before the experiment" to "I know it very well").

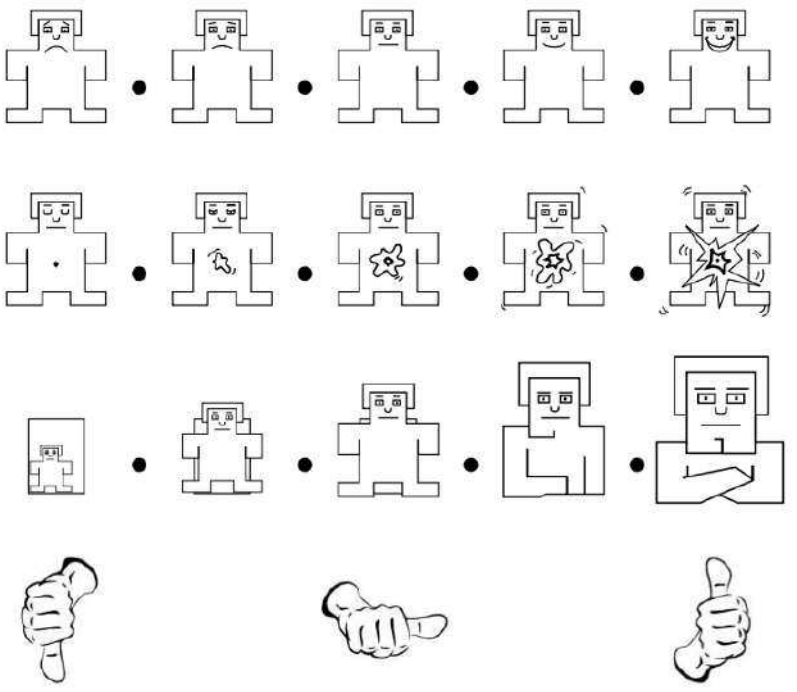


Fig. 5.30 Self-Assessment Manikins (SAMs) used in the experiment. The figures show the SAMs used by participants to rate their emotional responses across three dimensions: valence (ranging from unpleasant to pleasant), arousal (ranging from calm to excited), and dominance (ranging from submissive to dominant). Additionally, participants rated their liking for each video using a three-option scale: liked, neutral, or dislike.

5.7.3 System Architecture

We developed a ML framework to test the effectiveness of ML on the collected EEG data.

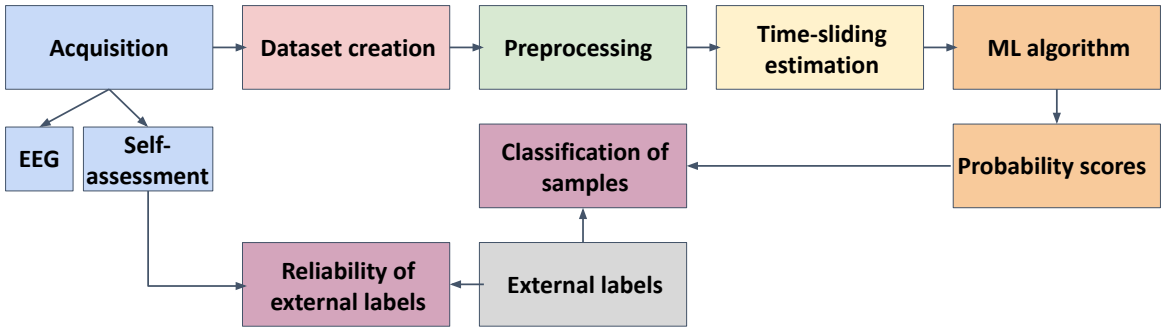


Fig. 5.31 Illustration of the workflow representing all the key steps in our proposed system for emotion recognition using EEG data. The figure outlines the following stages: data acquisition via the Muse 2 EEG device, data creation and organization, preprocessing of EEG signals (including noise reduction, filtering, and epoch segmentation), time-sliding estimation for identifying ROIs, machine learning-based classification of emotional states, and statistical analyses for correlation and participant reliability assessment.

The architecture of our ML framework, as illustrated in Figure 5.31, is composed of six principal components, each playing a crucial role in the processing and analysis of EEG data for emotion recognition:

- acquisition module: this component collects EEG data, utilizing the Muse device as the primary source.
- Data creation: this module systematically collects and organizes EEG data from each participant, ensuring precise alignment with the timestamps of baseline periods and stimuli presentations.
- Preprocessing module: upon receiving EEG signals from the Muse device, this module applies a series of processing techniques to refine the data for analysis. The preprocessing pipeline enhances signal quality by eliminating noise and artefacts commonly associated with EEG data. The steps involved are as follows:
  1. creation of an EEG epochs structure, utilizing the MNE framework [30] for efficient data segmentation and organization.
  2. Application of a Finite Impulse Response (FIR) filter with a cutoff frequency range of $1-45$ Hz with fir window='hamming' to isolate the relevant frequency bands for emotion recognition.

[30] https://mne.tools/stable/index.html

3. Implementation of ringing artifact reduction techniques, specifically devised to remove non-brain artefacts without compromising the integrity of the EEG signal.
4. Division of the processed EEG data into sub-epochs of 5 seconds each, facilitating a granular analysis of emotional responses over time.

- Time-sliding estimation: a dynamic strategy to pinpoint the most informative regions of interest (ROI) within the EEG signals as they evolve over time.
- ML module: this module is at the heart of our framework, where different ML algorithms are employed to interpret the preprocessed EEG signals. Leveraging state-of-the-art models, this component predicts the emotional state represented by the input data.
- Statistical analyses involve two main steps. First, the Pearson correlation between self-assessment scores and external labels for arousal and valence is calculated to evaluate label reliability. Second, the self-assessment and ML probability scores are analyzed to identify non-trustworthy or poorly performing participants.

A description of the operations performed is provided in the Algorithm 1. It details all the previously described steps.

Moreover, additional details about each step are provided in the following subsections.

Dataset Creation

The development of our dataset formed the foundational phase of our research. We meticulously collected EEG data files for each participant, augmented by timestamps indicating the onset of baseline periods and specific stimuli. This compilation was methodically structured into a dictionary format, integrating the EEG recordings with corresponding timestamps and markers for the pertinent stimuli.

After data collection, we delineated EEG epochs for each experimental condition, including baseline and stimulus-exposure intervals. These epochs were then subjected to a series of preprocessing steps. The dataset will be publicly available at https://sisinflab.poliba.it/neurosense-dataset-request/.

Preprocessing

The EEG signal preprocessing process is crucial due to its sensitivity to noise and artifacts. Initially, we extracted baseline and stimuli epochs and filtered the data

Algorithm 1: Pseudocode for subject-specific data processing and classification

Initialize a data table containing user data, with each user having a unique subject ID
for each unique subject ID **do**
Create subject-specific training dataset
Create subject-specific testing dataset
for each file in the subject's training dataset **do**
Extract trial epochs and corresponding labels from the training dataset
Extract baseline epochs from the training dataset
end for
for each file in the subject's testing dataset **do**
Extract trial epochs and corresponding labels from the testing dataset
Extract baseline epochs from the testing dataset
end for
end for
Initialize a processing pipeline with predefined configurations (e.g., feature extraction, normalization, classification)
Define hyperparameter grids for optimization (e.g., number of features, maximum dilations per feature, regularization parameters)
Initialize grid search for hyperparameter optimization based on the processing pipeline
Train the model using the training data $(X_{\text{train}}, y_{\text{train}})$
Generate predictions and classification probabilities using the testing data $(X_{\text{test}}, y_{\text{test}})$
Store the results in a dictionary for further analysis

between 1-45 Hz using MNE, minimizing artifacts by instructing participants to limit movement. Given the limited electrode setup, we utilized simple denoising methods, treating ocular artifacts as outliers and applying interpolation if samples exceeded thresholds determined by a K-NN algorithm.

For dynamic region of interest (ROI) detection over time, we employed the SlidingEstimator within MNE, fitting models at each time point to capture the temporal evolution of emotional responses. By analyzing EEG in 5-second intervals across the emotional quadrants, we identified the ROIs best distinguishing between baseline and target emotions, selecting intervals based on accuracy percentiles. This comprehensive pipeline was repeated for each emotion, refining our understanding of emotional response dynamics.

Machine Learning Algorithm

In this study, we utilized a binary logic framework to identify high-performing models across Russell's circumplex model's four emotional quadrants. This model categorizes emotions along two axes: arousal (high vs. low) and valence (positive vs. negative), providing a structured approach to understanding emotional states. By framing our task as a multi-class problem within this dimensional space, we sought to decode these emotional states from EEG data. Classification accuracy within these quadrants serves as a proxy for assessing how well EEG data can capture underlying emotional states.

Our methodology implemented a Leave-One-Subject-Out (LOSO) validation strategy to ensure robust model performance. Post-training, we selected the top 30 models based on predefined performance metrics for further analysis and prediction.

Feature extraction was performed using the MiniRocket algorithm, a streamlined version of Rocket designed for efficient time series feature extraction [73], as implemented in the sktime library [173]. This process involved convolving each time series with random kernels, performing global max pooling, and utilizing the proportion of positive outcomes from each convolution as features. These extracted features, capturing critical patterns in multivariate EEG data, supported the classification task using a Support Vector Machine (SVM).

Hyperparameter optimization was conducted within Python pipelines, initially adjusting kernel size and length in MiniRocket, followed by selecting an appropriate normalization method (MinMaxScaler, StandardScaler, or RobustScaler [218]) and fine-tuning the SVM's regularization parameter C. Evaluations were carried out through 3-fold cross-validation within each LOSO round, with hyperparameters optimized via Random Search across 50 iterations. We implemented the ML models and evaluations

in Python, utilizing the Scikit-learn library [218].

Statistical Analysis

We conducted an initial analysis to evaluate the correlation between the self-assessment scores and the external labels, providing a foundation for the subsequent ML model development. In particular, the Pearson correlation index was calculated for arousal and valence scores averaged across participants for each video.

Classification of Samples

After training the ML model to predict the quadrant location of each video on a bivariate plane defined by arousal and valence scores, we used 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^{th} 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.

5.7.4 Key Findings and Experiments

Reliability of External Labels

Figures 5.32, 5.33, and 5.34 display the results of self-assessment concerning valence, arousal, and dominance scales, respectively. The X-axis represents the 30 participants (user IDs), while the Y-axis depicts the 40 video IDs.

The correlation analysis reported in Figure 5.35 revealed varying degrees of alignment between self-assessment scores and external labels. Generally, there was a strong correlation for valence, while arousal and dominance correlations were comparatively weaker and not significant. This discrepancy underscores the challenge of matching subjective self-assessments with external labels obtained from a broader population database.

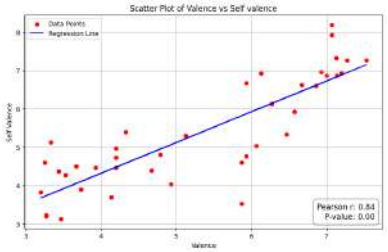
Nevertheless, external labels derived from a large-scale population study provided a standardized and objective measure of emotional response, essential for the consistency

required in ML model training and evaluation. This approach mitigated individual variability and biases inherent in self-assessment data, allowing for more robust model performance. Hence, we opted to utilize the Valence-Arousal Quadrant Estimate available in the DEAP dataset.

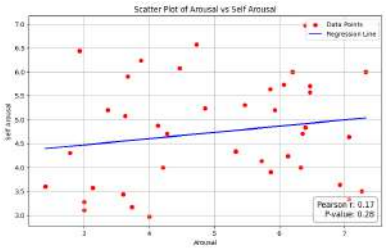
Fig. 5.32 Representation of participants' self-assessed *Valence* scores resulting from the SAM questionnaire.

Fig. 5.33 Representation of participants' self-assessed *Arousal* scores resulting from the SAM questionnaire.

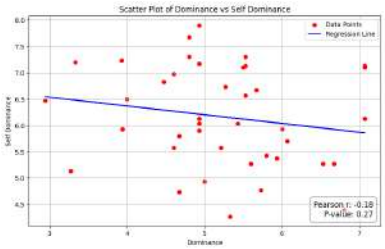
Fig. 5.34 Representation of participants' self-assessed *Dominance* scores resulting from the SAM questionnaire.



(a) Correlation between external and self-assessed valence scores.



(b) Correlation between external and self-assessed arousal scores.



(c) Correlation between external and self-assessed dominance scores.

Fig. 5.35 Correlation between external and self-assessed indexes.

Region of Interest Computation

Table 5.15 presents the outcomes of the ROI analysis in terms of scores associated with each of Russell's quadrants across different time segments. The results indicate that the most effective time segment for Russell's quadrants is the interval from 0 to 5 seconds. Hence, our pipeline enabled us to identify the EEG signal trial that optimally discriminates between EEG data corresponding to emotional stimulation and EEG data from the resting condition using binary classification methods.

Table 5.15 Time-sliding accuracy.

Time Segment [*seconds*]	**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	55.6
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) In detail, columns two, three, four, and five represent the average frequency with which the accuracy exceeds the threshold in the first, second, third, and fourth Russell quadrants. RQ = Russel's quadrant.

Classification Performance

Table 5.16 shows the average accuracy and standard deviation of each model in the LOSO training strategy for binary classification between stimuli and baseline conditions.

Table 5.16 Average model accuracy scores and standard deviations for binary classification between stimuli and baseline conditions.

Models	**Average Test score Accuracy**	**Std score**
1	0.77	0.035
2	0.78	0.022
3	0.76	0.024
4	0.80	0.028

The results described above highlight that for most users, we observe consistent accuracy levels above 75%, with standard deviations indicating relatively stable model performance across LOSO cross-validation splits. This suggests that our EEG-based ML framework reliably classifies emotional states and is able to distinguish between baseline conditions and stimuli with high accuracy, making it a promising tool for emotion recognition using EEG data from sparse electrode configurations.

5.7.5 Implications and Future Prospects of NeuroSense

NeuroSense introduces a novel dataset, accessible to EEG-based emotion recognition using the low-cost Muse 2 device with only four electrodes, challenging the standard reliance on dense EEG setups. By achieving an average accuracy above 75% in

classifying emotions across Russell's circumplex model, NeuroSense demonstrates the feasibility of accurate emotion detection with minimal hardware.

Our methodology combines an adaptive preprocessing pipeline tailored for sparse EEG signals and incorporates a dynamic time-sliding estimation to pinpoint regions of interest (ROI) within EEG epochs. This approach, alongside machine learning models such as MiniRocket and SVM, enables robust feature extraction and classification. Utilizing a Leave-One-Subject-Out (LOSO) validation strategy, we ensured model stability and generalizability across participants, a critical factor for practical applications.

The NeuroSense dataset was validated against self-assessment scores, revealing strong correlations for valence, underscoring its reliability for future affective computing research. By simplifying EEG setups without compromising accuracy, NeuroSense opens new avenues for emotion recognition in human-computer interaction and mental health applications where portability and affordability are crucial. This project highlights the potential of sparse EEG configurations and invites further exploration of continuous emotional state tracking in real-world settings.

5.8 Summary

In this chapter, we presented a progression of research aimed at harnessing the potential of EEG-based emotion recognition and machine learning within BCIs to address a range of challenges in emotional support, neurorehabilitation, and creative expression in music composition. The series of studies introduced here represents a comprehensive exploration of EEG and ML technologies in diverse applications, with a focus on the practical and theoretical advancements each project contributes to the field of affective computing and beyond.

Our initial study explored the classification of human emotions using EEG signals, demonstrating that ML techniques could achieve reliable, real-time emotion detection. This foundational work underscored the feasibility of capturing complex emotional states with high accuracy, establishing a robust basis for the development of responsive and interactive applications.

Building on this groundwork, we introduced ARIEL, an innovative framework that merges EEG-driven emotion recognition with large language models (LLMs) to create an emotionally aware conversational agent. ARIEL represents a notable advancement in affective computing, where real-time emotion analysis enables the system to dynamically adapt its conversational approach. This approach not only fosters emotionally supportive interactions but also sets a new standard for integrating

emotion detection into AI-driven dialogue systems, with potential applications in mental well-being support and therapeutic engagement.

Extending these advancements to the field of neurorehabilitation, we proposed a biofeedback system that leverages EEG signals to classify users' mental states into Focus and Relaxation categories. This system then tailors music-based feedback to enhance therapeutic engagement, aligning with research that emphasizes the importance of adaptive, emotion-sensitive interventions in accelerating recovery processes. This work demonstrates the potential of ML-driven, EEG-based biofeedback systems to personalize neurorehabilitation experiences and improve patient outcomes.

Our subsequent study, NeuralPMG, applies EEG-based emotion recognition to the field of music composition, presenting a novel framework that combines EEG-based mental state detection, Slonimsky's theoretical grammar, and interactive tools for professional composers. The system leverages EEG data to recognize Focused and Relaxed mental states, using this information to control aspects of music generation. While the system was validated in a laboratory setting, future research involving real-world scenarios is anticipated to assess its long-term usability and creative impact. Furthermore, incorporating additional mental state parameters, such as stress and interest, could expand the system's capabilities and offer deeper insights into the creative process.

Concluding the chapter, we introduced the NeuroSense dataset, a novel resource aimed at making emotion recognition more accessible and practical by using a low-cost, four-electrode EEG device (Muse 2). This open-source dataset demonstrates that even with a sparse electrode setup, it is possible to achieve competitive performance in recognizing emotional states, thus lowering barriers to entry for affective computing research. NeuroSense underscores the potential of low-cost EEG devices to expand applications in real-world settings, particularly in fields like mental health monitoring and human-computer interaction, where simplicity and accessibility are crucial. The dataset invites further research and innovation, serving as a foundational tool for developing regression models that could capture a continuous spectrum of emotions in line with the current trends in affective computing.

Looking forward, several avenues for future work emerge from these studies. For ARIEL, in-vivo studies in clinical settings could validate its effectiveness in providing emotional support. For NeuralPMG, deploying the system in authentic composition environments will allow for a more comprehensive evaluation of its effectiveness in the creative field. Finally, further exploration of the NeuroSense dataset using advanced ML models could enhance the understanding of emotion-related EEG patterns and

improve model accuracy. The potential for continuous emotion regression models in NeuroSense is also promising, aligning with trends in affective computing that emphasize the importance of capturing the full spectrum of human emotional states.

In summary, the works presented in this chapter highlight the transformative role of affective BCIs and AI in creating empathetic, responsive technologies. These technologies offer new possibilities for personalized healthcare, interactive assistance, and creative expression, underscoring their growing impact on mental well-being, neurorehabilitation, and artistic domains. Collectively, these contributions suggest promising directions for future research and practical applications that will continue to push the boundaries of affective computing.

Chapter 6

Applications of Artificial Intelligence in Bioelectrical Signals

6.1 Research Objectives and Chapter Contributions

This chapter explores the advanced use of AI and ML in biosignal analysis, extending the scope beyond EEG to include other physiological signals such as ECG and EMG. With applications in both health care and sports performance, addressing critical gaps in the current literature, particularly with regard to early diagnosis, movement analysis and personalised interventions. Through the development of interpretable ML models, we aim to support timely clinical decisions and enhance research in sports biomechanics by highlighting the potential of biosignal processing in different real-world contexts.

6.1.1 Objectives

Chapter 6 explores advanced applications of AI and ML in the analysis of biosignals, extending the methods used for EEG to other physiological signals such as ECG and EMG. The overarching aim is to harness machine learning for versatile applications in healthcare diagnostics and sports performance, addressing gaps identified in existing literature. Our research objectives are outlined below.

We first aim to develop and apply advanced ML and XAI techniques to ECG data, targeting the early detection and classification of cardiac conditions, including Normal Sinus Rhythm (NSR) and Arrhythmia (ARR). This goal emphasizes creating interpretable models that not only achieve high predictive accuracy but also provide actionable insights for clinicians, enabling more timely and informed medical decisions.

The second objective focuses on leveraging EMG data in the realm of sports biomechanics, specifically targeting performance analysis and injury prevention in fencing. By integrating ML models with sensor data, we aim to classify and distinguish movement patterns among novice and elite athletes, providing valuable feedback for optimizing training and refining athletic techniques.

Finally, we aim to evaluate the potential of wearable sensor technologies and real-time monitoring systems for dual applications in clinical and sports settings. This includes assessing the cross-domain adaptability of biosignal processing methods and their role in developing personalized, data-driven interventions.

6.1.2 Chapter Contributions

The contributions of this chapter are multifaceted, advancing the field of biosignal analysis through the following key innovations:

- **Advanced ECG Analysis:** We have created a robust machine learning framework that extracts and analyzes temporal and morphological features from ECG data. By incorporating XAI methods, we ensure model transparency, making our approach clinically relevant and interpretable for medical professionals.
- **EMG-Driven Biomechanical Insights:** We introduce ML models for classifying biomechanical performance in sports, with a particular emphasis on fencing. This work not only distinguishes between different levels of expertise but also provides actionable insights to improve athletic performance and prevent injuries.
- **Cross-Domain Applicability:** Our research demonstrates the flexibility and adaptability of biosignal processing techniques across various physiological data types. This highlights the potential for real-time, personalized applications in both healthcare and sports science.

6.1.3 Limitations of Current Methodologies

Despite significant advancements, current methodologies for biosignal analysis still encounter several critical limitations, which this thesis seeks to address.

ECG Analysis: Existing ECG-based approaches for heart disease detection frequently face challenges related to data quality and diversity. While high accuracy can often be achieved within individual datasets, models tend to exhibit substantial performance drops when applied to new, unseen datasets. This highlights issues with

data generalizability and robustness, as well as persistent problems like class imbalance that hinder the reliability of these systems across diverse populations.

EMG in Sports Biomechanics: In the realm of sports science, the precision of ML-driven motion analysis is highly dependent on factors such as sensor placement and the quality of data collected. Even slight misalignments or inconsistencies in sensor calibration can significantly impact model performance. Furthermore, many current methods lack explainability, making it difficult for athletes and coaches to interpret the results and apply them effectively for training and injury prevention.

Explainability and Interpretability: While ML models such as CNNs and Random Forests deliver impressive performance, their complexity can make them difficult to interpret, a critical limitation in healthcare and sports applications. The absence of XAI methods in most existing studies restricts the models' practical usability and acceptance in real-world scenarios.

By addressing these limitations, this chapter aims to contribute a comprehensive and interpretable framework for biosignal analysis, enhancing the usability and generalizability of AI-based models in both clinical and sports settings.

6.2 Background

To begin, it is essential to outline the background and guiding principles that shape the research focus highlighted in this chapter.

Machine Learning Applications in Sports Biomechanics and Fencing Training

ML and its applications have become of central interest to sports scientists, primarily due to the substantial impact on performance enhancement and injury prevention. In recent years, ML has significantly transformed various aspects of sports science, including:

- Data collection devices used in biomechanics;
- Insights derived from device-captured data, such as 3D kinematics and vertical ground reaction forces;
- Data processing, where classification methods can now efficiently separate and analyze data previously requiring extensive manual analysis;
- Enhanced understanding of athletic performance and predictive models for injury risk.

The increasing adoption of ML is expected to notably advance objectivity in decision-making within sports science over the coming decade. Research shows ML's effectiveness in predicting future injuries based on pre-season measurements, identifying movement strategies within athlete cohorts, and recognizing injury-prone movement patterns in otherwise healthy individuals [235].

Application of MLL in Fencing Biomechanics

Malawski et al. [186] investigated the effectiveness of inertial measurement units (IMUs) and Kinect for classifying fencing footwork to aid athlete and coach training. Using a SVM model, they achieved high recall and precision (100%) using Kinect, while the IMU-based technique yielded 99.38% recall and 98.77% precision. Although effective, this study suggested further improvements in accuracy for parameters like hand timing, potentially enhancing fencers' movement refinement. The authors noted that while Kinect is ideal for frontal postures, it is less suited for side-view tracking as required in fencing. They highlight that for IMU to match Kinect's performance, precise body positioning of sensors is crucial [186].

O'Reilly et al. [205] placed five IMUs on the lumbar spine, thighs, and shanks to classify correct and incorrect lunge techniques as well as variations in lunge execution. They evaluated SVM, k-Nearest Neighbors, Naïve Bayes, and Random Forest classifiers, finding Random Forest optimal due to its superior classification performance. Using permutation feature importance, they identified key factors contributing to lunge accuracy. For binary classification, the setup with five IMUs achieved 90% accuracy, while a reduced configuration using three IMUs (lumbar and both shanks) maintained similar classification reliability [205].

Another study by Malawski et al. [187] focused on classifying core fencing movements (step forward, step backward, rapid lunge, etc.) in a group of 10 fencers using a single IMU placed on the knee. The authors compared Dynamic Time Warping (DTW), SVM, and other models, concluding that SVM, based on time-domain features, provided high recognition accuracy and superior generalization potential. Their findings underscore the importance of appropriate feature selection and data segmentation techniques, especially when processing data for classification [187].

Machine Learning in Hearth disease

For the work related to ECG analysis, this bibliography has been reviewed as it provides a comprehensive overview of prominent methodologies, addressing diverse challenges

in signal preprocessing, feature extraction, and classification accuracy across different ECG datasets.

Merdjanovska et al. [194] developed a deep learning model trained on four different arrhythmia-annotated datasets. Their approach aimed to address class imbalance using Synthetic Minority Over-sampling Technique (SMOTE) and trained a CNN on segmented raw signals across two scenarios: (i) cross-dataset validation, where training occurred on one dataset and testing on another, and (ii) dataset-specific training for network validation. Cross-database testing revealed lower metrics for Positive Predictive Value (PPV), True Positive Rate (TPR), and f1-score compared to accuracy, indicating challenges with data diversity and model generalizability.

Sannino et al. [246] employed a deep learning method on the MIT-BIH Arrhythmia dataset, using denoising, peak detection, and segmentation to preprocess the data, followed by RR interval-based temporal feature extraction within a sliding window. They employed subsampling to address class imbalance, ensuring unique subjects across training and test sets. A trained CNN demonstrated accuracy, sensitivity, and specificity above 98%, underscoring the model's robust performance within the dataset.

Ali et al. [7] utilized a data mining approach with filtering to replace missing values and applied various supervised algorithms. High sensitivity, specificity, and accuracy metrics were achieved across K-nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), AdaboostM1 (ABM1), and Multilayer Perceptron (MLP) models. Using a heart disease dataset from Kaggle, they reported 100% accuracy, sensitivity, and specificity with the RF model. Their approach primarily utilized biological features instead of ECG signal characteristics.

Malakoutiica [190] explored multiple classifiers, including Naïve Bayes Gaussian (NBG), RF, LR, and Linear Discriminant Analysis (LDA), using 10-fold cross-validation to mitigate metric variance. NBG performed best with 96% accuracy and 97% precision, although none of these approaches analyzed explainability.

All this reviewed works was essential for the development of the proposals outlined in this chapter.

6.3 Explainable AI in Cardiac Health: Enhancing Transparency in Heart Disease Diagnosis

This section opens with an analysis of research focused on leveraging AI and ML in the detection of cardiovascular disease (Cardiovascular Diseases (CVD)), with a particular emphasis on Heart failure (HF). HF, a condition directly associated with CVD, is a

significant global health issue, ranking as a leading cause of morbidity and mortality. Given the limitations of traditional diagnostic methods—such as patient history and physical examination—there is an urgent need for advanced tools, particularly those involving ECG-based approaches, to enhance the early detection and management of HF.

In this work, the focus is on extracting temporal and morphological features from ECG signals to compare the performance of various ML classification models in HF detection. The authors propose a Light Gradient Boosting (LGBM) model, demonstrating its efficacy in distinguishing Normal Sinus Rhythm (NSR) from Arrhythmia (ARR) with a high accuracy of 0.99, alongside strong precision and recall metrics. To further enhance the interpretability of model predictions, the study integrates XAI techniques, which clarify the model's decision-making process and provide meaningful insights for healthcare professionals.

With the growth of big data and wearable devices, ML models in predictive medicine present significant potential for early HF detection and scalable healthcare solutions. This work has two primary objectives: creating a reliable classification model for ECG signals and using XAI for post-hoc analysis to understand the prediction process. The study examines these objectives through experimental evaluations using data from the MIT-BIH PhysioNet database, exploring the following Research Questions (RQs):

- **RQ1:** Can the newly extracted ECG signal-based features train a classification model to accurately predict heart failure?
- **RQ2:** Which model trained on our new ECG signal-based features is most effective in predicting heart failure?
- **RQ3:** Can each subject-based prediction related to the best model classification be explained to support physicians in diagnosis?

All these research questions focus on to the prediction of heart failure in a real-time environment.

To address these questions, a comprehensive experimental evaluation was conducted using signals from the MIT-BIH PhysioNet Database[1], widely recognized for its datasets containing cases of arrhythmia and normal sinus rhythm.

[1]https://physionet.org/content/mitdb/1.0.0/

6.3.1 Methods

Our work employs ECG signals from two datasets. Data went through a selection step with which only files containing information and annotations are kept. Then, the remaining signals are preprocessed with three steps: filtering and segmentation (section 6.3.1), and subsampling (section 6.3.1). Then, features are extracted based on the records sampling rate.

Dataset Description

ECG recordings are collected from the MIT-BIH physio-net database [99]. The considered conditions in order to build the dataset employed in this study are listed below:

- MIT-BIH Arrhythmia (ARR) database [197]: consists of 48 ambulatory two-channel ECG recordings sampled at 360Hz from 47 subjects. The subjects included in the study were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. The recordings are 30 min long and include both arrhythmia and normal sinus rhythm recordings as well as 4 recordings with a pace-maker. Also, rhythm and beat annotations by experts are provided for each record.
- MIT-BIH Normal Sinus Rhythm (NSR) database [125]: consists of 18 ambulatory long-term ECG recordings sampled at 128Hz. The recordings are from healthy adult subjects distributed as follows: 5 men aged 26 to 45 years, and 13 women aged 20 to 50 years. The duration of each recording is about 24 hours. The dataset includes both rhythm and beat annotations.

Since the classification is performed beat-wise, paced beats were excluded in order to maintain data consistency.

Data Preprocessing

The preprocessing pipeline was designed to ensure data consistency and optimize model performance. The dataset was parsed using the `pandas` library[2], removing inconsistent data to retain only entries containing raw signals and annotations. For computational efficiency, Normal Sinus Rhythm (NSR) signals were segmented into one-hour intervals. Noise filtering and segmentation were performed using NeuroKit2 [185],

[2]https://pandas.pydata.org/

a Python library specialized in neurophysiological signal processing, which was also employed to extract relevant ECG signal peaks (P, Q, R, S, T) and their corresponding segment boundaries.

The dataset was split into 74% training and 26% test sets to avoid overlap of subjects across the sets. Continuous features were standardized using the `RobustScaler` [3] technique [230] to reduce the impact of outliers.

Subsampling

Beat annotations centered on R peaks were used to segment beats for each sample in the ML models. NeuroKit2's segmentation functions helped extract beat segments as `pandas` DataFrames, labeling each beat according to R peak annotations. Normal beats were taken from NSR signals, while only abnormal beats were extracted from ARR records. Non-beat annotations were excluded to maintain dataset consistency.

Feature Extraction and Selection

From the segmented recordings, 26 initial features were computed, including:

- Duration of each segment in a beat (e.g., P, PQ, QRS, QR, QT, RS, ST, T, PT segments);
- Angles at key points (e.g., $\angle P_{on}PQ$, $\angle R_{off}QR$, $\angle QRS$, $\angle RST$, $\angle STT_{off}$);
- Segment slopes (e.g., PQ, QR, RS, ST slopes);
- Voltage values at each peak (e.g., P, Q, R in [mV]);
- Proportional values of QR/QS and RS/QS to represent durations within the QRS wave.

Each feature's duration was calculated as sample distances divided by the sampling frequency to obtain values in milliseconds. After removing samples with missing values, a recursive feature selection (Recursive Feature Elimination (RFE)) with 5-fold cross-validation was performed using `GridSearchCV`[4] across eXtreme Gradient Boosting (XGBoost) [52], RF [25], and LGBM [146] models to identify optimal model parameters.

[3] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

[4] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

The final feature set consisted of 13 selected features: PQ segment, QRS segment, RS segment, ST segment, $\angle QRS$, $\angle STT_{off}$, PQ slope, QR slope, RS slope, ST slope, and peak values for P, Q, and R. The dataset was undersampled with the RandomUnderSampler technique from the `imblearn` library [162] to balance the classes, resulting in 8259 samples each for NSR and ARR classes, and the 13 selected features as input for the models.

Classification Models

To identify the most effective classifier for predicting the HF condition, we compared several models: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting (LightGBM). These models were implemented in Python using the Scikit-learn[5] [219], xgboost libraries[6],and LightGBM[7]. We employed a GridSearchCV (5-fold) cross-validation to tune the classifiers, maximizing metrics including Precision, Recall and f1-score to achieve optimal classification performance. Table 6.1 presents the range of hyperparameters values explored for each model during the tuning process.

Table 6.1 Hyperparameter list and values for the classification models reported in this work. x_train.shape[1] is equal to 13

Algorithm	**Hyperparameter**	**Values**
Random Forest	seed	{42}
	n_estimators	{np.arange(50,201,50)}
	max_features	{np.arange(1, x_train.shape[1])}
	max_depth	[4, 6, 8]
eXtreme Gradient Boosting	seed	{42}
	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_train.shape[1])}
Light gradient boosting	seed	{42}
	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_train.shape[1])}

[5]http://scikit-learn.org
[6]https://xgboost.readthedocs.io/en/stable/
[7]https://lightgbm.readthedocs.io/en/stable/

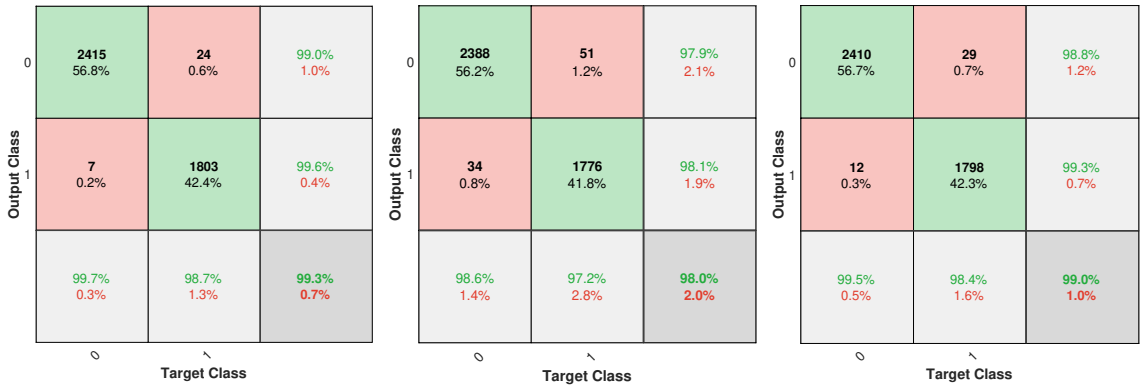
Explainability Techniques

This study adopts SHAP [178] approach to explain the predictions of the best model for the independent test set. SHAP represents the marginal contribution of each input variable in model decision making. This algorithm is based on game theory and in particular Shapley's approach to evaluate the contribution of each player in a cooperative game. SHAP introduces a variant of Shapley's approach through the use of a local contribution function, which calculates the contribution of each variable for each input instance of the test set.

6.3.2 ECG Model Analysis

In this section we detail the results obtained from our experimental approach. In particular, in Table 6.2, we report the results obtained from the identified best model. To determine the best, we also analyze the performance of these models in terms of precision, recall and f1-score for each class. From the table 6.2, we assert that LGBM with $learning\_rate = 1$, $max\_depth = 4$, $max\_features = 1$, $n\_estimator = 200$ is the best model, as it performs better than RF with respect to all considered metrics, while, compared to XGB, LGBM obtains equal performance except for precision and recall on the arrhythmia class where LGBM reaches better performance. In summary, LGBM, according to the *confusion matrix* in Figure 6.1, obtains the lowest number of misclassified samples for both classes. Indeed, the performance achieved by LGMB allows minimizing the number of misclassified, and reaches an accuracy of 0.99, the best value of precision (1.00), and recall (0.99), in the prediction of *Normal Sinus*, and a high value of f1-score in the prediction of *Arrhythmia* (0.99). Figure 6.2 reports a graphical representation of feature importance global explanation (in descending order from top to bottom). Moreover, the horizontal position of each point shows whether the feature moves the prediction towards positive or negative class and colors refer to the feature value, red for high values and blue for lower ones. Specifically, all features seem to have the same importance except for $\angle QRS$ $PQsegment$, $\angle STT_{off}$, $PPeakvalue$ and $QPeakvalue$ which contribution results to be marginal. This behaviour reflects on figures 6.3 and 6.4 which provide examples of local explanation on a negative and a positive sample, respectively. In detail, the negative sample in figure 6.3 shows that $RPeakvalue$ is the most important feature, furthermore $RSsegment$, $QRSsegment$ and $PQslope$ have almost the same cardinality. Whereas, all features from $\angle STT_{off}$ onwards appear to provide marginal contribution to the prediction, resulting less discriminating. On the other hand, a different trend is presented in figure 6.4 where features have different importance in classifying a positive

sample. Besides *QRSsegment* and *RSsegment*, *STslope* seems to provide a lot of information to the classification of positive samples rather than negative ones.



(a) LGBM (b) Random Forest (c) XGBoost

Fig. 6.1 Confusion matrix of the considered models.

Model	Condition	Precision	Recall	F1-score	Accuracy
LGBM	Normal sinus	1.00	0.99	0.99	0.99
	Arrhythmia	**0.99**	**1.00**	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 6.2 Evaluation metrics for all considered models.

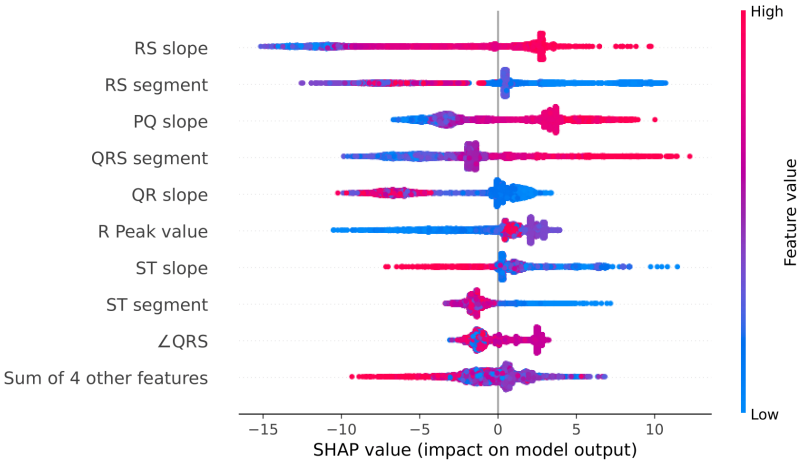


Fig. 6.2 Global interpretation using SHAP of ECG considered features.

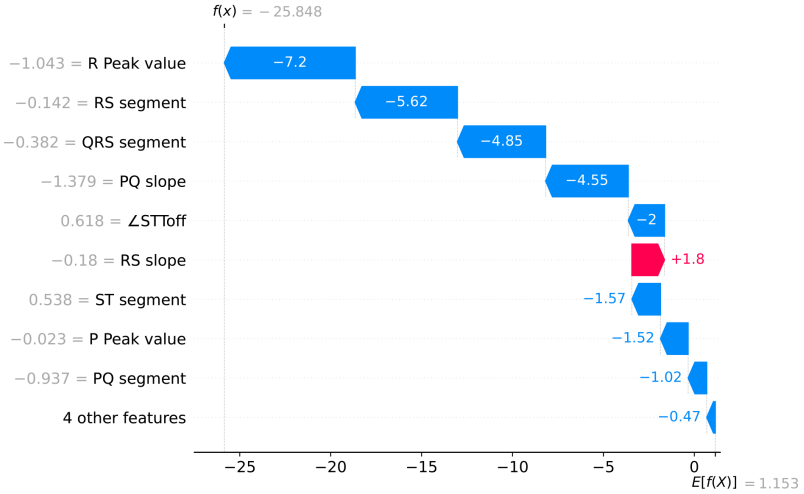


Fig. 6.3 Overview of individual feature contributions in the prediction of an NSR subject

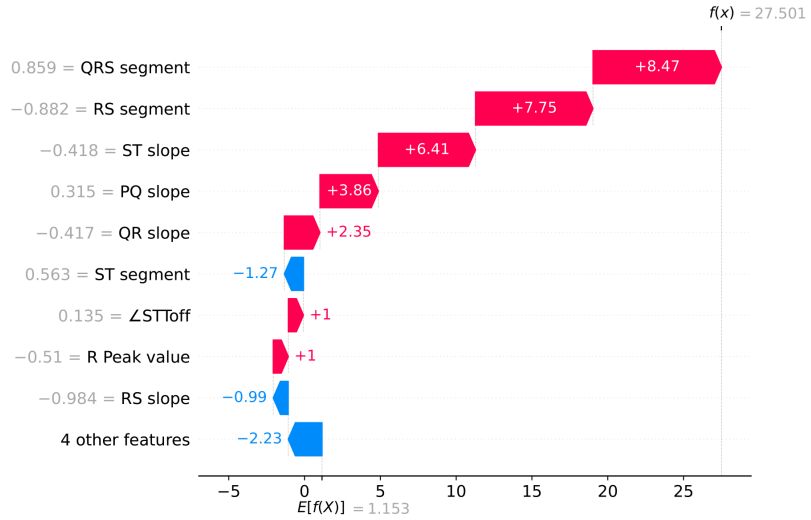


Fig. 6.4 Overview of individual feature contributions in the prediction of an ARR subject

6.3.3 Analysis Outcomes and Clinical Implications

In this work, we present an approach based on feature extraction, feature selection, and XAI to identify the optimal ML model for predicting HF in real-time. By training and testing various ML algorithms, we determined the model best suited for the task, utilizing only the most discriminative features. This approach is advantageous as it enhances model selection while ensuring interpretability of the prediction process.

Our HF classification approach shows high effectiveness in distinguishing arrhythmic cases, which is critical for medical applications. This framework could support both epidemiological studies and clinical screenings by prioritizing a reduced set of features derived from raw signals, potentially decreasing time and resource demands for further testing.

RQ1: Can extracted ECG signal-based features train a model to accurately predict HF? Our study demonstrates the feasibility of using specific ECG features for HF prediction. By leveraging these features, we identified informative characteristics within the signals that were used to train our models, underscoring the potential for further exploration to enhance predictive capabilities.

RQ2: Which model is most effective for HF prediction based on these ECG features? Through extensive evaluation, we found that Light Gradient Boosting (LGBM) achieved the best performance, with an accuracy of 0.99, a precision of 1.00, a recall of 0.99, and an F1-score of 0.99 for the NSR class, and similar metrics for the ARR class. These results align with the current state-of-the-art, highlighting the suitability of classification models for HF prediction tasks.

RQ3: Can the model predictions be explained in a clinically meaningful way? Integrating XAI techniques, our approach provides a transparent view of the model's predictions. The XAI analysis explains the feature contributions, offering clinicians interpretability of model predictions and enhancing the model's usability as a support tool in clinical decision-making.

Future Directions and Applications

While this study utilized a limited dataset, future work will focus on expanding the dataset to cover a broader range of heart diseases, enabling further testing with DL and anomaly detection models. In addition, this study serves as a basis for evaluating model interpretability both for ML engineers, in understanding feature relevance, and for clinical practitioners, who can benefit from a clinically validated, explainable model.

In conclusion, ML models based on carefully selected ECG features have shown substantial potential for HF classification tasks and could be further explored for applications in medical decision-making.

6.4 AI-Based Performance Analysis for Fencers

In line with the overarching themes of this chapter, which explores the application of artificial intelligence to bioelectrical signal processing, we now shift our focus to electromyographic (EMG) signal analysis. EMG provides a unique window into muscle activity and neuromuscular responses, offering valuable insights that can inform personalized therapeutic approaches.

This section introduces a study dedicated to analyzing EMG signals to distinguish different performance levels in fencers. Building on the foundational concepts discussed in the previous sections, this research utilizes machine learning algorithms to process EMG data from elite and novice fencers. The goal is to identify characteristic muscle activation patterns that correlate with skill levels, which could serve as a supportive tool for training and injury prevention in sports.

Through this work, we aim to extend the scope of AI applications in bioelectrical signals beyond EEG and ECG, showcasing the potential of EMG-based analysis for performance assessment and skill differentiation in sports settings. This study thus reinforces the central thesis of this chapter by demonstrating AI's capacity to yield actionable insights from bioelectrical signals across diverse clinical and performance-oriented applications.

Fencing, one of the oldest sports, places considerable demands on the body in terms of neuromuscular coordination, strength, and power, particularly during the execution

of movements like the lunge [54]. The effectiveness of a fencer's technique depends on quick reactions to visual, kinaesthetic, and auditory cues [269]. The rapid execution of "propulsion" and "dodge" actions in both offensive and defensive maneuvers subjects the body to varying forces, with notable impacts on musculoskeletal stability [30, 54]. These biomechanical complexities make fencing an ideal candidate for applied sports science and the analysis of muscle activity patterns.

With advancements in wearable sensor technology, sports biomechanics has greatly benefited from non-invasive, high-accuracy data collection. These sensors allow real-time analysis of athletic movements, which is instrumental in preventing injuries and optimizing performance [29, 41]. Combining these wearable technologies with ML techniques enables the extraction of meaningful biomechanical data, providing essential feedback for both athletes and trainers. Recent studies have shown that elite fencers exhibit distinct muscle activation patterns compared to novices, allowing for a nuanced differentiation of skill levels [54].

Through this work, we aim to extend the scope of AI applications in bioelectrical signals beyond EEG and ECG, showcasing the potential of EMG-based analysis for performance assessment and skill differentiation in sports settings. This study thus reinforces the central thesis of this chapter by demonstrating AI's capacity to yield actionable insights from bioelectrical signals across diverse clinical and performance-oriented applications.

6.4.1 Methods

Thanks to newly available technologies, biomechanical data acquisition in sports can be done using optical or non-optical systems. Optical systems include optoelectronic systems (MoCap), which can be marker-based or markerless, such as the Kinect (RGB camera and depth sensor). Non-optical systems include inertial systems. The following considerations guided the choice of experimental settings reported in this paper. Firstly, effective motion tracking using an RGB camera is highly difficult due to quick motions, the presence of several people, and challenging lighting conditions in training rooms. However, deep learning techniques accurately identified persons in RGB movies. Secondly, MoCaps are computationally too expensive to offer the necessary precision for real-time sports. They demand the use of numerous synchronized cameras and the wearing of many markers, making them expensive and impracticable for use in sporting events. Furthermore, to avoid occlusions, the area between the sensor and the tracked person must remain empty, which is a substantial restriction in training facilities. The athlete also needs to stay within the depth camera's field of

vision. Finally, monitoring joint rotation with depth sensors is challenging, which is crucial in sports [186]. On the other hand, IMUs, even though they must be placed on the athlete, they do not need a well-organized workspace to be used [41]. IMUs can measure acceleration and angular velocity, but because errors can accumulate during the integration of the acceleration data, they are far less helpful for monitoring position and velocity. For these reasons, magnetometer information is integrated using a sensor fusion technique, usually a Kalman filter. Moreover, IMUs have a greater sampling frequency, often between 50 Hz and 400 Hz, as opposed to a typical depth sensor's 30 Hz. When compared to the visual data from the Kinect (RGB camera), the information provided by IMUs regarding acceleration, angular velocity, and the magnetic field is very different. Nevertheless, previous studies have found that both modalities help support real-time sports training. Therefore, when selecting the sensor for sports motion analysis, ease of use may be a deciding factor. It is worth noting that tracking direction with an IMU may be essential for assessing sports actions that involve rotation, such as fencing during the lunge [186].

This section introduces (i) study population, (ii) study design, and (iii) algorithm used for preprocessing and classification purposes. Specifically, section 6.4.1 depicts the study population and inclusion criteria; section 6.4.1 depicts the instrumentation used during data collection; section 6.4.1 depicts the acquisition protocol; while section 6.4.1 depicts the biomechanical data preprocessing to identify the data associated with lunge movement; section 6.4.1 depicts the Principal Component Analysis algorithm used for dimensionality reduction; section 6.4.1 depicts the Machine Learning algorithm selected and used in this study and how Machine Learning algorithms were trained and tested. The data follows the logical flow represented in the Figure 6.5.

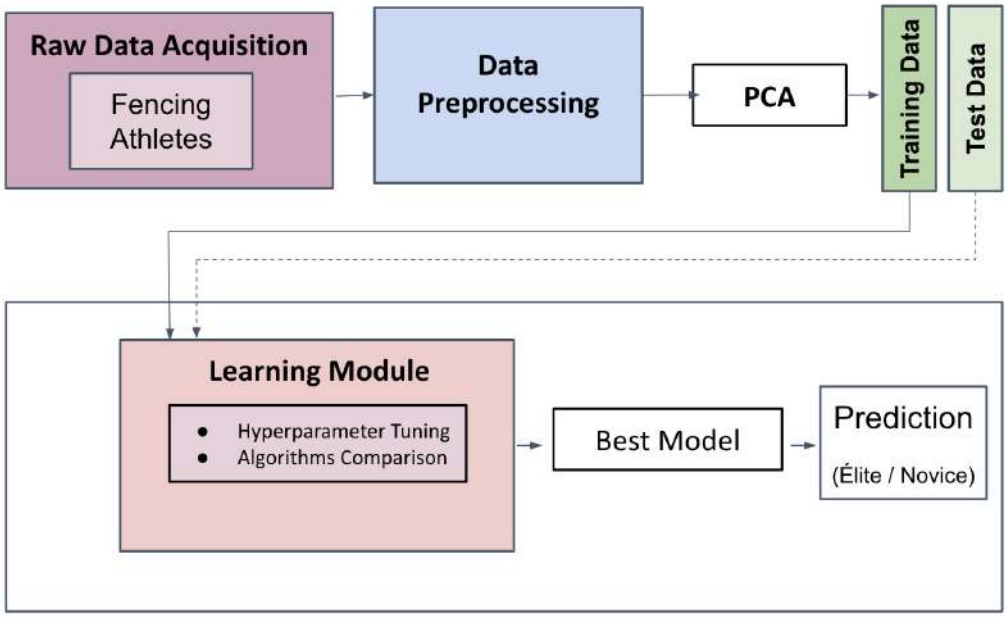


Fig. 6.5 Logical Data Flow.

Data Collection

Twenty-one male and female fencers (17 male and 4 female fencers) participated in this cross-sectional study, all belonging to the A.s.d. CLUB SCHERMA BARI team. Eight of them were classified as novice fencers, while the other thirteen fencers were élite fencers. They practised epée and foil. The age of fencers ranged from 8 to 35 years old. In the last six months before the test, they were free of lower extremity musculoskeletal injuries. Subjects unable to understand the required actions were excluded from the study. Study with its measurements and data collections before starting following the Helsinki Declaration of 1975. All participants provided written informed consent for the study before their trial.

Data collected from the athletes during the study, using the devices described in section 6.4.1, represented the dataset for the ML algorithms, which consisted of samples on the rows and features on the columns. Each subject performed all the different tasks at least two times, for a total of 6 trials per subject placed in rows within our dataset. There were 13 features used to train the different Machine Learning algorithms: tri-axial accelerations, tri-axial angular velocities, tri-axial pelvis angles, and four muscle envelopes. Each of these features is a two-dimensional signal over time each being represented by 500 records. Each column of the dataset is a one-record of the individual feature. Therefore, our dataset had 126 rows and 6500 columns.

Instruments

A professional fencing competition field setup was recreated with a platform measuring 1.80 m in length, equipped with a scoring system at one end to confirm successful touches. Kinematic data were collected using the BTS G-SENSOR 2 (BTS Bioengineering S.p.A., Italy), a wearable inertial motion system with a sampling frequency of 100 Hz. This device includes a triaxial accelerometer (sensitivity: ± 2, ± 4, ± 8, $\pm 16g$), a tri-axial gyroscope (sensitivity: ± 250, ± 500, ± 1000, $\pm 2000^{\circ}/s$), and a magnetometer with a dynamic range of $\pm 1200\mu T$. Positioned at the L5/S1 vertebrae with an elastic band, the IMU transmitted data via Bluetooth to a PC for analysis.

Surface electromyographic (sEMG) signals were gathered using four wearable BTS FREEEMG 1000 probes (BTS Bioengineering S.p.A., Italy), sampling at 1000 Hz with a 16-bit A/D converter. Following SENIAM guidelines, the probes were applied with Ag/AgCl adhesive electrodes on selected muscles: Deltoideus Anterior (DLTA) and RF on the armed side, and Erector Longissimus muscle (LONG) and Gastrocnemius Medialis (GAM) on the opposite side. To ensure stability during movement, an adhesive patch was applied to secure the probes. Both devices are lightweight and compact, thus minimally intrusive during motor tasks. Synchronization of all devices was managed by BTS EMG-Analyzer software, where the testing protocol was defined. Additionally, an iPhone camera on a tripod was positioned 3.80 m from the platform at a height of 1.14 m to capture sagittal plane movements.

Experimental Protocol

Each participant began with stretching and a five-minute warm-up. After familiarization with the test area, biographical (e.g., name, age) and anthropometric data were collected, including weight (W), height (H), height in *en garde* position (HGUARD), front leg length in *en garde* (LLL), thigh circumference (CLL), and armed arm circumference (CUL), alongside category, specialty, and weapon length (Leq). Table 6.3 presents sociodemographic and anthropometric data for novice and élite fencers, with Effect Size (ES) calculated using Cohen's criteria to determine the magnitude of differences between categories.

Table 6.3 Sociodemographic and anthropometric variables by fencers' category (novice and élite). Data are shown as mean ± standard deviation for continuous variables.

Variables	**Novice**	**élite**	**Effect Size (ES)**
Age (years)	10.50 ± 3.14	16.31 ± 5.85	$0.72(0.48, 0.97)$
BMI (kg/m^2)	18.87 ± 3.87	22.29 ± 3.82	$0.39(0.09, 0.73)$
W (kg)	39.30 ± 11.87	61.95 ± 11.86	$0.68(0.48, 0.9)$
H (m)	1.43 ± 0.08	1.66 ± 0.12	$0.75(0.63, 0.9)$
HGUARD (m)	1.32 ± 0.08	1.56 ± 0.13	$0.76(0.63, 0.91)$
LLL (cm)	73.66 ± 7.3	90.00 ± 8.36	$0.74(0.61, 0.9)$
CLL (cm)	40.70 ± 6.38	50.28 ± 9.58	$0.53(0.26, 0.83)$
CUL (cm)	22.70 ± 5.55	27.46 ± 7.67	$0.44(0.12, 0.77)$
Leq (cm)	79.5 ± 2.67	88.75 ± 2.23	$0.86(0.77, 0.95)$

After preparing the skin, sEMG probes were attached to the muscles. Each subject performed four motor tasks to record Maximum Voluntary Contraction (MVC) for each muscle, with each test lasting 30 seconds, repeated three times with 5 seconds rest in between. The inertial system was then secured to the fencer, who received instructions on the tasks.

The fencing test consisted of three exercises:

- *Explosive lunge*: A rapid lunge aimed at hitting the target as quickly as possible.
- *Step forward lunge*: The fencer steps forward to reach lunging distance before scoring a hit.
- *Step back lunge*: The fencer steps back to adjust to lunging distance before scoring.

Each task started from a static position; upon command, the fencer moved to the *en garde* stance, performed the lunge, and returned to *en garde*. The lunging cycle is shown in Figure 6.6, illustrating the start position (a), *en garde* (b), and final lunge (c). Each task was repeated twice, with distance self-adjusted by each fencer.

(a) *Representation of the Start position during Explosive lunge in the test environment.*

(b) *Representation of the en garde position during Explosive lunge in the test environment.*

(c) *Representation of the Final Lunge position during Explosive lunge in the test environment.*

Fig. 6.6 Cycle of movement for lunging during the Explosive lunge task: start (a), *en garde* (b), lunge (c).

Data Pre-processing

Collected data were imported to MATLAB R2020a[8]. The inertial system provided in output its orientation in time in the form of Euler angles referred to as the reference

[8]https://it.mathworks.com/products/new_products/release2020a.html

system of the sensor itself. These data made it possible to calculate the pelvis joint kinematics, also expressed as Euler angles, referred to as the pelvis reference system. The latter is obtained by making an anticlockwise rotation of the sensor reference system's 180°. EMG signals were processed with the following protocol: Butterworth bandpass filter with cut-off frequencies of 10 Hz and 450 Hz and Butterworth low pass filter with a cut-off frequency of 6 Hz. A threshold algorithm was applied to detect the start and the end of each lunge task. First, the Euclidean acceleration norm (Eq. 6.1) was calculated as follows:

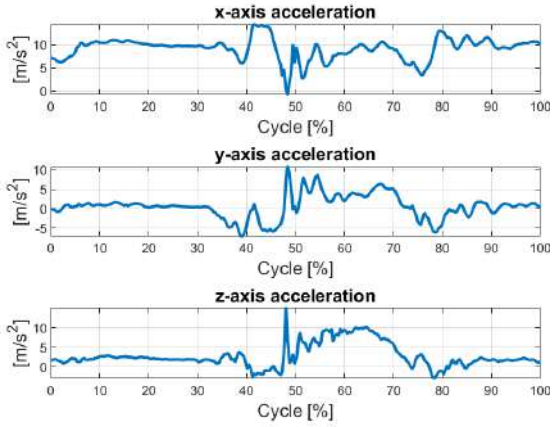
$$ACC = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \tag{6.1}$$

where ACC_x, ACC_y, and ACC_z are, respectively, the acceleration components expressed in the sensor reference system along the x, y, and z axes.

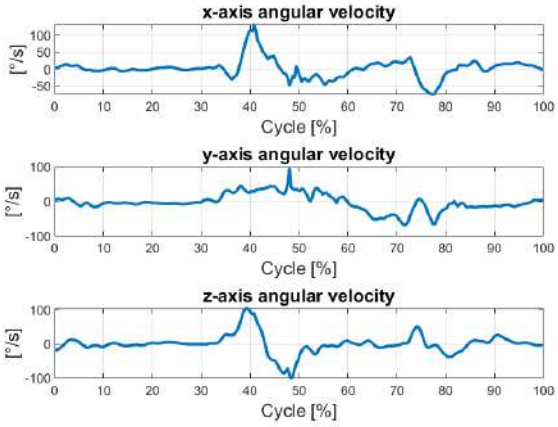
The subject was at rest in the first 3 seconds of acquisition due to the necessary sensor stabilization phase. Therefore, this time window (Eq. 6.2) was used to calculate the mean value and the standard deviation to obtain the threshold T value as follows:

$$T = \mu + J \times \sigma \tag{6.2}$$

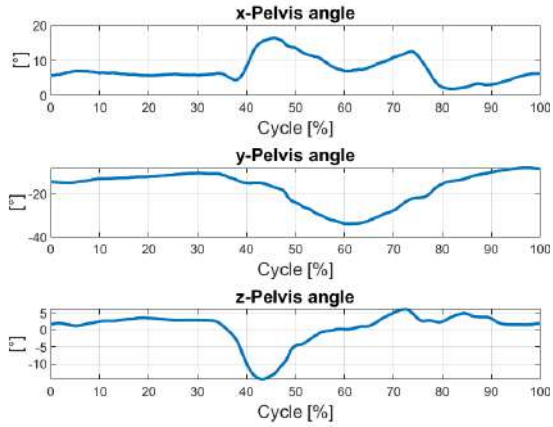
where μ and σ are the mean and standard deviation of the acceleration norm during a period of inactivity (time window), and J represents a numerical constant that takes on varying values depending on the case. In this study, $J = 3$. The signal was considered a "lunge movement" if its over-threshold duration was greater than 70 ms. The end of the lunge coincided with the index of the last over-threshold sample, 0.2 seconds before the first activation was considered as the beginning of the movement for the first task, 0.3 seconds for the other two tasks. Therefore, both kinematic data and EMG signals were aligned in the duration of the lunge cycle. The EMG signals were normalized using the maximum peak of the EMG envelope within the movement cycle to allow multiple subjects comparison. For each biomechanical data, a sampling of 500 samples was performed to standardize the number of samples of each signal, as this depends on how fast each subject was in performing the task. These biomechanical features were used to create a dataset consisting of 21 subjects and 13 features: the four EMG signals, the three components of accelerations, the three components of angular velocity, and the three components of pelvic angles, each expressed as a vector of 500 samples.



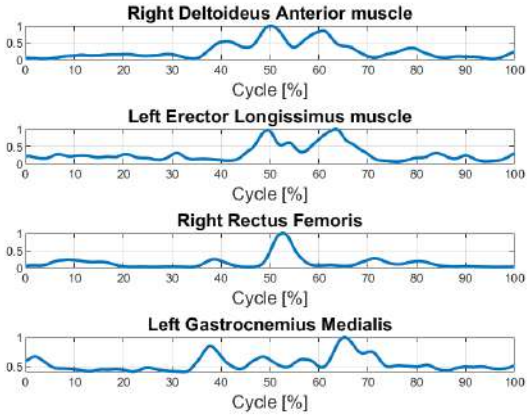
(a) *Accelerations signals along IMU reference system axes.*



(b) *Angular velocity signals along IMU reference system axes.*



(c) *Pelvis angle signals along Pelvis reference system axes.*



(d) *Muscles envelope signals of the four target muscles.*

Fig. 6.7 Feature dataset signals during Explosive lunge cycle: accelerations (a), angular velocities (b), pelvis angles (c), muscle envelopes (d).

In Figure 6.7 biomechanical data of an élite fencer during Explosive lunge are shown, in particular in Figure 6.7a accelerations on x, y and z-axis are shown, Figure 6.7b shows angular velocities on x, y and z-axis, Figure 6.7c shows pelvis angles on x, y and z-axis and in Figure 6.7d target muscles envelopes are shown.

Data Splitting and Dimensionality Reduction

The whole dataset consisted of 6500 columns, given by 13 features times 500 samples, for 126 records, given by 21 athletes executing two times the three exercises. Then, the dataset was split into training and test sets. The split percentages were 80% for the training set and 20% for the test set. The split has been developed by exploiting

the `train_test_split` function from `Scikit-learn` v1.0.2 library [219] with Python 3.9.0.

Since the dataset is affected by the curse of dimensionality, we applied a Principal Component Analysis (PCA) [86] to reduce the number of samples incrementally for each feature. Hence, we first made a train/test split, then fit the PCA with the training set and transformed both the training set and the test set with the fitted PCA. It is worth noting that, for each feature, every group of 500 samples has been individually taken and reduced with PCA as follows:

- $k = 50$, for an overall of 650 total features;
- $k = 25$, for an overall of 325 total features;
- $k = 10$, for an overall of 130 total features;
- $k = 5$, for an overall of 65 total features.

The reduction has been developed by exploiting the `Scikit-learn` v1.0.2 library, feeding in input `n_components` $= k$. Therefore, four training and test sets have been produced.

Machine Learning Algorithms

In order to determine the best classifier to predict the athlete class (Novice or élite), we analyzed the following models:

- **XGBoost Classifier** [52].
- **Multilayer Perceptron (MLP)** [121].
- **Random Forest (RF) Classifier** [36].
- **Support Vector Machine (SVM) Classifier** [66].

All four algorithms considered were supervised learning models. Afterward, a code was developed, and due to the interaction `Python`, it recalled the functions of the Machine Learning contained in the `scikit-learn` v1.0.2 library. This library was used to compare all models to identify the best suited to classify the élite and novice athletes and the model minimizing false-negative predictive values.

6.4.2 Evaluation and Model Performance for Fencer Classification

This section presents an in-depth analysis of the model performance after hyperparameter tuning and optimization. The primary goal was to identify the best performing model with an optimal balance between feature number and accuracy.

Hyperparameter Tuning of Best Model

The optimal model, identified as the MLP with $k = 50$, was fine-tuned using the GridSearchCV function in Python, utilizing a 5-fold cross-validation strategy to enhance its performance. GridSearchCV from the scikit-learn library was employed to optimize key parameters. The best hyperparameters obtained were:

- hidden_layer_sizes: (586,),
- activation: relu,
- solver: lbfgs,
- alpha: 0.1,
- learning_rate: constant.

Evaluation Setup and Experimental Protocol

The steps involved in the experiment setup and model evaluation included:

- Acquisition and preprocessing of raw data,
- Creation of a structured dataset for model training,
- Application of PCA for feature dimensionality reduction, with k selected based on the model's performance,
- Training the ML algorithm on the optimized dataset.

An 80/20 dataset split was used, with the training set subjected to 5-fold cross-validation to ensure robust model evaluation and to reduce sensitivity to data partitioning. Each fold's performance was independently evaluated, and the best hyperparameters for each classifier were chosen based on the cross-validation results.

Performance Metrics of the Final Model

The MLP model achieved a training accuracy of 100% and a validation accuracy of 96.0%. Additionally, the model's diagnostic capability was assessed using the Area Under the Curve (AUC) metric, shown in Figure 6.8, which reflects high discriminative power for classifying fencer expertise level. The confusion matrix in Figure 6.9 provides further insights into model reliability, highlighting a minimal misclassification rate between novice and élite fencers.

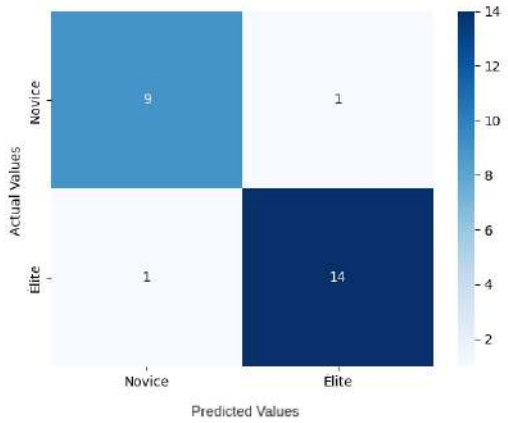


Fig. 6.8 ROC curves with the AUC value: 0 = novice, 1 = élite.

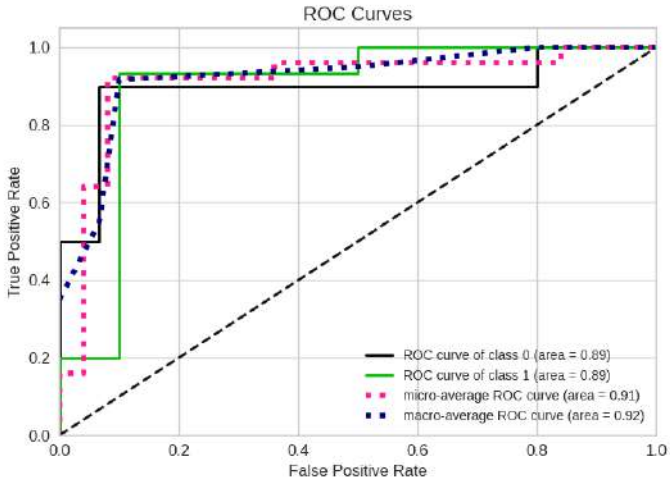


Fig. 6.9 Confusion matrix for MLP model during testing phase.

Overall Model Metrics

Table 6.4 provides the final precision, recall, and F1-score for both novice and élite categories during the test phase, demonstrating the model's high accuracy and balanced performance across classes.

Table 6.4 Performance metrics for MLP classifier.

Category	**Precision**	**Recall**	**F1-Score**
Novice (0)	90%	90%	90%
élite (1)	93%	93%	93%

In summary, the MLP model with $k = 50$ PCA components demonstrated excellent predictive capability, with accuracy, recall, and AUC values suggesting robust performance without overfitting. This high performance supports the model's suitability for reliably distinguishing between novice and élite fencers, offering a valuable tool for enhancing training strategies based on performance level assessment.

6.4.3 Key Findings and Insights

Effective analysis of sports movements requires a comprehensive understanding of their purpose, mechanics, and the factors influencing performance. In this study, we employed ML techniques to enhance fencing training by classifying fencers based on their skill level (élite vs. novice). Our approach utilized various data inputs, including acceleration, angular velocities, pelvis angles, and four surface sEMG signals. We compared multiple ML algorithms and identified the MLP model as the most suitable for this application. Through hyperparameter tuning, we optimized the MLP model, which was trained on a dataset comprising 650 features. The MLP demonstrated high classification accuracy, with a slight advantage in identifying élite fencers, a feature that can greatly aid in designing training regimens focused on improving critical movements, such as the lunge.

Structured and precise training plans are crucial for performance enhancement in all sports, and biomechanical analysis is especially significant in fencing to refine primary movements and prevent injuries [288]. While previous research has utilized ML algorithms to classify specific fencing tasks, such as footwork [187] or lunge phases [186], often reducing sensor requirements, our method advances these efforts. By integrating multiple biosignals, we provide a comprehensive framework that leverages wearable technology to capture and process data, delivering real-time feedback to athletes and coaches. This approach enhances training strategies and offers a holistic view of performance metrics.

Our methodology provides several key advantages: (i) it uncovers hidden patterns within training data, (ii) it efficiently manages large and complex datasets, and (iii) it is well-suited for integration into platforms that collect biomechanical data and supply instant feedback. Given the complexity and nonlinearity of human movement, predictive modeling, classification, and dimensionality reduction are crucial for extracting actionable insights [306]. The trend towards classifying athletic tasks using motion data from inertial sensors reflects the increasing demand for such advanced analytical tools.

A recent review by Hammes et al. [113] on artificial intelligence in élite sports emphasizes the growing emphasis on signal and image processing, as well as the challenges in sports analytics, including data acquisition, model explainability, and the practical controllability of AI. Addressing these challenges through AI-driven kinematic analysis can lead to standardized and optimized training protocols, enabling athletes and coaches to balance risks and opportunities effectively during training and competition.

6.5 Summary

To conclude, this chapter has presented two pivotal studies exploring the application of Machine Learning (ML) techniques in analyzing bioelectrical signals, specifically ECG and EMG, with substantial implications for clinical and sports applications. These works demonstrate the transformative potential of AI in harnessing bioelectrical signals to enhance medical diagnostics and improve athletic performance, underscoring how data-driven insights can be integrated into practice to benefit both fields.

The first study focused on ML applications in ECG analysis for early detection of cardiovascular abnormalities, particularly HF. By testing multiple classifiers, we identified LGBM as the most accurate model for distinguishing NSR from ARR, achieving high precision and recall. Beyond classification accuracy, XAI techniques provided interpretability to model predictions, making them more applicable and reliable for medical practitioners. These findings highlight the feasibility of AI-driven ECG analysis as a complementary tool in cardiovascular diagnostics, aiding in early detection, reducing diagnostic time, and supporting more targeted patient care.

The second study extended the analysis to EMG signals to differentiate skill levels among fencers by capturing neuromuscular responses during specific movements. Through rigorous testing of ML algorithms, the Multilayer Perceptron (MLP) model emerged as the best approach, demonstrating high accuracy in classifying élite and

novice fencers. Importantly, this work reveals the potential for EMG-based analysis to extend beyond sports training, providing a framework for rehabilitation applications. The same wearable sensors and data monitoring used to assess fencing movements, such as the lunge, could be repurposed in clinical settings as a rehabilitative tool. This approach could support young and elderly individuals alike in neuromuscular training, strength recovery, and movement precision—areas critical to injury prevention and rehabilitation.

In summary, these studies collectively validate the utility of ML in processing and interpreting bioelectrical signals, showing extensive applicability in healthcare and sports science. In cardiology, ECG-based models can support early-stage CVD detection and continuous patient monitoring, promoting preventative strategies and enhancing patient outcomes. Similarly, EMG-based analysis, initially applied to athletic training, could become a valuable component of personalized rehabilitation programs for various age groups, enabling clinicians to monitor neuromuscular function closely and optimize therapeutic exercises.

Overall, this chapter underscores that AI-driven bioelectrical signal analysis offers significant potential to bridge the gap between raw data and actionable insights for practitioners. These methods represent a forward-looking approach in clinical and sports domains, where improved diagnostics, performance assessments, and rehabilitative planning can be achieved through data-informed decision-making, setting a new standard in healthcare and athletic training.

Chapter 7

Conclusion

To conclude this doctoral thesis, we have thoroughly examined the transformative impact of AI on healthcare, specifically through its application in both intelligent diagnostic systems and the analysis of bioelectrical signals. The research was organized around two core themes: leveraging AI for improving diagnostic accuracy and developing innovative methods for analyzing bioelectrical signals, such as EEG, EMG, and ECG, to support personalized healthcare.

In Chapter 2 and Chapter 3, we established the foundational knowledge of intelligent diagnostics and bioelectrical signals. These chapters provided essential definitions, background concepts, and technological advancements in AI-based healthcare solutions, setting the stage for the subsequent research.

Chapter 4 explored the application of AI-based techniques for healthcare, particularly focusing on dietary and nutritional pattern analysis for predicting health outcomes. Using machine learning methodologies, we demonstrated how AI could enhance the prediction of health conditions like diabetes, frailty, and cognitive disorders, thereby offering more personalized healthcare interventions. The integration of XAI was crucial in ensuring transparency, which is essential for the adoption of these technologies by healthcare providers and patients alike.

Chapter 5 concentrated on enhancing devices for analyzing electroencephalographic (EEG) signals through AI. We developed systems for emotion recognition using deep learning and BCIs, providing real-time emotional assessments that can be applied in therapeutic contexts, such as mental health support. The chapter also covered the development of neurofeedback tools that help individuals monitor and adjust their emotional states. Additionally, the NeuroSense dataset was introduced to explore emotion recognition using EEG signals recorded from sparse, low-cost devices. The versatility of these approaches was further illustrated through the NeuralPMG system,

a framework for emotion-driven polyphonic music generation, demonstrating potential uses in both rehabilitation and creative applications.

Chapter 7 extended the application of AI in bioelectrical signals to practical performance and clinical contexts, including AI-based systems for analyzing cardiac signals using ECG data. The chapter also discussed the use of EMG signals for classifying expertise levels in fencing athletes, with potential applications in rehabilitation settings for both young and elderly individuals. These applications showcased the potential of combining machine learning algorithms and wearable sensor data to enhance training, rehabilitation, and sports performance.

The findings and methodologies presented in this thesis underscore the significant role of AI in revolutionizing healthcare. AI-assisted diagnostics have shown promising results in predicting and managing various health conditions, thereby contributing to personalized medicine and proactive healthcare. Similarly, AI-enhanced bioelectrical signal analysis provides deeper insights into physiological and emotional states, offering a range of applications from rehabilitation to emotional support systems.

The practical implications of this research are substantial. The methods and systems developed throughout this thesis bridge the gap between cutting-edge AI technology and clinical applications, with a particular emphasis on interpretability and real-world usability. By focusing on XAI principles, we aimed to make these advanced AI systems accessible and trustworthy for healthcare professionals and patients.

Looking ahead, this research lays a foundation for future advancements in AI-driven personalized medicine, mental health support, and rehabilitation. The integration of AI with bioelectrical signals offers a wide range of potential applications across healthcare, sports, and creative fields, ultimately aiming to improve patient outcomes and overall quality of life.

Key Contributions

Throughout this thesis, several key contributions were made to the field of AI in healthcare, particularly in the areas of bioelectrical signal analysis and intelligent diagnostics. These contributions are summarized as follows:

- **Advancements in AI for Medical Diagnostics:** Developed novel machine learning and deep learning models specifically designed for healthcare applications, including predictive models for chronic disease and personalized treatment recommendations, thereby enhancing clinical decision support systems in terms of accuracy and efficiency.

- **Integration of Explainable AI (XAI):** Applied XAI techniques to ensure transparency and interpretability in diagnostic models. This contribution enables healthcare professionals to gain insights into AI-based predictions, fostering trust and supporting critical decision-making processes.
- **Bioelectrical Signal Analysis:** Explored and implemented AI-driven methods for the analysis of bioelectrical signals, including EEG, ECG, and EMG, with applications spanning BCIs for emotion recognition, neurofeedback for mental health, and arrhythmia detection for cardiac care.
- **Applications in Public Health and Personalized Medicine:** Leveraged large-scale epidemiological data to propose data-driven strategies for chronic disease management and public health interventions, aligned with the goals of personalized medicine.
- **Applications Beyond Clinical Settings:** Demonstrated the potential of AI in fields beyond clinical healthcare, such as sports science and rehabilitation. This includes analyzing bioelectrical and biomechanical data for athlete performance assessment, injury prevention, and developing assistive technologies for rehabilitation in both young and elderly populations.

Future Directions

This thesis represents the beginning of a longer and increasingly interesting research path. Despite the numerous analyses and proposals presented, other research questions and open challenges naturally arose. Therefore, we have intentionally decided not to report the possible future work of the thesis in this chapter. Instead, several initial research directions and ideas have been formalized and tested over the last few months before the submission of the thesis. Their preliminary outcomes will be presented in the next (and last) chapter, following the same thematic structure provided in this chapter. Although it is probable that the majority of these ideas may not be successful upon further investigation, even a small percentage that yields valuable insights will be regarded as a significant achievement.

In conclusion, the contributions made in this thesis have laid a solid foundation for future research in the application of AI to healthcare. The methodologies and frameworks developed have the potential to significantly advance the field, offering new tools and strategies for improving healthcare outcomes. As we move forward, the

integration of AI with bioelectrical signal analysis and intelligent diagnostics promises to open new frontiers in personalized medicine, ultimately enhancing patient care and well-being.

Bibliography

[1] Ali Ahmadvand, Harshita Sahijwani, and Eugene Agichtein. "Would you like to talk about sports now? towards contextual topic suggestion for open-domain conversational agents." In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 2020, pp. 83–92.

[2] Zeeshan Ahmed, Khalid Mohamed, Saman Zeeshan, and XinQi Dong. "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine." In: *Database* 2020 (2020), baaa010.

[3] Mohammad Nazmul Alam, Mandeep Kaur, and Md Shahin Kabir. "Explainable AI in Healthcare: Enhancing transparency and trust upon legal and ethical consideration." In: *Int Res J Eng Technol* 10.6 (2023), pp. 1–9.

[4] Ahmed Shihab Albahri, Ali M Duhaim, Mohammed A Fadhel, Alhamzah Alnoor, Noor S Baqer, Laith Alzubaidi, Osamah Shihab Albahri, Abdullah Hussein Alamoodi, Jinshuai Bai, Asma Salhi, et al. "A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion." In: *Information Fusion* 96 (2023), pp. 156–191.

[5] Abdullah Alharbi, Wael Alosaimi, Radhya Sahal, and Hager Saleh. "Real-Time System Prediction for Heart Rate Using Deep Learning and Stream Processing Platforms." In: *Complexity* 2021.1 (2021), p. 5535734.

[6] 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.

[7] Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M.W. 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. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104672.

[8] Omar Ali, Anup Shrestha, Jeffrey Soar, and Samuel Fosso Wamba. "Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review." In: *International Journal of Information Management* 43 (2018), pp. 146–158.

[9] Ashik Mostafa Alvi, Siuly Siuly, Hua Wang, Kate Wang, and Frank Whittaker. "A deep learning based framework for diagnosis of mild cognitive impairment." In: *Knowledge-Based Systems* 248 (2022), p. 108815.

[10] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, et al. "Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review." In: *Medical Image Analysis* 67 (2021), p. 101848.

[11] Carla Aoun, Reine Bou Daher, Nada El Osta, Tatiana Papazian, and Lydia Rabbaa Khabbaz. "Reproducibility and relative validity of a food frequency questionnaire to assess dietary intake of adults living in a Mediterranean country." In: *PloS one* 14.6 (2019), e0218541.

[12] 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.

[13] 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). **To appear**.

[14] Miguel Areia, Yuichi Mori, Loredana Correale, Alessandro Repici, Michael Bretthauer, Prateek Sharma, Filipe Taveira, Marco Spadaccini, Giulio Antonelli, Alanna Ebigbo, et al. "Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study." In: *The Lancet Digital Health* 4.6 (2022), e436–e444.

[15] Simona Aresta, Ilaria Bortone, Francesco Bottiglione, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Mariapia Musci, Fedelucio Narducci, Andrea Pazienza, Rodolfo Sardone, et al. "Combining Biomechanical Features and Machine Learning Approaches to Identify Fencers' Levels for Training Support." In: *Applied Sciences* 12.23 (2022), p. 12350.

[16] Sohaib Asif, Yi Wenhui, Saif- ur-Rehman, Qurrat- ul-ain, Kamran Amjad, Yi Yueyang, Si Jinhai, and Muhammad Awais. "Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision." In: *Archives of Computational Methods in Engineering* (2024), pp. 1–31.

[17] 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.

[18] Adolfo Francesco Attili, N Carulli, E Roda, B Barbara, Livio Capocaccia, A Menotti, L Okoliksanyi, G Ricci, R Capocaccia, D Festi, et al. "Epidemiology of gallstone disease in Italy: prevalence data of the Multicenter Italian Study on Cholelithiasis (MI COL.)" In: *American journal of epidemiology* 141.2 (1995), pp. 158–165.

[19] Ulas Baran Baloglu, Muhammed Talo, Ozal Yildirim, Ru San Tan, and U Rajendra Acharya. "Classification of myocardial infarction with multi-lead ECG signals and deep CNN." In: *Pattern recognition letters* 122 (2019), pp. 23–30.

[20] Alexandre Barachant, Dano Morrison, Hubert Banville, Jason Kowaleski, Uri Shaked, Sylvain Chevallier, and Juan Jesús Torre Tresols. "muse-lsl." In: (May 2019). DOI: 10.5281/zenodo.3228861.

[21] N Barascud. *meegkit: EEG and Meg denoising in Python*.

[22] William G Baxt. "Application of artificial neural networks to clinical medicine." In: *The lancet* 346.8983 (1995), pp. 1135–1138.

[23] Luzheng Bi, Xin-An Fan, and Yili Liu. "EEG-based brain-controlled mobile robots: a survey." In: *IEEE transactions on human-machine systems* 43.2 (2013), pp. 161–176.

[24] Giovanni Maria Biancofiore, Tommaso Di Noia, Eugenio Di Sciascio, Fedelucio Narducci, and Paolo Pastore. "Aspect based sentiment analysis in music: a case study with spotify." In: *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. 2022, pp. 696–703.

[25] Gérard Biau and Erwan Scornet. "A random forest guided tour." In: *Test* 25 (2016), pp. 197–227.

[26] CD Binnie and PF Prior. "Electroencephalography." In: *Journal of Neurology, Neurosurgery & Psychiatry* 57.11 (1994), pp. 1308–1319.

[27] Jordan J. Bird, Luis J. Manso, Eduardo P. Ribeiro, Anikó Ekárt, and Diego R. Faria. "A Study on Mental State Classification using EEG-based Brain-Machine Interface." In: *2018 International Conference on Intelligent Systems (IS)*. 2018, pp. 795–800. DOI: 10.1109/IS.2018.8710576.

[28] Ashis Kumer Biswas, Nasimul Noman, and Abdur Rahman Sikder. "Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information." In: *BMC bioinformatics* 11.1 (2010), pp. 1–17.

[29] Ilaria Bortone, Lorenzo Moretti, Davide Bizzoca, Nuccio Caringella, Michelangelo Delmedico, Andrea Piazzolla, and Biagio Moretti. "The importance of biomechanical assessment after Return to Play in athletes with ACL-Reconstruction." In: *Gait & Posture* 88 (2021), pp. 240–246.

[30] Lindsay Bottoms, Andrew Greenhalgh, and Jonathan Sinclair. "Kinematic determinants of weapon velocity during the fencing lunge in experienced épée fencers." In: *Acta of bioengineering and biomechanics* 15.4 (2013).

[31] Ghaith Bouallegue, Ridha Djemal, Saleh A Alshebeili, and Hesham Aldhalaan. "A dynamic filtering DF-RNN deep-learning-based approach for EEG-based neurological disorders diagnosis." In: *IEEE Access* 8 (2020), pp. 206992–207007.

[32] Samar Bouazizi and Hela Ltifi. "Enhancing accuracy and interpretability in EEG-based medical decision making using an explainable ensemble learning framework application for stroke prediction." In: *Decision Support Systems* 178 (2024), p. 114126.

[33] Margaret M Bradley, Maurizio Codispoti, Bruce N Cuthbert, and Peter J Lang. "Emotion and motivation I: defensive and appetitive reactions in picture processing." In: *Emotion* 1.3 (2001), p. 276.

[34] 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.

[35] 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.

[36] L. Breiman. "Random forests." In: *Machine learning* 45.1 (2001), pp. 5–32.

[37] Joseph D Bronzino and Donald R Peterson. "Principles of electroencephalography." In: *Biomedical Engineering Fundamentals*. CRC press, 2006, pp. 445–456.

[38] Andrew Brouse. *The Interharmonium: an investigation into networked musical applications and brainwaves. Master of Arts thesis*. McGill University, 2001.

[39] Quinlan D Buchlak, Nazanin Esmaili, Jean-Christophe Leveque, Farrokh Farrokhi, Christine Bennett, Massimo Piccardi, and Rajiv K Sethi. "Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review." In: *Neurosurgical review* 43 (2020), pp. 1235–1253.

[40] 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.

[41] Valentina Camomilla, Elena Bergamini, Silvia Fantozzi, and Giuseppe Vannozzi. "Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review." In: *Sensors* 18.3 (2018), p. 873.

[42] Henry Candra, Mitchell Yuwono, Ardi Handojoseno, Rifai Chai, Steven Su, and Hung T Nguyen. "Recognizing emotions from EEG subbands using wavelet analysis." In: *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2015, pp. 6030–6033.

[43] 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.

[44] 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)*. IEEE. 2022, pp. 3228–3233.

[45] 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.

[46] Fabio Castellana, Roberta Zupo, Ilaria Bortone, Gianluigi Giannelli, Rossella Donghia, Luisa Lampignano, Chiara Griseta, Giovanni De Pergola, Heiner Boeing, Anna Maria Cisternino, et al. "Traditional Old Dietary Pattern of Castellana Grotte (Apulia) Is Associated with Healthy Outcomes." In: *Nutrients* 12.10 (2020), p. 3097.

[47] Fabio Castellana, Roberta Zupo, Ilaria Bortone, Gianluigi Giannelli, Rossella Donghia, Luisa Lampignano, Chiara Griseta, Giovanni De Pergola, Heiner Boeing, Anna Maria Cisternino, et al. "Traditional old dietary pattern of castellana grotte (apulia) is associated with healthy outcomes." In: *Nutrients* 12.10 (2020), p. 3097.

[48] Manisha Chandalia, Abhimanyu Garg, Dieter Lutjohann, Klaus Von Bergmann, Scott M Grundy, and Linda J Brinkley. "Beneficial effects of high dietary fiber intake in patients with type 2 diabetes mellitus." In: *New England Journal of Medicine* 342.19 (2000), pp. 1392–1398.

[49] Bahzad Charbuty and Adnan Abdulazeez. "Classification based on decision tree algorithm for machine learning." In: *Journal of Applied Science and Technology Trends* 2.01 (2021), pp. 20–28.

[50] A. Chawla, K.D. Nguyen, and Y.P. Sharon Goh. "Macrophage-Mediated Inflammation in Metabolic Disease." In: *Nat. Rev. Immunol.* 11 (2011), pp. 738–749. DOI: 10.1038/nri3071.

[51] Jonathan H Chen, Muthuraman Alagappan, Mary K Goldstein, Steven M Asch, and Russ B Altman. "Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets." In: *International journal of medical informatics* 102 (2017), pp. 71–79.

[52] 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.

[53] 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.

[54] Tony Lin-Wei Chen, Duo Wai-Chi Wong, Yan Wang, Sicong Ren, Fei Yan, and Ming Zhang. "Biomechanics of fencing sport: A scoping review." In: *PloS one* 12.2 (2017), e0171578.

[55] Weisi Chen, Zoran Milosevic, Fethi A Rabhi, and Andrew Berry. "Real-time analytics: Concepts, architectures and ML/AI considerations." In: *IEEE Access* (2023).

[56] Jinyong Cheng, Qingxu Zou, and Yunxiang Zhao. "ECG signal classification based on deep CNN and BiLSTM." In: *BMC medical informatics and decision making* 21 (2021), pp. 1–12.

[57] Erin Cherry and Celine Latulipe. "Quantifying the creativity support of digital tools through the creativity support index." In: *ACM Transactions on Computer-Human Interaction* 21.4 (2014), pp. 1–25.

[58] Francois Chollet et al. *Keras*. 2015. URL: https://github.com/fchollet/keras.

[59] Jan Pieter Clarys and Jan Cabri. "Electromyography and the study of sports movements: a review." In: *Journal of sports sciences* 11.5 (1993), pp. 379–448.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] C. Cortes and V. Vapnik. "Support-vector networks." In: *Machine Learning* 20.3 (1995), pp. 273–297.

[67] 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.

[68] Annunziata D'Alessandro, Luisa Lampignano, and Giovanni De Pergola. "Mediterranean diet pyramid: a proposal for Italian people. A systematic review of prospective studies to derive serving sizes." In: *Nutrients* 11.6 (2019), p. 1296.

[69] Danilo Danese, Tommaso Di Noia, Angela Lombardi, Domenico Lofù, Fatemeh Nazary, Rodolfo Sardone, and Paolo Sorino. "Integrating eXplainable AI in Healthcare: A Web Application Framework for Advancing the One Health Paradigm." 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). **To appear**.

[70] Yukun Dang, Zitong Liu, Xixin Yang, Linqiang Ge, and Sheng Miao. "A fatigue assessment method based on attention mechanism and surface electromyography." In: *Internet of Things and Cyber-Physical Systems* 3 (2023), pp. 112–120.

[71] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. "Big data in healthcare: management, analysis and future prospects." In: *Journal of big data* 6.1 (2019), pp. 1–25.

[72] 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).

[73] Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. "Minirocket: A very fast (almost) deterministic transform for time series classification." In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2021, pp. 248–257.

[74] Jiawen Deng and Fuji Ren. "A survey of textual emotion recognition and its challenges." In: *IEEE Transactions on Affective Computing* 14.1 (2021), pp. 49–67.

[75] Abhinav Dhall. "Emotiw 2019: Automatic emotion, engagement and cohesion prediction tasks." In: *2019 International Conference on Multimodal Interaction*. 2019, pp. 546–550.

[76] Gianluca Di Flumeri, Pietro Aricò, Gianluca Borghini, Alfredo Colosimo, and Fabio Babiloni. "A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel." In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 3187–3190.

[77] Suh-Yeon Dong, Bo-Kyeong Kim, Kyeongho Lee, and Soo-Young Lee. "A preliminary study on human trust measurements by EEG for human-machine interactions." In: *Proceedings of the 3rd International Conference on Human-Agent Interaction*. 2015, pp. 265–268.

[78] Xiaobing Du, Cuixia Ma, Guanhua Zhang, Jinyao Li, Yu-Kun Lai, Guozhen Zhao, Xiaoming Deng, Yong-Jin Liu, and Hongan Wang. "An efficient LSTM network for emotion recognition from multichannel EEG signals." In: *IEEE Transactions on Affective Computing* 13.3 (2020), pp. 1528–1540.

[79] Krishnamurthy Dvijotham, Jim Winkens, Melih Barsbey, Sumedh Ghaisas, Robert Stanforth, Nick Pawlowski, Patricia Strachan, Zahra Ahmed, Shekoofeh Azizi, Yoram Bachrach, et al. "Enhancing the reliability and accuracy of AI-enabled diagnosis via complementarity-driven deferral to clinicians." In: *Nature Medicine* 29.7 (2023), pp. 1814–1820.

[80] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. "Explainable AI (XAI): Core ideas, techniques, and solutions." In: *ACM Computing Surveys* 55.9 (2023), pp. 1–33.

[81] Sylwia Dziegielewska-Gesiak. "Metabolic syndrome in an aging society–role of oxidant-antioxidant imbalance and inflammation markers in disentangling atherosclerosis." In: *Clinical interventions in aging* (2021), pp. 1057–1070.

[82] Peter J Embi. "Algorithmovigilance—advancing methods to analyze and monitor artificial intelligence–driven health care for effectiveness and equity." In: *JAMA network open* 4.4 (2021), e214622–e214622.

[83] International Diabetes Federation. "IDF Diabetes Atlas." In: (2019).

[84] Marta Fernandes, Susana M Vieira, Francisca Leite, Carlos Palos, Stan Finkelstein, and João MC Sousa. "Clinical decision support systems for triage in the emergency department using intelligent systems: a review." In: *Artificial Intelligence in Medicine* 102 (2020), p. 101762.

[85] Mücahit Fındık, Şeyma Yılmaz, and Mehmet Koseoglu. "Random forest classification of finger movements using electromyogram (emg) signals." In: *2020 IEEE SENSORS*. IEEE. 2020, pp. 1–4.

[86] Imola K Fodor. *A survey of dimension reduction techniques*. Tech. rep. Lawrence Livermore National Lab., CA (US), 2002.

[87] Parisa Forouzannezhad, Alireza Abbaspour, Chunfei Li, Mercedes Cabrerizo, and Malek Adjouadi. "A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features." In: *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE. 2018, pp. 1341–1346.

[88] Amanda M Fretts, Jack L Follis, Jennifer A Nettleton, Rozenn N Lemaitre, Julius S Ngwa, Mary K Wojczynski, Ioanna Panagiota Kalafati, Tibor V Varga, Alexis C Frazier-Wood, Denise K Houston, et al. "Consumption of meat is associated with higher fasting glucose and insulin concentrations regardless of glucose and insulin genetic risk scores: a meta-analysis of 50,345 Caucasians." In: *The American journal of clinical nutrition* 102.5 (2015), pp. 1266–1278.

[89] William T Friedewald, Robert I Levy, and Donald S Fredrickson. "Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge." In: *Clinical chemistry* 18.6 (1972), pp. 499–502.

[90] Ryo Fujimori, Keibun Liu, Shoko Soeno, Hiromu Naraba, Kentaro Ogura, Konan Hara, Tomohiro Sonoo, Takayuki Ogura, Kensuke Nakamura, Tadahiro Goto, et al. "Acceptance, barriers, and facilitators to implementing artificial intelligence–based decision support systems in emergency departments: quantitative and qualitative evaluation." In: *JMIR formative research* 6.6 (2022), e36501.

[91] 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.

[92] Jianfeng Gao, Haoliang Qi, Xinsong Xia, and Jian-Yun Nie. "Linear discriminant model for information retrieval." In: *Proceedings of the International Conference on Research and Development in Information Retrieval (ACM SIGIR)*. 2005, pp. 290–297.

[93] Shuo Gao, Jing Gong, Bofei Chen, Bozhou Zhang, Fengrong Luo, Mustafa O Yerabakan, Yu Pan, and Boyi Hu. "Use of advanced materials and artificial intelligence in electromyography signal detection and interpretation." In: *Advanced Intelligent Systems* 4.10 (2022), p. 2200063.

[94] Zhongke Gao, Weidong Dang, Xinmin Wang, Xiaolin Hong, Linhua Hou, Kai Ma, and Matjaž Perc. "Complex networks and deep learning for EEG signal analysis." In: *Cognitive Neurodynamics* 15.3 (2021), pp. 369–388.

[95] David G Garson. "Interpreting neural network connection weights." In: (1991).

[96] Julie Gerlings, Millie Søndergaard Jensen, and Arisa Shollo. "Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare." In: *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects* (2022), pp. 169–198.

[97] Kosmas Glavas, Georgios Prapas, Katerina D. Tzimourta, Alexandros T. Tzallas, Nikolaos Giannakeas, and Markos G. Tsipouras. "Intra-User Analysis Based on Brain-Computer Interface Controlled Game." In: *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*. 2022, pp. 386–390. DOI: 10.1109/TSP55681.2022.9851336.

[98] Ercan Gokgoz and Abdulhamit Subasi. "Comparison of decision tree algorithms for EMG signal classification using DWT." In: *Biomedical signal processing and control* 18 (2015), pp. 138–144.

[99] Ary Goldberger et al. "Components of a new research resource for complex physiologic signals." In: *PhysioNet* 101 (Jan. 2000).

[100] Ary L Goldberger, Zachary D Goldberger, and Alexei Shvilkin. *Goldberger's Clinical Electrocardiography-E-Book: A Simplified Approach*. Elsevier Health Sciences, 2023.

[101] 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.

[102] Peiliang Gong, Pengpai Wang, Yueying Zhou, and Daoqiang Zhang. "A spiking neural network with adaptive graph convolution and lstm for eeg-based brain-computer interfaces." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), pp. 1440–1450.

[103] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[104] Susanna Yu Gordleeva, Sergey A Lobov, Nikita A Grigorev, Andrey O Savosenkov, Maxim O Shamshin, Maxim V Lukoyanov, Maxim A Khoruzhko, and Victor B Kazantsev. "Real-time EEG–EMG human–machine interface-based control system for a lower-limb exoskeleton." In: *Ieee Access* 8 (2020), pp. 84070–84081.

[105] 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.

[106] Juan M Górriz, Ignacio Álvarez-Illán, Agustín Álvarez-Marquina, Juan Eloy Arco, Martin Atzmueller, F Ballarini, Emilia Barakova, Guido Bologna, P Bonomini, Germán Castellanos-Dominguez, et al. "Computational approaches to explainable artificial intelligence: advances in theory, applications and trends." In: *Information Fusion* 100 (2023), p. 101945.

[107] Alexandre Gramfort et al. "MEG and EEG Data Analysis with MNE-Python." In: *Frontiers in Neuroscience* 7.267 (2013), pp. 1–13. DOI: 10.3389/fnins.2013.00267.

[108] Lena Griebel, Hans-Ulrich Prokosch, Felix Köpcke, Dennis Toddenroth, Jan Christoph, Ines Leb, Igor Engel, and Martin Sedlmayr. "A scoping review of cloud computing in healthcare." In: *BMC medical informatics and decision making* 15 (2015), pp. 1–16.

[109] Robert J Grissom and John J Kim. *Effect sizes for research: A broad practical approach.* Lawrence Erlbaum Associates Publishers, 2005.

[110] Sergio Grueso and Raquel Viejo-Sobera. "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review." In: *Alzheimer's research & therapy* 13 (2021), pp. 1–29.

[111] Andrzej Grzybowski, Piotr Brona, Gilbert Lim, Paisan Ruamviboonsuk, Gavin SW Tan, Michael Abramoff, and Daniel SW Ting. "Artificial intelligence for diabetic retinopathy screening: a review." In: *Eye* 34.3 (2020), pp. 451–460.

[112] Hongyan Gu, Jingbin Huang, Lauren Hung, and Xiang'Anthony' Chen. "Lessons learned from designing an AI-enabled diagnosis tool for pathologists." In: *Proceedings of the ACM on Human-computer Interaction* 5.CSCW1 (2021), pp. 1–25.

[113] Fabian Hammes, Alexander Hagg, Alexander Asteroth, and Daniel Link. "Artificial intelligence in elite sports—a narrative review of success stories and challenges." In: *Frontiers in sports and active living* 4 (2022).

[114] Alan Hanjalic and Li-Qun Xu. "Affective video content representation and modeling." In: *IEEE transactions on multimedia* 7.1 (2005), pp. 143–154.

[115] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In: *Advances in Psychology* 52 (1988), pp. 139–183. ISSN: 0166-4115.

[116] Rex Hartson and Pardha S Pyla. *The UX book: Agile UX design for a quality user experience.* Morgan Kaufmann, 2018.

[117] Marc Hassenzahl, Michael Burmester, and Franz Koller. "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität." In: *Mensch & Computer 2003: Interaktion in Bewegung* (2003), pp. 187–196.

[118] Ahmed Rachid Hazourli, Amine Djeghri, Hanan Salam, and Alice Othmani. "Multi-facial patches aggregation network for facial expression recognition and facial regions contributions to emotion display." In: *Multimedia Tools and Applications* 80.9 (2021), pp. 13639–13662.

[119] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[120] Hadeer A Helaly, Mahmoud Badawy, and Amira Y Haikal. "Deep learning approach for early detection of Alzheimer's disease." In: *Cognitive computation* 14.5 (2022), pp. 1711–1727.

[121] G. E. Hinton. "Connectionist learning procedures." In: *Machine Learning, Volume III*. Elsevier, 1990, pp. 555–610.

[122] Richard 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.

[123] Md Imran Hossain, Ghada Zamzmi, Peter R Mouton, Md Sirajus Salekin, Yu Sun, and Dmitry Goldgof. "Explainable AI for Medical Data: Current Methods, Limitations, and Future Directions." In: *ACM Computing Surveys* (2023).

[124] Chin-Chuan Hsu, Yuan Kao, Chien-Chin Hsu, Chia-Jung Chen, Shu-Lien Hsu, Tzu-Lan Liu, Hung-Jung Lin, Jhi-Joung Wang, Chung-Feng Liu, and Chien-Cheng Huang. "Using artificial intelligence to predict adverse outcomes in emergency department patients with hyperglycemic crises in real time." In: *BMC Endocrine Disorders* 23.1 (2023), p. 234.

[125] 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.

[126] Shigao Huang, Jie Yang, Simon Fong, and Qi Zhao. "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges." In: *Cancer letters* 471 (2020), pp. 61–71.

[127] Lucas Hulen. "A musical scale in simple ratios of the harmonic series converted to cents of twelve-tone equal temperament for digital synthesis." In: *WSEAS Transactions on Computers* 5.8 (2006), pp. 1713–1719.

[128] Cosimo Ieracitano, Nadia Mammone, Amir Hussain, and Francesco Carlo Morabito. "A novel explainable machine learning approach for EEG-based brain-computer interface systems." In: *Neural Computing and Applications* 34.14 (2022), pp. 11347–11360.

[129] F Imamura, S Brage, SJ Griffin, NJ Wareham, N Forouhi, et al. "Intakes and sources of dietary sugars and their association with metabolic and inflammatory markers." In: (2018).

[130] 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.

[131] Istat. *Annuario Statistico Italiano*. Rome, Italy: Istat, 2017. ISBN: 9788845819650.

[132] Istat, Italia, and Direzione generale della statistica. "Annuario statistico italiano." In: (1918).

[133] Mahboobeh Jafari, Delaram Sadeghi, Afshin Shoeibi, Hamid Alinejad-Rokny, Amin Beheshti, David López García, Zhaolin Chen, U Rajendra Acharya, and Juan M Gorriz. "Empowering precision medicine: AI-driven schizophrenia diagnosis via EEG signals: A comprehensive review from 2002–2023." In: *Applied Intelligence* 54.1 (2024), pp. 35–79.

[134] Vicneswary Jahmunah, Eddie YK Ng, Ru-San Tan, Shu Lih Oh, and U Rajendra Acharya. "Explainable detection of myocardial infarction using deep learning models with Grad-CAM technique on ECG signals." In: *Computers in Biology and Medicine* 146 (2022), p. 105550.

[135] Ankush D Jamthikar, Deep Gupta, Luca Saba, Narendra N Khanna, Klaudija Viskovic, Sophie Mavrogeni, John R Laird, Naveed Sattar, Amer M Johri, Gyan Pareek, et al. "Artificial intelligence framework for predictive cardiovascular and stroke risk assessment models: A narrative review of integrated approaches using carotid ultrasound." In: *Computers in biology and medicine* 126 (2020), p. 104043.

[136] Kashif Javed, Haroon A Babri, and Mehreen Saeed. "Feature selection based on class-dependent densities for high-dimensional binary data." In: *IEEE Transactions on Knowledge and Data Engineering* 24.3 (2010), pp. 465–477.

[137] Xiaofeng Jin, Andrew Cai, Tailin Xu, and Xueji Zhang. "Artificial intelligence biosensors for continuous glucose monitoring." In: *Interdisciplinary Materials* 2.2 (2023), pp. 290–307.

[138] Owen Jorgensen. *Tuning: containing the perfection of eighteenth-century temperament, the lost art of nineteenth-century temperament, and the science of equal temperament, complete with instructions for aural and electronic tuning*. Vol. 4. Michigan State University Press East Lansing, MI, 1991.

[139] Vaishali M Joshi and Rajesh B Ghongade. "IDEA: Intellect database for emotion analysis using EEG signal." In: *Journal of King Saud University-Computer and Information Sciences* (2020).

[140] Kaumudi J Joshipura, Frank B Hu, JoAnn E Manson, Meir J Stampfer, Eric B Rimm, Frank E Speizer, Graham Colditz, Alberto Ascherio, Bernard Rosner, Donna Spiegelman, et al. "The effect of fruit and vegetable intake on risk for coronary heart disease." In: *Annals of internal medicine* 134.12 (2001), pp. 1106–1114.

[141] Tamás Karácsony, John Paulin Hansen, Helle Klingenberg Iversen, and Sadasivan Puthusserypady. "Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback." In: *Proceedings of the 10th Augmented Human International Conference 2019*. 2019, pp. 1–8.

[142] Kalliopi Karatzi, George Moschonis, Afroditi-Alexandra Barouti, Christos Lionis, George P Chrousos, and Yannis Manios. "Dietary patterns and breakfast consumption in relation to insulin resistance in children. The Healthy Growth Study." In: *Public health nutrition* 17.12 (2014), pp. 2790–2797.

[143] Eva Kassi, Panagiota Pervanidou, Gregory Kaltsas, and George Chrousos. "Metabolic syndrome: definitions and controversies." In: *BMC medicine* 9 (2011), pp. 1–13.

[144] Stamos Katsigiannis and Naeem Ramzan. "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices." In: *IEEE journal of biomedical and health informatics* 22.1 (2017), pp. 98–107.

[145] Simarjeet Kaur, Jimmy Singla, Lewis Nkenyereye, Sudan Jha, Deepak Prashar, Gyanendra Prasad Joshi, Shaker El-Sappagh, Md Saiful Islam, and SM Riazul Islam. "Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives." In: *IEEE Access* 8 (2020), pp. 228049–228069.

[146] 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).

[147] Mohamed Khalifa and Mona Albadawy. "AI in diagnostic imaging: Revolutionising accuracy and efficiency." In: *Computer Methods and Programs in Biomedicine Update* (2024), p. 100146.

[148] Shaan Khurshid, Samuel Friedman, Christopher Reeder, Paolo Di Achille, Nathaniel Diamant, Pulkit Singh, Lia X Harrington, Xin Wang, Mostafa A Al-Alusi, Gopal Sarma, et al. "ECG-based deep learning and clinical risk factors to predict atrial fibrillation." In: *Circulation* 145.2 (2022), pp. 122–133.

[149] Jemin Kim, Changyoon Lee, Sungchul Choi, Da-In Sung, Jeonga Seo, Yun Na Lee, Joo Hee Lee, Eun Jin Han, Ah Young Kim, Hyun Suk Park, et al. "Augmented decision-making in wound care: evaluating the clinical utility of a deep-learning model for pressure injury staging." In: *International Journal of Medical Informatics* 180 (2023), p. 105266.

[150] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization." In: *arXiv preprint arXiv:1412.6980* (2014).

[151] Daniel Kirk, Esther Kok, Michele Tufano, Bedir Tekinerdogan, Edith JM Feskens, and Guido Camps. "Machine learning in nutrition research." In: *Advances in Nutrition* 13.6 (2022), pp. 2573–2589.

[152] R Benjamin Knapp and Hugh S Lusted. "A bioelectric controller for computer music applications." In: *Computer music journal* 14.1 (1990), pp. 42–47.

[153] 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 (2011), pp. 18–31.

[154] Petra Korica, Neamat El Gayar, and Wei Pang. "Explainable artificial intelligence in healthcare: Opportunities, gaps and challenges and a novel way to look at the problem space." In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer. 2021, pp. 333–342.

[155] Shiba Kuanar, Vassilis Athitsos, Nityananda Pradhan, Arabinda Mishra, and Kamisetty R Rao. "Cognitive analysis of working memory load from EEG, by a deep recurrent neural network." In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 2576–2580.

[156] Sajeev Kunjan, Tyler S Grummett, Kenneth J Pope, David MW Powers, Sean P Fitzgibbon, T Bastiampillai, M Battersby, and Trent W Lewis. "The necessity of leave one subject out (LOSO) cross validation for EEG disease diagnosis." In: *Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14*. Springer. 2021, pp. 558–567.

[157] Tomoyuki Kuroiwa, Aida Sarcon, Takuya Ibara, Eriku Yamada, Akiko Yamamoto, Kazuya Tsukamoto, and Koji Fujita. "The potential of ChatGPT as a self-diagnostic tool in common orthopedic diseases: exploratory study." In: *Journal of medical Internet research* 25 (2023), e47621.

[158] Seulki Kyeong, Jirou Feng, Jae Kwan Ryu, Jung Jae Park, Kyeong Ha Lee, and Jung Kim. "Surface electromyography characteristics for motion intention recognition and implementation issues in lower-limb exoskeletons." In: *International Journal of Control, Automation and Systems* 20.3 (2022), pp. 1018–1028.

[159] Luisa Lampignano, Nicola Quaranta, Ilaria Bortone, Sarah Tirelli, Roberta Zupo, Fabio Castellana, Rossella Donghia, Vito Guerra, Chiara Griseta, Pasqua Letizia Pesole, et al. "Dietary habits and nutrient intakes are associated to age-related central auditory processing disorder in a cohort from Southern Italy." In: *Frontiers in Aging Neuroscience* 13 (2021), p. 629017.

[160] Peter Lang and Margaret M Bradley. "The International Affective Picture System (IAPS) in the study of emotion and attention." In: *Handbook of emotion elicitation and assessment* 29 (2007), pp. 70–73.

[161] Eunjung Lee, Saki Ito, William R Miranda, Francisco Lopez-Jimenez, Garvan C Kane, Samuel J Asirvatham, Peter A Noseworthy, Paul A Friedman, Rickey E Carter, Barry A Borlaug, et al. "Artificial intelligence-enabled ECG for left ventricular diastolic function and filling pressure." In: *npj Digital Medicine* 7.1 (2024), p. 4.

[162] 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.

[163] Luigi Lera. "Grammatica della notazione di Notre-Dame." In: *Acta musicologica* 61.Fasc. 2 (1989), pp. 150–174.

[164] Gen Li, Chang Ha Lee, Jason J Jung, Young Chul Youn, and David Camacho. "Deep learning for EEG data analytics: A survey." In: *Concurrency and Computation: Practice and Experience* 32.18 (2020), e5199.

[165] Chengwei Liang, Fan Yang, Xiaobing Huang, Lijuan Zhang, and Ying Wang. "Deep learning assists early-detection of hypertension-mediated heart change on ECG signals." In: *Hypertension Research* (2024), pp. 1–12.

[166] Otavio G Lins, Terence W Picton, Patrick Berg, and Michael Scherg. "Ocular artifacts in EEG and event-related potentials I: Scalp topography." In: *Brain topography* 6.1 (1993), pp. 51–63.

[167] 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.

[168] Saskia Locke, Anthony Bashall, Sarah Al-Adely, John Moore, Anthony Wilson, and Gareth B Kitchen. "Natural language processing in medicine: a review." In: *Trends in Anaesthesia and Critical Care* 38 (2021), pp. 4–9.

[169] Domenico Lofù, Pietro Di Gennaro, Paolo Sorino, Tommaso Di Noia, and Eugenio Di Sciascio. "CPU-side comparison for Key Agreement between Tree Parity Machines and standard Cryptographic Primitives." In: *2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT)*. IEEE. 2022, pp. 1–6.

[170] Domenico Lofù, Paolo Sorino, Tommaso Di Noia, and Eugenio Di Sciascio. "Towards a Federated Intrusion Detection System based on Neuromorphic Computing." In: *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE. 2024, pp. 1–5.

[171] Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U Rajendra Acharya. "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.

[172] 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.

[173] 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).

[174] Yang Lu. "Artificial intelligence: a survey on evolution, models, applications and future trends." In: *Journal of Management Analytics* 6.1 (2019), pp. 1–29.

[175] Alvin Lucier and Douglas Simon. *Chambers: Scores by Alvin Lucier*. Wesleyan University Press, 2012.

[176] Christian von Lücken, 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.

[177] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. "From local explanations to global understanding with explainable AI for trees." In: *Nature machine intelligence* 2.1 (2020), pp. 56–67.

[178] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions." In: *Advances in neural information processing systems* 30 (2017).

[179] Tamra Lysaght, Hannah Yeefen Lim, Vicki Xafis, and Kee Yuan Ngiam. "AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research." In: *Asian Bioethics Review* 11 (2019), pp. 299–314.

[180] Jiantao Ma, Paul F Jacques, James B Meigs, Caroline S Fox, Gail T Rogers, Caren E Smith, Adela Hruby, Edward Saltzman, and Nicola M McKeown. "Sugar-sweetened beverage but not diet soda consumption is positively associated with progression of insulin resistance and prediabetes." In: *The Journal of nutrition* 146.12 (2016), pp. 2544–2550.

[181] Batta Mahesh. "Machine learning algorithms-a review." In: *International Journal of Science and Research (IJSR).[Internet]* 9.1 (2020), pp. 381–386.

[182] Normadiah Binti Mahiddin, Zulaiha Ali Othman, Azuraliza Abu Bakar, and Nur Arzuar Abdul Rahim. "An interrelated decision-making model for an intelligent decision support system in healthcare." In: *IEEE Access* 10 (2022), pp. 31660–31676.

[183] 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.

[184] Babita Majhi and Aarti Kashyap. "Explainable AI-driven machine learning for heart disease detection using ECG signal." In: *Applied Soft Computing* 167 (2024), p. 112225.

[185] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. "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.

[186] Filip Malawski. "Depth versus inertial sensors in real-time sports analysis: a case study on fencing." In: *IEEE sensors journal* 21.4 (2020), pp. 5133–5142.

[187] Filip Malawski and Bogdan Kwolek. "Classification of basic footwork in fencing using accelerometer." In: *2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE. 2016, pp. 51–55.

[188] 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.

[189] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

[190] Seyed Matin Malakouti. "Heart disease classification based on ECG using machine learning models." In: *Biomedical Signal Processing and Control* 84 (2023), p. 104796. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2023.104796.

[191] Barbara Mukami Maweu, Sagnik Dakshit, Rittika Shamsuddin, and Balakrishnan Prabhakaran. "CEFEs: a CNN explainable framework for ECG signals." In: *Artificial Intelligence in Medicine* 115 (2021), p. 102059.

[192] Michael McMahon and Michael Schukat. "A low-cost, open-source, BCI-VR game control development environment prototype for game based neurorehabilitation." In: *2018 ieee games, entertainment, media conference (gem)*. IEEE. 2018, pp. 1–9.

[193] Mickael Ménard, Paul Richard, Hamza Hamdi, Bruno Daucé, and Takehiko Yamaguchi. "Emotion Recognition based on Heart Rate and Skin Conductance." In: *PhyCS*. 2015, pp. 26–32.

[194] E. Merdjanovska and A. Rashkovska. "Cross-Database Generalization of Deep Learning Models for Arrhythmia Classification." In: *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. 2021, pp. 346–351. DOI: 10.23919/MIPRO52101.2021.9596930.

[195] Ahmad Moniri, Dan Terracina, Jesus Rodriguez-Manzano, Paul H Strutton, and Pantelis Georgiou. "Real-time forecasting of sEMG features for trunk muscle fatigue using machine learning." In: *IEEE Transactions on Biomedical Engineering* 68.2 (2020), pp. 718–727.

[196] Stefania Montani and Manuel Striani. "Artificial intelligence in clinical decision support: a focused literature survey." In: *Yearbook of medical informatics* 28.01 (2019), pp. 120–127.

[197] 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.

[198] Pedro A Moreno-Sánchez, Guadalupe García-Isla, Valentina DA Corino, Antti Vehkaoja, Kirsten Brukamp, Mark van Gils, and Luca Mainardi. "ECG-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review." In: *Computers in Biology and Medicine* (2024), p. 108235.

[199] Andrew Morley, L Hill, and AG Kaditis. "10-20 system EEG Placement." In: *European Respiratory Society, European Respiratory Society: Lausanne, Brussels and Sheffield* (2016).

[200] Shadman Nashif, Md Rakib Raihan, Md Rasedul Islam, and Mohammad Hasan Imam. "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system." In: *World Journal of Engineering and Technology* 6.4 (2018), pp. 854–873.

[201] Mobeen Nazar, Muhammad Mansoor Alam, Eiad Yafi, and Mazliham Mohd Su'ud. "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques." In: *IEEE Access* 9 (2021), pp. 153316–153348.

[202] William S Noble. "What is a support vector machine?" In: *Nature biotechnology* 24.12 (2006), pp. 1565–1567.

[203] CENTRO DI RICERCA ALIMENTI E NUTRIZIONE. "LINEE GUIDA PER UNA SANA ALIMENTAZIONE." In: ().

[204] Heather O'Brien. "Theoretical perspectives on user engagement." In: *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media*. Ed. by Heather O'Brien and Paul Cairns. Springer, 2016, pp. 1–26.

[205] Martin A O'Reilly, Darragh F Whelan, Tomas E Ward, Eamonn Delahunt, and Brian Caulfield. "Classification of lunge biomechanics with multiple and individual inertial measurement units." In: *Sports biomechanics* 16.3 (2017), pp. 342–360.

[206] Olugbemi T Olaniyan, Charles O Adetunji, Ayobami Dare, Olorunsola Adeyomoye, Mayowa J Adeniyi, and Alex Enoch. "Clinical applications of deep learning in neurology and its enhancements with future predictions." In: *Artificial Intelligence for Neurological Disorders*. Elsevier, 2023, pp. 209–224.

[207] Julian D Olden, Michael K Joy, and Russell G Death. "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data." In: *Ecological modelling* 178.3-4 (2004), pp. 389–397.

[208] Tolulope Olorunsogo, Adekunle Oyeyemi Adenyi, Chioma Anthonia Okolo, and Oloruntoba Babawarun. "Ethical considerations in AI-enhanced medical decision support systems: A review." In: *World Journal of Advanced Engineering Technology and Sciences* 11.1 (2024), pp. 329–336.

[209] Mahesh Pal. "Random forest classifier for remote sensing classification." In: *International journal of remote sensing* 26.1 (2005), pp. 217–222.

[210] Trishan Panch, Peter Szolovits, and Rifat Atun. "Artificial intelligence, machine learning and health systems." In: *Journal of global health* 8.2 (2018).

[211] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. "Co-design of human-centered, explainable AI for clinical decision support." In: *ACM Transactions on Interactive Intelligent Systems* 13.4 (2023), pp. 1–35.

[212] Michalis Papakostas, Varun Kanal, Maher Abujelala, Konstantinos Tsiakas, and Fillia Makedon. "Physical fatigue detection through EMG wearables and subjective user reports: a machine learning approach towards adaptive rehabilitation." In: *Proceedings of the 12th ACM international conference on pervasive technologies related to assistive environments*. 2019, pp. 475–481.

[213] 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.

[214] Farhad Parivash, Leila Amuzadeh, and Alireza Fallahi. "Design expanded BCI with improved efficiency for VR-embedded neurorehabilitation systems." In: *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE. 2017, pp. 230–235.

[215] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. psychology press, 2001.

[216] Sandeep Pushyamitra Pattyam. "Advanced AI Algorithms for Predictive Analytics: Techniques and Applications in Real-Time Data Processing and Decision Making." In: *Distributed Learning and Broad Applications in Scientific Research* 5 (2019), pp. 359–384.

[217] Fernando Pazos. "Range of adiposity and cardiorenal syndrome." In: *World Journal of Diabetes* 11.8 (2020), p. 322.

[218] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[219] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. "Scikit-learn: Machine learning in Python." In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[220] Mor Peleg and Samson Tu. "Section 5: Decision support, knowledge representation and management: Decision support, knowledge representation and management in medicine." In: *Yearbook of medical informatics* 15.01 (2006), pp. 72–80.

[221] 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.

[222] Sergio Pérez-Velasco, Diego Marcos-Martínez, Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Selene Moreno-Calderón, and Roberto Hornero. "Unraveling motor imagery brain patterns using explainable artificial intelligence based on Shapley values." In: *Computer Methods and Programs in Biomedicine* 246 (2024), p. 108048.

[223] 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.

[224] Mihaela Porumb, Saverio Stranges, Antonio Pescapè, and Leandro Pecchia. "Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG." In: *Scientific reports* 10.1 (2020), p. 170.

[225] Bikash Pradhan, Saugat Bhattacharyya, and Kunal Pal. "IoT-based applications in healthcare devices." In: *Journal of healthcare engineering* 2021.1 (2021), p. 6632599.

[226] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-computer interaction*. Addison-Wesley Longman Ltd., 1994.

[227] Rizwan Qureshi, Muhammad Irfan, Hazrat Ali, Arshad Khan, Aditya Shekhar Nittala, Shawkat Ali, Abbas Shah, Taimoor Muzaffar Gondal, Ferhat Sadak, Zubair Shah, et al. "Artificial intelligence and biosensors in healthcare and its clinical relevance: A review." In: *IEEE Access* 11 (2023), pp. 61600–61620.

[228] Sejuti Rahman, Sujan Sarker, AKM Nadimul Haque, Monisha Mushtary Uttsha, Md Fokhrul Islam, and Swakshar Deb. "AI-driven stroke rehabilitation systems and assessment: A systematic review." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2022), pp. 192–207.

[229] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. "AI in health and medicine." In: *Nature medicine* 28.1 (2022), pp. 31–38.

[230] 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.

[231] Arnab Rakshit, Amit Konar, and Atulya K Nagar. "A hybrid brain-computer interface for closed-loop position control of a robot arm." In: *IEEE/CAA Journal of Automatica Sinica* 7.5 (2020), pp. 1344–1360.

[232] Gundala Jhansi Rani, Mohammad Farukh Hashmi, and Aditya Gupta. "Surface electromyography and artificial intelligence for human activity recognition-A systematic review on methods, emerging trends applications, challenges, and future implementation." In: *IEEE Access* (2023).

[233] 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.

[234] Iain E Richardson. *Video codec design: developing image and video compression systems*. John Wiley & Sons, 2002.

[235] Chris Richter, Martin O'Reilly, and Eamonn Delahunt. "Machine learning in sports science: challenges and opportunities." In: *Sports Biomechanics* (2021), pp. 1–7.

[236] Irina Rish et al. "An empirical study of the naive Bayes classifier." In: *Workshop on Empirical Methods in Artificial Intelligence, in conjunction with International Joint Conference on Artificial Intelligence*. 2001, pp. 41–46.

[237] Matthew WM Rodger, William R Young, and Cathy M Craig. "Synthesis of walking sounds for alleviating gait disturbances in Parkinson's disease." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.3 (2013), pp. 543–548.

[238] D. Rosenboom. *Biofeedback and the Arts, Results of Early Experiments*. Aesthetic Research Centre of Canada, 1976. ISBN: 9780889850026.

[239] Ronald Rosenfeld. "Two decades of statistical language modeling: where do we go from here?" In: *Proc. IEEE* 88.8 (2000), pp. 1270–1278.

[240] Giulio Ruffini, David Ibanez, Marta Castellano, Stephen Dunne, and Aureli Soria-Frisch. "EEG-driven RNN classification for prognosis of neurodegeneration in at-risk patients." In: *Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part I 25*. Springer. 2016, pp. 306–313.

[241] James A Russell. "A circumplex model of affect." In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.

[242] Mirka Saarela and Susanne Jauhiainen. "Comparison of feature importance measures as explanations for classification models." In: *SN Applied Sciences* 3.2 (2021), p. 272.

[243] Manju Lata Sahu, Mithilesh Atulkar, Mitul Kumar Ahirwal, and Afsar Ahamad. "IoT-enabled cloud-based real-time remote ECG monitoring system." In: *Journal of medical engineering & technology* 45.6 (2021), pp. 473–485.

[244] Omar H Salman, Zahraa Taha, Muntadher Q Alsabah, Yaseein S Hussein, Ahmed S Mohammed, and Mohammed Aal-Nouman. "A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work." In: *Computer Methods and Programs in Biomedicine* 209 (2021), p. 106357.

[245] L'epidemiologia per la Sanità Pubblica. *La Qualità della Vita Vista dalle Persone con 65 Anni e Più*. Available online: https://www.epicentro.iss.it/passi-argento/ (accessed on 7 February 2022), 2022.

[246] G. Sannino and G. De Pietro. "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection." In: *Future Generation Computer Systems* 86 (2018), pp. 446–455. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2018.03.057.

[247] A Saranya and R Subhashini. "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends." In: *Decision analytics journal* 7 (2023), p. 100230.

[248] Deepti Saraswat, Pronaya Bhattacharya, Ashwin Verma, Vivek Kumar Prasad, Sudeep Tanwar, Gulshan Sharma, Pitshou N Bokoro, and Ravi Sharma. "Explainable AI for healthcare 5.0: opportunities and challenges." In: *IEEE Access* 10 (2022), pp. 84486–84517.

[249] Rodolfo Sardone, Luisa Lampignano, Vito Guerra, Roberta Zupo, Rossella Donghia, Fabio Castellana, Petronilla Battista, Ilaria Bortone, Filippo Procino, Marco Castellana, et al. "Relationship between inflammatory food consumption and age-related hearing loss in a prospective observational cohort: results from the Salus in Apulia Study." In: *Nutrients* 12.2 (2020), p. 426.

[250] LA Sargeant, KT Khaw, S Bingham, NE Day, RN Luben, S Oakes, A Welch, and NJ Wareham. "Fruit and vegetable intake and population glycosylated haemoglobin levels: the EPIC-Norfolk Study." In: *European journal of clinical nutrition* 55.5 (2001), pp. 342–348.

[251] Udit Satija, Barathram Ramkumar, and M Sabarimalai Manikandan. "Real-time signal quality-aware ECG telemetry system for IoT-based health care monitoring." In: *IEEE Internet of Things Journal* 4.3 (2017), pp. 815–823.

[252] Nina Schaffert, Thenille Braun Janzen, Klaus Mattes, and Michael H Thaut. "A review on the relationship between sound and movement in sports and rehabilitation." In: *Frontiers in psychology* 10 (2019), p. 244.

[253] Albrecht Schmidt. "Interactive human centered artificial intelligence: a definition and research challenges." In: *Proceedings of the International Conference on Advanced Visual Interfaces (AVI)*. 2020, pp. 1–4.

[254] Tjeerd AJ Schoonderwoerd, Wiard Jorritsma, Mark A Neerincx, and Karel Van Den Bosch. "Human-centered XAI: Developing design patterns for explanations of clinical decision support systems." In: *International Journal of Human-Computer Studies* 154 (2021), p. 102684.

[255] Carolina Schwedhelm, Khalid Iqbal, Sven Knüppel, Lukas Schwingshackl, and Heiner Boeing. "Contribution to the understanding of how principal component analysis–derived dietary patterns emerge from habitual data on food consumption." In: *The American journal of clinical nutrition* 107.2 (2018), pp. 227–235.

[256] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: visual explanations from deep networks via gradient-based localization." In: *International journal of computer vision* 128 (2020), pp. 336–359.

[257] Hassan Serhal, Nassib Abdallah, Jean-Marie Marion, Pierre Chauvet, Mohamad Oueidat, and Anne Humeau-Heurtier. "Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG." In: *Computers in Biology and Medicine* 142 (2022), p. 105168.

[258] Tayyebe Shabaniyan, Hossein Parsaei, Alireza Aminsharifi, Mohammad Mehdi Movahedi, Amin Torabi Jahromi, Shima Pouyesh, and Hamid Parvin. "An artificial intelligence-based clinical decision support system for large kidney stone treatment." In: *Australasian physical & engineering sciences in medicine* 42 (2019), pp. 771–779.

[259] Ben Shneiderman. "Human-centered artificial intelligence: Reliable, safe & trustworthy." In: *International Journal of Human–Computer Interaction* 36.6 (2020), pp. 495–504.

[260] Ben Shneiderman. *Human-centered AI*. Oxford University Press, 2022.

[261] Afshin Shoeibi, Delaram Sadeghi, Parisa Moridian, Navid Ghassemi, Jónathan Heras, Roohallah Alizadehsani, Ali Khadem, Yinan Kong, Saeid Nahavandi, Yu-Dong Zhang, et al. "Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models." In: *Frontiers in neuroinformatics* 15 (2021), p. 777977.

[262] Edward H Shortliffe and Martin J Sepúlveda. "Clinical decision support in the era of artificial intelligence." In: *Jama* 320.21 (2018), pp. 2199–2200.

[263] Ajay Shrestha and Ausif Mahmood. "Review of deep learning algorithms and architectures." In: *IEEE access* 7 (2019), pp. 53040–53065.

[264] Vanderlei Carneiro Silva, Bartira Gorgulho, Dirce Maria Marchioni, Tânia Aparecida de Araujo, Itamar de Souza Santos, Paulo Andrade Lotufo, and Isabela Martins Benseñor. "Clustering analysis and machine learning algorithms in the prediction of dietary patterns: Cross-sectional results of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil)." In: *Journal of Human Nutrition and Dietetics* 35.5 (2022), pp. 883–894.

[265] Mohammad Soleymani, Joep JM Kierkels, Guillaume Chanel, and Thierry Pun. "A bayesian framework for video affective representation." In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE. 2009, pp. 1–7.

[266] 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.

[267] Nisha Soms, David Samuel Azariya, J Jeba Emilyn, and Abhinaya Saravanan. "Ensuring Ethical Standards and Equity in Explainable Artificial Intelligence Applications Within Healthcare." In: *International Conference on Artificial Intelligence and Smart Energy*. Springer. 2024, pp. 369–380.

[268] 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.

[269] Anthony Sorel, Pierre Plantard, Nicolas Bideau, and Charles Pontonnier. "Studying fencing lunge accuracy and response time in uncertain conditions with an innovative simulator." In: *Plos One* 14.7 (2019), e0218959.

[270] Paolo Sorino. "Blockchain and AI to Build an Alzheimer's Risk Calculator." In: *International Conference on Web Engineering*. Springer. 2022, pp. 432–436.

[271] 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.

[272] 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)*. **To appear**. 2024.

[273] 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.

[274] Shiliang Sun and Rongqing Huang. "An adaptive k-nearest neighbor algorithm." In: *Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE. 2010, pp. 91–94.

[275] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. "An overview of clinical decision support systems: benefits, risks, and strategies for success." In: *NPJ digital medicine* 3.1 (2020), p. 17.

[276] Kenji Suzuki and Yisong Chen. *Artificial intelligence in decision support systems for diagnosis in medical imaging*. Vol. 140. Springer, 2018.

[277] 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.

[278] Peter D Symes. *Video Compression*. McGraw-Hill Professional, 1998.

[279] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[280] Scott D Tagliaferri, Maia Angelova, Xiaohui Zhao, Patrick J Owen, Clint T Miller, Tim Wilkin, and Daniel L Belavy. "Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: three systematic reviews." In: *NPJ digital medicine* 3.1 (2020), p. 93.

[281] K Song Tan, Reza Saatchi, Heather Elphick, and Derek Burke. "Real-time vision based respiration monitoring system." In: *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*. IEEE. 2010, pp. 770–774.

[282] Tien-En Tan, Ayesha Anees, Cheng Chen, Shaohua Li, Xinxing Xu, Zengxiang Li, Zhe Xiao, Yechao Yang, Xiaofeng Lei, Marcus Ang, et al. "Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study." In: *The Lancet Digital Health* 3.5 (2021), e317–e329.

[283] Rossella Tatoli, Luisa Lampignano, Ilaria Bortone, Rossella Donghia, Fabio Castellana, Roberta Zupo, Sarah Tirelli, Sara De Nucci, Annamaria Sila, Annalidia Natuzzi, et al. "Dietary patterns associated with diabetes in an older population from Southern Italy using an unsupervised learning approach." In: *Sensors* 22.6 (2022), p. 2193.

[284] Michal Teplan et al. "Fundamentals of EEG measurement." In: *Measurement science review* 2.2 (2002), pp. 1–11.

[285] Tagne Poupi Theodore Armand, Kintoh Allen Nfor, Jung-In Kim, and Hee-Cheol Kim. "Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review." In: *Nutrients* 16.7 (2024), p. 1073.

[286] Michael Tipping. "The relevance vector machine." In: *Advances in neural information processing systems* 12 (1999).

[287] 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.

[288] Anthony Turner, Nic James, Lygeri Dimitriou, Andy Greenhalgh, Jeremy Moody, David Fulcher, Eduard Mias, and Liam Kilduff. "Determinants of Olympic fencing performance and implications for strength and conditioning training." In: *The journal of strength & conditioning research* 28.10 (2014), pp. 3001–3011.

[289] Arijit Ukil, Soma Bandyoapdhyay, Chetanya Puri, and Arpan Pal. "IoT healthcare analytics: The importance of anomaly detection." In: *2016 IEEE 30th international conference on advanced information networking and applications (AINA)*. IEEE. 2016, pp. 994–997.

[290] Esther Van Eekelen, Anouk Geelen, Marjan Alssema, Hildo J Lamb, Albert De Roos, Frits R Rosendaal, and Renée de Mutsert. "Sweet snacks are positively and fruits and vegetables are negatively associated with visceral or liver fat content in middle-aged men and women." In: *The Journal of Nutrition* 149.2 (2019), pp. 304–313.

[291] Damini Verma, Kshitij RB Singh, Amit K Yadav, Vanya Nayak, Jay Singh, Pratima R Solanki, and Ravindra Pratap Singh. "Internet of things (IoT) in nano-integrated wearable biosensor devices for healthcare applications." In: *Biosensors and Bioelectronics: X* 11 (2022), p. 100153.

[292] Ankit Vijayvargiya, Puneet Singh, Rajesh Kumar, and Nilanjan Dey. "Hardware implementation for lower limb surface EMG measurement and analysis using explainable AI for activity recognition." In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–9.

[293] Filipa Campos Viola, Stefan Debener, Jeremy Thorne, and Till R Schneider. "Using ICA for the analysis of multi-channel EEG data." In: *Simultaneous EEG and fMRI: Recording, Analysis, and Application: Recording, Analysis, and Application* (2010), pp. 121–133.

[294] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. ""Brilliant AI doctor" in rural clinics: Challenges in AI-powered clinical decision support system deployment." In: *Proceedings of the 2021 CHI conference on human factors in computing systems*. 2021, pp. 1–18.

[295] Ping Wang, Aimin Jiang, Xiaofeng Liu, Jing Shang, and Li Zhang. "LSTM-based EEG classification in motor imagery tasks." In: *IEEE transactions on neural systems and rehabilitation engineering* 26.11 (2018), pp. 2086–2095.

[296] Xinru Wang and Ming Yin. "Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making." In: *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 2021, pp. 318–328.

[297] Yixin Wang, Shuang Qiu, Dan Li, Changde Du, Bao-Liang Lu, and Huiguang He. "Multi-Modal Domain Adaptation Variational Autoencoder for EEG-Based Emotion Recognition." In: *IEEE/CAA Journal of Automatica Sinica* (2022).

[298] Niyaz Ahmad Wani, Ravinder Kumar, Jatin Bedi, Imad Rida, et al. "Explainable AI-driven IoMT fusion: Unravelling techniques, opportunities, and challenges with Explainable AI in healthcare." In: *Information Fusion* (2024), p. 102472.

[299] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges." In: *Artificial Intelligence Review* 55.7 (2022), pp. 5731–5780.

[300] Shalini Wankhade and S Vigneshwari. "A novel hybrid deep learning method for early detection of lung cancer using neural networks." In: *Healthcare Analytics* 3 (2023), p. 100195.

[301] Drishti Yadav and Karan Veer. "Recent trends and challenges of surface electromyography in prosthetic applications." In: *Biomedical Engineering Letters* 13.3 (2023), pp. 353–373.

[302] Yujun Yang, Jianping Li, and Yimei Yang. "The research of the fast SVM classifier method." In: *Proceedings of the International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE. 2015, pp. 121–124.

[303] Safoora Yousefi, Fatemeh Amrollahi, Mohamed Amgad, Chengliang Dong, Joshua E Lewis, Congzheng Song, David A Gutman, Sameer H Halani, Jose Enrique Velazquez Vega, Daniel J Brat, et al. "Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models." In: *Scientific reports* 7.1 (2017), pp. 1–11.

[304] Jinyang Yu, Anders Austlid Taskén, Erik Andreas Rye Berg, Tomas Dybos Tannvik, Katrine Hordnes Slagsvold, Idar Kirkeby-Garstad, Bjørnar Grenne, Gabriel Kiss, and Svend Aakhus. "Continuous monitoring of left ventricular function in postoperative intensive care patients using artificial intelligence and transesophageal echocardiography." In: *Intensive Care Medicine Experimental* 12.1 (2024), p. 54.

[305] 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.

[306] Matteo Zago, Ana Francisca Rozin Kleiner, and Peter Andreas Federolf. "Machine learning approaches to human movement analysis." In: *Frontiers in Bioengineering and Biotechnology* (2021), p. 1573.

[307] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. "Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–21.

[308] Ce Zhang and Azim Eskandarian. "A survey and tutorial of EEG-based brain monitoring for driver state analysis." In: *arXiv preprint arXiv:2008.11226* (2020).

[309] Xiao Zhang and Dongrui Wu. "On the vulnerability of CNN classifiers in EEG-based BCIs." In: *IEEE transactions on neural systems and rehabilitation engineering* 27.5 (2019), pp. 814–825.

[310] 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.

[311] 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.

[312] Paola Zinno, Francesco Maria Calabrese, Emily Schifano, Paolo Sorino, Raffaella Di Cagno, Marco Gobbetti, Eugenio Parente, Maria De Angelis, and Chiara Devirgiliis. "FDF-DB: A Database of Traditional Fermented Dairy Foods and Their Associated Microbiota." In: *Nutrients* 14.21 (2022), p. 4581.

[313] 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.

[314] Roberta Zupo, Rodolfo Sardone, Rossella Donghia, Fabio Castellana, Luisa Lampignano, Ilaria Bortone, Giovanni Misciagna, Giovanni De Pergola, Francesco Panza, Madia Lozupone, et al. "Traditional dietary patterns and risk of mortality in a longitudinal cohort of the Salus in Apulia Study." In: *Nutrients* 12.4 (2020), p. 1070.

[315] Roberta Zupo, Rodolfo Sardone, Rossella Donghia, Fabio Castellana, Luisa Lampignano, Ilaria Bortone, Giovanni Misciagna, Giovanni De Pergola, Francesco Panza, Madia Lozupone, et al. "Traditional dietary patterns and risk of mortality in a longitudinal cohort of the Salus in Apulia Study." In: *Nutrients* 12.4 (2020), p. 1070.

Appendix

Supplementary table 1

Concordance of the single foods in the questionnaire and the food grouping used in the analyses.

Final Food Groups	Single Foods from Questionnaire
1. DAIRY	Latte intero (whole-fat milk), Scamorza-Caciottina fresca-Stracchino-Fontina (semi-seasoned italian cheese), Bel Paese-Gorgonzola (italian blue cheese), Provolone-Caciocavallo (seasoned italian cheese), Grana-Parmigiano, Svizzero (swiss cheese), Pecorino-Vacchino (goat cheese, cow cheese), Formaggino (cheese spread), Mozzarella (mozzarella cheese), Gelato (ice cream), Yogurt
2. LOW FAT DAIRY	Latte scremato–parzialmente scremato (skimmed and semi-skimmed milk), Ricotta (cottage cheese)
3. EGGS	Uova (eggs)
4. WHITE MEAT	Pollo (chicken), Coniglio (rabbit)
5. RED MEAT	Vitello (veal), Cavallo (horse), Maiale (Pork), Fegato (liver), Agnello (lamb)
6. PROCESSED MEAT	Salsiccia fresca (fresh sausages), Prosciutto crudo (raw ham), Mortadella (a typical Italian cured meat), Prosciutto cotto (ham), Salame (salami)
7. FISH	Sogliola-Orata-Dentice-Spigola-Cernia (sole, sea bream, snapper, sea bass, grouper), Merluzzo-Razza-Palombo (codfish, stingray, dogfish), Triglia-Cefalo-Sgombro (goatfish, mullet, mackerel), Acciughe-Sarde (anchovies, sardines), Tonno sott'olio (tuna in oil)
8. SEAFOOD/SHELLFISH	Polpo-Seppie-Calamari-Gamberi (octopus, cuttlefish, squid, prawns), Cozze-Altri frutti di mare (mussels, other seafoods)
9. LEAFY VEGETABLES	Spinaci (spinach), Bietole-Cicorie (chard, chicory), Insalata (salad)
10. FRUITING VEGETABLES	Pomodori (tomatoes), Zucchine-Melanzane (zucchini, eggplants), Peperoni (peppers), Carciofi (artichokes), Cetrioli-cocomeri (cucumbers)
11. ROOT VEGETABLES	Carote (carrots)
12. OTHER VEGETABLES	Minestrone (vegetable soup), Cavoli–Cavolfiori-Cime di Rape-Rape (cabbage, cauliflower, broccoli, green turnips), Finocchi-Sedano (fennels, celery)
13. LEGUMES	Ceci - Lenticchie–Fagioli (chickpeas, lentils, beans), Piselli (peas), Fagiolini (green beans), Fave con Verdura (broad beans with vegetables)
14. POTATOES	Patate (potatoes)
15. FRUITS	Arance-Mandarini-Pompelmi (oranges, tangerines, grapefruits), Pesche (peaches), Fichi (figs), Albicocche (apricots), Uva (grapes), Anguria (watermelon), Melone giallo (melon), Mele-Pere (apples, pears), Kiwi, Ciliege (cherries), Banane (bananas)
16. NUTS	Frutta secca (nuts)
17. GRAINS	Pane (bread), Pasta asciutta (pasta), Riso o risotti (rice or risotti), Pastina o riso in brodo (pasta or rice in broth)
18. OLIVES AND VEGETABLE OIL	Olive da tavola (olives), Olio di oliva (olive oil)
19. COOKING EDIBLE FATS	Olio di semi (seed oil), Olio di oliva per frittura (olive oil for frying), Olio di semi per frittura (seeds oil for frying), Olio di oliva per cucinare (olive oil for cooking), Olio di semi per cucinare (seeds oil for cooking), Burro (butter), Margarina (margarine), Burro per frittura (butter for frying), Margarina per frittura (margarine for frying), Burro per cucinare (butter for cooking), Margarina per cucinare (margarine for cooking)
20. SWEETS	Caramelle (sweets), Cioccolata (chocolate), Pasticceria (pastries), Biscotti–Paste secche (cookies, biscuits)
21. SUGARY	Zucchero (sugar), Frutta sciroppata (fruit in syrup)
22. JUICES	Succhi di frutta (fruit juice)
23. CALORIC DRINKS	Coca Cola–Aranciata–Chinotto (coke, orange juice, chinotto)
24. READY TO EAT DISH	Pizza, Focaccia (a typical Apulian bakery product)
25. COFFEE	Caffè (coffee), Caffè d'orzo (barley coffee)
26. WINE	Vino (wine)
27. BEER	Birra (beer)
28. SPIRITS	Liquore (liquor)
29. WATER	Acqua (water)

Supplementary Table 2

A dataset snippet, relating to a single user (i.e. User1), acquired during training of the mental state classification system. The first 50 items were collected while the user was stimulated to the Focused state by reading sequences of numbers, the next 50 while listening to natural environment sounds to induce the Relaxed state.

User	x_i	*Delta*				*Theta*				*Alpha*						*Beta*					*Gamma*					Target
		AF3	T7	PZ	T8	AF4	AF3	T7	PZ	T8	AF4	AF3	T7	PZ	T8	AF4	AF3	T7	PZ	T8	AF4	AF3	T7	PZ	T8	AF4	
User1	x_1	1.643	0.967	0.255	0.386	0.595	0.111	1.026	0.603	1.474	0.603	0.931	0.231	0.572	1.137	1.306	0.526	1.028	0.441	0.795	1.645	1.479	1.639	1.544	1.647	1.543	Focused
	x_2	0.765	0.504	0.836	0.627	0.524	0.733	0.891	0.437	0.415	0.468	0.444	0.400	1.127	1.398	1.127	1.192	1.224	0.476	1.221	1.183	1.278	0.758	0.953	1.426	0.665	Focused
	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
	x_{50}	1.174	1.040	0.014	1.264	1.383	1.156	0.064	0.567	1.056	1.094	0.336	0.403	1.009	0.716	0.935	0.354	0.265	1.244	1.150	1.401	0.161	0.273	0.148	0.734	0.289	Focused
	x_{51}	1.343	0.148	0.066	0.835	1.158	0.467	0.268	0.508	0.315	1.359	0.943	0.152	0.586	0.081	0.751	0.647	1.496	1.217	0.728	1.341	0.206	0.585	1.391	1.376	1.070	Relaxed
	x_{52}	0.927	0.514	1.404	0.187	1.095	0.969	1.249	0.597	1.124	1.252	0.483	0.828	1.468	0.824	0.495	0.929	0.541	1.134	0.620	0.738	1.042	1.459	0.491	1.256	1.108	Relaxed
	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
	x_{100}	1.431	0.047	0.535	0.994	0.422	0.345	1.066	0.936	0.885	0.990	0.071	0.523	0.677	0.361	1.072	1.284	0.422	1.096	0.206	1.255	0.207	0.882	0.549	1.210	1.002	Relaxed