Fall 12-2021

Machine-Learning-Based Approach to Decoding Physiological and Neural Signals

Elnaz Lashgari
Chapman University, lashgari@chapman.edu

Machine-Learning-Based Approach to Decoding Physiological and Neural Signals

A Dissertation by

Elnaz Lashgari

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Data Sciences

December 2021

Committee in charge:

Uri Maoz, Ph.D., Chair

Erik Linstead, Ph.D.

Aaron Schurger, Ph.D.

CHAPMAN UNIVERSITY
SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY
Computational and Data Sciences

The dissertation of Elnaz Lashgari is approved.

Uri Maoz, Ph.D., Chair

Erik Linstead, Ph.D.

Aaron Schurger, Ph.D.

November 2021

Machine-Learning-Based Approach to Decoding Physiological and Neural Signals

by Elnaz Lashgari

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my supervisor, Dr. Uri Maoz for his invaluable advice, continuous support, and patience during my PhD study. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I would also like to thank the rest of my thesis committee, Dr. Aaron Schurger and Dr. Erik Linstead for their technical support and insightful comments on my study. I would like to thank Prof. Amir Raz and all the members in the Institute for interdisciplinary brain and behavioral sciences. It is their kind help and support that have made my study and life in the US a wonderful time. Finally, I would like to express my gratitude to my parents, my beloved husband, and my family. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

LIST OF PUBLICATIONS

Lashgari, E., Liang, D. and Maoz, U., 2020. Data augmentation for deep-learning-based electroencephalography. Journal of Neuroscience Methods, p.108885.

Omura, Y., Kipke, J.P., Salavatian, S., Afyouni, A.S., Wooten, C., Herkenham, R.F., Maoz, U., Lashgari, E., Dale, E.A., Howard-Quijano, K. and Mahajan, A., 2021. Spinal Anesthesia Reduces Myocardial Ischemia–triggered Ventricular Arrhythmias by Suppressing Spinal Cord Neuronal Network Interactions in Pigs. Anesthesiology, 134(3), pp.405-420.

Lashgari, E., Ott, J., Connelly, A., Baldi, P. and Maoz, U., 2021. An end-to-end CNN with attentional mechanism applied to raw EEG in a BCI classification task. Journal of Neural Engineering, 18(4), p.0460e3.

Lashgari, E. and Maoz, U., 2021. Dimensionality reduction for classification of object weight from electromyography. Plos one, 16(8), p.e0255926.

ABSTRACT

Machine-Learning-Based Approach to Decoding Physiological and Neural Signals

by Elnaz Lashgari

In recent years, machine learning algorithms have been developing rapidly, becoming increasingly powerful tools in decoding physiological and neural signals. The aim of this dissertation is to develop computational tools, and especially machine learning techniques, to identify the most effective methods for feature extraction and classification of these signals. This is particularly challenging due to the highly non-linear, non-stationery, and artifact- and noise-prone nature of these signals.

Among basic human-control tasks, reaching and grasping are ubiquitous in everyday life. I investigated different linear and non-linear dimensionality reduction techniques for feature extraction and classification of electromyography (EMG) during a reach-grasp-lift task. The results highlighted the advantages of completely automated feature-learning by Laplacian Eigenmaps over manual feature engineering, especially when combined with classification, to achieve high accuracy with few training samples. The ability to decode and reduce the complexity of EMG could enable new practical applications for EMG in basic science and in the clinic. It could also help design humanoid and other robots.

Beyond EMG, a key objective of my dissertation was to decode brain activity during the decision-making processes that lead to voluntary action. This was based on electroencephalography (EEG) and holds the promise for improved brain-computer interfaces (BCIs), particularly related to motor imagery (MI). We developed an end-to-end convolutional neural network with attentional

mechanism together with different data augmentation techniques on two benchmark MI datasets. I also collected a new dataset, recorded using high-density EEG, and containing both MI and motor execution (ME) tasks. This enabled us to directly compare the decodability of MI and ME, investigate optimal channel configurations, and much more. In particular, this facilitates the analysis and decoding of MI on the fly—online and in in real time.

Another potential use of EEG is measuring brain activity of people who are floating in body-temperature water in a sensory-deprivation-tank float pod. We compared lying in the float pod versus lying in bed (a control condition). And we found differences between the two, especially in the gamma band. More research is required to understand what these findings mean for levels of stress in the float pod.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS .... **IV**

LIST OF PUBLICATIONS .... **V**

ABSTRACT .... **VI**

LIST OF TABLES .... **XII**

LIST OF FIGURES .... **XIV**

1 INTRODUCTION **1**

2 DIMENSIONALITY REDUCTION FOR CLASSIFICATION OF OBJECT WEIGHT FROM ELECTROMYOGRAPHY **7**

2.1 Introduction **7**

2.2 Materials and Method **10**

2.2.1 Dataset 12

2.2.2 Pre-processing 14

2.2.3 Segmentation and feature selection 14

2.2.4 Feature Extraction using dimensionality reduction 14

2.2.5 The Laplacian Eigenmap algorithm 16

2.2.6 Classification 21

2.2.7 Running-window analysis 22

2.3 Results **23**

2.3.1 Parameter settings 23

2.3.2 Classifier Performance 27

2.3.3 Evolution of classification accuracy over time 28

2.4 Discussion **32**

2.5 Conclusion, Limitations and Future Work **35**

3 DATA AUGMENTATION FOR DEEP-LEARNING-BASED ELECTROENCEPHALOGRAPHY **36**

3.1 Introduction **36**

3.2 Methods **40**

3.2.1 Search method for identification of related studies 40

3.2.2 Data Extraction and presentation 42

3.3 Results 42
3.3.1 Origin of the selected studies 42
3.3.2 EEG classification task 43
3.3.3 Data and Reproducibility 45
3.3.4 Pre-processing and Feature extraction 45
3.3.5 Input Formulation 47
3.3.6 Deep learning architectures 48
3.4 Data Augmentation methods 51
3.4.1 Noise addition 52
3.4.2 Generative adversarial network 58
3.4.3 Sliding window or overlapping window 71
3.4.4 Sampling 75
3.4.5 Fourier Transform 78
3.4.6 Recombination of Segmentation 80
3.4.7 Other 81
3.5 Accuracy gains of data augmentation 85
3.6 Discussion 87
3.6.1 Rationale 88
3.6.2 Data 88
3.6.3 EEG pre-processing 89
3.6.4 Deep-learning methodology 90
3.6.5 Data augmentation 90
3.6.6 Guidelines for reporting results in papers 93
3.7 Limitations 94
3.8 Conclusions 95

4 AN END-TO-END CNN WITH ATTENTIONAL MECHANISM APPLIED TO RAW EEG IN A BCI CLASSIFICATION TASK 122
4.1 Introduction 122
4.2 Related work 124
4.3 Methods 127
4.3.1 Proposed CNN-based neural-network architecture 127
4.3.2 Hyperparameter Optimization and Training 131
4.3.3 Data augmentation 135
4.4 Dataset and experimental protocol 138
4.5 Channel selection 142
4.6 Results 144
4.6.1 Performance of the proposed CNN (Neural architectures vs. Neural architectures) 144
4.6.2 Properties of our collected dataset 149
4.6.3 Motor Imagery vs. Motor Execution 150
4.6.4 Channel selection 153

4.6.5 Data augmentation 154
4.6.6 Different portions of dataset 158
4.6.7 Combination of participants' EEG signals 159
4.6.8 Leave-one-participant out and transfer learning 161
4.7 Discussion 163

5 A SYSTEMATIC REVIEW ON RESTRICTED ENVIRONMENTAL STIMULATION THERAPY (REST) FLOATATION 174
5.1 Introduction 174
5.2 Methods 176
5.2.1 Search method for identification of related studies 176
5.2.2 Data extraction and presentation 178
5.3 Results 178
5.3.1 Origin of the selected studies and different applications of float pod 178
5.3.2 Sample characteristics 180
5.3.3 Experiment design 181
5.3.4 Treatment delivery 182
5.3.5 Questionnaire 184
5.4 Float pod application 186
5.4.1 Positive effects of float pod 187
5.4.2 Pain 200
5.4.3 Anxiety 205
5.4.4 Physiological 207
5.4.5 Consciousness 213
5.4.6 Sleep 217
5.4.7 Psychology 217
5.5 Discussion 218
5.6 Limitations and strengths 221
5.7 Future research and potential implications 222

6 SHORT-TERM FLOATATION-REST (RESTRICTED ENVIRONMENTAL STIMULATION TECHNIQUE) REDUCES STRESS: ANALYSES BASED ON NEURAL AND CARDIAC COMPONENTS 274
6.1 Introduction 274
6.2 Methods and Materials 277
6.2.1 Participants Recruitment and Randomization 277
6.2.2 Data acquisition and pre-processing 277
6.2.3 Heart-beat detection and the RR time-series 278
6.2.4 Extraction of the root mean of successive heart beats differences 279
6.2.5 Power Spectrum of RR time-series 280
6.2.6 EEG signal processing 280
6.2.7 Statistical Analysis 281
6.2.8 Sliding-window Analysis 282

6.3 Results **282**
6.3.1 Temporal and spectral analysis of RR intervals 283
6.3.2 Spectral Analysis of the EEG signal 286
6.3.3 Sliding-window analysis 289
6.4 Discussion **292**

REFERENCE **296**

APPENDICES **316**

LIST OF TABLES

Page

Table 1. The number of trials for each class and each subject 13

Table 2. The number of dimensions that leads to the highest F1-score (obtained using grid search; for k-NN, which was the best classifier; see also Figure 8) vs. different dimensionality reduction techniques for each subject. The mean and standard error (SE) over all subjects for the different methods are also given. 25

Table 3. Performance of Laplacian Eigenmaps (simple-minded) and different classifiers vs. different number of neighbors (k) on the EMG signals (F1 score $\pm$ SE) 31

Table 4. Performance of different classifiers vs. different dimensionality-reduction methods on EMG signals (F1 score $\pm$ SE). See S1 Table 4 for post-hoc t-tests for this table. 31

Table 5. Data items extracted for each article selected 42

Table 6. All reviewed papers that used noise addition as their DA technique 56

Table 7. All reviewed papers that used GAN as their DA technique 67

Table 8. Details of all the papers that we found for our review paper. In Data Augmentation column: NA: Noise addition, SW: Sliding window, S: Sampling, FT: Fourier-transform, Recombination of Segmentation: RS, O: Others and in EEG task column: ER: Emotion recognition, MW: Mental workload, MI: Motor imagery, S: Seizure, SS: Sleep stages, IT: Imagery task, VT: Visual task and in input formulation column: S: signal, I: Images and CF: Calculated features. Some studies used different dataset or different DA techniques and we show them separately. 97

Table 9 Summary of the proposed CNN with attentional mechanism parameters (“-1” represents a flexible shape, essentially the batch size) 130

Table 10. The hyperparameter space 134

Table 11. Hyperparameter tuning by SHERPA for 3 the datasets (BCI 2a, BCI 2b and our experimental dataset) for 3 different models (Dense, CNN dense, and CNN attention dense) 135

Table 12. DA techniques that are used on the MI task 136

Table 13. Summary of the 3 datasets used in this study: Our experimental dataset, BCI 2a, and BCI 2b 140

Table 14. Participant-by participant comparison of the proposed CNN with attentional mechanism—with and without DA—against Dai et al. [145] results on the BCI 2a and BCI 2b datasets..147

Table 15. Comparison of our proposed method (with and without data augmentation) with other state-of-the-art methods. All methods were run on the same dataset (BCI 2a and/or BCI 2b). 148

Table 16. Validation accuracy for different channel selections on our dataset for single participants and the average over all participants. For each participant, we present mean ± SE over trials. In the bottom row, we present mean ± SE over participants.........................154

Table 17. Comparison of different DA techniques with different magnification factors and hyperparameters for BCI 2a, BCI 2b, and our experimental dataset (for 64 channels and 18 channels)..157

Table 18. Accuracies for different proportion of our dataset with different DA techniques158

Table 19. Leave-one-out and transfer-learning validation accuracy for BCI 2a, BCI 2b, and our dataset (64 and 18 channels)..162

Table 20. Data items extracted from each selected article 178

Table 21. Treatment delivery for different application of float pod..............................182

Table 22. Different questionnaires used in this review paper 184

Table 23. Details of all the papers that we found for our review paper. 226

LIST OF FIGURES

Page

Figure 1. Processing pipeline for EMG signal classification 11

Figure 2. EMG preprocessing. (Top) Pre-processed EMG signals of 5 muscles (Anterior Deltoid, Brachoradial, Flexor Digitorum, Common Extensor Digitorum, and First Dorsal Interosseus). (Bottom) Rectified and filtered EMG signals by band-pass Butterworth filter (4th order) in the 5-450 Hz range on the full wave rectified and normalized signals from each muscle. 11

Figure 3. Raw EMG signals of 5 muscles (Anterior Deltoid, Brachioradialis, Flexor Digitorum, Common Extensor Digitorum, and First Dorsal Interosseous) for 3 different weights (165, 330, and 660 g) 13

Figure 4. Manifold learning techniques. MDS, ISOMAP, LLE, t-SNE, and Spectral embedding (SE) or Laplacian Eigenmaps on 2000 points randomly distributed on the surface of a sphere. Computation time in seconds is given after each method's name in parentheses. The first column for SE is simple-minded constructing weight matrix and the last column is for heat kernel. 20

Figure 5. Example of an out-of-sample extension via the Nyström approximation in embedded space by Laplacian Eigenmaps with k = 8 and $\sigma = 1$. The train / test ratio was 90% / 10%. 21

Figure 6. Visualization of the effect of k (top row) and σ (bottom row) on the training dataset for Subject 1 in the embedded space 24

Figure 7. Visualization of the embedding process: EMG visualization using the 2 most prominent components of different dimensionality-reduction technique (The x-axis is the most prominent component, and the y-axis is the second most prominent component) 26

Figure 8. Average accuracy (across all 12 subject) as a function of the number of eigenvectors obtained with different dimensionality reduction techniques for the best classifier—k-NN (see also Table 3). We included PCA, ISOMAP, LLE, Laplacian Eigenmaps, and t-SNE. The LDA dimensionality-reduction technique was not included in these curves because the maximal dimension for LDA is equal to the number of classes minus one. 27

Figure 9. Sliding-window analysis of Laplacian Eigenmap dimensionality reduction and k-NN classification before touching the object (a) and after touching it (b). We used a 100 ms sliding window with a step size of 40 ms. In both panels, the onset of touching the object is designated by a vertical red line at time 0. Table 3. Performance of Laplacian Eigenmaps (simple-minded) and different classifiers vs. different number of neighbors (k) on the EMG signals (F1 score $\pm$ SE) 30

Figure 10. Selection process for the papers 41

Figure 11. EEG classification task. (A) Number of publications per domain of EEG task per year. (B)The percentage of different EEG classification task across all studies. 44

Figure 12. Frequency range used in EEG analysis for each identified study, organized by EEG task type. 46

Figure 13. Input formulation across all reviewed papers. (A) The inner circle shows the general input formulation, while the outer circle shows more specific details. (B) Number of papers for general input formulation compared across different tasks. 48

Figure 14. Deep learning architecture across all studies 49

Figure 15. Aggregated information of deep learning architectures. The inner circle shows the general DL architecture, the middle circle, shows the primary design features, such as the hidden layers or convolutional layers, and the outer circle shows the last layer of DL architecture. FC: Fully connected, hid: Hidden layers, softmax: Softmax function. 50

Figure 16. The percentage of input formulation by chosen DL architecture 51

Figure 17. DA across all studies. (A) Number of publications per domain of DA per year. (B) The percentage of different DA methods across all studies. Note that we only collected data until January 2020. 53

Figure 18. Diagram of Generative Adversarial Network 60

Figure 19. Data augmentation methods across all reviewed papers. The inner circle shows the general DA methods, and the outer circle shows the deep learning architecture strategy used. 82

Figure 20. Number of papers for general EEG tasks compared across different DA techniques 85

Figure 21. (A) The improvement score, or the fraction of variance left unexplained by the original DL method that was explained when training the model using DA, for different DA techniques (mean ± 95% confidence intervals). Here is the number of studies everywhere except GAN, Sampling, and Other, where there were more than one analysis, with different accuracies, reported in each study; hence there 'n' is the number of accuracies. (B) Same as A but the improvement is over EEG tasks. Here "n" is the number of studies everywhere except "emotion recognition", where there were 9 studies, 2 of which ran multiple DA analyses; hence "n" there is the number of analyses. No motor-task studies included accuracy before and after DA, so that task is not included in this figure. 87

Figure 22. Our proposed CNN with attentional mechanism. (A) The sliding window (length is 1000 ms and step-size is 100 ms) applied to 64 EEG channels. (B) The 64 segments of raw EEG signal, depicted in orange in (A). Each time window and channel are separately sent through shared convolution layers. The embedded features I (C x E) applied to self-attention. The output of self-attention passes through 2 dense layers. (C) An expansion of the self-attention block. 130

Figure 23. The experimental paradigms for our experimental dataset, BCI 2a, and BCI 2b. 142

Figure 24. Four different electrode configurations on the actiCAP—which included 3, 7, 18, and all 64 electrodes........144

Figure 25. Comparison the average validation accuracy (±SE) on the BCI 2a and BCI 2b datasets with Dense, CNN, Dai et al. (2020), and CNN-Attention-Dense (See section 2.1 and 2.2).........145

Figure 26. Validation accuracy of sliding-window analysis in ME (top), MI (middle), and ME and MI combined (bottom). The left column is the accuracy over time averaged across all 7 participants. The right column depicts the accuracy for the participant with the highest overall accuracy in the ME condition (Participant 4).........152

Figure 27. Validation accuracy for different channel configurations on the 7 participants of our dataset........154

Figure 28. Our proposed cGAN model. In the generator (G), the prior input noise and label are combined into a hidden representation. In the discriminator (D), Real Data (i.e., raw EEG data) and the Label are presented as inputs to a discriminative function. The contents of all purple boxes in the architecture are the same and are expanded at the bottom left........156

Figure 29. (Top)Validation accuracies for combinations of participants for BCI 2a, BCI 2b, and our experimental dataset. (Bottom) line plots of differences between mean validation accuracies of consecutive groups for the 3 datasets. The x axis labels are the smaller groups; so, differences between 2 participants and one are plotted above the label "1 participant", between 3 and 2 participants above "2 participants", and so on. 161

Figure 30. Validation accuracy for BCI 2a, BCI 2b, and our dataset (64 and 18 channels) with and without transfer learning........163

Figure 31. Selection process for the papers........177

Figure 32. The percentage of different journals categories across all reviewed studies .179

Figure 33. Different applications of float pods. (Left) Number of publications per domain for different applications of float pods. (Right) Percentage of different applications of float pods across all studies........180

Figure 34. Total number of participants in each application of float pod across 46 studies181

Figure 35. Experiment design for 46 studies........182

Figure 36. Aggregated information of different questionnaires across different types of float pod applications........186

Figure 37. (Left) A customized 6-channel LiveAmp mobile EEG amplifier from Brain Products GmbH was used to collect EEG signal from the frontal lobe and bipolar ECG signal. (Right) Sensory deprivation tank. 278

Figure 38. (Top) Raw ECG signal, (Middle) RR tachograms, (Bottom) $RRLF$ and $RRHF$ components for 40 seconds RR time series. 279

Figure 39. Power spectrum of RR time series for subject 4, interval 3 in pod (left) and bed (right) for 3 frequency bands (VLF: 0.0033–0.04 Hz, LF: 0.04–0.15 Hz, and HF: 0.15–0.40 Hz)280

Figure 40. Point estimates of the RR intervals in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals. 284

Figure 41. Point estimates of the RMSSD in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals. 285

Figure 42. Point estimates of the LF/HF ratio in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals. 286

Figure 43. (a) Distribution of the dependent variable. (b) Distribution of the model's residuals. (c) Residual distribution vs. predicted values suggesting residuals' heteroscedasticity. 287

Figure 44. (a) Differences in total PSD between the bed and pod. (b) Differences in total PSD across the six channels. (c) Differences in total PSD between the three frequency bands. The vertical lines denote 95% confidence intervals. 288

Figure 45. (a) Differences in total PSD between the six channels and the condition. (b) Differences in the PSD between alpha, beta, and gamma bands in the bed and pod conditions. The vertical lines denote 95% confidence intervals. 289

Figure 46. Sliding window with length 60 s and step size of 10 s on the EEG signals of 17 subjects. The plot represents three time intervals and two EEG channels (AF7 and AF8). Each column represents specific frequency band (left to right: alpha, beta and gamma). The shaded area depicts the 95% confidence band. 291

Abbreviation	**Meaning**
AASM	American Academy of Sleep Medicine
ANS	Autonomic Nervous System
BCI	Brain-Computer Interfaces
BP	Blood Pressure
cBEGAN	Conditional Boundary Equilibrium GAN
cDCGAN	Conditional Deep Convolutional Generative Adversarial Network
CNN	Convolutional Neural Networks
CSF	Cerebrospinal Fluid
CSF	Cerebrospinal Fluid
CSP	Common Spatial Pattern
DA	Data Augmentation
DADA	Deep Adversarial Data Augmentation
DL	Deep-Learning
ED	Euclidean Distance
EEG	Electroencephalography
EMG	Electromyograms
ERS	Event-Related Synchronization
FBCSP	Filter-Bank Common Spatial-Patterns
FFT	Fast Fourier Transform
FID	Frechest Inception Distance
fMRI	Functional Magnetic Resonance Imaging
FN	False Negative

FP	False Positive
GAN	Generative Adversarial Network
HCI	Human-Computer Interaction
HF	High Frequency
HRV	Heart Rate Variability
ICU	Intensive Care Unit
k-NN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LE	Laplacian Eigenmaps
LF	Low Frequency
LLE	Locally Linear Embedding
LPD	Lateralized Periodic Discharges
LRDA	Lateralized Rhythmic Delta Activity
LSTM	Long Short-Term Memory
MDS	Multi-Dimensional Scaling
ME	Motor Execution
MI	Motor Imagery
MLP	Multi-Layer Perceptron
MW	Mental Workload
NN	Neural Networks
PCA	Principal Component Analysis
PNS	Parasympathetic Nervous System
PSD	Power Spectral Density

PSD	Power Spectral Density
PSG	Polysomnography
REM	Rapid Eye Movements
REST	Restricted Environmental Stimulation Technique
RMSSD	Root Mean Square of Successive Differences
RNN	Recurrent Neural Network
SAE	Stacked Auto Encoders
SMR	Sensorimotor Rhythms
SNR	Signal-To-Noise Ratio
SNS	Sympathetic Nervous System
SSVEP	Steady State Visual Evoked Potential
STD	Standard Deviation
SVM	Support Vector Machine
t-SNE	T-Distributed Stochastic Neighbor Embedding
TN	True Negative
TP	True Positive
VAE	Variational Auto-Encoder
VLF	Very Low Frequency
WAD	Whiplash Associated Disorders
WD	Wasserstein Distance
WGAN	Wasserstein Generative Adversarial Network

1 Introduction

Neural decoding is an important tool for understanding how neural activity relates to the outside world and for engineering applications such as BCI. Advances in BCI technology in the past decade have led to exciting developments and made BCI a key research area in applied neuroscience and neuro-engineering. Non-invasive BCI facilitates new methods of neurorehabilitation for physically disabled people (e.g., paralyzed patients and amputees) and patients with brain injuries (e.g., stroke patients). BCI systems utilize recorded brain activity to directly communicate between the brain and computers to control the environment in a manner compatible with the individual's intentions. However, the ability to decode intentions is also an important tool for basic neuroscientific research. And, more specifically, decoding intentions in real time would open the door to interesting experimental possibilities, such as interventions to facilitate or frustrate intentions and intention-contingent stimulation. Technological advances of recent decades—such as untethered, wireless recording, machine-learning-based analysis, and real-time analysis of raw EEG signals—have increased the interest in BCI based on electroencephalography (EEG). EEG has proved to be the most popular brain-imaging method for BCI because it is inexpensive, noninvasive, directly measures neural activity (as opposed to Functional magnetic resonance imaging (fMRI) for example) and can facilitate portability to clinical use. EEG signals thus serve as pathways from the brain to various external devices, resulting in brain-controlled assistive devices for disabled people and brain-controlled rehabilitation devices for patients with strokes and other neurological deficits. One of the most challenging topics in BCI is finding and analyzing the relations between recorded brain activity and underlying models of the human body, of biomechanics, and of cognitive processing.

The investigation of relations between EEG signals and upper limb movement has gained more attention in recent years. Think back to this morning: turning off the alarm, getting dressed, brushing your teeth, making coffee, drinking coffee, and locking the door as you left for work. Now imagine doing all those things again, without the use of your hands. Patients who have lost hand function due to amputation or neurological disabilities wake up to this reality every day. Restoring a patient's ability to perform these basic activities of daily life with a (BCI) prosthetic device would greatly increase their independence and quality of life. Currently, there are no realistic, affordable, or low-risk options for neurologically disabled patients to directly control external prosthetics with their brain activity. Better understanding the relations between EEG signals and hand movements is critical to developing a BCI device that would give patients with neurological disabilities the ability to move through the world with greater autonomy. The neuromuscular activations associated with hand movement can be noninvasively recorded by surface electromyograms (sEMG). Therefore, decoding and finding optimal feature vectors therefore plays an important role in EMG classification and hand movement.

In Chapter 2, I used the WAY_EEG_GAL open public dataset, which is freely available and commonly used to test techniques for decoding during a reach-grasp-lift task. In particular, our aim was to decode the weight of an object (165, 330, or 660 g) from the time-domain EMG data of twelve subjects, who reached for and lifted the object. A key objective of our study was to compare different—linear and nonlinear—dimensionality reduction techniques and different classification techniques over the EMG data. In addition, previous work on the WAY_EEG_GAL dataset included either EEG alone or EEG together with EMG, whereas we wanted to investigate to what extent we could classify the weights in this reach-grasp-lift task using EMG alone. Although deep learning (DL) does enable automatic end-to-end learning of preprocessing, feature extraction, and classification modules, DL models are also typically complex—i.e., have many free parameters (or degrees of freedom) to fit—and therefore require large amounts of data to overcome the risk of overfitting those models to specific quirks of the training set. They thus limit the generalizability of the model to an independent test set (although data augmentation might ameliorate these issues). By directly manipulating EMG signals, our study therefore shifts the focus from manual (human-based) feature engineering to completely automated feature-learning even when only few training samples are available. To better understand the suitability of this method for real-time decoding—for example to control a powered prosthesis—we also tracked the evolution of the classification accuracy over time.

Chapter 3, unfortunately, high-grade data collection requires relatively expensive hardware and a lot of participant time. At the same time, access to large, especially clinical, dataset is often limited by privacy and proprietaries concerns. Therefore, large, openly available datasets are uncommon. One general challenge of physiological and neural signal decoding, especially with DL models, is obtaining enough data to train the numerous parameters in these large statistical models. But participants are easily fatigued and thus cannot produce a large amount of data in each experimental session. Bringing participants in for multiple sessions runs into issues of participant attrition for example. We addressed the above issue by carrying out a systematic review on data augmentation (DA) for deep-learning-based electroencephalography. Applying DL to EEG has shown great promise in processing these complex signals due to its capacity to learn good feature representations from raw data through successive non-linear transformations. DA comprises the generation of new samples to augment an existing dataset by transforming existing samples in a manner that increases the accuracy and stability of the classification. This review strived to identify trends and highlight available approaches in DA for DL in EEG to address the following critical questions: (1) What DA approaches exist for EEG? (2) Which dataset and EEG classification tasks have been explored with DA? (3) Are there specific DA methods suitable for specific tasks measured by EEG? (4) Which of the input features in EEG are used for training the deep neural networks with DA?

In Chapter 4, after getting insight on DA for DL-based EEG, I focused on decoding activity for the purpose of decision-making processes that lead to voluntary movement. The prediction in movement onset is promising in EEG-based BCIs, particularly motor-imagery (MI), which have the potential to become groundbreaking technologies in both clinical and entertainment settings. We proposed an end-to-end CNN-based neural network with an attentional mechanism together with different DA techniques on two benchmark MI datasets. In addition, I collected a new dataset, recorded using high-density EEG, and containing both MI and motor execution (ME) tasks to investigate various aspects of EEG decoding critical for neuroscience and BCI, such as finding optimal channel configurations and the best DA techniques, as well as combining data across participants and the role of transfer learning.

Another potential use of EEG is measuring brain activity of people who are floating in body-temperature water in a sensory-deprivation-tank float pod. The literature suggests that floatation, or Restricted Environmental Stimulation Technique (REST) tanks may increase originality, imagination, intuition, and creativity as well as reduce stress. However, proper measurement of physiological and neural signals during floatation, which may facilitate more objective tests of claims of the benefits of floatation, has been lacking. Hence, Chapter 5 offers a systematic review on the available studies on float pods to obtain some insight about recent trends and application as well as to better understand the potential of this possible therapy. What is more, we have been developing physiological and neural signals during floatation, measuring EEG and electrocardiography (ECG) data during REST. Therefore, in Chapter 6, we analyzed the ECG and EEG in the float-pod, and we compare the results with a control condition, lying in bed. We found some differences between the conditions, especially in the gamma band. However, more research is required to understand what these findings mean for claims about reduced stress levels in the float pod.

2 Dimensionality Reduction for Classification of Object Weight from Electromyography

2.1 Introduction

The neuromuscular activations associated with the contraction potentials of the skeletal muscles generate electrical fields that can be noninvasively recorded and are termed surface electromyograms (EMG) [1]. EMG signals are non-stationary in nature and are affected by the structural and functional characteristics of muscles [2]. They have been widely used in various research, industrial, and clinical settings [3, 4]. Potential applications for signal classification and surface EMG include control of robotic arms and fingers, electric wheelchairs, multifunction prostheses and in particular neural prostheses, virtual keyboard and mouse, navigation in virtual worlds, and more [3].

Among the above basic human-control tasks, reaching and grasping are ubiquitous in everyday life and also serve as human interfaces for controlling robotic systems [5-7]. Identification of hand movements from EMG measurements has been used in video games, robotic exoskeleton devices, power prostheses and more [8-11]. A large number of these studies focus on feature selection for EMG movement classification and include a dimensionality-reduction step followed by machine-learning-based classification.

These studies have suggested that successful classification and pattern recognition of EMG signals require three main steps in the following order: (i) data preprocessing, (ii) feature extraction, and (iii) classification. Common EMG data preprcoessing steps include low- and high-pass filtering, whereas feature extraction is a method of finding intrinsic and meaningful information that may be latent in the EMG signal [12, 13]. Over the past few decades, various manual EMG feature-extraction methods have explored in the time and/or frequency domains [14]. Finding optimal feature vectors therefore plays an important role in EMG classification because appropriate feature extraction tends to result in considerably high classification accuracy [12, 15].

A common method to extract features from signals is dimensionality reduction, or learning low-dimensional embeddings from samples in high dimensional space [16-18]. Most dimensionality-reduction techniques are linear and relate to Principal Component Analysis (PCA) [19] or Multi-Dimensional Scaling (MDS) [20]. While applying PCA may result in a lower-dimensional representation that captures more relevant information, such linear techniques have various limitations when applied to EMG. They are often less reliable and more sensitive to the number of samples in the training set. In addition, linear techniques by nature model linear relations, which may not describe EMG signals well. Last, linear techniques are global by nature, which means that they cannot preserve local structures in the original feature space [21].

More modern, non-linear dimensionality-reduction techniques include Locally Linear Embedding (LLE) [22]. The LLE algorithm computes the basis of a low-dimensional space, though the dimensionality of the embedding often needs be given as a parameter [23]. Moreover, the output is an embedding for the specific given dataset and not a general mapping from the original to the lower-dimensional space. LLE is also not isometric and often fails by mapping distant points close to each other. Another non-linear technique, ISOMAP, is an extension of MDS that uses geodesic instead of Euclidean distances and can therefore be applied to non-linear manifolds [24]. The geodesic distances between points are approximated by graph distances. Then, MDS is applied on the geodesic distances to compute an embedding that strives to preserve distance between points.

Here we used the Laplacian Eigenmaps algorithm [25]. It computes the normalized graph Laplacian of the adjacency graph of the input data, which is an approximation of the Laplace-Beltrami operator on the manifold. It exploits locality-preserving properties that were first observed in clustering. The Laplacian Eigenmaps algorithm can be viewed as a generalization of LLE, as the two are identical when the weights of the graph are chosen according to the criteria of the latter. Much like LLE, the dimensionality of the manifold also needs to be provided; the computed embeddings are not isometric, and a general mapping between the two spaces is not output. In the past, EMG-based classification using non-linear dimensionality reduction techniques was more often applied to human gait [26, 27] than to the more complex reach-and-grasp movements, which also utilize more degrees of freedom [21].

In this study, we used the WAY_EEG_GAL open public dataset, which is freely available (see Materials and Methods) and commonly used to test techniques for decoding during a reach-grasp-lift task. In particular, our aim was to decode the weight of an object (165, 330, or 660 g) from the time-domain EMG data of twelve subjects, who reached and lifed the object. After preprocessing, we automatically extracted the features, reduced the dimensionality, and fed the resulting data into a machine-learning classifier. A key objective of our study was to compare different—linear and nonlinear—dimensionality reduction techniques and different classification techniques over the EMG data. In addition, previous work on the WAY_EEG_GAL dataset included either EEG alone or EEG together with EMG, whereas we wanted to investigate to what extent we could classify the weights in this reach-grasp-lift task using EMG alone. However, we did not focus our study on deep-learning (DL) techniques. Although DL does enable automatic end-to-end learning of preprocessing, feature extraction, and classification modules, DL models are also typically complex—i.e., have many free parameters (or degrees of freedom) to fit—and therefore require large amounts of data to overcome the risk of overfitting those models to specific quirks of the training set. Therfore, they limit the generalizability of the model to an independent test set (although data augmentation might emiliorate these issues) [28].

By directly manipulating EMG signals, our study therefore shifts the focus from manual (human-based) feature engineering to completely automated feature-learning even when only few training samples are available [29-31].

2.2 Materials and Method

Our methodology for EMG signal classification is illustrated in Fig 1 and detailed below. Briefly, EMG signals were first preprocessed and segmented into the first 8 s of each trial before feature extraction. This segmentation ensured that the subject started from home position and returned to the home position, removing noise after returning to the home position (Fig 2). The components corresponding to the highest eigenvalues from the output of the dimensionality-reduction algorithms were extracted as the dominant features. Thereafter, these intrinsic features were used for classification.

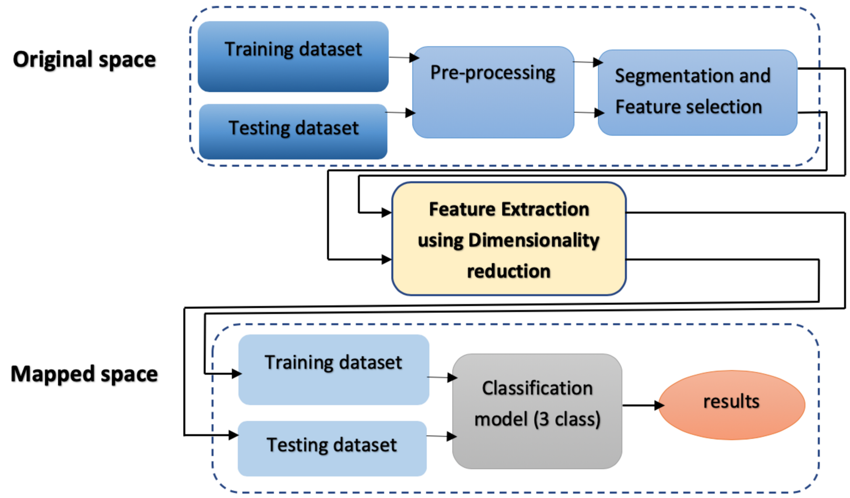

Figure 1. Processing pipeline for EMG signal classification

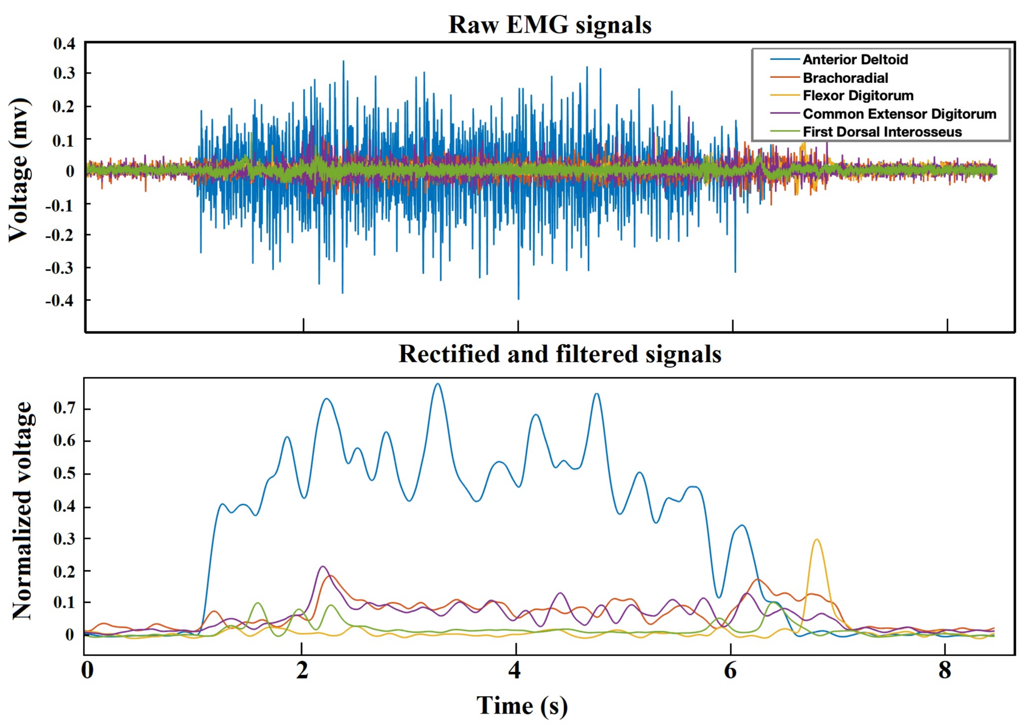

Figure 2. EMG preprocessing. (Top) Pre-processed EMG signals of 5 muscles (Anterior Deltoid, Brachoradial, Flexor Digitorum, Common Extensor Digitorum, and First Dorsal Interosseus). (Bottom) Rectified and filtered EMG signals by band-pass Butterworth filter (4th order) in the 5-450 Hz range on the full wave rectified and normalized signals from each muscle.

2.2.1 Dataset

The WAY_EEG_GAL dataset is freely available and has become somewhat of a benchmark to test techniques that decode sensation, intention, and action from surface EMG and scalp EEG in humans performing a reach-grasp-lift task (https://doi.org/10.6084/m9.figshare.c.988376) [32]. Here we focus exclusively on EMG data. The EMG signals were sampled at 4 kHz. In each trial, the participants rested their hand in the home position. Then they were cued to reach for the object, grasp it with the thumb and index finger, lift it straight up in the air and hold it for a few of seconds. They were then instructed to put the object back on the support surface, let go of it, and return the hand to a designated home position [32]. The state of the LED indicated to the participant to start and terminate a trial. The object's weight varied between 165, 330, and 660 g and the surface material varied between sandpaper, suede, or silk. We used all available 2,645 trials of EMG signals, across all 12 subjects, including trials with different weights (840 trials for 165 g, 1122 trials for 330 g, and 683 trials for 660 g). The number of trials for each subject was 220 or 221, and the highest imbalance-ratio between classes for any subject was 0.61 (Table 1). The material in all trials was always sandpaper, as per the original design of the experiment [32]. Five EMG electrodes recorded the activity from 5 muscles (Figs. 2, 3).

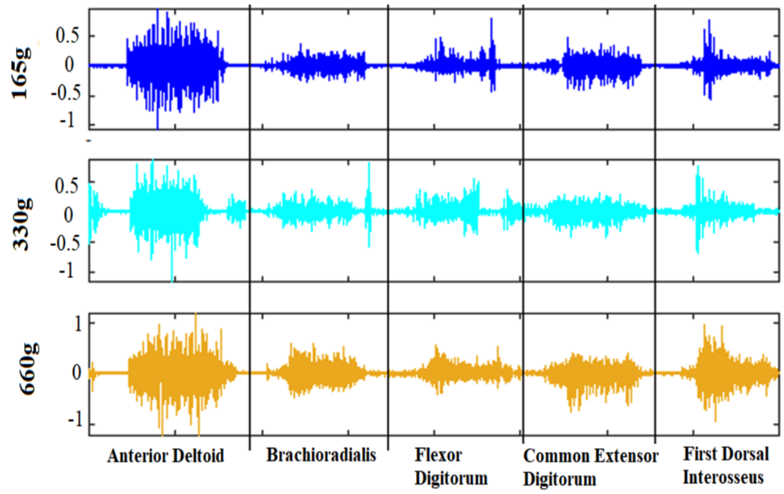

Figure 3. Raw EMG signals of 5 muscles (Anterior Deltoid, Brachioradialis, Flexor Digitorum, Common Extensor Digitorum, and First Dorsal Interosseous) for 3 different weights (165, 330, and 660 g)

Table 1. The number of trials for each class and each subject

ID	**165g**	**330g**	**660g**	**imbalance-ratio**	**Total trials for each subject**
Subject 1	70	93	57	0.61	220
Subject 2	70	94	57	0.61	221
Subject 3	70	93	57	0.61	220
Subject 4	70	94	57	0.61	221
Subject 5	70	94	57	0.61	221
Subject 6	70	93	56	0.60	219
Subject 7	70	94	57	0.61	221
Subject 8	70	93	57	0.61	220
Subject 9	70	93	57	0.61	220
Subject 10	70	94	57	0.61	221
Subject 11	70	94	57	0.61	221
Subject 12	70	93	57	0.61	220
Total	840	1122	683	Mean: 0.61	2645

2.2.2 Pre-processing

All processing was carried out on a PC (3.4 GHz Intel® CoreTM i7-6700 CPU) using Python 3 and MATLAB 2019b.

EMG signals are typically contaminated by various types of noise and artifacts. Therefore, preprocessing prior to feature extraction was important. We used a band-pass Butterworth filter (4th order) in the 5-450 Hz range on the full-wave rectified and normalized signals from each muscle (Fig 2).

2.2.3 Segmentation and feature selection

The time required to reach, grasp, and lift varied among trials and subjects. So, we focused on the first 8 seconds for every trial. Doing so also removed noise that appeared at the end of the trial, after the subject returned their hand to the home position. For feature selection, we concatenated the signals of the 5 muscles (as in Fig. 3). We then subsampled, taking every 5th sample for increased processing speed (lowpass filtering was already carried out before the subsampling, as part of the band-pass filter during preprocessing). We ended up with 5 x 8 x 800 (muscle x time (second) x samples) = 32,000 features.

2.2.4 Feature Extraction using dimensionality reduction

EMG signals are complex, high-dimensional, and non-linear and hence hard to study in their original form. Effort has therefore been put into finding meaningful, low-dimensional features of these signals. Classical dimensionality-reduction techniques include linear methods, such as principal component analysis (PCA) [33] and linear discriminant analysis (LDA) [34]. These techniques preserve global structure of the data but at the cost of obscuring local features and preventing any local manipulation of the data.

In contrast, manifold learning is a non-linear technique for recovering a low-dimensional representation from high-dimensional data [23, 35]. The literature on manifold learning is dominated by spectral methods. These have a characteristic computational pattern. The first step involves the computation of the k-nearest neighbors (k-NN) of all N data points. Then, an N×N square matrix is populated using some geometric principle. This characterizes the nature of the desired low-dimensional embedding. The eigenvalue decomposition of this matrix is then used to obtain the low-dimensional representation of the manifold.

A trade-off between preserving local and global structures must often be made when inferring the low-dimensional representation. Manifold learning techniques such as Locally Linear Embedding (LLE) [22], Laplacian Eigenmaps [25], and t-Distributed Stochastic Neighbor Embedding (t-SNE) [36] are considered to be local methods because they are designed to minimize some form of local distortion and hence result in an embedding that preserves locality. Methods such as ISOMAP [23] are considered global because they preserve all geodesic distances in the low-dimensional embedding. All spectral techniques are parameter less (except for neighborhood size; see below) and hence do not characterize the map that generates them. In this study, we compared different algorithms for manifold learning—Global: ISOMAP; and local: LLE, t-SNE, Laplacian Eigenmaps—and further compared them with linear dimensionality-reduction techniques, PCA and LDA (the latter is the only supervised dimensionality-reduction technique). In the next section, we explain the Laplacian Eigenmaps algorithm in more details.

2.2.5 The Laplacian Eigenmap algorithm

The Laplacian Eigenmap algorithm plays a larger role in this study, hence we describe it in more detail, following Belkin et al. (see [25]). Given k points $x_1, \ldots, x_k$ in R^l, it finds a set of points $y_1, \ldots, y_k$ in $\mathbb{R}^m$ ($m \ll l$) such that y_i represents x_i. Therefore, $x_1, \ldots, x_k \in M$ and M is a manifold embedded in $\mathbb{R}^l$. The Laplacian Eigenmaps (spectral embedding) is based on the following steps:

Algorithm 1. Laplacian Eigenmaps

Input: High-dimensional data-points of the manifold:

$$\{x_i \in \mathbb{R}^l\}, \qquad i = 1, 2, \ldots, k$$

Output: Low-dimensional embeddings of data points:

$$\{y_i \in \mathbb{R}^m\}, \qquad m \ll l, \qquad i = 1, 2, \ldots, k$$

Step1. Constructing the graph:

We put an edge between nodes i and j if x_i and x_j are n-nearest neighbors. Thus, nodes i and j are connected by an edge if i is among the n-nearest neighbors of j, or j is among n-nearest neighbors of i. This then leaves us with a connected graph.

Step 2. Choosing the weights. There are two possible ways for choosing the weights:

a. Heat Kernel: $W_{ji} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$, if vertices i and j are connected by an edge; and $W_{ji} = 0$, if vertices i and j are not connected by an edge. The only parameter in the Heat-Kernel equation is σ, which defines the extent to which distant neighbors influence the embedding of each point. The choice of parameter σ is data-dependent and is typically tuned empirically.

b. Simple-Minded: $Wij = 1,$ if vertices i and j are connected by an edge and $W_{ji} = 0$ if vertices i and j are not connected by an edge.

Step 3. Eigenmaps: Compute eigenvalues (λ) and eigenvectors (f) for the generalized eigenvector problem:

$Lf = \lambda Df,$

where D is a diagonal weight matrix, and its elements are column (or row, since W is symmetric) sums of W.

$D_{ii} = \sum_j W_{ji}$, $L = D - W$ is the Laplacian matrix (symmetric, positive semidefinite).

We leave out the eigenvector corresponding to eigenvalue 0 and use the next m eigenvectors for embedding in m-dimensional Euclidean space: $x_i \to f_1(i), \ldots, f_m(i)$. The m eigenvectors will be considered features of the dataset.

The core algorithm is relatively simple. It has a few local computations (in the matrix) and one solution to the sparse eigenvalue problem. The solution reflects the intrinsic geometric structure of the manifold. It requires a search for neighboring points in a high-dimensional space. The justification for the algorithm comes from the role of the Laplace Beltrami operator in providing an optimal embedding for the manifold. The manifold is approximated by the adjacency graph computed from the data points. The Laplace Beltrami operator is approximated by the weighted Laplacian of the adjacency graph, with weights chosen appropriately. The key role of the Laplace Beltrami operator in the heat equation enables us to use the heat kernel to choose the weight decay function in a principled manner. Thus, the embedding maps for the data approximate the eigenmaps of the Laplace Beltrami operator, which are maps intrinsically defined on the entire manifold. For more information about the justification for Laplacian algorithm and the role of the Laplace Beltrami operator in providing an optimal embedding, see Supplementary Methods. The low dimensional representation of the data set that optimally preserves local neighborhood-information may be viewed as a discrete approximation to a continuous map that naturally arises from the geometry of the manifold. It is worth highlighting some aspects of Laplacian Eigenmaps here: 1) The algorithm reflects the intrinsic geometric structure of the manifold, which is simple with few local computations and one sparse eigenvalue problem. 2) The justification for the algorithm comes from the role of the Laplace-Beltrami operator in providing an optimal embedding for the manifold. The key role of the Laplace-Beltrami operator in the heat equation that enables us to use the heat kernel is to choose the weight decay function in a principled manner. Thus, the embedding maps for the data approximate the Eigenmaps of the Laplace-Beltrami operator, which are maps that intrinsically depend on the entire manifold. 3) The locality preserving character of the Laplacian Eigenmap algorithm makes it relatively insensitive to outliers and noise. Close connections to spectral clustering algorithms were developed in machine learning and computer vision. To help gain intuition about manifold-learning algorithms, we demonstrate their use on a simple, spherical dataset (2000 random points on the surface of a 3D sphere) and on a "Swiss roll" (The 2000 points chosen at random from the Swiss roll; Fig 4). We used the Scikit-learn Python package [37] and Matlab toolbox [38, 39] for dimensionality reduction. Laplacian Eigenmaps are termed Spectral Embedding (SE) in Scikit-learn and the embedding is not strictly the adjacency matrix of a graph but more generally an affinity or similarity matrix between samples [40]. It has 2 different methods (heat kernel and simple-minded) for constructing the weight matrix. The kernel function for the Heat-Kernel ($W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$) in this package is a Gaussian radial basis function kernel (RBF) with $\gamma = \frac{1}{2\sigma^2}$, where γ is a parameter that sets the "spread" of the kernel. The results of various manifold-learning techniques for 8 neighbors in 2D space are shown in Fig 4. Laplacian Eigenmaps (simple-minded), or SE, is the fastest algorithm; the computation time for SE-rbf is 5.5 times longer for a sphere and 31.7 times longer for a Swiss roll. It appears that the construction of the weight matrix drives this difference in computation time. For more information about the properties of techniques for dimensionality reduction, see Supplementary Methods.

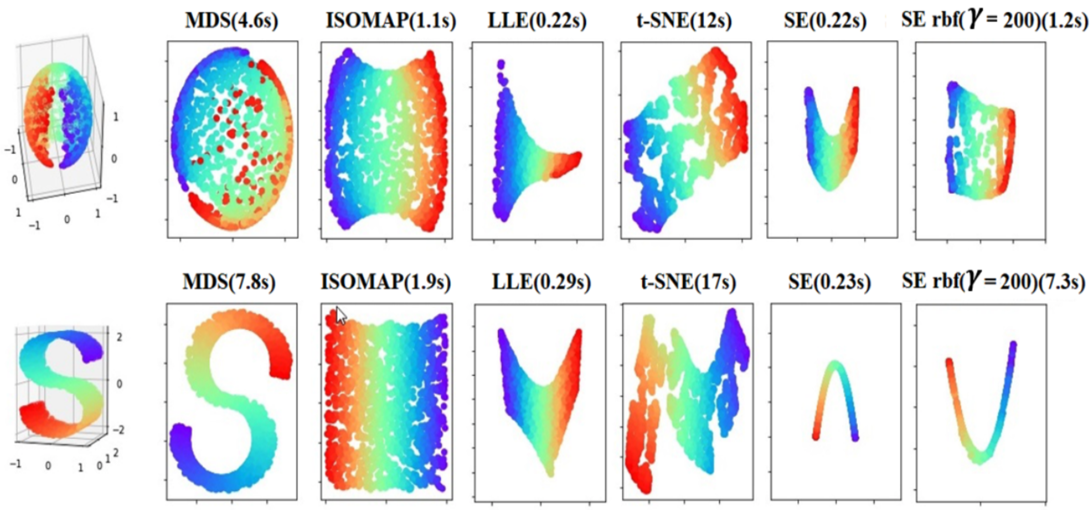

Figure 4. Manifold learning techniques. MDS, ISOMAP, LLE, t-SNE, and Spectral embedding (SE) or Laplacian Eigenmaps on 2000 points randomly distributed on the surface of a sphere. Computation time in seconds is given after each method's name in parentheses. The first column for SE is simple-minded constructing weight matrix and the last column is for heat kernel.

An important requirement for dimensionality reduction techniques is the ability to embed new high-dimensional datapoints into an existing low-dimensional data representation. However, there is no explicit projection function between the original data and their low dimensional representations in the original LE algorithm, which makes out-of-sample extension difficult. To find projection of any additional samples, LE needs to be run on all the data together with the additional samples, resulting in considerable computational cost, especially when applying it to large scale data pattern recognition. Fortunately, various methods have been developed to mitigate the out-of-sample problem [5]. Nyström approximation supports out-of-sample extensions for spectral techniques such as ISOMAP, LLE, and Laplacian Eigenmaps. In Supplementary Methods, we explain Nyström approximation in greater details. In Fig 5 we depict the embedding of additional, out-of-sample points (there termed "test dataset"). As is apparent, the out-of-sample points are mapped to plausible locations in the low-dimensional space. For more information about the out-of-sample extension, see Supplementary Methods.

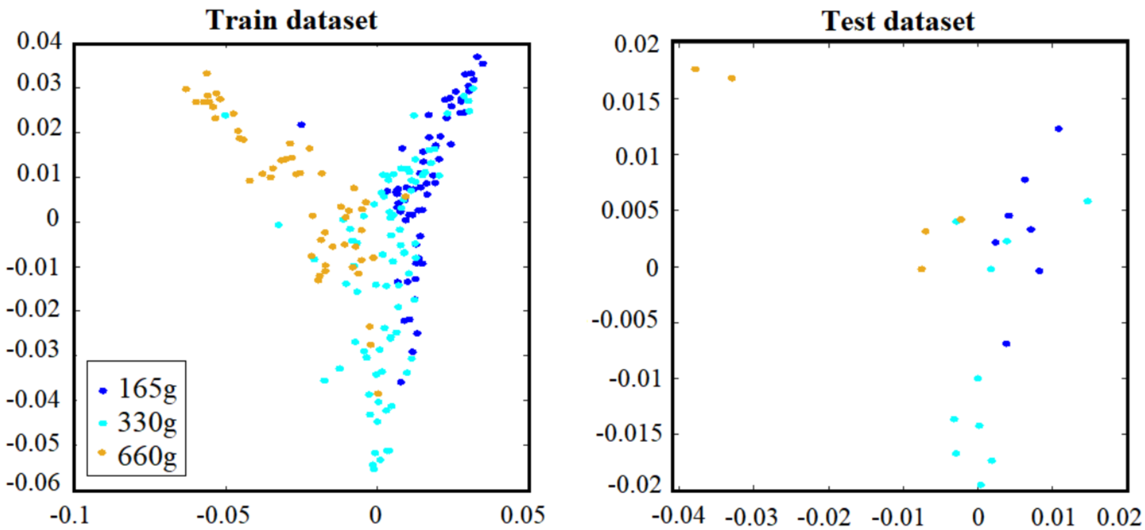

Figure 5. Example of an out-of-sample extension via the Nyström approximation in embedded space by Laplacian Eigenmaps with k = 8 and $\sigma = 1$. The train / test ratio was 90% / 10%.

2.2.6 Classification

After finding an optimal feature set, we tested commonly used classification algorithms: k-NN [41], linear and RBF SVM [30] (C=32, $\gamma = 0.01$ for RBF SVM), and Random Forest [42]. We evaluated the performance of each classifier on the data after running the above dimensionality-reduction techniques. We ran the analysis on each subject separately. The dataset was divided into disjoint training and testing sets, which consisted of 90% and 10% of the total trials, respectively. For each subject, we further ran 10-fold cross-validation on the training dataset. Table 1 shows the details of trials for each class and subject. The results we report are therefore averaged over all subjects on the testing dataset.

Another common method for EMG classification, on top of those above, is deep learning [43]. We tested several deep learning architectures on our selected feature set. However, we ran into severe overfitting issues resulting in accuracies very close to chance level. This result is likely due to the relatively small features-to-samples ratio for our dataset. Consequently, we did not include deep-learning results in our analyses.

As the dataset was imbalanced (the highest imbalanced ratio was 0.6), we used the F1-score as a metric of accuracy [44]. The F1-score provides a way to combine both precision and recall into a single measure that captures both properties. Once precision and recall have been calculated for a binary or multiclass classification problem, the two scores can be combined into the calculation of the F1-score.

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

where $Precision = \frac{TP}{(TP+FP)}$ and $Recall = \frac{TP}{(TP+FN)}$. Here TP is number of true positives, FP is number of false positives, and FN is number of false negatives. This is the harmonic mean of the two fractions. The F1-Score is a very common metric for imbalanced classification problems [45].

2.2.7 Running-window analysis

In this section we describe an additional, running-window analysis that we performed on this dataset. This section helps to get more insight into the temporal dynamics of the current model's classification accuracy. In the analysis, we used a 100 ms sliding window with a step size of 40 ms. We tested various step sizes (between 10 and 50 ms) and 40 ms resulted in the best visualization (though the visual differences between the step sizes were minute).

We applied the proposed pipeline in a sliding-window manner to estimate the extent to which the prediction accuracy would be stable over consecutive time windows. Laplacian Eigenmaps (simple-minded k=8), 120 dimensions after embedding with k-NN (number of neighbors k=8), was applied on the preprocessed EMG signal. The minimum length of the sliding window that is possible to run on this dataset is 100 ms. Below this length, the adjacency graph of the input data appears not to be fully connected (Supplementary Methods). For example, a row (and a column, since this is symmetric) can be all zeros and therefore one of the nodes will not be connected, resulting in a warning.

We therefore segmented the dataset based on the events (onset of touching the object, LED on, LED off). There was variability among subjects' speed in this task. The minimum time across subjects was 0.74 s before touching the object and 1.82 s after touching the object.

2.3 Results

2.3.1 Parameter settings

The proposed framework has 2 parameters: the number of nearest neighbors in Laplacian Eigenmaps to construct the Laplacian matrix (either using the direct number of neighbors, k, or using a heat kernel approach, $\boldsymbol{\sigma}$) and the number of eigenvectors used for data mapping, i.e. the dimensions of the mapped space. The number of nearest neighbors in Laplacian eigenmaps, k, was tuned to 4, 5, 6, …, 20—i.e., using a grid search. We also tested values of $\boldsymbol{\sigma}$ in the range 0.1, 1, 10, 100,1000, again using a grid search. Figure. 6 shows the effect of different k and $\boldsymbol{\sigma}$ values on the training dataset for Subject1. The number of eigenvectors or dimensions is tuned in the range of 1,5, 10, …, length (training trial), once more using a grid search. (see S1 Table 1. Properties of techniques for dimensionality reduction)

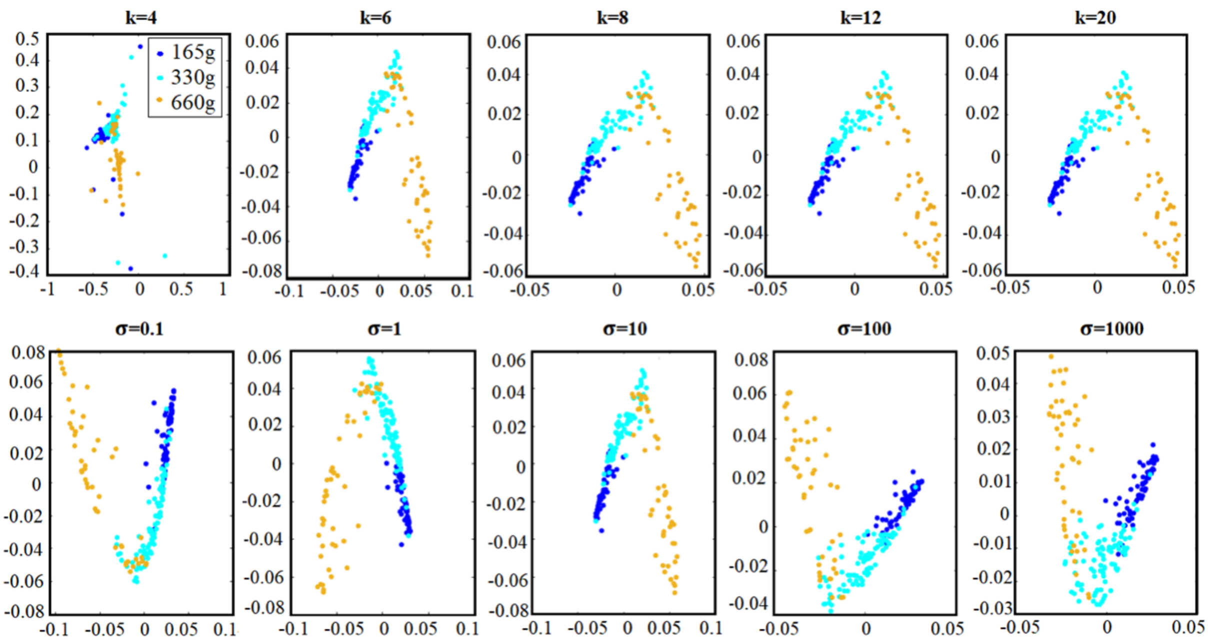

Figure 6. Visualization of the effect of k (top row) and σ (bottom row) on the training dataset for Subject 1 in the embedded space

Table 2 shows the optimal number of eigenvectors for each dimensionality-reduction method across all subjects. As a sanity check, we also used the maximum likelihood estimator (MLE) as the intrinsic dimensionality estimator in the Matlab toolbox for dimensionality reduction [39]. The number of eigenvectors varied between 110 and 170 over the 12 subjects for different dimensionality-reduction techniques. We also visualized the embedded EMG using our six dimensionality-redution methods (PCA, LDA, ISOMAP, LLE, Laplacian Eigenmaps, and t-SNE). Figure 7 shows the 2 most prominent components for each of these methods.

Table 2. The number of dimensions that leads to the highest F1-score (obtained using grid search; for k-NN, which was the best classifier; see also Figure 8) vs. different dimensionality reduction techniques for each subject. The mean and standard error (SE) over all subjects for the different methods are also given.

ID	PCA	ISOMAP	LLE	LE(simple minded)	LE(rbf)	t-SNE
Subject 1	160	180	135	115	120	180
Subject 2	170	155	110	115	110	155
Subject 3	165	175	125	125	125	170
Subject 4	190	180	90	115	90	190
Subject 5	170	180	120	120	105	175
Subject 6	155	130	110	115	145	130
Subject 7	175	120	100	140	120	175
Subject 8	125	145	120	120	125	185
Subject 9	110	100	95	125	95	140
Subject 10	170	160	135	140	155	185
Subject 11	170	185	110	125	105	170
Subject 12	165	190	95	115	110	140
[Min Max] Mean± SE	[110 190] 160.4 ± 6.4	[100 190] 158.3 ± 8.4	[90 135] 112.1 ± 4.4	[115 140] 122.5 ± 2.6	[90 155] 117.9 ± 5.5	[130 190] 166.3 ± 5.8

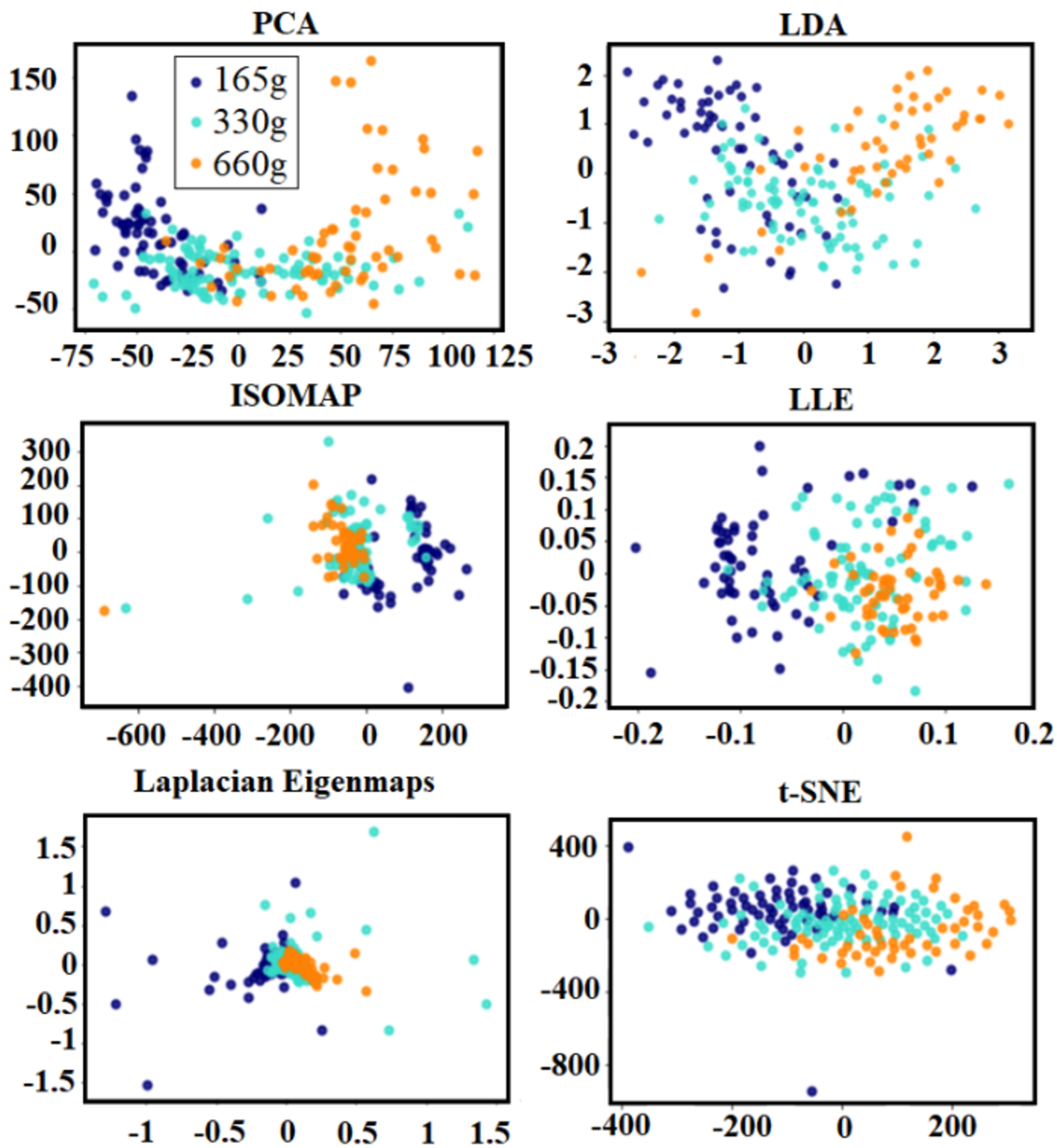

Figure 7. Visualization of the embedding process: EMG visualization using the 2 most prominent components of different dimensionality-reduction technique (The x-axis is the most prominent component, and the y-axis is the second most prominent component)

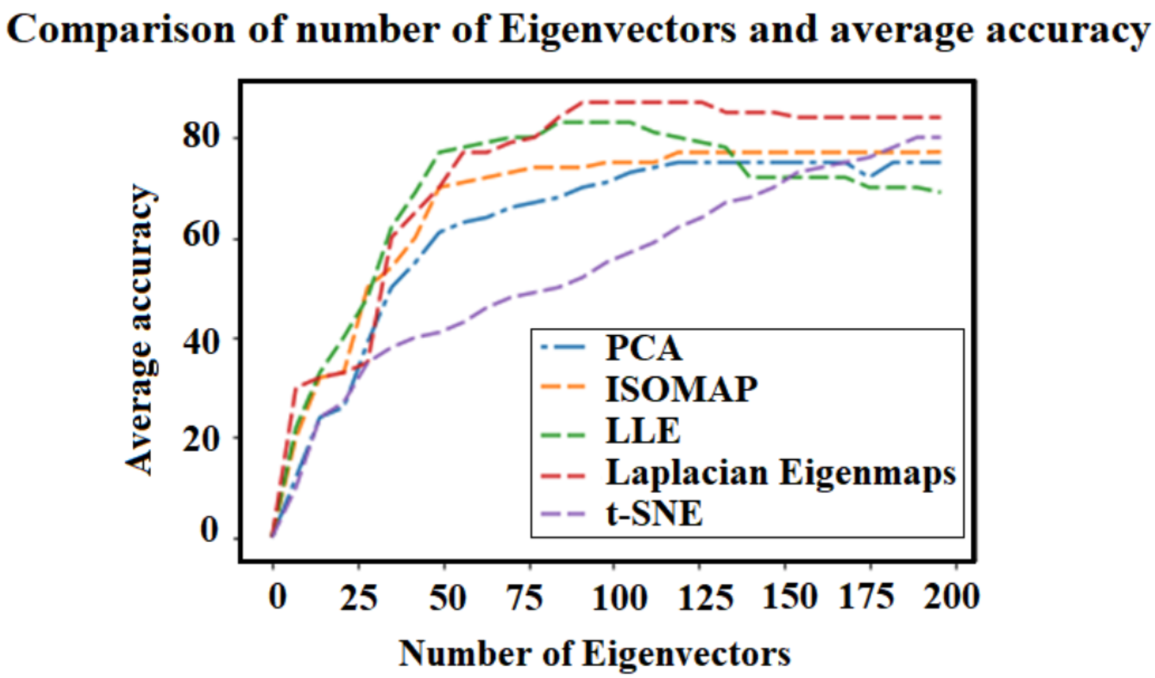

Figure 8. Average accuracy (across all 12 subject) as a function of the number of eigenvectors obtained with different dimensionality reduction techniques for the best classifier—k-NN (see also Table 3). We included PCA, ISOMAP, LLE, Laplacian Eigenmaps, and t-SNE. The LDA dimensionality-reduction technique was not included in these curves because the maximal dimension for LDA is equal to the number of classes minus one.

2.3.2 Classifier Performance

We computed the average accuracy using PCA, ISOMAP, LLE, Laplacian Eigenmaps, and t-SNE for the classifier that produced the highest accuracy—k-NN (Figure. 8). In Table 3, the performance of Laplacian Eigenmaps (simple-minded) and different classifiers vs. different number of neighbors (k) is shown. It demonstrates that k=8 fits well for this dataset.

Table 4 details the prediction accuracies of the different classification algorithms on the test set for the various dimensionality-reduction techniques over all 12 subjects. On average, Laplacian Eigenmaps (especially with a heat kernel) is the algorithm with the highest accuracy across all dimensionality-reduction methods—78.15%. And k-NN is the classification method resulting in the highest mean accuracy across all classification algorithms—80% on average—and significantly higher than linear SVM and RBF-SVM and marginally higher than Random Forest (repeated-measures ANOVA $F(3)=15.5$, $p<0.001$; post-hoc t-tests suggest all comparisons are significant at the 0.05 level except Random Forest vs. k-NN, which was $p=0.053$; and RBF-SVM vs. Random Forest, $p=0.685$—see S1 Table 4 and S1 Table 5; we further found no evidence that the pairwise differences are not normally distributed—Shapiro-Wilk test was not significant). Interestingly, the intersection of Laplacian Eigenmaps and k-NN has the highest overall accuracy, at 88%. We also ran a statistical analysis on the different dimension reduction techniques to compare the linear and non-linear techniques. However, there wasn't significant difference between them, maybe because of the low number of subjects (see S1 Table 2 and S1 Table 3).

2.3.3 Evolution of classification accuracy over time

So far, we focused on optimizing the dimensionality reduction and classification accuracy on the entire movement duration. However, another interesting aspect of this dataset is the evolution of the dimensionality reduction and classification accuracy over time within each trial. A running-window analysis of our best combination of dimensionality reduction technique and classification method (Laplacian Eigenmap and k-NN) suggests that there is little to no information in the EMG of the muscles before the subject touches the object (Figure. 9a). The mean accuracy over our 100 ms window during that time was 43.39% (±9.79). It is not surprising that the accuracy is slightly above chance, as some of the experiment was carried out in a blocked design; hence, the weight often did not change between consecutive trials [46]. So, subjects may therefore have begun preparing their hand posture while reaching for the object based on the weight they anticipated from the previous trial. And we were able to capture this preparatory muscle activity with our algorithm.

Perhaps more interesting is the running-window analysis *after* the subjects grasped the object. With such a small window, we expected a much lower accuracy than that over the entire movement window. Indeed, the mean accuracy was only 57.65% (±11.59). But interestingly, the accuracy was above chance level already in the first 100 ms window (Figure. 9b). And it was generally stable throughout much of the duration when the subject held the object, though there appears to have been a small decrease in accuracy toward the end of that time duration.

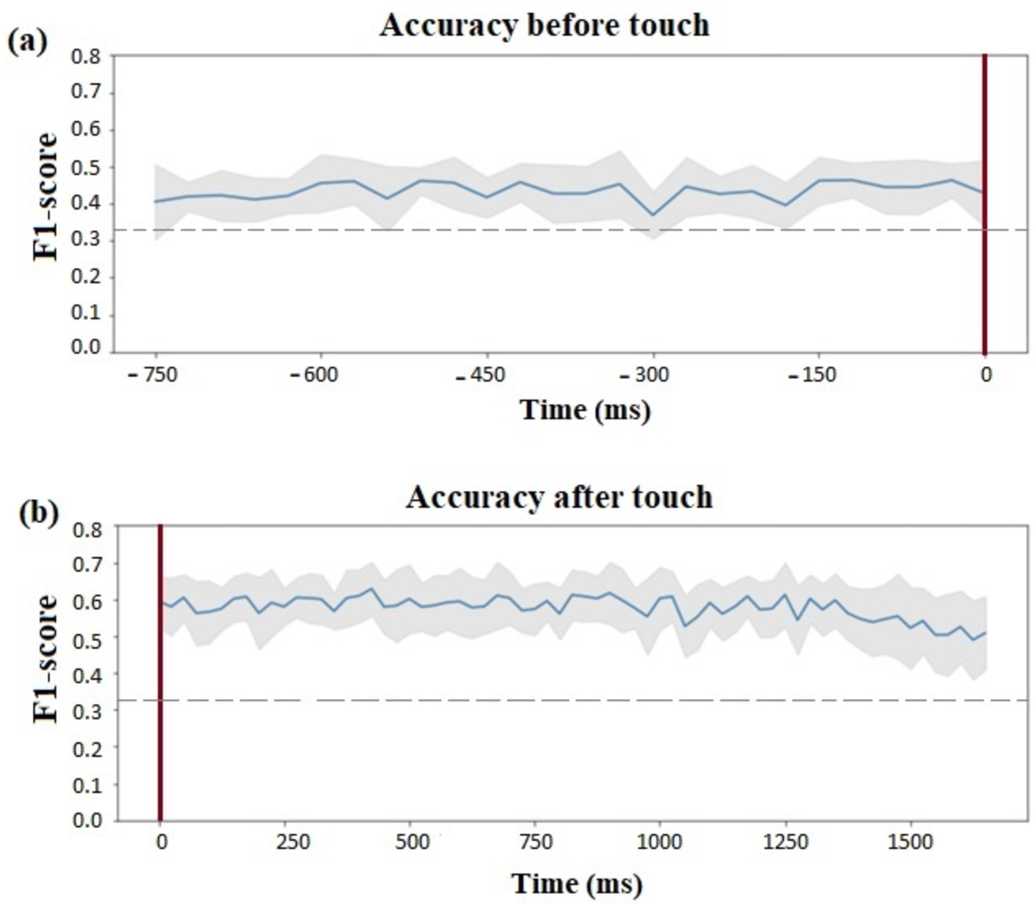

Figure 9. Sliding-window analysis of Laplacian Eigenmap dimensionality reduction and k-NN classification before touching the object (a) and after touching it (b). We used a 100 ms sliding window with a step size of 40 ms. In both panels, the onset of touching the object is designated by a vertical red line at time 0. Table 3. Performance of Laplacian Eigenmaps (simple-minded) and different classifiers vs. different number of neighbors (k) on the EMG signals (F1 score ± SE)

Table 3. Performance of Laplacian Eigenmaps (simple-minded) and different classifiers vs. different number of neighbors (k) on the EMG signals (F1 score $\pm$ SE)

Different number of neighbors(k)	4	6	8	10	12	15	20
Laplacian Eigenmaps (simple-minded) + k-NN	43.42(±9.2)%	71.98(±4.9)%	88.2(±3.5)%	79.32(±6.7)%	71.76(±4.5)%	63.64(±7.8)%	58.31(±6.3)%
Laplacian Eigenmaps (simple-minded) + RBF SVM	32.88(±6.4)%	74.68(±2.7)%	77.6(±2.3)%	76.98(±3.1)%	76.46(±8.5)%	73.52(±4.3)%	68.23(±8.1)%
Laplacian Eigenmaps (simple-minded) + Linear SVM	43.42(±5.3)%	62.98(±3.9)%	63.6(±2.2)%	63.59(±2.7)%	63.06(±1.5)%	63.64(±7.8)%	58.31(±6.2)%
Laplacian Eigenmaps (simple-minded) + Random Forest	59.21(±3.2)%	72.28(±4.3)%	79.5(±2.8)%	78.54(±2.4)%	77.76(±4.5)%	77.64(±3.9)%	68.22(±2.3)%

Table 4. Performance of different classifiers vs. different dimensionality-reduction methods on EMG signals (F1 score $\pm$ SE). See S1 Table 4 for post-hoc t-tests for this table.

	k-NN	RBF SVM	Linear SVM	Random Forest	Average (%)
PCA	75.3(±2.8) %	64.3(±1.2)%	63.5(±4.9)%	75.4(±3.2)%	69.62(±3.3)%
LDA	78.2(±15.3)%	72.7(±13.2)%	67.2(±12.2)%	76.2(±9.3)%	73.57(±2.4)%
ISOMAP	77.4(±7.2)%	74.4(±2.2)%	57.7(±3.9)%	73.9(±4.9)%	70.85(±4.4)%
LLE	84.6(±7.9)%	82.3(±4.1)%	58.5(±4.1)%	76.7(±3.8)%	75.52(±5.9)%
Laplacian Eigenmaps (simple-minded k=8)	88.2(±3.5)%	78.2(±2.3)%	63.6(±2.2)%	72.6(±2.8)%	77.7(±3.8)%
Laplacian Eigenmaps (rbf, $\sigma = 10$)	84.2(±3.9)%	71.2(±5.3)%	61.3(±4.7)%	79.9(±2.9)%	78.15(±3.7)%
t-SNE	75.8(±4.2)%	73.2(±2.1)%	71.1(±8.3)%	69.3(±8.7)%	72.35(±1.3)%
Average ($\pm SE$))%	80.53(±1.9)%	75.24(±2.3)%	65.24(±2.1)%	74.85(±1.2) %	

2.4 Discussion

Our goal in this study was to decode to which of 3 weight classes an object in a reach-grasp-lift task belonged using only EMG data from the arm and hand. In particular, we compared the performances of various linear and non-linear dimensionality-reduction techniques, combined with several classification methods. We worked on pre-processed EMG signals directly, automatically extracting the features for the classification phase. The dimensionality-reduction algorithms we used lowered the dimensionality of our data from 32,000 to less than 200—i.e., more than 160-fold. We then applied various classification techniques on this 3-way classification problem and discovered that the combination of Laplacian Eigenmaps (simple-minded, k=8) with the k-NN classifier resulted in the highest classification accuracy (F1 score 88.2±3.5%). As a result, we used automatic feature-extraction directly from the pre-processed EMG time-domain signal [20, 47-52]. Importantly, our approach to extract features from EMG signal resulted in relatively high decoding accuracy.

Other studies that relied on the same dataset that we used mostly focused on EEG [47-50]. However, Cisotto et al. used both EEG and EMG to classify the same dataset [51]; though they attempted classification of only 2 of the 3 available classes (the most extreme weights: 165 and 660 gr). They also reported their results in terms of accuracy, even though their classes were imbalanced (imbalance ratio of 0.81 between the number of trials in the 2 classes). They reported a maximal accuracy of 94% (using only the Brachoradial muscle). Running our analysis as is (using all muscles and without any parameter optimization) with only the 2 weight classes they used, and a reporting accuracy instead of F1 score, we get an accuracy of $90.9 \pm 2.5\%$. This accuracy is statistically indistinguishable from theirs (t-test: t(11)=-1.33, p=0.21). Therefore, even though we used only EMG and not EEG, and we did not focus our analysis on a binary classification problem, we were able to achieve comparable results. These might be due to the superiority of our method—perhaps our automatic feature extraction or our dimensionality-reduction algorithm. Another, not mutually exclusive, reason might be that the high classification accuracy that Cisotto et al. we were able to achieve owes much to the EMG signals that they used in conjunction with the EEG signals [51]. Hence, at least for this dataset, the addition of the EEG signals may not have added that much to the decoding accuracy.

The increasing adoption of DL tehcniques in machine learning is shifting the focus from feature engineering to feature learning [8, 52]. Nevertheless, the black-box nature of DL makes it hard to understand what information is learned by the network and how it relates to handcrafted features. At the same time, the application of DL on insufficiently large datasets risks overfitting. In additional, the high variability of EMG recordings between participants often makes deep-learned features generalize poorly across subjects.

The range of mean accuracies among the dimensionality reduction algorithms we used was 70-78% (Table 4). Interestingly, Laplacian Eigenmaps not only performed best on average; its simple-minded version also generally required the shortest computing time (see Materials and Methods). The average accuracies of the different classifiers varied from 65.24% to 80.53%. It appears that the linearity of linear-SVM was detrimental for EMG signal decoding, while the most non-linear technique, k-NN, faired best.

It also appears that dimensionality-reduction techniques relying on local embedding were better for this dataset than those that used global embedding. Such local methods strive to map nearby samples on the original manifold to nearby samples in the low-dimensional space (and vice versa for far away samples). Global methods, in contrast, strive for a faithful representation of the data's global structure. As reach-grasp-lift motion is composed of different phases of movement, local methods may better preserve the varying geometry across phases. Local methods are also computationally more efficient, involving only sparse matrix computations. It may further not be surprising that k-NN works best with local dimensionality-reduction methods. These methods keep nearby samples close to each other, facilitating nearest-neighbor approaches like k-NN.

Our method, therefore, resulted in relatively high accuracy on 3-way classification while maintaining automatic feature extraction. What is more, the methodology proposed in this paper is well suited to real-time operation, potentially in combination with EEG [53], because the computational load in training and testing the model is relatively low. In addition, the variability in many datasets could be due to just a small number of factors. If that is the case, the samples from these datasets may well lie on or near some low-dimensional manifold embedded in the high dimensional space. For instance, natural signal variation among different subjects, fatigue, and delay in performing the tasks are very poorly approximated by changes in linear basic functions. However, previous studies suggest that manifold learning could capture these changes, and, using affine transformations, may even tolerate the effect of variations [54, 55].

To better understand the suitability of this method for real-time decoding—for example to control a powered prosthesis—we needed to better understand the evolution of the classification accuracy over time. One pertinent question is how soon after touching the object would there be information in the muscle about the weight of the object that is decodable using this technique. For a running-window analysis, the shortest time-window possible using our technique (100 ms) suggested that the information exists in the muscle already within the first 100 ms after the subjects touch the object (Figure. 9b). We also saw that the accuracy of our method was generally stable over the time duration when the subject grasped and moved the object. Achieving a stable decoding accuracy quickly after touching the object bodes well for the use of this technique in real time, though the relatively low accuracy over small time windows is a limitation worth noting. Therefore, constructing a combination of dimensionality reduction and classification techniques to specifically manage classification over small time windows is an interesting area of investigation for future studies.

2.5 Conclusion, Limitations and Future Work

This study proposes a complete, automated pipeline for the preprocessing, feature selection, feature extraction, and classification of objects of 3 different weights in a reach-grasp-lift task, where the only input was pre-processed EMG data from 5 muscles. Besides showcasing relatively high classification accuracy (F1 score 88.2±3.5%), our study highlights the importance of properly combining feature selection and classification algorithms to achieve this high accuracy.

The findings of our study are limited by a few factors. First, we used the open-source dataset, which has only 12 subjects, so the results of our statistical analyses should be interpreted cautiously. We have also left an analysis of the effect of fatigue on weight decoding for future studies. Nevertheless, given the high accuracy of our method overall, it is likely that the effect of fatigue on decoding accuracy is not dramatic. Similarly, the lower decoding accuracy of our method on smaller time windows deserves additional scrutiny.

3 Data Augmentation for Deep-Learning-Based Electroencephalography

3.1 Introduction

Electroencephalography (EEG) measures electric fluctuations in the brain. One use of EEG is to measure rhythmic oscillations, which reflect synchronized activity of substantial populations of neurons. Changes in these rhythmic oscillations during cognitive tasks correlate with task conditions, including perceptual, cognitive, motor, emotional, and other functional processes. This renders such task monitoring tractable using EEG [56]. Several reasons make EEG a useful tool for studying neurocognitive processes. First, it captures cognitive dynamics in the time scale at which cognition occurs—tens to hundreds of milliseconds. Second, EEG can directly measure complex patterns of neural activity within small fractions of a second after stimulus onset. Third, the EEG signal is multidimensional, comprising time and frequency, power, and phase, across many electrodes over the scalp. This multidimensionality facilitates specifying and testing hypotheses that are rooted both in neurophysiology and in psychology [57]. Nevertheless, EEG also suffers from several limitations. First, it is an aggregate signal emanating from the aggregated neuronal activity of millions or more cells, which has been transduced through several layers of tissue, fluid, bone, etc. EEG also suffers from low signal-to-noise ratio (SNR) [56-59]. Though various filtering and de-noising techniques strive to decrease the noise in favor of the underlying neural activity. What is more, EEG is a non-stationary signal—its statistics varying over time [56, 60, 61]. This is especially problematic for online, real-time analysis, where it is inherently models that were trained on past neural data that are used to decode present neural activity. Further, for complex machine-learning models, model training time might be lengthy. Hence, not only is it well outside the scope of real-time analysis, necessitating off-line training, but the statistics of the relevant brain activity may change considerably by the time the model is trained. There have been some attempts at adaptive machine-learning techniques to better track the changing statistics of the signal [62-64]. If that is not enough, EEG is generally recorded using tens to hundreds of electrodes recording simultaneously at hundreds or thousands of samples per electrode, whereas a typical dataset, at least in cognitive neuroscience, contains only some hundred to a few thousand samples (i.e., experimental trials) at the most. Hence, the initial ratio of samples to features is low. This problem is only exacerbated for datasets involving rare events, which tend to result in highly unbalanced classes of events versus non-events (e.g., for seizure detection or transitional sleep stages) [56]. Due to the above, classifiers trained on EEG datasets tend to generalize poorly to data recorded at different times, even on the same individual.

Unfortunately, there are additional challenges: inherent variabilities in brain anatomy and dynamics across subjects considerably limit the generalizability of EEG analyses across individuals [56, 65]. In other words, even if a model is well trained on one experimental subject, it would tend to generalize poorly to other subjects. Thus, most EEG classifiers tend to be subject-specific. Yet, even for a single subject, many time-consuming experimental sessions must be gathered to train the machine-learning models well enough to be useful. To overcome some of the above-mentioned limitations, processing pipelines with domain-specific approaches are often used to clean, extract relevant features from, and then classify, EEG data.

Deep Learning (DL) is a subfield of machine learning that focuses on computational models that typically learn hierarchical representations of the input data through successive non-linear transformations—termed neural networks (NN) (because of their superficial resemblance to biological neural networks in the nervous system) [56, 66, 67]. In the past few years, DL has achieved breakthrough accuracies and discovered intricate structures in complex and high-dimensional data such as image classification [68-70], speech recognition [71-73], machine translation, and more [56]. The architecture of the neural networks, their training procedure, regularization, optimization, and hyper-parameter searches are all active research topics in DL, with advances often resulting in dramatic increases in decoding accuracy. DL typically thrives on problems where (1) there is a lot of data, and (2) The basic unit of information (e.g., a pixel, a letter) has little overall meaning; but potentially complex, hierarchical combinations of such units are useful in understanding the sample. Successful machine-learning classification also at least has the potential to make considerable impact on EEG decoding, remarkably simplifying its processing pipelines for example. It could possibly enable automatic end-to-end learning of preprocessing, feature extraction, and classification modules, while also reaching competitive performance on the target task [74, 75]. DL in particular has shown some promise for inter-subject generalization [76], which is especially important when only little data is available per subject. A critical question concerning the application of DL to EEG data is therefore "How much EEG data is enough for a desired accuracy level?" Unfortunately, high-grade EEG data collection requires relatively expensive hardware and a lot of participant time. At the same time, access to large, especially clinical, dataset is often limited by privacy and proprietariness concerns. Therefore, large, openly available EEG datasets are uncommon.

Data augmentation (DA) comprises the generation of new samples to augment an existing dataset by transforming existing samples in a manner that increases the accuracy and stability of the classification. Exposing the classifiers to varied representations of its training samples makes the model more invariant and robust to such transformations when attempting to generalize the model to new datasets [77-79]. DA has proven effective in many fields, such as image processing and object recognition. It has even been demonstrated that can give a higher accuracy boost very deep neural networks than the other standard approach, which is regularization [79].

New, augmented data is typically generated using two approaches. The first is by applying geometric transformations: translations, rotations, cropping, flipping, scaling, etc. The second is via the addition of noise to the existing training data. Note that increasing the size of the training set also facilitates training more complex models with additional parameters and/or reducing overfitting. However, unlike images, EEG is a collection of very noisy, somewhat correlated (in time and space), non-stationary time-series from different electrodes. And even if feature extraction is performed, geometric transformations are not directly suitable for EEG data because those may destroy time-domain features [78]. Also, while a human can easily decide whether an augmented dataset (e.g., of cats or other images) still resembles the original class, the same is not true of augmented EEG signals. In other words, correctly labeling augmented datasets can be difficult. Nevertheless, in recent years DA techniques have received widespread attention and achieved appreciable performance boosts when using DL on EEG signals.

We ran a systematic review on DA in EEG and collected all the papers that we were able to find up to and including 2019. The earliest paper we could find was in 2015. And a testament to the growing importance of DA for EEG is that 37 out of 53 papers we found (70%) were from 2018 and 2019 and 21 (40%) were from 2019 alone. This review paper strives to identify trends and highlight available approches in DA for DL in EEG to address the following critical questions: (1) What DA approaches exist for EEG? (2) Which dataset and EEG classification tasks have been explored with DA? (3) Are there specific DA methods suitable for specific tasks measured by EEG? (4) Which of the input features in EEG are used for training the deep NNs with DA?

3.2 Methods

3.2.1 Search method for identification of related studies

The search was conducted on 3rd January 2020 within the Google Scholar, Web of Science, and PubMed databases using the following group of keywords: ('Data Augmentation') AND ('Deep Neural Network' OR 'Deep Learning' OR 'Deep Machine Learning' OR 'Deep Convolutional' OR 'Representation Learning' OR 'Deep Recurrent' OR 'Deep LSTM') AND ('EEG' OR 'Electroencephalography'). Only studies within the inclusion criteria are included below. Further, duplicates among these databases were removed from the search results. Full texts of the remaining studies were then screened.

Inclusion criteria	Exclusion criteria
EEG classification—This review focused solely on classification based on EEG signals. Deep learning—In this review, DL is defined as learning using a neural network with at least one hidden layer EEG augmentation— This review focused on the augmentation of EEG signals.	Other studies, such as power analysis and feature selection with no end classification, were excluded. review papers were excluded

English Journal and conference papers, as well as electronic preprints, published were chosen as the target of this review.	
Studies focusing only on “EEG” AND “DL” and “DA”	

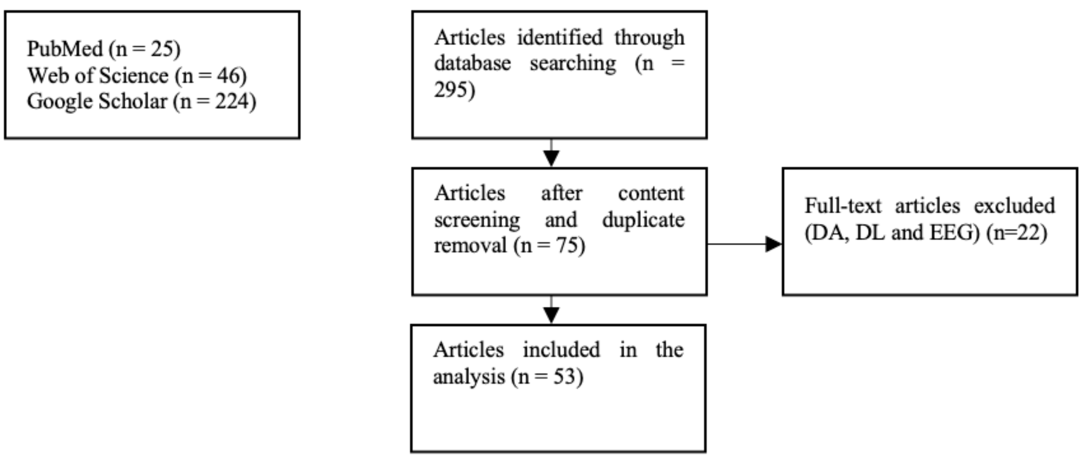

Figure 10. Selection process for the papers

The database queries yielded 295 matching results. Of those, 32 were duplicated. After screening the others, we ended up with 75 papers. Based on our inclusion and exclusion criteria, 53 papers were selected for inclusion in this analysis, as shown in Figure 10.

3.2.2 Data Extraction and presentation

For each selected paper, around 40 features were extracted covering 7 categories: Origin of the article, DA types, Dataset, Task information, Preprocessing, DL strategy, Results (Table 5).

3.3 Results

3.3.1 Origin of the selected studies

Table 5. Data items extracted for each article selected

Category	Data item
Article origin	Type of publication (Journal article, conference article, or in an electronic preprint repository)
Data augmentation (DA)	DA technique used to generate new samples Parameters for DA Magnification factor (m)
Dataset	Quantity of data, subjects, classes, channels
Task information	Task type
Preprocessing	Frequency range used for analysis EEG signal features
Deep-learning strategy	Main characteristics of NN, such as number of convolutional layers, hidden layers, activation function of hidden layers and output.
Results	Decoding accuracy

Our research methodology returned 26 journal papers,16 conference and workshop papers, and 11 preprints (arXiv or bioRxiv) that met our inclusion criteria. There were 4 papers in IEEE Transactions on Neural Systems and Rehabilitation Engineering, 3 papers in Biomedical Signal Processing and Control and the rest of the papers were each in a different journal (see Table 8 for details). Interestingly, we found no papers that fulfilled our search criteria before 2015. Further, testament to the growing importance of DA for EEG is the clear year-by-year rise in the number of papers answering our search criteria from 2015 to 2019 (Figure 11 A).

3.3.2 EEG classification task

The EEG tasks in these papers fell into 7 groups: seizure-detection (24%), motor imagery (21%), sleep stages (15%), emotion recognition (15%), mental workload (9%), motor task (8%), and visual task (8%) (Figure 11 11B). The following describes these EEG tasks (see Table 8 for more details):

Seizure-detection studies. A seizure is a sudden, uncontrolled disturbance in the electrical activity of the brain. For seizure detection in epilepsy, EEG signals are recorded during seizure and non-seizure periods. The goal of these studies is to detect upcoming seizure and preemptive notification to the patients [80, 81]. Seizure manifestations on EEG are extremely variable both inter- and intra-patient. Naturally, non-seizure events are easy enough to record. But seizures tend to be rare. DA has been successful at increasing the number of rare events (seizures) in the dataset and thus at increasing the accuracy of seizure-detection algorithms.

Motor imagery tasks. These studies instruct subjects to imagine moving their limbs, tongue, or other body parts. Motor imagery EEG decoding is an important method in brain-computer interfaces (BCI) that has the potential to help highly disabled people communicate with the outside world without relying on muscle activity (e.g. [82]).

Sleep stages scoring tasks. Studies on sleep-stage classification record the EEG signal of subjects overnight. These signals are then scored and classified to wakefulness (W) and then 4 stages of sleep based on the American Academy of Sleep Medicine(AASM) scoring manual: Rapid eye movements, or REM (R) and 3 non-REM stages (N1, N2, and N3) [83, 84]. The eventual application of this research focuses on sleep related disorders, such as sleep apnea, insomnia, and narcolepsy (e.g. [85, 86]).

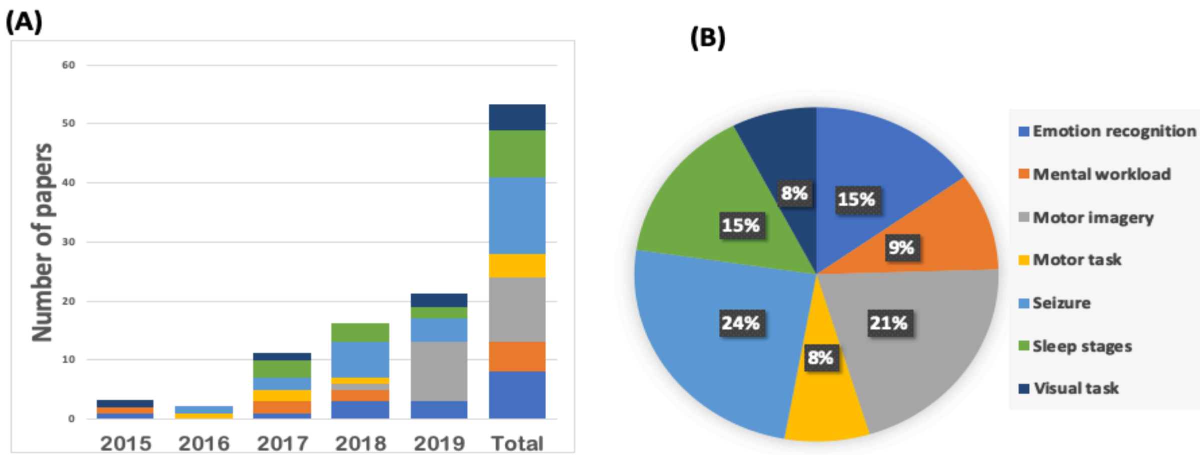

Figure 11. EEG classification task. (A) Number of publications per domain of EEG task per year. (B)The percentage of different EEG classification task across all studies.

Emotion recognition tasks. Here subjects watch video clips, which have been categorized by experts as eliciting various emotions. Facial expressions and EEG signals are then recorded from the subjects. However, some subjects may hide their real emotions using misleading facial expressions. Therefore, EEG signals and emotion self-assessment typically follows. The result can be parsed into valence and arousal scales. Emotion recognition is a crucial problem in human-computer interaction (HCI) for example: virtual reality, video games, and educational systems (e.g. [78]).

Mental workload tasks. Subjects are here instructed to carry out different mental tasks of varying complexity. The results of these studies reflect the interaction between the human inner cognitive capacity and the level of task complexity. Research into mental workload has applications in BCI performance monitoring and in cognitive stress monitoring (e.g. [87]).

Motor tasks. Here, subjects are instructed to either rest or move some parts of their bodies. Researchers use such tasks to design, modify or improve classification methods for different applications (e.g. [88]).

Visual tasks. These studies focus on the detection and classification of the intentions and decisions of subjects while they watch rapidly changing sequences of pictures or letters. This helps to improved non-verbal communication systems and BCI (e.g. [89]).

3.3.3 Data and Reproducibility

We collected dataset information for 53 papers. This information included:

- Data quantity: Amount of data in the study (total hours of recording or number of samples)
- Number of Channels: Number of channels recorded and which of them were used for analysis
- Subjects: Number of recorded participants and which of them were analyzed
- Dataset: Publicly available, proprietary, etc.

See Table 8 for more details.

3.3.4 Pre-processing and Feature extraction

The analysis of EEG signals is typically carried out by one of two methods. The first is event-related potentials, which are fluctuations of the potentials over time that are locked to an event (e.g., to 'stimulus onset' or 'button press'). The second is spectral analysis of rhythmic oscillations, which reflect the synchronized activity of very large populations of neurons. Regardless of the analysis method, it is an aggregate signal emanating from the neuronal activity of millions or more brain cells, which has been transduced through several layers of tissue, fluid, bone, etc. It also potentially includes undesired electrophysiological signals, such as electromyograms (EMG) of muscle contractions specifically eye blinks, heart beats, and others. Therefore, the EEG signal is inherently noisy. Though various filtering and de-noising techniques strive to decrease the noise in favor of the underlying neural activity. In the 53 studies we found, 85% (45 studies) removed the artifacts manually—mainly using high, low, and band pass filtering. A further 13% (7) of studies did not take any action to remove artifacts, and the remaining study (2%) did not address artifact removal.

Most studies used frequency domain filters to limit the bandwidth of the EEG signals. This enabled them to focus on a certain frequency range that was of interest. Roughly, half of the reviewed papers low pass filtered the signal below low gamma band or 40 Hz. The filtered frequency ranges, organized by task type (Figure 12 12). We found that there were no studies that specifically check the role of this filtering for NN [75].

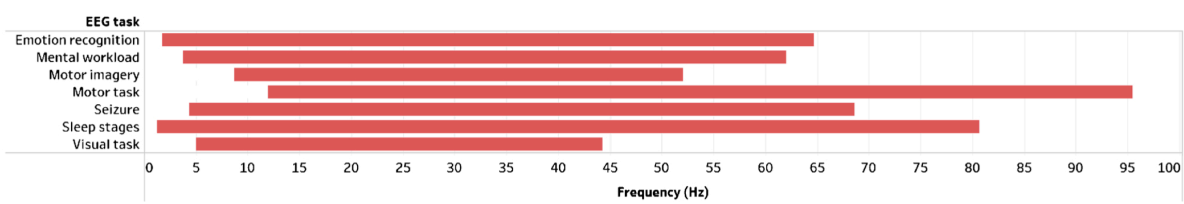

Figure 12. Frequency range used in EEG analysis for each identified study, organized by EEG task type.

3.3.5 Input Formulation

The inputs to the NNs in the studies that fulfilled our inclusion criteria, fell into three categories. The first included raw EEG signals (in the time domain) (36%). The second calculated features from the raw signals and used those as inputs (49%). And the third used spectrograms, processed as images (15%). The selection of input formulation heavily depended on the task type and deep-learning architecture. Thus, we can see that most of the studies used calculated features to train their proposed NNs. When attempting to find behavioral patterns, it is common to analyze specific frequency ranges of EEG signals. Wavelet, entropy, spatial filter, short-time Fourier transform (STFT), spatio-temporal features, and power spectral density were used in the reviewed papers to calculate the features of EEG the signals. Raw EEG values was another popular feature for training NN. It's interesting that NNs can learn complicated features from large amount of raw data. Many NNs, especially RNN, used spectrogram and fast Fourier transform (FFT) to convert EEG signals to images (Figure 15). When we analyzed the studies that fulfilled our inclusion criteria based on the input formulation and on the EEG task, we found that (Emotion recognition, mental workload, motor imagery, and seizure) mostly used calculated features. Motor task, sleep stages, and visual task chose signal values as their input primarily (Figure 15).

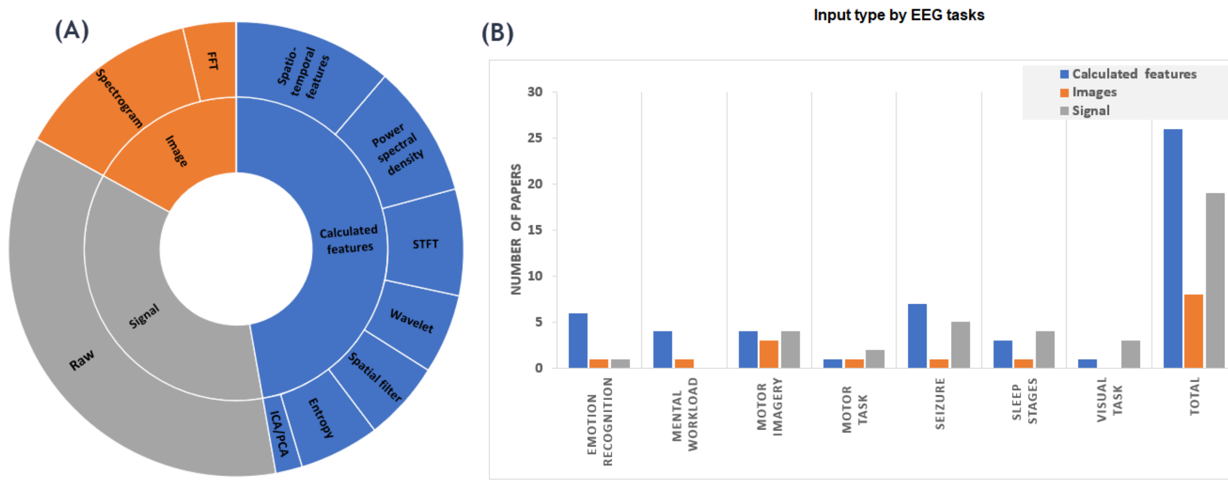

Figure 13. Input formulation across all reviewed papers. (A) The inner circle shows the general input formulation, while the outer circle shows more specific details. (B) Number of papers for general input formulation compared across different tasks.

3.3.6 Deep learning architectures

Deep learning is a subfield of machine learning based on artificial neural networks, which can be thought of as learn hierarchical representations of the input data through non-linear transformations. While beginning to rise to prominence in the late 2000’s, in the few years since, it has arguably revolutionized the field, achieving remarkable accuracy on, and discovering intricate structures in complex and high-dimensional data, such as image classification, speech recognition, and automated translation. Various deep learning architectures have been developed since, with this fast-moving research field routinely producing new architectures. We discerned 6 different categories in deep learning: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Multi-layer perceptron (MLP), Stacked Auto Encoders (SAE), Long Short-Term Memory (LSTM), and hybrid combinations of the above. By order of prevalence these were: CNN (62%), Hybrid (16%), MLP (8%), SAE (6%), LSTM (6%), and RNN (2%) (Figure 17).

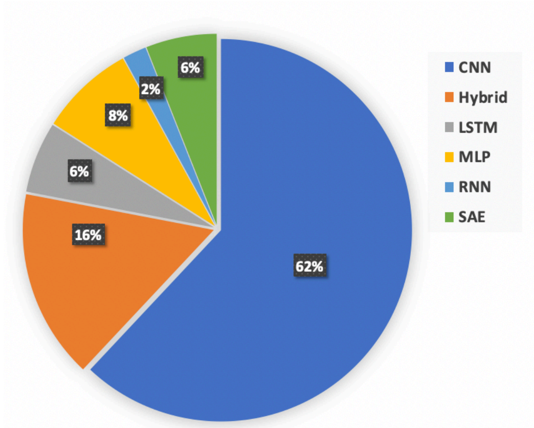

Figure 14. Deep learning architecture across all studies

Figure 15 visualizes the aggregated information about DL architecture of reviewed studies. This figure helps to understanding the trends in the formation of specific deep-learning architectures. For more details see Table 8.

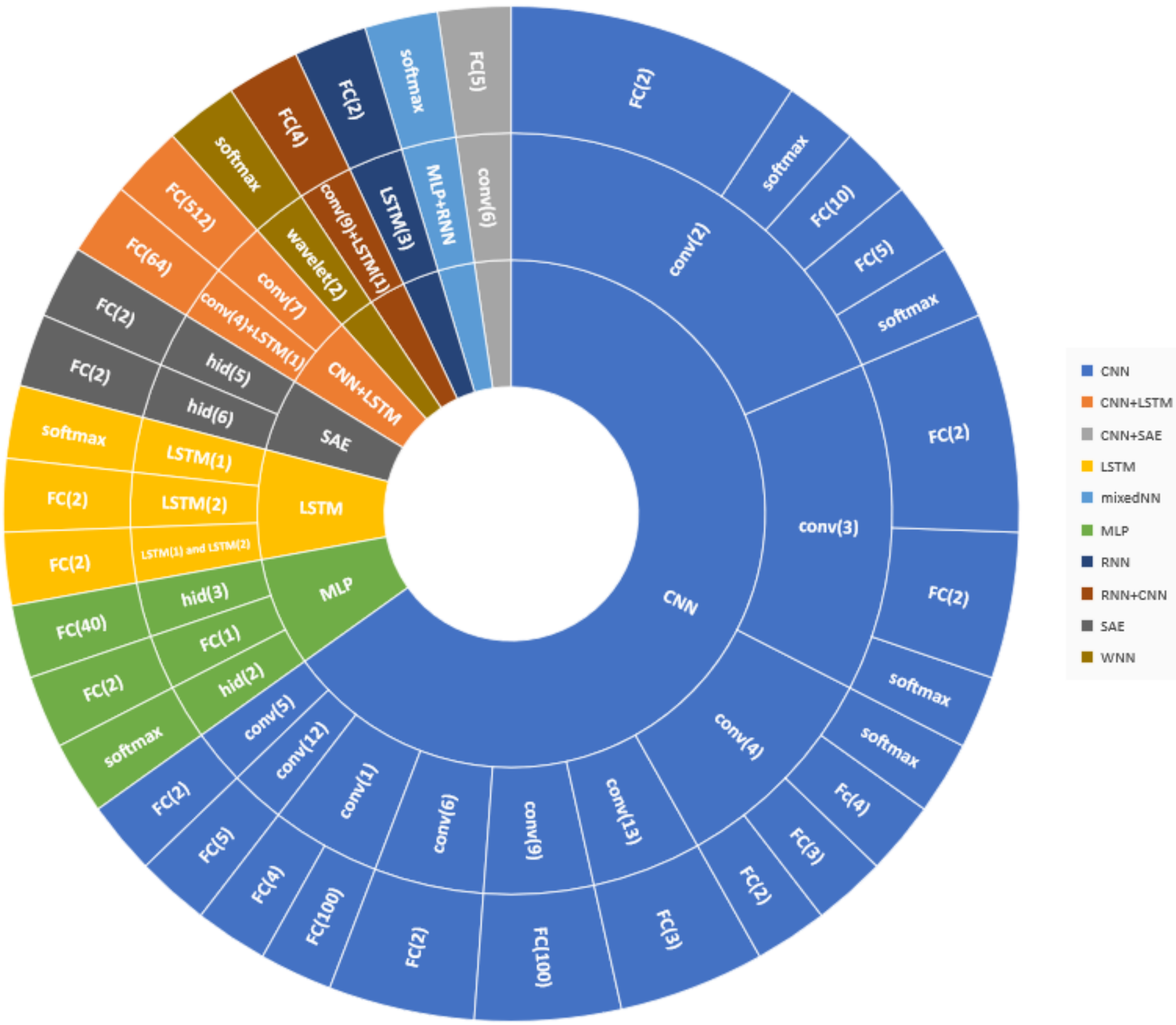

Figure 15. Aggregated information of deep learning architectures. The inner circle shows the general DL architecture, the middle circle, shows the primary design features, such as the hidden layers or convolutional layers, and the outer circle shows the last layer of DL architecture. FC: Fully connected, hid: Hidden layers, softmax: Softmax function.

Figure 16 visualizes the proportion of input formulation by DL architecture. As is apparent, the specific input formulation strategies varied significantly as a function of the type of the deep learning architecture. While there was not a clear consensus for all studies together, RNN and SAE architectures used only images and calculated features as inputs, respectively. Hybrid, CNN and MLP studies included instances of all 3 information types. Interestingly MLP and CNN used directly signal values as inputs.

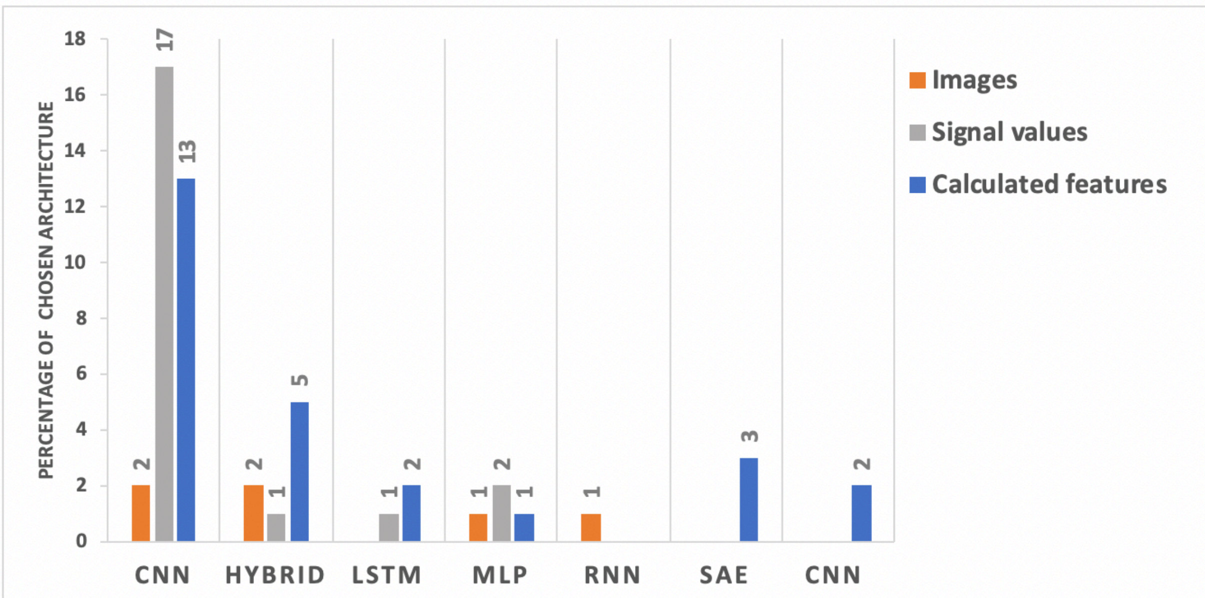

Figure 16. The percentage of input formulation by chosen DL architecture

3.4 Data Augmentation methods

This section details the methods found for methods that have so far been used to augment EEG signal for machine learning. Data augmentation (DA) comprises the generation of new samples to augment an existing dataset by transforming the existing samples in a manner that increases the accuracy and stability of the classification or regression. Exposing the classifier to more variable representations of its training samples make the model more invariant and robust to transformations of the type that it is likely to encounter when attempting to generalize to unseen samples. Further, increasing the size of the training set facilitates training more complex models with additional parameters and/or reducing overfitting. In recent years, DA techniques have received widespread attention and achieved appreciable performance boosts for DL on EEG signals. Here we cover all the papers that we were able to find up to and including 2019. The first paper was found in 2015. The testament to the growing importance of DA for EEG is that 37 out of 53 papers (72%) we found are from 2018 and 2019 (Figure 17).

The DA for DL-based EEG in 53 papers fell into 7 categories in our analysis: noise addition (17%), GAN (21%), sliding window (24%), sampling (17%), Fourier transform (4%), recombination of segmentation (6%) and other (11%) (Figure 17). Below we discuss each DA method in much more detail.

3.4.1 Noise addition

In our research, we found two main categories for adding noise to the EEG signals in purpose of DA: (1) Add various types of noise such as Gaussian, Poisson, salt and pepper noise, etc. with different parameters (for instance: mean ($\boldsymbol{\mu}$) and standard deviation ($\boldsymbol{\sigma}$)) to the raw signal (2) Convert EEG signals to sequences of images and add noise to the images. Nine papers used noise addition method to increase training dataset.

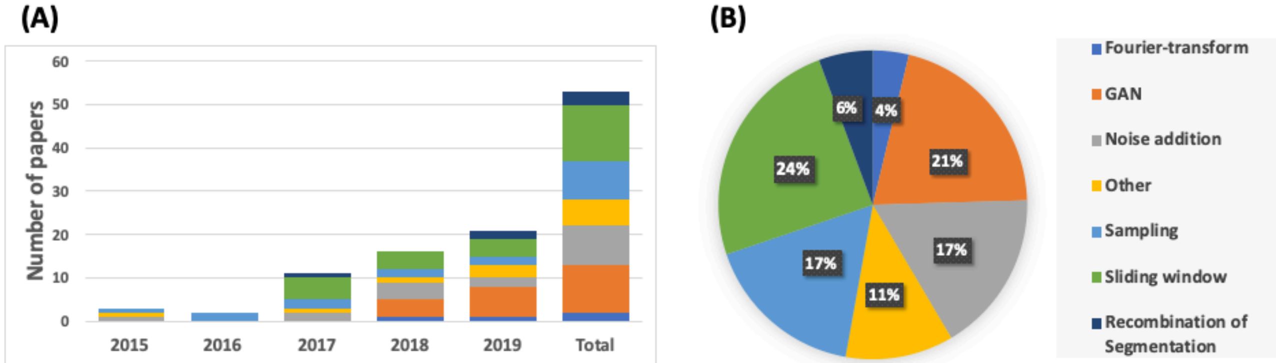

Figure 17. DA across all studies. (A) Number of publications per domain of DA per year. (B) The percentage of different DA methods across all studies. Note that we only collected data until January 2020.

In 2015, Bashivan et al. transformed EEG signals into a sequence of topology-preserving multi-spectral (2D feature images) in a specific time interval [90]. FFT was performed on the mental load EEG signals to estimate the power spectrum of the signal in three frequency bands of theta (4-7Hz), alpha (8,13Hz), and beta (13-30Hz). A single image was constructed from spectral power within three prominent frequency band which is extracted from each electrode location. The sequences of image representations fed into the LSTM and CNN for the EEG classification. For addressing the unbalanced ratio between number of samples and number of model parameters, they randomly added various noise level to the images. However, augmenting the dataset did not improve the classification performance and even for higher value of noise, the error rate increased.

Z. Yin et al. (2017) proposed an adaptive DL model based on Stacked Denoising AutoEncoders (SDAE), which was designed for cross-session Mental Workload (MW) classification using EEG [91, 92]. They could increase the accuracy of their model by adding Gaussian white noise to the EEG feature vector ($\mu = 0.01, m = 2,3,4,5,6$). This vector contains centroid frequency, log-energy entropy, mean, five power components, Shannon entropy, sum of energy, variance, zerocrossing rate of each channel and power differences between four selected channel pairs. Their classification accuracy on an independent dataset improved from 76.5% (without DA) to 85.5% (with DA). The highest classification accuracy was achieved with m=6 and the lowest with m=0 (without DA). They concluded that the number of samples (trials) in the original dataset was insufficient for training the NN.

Wang et al. (2018) added Gaussian white noise to their training data (in the time domain) to obtain new samples for an emotion-recognition task [78]. In their experiments, EEG signals were recorded while subjects were watching emotionally loaded videos. They used differential entropy (DE) features to train their proposed classifiers. For EEG signals, the DE feature is equivalent to the logarithm of the energy spectrum in the delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz) frequency bands. The authors opted for Gaussian noise due to concerns that adding some local noise (i.e., noise that affects EEG data locally) such as Poisson or salt-and-pepper may change the intrinsic features of EEG signals. The probability density function, P, of a Gaussian random variable, z, is defined by:

$$\boldsymbol{P_G(z)} = \frac{\mathbf{1}}{\boldsymbol{\sigma\sqrt{2\pi}}} \boldsymbol{e}^{\frac{-(z-\mu)^2}{2\sigma^2}}$$

where z represents the density level, μ is the mean value and σ is the standard deviation. The experimental results on SEED dataset showed that by augmenting training dataset 30 times, the accuracy of ResNet improved from 34.2% to 75%, better than LeNet (from 49.6% to 74.3%).

R. Hussein et al. (2018) used another DL technique, using a recurrent neural network (RNN) and Long Short-Term Memory (LSTM) network. Their goal was automatic detection of epileptic seizures using EEG signals [93]. And they reported that they improved the robustness of their model by adding Gaussian white noise, muscle artifacts and eye-blinking. Though they did not give any specific details about the DA methods they used.

S. Kuanar et al. (2018) also used an LSTM with a convolutional neural network (CNN) to learn robust features and predict the levels of cognitive load from EEG recordings [87]. They transformed the EEG time-series into a sequence of multispectral images that carried spatial information—theta (4-7Hz), alpha (8-13Hz), and beta (13-30 Hz). The data was once again augmented by adding various Gaussian noise level to the images. Though they did not compare any specific details about the DA methods they used.

E. Salama et al. (2018) generated the noisy EEG signals, by adding Gaussian noise with zero mean and unit variance to the original input EEG training dataset [94]. They set the signal-to-noise ratio (SNR) between original EEG signal and the noisy to 5. The DA phase enhanced the performance of the proposed 3D-CNN on emotion recognition dataset. For valence and arousal classification, they achieved 79.11% (without DA) and 88.49%(with DA). For 4 combinations of valence and arousal — (low valence-low arousal), (low valence-high arousal), (high valence-low arousal) and (high valence-high arousal) they obtained 79.11%(without DA) and 87.44%(with DA).

Parvan et al. (2019) doubled the number of trials of BCI competition IV dataset 2b by adding gaussian noise with zero mean and a standard deviation of 0.15 to ovoid overfitting [95]. Their proposed CNN had 4 convolution layers as well as data augmentation and resulted in a 0.07 improvement in the kappa coefficient [95].

Y. Li et al. (2019) emphasized the fact that increasing depth of CNN causes a higher classification accuracy. However, doing so may aggravate the vanishing-gradient problem and substantially increase the number of trainable parameters to be tuned, and these models may tend to be overfitting easily [96]. For four-class motor imagery task, they exploit the standard deviation of Gaussian noise in the DA affects the classification result. The optimal standard deviation is 0.001with zero mean on 2 imagery task datasets (Table 6). It is noticeable that for almost all subjects, the performance has been significantly improved after DA. Furthermore, by comparing confusion matrix before and after DA, they showed that for a specific imagery task, DA worked well except for one task(feet). Table 6 shows all the papers used noise addition as their DA technique. From this table, we can see that there is lack of information about noise addition parameters (**μ**: mean, ***σ***: standard deviation), magnification factor (m) and reported accuracy before and after DA. Maybe this is because that their problem wasn't DA topic and they wanted to increase just performance accuracy.

Table 6. All reviewed papers that used noise addition as their DA technique

Dataset	Task information	Input formulation	Deep learning strategy	Noise Addition parameters	Accuracy (without DA)	Accuracy (with DA)
University of Memphis Institutional review board 13 subjects 2670 trials 64 channels	Mental workload 4-13 Hz	Images, FFT	CNN+LSTM Conv(7) + FC(512) Relu, softmax	NA	NA	Did not improve
SDAE 8 subjects 180 min/subject 1 channel	Mental workload 1.5-40Hz	Calculated features, Power	SAE Hid(5) + FC(2) Sigmoid, NA	$\sigma = 0.01$, $\mu = 0$, $m = 6$	NA	93%

2 class		spectral density				
AutoCAM 7 subjects 1h 11 channel	Mental workload 1-40 Hz	Calculated features, FFT and power spectral	SAE Hid(6) + FC(2) Sigmoid, sigmoid	$\sigma = [0.1, 0.2, \ldots 1.5, \mu = 0, m = 6$	76.5%	85.5%
SEED 14 subjects 1890 trials 62 channels 3 class	Emotional recognition 1-50 Hz	Calculated features, Entropy	CNN Conv(4) + FC(3) sigmoid	$\sigma = 0.2, \mu = 0, m = 30$	49.6%	74.3%
SEED 14 subjects 1890 trials 62 channels 3 class	Emotional recognition 1-50 Hz	Calculated features, Entropy	CNN Conv(13)+FC(3) sigmoid	$\sigma = 0.2, \mu = 0, m = 30$	34.2%	75%
MAHNOB-HCI 30 subjects 527 trials 32 channel 3 class	Emotional recognition 1-50Hz	Calculated features, Entropy	CNN Conv(13)+FC(3)	$\sigma = 0.2, \mu = 0, m = 30$	40.8%	45.4%
Bonn University 5 subject [2,3,5] class	Seizure [0.53, 40] Hz	Raw signal	LSTM softmax	Gaussian white noise+(muscle and eye blink)	NA	2 class: 99%
NIMHANS 22 subject, 6490 trials(8 hours)	Mental workload 4-30 Hz	Calculated features, power	RNN+CNN Conv(9)+LSTM(1) Relu, softmax	NA Add noise to image M: NA	NA	93%

64 channels 4 class		spectral density				
DEAP 32 subject 40min/subject 32channels 2 and 4 class	Emotion recognition [1,50]U[60,end) Hz	Calculated features Spatio-temporal	CNN Conv(2) Relu, softmax	$\sigma = 1$, $\mu = 0$, m=[10,30,50]	79.11% 79.12%	88.49% 87.44%
BCI competition IV 2b 9 subjects 5 sessions 3 channels	Motor Imagery [0.5,100] Hz	Raw signal	CNN Conv(4)+FC(2) Elu, softmax	$\sigma = 0.15$, $\mu = 0$, $m = 2$	NA	NA
BCI competition IV 2a 9 subject 72trials/subject 22 channel 4 class	Motor imagery 7-125 Hz	Calculated features Spatio-temporal	CNN Conv(1) FC(4)+softmax Relu, sigmoid	$\sigma = 0.001$, $\mu = 0$,	Reported subject by subject e.g. 70%	Increased 77.9%
High Gamma dataset(HGD) 30 subjects 7000trials/subject 1channel 2 class	Motor imagery 7-125 Hz	Calculated features Spatio-temporal	CNN Conv(1) FC(4)+softmax Relu, sigmoid	$\sigma = 0.001$, $\mu = 0$,	NA	NA

3.4.2 Generative adversarial network

The term Generative Adversarial Network (GAN) was first demonstrated by Goodfellow, et al. as a new framework to learn the underlying distribution of data from two competing networks: the generator (G) and the discriminator (D). While the generator makes “fake data”, the discriminator classifies the “fake data” as real or fake using the given label as if they were playing a minimax game [97].

During the process, the generator gets better at generating data that are similar to the real data, until the discriminator fails to distinguish real from fake data Figure 18. The minimax game of a GAN is given by:

$$\min_{D}\max_{G} V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_{z}(z)}[1 - \log D(G(z))],$$

where p_{data} is the distribution of the real data and p_z is gaussian noise. $D(x)$ gives a probability of an input x belonging to the real data, while $G(z)$ produces fake samples that strive to trick D by learning how to produce data that appears to come from the distribution of the real samples, p_{data}. The optimization process utilized the Jensen-Shannon (JS) Divergence to find the minimum of the function [97].

GANs have been widely applied for generating data in many disciplines outside neuroscience and EEG. For example, In the method that Zhang et al. (2017) proposed a GAN was used to generate images from text [98]. Bousmalis, K., et al. (2017) strives to generate rendered images that are similar to images in a dataset [99]. Antoniou et al. (2017)used a GAN to create new data from three different popular image datasets: Omniglot, EMNIST, and VGG-face [100].

Specifically for augmenting EEG signals, Zhang et al. (2018) proposed a conditional deep convolutional generative adversarial network (cDCGAN) [101]. The cDCGAN is an improved version of the GAN that uses information from the labels and adds them to the model as conditional properties:

$$\min_{D} \max_{G} V(D,G) = E_{x \sim p_{data}(x)}[\log D(x|y_z)] + E_{z \sim p_z(z)}[1 - \log D(G(z|y_d))],$$

where y_z and y_d is the information from the corresponding labels. The dataset contained EEG signals recorded over 3 electrodes, and composed of 7 sessions, with 40 trials per second, each lasting 9 seconds. It was collected while subjects were asked to imagine moving either left or right. A CNN was trained to classify each EEG signal as Left or Right.

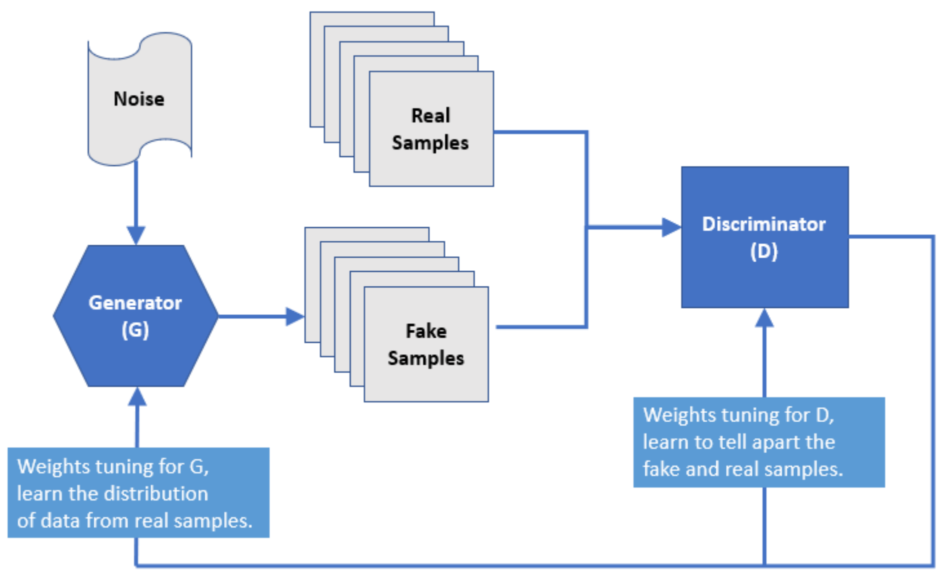

Figure 18. Diagram of Generative Adversarial Network

The EEG signals were preprocessed before feeding them into the CNN. Only 5 out of the 9 seconds of EEG in each trial were selected for processing and only alpha (7-15 Hz) frequency components were extracted as time-frequency features. Using data generated from the cDCGAN, classification accuracy increased from 83% to 86%. The authors compare the accuracies for models trained using different proportions of artificial data. However, the largest dataset only doubles the original dataset in the experiment (i.e. m=2), while others have used larger augmentation.

Piplani et al. (2018) used a GAN to generate more EEG data to increase the robustness of a "passthought authentication system" that uses the user's EEG signals to securely log into devices [102]. The EEG signals were collected using a device with only a single channel at a sampling rate of 500Hz. The 'negative' samples were collected from 30 subjects who were asked to perform a series of mental task for 5 minutes while EEG was recorded. The 'positive' samples were collected from one subject while the subject was doing the same mental tasks for 5 mins and free to do any tasks for another 5 minutes. The dataset that trains the selected model, XGBoost, consists of 30,000 negative samples and 40,000 positive samples. Each sample is a segment of the EEG signal. These data were augmented with 10,000 artificial EEG signals that were generated from a GAN. This increased the accuracy of the model from 90.8% to 95.0%, which is noteworthy for such high accuracies.

Zhang et al. (2018) proposed a framework called Deep Adversarial Data Augmentation (DADA) for generating new data, allowing deep network classifiers to be trained on small datasets [103]. They further investigated and compared different traditional approaches for dealing with small datasets in DL applications—such as dimensionality reduction, semi-supervised learning, transfer learning, and data augmentation. DA was widely used for image data because images can be altered easily—maintaining their content on the one hand while increasing the variance of the representation of that content by rotating, cropping, scaling or just adding noise to the original dataset. However, these techniques are usually not suitable for non-image data such as EEG signals. One of the examples in this study focuses on increasing the size of an EEG dataset from a BCI competition [104]. This dataset contained 3 channels (C3, Cz, and C4) of EEG collected from 400 trials of motor imaginary tasks. Time-frequency features were extracted from these EEG signals, which formed a 32 x 32 x 3 image for each EEG signal, which was in turn used for training a CNN classifier. Compared to the traditional GAN, DADA was able to generate more diverse artificial data because of its redesigned loss function. A traditional GAN trains the discriminator on only 2 classes. In contrast, DADA uses 2K classes for the discriminator, K for each of the real and artificial datasets. They found that accuracy increased from 74.8%, using a traditional CNN as a benchmark, to 79.3%, using the DADA model.

Hartmann et al. (2018) used a slightly modified version of the Wasserstein Generative Adversarial Network (WGAN) to generate new EEG signals [105]. Training an original GAN suffered from vanishing gradients while optimizing the JS Divergence [97]. A WGAN solved this problem by minimizing the Wasserstein distance:

$$W\left(p_{data}, p_{fake}\right) = E_{x\sim p_{data}}[D(x)] - E_{x\sim p_{fake}}[D(x)],$$

where p_{fake} is the distribution of the generator that generates fake (or artificial) samples. In addition, a gradient penalty term $P(p_{\hat{x}}) = \lambda \cdot E_{\hat{x}\sim p_x}[max(0, \|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$ was also added to produce a useful gradient, where $p_{\hat{x}}$ is the distribution of $\hat{x}$ that are points on a line connecting the real and fake data. Hartmann et al. improved the model by scaling λ, allowing the parameter to adjust its impact based on different Wasserstein distances [105]:

$$L = -W\left(p_{data}, p_{fake}\right) + \max\left(W\left(p_{data}, p_{fake}\right)\right) \cdot\ P(p_{\hat{x}})$$

The EEG signals were collected from a simple motor task experiment, in which subjects were asked to raise their left hand or to rest. There were 438 trials in total—286 were used for training, 72 for validation, and 80 for testing. Only one channel, FCC4h, was included in this experiment. All total 438 signals were used to train the WGAN model. Unlike other studies that only used classification accuracies to compare the quality of new generated data and the original data, four other evaluation metrics were used in this study: the inception score (IS) [106], Frechest inception distance (FID) [107], Euclidean distance (ED), and sliced Wasserstein distance (SWD) [108]. After comparing the different metrics, optimizing the GAN for good IS and FID produced the best EEG data approximations [105]. This method did not use any classification algorithms to validate the accuracy, therefore it is not included in Table 7 for accuracy comparison.

A conditional version of the WGAN was used by Luo and Lu (2018) to augment EEG data. Similar to cDCGAN, WGAN also utilized the label information to infer the distribution of the real data [109]. The datasets used to test the WGAN model were SEED [110] and DEAP [111]; two popular public EEG datasets for emotion recognition. The EEG signals from the SEED dataset had 62 channels. They were collected from 15 subjects while they were watching film clips selected to induce positive, negative, or neutral emotions. For each subject, 3394 epochs were recorded. The DEAP dataset had 32 channels of EEG signals recorded from 32 subjects, with 2400 epochs each while they were watching music videos. There were 2 classification tasks for the DEAP dataset: high vs low arousal and high vs low valence. Luo and Lu tried different sizes for the augmented data and found that doubling the data (m=2) provided the highest accuracy comparing to other attempts (m=0.5, 1.0, 1.5). An SVM classifier trained on the augmented dataset improved 2.97% for the SEED dataset from 83.99% to 86.96%. DA seemed to have a larger effect on the DEAP dataset. While classifying arousal, there was a 9.15% improvement in classification accuracy from 69.02% to 78.71%. For valence classification, the improvement was even larger with a 20.13% increase from 53.76% to 73.89%. The method did not specifically mention the chance level accuracy for both datasets. For the SEED dataset, since there are three classes, we are assuming that the chance level accuracy is 33.33%. For DEAP dataset, the chance level accuracy is 50% for binary classification.

In 2019, Luo et al. adopted a conditional Boundary Equilibrium GAN (cBEGAN) to generate artificial differential entropy features of EEG signals on 2 popular emotion recognition dataset (SEED, SEED V) [112]. cBEGAN used the Wasserstein distance to measure the difference between two reconstruction loss distributions. The main advantage of cBEGAN is that it can overcome the instability of conventional GAN and has very quick convergence speed. They generated 50 to 2000 artifacts samples and added them to the original training dataset. With 2000 added samples, the accuracy increases from 81.9% to 87.56% for SEED; and with 1000 samples, the accuracy increases from 54.3% to 62.8% for SEED V, respectively.

Wei et al. (2019) used WGAN with gradient penalty to increase the sample diversity in seizure detection in the CHB-MIT Scalp EEG database (with 23 subjects) [113]. Testing the performance on one patient, they used generated data from the other 22 patients involved in the training. They employed a 12-layers CNN and achieved 81% accuracy (without DA) and 84% (with DA).

Chang et al. (2019) used GAN to increase the size of dataset for a 2-class emotion recognition task [114]. The generator and discriminator of the GAN consists of three hidden layers, which consists of 50, 100, and 50 nodes, respectively. The number of nodes in each layer was determined after evaluations with multiple combinations of hyper parameters that showed the highest training speeds. The generator received random values between 0 and 1 and generated virtual EEG data. The discriminator received EEG collected through experiments and virtual data and distinguished the original data from the virtual data. Once the training was complete, the EEG data generated by the generator were saved. The authors increased the number of trials from 32,000 to 92,000 and by that raised the final accuracy from 97.9% to 98.4%.

Yang et at. (2019) augmentated dataset 2b competition IV BCI using a GAN network [115]. They used CNN-LSTM to classify left and right hand motor imagery task. The average accuracy for 9 subjects was 76.4%. Unfortunately, they did not report the results without DA.

Panwar et al. (2019) proposed using a class conditioned Wasserstein Generative Adversarial Network with gradient penalty (cWGAN-GP) to generate synthetic EEG data of a single channel [116]. The study claims that the cWGAN-GP method is able to counter instability and frequency artifacts problems while training an ordinary GAN [105]. The Wasserstein distance and the gradient penalty stabilized the training process [117]. The class conditioned implementation allowed the generator and discriminator to avoid mode collapsing, which is responsible for trapping the data generated from the GAN in some specific modes [105]. The proposed architecture had two fully connected layers and two convolutional layers for the generators well as three convolutional layers and two fully connected layers for the discriminator. The dataset that the paper used to train the cWGAN-GP was collected during the BCIT X2 Rapid Series Visual Presentation (RSVP) experiment, where subjects were asked to identify target images in an image stream presented at 5Hz [118]. The dataset contained EEG signals from 10 subjects, with 5 sessions and 1 hour of recording per session using a 256-channel BioSemi system. It had two classes, target and non-target, 967 samples each, which were pre-processed using the PREP pipeline [58]. The pipeline performed band-pass filtering from 0.1 to 55Hz, referencing, bad channel interpolation and baselining. One second of signal from each trial after image onset was extracted, down sampled to 64Hz and normalized using the mean and standard deviation from each epoch. The paper used three different methods to evaluate the performance of the data generated from the cWGAN-GP: visual inspection, log-likelihood distance from Gaussian mixture models (GMMs), and classifier performance. The visual inspection and GMM results both showed that the generated data was of high quality. In classifier performance evaluation, the synthetic data size was 3828, and it was added to the training dataset during training. The classifier trained with the synthetic data shows an improvement of 5.18% (from 50.02% to 55.2%) on cross subject evaluation and 3.12% (from 60.8% to 64.08%) on same subject evaluation using a CNN with 3 convolutional layers.

Aznan et al. (2019) had subjects look at one of three different objects, each flickering at 10, 12, or 15 Hz (each at a different frequency). Their goal was to detect which object the subject was looking at using BCI technology and then direct a humanoid robot toward that object. They compared three different methods: Deep Convolutional Generative Adversarial Network (DCGAN) [119], gradient panelized Wasserstein Generative Adversarial Network (WGAN-GP) [117], and Variational Auto-encoder (VAE) [120]. They then used those methods to generate synthetic EEG data to improve the classification accuracy on their Steady State Visual Evoked Potential (SSVEP) based BCI system [121]. The SSVEP-based classifier was able to pick up the corresponding frequency from the EEG. The dataset used to train the generative models is the video-stimuli dataset [121] that contains 50 samples of EEG signals collected from offline videos played to one subject, referring to subject 1 in the NAO dataset [121]. The NAO dataset has two portions—offline and online—collected from tasks the same as in the Video-Stimuli dataset using a dry EEG device with 20 channels. The three generative models were trained only using the video-stimuli dataset, while the SSVEP classifier was tested on the NAO dataset. The generated EEG samples were used to pre-train the SSVEP classifier. The offline portion of the NAO dataset for each subject was fed into the pre-train model to fine tune for that particular subject. After the classifier was trained, the online portion of the NAO dataset was used to test the performance of the classifier. Different sizes of augmentation were empirically tested and compared. The result showed that for all three methods, a sample size of 500 resulted in the best classification accuracy. Table 7 shows the performances of different methods.

Table 7 shows all the papers used GAN as their DA technique. By reporting the magnification factor and accuracy before and after DA, we think that GAN technique is trending to use as DA technique for EEG signal.

Table 7. All reviewed papers that used GAN as their DA technique

Study	Dataset	Task information	Input formulation	Deep learning strategy	Best Performance Augmented Size	Accuracy (without DA)	Accuracy (with DA)
[101]	BCI competition II dataset III 1 subject 280trials 3 channels 2 class	Motor Imaginary EEG 7-15 Hz	Calculated features spectrogram	CNN cDCGAN	Doubled	83%	86%
[102]	30 subject 70000 trials 1 channel 2class	Mental task(EEG-Based Login Authentication)	Calculated feature Power spectral	GAN ,XGBoost	Added 10,000 samples	90.8%	95%

[103]	BCI competition IV dataset 2b 1subject, 400 trials, 3 channel 2 class	Motor imagery	Calculated features spectrogram	GAN+CNN	10 times	77.6%	79.3%
[105]	1 subject, 438 trials, 1 electrode 2 class	Motor task	Calculated features-spectrogram	GAN-SWD	NA	NA	NA
[109]	SEED 15 subjects, 62 channels, 3394 samples per subject	Emotion recognition 1-50 Hz	Calculated features spectrogram	CWGAN	Doubled	83.99 % Arousal 69.02 %, Valence 53.76 %	86.96 % Arousal 78.17 %, Valence 73.89 %
[109]	32 subjects, 32 channels, 2400 trials each subject	Emotion recognition 1-50 Hz	Calculated features spectrogram	CWGAN	Doubled	NA	NA
[112]	SEED: 9 subjects, 62 channels, 3classes, 45 videos/subject SEED V: 16 subjects, 62	Emotion recognition 1-50 Hz	Calculated features Differential entropy	cBEGAN	2000 1000	81.9% 54.3%	87.56 % 62.8%

	channels, 5 classes						
[113]	23 subjects 5085 trials more than 23 channels	Seizure	Raw signal	CNN	NA	81%	84%
[114]	18 subjects 32000 samples 14 channels	Emotion recognition	Raw signal	GAN	~triple	97.9%	98.4%
[115]	9 subjects, 32 channels, 500samples/s ubject	Motor imagery 0.5-100 Hz	Raw signal	CNN+LS TM	NA	NA	76.4%
[116]	10 subjects, 256 channels, 5 hours/subject	RSVP 0.1-55 Hz	Raw signal	CNN	3828	NA	NA
[121]	video stimuli dataset: 1 subject, 20 channels, 50 unique samples for each of the three class NAO dataset: 3 subjects, 20 channels 50 samples per class offline, 30 samples per class online	Steady state visual evoked 9-60 Hz	Raw signal	CNN	500	NA	NA

Study	Dataset	Task information	Input formulation	Deep learning strategy	Best Performance Augmented Size	Accuracy (without DA)	Accuracy (with DA)
[101]	BCI competition II dataset III 1 subject 280trials 3 channels 2 class	Motor Imaginary EEG 7-15 Hz	Calculated features spectrogram	CNN cDCGAN	Doubled	83%	86%
[102]	30 subject 70000 trials 1 channel 2class	Mental task(EEG-Based Login Authentication)	Calculated feature Power spectral	GAN ,XGBoost	Added 10,000 samples	90.8%	95%
[103]	BCI competition IV dataset 2b 1subject, 400 trials, 3 channel 2 class	Motor imagery	Calculated features spectrogram	GAN+CNN	10 times	77.6%	79.3%
[105]	1 subject, 438 trials, 1 electrode 2 class	Motor task	Calculated features-spectrogram	GAN-SWD	NA	NA	NA
[109]	SEED 15 subjects, 62 channels, 3394 samples per subject	Emotion recognition 1-50 Hz	Calculated features spectrogram	CWGAN	Doubled	83.99% Arousal 69.02%,	86.96% Arousal 78.17%,

						Valence 53.76 %	Valence 73.89 %
[109]	32 subjects, 32 channels, 2400 trials each subject	Emotion recognition 1-50 Hz	Calculated features spectrogram	CWGAN	Doubled	NA	NA
[112]	SEED: 9 subjects, 62 channels, 3classes, 45 videos/subject SEED V: 16 subjects, 62 channels, 5 classes	Emotion recognition 1-50 Hz	Calculated features Differential entropy	cBEGAN	2000 1000	81.9% 54.3%	87.56 % 62.8%

3.4.3 Sliding window or overlapping window

O'shea et al. (2017) presented a novel end-to-end architecture that learns representations from raw EEG signal by CNN for the task of neonatal seizure detection [122]. Interpretation of neonatal EEG requires highly trained healthcare professionals, and it is limited to specialized units. They used overlapping window to augment 1389 seizures during 835 hours of EEG signal. Each trial split into 8s epochs with 50% overlapping to have more training sample for their proposed CNN. They obtained 97.1% accuracy; however, they didn't evaluate their result without overlapping or different shift lengths.

N. Kwak et al. (2017) used CNN for the robust classification of a steady-state visual evoked potentials paradigm [89]. They recorded EEG for the brain-controlled exoskeleton under ambulatory conditions. For generating more training samples, they used overlapping window. In their results, different shift lengths from 10 ms to 60 ms out of 2-s window were compared. They found the training samples with smaller shifts, performed much better than larger ones. The highest accuracy was 99.28% for 5-class visual evoked potential task.

For Schirrmeister et al. (2017), a key question was the impact of CNN training (e.g., training on entire trials or cropping within trials) on decoding accuracies [123]. The concept of overlapping window was pushed even further in this study: First, DA by overlapping windows share information was used to design an additional term to the cost function, which further regularizes the model by penalizing decisions that are not the same while being close in time. Second, redundant computations due to EEG samples being in more than one window were simplified, which ensured these computations were done once, thereby speeding up training. As a result, cropped training (segments of about 2 s length) increased the accuracy to 95% for CNN on high pass filtered data (The authors did not report the accuracies before DA).

Ullah et al. used a 1D-CNN for research on epilepsy detection [124]. The number of trials collected in this study was not enough to train the CNN. And obtaining a large-enough dataset during seizure activity was not practical. At the same time, the available, small dataset resulted in overfitting. To overcome this problem, the authors proposed 2 methods for DA: (Note that the EEG signal length in this dataset was 4097): (Sliding window of length 512, stride 64, leading to 87.5% overlap. Each of these windowed signals was treated as an independent instance. Therefore, each trial was divided to 57 sub-signals (Sliding window of length 512 with stride 128, leading to overlap of 75%, leading to 29 sub-signals).

The average accuracies were 96.45±0.13 and 95.40±0.35 using DA with 87.5% and 75% overlap, respectively (The authors did not report the accuracies before DA).

N Truong et al. (2018) used GAN for semi-supervised seizure prediction [125]. They generated extra samples to balance the Freiburg and CHB-MIT datasets. As a result, training sets are 10 times larger than original one by using overlapping window. The extra generated training dataset is by sliding a 30-s window along the time with different shift length. However, they didn't report the accuracy achieved by different shifting length.

They achieved 60.91% and 72.63% accuracy (without DA) and 74.33%and 75.33% (with DA) for Freiburg hospital and CHB-MIT, respectively when training GAN on individual subjects.

Majidov et al. (2019) proposed an efficient classification of Motor imagery EEG task by using CNN [126]. For DA, they used sliding window with different shifting length. However, their result lacks more details about DA.

Z. Mousavi et al. (2019) proposed a single-channel EEG-based automatic sleep stage classification (2 to 6 classes) algorithm which processes the raw signals in order to learn features and automatically diagnose sleep stages using CNN [127]. The lack of balance between the data of each class was challenging situation which caused biasedness of classification results and degraded accuracy. Therefore, they used overlapping technique to augment their dataset. The training set was 50% of the dataset included 7592 epochs (30s), however after DA, they had 24162 epochs (3s). They achieved to 93.55% accuracy for classification 6 classes of sleep stages. In addition, to evaluate the performance of the proposed DA, GAN was also implemented. However, according to their results, using GAN for the 6 sleep stages classification had achieved 72.33%, which is lower than overlapping window.

Avcu et al. (2019) developed an end-to-end CNN for seizure detection [128]. They strove to minimize the number of channels used (just 2 channels—Fp1 and Fp2) and compared that to the result with all channels. EEG data of 29 pediatric patients diagnosed with a typical absence seizure were included in this study. In total, the data contained 1037 minutes of EEG with 25 minutes of seizure data distributed among 120 seizure onsets. To overcome the imbalance in the dataset, they applied different overlapping proportions according to existence or absence of seizures. Namely, while shifting with 5 seconds (no overlapping) was implemented to create interictal class, 0.075 second shifting was used for ictal class to create balanced input for the CNN. The sensitivity for 2-channel was 93.3% and for 18-channel was 95.8%. However, the result of DA was not reported in this study.

Tayeb at al. (2019) developed three deep-learning models: LSTM, CNN, and RNN for decoding motor imagery [129]. This group used shifting window with 4s length to reflect the partial time invariance of the data and overcome the problem of overfitting. This cropping strategy increased the training dataset by a factor of 25. The CNN architecture showed better performance and achieved a mean accuracy higher than 84% over all the 20 participants. However, their result lacks more details about DA.

Also, we found more papers which segmented the dataset to create more training data: Chambon et al. (2017) segmented the input data to 30s segment to create more dataset for each class of sleep stage [86]. Tsiouris et al. (2018) used LSTM for the prediction of epileptic seizure [130]. To overcome unbalance problem of rare seizure event, the EEG segment from the interictal class were split into smaller subgroups of equal size to the preictal class. Tang et al. (2017), proposed CNN for the failure prediction [131]. To avoid multiple instances learning issue for their CNN, they used segmentation window to have sufficient new training dataset. The length of each segment was found by adaptive multi-scale sampling. Their result was improved from 70.9% (without DA) to 77.9% (with DA) on seizure dataset.

Although many studies used this method, there seems to be no consensus on the best overlapping percentage to use, e.g., the impact of using a sliding window with 10% overlap versus 90% overlap. Some studies tried different shifting length; but this issue still is not clear. For more information refer to Table 8.

3.4.4 Sampling

Oversampling: R. Manor et al. (2015), presented a CNN model for the use of single trial EEG classification in five category rapid serial visual tasks [132]. They used oversampling of the minor class (bootstrapping) to balance the dataset. They mentioned that although this method caused some overfitting on the minor class, however, it provided a more balanced classification performance in their experiment.

Drouin-Picaro et al. (2016), proposed a CNN model to classify saccades from frontal EEG signals to aim cursor control without the need for a separate eye tracking device in provide brain-computer interfaces [133]. In order to have a balanced dataset, horizontal saccades were sampled from without replacement so that the number of horizontal saccades in the dataset was the same as the highest number of vertical saccade (either up or down). The other vertical direction was then augmented by sampling from it with replacement, to make the number of data points in each direction equal. Hence, the dataset contained roughly 3000 examples of each saccade direction.

Supratak et al. (2017), used a CNN model, named DeepSleepNet, for automatic sleep stage scoring based on raw single-channel EEG. They extracted time-invariant features and used LSTM to learn transition rules among sleep stages automatically [134]. By duplicating the minority sleep stages in the original training set such that all sleep stage has the same number of samples they avoided overfitting.

Dong et al. (2017), proposed a Mixed NN for temporal sleep stage classification [135]. Because of the inherent imbalance in occurrence of the different sleep stages, the authors used oversampling to generate a new balance dataset which every sleep stage is equally presented.

Sors et al. (2018) used a CNN on raw single-channel EEG signal for scoring 5 class sleep stage [85]. They mentioned their dataset (SHHS) has a very imbalanced class distribution. In order to account for this, they tried cost-sensitive learning or oversampling but the overall performance using this approach did not improve.

Ruffini et al. (2019), randomly replicated subjects from the minority class to balance their classes [136]. Their proposed model helps for diagnosis derived from a few minutes of eye-close resting EEG signal collected at baseline idiopathic patients. They didn't compare the result with and without DA.

Sun et al. (2019) scored the sleep stage automatically. This study presents a stage-classification method based on a two-stage neural network [137]. The first, feature learning stage can fuse network-trained features with traditional hand-crafted features. A second, RNN stage is fully utilized for learning temporal information between sleep epochs and obtaining classification results. Oversampling was used to solve a serious sample imbalance problem. Sadly, the result lacked more details about DA.

Subsampling: Thodoroff et al. (2016), evaluated the capacity of a deep NN to learn robust features of EEG to automatically detect seizures [138]. They randomly subsampled the majority samples of the dataset to re-balance the ratio between seizure and non-seizure data (from 1000/1 to 80/20) which facilitate the training. However, because seizure manifestations on EEG are extremely variable both intra- and intra-patients, a second challenge was the overlack of data for each patient (average of 8 seizures per patient). They trained the CNN by using 0.5 s window instead on 1 s. Using transfer learning, the general representation of a seizure on other patients learned first and then they trained the model to the specific patient using the weights previously learned as initialization.

Sengur et al. (2019) employed deep feature extraction for focal EEG signals [139]. The deep features were extracted from spectrogram images using the AlexNet, VGG16, VGG19, and ResNet50 CNN models. The FC6 and FC7 activation layers were used for feature extraction resulting in 4096-dimensional feature vectors. The obtained feature vectors were used as input to various k-NN classification models. Random subsampling was performed as the DA technique (no other details were provided about the parameters). See Table 8 for more details about these studies.

3.4.5 Fourier Transform

J. Schwabedal et al. (2018) proposed a new method for augmenting EEG signals when attempting sleep-stage classification [140]. They focused on imbalanced dataset in transitional sleep stages, such as S1 and S3, which are rare events with respect to more stable stages such as wakefulness or Rapid Eye Movement (REM) sleep. Cost-sensitive learning [85], oversampling of the minority class [132, 134, 141], and subsampling the majority class [138, 142] are common techniques to address imbalanced classes. But the overall performance using these approaches resulted in some biases in prediction and did not improve the accuracy [56]. Therefore, they used Fourier Transform Surrogates to augment the EEG data. The complex Fourier components of a signal x_n can be decomposed into amplitudes a_n and phases φ_n:

$$x_n = a_n e^{i\varphi_n}$$

Under the assumptions of linearity and stationarity of the signal, they generated a new signal which is statistically independent from the original signal. This happened by randomizing the Fourier-transform phases [0, 2π] and then applying the inverse Fourier transform. The authors processed the CAPSLPDB sleep database, consisting of 101 overnight Polysomnography's (PSGs), using a CNN for 6 sleep stage classification. They then used the above method to balance and augment the database to achieve better generalization. They improved the mean F1-score by 7% for sleep-stage classification.

Zhang et al. (2019), proposed a novel DL approach with DA to improve classification of motor imagery EEG signals [82]. They applied the empirical mode decomposition on the EEG frames and mixed their intrinsic mode functions to create new artificial EEG frames, followed by transforming all EEG signals into tensors as input for the NN by complex Morlet wavelet s. Complex Morlet wavelets transformation of the EEG signals has been proved effective in recent motor imagery research, including tensor decomposition and wavelet-based combined feature vectors method. Their algorithm decomposes the original signals into a finite number of functions called intrinsic mode functions (IMFs). Each of these IMFs, represents a non-linear oscillation of the signal. Once the signal has been decomposed, it can be recovered by adding all IMFs and the residue without loss. The main idea in this study is that by mixing IMFs of the same class we can generate new samples from this class by preserving all intrinsic characteristics. This aims to decrease overfitting problem in training NN and eventually improves classification results. They used CNN and WNN (Wavelet Neural Networks) models to evaluate their results.

They found that magnifying two times the original training sets had highest mean value and better stability in CNN. And as for the WNN, the highest magnification was achieved by 5. The average of the accuracy for CNN was better than WNN. They evaluated their method on BCI competition dataset. By magnification factor 5, the CNN accuracy was 77.9% without DA and 82.9% with DA and WNN reached 88% without DA and 84.3% with DA. The relatively low computational efficiency of the WNN was the limitation in their proposed work. This group worked very well on DA details. They used two more big motor imagery datasets to evaluate their methods. They found that WNN has better classification performance and smaller loss than the CNN. However, each iteration of the WNN model takes almost five times as long as the CNN. And they speculated that it’s because the WNN lacks the consideration of parallel computing.

3.4.6 Recombination of Segmentation

Said et al. (2017) presented a joint compression and classification method for EEG and electromyogram (EMG) using a multimodal auto encoder [143]. They conducted their experiments on the DEAP dataset. It included the modalities of EEG, EMG, and multiple physiological signals recorded from 32 participants during 63 seconds at 128 Hz. During experiments, volunteers watched 40 music videos and rated them on a scale of 1 to 9 with respect to four criteria: likeness (dislike, like), valence (unpleasant to pleasant), arousal (uninterested or bored to excited) and dominance (helpless and weak feelings to empowered feelings). Signals were normalized and segmented into 6 seconds segments. EEG and EMG modalities contained 23040 samples of 896 features. They trained the multimodal auto encoder by adding zero values to one modality while keeping the original values for the other modality and vice-versa. Thus, one third of the training data was EEG only, another one third was EMG only, and the rest had both EEG and EMG data.

Zhang et al. (2019) used common spatial pattern (CSP) and CNN to detect seizures [144]. They first split each training EEG trial into three segments, and then generate new artificial trials as a combination of segments coming from various, randomly selected trials. They achieved 90% average accuracy, but did not report their multiplication factor or the accuracy before DA.

Dai et al. (2019) employed hybrid scaling CNN (HS-CNN) for motor imagery classification [145]. They varied the CNN kernel size between subjects and even between sessions. They found three kernel sizes for each selected frequency band: theta, mu, and beta. To improve the accuracy of HS-CNN, they used a 3-stage DA method: (1) segment each trial to 3 segments; (2) recombine the segments within different trials in the time domain; (3) swap frequencies: after band-pass filtering, the filtered trials (theta, mu, and beta) in the same frequency band were randomly swapped. Step2 and 3 were repeated multiple times for a multiplication factor of 3. The average accuracy for dataset 2b of BCI competition IV increased from 86% to 87.6%. They tried other DA techniques, such as noise addition and sliding window resulting in average accuracies of 86.1% and 80.1%, respectively.

3.4.7 Other

Frydenlund et al. (2015) used video and EEG data from subjects to estimate emotional response to music video (120 one minute music videos) [146]. To reduce computational cost, the researchers often throw away part of the signal by down sampling. In this experiment, authors reused the data thrown away during down sampling as new trials. Down sampling by a factor of N would therefore allow an augmentation of N times. However, the authors did not explicitly frame this as a DA method. So, no direct comparison was made of the accuracy with and without using the down sampled data.

Sakai et al. (2017), published a paper about DA methods for ML-based classification of bio-signals [88]. Their proposed DA methods for EEG signals includes: a) Shifting all-time data (±10ms) b) Amplifying all-time data (90% and 110%) c) Shifting near-peak value (±10ms) d) Amplifying near peak value (90% and 110%). Multiplication factors ranged from (±5% to ±50% every ±5% in b and d and ±5ms to ±5ms every 5ms).

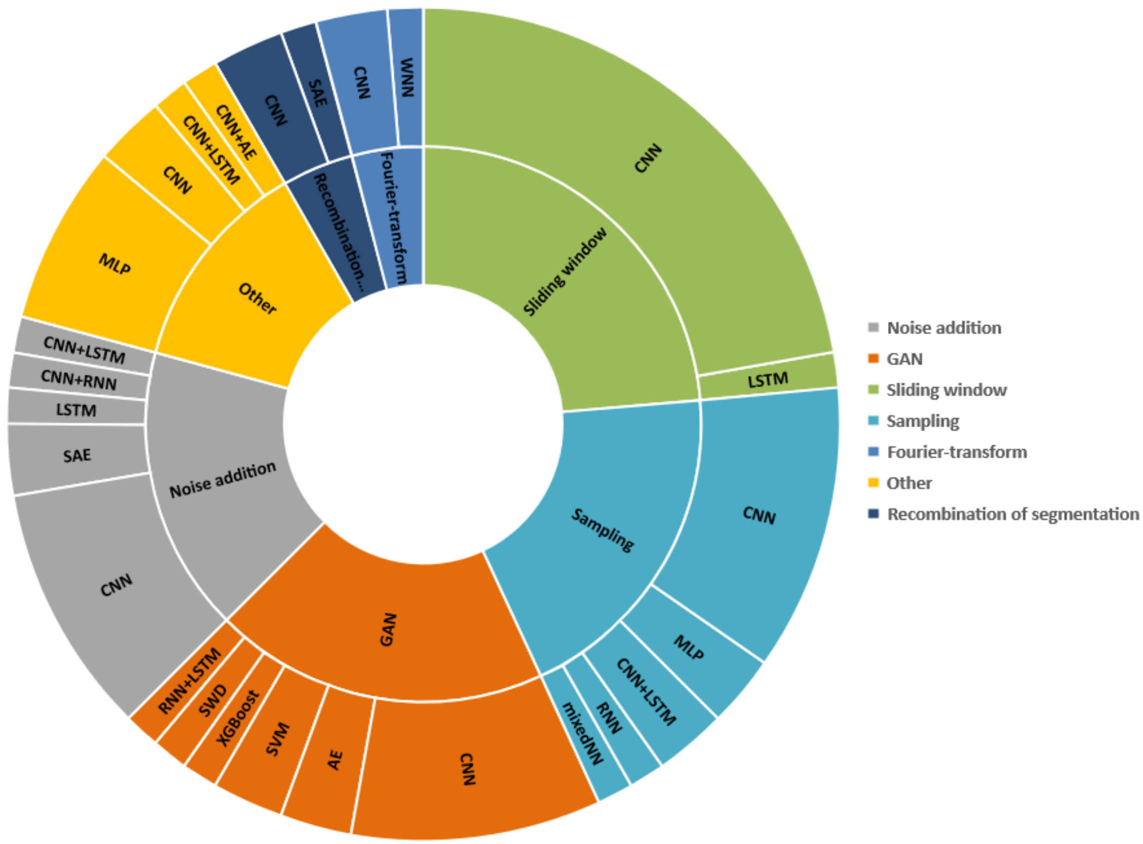

Figure 19. Data augmentation methods across all reviewed papers. The inner circle shows the general DA methods, and the outer circle shows the deep learning architecture strategy used.

Deiss et al. (2018) suggested swapping right and left electrodes to double the size of the dataset. They utilized a dataset of brain monitoring in an intensive care unit (ICU) for 5-way classification (Seizure, Lateralized Periodic Discharges (LPD), Generalized Periodic Discharges (GPD), Generalized Rhythmic Delta Activity (GRDA), Lateralized Rhythmic Delta Activity (LRDA)), and the last one corresponds to Other/Artifacts (O/A))on 155 patients [147]. The most challenging issue in their experiment was to make the model learn how to generalize to new patients. To simulate different patients, they kept three reference electrodes in the middle of the scalp unchanged and left/right flipped the remaining electrodes. Swapping electrodes in this manner doubles the amount of data. The authors reported that this DA method did not affect classification for tasks with symmetrical signals between the brain hemispheres (The authors did not report the accuracies before DA).

Shovon et al. (2019) applied STFT on EEG signals to transform signal to images for binary classification of motor-imagery signals [148]. They used rotation, flipping, zoom in and zoom out as DA techniques to overcome the overfitting problem in their proposed CNN model. Additional 1000 augmented images increased the average accuracy to 89.19% (no accuracy before DA was reported).

Freer et al. (2019) constructed a convolutional LSTM (C-LSTM) network based on filter bank common spatial patterns (FBCSP) for 4-way classification in a motor-imagery task [149]. The effects of several DA methods of data augmentation on different classifiers were explored, combining noise addition, multiplication, frequency shift, and phase shift. These DA methods improved the average overall accuracy of the classifiers by 5.3%.

Finally, Mokatren et al. (2019) applied the discrete-wavelet transform to extract energy and entropy of 4 frequency bands: theta, alpha, beta, and gamma in an emotion-recognition task [150]. A 3-D array of size KxKxB was created, where the first two dimensions represent an image of KxK pixels corresponding with the channels positioning over the scalp, while the third dimension represents the number of features: energy and entropy for 4 frequency bands(B=8). They used image augmentation techniques, such as horizontal and vertical shifting, to improve the accuracy of their CNN. Their classification accuracy on the DEAP dataset improved from 86.47% (without DA) to 90.87% (with DA) for Arousal and 88.34% (without DA) to 91.33% (with DA) for valence.

This section shows that these authors tried to improve the accuracy of their classification method with different techniques but because we found just one case from each of this innovation, we grouped them together. Figure 19 displays the aggregated information on DA methods and DL architecture strategy. While there is no clear consensus when looking to all 53 studies together, studies that employed sliding window, sampling, and noise addition as DA method, mostly used CNN. We investigated the EEG task compared across different DA techniques Figure 24. Following our review, we conclude that, for seizure task, the sliding window method should be used. For Mental workload, noise addition achieved the best results. And for deciphering sleep stages, the sampling method is the best fit. In sum, we recommend that sliding windows should be used for seizure detection. We also found that noise addition works best for mental workload. And the sampling method appears optimal to classify sleep stages.

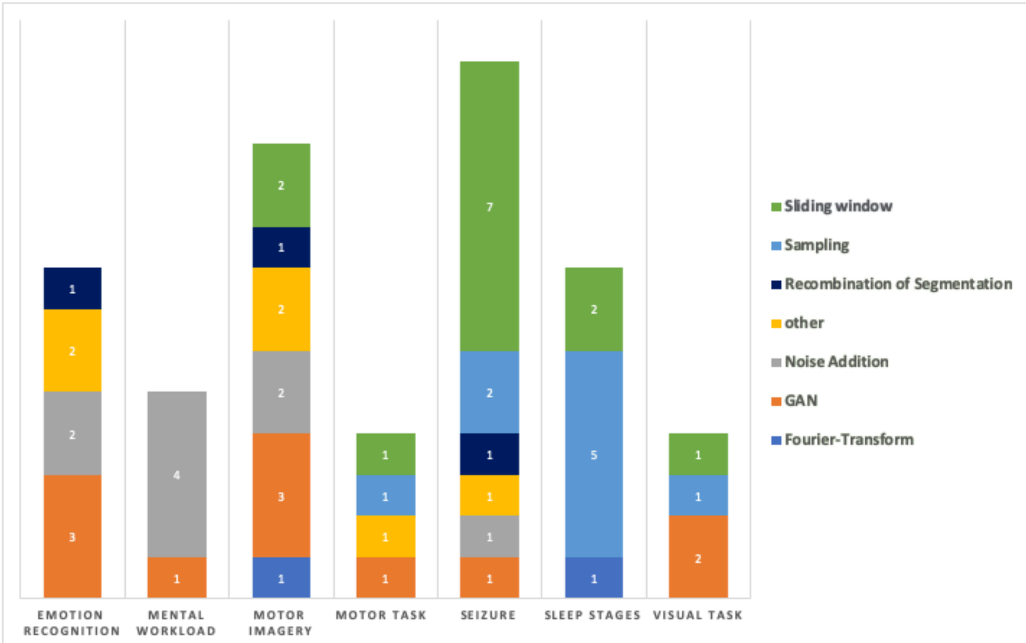

Figure 20. Number of papers for general EEG tasks compared across different DA techniques

3.5 Accuracy gains of data augmentation

The application of DA for DL on EEG is still nacent, with relatively few studies having been conducted. What is more, many of those studies unfortunately do not report the gain in accuracy that the DA method brought about or which parameters were used exactly (Table 8). Nevertheless, 29 of the 53 papers we surveyed included a measure of accuracy before and after DA. We therefore computed an improvement score on those for each DA analysis, $\frac{\boldsymbol{a}-\boldsymbol{o}}{\mathbf{1}-\boldsymbol{o}}$. Here, "$\boldsymbol{a}$" stands for the accuracy of the model when trained on the augmented dataset and "$\boldsymbol{o}$" stands for the accuracy on the initial, non-augmented dataset. Hence, an improvement score of $\boldsymbol{s}$ suggests that, by training also on the augmented dataset, a fraction $\boldsymbol{s}$ of the gap between initial accuracy and perfect accuracy was covered by the model traiedn on the augmented dataset. The overall improvement score was 0.29±0.08 (mean±s.e.m.). Though the score varied among the different DA techniques—from 0.08

for recombination of segmentation to 0.36 for noise addition (Figure 27 A). For tasks, it avaried from 0.14 for motor imagery to 0.56 for mental workload (Figure 27 B). The 95% confidence intervals for all tasks (except “visual task”) and DA tecniques did not include 0. It should be noted though that these statistics rely on relatively small number of analyses. And thus more studies are required to establish reliable DA imporvement score for different techniques and tasks.

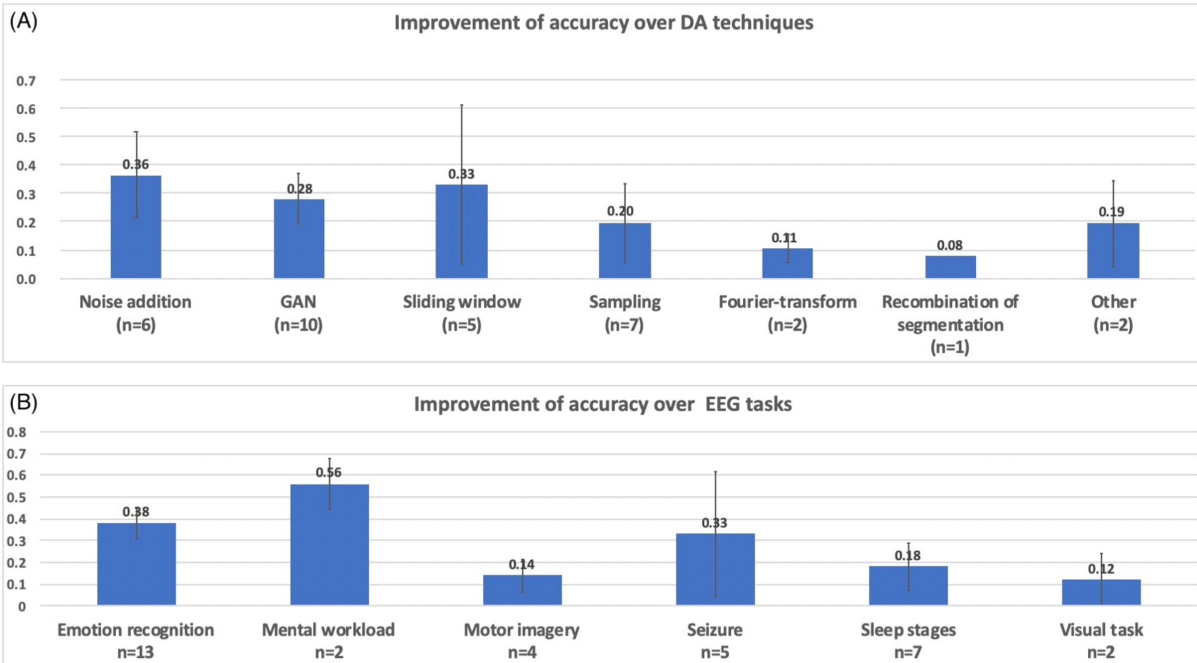

Figure 21. (A) The improvement score, or the fraction of variance left unexplained by the original DL method that was explained when training the model using DA, for different DA techniques (mean ± 95% confidence intervals). Here is the number of studies everywhere except GAN, Sampling, and Other, where there were more than one analysis, with different accuracies, reported in each study; hence there 'n' is the number of accuracies. (B) Same as A but the improvement is over EEG tasks. Here "n" is the number of studies everywhere except "emotion recognition", where there were 9 studies, 2 of which ran multiple DA analyses; hence "n" there is the number of analyses. No motor-task studies included accuracy before and after DA, so that task is not included in this figure.

3.6 Discussion

Here we review the most important findings from our results section and discuss the significance and impact of various trends highlighted in the results. We also provide some recommendations for the 7 tasks on which we analyzed DL_EEG: seizure detection, sleep stages, motor imagery, mental workload, emotion recognition, motor tasks, and visual tasks.

3.6.1 Rationale

The relatively small size of EEG datasets drastically decreases the effectiveness of DL. In the past few years, DA techniques have received widespread attention and achieved considerable performance gains for DL. Therefore, we focused our review on available DA methods for DL-based EEG. Such augmented datasets facilitate training more complex models, with more parameters, while at the same time potentially reducing overfitting. We only considered papers that focused on DA in DL-based analysis of EEG.

Previous review papers recommended that more targeted work be carried out to fully exploit the potential advantages of DL in EEG processing [56, 75]. It thus appears natural to explore the relation between performance and DA. Toward this goal, we carried out a systematic review of DA for DL-based EEG. Our goal was to address the following critical questions: (1) What DA approaches exist for EEG? (2) Which datasets and EEG classification tasks have been explored with DA? (3) Are specific DA approaches more suitable for particular tasks? (4) What input features are used for training deep networks with DA?

3.6.2 Data

A lingering critical question in machine learning is “how much data is enough data?”, and it is of special relevance when applying sophisticated DL techniques on limited size EEG datasets. Naturally, the amount of data is critical in achieving high DL performance. But, needless to say, the quality of the data is also very important. To analyze this, we looked at dataset features such as the number of subjects, amount of EEG recorded (in trials or time), and the DA schemes used. We found that, for noise addition, the best results were obtained when a lower standard deviation was used. However, more generally, we did not find one specific, definitive answer to the data quantity question. That said, our analysis clearly suggests that DA techniques are typically successfully able to increase the performance of DL (Table 8).

3.6.3 EEG pre-processing

Most studies used frequency-domain filters to limit the bandwidth of the EEG signals. This enabled them to focus on specific frequency ranges that were of interest (Figure 12). The filtered frequency ranges were organized by EEG task type. We found no studies that specifically tested the role of this filtering on NN. (This lacuna is discussed in other review papers [75].) The great majority of the reviewed papers preprocessed the EEG data before feeding it into NNs. Based on Figure 13, 49% of the reviewed papers used calculated features such as wavelet, entropy, spatial filter, or STFT as the input to NNs. On top of that, 36% simply used the raw EEG time-series signal as the only input to the NN. This is not surprising as a key motivation for using NN for EEG processing is to automatically learn features. An analysis of the sort that we carried out could in principle give some sense of which input types should be used for these purposes. But a complete answer depends on many factors, including the EEG task. And it is therefore difficult to draw definitive conclusions when only 53 studies using DL and DA are currently available.

3.6.4 Deep-learning methodology

Our analysis focused on architecture trends and input formulations for each architecture. However, the EEG task too is of importance, of course. CNN was the most popular NN architecture—likely because it is well suited for end-to-end learning, scales well to large datasets, and can exploit hierarchical structure in natural signals. The number of hidden layers in the different NN architectures varied case by case. Given the relatively small number of papers so far, we were able to aggregate information about DL architectures into a single figure, which we hope would help our colleagues gain some intuition about this nontrivial issue. Thus, the input formulation for RNN is images while for SAE it is calculated features. For LSTM, there are two input categories: signal and calculated features. CNN and MLP studies included instances of all input formulations, but the signal formulation was used most often for their inputs. There were also hybrid architectures that used a combination of two standard NN. The papers relying on such hybrid NNs commonly used calculated features and images as their inputs.

3.6.5 Data augmentation

Figure 14 is a testament to the importance of DA for EEG processing with DL. DA techniques have received widespread attention and achieved appreciable performance boosts for DL techniques on EEG. However, more work is required to clearly assess their advantages as well as their potential disadvantages. Here we covered all the available DA techniques that we could systematically source and grouped them into 7 categories: noise addition, GAN, sliding window, sampling, Fourier transform, recombination of segmentation, and other. Sliding windows, at 24%, was the most common. Nevertheless, there seems to be no consensus on the best overlapping percentage to use between consecutive windows—e.g., the impact of using a sliding window with 10% versus 90% overlap. Some studies tried different shifting lengths **[89]** [124] , but this issue remains unsettled.

We found two main approaches for adding noise to EEG signals for DA: (1) adding various types of noise (Gaussian, Poisson, salt, and pepper, etc.) to the raw signal; (2) converting the EEG signals to sequences of images (spectrograms) and adding noise to these images. Though it has been reported that adding noise to the images did not improve the classification accuracy [90]. Unfortunately, some authors did not provide details about the accuracy before and after DA. In a similar vein, critical noise parameters (e.g., mean, standard deviation, the magnification factor of training dataset) were sometimes not reported. This made it more difficult to compare techniques and parameters across studies.

Since 2018, GAN has become very popular for generating EEG signals that mimic real ones. Though GAN and related DL algorithms were used and discussed more for generating synthetic images for image classification tasks. EEG can often be analyzed and visualized in the frequency domain over time as spectrograms (through a Fourier or wavelet transformation). These spectrograms can then be treated like any other image, and therefore data augmentation methods that were developed for images can, at least in the technical sense, be directly applied to them. The spectrograms generated via the DA process is then converted back to an EEG signal of course.

While GAN data augmentation for EEG shows some improvement of classification accuracy, it has still not been clearly demonstrated to be better than other, simpler methods—like noise addition. For example, from our results (Table 8), it appears that the mean increase in accuracy when using GANs is 5.7% (STD 5%) while for noise addition, the increase is 14.2% (STD 13%).

The GANs that we covered here mostly learned the underlying distribution of the training EEG signals and generated fake signals that are within the distribution. In this sense, it might be argued that it is not that different from adding noise to the original signals. In addition, whether we can treat a spectrogram as an image and simply apply image-based data augmentation techniques remains an open question. First, the "pixels" in spectrograms relay temporal and frequency information that images do not. Second, at least when using CNNs, the invariant filters that work well across images would often not be expected to work on the spectrograms. For example, there is no a priori reason to expect to find the same pattern in high gamma early in time and in theta later. Third, when using real images, the developer of a GAN can rely on her visual system to judge how well the GAN works for generating fake images that are like the real ones on which it is based. However, the same cannot be said for a GAN developed for EEG. What is more, we know too little about the characteristic of EEG for specific tasks (certainly across subjects and variants of the task) to develop a method that would judge the quality of an EEG GAN. So, this technique should be used with proper caution.

Fourier transform was used in 2018 to augment EEG signals, very successfully. This method assumes linearity and stationarity of the EEG signals [140] . In 2019, Zhang and colleagues used these intrinsic features of EEG and decomposed the signal to its IMFs. By mixing IMFs, they generated new samples and decreased overfitting [82]. Sampling was used in many studies to better balance imbalanced datasets. Balancing the number of samples among classes may drastically improve the usefulness of a dataset.

Figure 20 enables us to draw a few trends. We see that the sliding-window technique is used for the majority of papers that analyze seizure detection. Similarly, we found that noise addition is the most common technique for mental workload. And sampling methods appear most often for classifying sleep stages. To what extent the more popular techniques are also more optimal is still unclear given the relatively small number of papers so far.

But how useful is DA for DL-based EEG analysis? How much does it improve classifier accuracy? Before delving into the numbers, it is also worth bearing in mind that publication bias might be driving these numbers up. Unsuccessful attempts at improving DL accuracy with DA may be less likely to be published. That said, on average, training on the DA dataset helped make up almost 3/10 of the gap that was left in accuracy between the original analysis and perfect accuracy. Though this improvement score varied widely among DA techniques and tasks (Figure 21). Too few studies reported both accuracy before and after DA and the parameters of their DA method for us to be able to carry out more in-depth statistical analyses. Those will be possible with additional publications.

3.6.6 Guidelines for reporting results in papers

Some papers clearly explained their methodology with respect to DA (e.g., [149]). Unfortunately, these were the exception. Of the reviewed papers, 45% did not report the accuracy before DA and 38% did not report the parameters they modified in their DA method. It is also noteworthy that 41% did not mention the magnification factor they used. This made surveying and comparing the literature rather difficult.

Therefore, in order to improve the quality and reproducibility of the work in the field of DA on DL-based EEG, we recommend that authors follow the guidelines below when reporting their results in their studies.

Clearly describe the data augmentation method they used	Method, parameters they change in their method, Magnification factor
Clearly describe the dataset	#subject, #trials, #classes
Test their proposed method on an existing dataset	Compare model performance and evaluate their results on a public dataset
Clearly describe the architecture	#layers, their widths, the activation functions used
Report the accuracy and results	Report the accuracy before and after using DA, report the results when changing the parameters of DA
Share internal recording and reproducible code	Whenever possible, including hyperparameter choices

3.7 Limitations

One clear limitation of our study is the relatively small number of papers published so far on this topic. This procludes us from carrying out mode detailed analyses than the above. What is more, another obvious limitation of our methodology, already discussed above, is that our analysis is only as good as the data on which it is founded. When little information is provided about the DL or DA methods, it directly and immediately limits our ability to analyze those data, as discussed above.

In addition, although the search methodology we used to identify relevant studies is well-founded, it undeniably did not capture all of the existing literature on the topic. Since the field of DA for DL-based EEG is still young and the number of publications available at the time of writing this manuscript was limited, we decided to include all the papers we could find (note that some of the newer trends are more visible in repositories such as arXiv and bioRxiv, as those manuscripts may be going through the publication process). They have been adopted by the DL community to quickly disseminate results and encourage a fast research-iteration cycle. Our goal was to provide a transparent and objective analysis of the trends in DA for DL-based EEG.

We focused our analysis on the points that we thought would be most interesting, valuable, and impactful for the performance of DA on DL-based EEG. Therefore, we didn't include normalization procedures, software toolboxes, loss function, training time etc., in this analysis.

3.8 Conclusions

DL has been successfully applied to many EEG tasks such as: sleep stages, motor imagery, mental workload, and emotion recognition tasks. Applying DL to EEG has shown great promise in processing these complex signals due to its capacity to learn good feature representations from raw data through successive non-linear transformations. However, DL is inherently limited over EEG datasets because of their relative smaller size. DA, in turn, increases the available training data, facilitating the use of more complex DL models. It can also reduce overfitting and increase the accuracy and stability of the classifiers.

Looking at the inputs to the DL architectures, the most common technique is still to calculate features (49%) outside the NN and feed it into the network, though a sizable fraction of papers input the raw signals (36%) into the NN and let it extract features itself. In addition, while various architectures have been used successfully on EEG datasets, CNN is most often used (62%). Taking all of the above into account, our analysis of the literature suggests that DA was mainly used for seizure detection(24%) and motor imagery(21%). In particular, sliding windows are favored for seizure detection. Noise addition is most common for mental workload. And sampling methods are the procedures of choice to classify sleep stages.

Our attempt to compare results between different studies highlighted for us the high degree of variability in how results were reported across studies. We therefore made specific recommendations to ensure reproducibility and better comparison of the results when the authors use DA and DL. It is key to clearly describe the DA method, its parameters and their role in achieving the accuracy that the paper boasts. It is also critical to report the magnification factor as well as the accuracy before and after DA.

In sum, we hope this review will constitute a good entry point for EEG researcher looking to apply DA for training DL algorithms on their datasets and will assist the field to produce high-quality, reproducible results.

Table 8. Details of all the papers that we found for our review paper. In Data Augmentation column: NA: Noise addition, SW: Sliding window, S: Sampling, FT: Fourier-transform, Recombination of Segmentation: RS, O: Others and in EEG task column: ER: Emotion recognition, MW: Mental workload, MI: Motor imagery, S: Seizure, SS: Sleep stages, IT: Imagery task, VT: Visual task and in input formulation column: S: signal, I: Images and CF: Calculated features. Some studies used different dataset or different DA techniques and we show them separately.

Study	Data Augmentation	EEG task	# Subjects	# Trials	# Channels	# Classes	Frequency Range	Input Formulation	Type (input formulation)	NN Architecture	Hidden Layers	Last Layer (# Classes)	Activation Function in	Activation function in Output	Accuracy (before DA)	Accuracy (after DA)	Dataset
[90]	NA	MW	13	2670 trials	64		4-13 HZ	I	FFT	CNN+LSTM	conv (7)	FC(512)	Relu	softmax	NA	NA	University of Memphis Institutional review board
[91]	NA	MW	8	180 min/subject	1	2	1.5-40 HZ	CF	power spectral	SAE	hid(5)	FC(2)	Relu	softmax	NA	0.93	SDAE
[92]	NA	MW	7	1h	11		1-40 HZ	CF	FFT, power	SAE	hid(6)	FC(2)	NA	NA	0.342	0.75	AUTOCAM

									spectral								
[78]	NA	ER	14	1890 trials	62	3	1-50 HZ	CF	Entropy	CNN	conv (4)	FC(3)	Relu	softmax	0.49	0.74	SEED
[78]	NA	ER	14	1890 trials	62	3	1-50 HZ	CF	Entropy	CNN	conv (13)	FC(3)	sigmoid	NA	0.765	0.855	SEED
[78]	NA	ER	30	527 trials	32	3	1-50 HZ	CF	Entropy	CNN	conv (13)	FC(3)	sigmoid	sigmoid	0.408	0.45	MAHNOB-HCI
[93]	NA	S	5	16.3 h	100	2 & 3 & 5	0.53-40	S	Raw	LSTM	LSTM(1)	Softmax	ELU	softmax	subject by subject	increased	Bonn University
[87]	NA	MW	22	6490 trials (8h)	64	4	4-30HZ	CF	power spectral	RNN+CNN	hybrid=conv(9)+LSTM (1)	FC(4)	Relu	softmax	NA	0.93	NIMHANS
[94]	NA	ER	32	40(video 1 min)	32	2 & 4	[1,50]U[60,end]	CF	spatio-temporal	CNN	conv (2)	softmax	NA	sigmoid	0.79 0.79	0.88 0.87	DEAP

				/subject													
[95]	NA	MI	5	5 sessions	3	2	0.5-100 Hz	CF	Raw	CNN	Conv(4)	FC(2)	ELU	softmax	NA	NA	BCI competition IV 2b
[96]	NA	MI	9	72 trials /task	22	4	7-125 HZ	CF	spatio-temporal	CNN	conv(1)	FC(4)	Relu	sigmoid	0.709	0.779	BCI competition IV 2a
[96]	NA	MI	14	880 trials (160 trials include for test and train)	128	4	7-125 HZ	CF	spatio-temporal	CNN	conv(1)	FC(4)	Relu	softmax	NA	NA	High Gamma dataset(HGD)

[102]	GAN	MW	30	70000 trials/subject	1	2	NA	CF	power spectral	XGBoost	Unspecified	Unspecified	logestic sigmoid	softmax	0.90	0.95	NA
[101]	GAN	MI	1	280	3	2	7-15HZ	I	spectrogram	CNN	Unspecified	Unspecified	Relu	softmax	0.83	0.86	BCI compititionII dataset III
[105]	GAN	MT	1	438	1	2	8-13HZ	I	spectrogram	wasserstein distance(SWD)	Unspecified	Unspecified	leaky Relu	softmax	NA	Did not improved results	not mentioned
[109]	GAN	ER	15	3394 per subject	62	3	1-50HZ	I	spectrogram	CWGAN	Unspecified	Unspecified	Relu	softmax	0.839% Arousal 0.69%, Valence	0.869% Arousal 0.78%, Valence	SEED

															0.53 %	0.73 %	
[109]	GAN	ER	32	2400 per subject	32	2	1-50HZ	I	spectrogram	CWGAN	Unspecified	Unspecified	softmax	softmax	NA	NA	DEAP
[103]	GAN	MI	1	400trials	3	2	NA	I	spectrogram	CNN	Unspecified	Unspecified	NA	softmax	0.77	0.79	BCI cometitionIV dataset 2b
[57]	GAN	ER	9	45video/subject	62	3	1-50Hz	CF	Entropy	cBEGAN	Unspecified	Unspecified	Relu	NA	0.81	.87	SEED
[57]	GAN	ER	16	48 experiment	62	5	1-50Hz	CF	Entropy	cBEGAN	Unspecified	Unspecified	Relu	NA	0.54	0.62	SEED V
[113]	GAN	S	23	5085 trials	>23	2	NA	S	Raw	CNN	Conv(5)	FC(2)	Relu	softmax	0.81	0.84	CHB-MIT
[114]	GAN	ER	18	32000 trials	14	2	NA	S	Raw	GAN	NA	NA	NA	NA	0.97	0.98	Not public

[115]	GAN	MI	9	500trials/subject	32	2	0.5-100 Hz	S	Raw	GAN	CNN+LSTM	Unspecified	Relu	softmax	NA	0.76	BCI cometitionIV dataset 2b
[116]	GAN	VT	10	5 hours/subject	256	2	0.1-55Hz	S	Raw	CNN	Conv(2)	Unspecified	NA	NA	0.50 (cross subject) 0.62(same subject)	0.53 (cross subject) 0.627(same subject)	BCIT X2 rapid series visual presentation
[121]	GAN	VT	1	50trials/class	20	3	9-60Hz	S	Raw	CNN	Conv(1)	FC(3)	Relu	softmax	NA	NA	Video-Stimuli Dataset for augmentation

																	, NAO Dataset for testing
[121]	GAN	VT	3	50 offline trials/class+ 30 online trials/class	20	3	9-60Hz	S	Raw	CNN	Conv(1)	FC(3)	Relu	softmax	0.91 for S1; 0.87 for S2; 0.84 for S3; 0.69 for across subjects	DCGAN : 0.97 for S1, 0.93 for S2, 0.87 for S3; VAE : 0.73 across	Video-Stimuli Dataset for augmentation, NAO Dataset for testing

																subjects	
[125]	SW	S	13	311.4h	6	2	57-63 &117-123	CF	STFT+GAN	GAN+CNN	conv(3)+conv(3)	FC(256)	sigmoid	softmax	0.60	0.72	Freiburg Hospital intracranial EEG dataset
[57]	SW	S	13	209h	22	2	47-53 & 97-103	CF	STFT+GAN	GAN+CNN	conv(3)+conv(3)	FC(256)	NA	NA	0.74	0.75	CHB-MIT database
[122]	SW	S	18	835 hours 1389 seizures	100	2	0.5-12.8	CF	spatio-temporal +information theory	CNN	conv(6)	Softmax	sigmoid	softmax	NA	0.97	Neonatal intensive Care Unit of Cork Univ

																	ersity(NICU) Maternity Hospital
[89]	SW	VT	7	varied	8	5	4.0-40	S	Raw	CNN	conv(2)	FC(5)	ELU	softmax	subject by subject	Increased 0.99	Due to ethical restrictions imposed by Korea University Institutional Revi

																	ew Board, data cannot be made publicly available
[123]	SW	MT	NA	288 trials /subject	NA	4	4-end	CF	spatial filter	CNN	conv (4)	FC(4)	Relu	softmax	NA	0.84	BCI competition IV-2a
[123]	SW	MI	NA	880 trials /subject	NA		4-end	S	Raw	NA	NA	NA	Relu	softmax	NA	0.95	High Gamma Dataset
[124]	SW	S	10	4090 trials	100	2 & 3	NA	S	Raw	CNN	conv (3)	FC(2)	NA	NA	NA	0.96	Bonn Univ

																	ersity
[124]	SW	S	10	4090 trials	100	2 & 3	NA	S	Raw	CNN	conv (3)	FC(2)	NA	NA	NA	0.95	Bonn University
[125]	SW	S	13	311.4h	6	2	57-63 &117-123	CF	STFT	CNN	conv (3)	FC(2)	NA	FC(10)	NA	NA	Freiburg Hospital intracranial EEG dataset
[125]	SW	S	13	209h	22	2	47-53 & 97-103	CF	STFT	CNN	conv (3)	FC(2)	Relu	soft max	NA	(+7% incraesed)	the Boston Children's Hospital -

																	MIT scalp EEG dataset,
[125]	SW	S	2	15.5 h	16	2	47-53 & 97-103	CF	STFT	CNN	conv (3)	FC(2)	Relu	softmax	0.89%%	0.889	American Epilepsy Society Seizure Prediction Challenge dataset(kaggle)
[126]	SW	IT	9	NA	22	2	8-36HZ	CF	spatial filter +PS	CNN	NA	FC(100)	Relu	softmax	NA	NA	BCI-2a and 2b

									D+covariance								competition IV
[126]	SW	IT	9	NA	3	2	8-36HZ	CF	spatial filter	CNN	NA	FC(100)	Relu	softmax	NA	0.75	BCI-2b
[127]	SW	SS	NA	15188trials	4	2-6 class	NA	S	Raw	CNN	conv (9)	FC(100)	NA	NA	0.829	0.857	Sleep EDF bank (Physionet database)
[86]	SW	SS	NA	62 nights	20	5	0-30HZ	CF	spatial filter	CNN	conv (3)	softmax	Relu	softmax	NA	0.812	MASS
[128]	SW	S	29	1037 minutes	2 & 18	2	0.5-70HZ	S	Raw	CNN	conv (4)	FC(2)	Relu	Relu	NA	0.93 for 2 channels	KK women and children

																0.95 for 18 channels	hospital
[129]	SW	MI	20	750 trials	3	2	2-60HZ	I	Spectrum	CNN	conv (3)	FC(2)	Relu	softmax	NA	0.84	Public dataset and BCI competitionIV dataset 2b
[130]	SW	S	24	980h	18	2	1-110 HZ	CF	spatio-temporal +Graph theory+correlation	lstm	LSTM(1) & LSTM(2)	FC(2)	NA	NA	0.70	0.78	CHB-MIT database

[131]	SW	S	2	15min	varied	2	NA	S	Raw	CNN	conv (3)	FC(2)	NA	NA	NA	NA	kaggle
[132]	S	VT	12	NA	64	2	0-51HZ	S	Raw	conv	conv (3)	FC(2)	NA	NA	0.8399	0.8696	Hebrew university
[133]	S	MT	NA	3000 trials	2	4	0.1-36HZ	CF	spatial filter	MLP	hid(3)	FC(40)	Relu	soft max	NA	0.82	MAHNOB HCI-Tagging database
[133]	S	MT	NA	3000 trials	2	4	0.1-36HZ	CF	spatial filter	CNN	conv (2)	FC(10)	Relu	soft max	NA	0.93(6class)	MAHNOB HCI-Tagging database
[134]	S	SS	62	58600tria	20	5	0.3-100 HZ	S	Raw	CNN	conv (4)	soft max	NA	NA	Arousal 0.69	Arousal 0.78	Advanced

				ls(30s)											%, Valence 0.53 %	%, Valence 0.73 %	research in sleep medicine of the hopital du sacre coeur de montreal
[134]	S	SS	20	41950trials(30s)	2	8	0.3-100HZ	S	Raw	CNN	conv (4)	softmax	NA	NA	0.776	0.793	Sleep-EDF
[135]	S	SS	62	494 hours	20	5	0-30HZ	CF	STFT	mixedNN	MLP +RNN	softmax	Relu	softmax	0.893	0.901	NA
[85]	S	SS	NA	5793 trials	2	5	NA	S	Raw	CNN	conv (12)	FC(5)	sigmoid	softmax	0.60	0.74	SHHS(SHHS-1)

[136]	S	SS	121	NA	14	2	0.3-100 HZ	I	spectrogram	CNN	conv (2)	FC(2)	NA	sigmoid	NA	0.70	Advanced research in sleep medicine of the hopital du sacre coeur de montreal
[136]	S	SS	122	NA	14	2	0.3-100 HZ	I	spectrogram	NN	FC(1)	FC(2)	Relu	softmax	NA	NA	Advanced research in sleep medi

																	cine of the hopital du sacre coeur de montreal
[136]	S	SS	123	NA	14	2	0.3-100 HZ	I	spectrogram	RNN	LSTM(3)	FC(2)	Relu	softmax	NA	NA	Advanced research in sleep medicine of the hopital du sacre coeur de

																	montreal
[137]	S	SS	20	43836 trials	2	5	0.3-35	CF	Energy, Power, Window deep belief window	LSTM	LSTM(2)	FC(5)	NA	softmax	0.846	0.855	Sleep-EDF and Sleep Apnea
[138]	S	S	23	NA	23	NA	0-49HZ	CF	spatio-temporal	CNN+LSTM	conv(4)+LSTM(1)	FC(64)	Relu	softmax	NA	NA	CHB_MIT
[139]	S	S	NA	10240 trials	NA	2	0.5-150 Hz	I	Spectrogram	CNN	Conv(4)	FC(2)	Relu	softmax	NA	0.99	www.upf.edu
[140]	FT	SS	NA	8h	16	6	0-13 HZ	S	Raw	CNN	conv(5)	softmax	sigmoid	softmax	0.72	0.75	CAP SLP DB

[82]	FT	MI	5	240 Trials/subject	14	2	8HZ-30HZ	CF	spatio-temporal	CNN	conv(2)	FC(2)	Relu	softmax	NA	.9	Not available
[82]	FT	MI	5	240 Trials/subject	14	2	8HZ-30HZ	CF	spatio-temporal	WNN	FC(2)	Relu	softmax	FC(2)	NA	0.85	Not available
[82]	FT	MI	1	280 trial	3	2	8HZ-30HZ	CF	spatio-temporal	CNN	conv(2)	FC(2)	Relu	softmax	0.88	0.82	BCI competition II, dataset III
[82]	FT	MI	1	280 trial	3	2	8HZ-30HZ	CF	spatio-temporal	WNN	FC(2)	Relu	softmax	FC(2)	0.88	0.84	BCI competition II, dataset III
[146]	O	ER	22	120 min	32	NA	4-100 HZ	CF	ICA/PCA	NN	hid(2)	Softmax	NA	NA	0.408	0.454	DEAP

[88]	O	MT	5	20 to 60/subject	1	2	1-30HZ	S	Raw	NN	hid(2)	NA	Relu	softmax	NA	0.97	NA
[88]	O	MT	5	20 to 60/subject	1	2	1-30HZ	S	Raw	NN	hid(2)	NA	sigmoid	sigmoid	NA	0.99 static 0.94 ambulatory conditions	NA
[88]	O	MT	5	20 to 60/subject	1	2	1-30HZ	S	Raw	NN	hid(2)	NA	ELU	softmax	NA	0.92	NA
[88]	O	MT	5	20 to 60/subject	1	2	1-30HZ	S	Raw	NN	hid(2)	NA	NA	NA	NA	NA	NA
[147]	O	S	155	~24 hours /subj	19	5	0-60 HZ	CF	Entropy	Hybrid(CNN+AE)	conv(6)	FC(5)	Relu	softmax	NA	0.814	Neuroscience ICU

																	at Massachusetts
[148]	O	MI	1	280	3	2	NA	CF	STFT	CNN	conv(6)	FC(2)	Relu	softmax	NA	0.89	BCI competition II ,datset III
[148]	O	MI	9	400	3	2	NA	CF	STFT	CNN	conv(6)	FC(2)	Relu	softmax	NA	NA	BCI competition IV, dataset 2b
[149]	O	MI	9	4session 72trials/session	3	4	7-30Hz	S	Raw	CNN+LSTM	varied	varied	varied	varied	NA	Overal 5.3% improved accuracy	BCI competition IV, dataset 2a

[150]	O	ER	32	40minutes/subject	32	2	4-45Hz	CF	Wavelet	CNN	conv(2)	FC(2)	Relu	Relu	Arousal:0.86 Valence: 0.88	Arousal: 0.91 Valence: 0.91	DEAP
[143]	RS	ER	32	23040trials	32	4	NA	S	Wavelet	SAE	Unspecified	Unspecified	Exponential Linear Units	softmax	NA	0.6875	DEAP
[89]	RS	S	23	Varied for each subject	18	2	5-50Hz	CF	Wavelet	CNN	Conv(2)	FC(2)	Leaky Relu	softmax	NA	0.9	CHB-MIT
[145]	RS	MI	9	6520 trials	3	2	4-32Hz	S	Raw	CNN	Conv(2)	FC(2)	ELU	softmax	0.86	0.87	BCI competition IV,

																	datas et 2b

4 An end-to-end CNN with attentional mechanism applied to raw EEG in a BCI classification task

4.1 Introduction

Advances in brain science and computer technology in the past decade have led to exciting developments in Brain-Computer Interfaces (BCI), thereby making BCI a key research area in applied neuroscience and neuro-engineering [151]. Non-invasive BCI facilitates new methods of neurorehabilitation for physically disabled people (e.g., paralyzed patients and amputees) and patients with brain injuries (e.g., stroke patients) [151]. BCI systems utilize recorded brain activity to directly communicate between the brain and computers to control the environment in a manner compatible with the individual's intentions [152].

However, the ability to decode intentions is also an important tool for basic neuroscientific research. In particular, it strongly enhances the scientific armamentarium used to investigate volition [153, 154]. And, more specifically, decoding intention in real time would open the door to interesting experimental possibilities, such as interventions to facilitate or frustrate intentions [13, 155, 156], and intention-contingent stimulation [153]. Technological advances of recent decades—such as untethered, wireless recording, machine-learning-based analysis, and real-time analysis of raw EEG signal have increased the interest in electroencephalography (EEG) based BCI approaches [157].

EEG has proved to be the most popular brain-imaging method for BCI because it is inexpensive, noninvasive, directly measures neural activity (as opposed to fMRI for example), and can facilitate portability to clinical use [152]. EEG signals thus serve as pathways from the brain to various external devices, resulting in brain-controlled assistive devices for disabled people and brain-controlled rehabilitation devices for patients with strokes and other neurological deficits [151, 158, 159]. One of the most challenging topics in BCI is finding and analyzing the relations between recorded brain activity and underlying models of the human body, of biomechanics, and of cognitive processing. The investigation of relations between EEG signals and—real and imagined—upper limb movement has gained more attention in recent years [160, 161].

To implement an EEG-based BCI system for a particular application, a specific experimental protocol and paradigm must be chosen for all phases of the experiment. Typically, the participant first performs a particular task (e.g., a motor-imagery task, a visual task) to learn how to modulate their brain activity, while EEG signals are simultaneously recorded from their scalp. Using the recorded EEG as training data, a machine-learning-based neural decoder for the paradigm is then constructed [151]. Finally, the participant performs the task again, and the neural decoder is used for BCI control.

The process for BCI systems based on motor imagery (MI) is similar. Though, in this case, the participant imagines the movement rather than actually executing it [160]. Previous studies have confirmed that imagination activates areas of the brain that are responsible for generating actual movement [151, 162]. The most common MI paradigms reported in literature are based on sensorimotor rhythms (SMR) and imagined body kinematics. In the SMR paradigm (e.g., [163, 164] participants imagined kinesthetic movements of some body part—such as hands, feet, or tongue—which result in modulations of brain activity that are trackable using EEG [165]. Imagined movement in such SMR paradigms often causes event-related desynchronization (ERD) in mu (typically 8-12 Hz) and beta rhythms (roughly 12-30 Hz). In contrast, relaxing after MI results in event-related synchronization (ERS) [166]. The ERD and ERS modulations are most prominent in EEG signals acquired from electrode locations C3 and C4 (in the 10/20 international system); these electrodes are approximately above the motor cortices of both brain hemispheres.

MI classification is one of the most popular EEG-based BCI paradigms. EEG MI classification generally consists of four parts: signal acquisition, feature extraction, classification, and control. Most existing feature-extraction methods depend on manually designed features, based on human knowledge. Feature extraction and classification of EEG signals for MI tasks have been attempted in the time, frequency, and space (electrodes) domains—not necessarily mutually exclusively. Time-frequency feature extraction in EEG has focused mostly on short-time Fourier transform [167, 168] or wavelets [169, 170]. In the space domain, filter-bank common spatial-patterns (FBCSP) has achieved notable performance [171, 172]. However, FBCSP uses a fixed temporal duration, ignoring difference between participants. As such, it does not make full use of time-domain information. Moreover, these methods generally use handcrafted features and require heuristic parameter setting—e.g., predefined frequency bands—which often do not generalize well across tasks and participants [145]. As such, they often result in limited classification accuracy [169, 173, 174].

4.2 Related work

Recently, researchers have successfully used deep learning (DL) to perform automatic feature extraction [175] and classification [123, 145, 176, 177]. DL has achieved breakthrough accuracies and discovered intricate structures in various complex and high-dimensional data [178, 179]. In particular, it has provided promising results in the analysis and decoding of EEG signals [28]. Thus, NN architectures, their training procedures, regularization, optimization, and hyperparameter settings are all active area of research in DL-based analysis of EEG, with advances often resulting in dramatic increases in decoding accuracy [28].

Recently, Zhang et al., proposed a hybrid DL architecture, which combined convolutional neural networks (CNNs) and long short-term memory (LSTM) models to handle sequential time domain data [180]. Even more recently, Dai et al., proposed an architecture composed of a CNN with a hybrid convolution scale (HS-CNN), which separates a signal into three frequency bands using bandpass filters at 4∼7 Hz, 8∼13 Hz, and 13∼32 Hz. The three frequency bands are then fed into the convolutional layers with different filter sizes [145]. The features, including different semantic information, were concatenated and then MI classification was carried out. In another study, Zhang et al., applied an attention module to LSTM to utilize long-range information for EEG-based hand-movement classification [181].

Despite their promise, these deep NN architectures are not easy to train from scratch, because they require large amounts of training data to achieve high classification accuracy. However, it is particularly challenging to obtain a large amount of training samples for MI classification. This is because gathering high-quality data requires training and experience as well as a state-of-the-art EEG machine and a noise-free environment. MI tasks are also time consuming and fatigue-inducing for the participants. For example, during the task, participants must minimize, if not altogether avoid, eye movements and other muscle contractions, especially around the head. At the same time, they typically need to employ a great deal of concentration and attention during MI tasks. Thus, participants can only produce a limited amount of data at each session and must come in for multiple sessions to construct a large dataset of EEG MI. This often results in attrition over the course of multiple sessions.

Data augmentation (DA) can lead to considerable performance gains for DL, reducing overfitting and increasing overall accuracy and stability. DA generates new samples to augment an existing dataset by transforming existing samples in some systematic manner. Exposing the classifiers to various transformations of the training samples, as DA does, makes the models more robust and invariant to these and potentially other transformations when attempting to generalize beyond the training set [79, 82, 182].

DA is an especially important technique for EEG-based BCI because of its specific combination of two factors: the dimensionality of EEG signals tends to be high, while the number of available training samples tends to be low. In a recent systematic review on DA in EEG, Lashgari et al. collected all the papers that used DA for NN-based analysis of EEG up to and including 2019 [28]. They showed that convolutional neural networks (CNN) were the most popular NN architectures for EEG MI classification and typically resulted in accurate decoding. This is likely because CNNs are well suited to end-to-end learning, scale well to large datasets, and can exploit hierarchical structure in natural signals. The review also found that the most common input formulation for motor tasks and MI was raw EEG signals [28].

With these elements in mind, here we investigated the efficacy and generalizability of deep learning on EEG-based decoding of MI. We designed an end-to-end CNN with an attentional mechanism [183]. This is because a CNN with an attention-mechanism architecture can improve classification performance using EEG signals by focusing on essential, task-relevant features on different time-steps.

We begin by testing this architecture on 2 benchmark datasets (BCI Competition IV 2a and 2b) as well as on the dataset that we collected, which we share with the community. Then, we compare MI to ME on the dataset that we collected. Next, we tackled a common question when collecting EEG data: how many channels to record for optimal decoding accuracy? We thus compared the decoding accuracy for different numbers of channels. It has also been demonstrated that DA techniques hold promise for EEG decoding. So, we also tested how much DA can boost the accuracy of our method across the datasets. How much EEG data is needed to train deep NN is also not well understood, especially in relation to DA techniques. We therefore next investigate how the accuracy of our model depends on the amount of data on which we train and the type and amount of DA we use. Of course, structure and anatomical features vary across brains. So, we further investigated what happens to the decoding accuracy when we train and test it on EEG from single participants, on pair of participants, triplets, and so on. In the interest of understanding how well models of EEG decoding generalize to previously unseen participants, we also investigated what happens when we train the model on all but one participant and then test on that remaining participant, with and without transfer learning.

4.3 Methods

4.3.1 Proposed CNN-based neural-network architecture

Convolutional models have been successful in many signal processing applications, as they allow temporally related inputs to be processed together via a sliding-window approach (Figure 22). This produces shared weights, where the same weight kernel is applied across the temporal domain (for a 1D convolutional model over time). In our architecture (Figure 22), this reduces the number of parameters needed in such a model and enables the signal to maintain its spatial relations—across time within each electrode and across electrodes over the head. The signal from each electrode channel is fed through the same convolutional base to produce an output matrix of dimension $C \times E$, where C is the number of electrodes (or channels) and E is the size of the embedding dimension (Figure 22). Hence, the convolutional layers in effect reduce the dimension of the input to the embedding dimension, E.

Now, in the self-attention part of the network [183, 184], we first initialize the weights for the Query (Q), Key (K), and Value (V). The magnitudes of Q, K, and V are derived by the product of the input (I) and the weights. The second step is to calculate the attentional score (S): $S = QK^T$. The shape of S will be $C \times C$. The Softmax (W) of S is calculated to return a vector of C x 1. The third step is to find the weighted values (M), $M = WV^T$. Each input's value for M is concatenated to return a shape of C x C, which will be the value for the final Attention. $Tanh$ was used to produce the alignment score. In the following, the equations show more details:

I	Input for self-attention, shape (number of channels (C) x the size of the embedding dimension (E))
Key, Value, Query	Initialize weights for key, value and query with shape of input size (C x E)
K = I x Key^T	Derive key, query, and value
V = I x $Value^T$	Shape (C x C)
Q = I x $Query^T$	
$S = Q \cdot K^T$	Calculate attention score by dot product (C x 1)
$W = Softmax(S)$	Calculate Softmax (C x 1)
$M = W \times V$	Multiply scores with value
$O = tanh\,(M \times W^T)$	Linear transformation of M

The attention layer discussed above is added after the convolutional base (Figure 22), so that each electrode channel is computed with every other channel to produce a matrix of scalar values. Summing across rows and normalizing these scalars produces a vector of attention scores. These scores are used to create a linear combination of all the electrode channel vectors, which is passed to the fully connected layers of the network for classification. A valuable part of this model is therefore its interpretability [155, 185, 186]. The attention scores for each electrode channel can be examined to determine the importance of each electrode in the model's prediction. However, in this study we were not interested in the added interpretability that the attentional mechanism affords us. Instead, we relied on the attentional mechanism to improve the prediction accuracy of our architecture. This is because a CNN with attention-mechanism architecture can improve classification performance using EEG signals by focusing on essential, task-relevant features on different time-steps, via the sliding windows. Table 1 shows the summary of the proposed NN parameter.

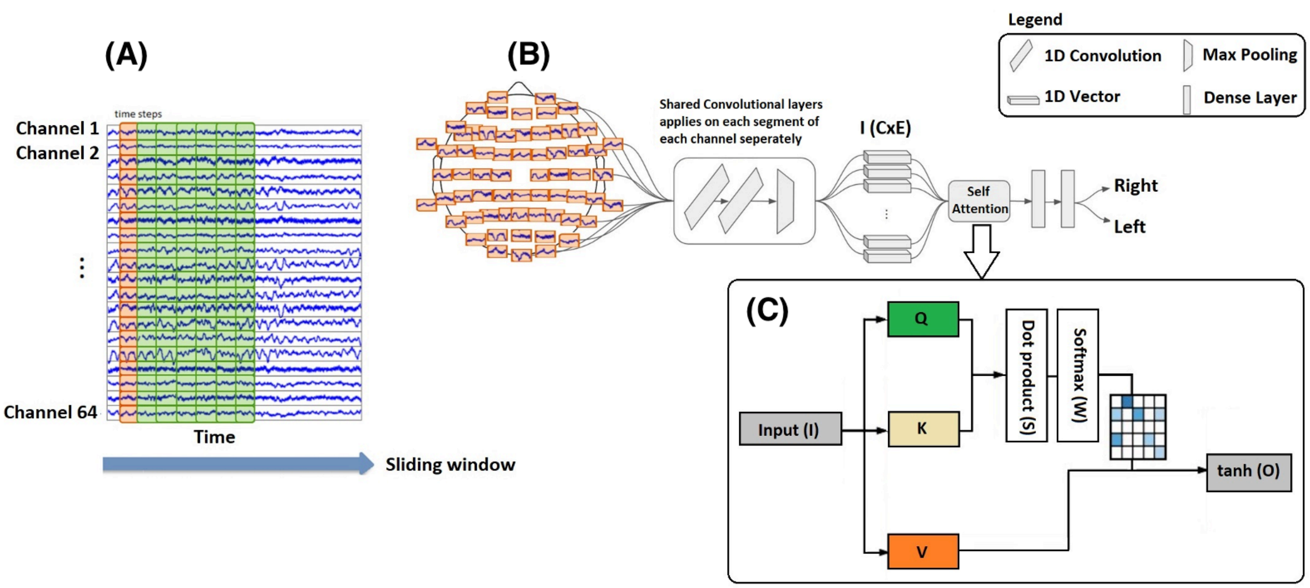

Figure 22. Our proposed CNN with attentional mechanism. (A) The sliding window (length is 1000 ms and step-size is 100 ms) applied to 64 EEG channels. (B) The 64 segments of raw EEG signal, depicted in orange in (A). Each time window and channel are separately sent through shared convolution layers. The embedded features I (C x E) applied to self-attention. The output of self-attention passes through 2 dense layers. (C) An expansion of the self-attention block.

Table 9 Summary of the proposed CNN with attentional mechanism parameters (“-1” represents a flexible shape, essentially the batch size)

Layer (Type)	Output Shape	Param #	Shared convolutional layer
Convolution 1D	[-1, 16, 64]	816	x64

Convolution 1D	[-1, 16, 6]	12,816	x64
Max Pooling 1D	[-1, 16, 3]	0	x64
1D Vector	[-1, 48]	2,304	0
Attention	[[-1, 64, 48], [-1, 64, 64]]	4,608	0
Dense	[-1, 32]	98,336	0
Dense	[-1, 2]	66	0
Total parameters	**977,762**		

4.3.2 Hyperparameter Optimization and Training

When implementing NN there are several choices (or hyperparameters) that must be set prior to training—those range from the type of architecture to the depth and width of the layers, through to the neuronal activation-function in the different layers, and so on. Choosing hyperparameters arbitrarily is likely to lead to suboptimal results. To address this, we first created a 3-way split of our data into a training, validation, and test sets to identify reasonable architectures and parameter ranges. Then, guided by those preestablished ranges, we conducted NN optimization via a Bayesian hyperparameter search using SHERPA [187], a Python library for hyperparameter tuning. The Bayesian search has the advantage of learning a distribution over the hyperparameters of the network architecture, in relation to the task to be optimized. By employing this procedure, we were able to evaluate a large space of possible models and test many configurations.

We detail the hyperparameters of interest in Table 10, as well as the range of available options during the search. The hyperparameters of interest consisted of the activation function, dropout percentage, learning rate, learning rate decay, nodes per layer, and the optimizer. Additional hyperparameters for convolutional models included the number of filters and the kernel size. We tried 250 different hyperparameter settings for each network architecture (Dense NN, Conv Net-

Dense NN, Conv Net-Attention-Dense NN), for a total of 750 models over 3 different NN (Dense NN, Conv Net-Dense NN, Conv Net-Attention-Dense NN). Table 11 present the result of best hyperparameters tuning by SHERPA for the 3 datasets: BCI competition IV 2a (BCI 2a), BCI competition IV 2b (BCI 2b), and our dataset and for 3 different models (Dense, CNN-Dense, and CNN-Attention-Dense).

For the 3 datasets examined in this study, we adhered to the following procedure. For each set of hyperparameters sampled in the search, we partitioned each subject's data into a training and validation set. The proposed architecture was thus trained on each subject separately. Then, to evaluate the architecture, we averaged the validation accuracy scores across subjects. We then selected the network architecture with the highest average accuracy score across all subjects. Critically, this process ensures that we find architectures that perform well across subjects, but which are not tailored to specific subjects or tasks.

All networks were trained for 250 epochs using an early stopping condition—i.e., when the accuracy on the validation set did not improve for 25 epochs, training stopped. All models were trained using 10-fold cross-validation. The partitioning was stratified to ensure a constant ratio of representation amongst right and left examples—roughly 50/50—in keeping with the ratio in the data overall. This cross-validation procedure requires a given model to be trained 10 distinct times (re-initializing the network parameters each time) and ensures that, on the one hand, different subsets of the data are used for training and testing, while on the other hand, each datapoint serves as part of the training set (9 times) and in the test set (once). To be clear, when we performed cross validation, we used data partitions that were not used during the hyperparameter search. The accuracies reported below are therefore always the average accuracies across the 10 validation sets described above.

To double check our results, we carried one additional train/validation/test split of 75/15/10%, respectively. After this train/validation/test procedure, we ended up with neural architectures that were the same as those selected by the cross-validation procedure above—both in terms of the number of layers and the kernel size. This gave us confidence that our results are not due to some leakage between the training and test sets. Our cross-validation procedure allowed us to report confidence scores, in the form of average accuracies and standard deviations. It also demonstrated that we did not cherry pick a data partition in which the proposed architectures happened to perform well; rather, our models were robust across partitions.

Training took place on NVIDA Titan V GPUs with 12GB of memory. Each epoch took less than a minute to complete. Training for a single fold typically completed within 30 minutes.

Table 10. The hyperparameter space

Name	Range	Type
Activation	(ReLU, ELU)	Choice
Dropout	(0, 0.9)	Continuous
Kernel Size	(25, 50, 75)	Choice
Learning Rate	(0.0001, 0.1)	Continuous
Learning Rate Decay	(0.5, 1.0)	Continuous
Number of Dense Nodes	(8, 512)	Discrete
Number of Filters	(16, 32, 64)	Choice
Optimizer	Adam, SGD, RMSProp	Choice

Table 11. Hyperparameter tuning by SHERPA for 3 the datasets (BCI 2a, BCI 2b and our experimental dataset) for 3 different models (Dense, CNN dense, and CNN attention dense)

dataset	Model	Kernel Size	Activation	Dropout	Learning Rate	Learning Rate Decay	Number of filters	Dense Nodes	Optimizer
BCI 2a	Dense NN	NAN	ReLU	0.171	0.017	1	NAN	27	Adam
	Conv Net-Dense NN	25	ELU	0.092	0.052	1	64	303	SGD
	Conv Net-Attention-Dense NN	25	ELU	0.9	0.1	1	32	91	SGD
BCI 2b	Dense NN	NAN	ReLU	0.845	0.001	0.864	NAN	289	Adam
	Conv Net-Dense NN	50	ReLU	0	0.1	1	16	15	SGD
	Conv Net-Attention-Dense NN	25	ELU	0	0.1	1	64	263	SGD
Our dataset	Dense NN	NAN	ReLU	0.687	0.037	1	NAN	369	SGD
	Conv Net-Dense NN	25	ReLU	0.68	0.034	0.989	32	196	SGD
	Conv Net-Attention-Dense NN	50	ELU	0.807	0.1	0.978	32	183	SGD

4.3.3 Data augmentation

Generally, in machine learning, but especially for NN, the classification accuracy tends to critically depend on the amount of training data; limited training data typically leads to low accuracy. DA comprises the systematic generation of new samples to augment an existing dataset by transforming existing samples in a manner that increases the accuracy and stability of classification

[28]. Exposing the classifiers to varied representations of its training samples typically makes the model more invariant and robust to such transformations when attempting to generalize the model to new datasets. DA for the MI task fell into 5 categories in our analysis: noise addition [95, 188], GAN [101, 105, 115, 189], sliding window [123, 126, 129], Fourier transform [82], and recombination of segmentation [145]. Table 12. DA techniques that are used on the MI task shows more details about each of these methods. We evaluate all DA techniques with a magnification factor $m = (2, 5, 10, 15, 20, 30, 50)$ for our proposed CNN.

Table 12. DA techniques that are used on the MI task

DA methods	Details of the method
Sliding window [123, 126, 129]	Sliding window over the input of each trial, which leads to many more training examples for the network compared to using than the entire. More formally, given an original trial $X^j \in \mathbb{R}^{E \times T}$, with E electrodes and T timesteps, we create a set of crops with crop size T' as time slices of the trial: $C^j = (X^j_{1,\ldots,E;\, t,\ldots t+T'} \mid t \in 1, \ldots T - T')$. All of these $T - T'$ crops then become training examples for our CNN and will get the same label, y_j, as the original trial. The best results in the BCI dataset are for 1s window length. In this study, we tried to evaluate this technique with different m and 100 ms step-size.
Noise Addition [95, 188]	We found two main categories for adding noise to the EEG signals in purpose of DA: (1) Add various types of noise such as Gaussian, Poisson, Salt and pepper noise, etc. with different parameters (for instance: mean (μ) and standard deviation (σ) to the raw signal (2) Convert EEG signals to sequences of images and add noise to the images [28]. Our proposed end-to-end CNN is for raw EEG. Therefore, we add noise just on the raw EEG signal. We add Gaussian noise with different parameters (mean = 0, standard deviation $\sigma = (0.01, 0.1, 0.2, 0.5)$ to all channels of raw EEG signal.
GANs [101, 105, 115, 189]s	The GAN framework consists of two opposing networks trying to outplay each other [190]. The discriminator (D) is trained to distinguish

	between real and fake input data. The generator (G) takes a latent noise variable z as input and tries to generate fake samples that would not be recognized as fake by the discriminator. To learn a generator distribution p_g over data x, the generator builds a mapping function from a prior noise distribution $p_z(z)$ to data space as $G(z; \theta_g)$. And the discriminator, $D(x;\ \theta_d)$, outputs a single scalar representing the probability that x came from training data rather than pg. G and D are both trained simultaneously: we adjust parameters for G to minimize $log(1 - D(G(z))$ and adjust parameters for D to minimize $logD(x)$ [190]. This results in a minimax game in which the generator is forced by the discriminator to produce ever better samples with value function $V(G, D)$: $$min_G max_D\, V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))].$$ GAN can be extended to a conditional model if both generator and discriminator are conditioned on some extra information such as y. In conditional generative adversarial nets (cGANs) y could be any kind of auxiliary information, such as class labels or data from other modalities. We can perform the conditioning by feeding y into the both the discriminator and generator as additional input layer. In the generator the prior input noise $p_z(z)$, and y are combined in joint hidden representation, and the adversarial training framework allows for considerable flexibly in how this hidden representation is composed. In the discriminator x and y are presented as inputs and to a discriminative function. The objective function of a two-player minmax game would be as: $$min_G max_D\, V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[logD(x\|y)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z\|y)))].$$
Recombination of segmentation [145]	Perform segmentation on the input trials (i.e., left-/right-hand MI) with the same label. Each trial is segmented into three crops. The crops with the same labels are then recombined to generate new trials. For the same person and the same class, the crops at the same position from

	multiple trials are randomly swapped and recombined in the time/frequency domain to generate recombined trials [145].
Fourier Transform/Wavelet [82]	Apply the empirical mode-decomposition algorithm on the EEG frames and mixed their intrinsic mode functions to create new, artificial EEG frames [82]. The algorithm decomposes the original EEG signals into a finite number of functions called "intrinsic mode functions" (IMFs). Once the signal has been decomposed, we can recover it by adding all the IMFs and the residue without loss. To generate the new samples, we swapped the IMFs of the decompositions. Moreover, the intrinsic characteristics of each class (left/right) will be preserved because we mixed the IMFs of the same class. We randomly select the trials that contribute with their IMFs to generate samples for specific class.

4.4 Dataset and experimental protocol

We used three datasets in this study: (1) A dataset that we collected ourselves, (2) the BCI 2a dataset [191], and (3) the BCI 2b dataset [192] (The experimental paradigms for our experimental dataset, BCI 2a, and BCI 2b.Figure 23. The experimental paradigms for our experimental dataset, BCI 2a, and BCI 2b.).

Our dataset: Seven healthy volunteers (3 male and 4 female) participated in the study, all were right-handed and between the ages of 23 to 30 (mean age 28). All participants gave written, informed consent to participate in the study. Participants were seated in a chair at a distance of 80 cm from an LCD screen with both hands resting on a Table. They held a tennis ball in each hand and were told to remain relaxed and strive to minimize movement and eye blinks. When required to respond, they were to squeeze the tennis ball in their hand but try to avoid tensing their arms or shoulders. Each session (ME and MI— Figure 23) was repeated twice. The whole experiment thus consisted of four sessions. Every session lasted 30–40 minutes with 10 to 15 minutes breaks between sessions. The duration of the whole experiment, including setup, was kept below 3 hours to minimize fatigue. EEG data was recorded and sampled at 250 Hz using 64 active electrodes (BrainVision actiCHamp) placed according to the 10/20 montage. Bipolar electromyography (EMG) electrodes were placed on the Brachioradialis for both hands as a sanity check for any movement in MI session.

Sessions 1 and 3 were designed to identify EEG signals related to ME. Participants were instructed to squeeze the tennis ball with their right or left hand while fixating on the cross displayed on the screen. They were encouraged to minimize all other movement and to only use the designated hand. One hundred trials were collected for each hand.

Session 2 and 4 aimed to show that a decoding model based on actual ME, derived from the first session, could be used to decode EEG activity in the absence of execution. Participants were instructed to carry out MI of the repetitive hand movement instructed in session 1 while fixating on the cross displayed on the screen. One hundred trials were collected for both left and right imagination per each session. All other aspects of the task were identical to session 1. This session also allowed us to screen participants for the presence of motor-related EEG oscillations, and at least minimal voluntary control over these oscillations. Hence, overall, we collected 200 trials of ME and 200 trials of MI for each subject. The data underlying this study have been uploaded to figshare.

Data are available from the following link: https://doi.org/10.6084/m9.figshare.14721297.v1

BCI 2a: BCI 2a contains EEG data from 9 healthy participants [191], 2 sessions per participant. Each session is made up of 288 trials, resulting in 5184 trials overall. No feedback was provided.

Twenty-two Ag/AGCL channels were used to record EEG. The signals were sampled with 250 Hz and bandpass filtered between 0.5-100 Hz. To compare our results with previous studies ([145], [193], [194] etc.) we focused on the C3, CZ, and C4 electrodes.

BCI 2b: BCI 2b contains EEG data from another 9 healthy participants [192]. For each participant, 5 sessions of data are collected. Each of the first 2 sessions has 120 trials and each of the last 3 sessions has 160 trials. The total number of trials is thus 6480. Two types of trials are included in these datasets: left- and right-hand MI. The first 2 sessions contain training data without feedback, while the last three sessions gave a smiley face as feedback. The EEG data is again collected over the C3, CZ, and C4 electrodes, which were placed following the international 10–20 system. The sampling frequency was 250 Hz. Table 13 presents the summary of three datasets.

Table 13. Summary of the 3 datasets used in this study: Our experimental dataset, BCI 2a, and BCI 2b

	Experimental dataset	**BCI 2a**	**BCI 2b**
The dataset provided by	The Institute for Interdisciplinary Brain and Behavioral Sciences, Chapman University	The Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology	The Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology
Open-source dataset	Yes	Yes	Yes
Description of dataset	2-class MI and ME (left hand and right hand). Session 1 and 3 are ME and 2and 4 MI, No feedback.	4-class MI (left hand, right hand, both feet, and tongue. No feedback	2-class MI (right hand, left hand). The first two sessions contain training data without feedback, and the last three sessions with smiley feedback.

# Channels	64 EEG channels (0.5-100Hz - BrainVision actiCHamp	22 bipolar EEG channels (0.5-100Hz; notch filtered)	3 bipolar EEG channels (0.5-100Hz; notch filtered)
Sampling frequency	250 Hz	250 Hz	250 Hz
# Subjects	7	9	9
# Sessions per subject	4	2	5
# Trials per session	100	288	120 for first 2 sessions and 160 trials for last 3 session
Total trials for each subject	400	576	720
Total trials in the dataset	2800	5184	6480

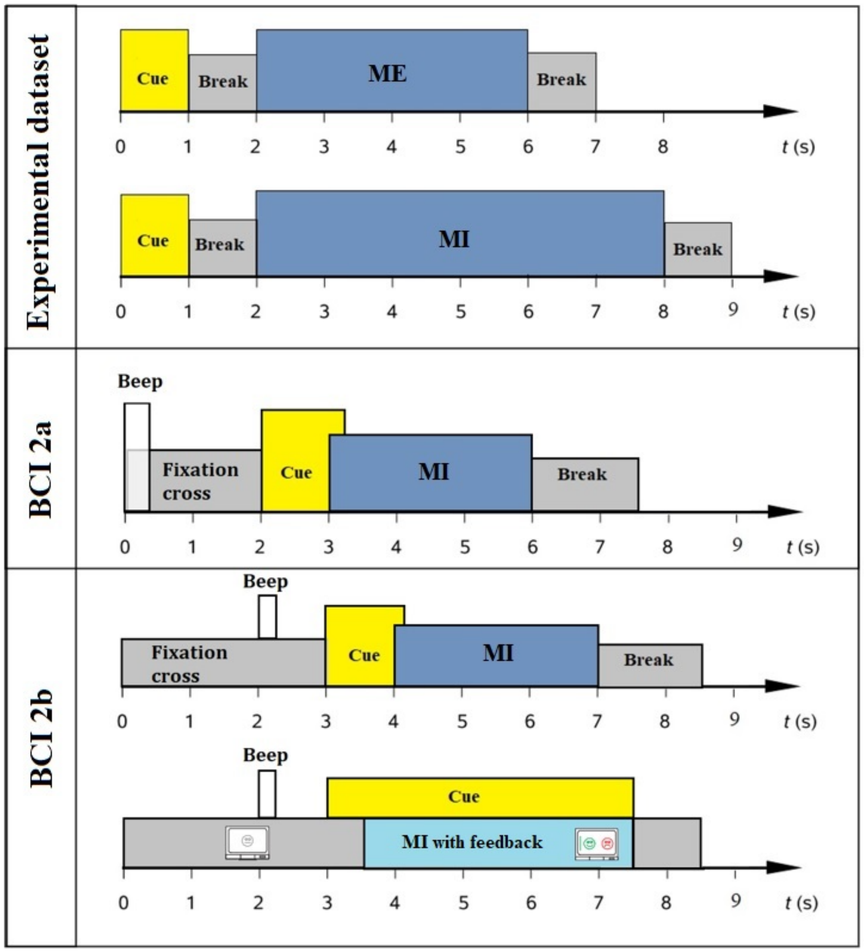

Figure 23. The experimental paradigms for our experimental dataset, BCI 2a, and BCI 2b.

4.5 Channel selection

Analyzing dense-array EEG is computationally expensive and complex; it also typically requires more expensive EEG systems than those with sparser electrodes. We therefore tested 4 different electrode configurations on our participants—which included 3, 7, 18, or all 64 electrodes (see

Methods)—to further test the effect of channel selection on classification accuracy for MI in our own dataset.

Configuration (1) C3, CZ, and C4 electrodes were chosen in accordance with the 10-20 framework [195] since these electrodes have been shown to be especially discriminatory in hand and foot movements data [196]. It should be noted that right (left) hand's MI operation is usually detected above the left (right) motor cortex underneath the C3 (C4) electrode, and the foot's MI action is typically captured by the CZ electrode.

Configuration (2) The brain's frontal, central and parietal lobes are important from a neurological perspective for MI commands. We therefore also focused on these 7 electrodes (i.e. F3, F4, C3, CZ, C4, P3 and P4), which reside above these lobes of interest according to the 10-20 standard are considered in criteria 2 [195].

Configuration (3) Electrodes that are generally placed around the left and right motor cortices are included in this configuration because they are related to MI. According to 10-20 electrode montage [195], 18 electrodes lie around motor cortex. These are labelled C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4 and P6 [197, 198].

Configuration (4) We used all 64 EEG channels.

In Figure 24, we showed these four configurations.

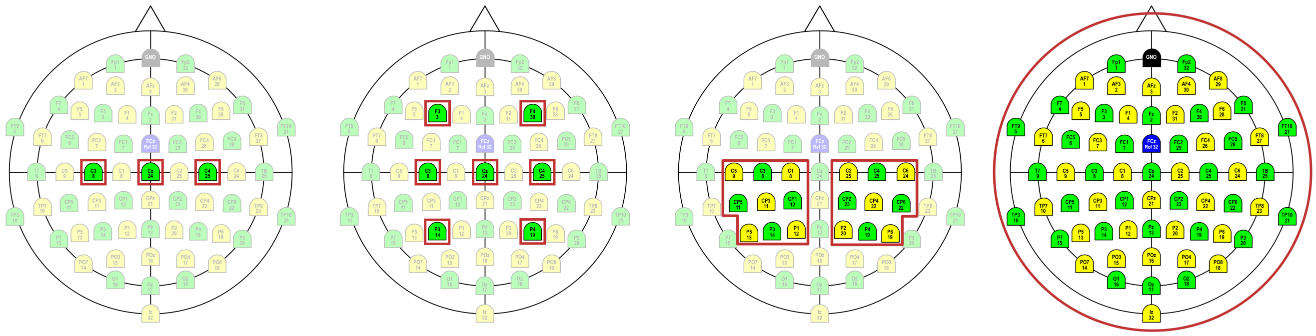

Figure 24. Four different electrode configurations on the actiCAP—which included 3, 7, 18, and all 64 electrodes

4.6 Results

4.6.1 Performance of the proposed CNN (Neural architectures vs. Neural architectures)

To evaluate the performance of our proposed CNN, we conducted comparisons between the Dense NN, Conv Net-Dense NN, Dai et al. (2020), and Conv Net-Attention-Dense NN (Figure 25). The baseline Conv Net is identical to the Conv Net-Attention-Dense NN but lacks the attention module (see Methods). The dense network sends all channels through 2 dense layers, then it concatenates all the vectors into a single one and sends that through 2 more dense layers. We used SHERPA for hyperparameter optimization for all 4 types of networks [187]. We also reproduced the proposed

NN in Dai et al. (2020) [145] without the use of DA to compare it with the proposed CNN with the attentional mechanism.

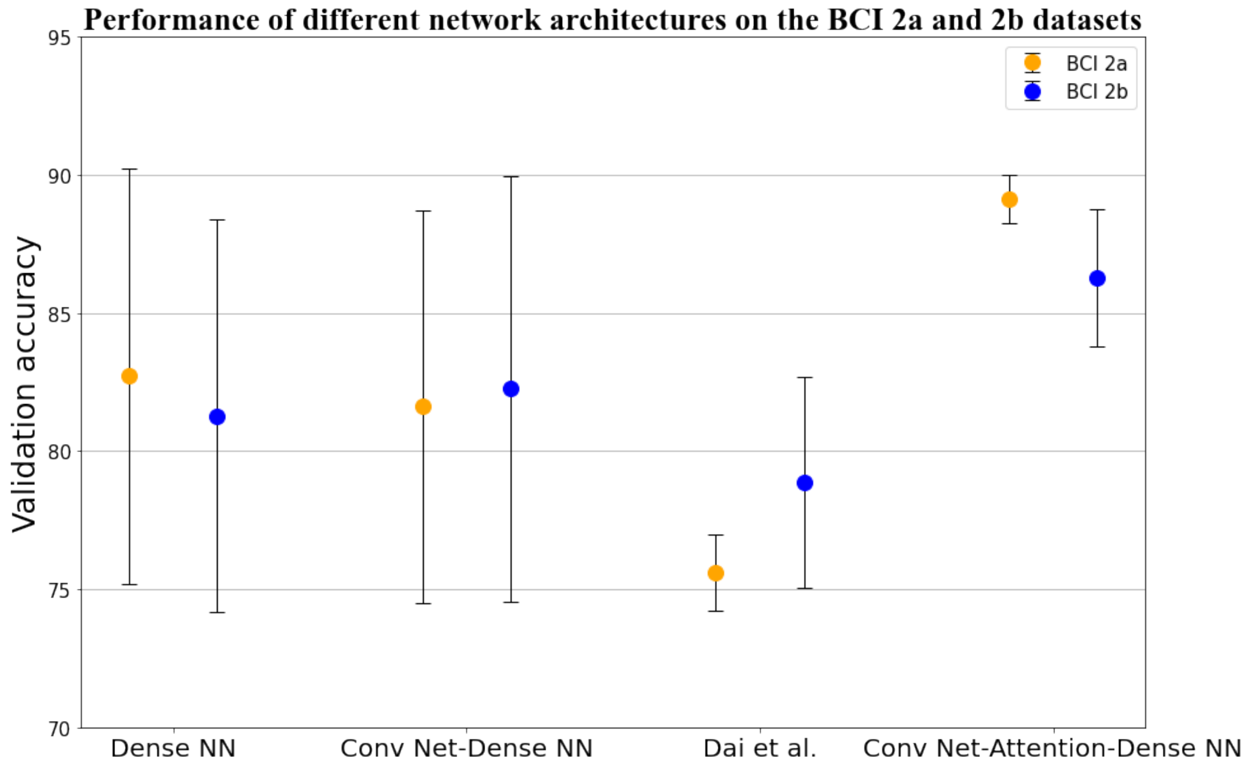

Figure 25. Comparison the average validation accuracy (±SE) on the BCI 2a and BCI 2b datasets with Dense, CNN, Dai et al. (2020), and CNN-Attention-Dense (See section 2.1 and 2.2).

Table 14 represents the classification results of our proposed CNN (with the attentional mechanism) without DA and with DA, which resulted in the highest accuracy for both datasets. Those are further compared against the results of Dai et al. [145]. All classifications were carried out on the BCI 2a and BCI 2b datasets. The average accuracy in Dai et al. (2020) for BCI 2a and BCI 2b were 91.57% (±5.73) and 87.6% (±8.48), respectively. In comparison, our proposed method with DA (GAN and m=15) achieved an average accuracy of 93.6% (±2.59) for BCI 2a and 87.83% (±6.34) for BCI 2b. Hence, our method has a higher average accuracy than Dai et al.

(2020) while maintaining less variability in the accuracy across participants for both datasets. For the BCI 2a, our proposed method was 90.54% or higher for all participants while Dai et al. (2020) got this accuracy just for 5 of 9 participants (56%). Furthermore, we reproduced the NN described in [145] without the use of DA to compare with our proposed CNN with the attentional mechanism without DA. Our results on the BCI 2a and 2b datasets were 89.11% (±3.77) and 86.28% (±7.41), respectively, outperforming those of [145] at 75.61% (±14.63) and 78.88% (±11.42), respectively. Again, our results were also less variable than theirs.

Table 15 further compares our results with various other state-of-the-art methods. As is apparent from the Table, our results outperform all others, typically by a wide margin. On average, our method is 16.44 % and 7.21% more accurate than the other method for the 2a and 2b datasets, respectively. What is more, even without DA, our method has a higher average accuracy than all other methods except for Dai et al. (2020). And, with DA, our method beats all other methods, including Da. et al.'s.

Table 14. Participant-by participant comparison of the proposed CNN with attentional mechanism—with and without DA—against Dai et al. [145] results on the BCI 2a and BCI 2b datasets.

	BCI 2a				BCI 2b			
Participant	**Dai et al. (2020)[145]**	**Reproduced the result in [145] (without DA)**	**Proposed method without DA**	**Proposed method with DA (GAN m=15)**	**[145]**	**Reproduced the result (without DA)**	**Proposed method without DA**	**Proposed method with DA (sliding window m=2)**
1	90.07%	69.77%	91.58%	95.38%	80.50%	70.83%	81.64%	84.13%
2	80.28%	65.62%	89.67%	91.25%	70.60%	63.24%	73.17%	77.92%
3	97.08%	97.91%	91.89%	91.25%	85.60%	62.64%	81.50%	83.64%
4	89.66%	69.45%	90.05%	96.12%	94.60%	97.84%	98.61%	99.18%
5	97.04%	62.51%	91.28%	95.05%	98.30%	80.95%	93.83%	94.97%
6	87.04%	62.48%	90.97%	94.62%	86.60%	80.28%	85.22%	85.83%
7	92.14%	66.66%	81.38%	91.22%	89.60%	84.58%	86.57%	86.57%
8	98.51%	90.64%	91.20%	90.54%	95.60%	86.05%	89.90%	90.50%
9	92.31%	95.46%	83.95%	97.50%	87.40%	83.47%	86.05%	87.73%
AVG	91.57%	75.61%	89.11%	93.60%	87.60%	78.88%	86.28%	87.83%
S.D.	5.73	14.63	3.77	2.59	8.48	11.42	7.41	6.34
S.E.	1.91	4.87	1.26	0.87	2.83	3.81	2.47	2.11

Table 15. Comparison of our proposed method (with and without data augmentation) with other state-of-the-art methods. All methods were run on the same dataset (BCI 2a and/or BCI 2b).

	[199]	[200]	[201]	[194]	[202]	[177]	[167]	[203]	[204]	[205]	[193]	[206]	[145]	Proposed method	Proposed method
Data set	**2b**	**2b**	**2b**	**2a/2b**	**2b**	**2b**	**2b**	**2b**	**2a**	**2a/2b**	**2a/2b**	**2a**	**2a/2b**	**2a/2b**	**2a/2b**
S1	77.0	70.0	80.0	63.69/73.2	84.6	81.0	76.0	72.5	88.9	90.28/70.3	66.7/62.8	91.5	90.07/80.5	91.58/81.64	95.38/84.13
S2	64.5	60.0	66.0	61.97/67.5	66.3	65.0	65.8	56.4	51.4	57.64/50.6	63.9/67.1	60.6	80.28/70.6	89.67/73.17	91.25/77.92
S3	61.0	61.0	53.0	91.09/63	62.9	66.0	75.3	55.6	96.5	95.14/52.8	77.8/98.7	94.2	97.08/85.6	91.89/81.50	91.25/83.64
S4	96.5	97.5	98.5	61.72/97.4	95.8	98.0	95.3	97.2	70.1	65.97/93.8	63.2/88.4	76.7	89.66/94.6	90.05/98.61	96.12/99.18
S5	82.0	92.8	93.5	63.41/95.5	89.2	93.0	83.0	88.4	54.9	61.11/63.8	72.2/96.3	58.5	97.04/98.3	91.28/93.83	95.05/94.97
S6	84.5	81.0	89.0	66.11/86.7	97.9	88.0	79.5	78.7	71.5	65.28/74.1	70.1/75.3	68.5	87.04/86.6	90.97/85.22	94.62/85.83
S7	75.0	77.5	81.5	59.57/84.7	82.1	82.0	74.5	77.5	81.3	61.11/61.9	64.6/72.2	78.6	92.14/89.6	81.38/86.57	91.22/86.57
S8	91.0	92.5	94.0	62.84/95.9	86.3	94.0	75.3	91.9	93.8	91.67/83.1	76.4/87.8	97.0	98.51/95.6	91.20/89.90	90.54/90.50
S9	87.0	87.2	90.5	84.46/92.6	97.1	91.0	73.3	83.4	93.8	86.11/77.2	77.1/85.3	93.9	92.31/87.4	83.95/86.05	97.50/87.73
AVG	80	80	83	68.32/84.1	84.7	84	77.6	78	78.01	74.92/69.7	70.2/81.6	79.93	91.57/87.6	89.11/86.28	93.60/87.83
S.D.	1.3	1.5	1.6	1.3/1.5	1.4	1.3	0.9	1.6	1.9	1.7/1.6	0.7/1.4	1.7	0.6/0.9	0.4/0.8	0.3/0.7

4.6.2 Properties of our collected dataset

There are several available BCI datasets [191, 192, 207]. However, we wanted to investigate several open questions in neuroscience and BCI that were outside the scope of the available datasets. So, we took the time and effort to collect our own dataset, which we are now sharing with the community. First, we wanted to test and directly compare the performance of our proposed attentional CNN on ME, MI, and their combination. In particular, we wanted to track the decoding accuracy over time via a sliding-window approach. We therefore increased the duration of the motor-imagination period from 2-3s to 4-6s to gain more insight and track the changes in decoding accuracy over time.

Second, BCI datasets typically instruct subjects to make trivial movements, such as pressing a button. We wanted to test our subjects on a less trivial paradigm, that requires them to exert some force. We therefore had our subjects squeeze a tennis ball (ME) or imagine doing that (MI). We expected this to make our classifier more robust against variety of MI tasks. This is vindicated by recent evidence that decoding attempted handwriting movements results in much higher accuracy than attempted typing [208].

Third, most of the BCI datasets for MI focused on electrodes above the motor region—such as C3, C4, and Cz [192]. We wanted to test to what degree general, high-density EEG recordings across the cortex (to the extent that those brain regions are accessible to EEG) contribute to the performance of an MI classifier. This also let us investigate the extent to which channel selection is useful in MI classification. Forth, an additional goal of our study was to evaluate the role of DA in MI classification. So, we needed a large enough dataset to be able to compare classification results when training our classifier on only a portion of the dataset. Altogether we recorded 400 trials pers subject (200 each for ME and MI, see Methods).

4.6.3 Motor Imagery vs. Motor Execution

MI could be described as kinesthetic anticipation of corresponding overt ME without producing an actual motor output. Jeannerod stated that MI is functionally equivalent to its ME counterpart [209]. More specifically, MI is related to the preparation of ME and represents meaningful neurophysiological dynamics of human motor functions [210]. Consequently, both MI and ME are accompanied by activation in common sensorimotor areas, such as the primary motor area (M1), supplementary motor area (SMA), and premotor cortex (PMC) [209, 210]. The neurophysiology underlying MI may differ in healthy people and patients with motor-impairing conditions [211]. MI-based BCI may further augment the motor learning process in healthy participants [212]. What is more, in patients with impaired motor functions, MI is often the only viable option to drive rehabilitative BCI, because these patients cannot perform overt ME [211]. The individuality and severity of motor impairments impact the underlying neurophysiology; for example, post-stroke neurophysiology relies on lesion locations [213]. Additional work is needed to further delineate the roles of MI and ME in motor learning or relearning for both healthy and impaired participants to refine the design of BCI for supplementing the motor learning process.

Our own dataset enables us to directly compare ME and MI within each participant. In our task, the participants were presented with the cue for 1 s, then saw a blank screen for 1 s, and finally began ME or MI for 4 s (see Methods). However, Dai et al. (2020), only used 2 s of MI. To better compare our results to theirs, we ran a sliding window analysis only for the first 2 s of the 4-s-long

ME or MI period. We used window sizes of 100 ms, 300 ms, 500 ms, 1 s, and 2 s, with the step size fixed at 100 ms (see Methods) on the data from all 64 channels. With this analysis, we would expect to see a rise in the accuracy leading up to the moment when the participants needed to begin ME or MI. Further, as participants were supposed to execute or imagine the movement for 4s, we expected the accuracy to then generally plateau over this after the above rise (similarly to Salvaris & Haggard, 2014 for example).

The left column in Figure 26. Validation accuracy of sliding-window analysis in ME (top), MI (middle), and ME and MI combined (bottom). The left column is the accuracy over time averaged across all 7 participants. The right column depicts the accuracy for the participant with the highest overall accuracy in the ME condition (Participant 4). represents the average validation accuracy over all 7 participants and the right column is specifically for Participant 4. Both show the accuracy of the running-window analysis and over the first 4 s after cue onset for 3 analyses: ME only, MI only, and the combination of ME and MI trials. The window shown at the 4 s mark is from 3900 to 4000 ms for the 100 ms window, for 3700 to 4000 ms for the 300 ms window, and so on.

Our method's accuracy on ME is greater than on MI (Figure 26. Validation accuracy of sliding-window analysis in ME (top), MI (middle), and ME and MI combined (bottom). The left column is the accuracy over time averaged across all 7 participants. The right column depicts the accuracy for the participant with the highest overall accuracy in the ME condition (Participant 4).), which is consistent with previous findings about ME versus MI [214]. The average validation accuracy for the combination of MI and ME (All) is also greater than MI. Looking at the variability among the different window sizes, we see more variability in the ME condition than the MI or combined condition, on average. Our averaged results over all participants also align with our expectations, in that the accuracy rises from chance toward the beginning of the ME and MI periods and then generally plateaus (again, compare with Salvaris & Haggard, 2014).

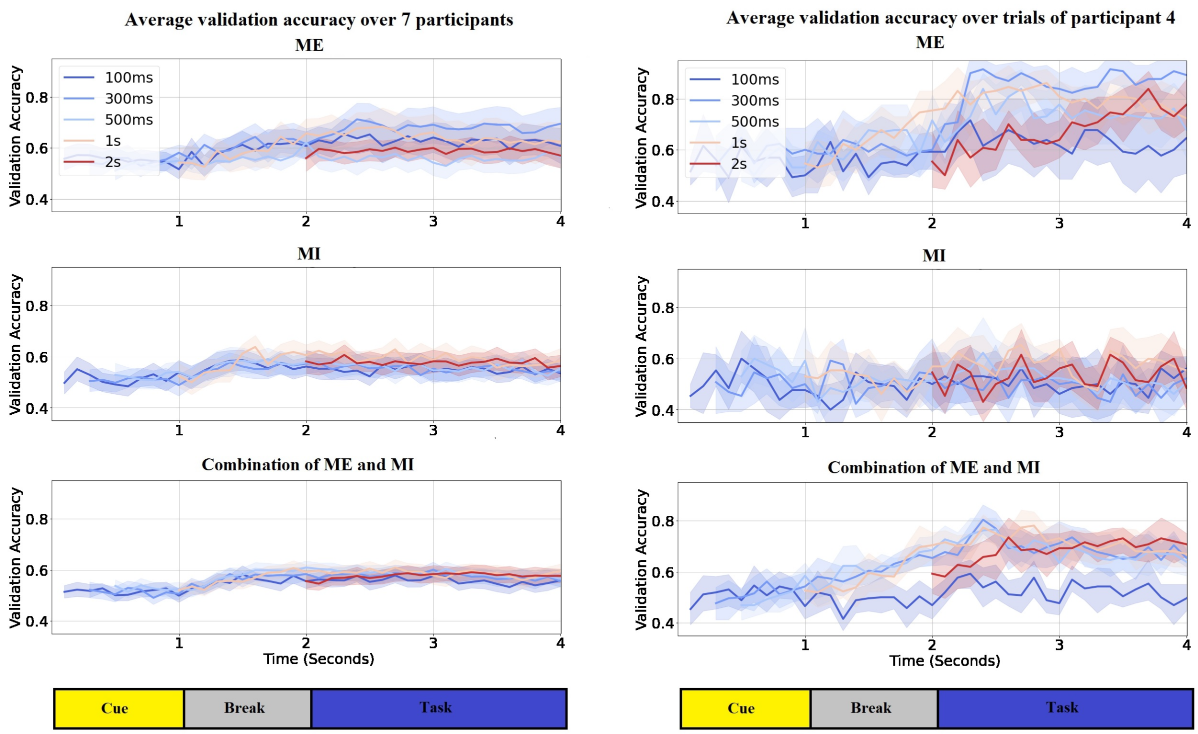

Figure 26. Validation accuracy of sliding-window analysis in ME (top), MI (middle), and ME and MI combined (bottom). The left column is the accuracy over time averaged across all 7 participants. The right column depicts the accuracy for the participant with the highest overall accuracy in the ME condition (Participant 4).

4.6.4 Channel selection

Analyzing dense-array EEG is computationally expensive and complex; it also typically requires more expensive EEG systems than those with sparser electrodes. Therefore, in this study we tested 4 different electrode configurations on our participants—which included 3, 7, 18, or all 64 electrodes (see Methods)—to further test the effect of channel selection on classification accuracy for MI in our own dataset.

The validation accuracy of the 7 participants for the 4 different channel-configurations are shown in Figure 27. Validation accuracy for different channel configurations on the 7 participants of our dataset. In Table 16, the validation accuracy for each participant and the average accuracy across all participants are shown. The 18-channel layout had the highest accuracy, at 81.73% ($\pm$2.5).

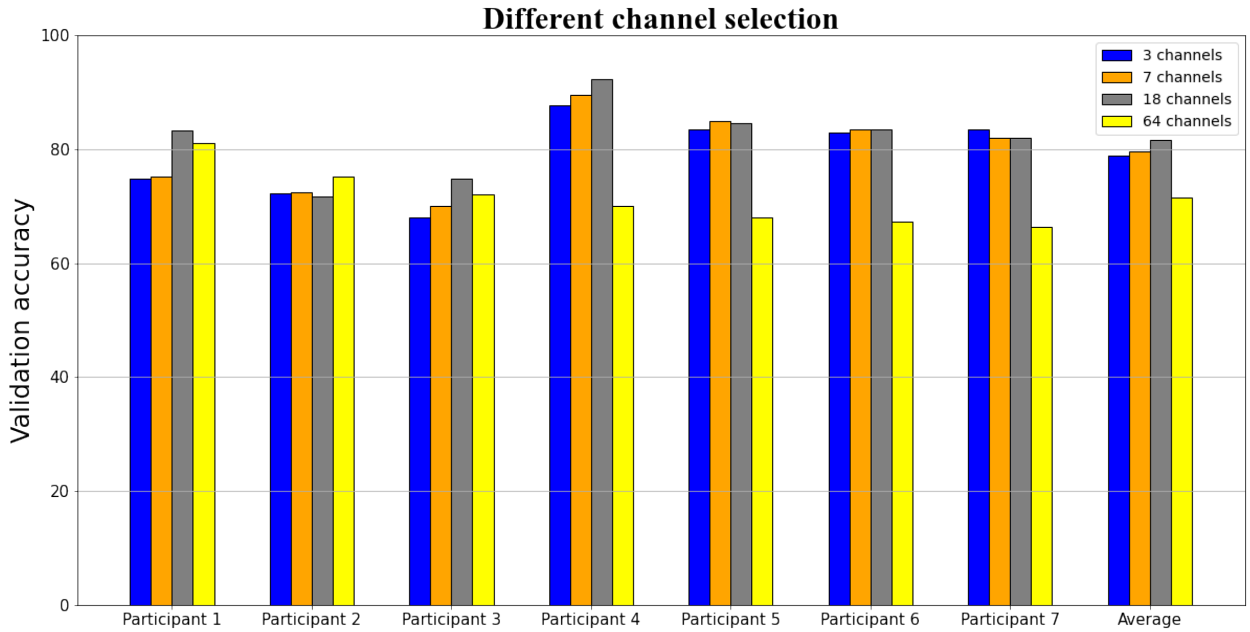

Figure 27. Validation accuracy for different channel configurations on the 7 participants of our dataset

Table 16. Validation accuracy for different channel selections on our dataset for single participants and the average over all participants. For each participant, we present mean ± SE over trials. In the bottom row, we present mean ± SE over participants.

Participant	3 channels	7 channels	18 channels	64 channels
1	74.75(±4.3)	75.25(±2.2)	83.25(±4.1)	81.18(±8.9)
2	72.25(±4.2)	72.50(±4.9)	71.75(±4.1)	75.22(±4.3)
3	68.01(±3.9)	70.01(±4.1)	74.75(±4.3)	72.05(±3.2)
4	87.69(±5.4)	89.62(±3.2)	92.31(±3.6)	70.03(±3.1)
5	83.50(±6.3)	85.01(±3.3)	84.50(±5.7)	68.08(±2.2)
6	83.00(±4.2)	83.50(±6.7)	83.51(±5.8)	67.33(±1.6)
7	83.50(±3.4)	82.01(±6.7)	82.01(±5.9)	66.41(±2.4)
AVG (±$S.E.$)	78.95(±**2.7**)	79.70(±**2.7**)	81.73(±**2.5**)	71.47(±**1.9**)

4.6.5 Data augmentation

We used 5 types of DA for the MI task: noise addition [95, 188], GAN [101, 105, 115, 189], sliding window [123, 126, 129], Fourier transform [82], and recombination of segmentation [145]. Table

9 represents the result of different DA techniques on the BCI 2a, BCI 2b and our dataset for 64 channels and 18 channels. We evaluate all DA techniques with magnification factor $m = (2, 5, 10, 15, 20, 30, 50)$ for the proposed CNN. For Fourier transform, we used the same technique as in [82]. For noise addition, we opted for Gaussian noise with $\mu = 0, \sigma = (0.1, 0.2, 0.5)$.

cGANs allow generation based on a class assignment [187]. In this study, the GAN had 2 different conditions that were implemented: In order to provide context about the task, the first GAN model generates a sample conditioned on the participant's decision—i.e., left vs. right. The second GAN model applies finer granularity by conditioning not only on left vs right but also the electrode channel. When generating data, the conditional inputs provide additional information and allow the model to tailor its outputs with greater detail. Figure 28. Our proposed cGAN model. In the generator (G), the prior input noise and label are combined into a hidden representation. In the discriminator (D), Real Data (i.e., raw EEG data) and the Label are presented as inputs to a discriminative function. The contents of all purple boxes in the architecture are the same and are expanded at the bottom left.illustrates the architecture of cGAN in our work:

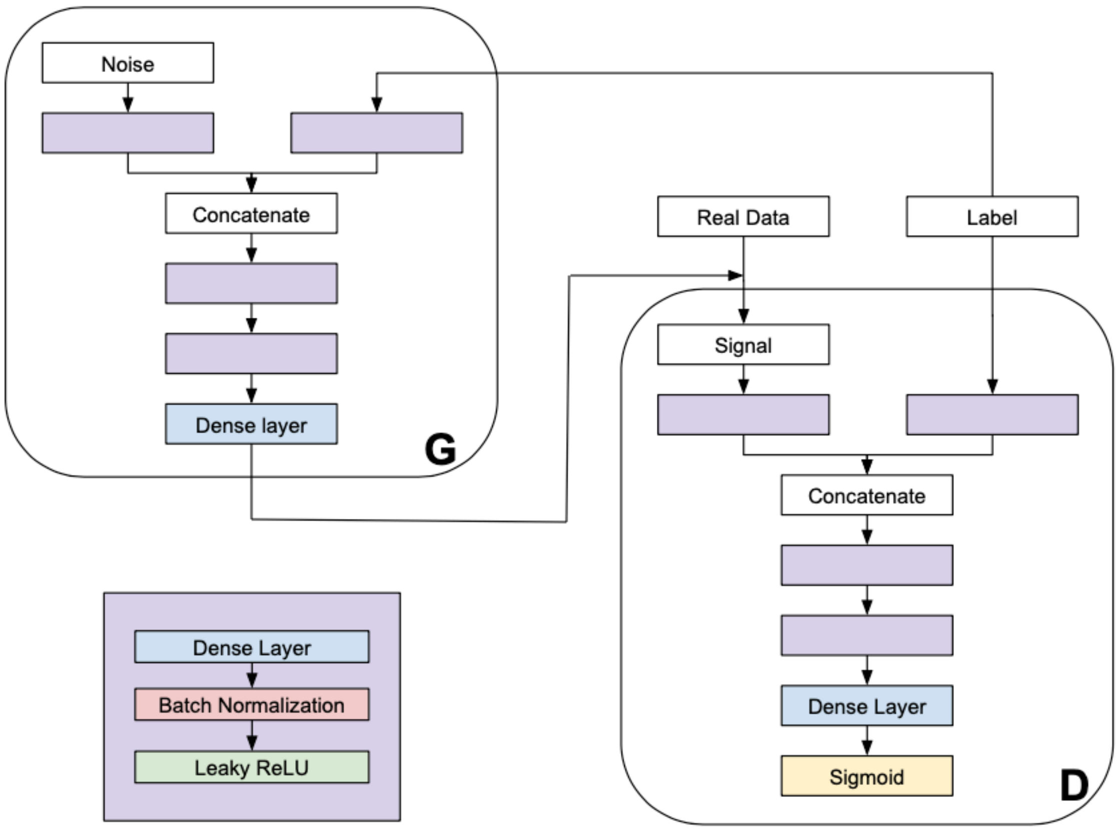

Figure 28. Our proposed cGAN model. In the generator (G), the prior input noise and label are combined into a hidden representation. In the discriminator (D), Real Data (i.e., raw EEG data) and the Label are presented as inputs to a discriminative function. The contents of all purple boxes in the architecture are the same and are expanded at the bottom left.

We also evaluated sliding-window technique (lengths $l = 1000\ ms$ with sampling frequency 250 Hz and step-size $100\ ms$). Table 17 demonstrated that GAN (conditional left vs. right and channels) with m=15 resulted in the best accuracy (93.6%) for BCI 2a dataset while Sliding Window (500 ms windows and 100 ms step size) with m=2 achieved the best accuracy (87.83%)

for BCI 2b dataset. For our dataset, Fourier Transform with m=15 for 64 (86.61%) and 18 (83.42%) channels, respectively. The BCI 2a dataset had a magnification factor of 15 for the best result compared to a magnification factor of only 2 for BCI 2b. This might be because we did not include neurofeedback within our experimental paradigm. Decoding neurofeedback dataset has less complexity which is why BCI 2b dataset was seen to have a smaller magnification factor of 2. Our dataset did not include neurofeedback in the paradigm similarly to the BCI 2a dataset.

Table 17. Comparison of different DA techniques with different magnification factors and hyperparameters for BCI 2a, BCI 2b, and our experimental dataset (for 64 channels and 18 channels)

Dataset	DA techniques		Fourier-Transform	Noise Addition			GAN		Sliding Window
	parameter for each DA		(EMD)	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.5$	Conditional (left vs. right)	Conditional (left vs. right and channels)	Sliding window of length 1s (step-size: 100 ms)
BCI 2a	Magnification factor	2	0.8671	0.9056	0.8982	0.8768	0.9133	0.9025	0.8948
		5	0.8652	0.8999	0.8849	0.8908	0.9240	0.9092	0.8904
		10	0.8822	0.8902	0.8920	0.8721	0.9087	0.9217	0.8992
		15	0.8858	0.8988	0.8756	0.8750	0.9358	0.9360	0.8949
		20	0.8932	0.8898	0.8975	0.8904	0.9193	0.9300	0.9092
BCI 2b	Magnification factor	2	0.8535	0.8647	0.8614	0.8575	0.7939	0.8511	0.8783
		5	0.8391	0.8746	0.8696	0.8558	0.7747	0.8624	0.8747
		10	0.8339	0.8677	0.8668	0.8560	0.7733	0.8582	0.8726
		15	0.8228	0.8660	0.8717	0.8551	0.7601	0.8646	0.8749
		20	0.8217	0.8736	0.8677	0.8535	0.7611	0.8708	0.8691
Our dataset (64 channels)	Magnification factor	2	0.8442	0.8146	0.7548	0.7720	0.7914	0.8159	0.7904
		5	0.8305	0.7743	0.7844	0.7897	0.8377	0.7945	0.7933
		10	0.8377	0.7907	0.7885	0.7793	0.8024	0.8044	0.8033
		15	0.8661	0.7775	0.7541	0.7556	0.8184	0.7824	0.8362
		20	0.8560	0.7521	0.7826	0.7886	0.7994	0.8052	0.7990
Our dataset	Magnification	2	0.8124	0.8051	0.8056	0.8079	0.8045	0.8174	0.8190
		5	0.8010	0.8179	0.8121	0.8090	0.7969	0.8156	0.8224

(18 channels)		10	0.7988	0.8123	0.8162	0.8048	0.7965	0.8020	0.8312
		15	0.7954	0.8203	0.8141	0.8047	0.7842	0.8015	0.8342
		20	0.7963	0.8209	0.8051	0.8048	0.7875	0.8102	0.8277

4.6.6 Different portions of dataset

A dearth of data is a common problem when training machine-learning models on neuroimaging data. We therefore wanted to systematically test to what degree DA can compensate for the reduced availability of data. We thus randomly selected 100%, 75%, 50%, or 25% of the samples in our dataset. And we tested the accuracy of DA on these different proportions of our dataset for different DA techniques and magnification factors (Table 18). Fourier transform resulted in the best accuracy for 100%, 75%, and 50% of the data, with 86.61%, 88.26%, and 86.18% accuracy, under magnification factors 15, 5, and 10, respectively. When using only 25% of the data, GAN (conditional left vs. right and channels) was the best DA technique in terms of accuracy, with 82.18% and a magnification factor of 15.

Table 18. Accuracies for different proportion of our dataset with different DA techniques

		DA techniques	Fourier-Transform	Noise Addition			GAN		Sliding window
Proportion of dataset		Parameter for each DA	(EMD)	σ =0.1	σ =0.2	σ =0.5	Conditional (left vs. right)	Conditional (left vs. right and channels)	Sliding window of length
100 %	Magnification factor	2	0.8442	0.8146	0.7548	0.772	0.7914	0.8159	0.7904
		5	0.8305	0.7743	0.7844	0.7897	0.8377	0.7945	0.7933
		10	0.8377	0.7907	0.7885	0.7793	0.8024	0.8044	0.8033
		15	0.8661	0.7775	0.7541	0.7556	0.8184	0.7824	0.8362
		20	0.856	0.7521	0.7826	0.7886	0.7994	0.8052	0.799
75%	Magnification factor	2	0.8644	0.7975	0.7886	0.8129	0.7772	0.7927	0.7695
		5	0.8826	0.7856	0.7877	0.7987	0.7997	0.8045	0.7998
		10	0.8707	0.8096	0.7743	0.7921	0.804	0.795	0.798

		15	0.8732	0.7735	0.8013	0.7741	0.778	0.8057	0.8104
		20	0.8625	0.8066	0.7838	0.7814	0.8223	0.8159	0.8158
50%	**Magnification factor**	2	0.8346	0.8116	0.7957	0.7909	0.7743	0.756	0.7669
		5	0.8536	0.7672	0.7687	0.782	0.7754	0.8063	0.7656
		10	0.8618	0.8067	0.8222	0.7695	0.8034	0.7943	0.7503
		15	0.8474	0.8037	0.7969	0.7687	0.7671	0.8151	0.7426
		20	0.8128	0.756	0.801	0.7539	0.8247	0.8069	0.8039
25%	**Magnification factor**	2	0.7422	0.798	0.7868	0.8057	0.7595	0.7731	0.7387
		5	0.7683	0.8016	0.7569	0.7755	0.7714	0.7821	0.7202
		10	0.7417	0.7838	0.7767	0.8087	0.7643	0.8204	0.7256
		15	0.7909	0.7643	0.8187	0.7584	0.7737	0.8218	0.7138
		20	0.7826	0.7643	0.7814	0.7513	0.7731	0.7982	0.7501

4.6.7 Combination of participants' EEG signals

The variability in brain anatomy and even more so functionality among different individuals is well known [e.g., 215]. Strong structure-function correspondences is therefore typically derived only at the aggregate level [216]. For example, Smith et al. delineated structural differences, suggesting that the number of folds and thickness of the cortex could be associated with whole-brain functional network [217]. Furthermore, inter-participant variability in brain topography may also occurs due to participant-specific cognitive styles and the strategies that different participants use to perform the task [218]. This might augment the underlying learning processes—e.g., motor and perceptual learning [219]. Intra- and inter-participant variability might be explained by scale-dependent brain networks in spatial, temporal and topological domains [220].

Motor variability due to variability in human kinematic parameters—e.g., force field adaptation, speed and trajectory, and motivational factors such as level of user engagement, arousal and feelings of competence, necessary for performing a motor task—is an integral part of the motor learning process [221-223]. What is more, EEG signals are of course measured from the scalp rather than directly inside the brain, so they suffer from various signal distortions and technical limitations [224]. Given the above, the extent to which machine-learning models can be transferred between participants is not completely understood. The EEG patterns associated with motor variability could partly explain intra-individual variability in SMR-based BCI [225]. The neurophysiological processes underpinning the SMR often vary over time and across participants. Inherent intra- and inter-participant variability causes covariate shift in data distributions that impede the transferability of model parameters among sessions/participants.

Given the above, we evaluate the performance of the proposed NN on combinations of data across participants. The validation accuracy was averaged over every possible combination for each dataset—e.g., all participant pairs, all triplets, etc. After finding all the possible combinations, the data was split into training and test for each combination to compute the validation accuracy. The averages of the validation accuracy over all the states for the three datasets are reported in Figure 29(Top) and differences between group (bottom). As we add more participants, the accuracy decreases—but the decreases become smaller. In Figure 29 (bottom), for the BCI 2a and 2b datasets, after combining 6 or more participants, we can see the curves plateau. This suggest that our proposed CNN was able to learn the important variations of the different EEG signals among the different subjects thus achieving stable accuracy.

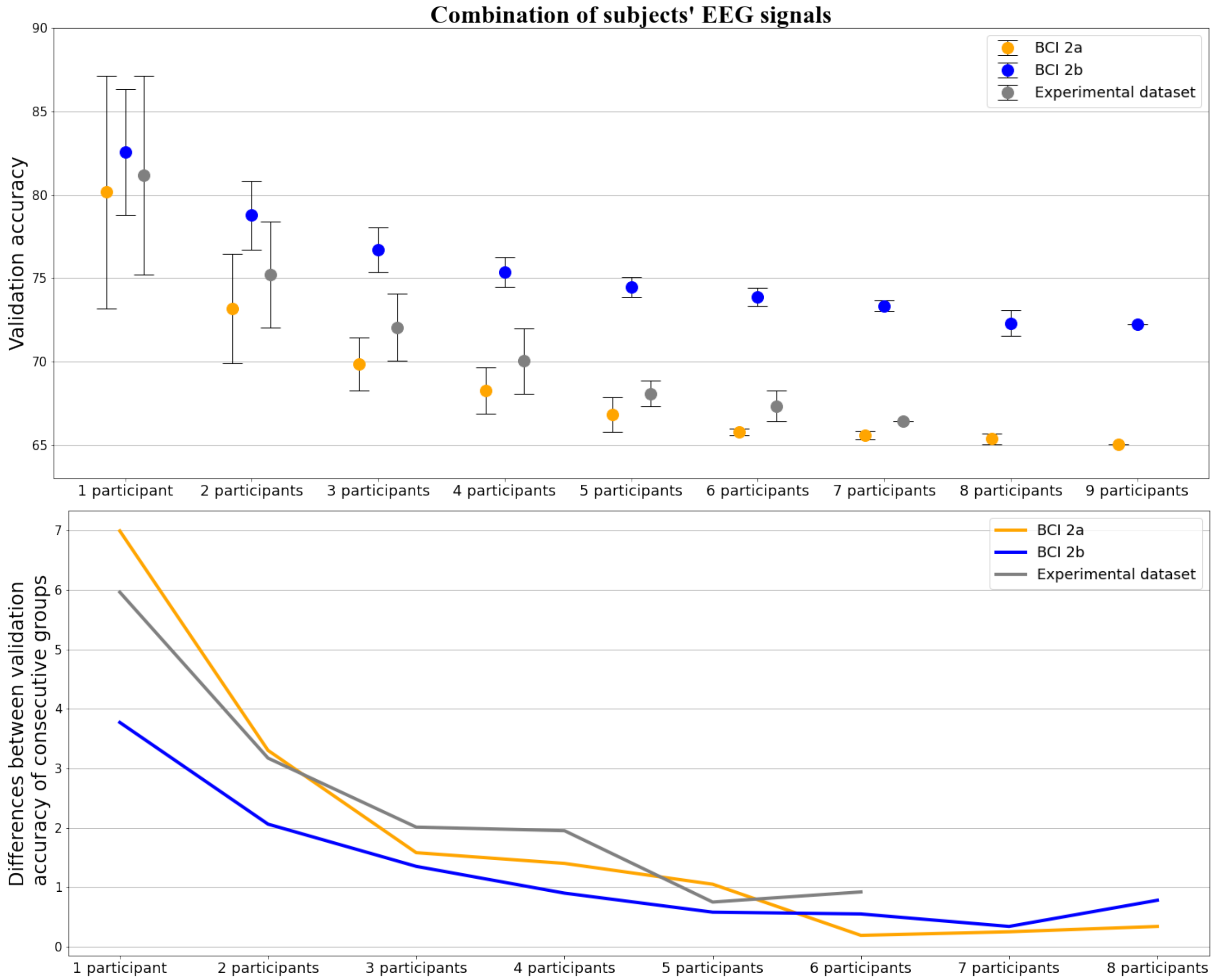

Figure 29. (Top)Validation accuracies for combinations of participants for BCI 2a, BCI 2b, and our experimental dataset. (Bottom) line plots of differences between mean validation accuracies of consecutive groups for the 3 datasets. The x axis labels are the smaller groups; so, differences between 2 participants and one are plotted above the label “1 participant”, between 3 and 2 participants above “2 participants”, and so on.

4.6.8 Leave-one-participant out and transfer learning

This subsection addresses two separates but closely related tasks. The first, leave-one-out, trains a NN on n-1 participants and tests on the remaining nth participant. This task addresses the question of how information is shared between different participants' EEG signals (see section 3.7, on the x-axis, 8 participants for BCI 2a, BCI 2b and 6 participants for our dataset).

The second task, transfer learning, pretrains a NN on $n-1$ participants and fine-tunes to the n^{th} participant [226]. The pre-training phase orients the network weights to extract meaningful representations from the data. Then the fine-tuning, where the learning rate is decreased, adjusts to the task of interest, the n^{th} participant. For transfer learning, 10-fold cross validation over the n^{th} participant was used. Each fold fine-tunes on 9 folds and tests on the held-out 10^{th} fold. Table 19 shows the result of transfer learning on the BCI 2a, BCI 2b, and our dataset (64 channels and 18 channels). Figure 30 compared the result with and without transfer learning for all 3 datasets. For instance, the validation accuracy without transfer learning on participant n is defined by the trained model based on combination of the other $n-1$ participants and is tested on the complete dataset of participant 9. However, the validation accuracy with transfer learning on participant n is tuned to the trained model based on combination of the other $n-1$ participants based on 10% of the n^{th} participant and is tested on 90% of participant n.

Table 19. Leave-one-out and transfer-learning validation accuracy for BCI 2a, BCI 2b, and our dataset (64 and 18 channels)

Train (participants index)	**Finetune (participant index)**	**BCI 2a (with transfer learning for different participants)**	**BCI 2b (with transfer learning for different participants)**
2-3-4-5-6-7-8-9	1	78.12	78.75
3-4-5-6-7-8-9-1	2	76.38	71.62
4-5-6-7-8-9-1-2	3	89.53	79.17
5-6-7-8-9-1-2-3	4	77.77	97.02
6-7-8-9-1-2-3-4	5	77.41	83.10
7-8-9-1-2-3-4-5	6	78.83	81.94
8-9-1-2-3-4-5-6	7	80.58	81.67
9-1-2-3-4-5-6-7	8	81.60	87.36

1-2-3-4-5-6-7-8	9	90.63	84.44
Train (participants index)	**Finetune (participant index)**	**Our dataset, 64 channels**	**Our dataset, 18 channels**
2-3-4-5-7-6	1	83.25	83.75
3-4-5-6-7-1	2	73.01	87.25
4-5-6-7-1-2	3	76.50	77.50
5-6-7-1-2-3	4	91.15	92.70
6-7-1-2-3-4	5	91.01	85.50
1-2-3-4-5-7	6	82.10	82.50
1-2-3-4-5-6	7	84.50	86.50

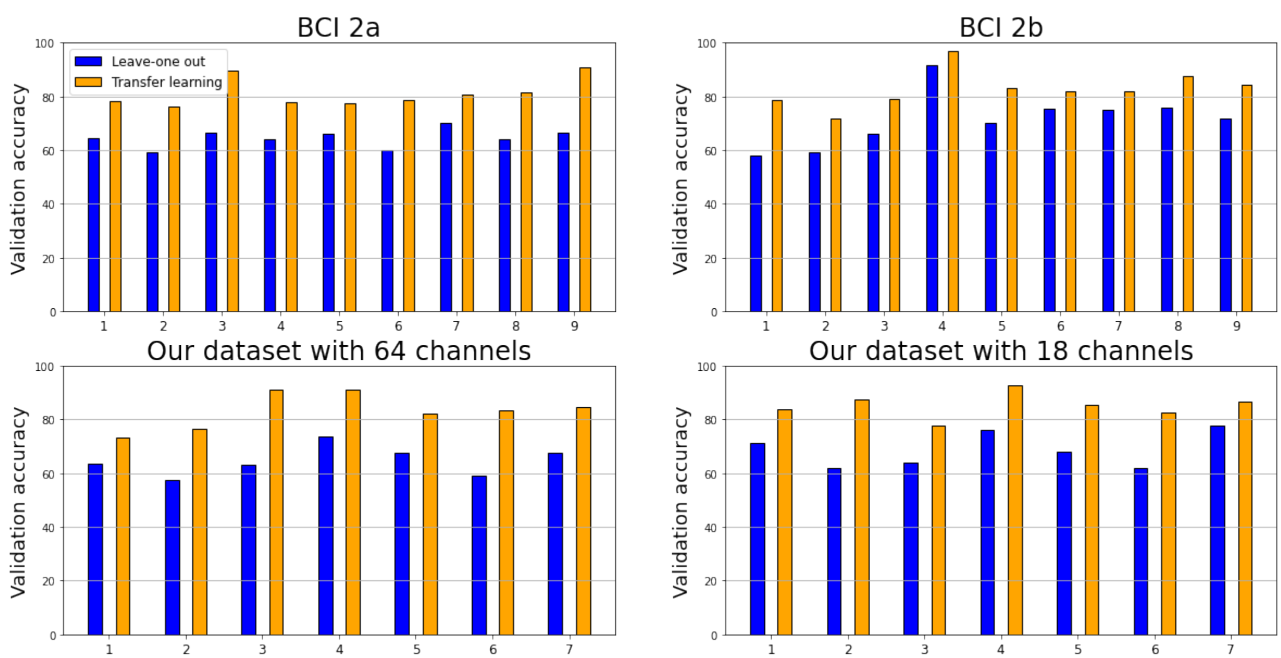

Figure 30. Validation accuracy for BCI 2a, BCI 2b, and our dataset (64 and 18 channels) with and without transfer learning

4.7 Discussion

In this study we proposed an end-to-end CNN architecture for EEG-based MI classification. This proposed mechanism is used to automatically extract features from raw EEG data (Figure 22). The

NN optimization used the SHERPA Bayesian hyperparameter search on 3 datasets: the BCI Competition IV 2a and BCI Competition IV 2b, which have become benchmarks in the field, and a dataset that we collected ourselves (Figure 23; see Methods).

We began by comparing the architecture we favored, Conv Net-Attention-Dense NN, to two other baseline architectures—a Dense NN and Conv Net-Dense NN —as well as to what was, to the best of our knowledge, the top result in the field on the benchmark datasets—the architecture described in Dai et al. (2020) (see Figure 32). Our CNN-Attention-Dense achieved 93.6% (S.E.: ±0.87) and 87.8% (S.E.: ±2.11) accuracy over the BCI 2a and 2b datasets, respectively (Table 15). That is 6.4% to 13.5% and 4.03% to 5% better than the other architectures for BCI 2a and 2b, respectively (Figure 25). We further compared our results with all the papers we could find that classified the BCI 2a and 2b datasets and reported participant-by-participant results. For the BCI 2a dataset, our proposed EEG MI classification method achieved an improvement of 2.03% to 25.28% over all other methods (Table 14 and Table 15). For the BCI 2b dataset, our proposed method achieved an average improvement of 0.23% to 18.13% over previous methods (Table 15).

To the best of our knowledge, our CNN-Attention-Dense architecture achieved the highest accuracy thus far for the 2 benchmark datasets—BCI 2a and 2b. On top of that, an additional strength of our approach is its automated features extraction, directly from raw EEG. This contrasts with most methods, which tend to use handcrafted features and require heuristic parameter setting (e.g., predefined frequency bands). Automated features have the advantage of often generalizing better across tasks and participants [145]. Another potential advantage of our architecture is that the attentional mechanism could potentially lead to more interpretable results. However, we leave the explainable-AI facet of our architecture for further, future research.

The dataset that we collected for this study used 64 electrodes (according to the 10/20 montage; Figure 24). It included both ME and MI tasks and enabled us to compare the two tasks. Having all 3 datasets further enabled us to compare MI with and without neuro-feedback training (datasets 2b and 2a, respectively) as well as imagining button presses versus squeezing tennis balls (datasets 2a and 2b versus our own dataset, respectively).

A long-standing question in neuroscience and motor control is the extent of shared neural mechanisms between MI and ME [214]; though there is a general consensus that MI and ME at least share some important neural mechanisms. This similarity has been used in the MI-decoding literature, where some attempts to decode MI have relied on ME as training data [153]. Our results suggest that it is easier to decode ME than MI, at least when using EEG and relying on our decoding methods (Figure 26). Furthermore, we found that, on average, the decoding accuracy started at chance and then rose toward the time that participants were required to move or to imagine moving. After that it more or less plateaued. Interestingly, though perhaps not surprisingly, the accuracy level at the plateau, when using sliding windows, was lower than the accuracy for the full 4 s of ME (Figure 26). A likely contributing factor to this is that the sliding-window analysis decoded the EEG over shorter time windows than the full 4 s.

Another long-standing question when decoding EEG, and especially dense-array EEG, relates to how many and which electrodes (or channels) to use when recording the task. On the one hand, when using all channels (64, in our case), the set-up time for the task is longer, analyzing the larger dataset is more complex and computationally expensive, and brain signals unrelated to the task and noise are perhaps more often introduced. On the other hand, using only a limited number of channels, there may not be full coverage of brain regions that may be involved in the decision-making and action-preparation processes. We therefore wanted to identify the appropriate channels relevant to the MI task. We thus selected different combinations of channels, according to 10-20 system standard, based on what is known about the neurophysiology of decision making and action formation, [153, 155, 227]. Hence we included different EEG configurations in our study (see Results), with 3, 7, or 18, channels around the motor cortex (see Methods), or with all 64 channels [195]. Our analysis suggests that, without DA, the 18-channels configuration had the best average accuracy (81.73±2.5), at least on our dataset (Figure 30), while using all 64 channels resulted in the worse accuracy (71.47±1.9). Our results therefore suggest that, for MI decoding, it may be best to use only the 18 channels around the left and right motor region rather than all the channels. However, that result should be taken with a grain of salt, because when including DA, the tables were flipped, and it was the 64-channel configuration that did best, as described above.

One of the EEG configurations we tested included only 3 channels (C3, Cz, and C4)—this thus let us more directly compare our dataset to the two benchmark ones and the results of other studies. On those 3 electrodes, we achieved a mean accuracy of 79.95% for our dataset, while our analysis resulted in an accuracy of 89.11% and 86.28% for BCI 2a and 2b, respectively—all without DA. The higher accuracy for the benchmark datasets over our dataset might be due to the difference in tasks, the inclusion of neurofeedback (in BCI 2b), or that they perhaps ran participants who were better able to elicit good EEG data.

One general challenge of EEG decoding, especially with deep NNs, is obtaining enough data to train the numerous parameters in these large statistical models. The problem is compounded for MI tasks, because they are highly cognitively demanding. So, participants are easily fatigued and thus cannot produce a large amount of data in each experimental session. Bringing participants in for multiple sessions runs into issues of participant attrition for example. Another issue with collecting EEG over multiple session is the non-stationary nature of EEG signals [75]—i.e., the statistics of the EEG signals vary across time. As a result, a classifier trained at a specific time would tend to generalize increasingly poorly to data recorded at another time that was increasingly temporally removed—even for the same participant. This is a challenge for real-life applications of EEG, which must often work train on only limited amounts of data.

Some studies indeed strived for very lengthy data collection paradigms. One study, investigating MI control of 3D movement, had participants come back for up to 50 experimental sessions, which amounted to more than 20 hours of training per participant in some cases [228]. In another study, focusing on an EEG-based stroke-rehabilitation system [229], it took 12 weeks to collect enough data for three MI tasks, with each participant participating in 2 sessions per week [229]. While these are extreme examples, they highlight how common it is for participants to become fatigued after as little as 1 hour or less of data collection [230-232].

A promising solution to this dearth of data is to use DA, especially when using DL models on EEG data [28]. We therefore tested 5 disfferent DA techniques: sliding window, noise addition, GAN, Recombination of segmentation, and Fourier transform/wavelet. We further tested different magnification factors and hyperparameters (e.g., different window sizes for sliding window, various standard deviations for noise addition) for each technique we evaluated. Based on the guidelines in Lashgari et al. (2020) we evaluated the accuracy of the proposed method before and after DA. Our main objective was to find the best DA technique for each of the 3 datasets above. As far as we know, this is the first study to compare these various DA techniques as well as the different hyperparameters of the various techniques on benchmark datasets BCI 2a and 2b (Table 17). We found that different techniques work best for different datasets. For BCI 2a, GAN (conditional left vs. right and channels, m = 15) achieved the best accuracy, 93.6%. In contrast, sliding window (m = 2) gave the best accuracy for BCI 2b, at 87.83%. The DA step thus clearly boosted the performance of our proposed CNN as discussed below.

Interestingly, the BCI 2a dataset did not include neurofeedback training for the participants, while BCI 2b did. At the same time, the DA method that worked best for BCI 2a was a highly complex GAN with a large magnification factor, while that for BCI 2b it was a simple sliding window with a small magnification factor. So, one possible conclusion is that the neurofeedback training in BCI 2b, which effectively trained the participants to emit neural activity that would be better classified by the classifier, may have led to the superior accuracy from a simpler DA technique.

We also tested different DA techniques on our own dataset, which included 64 channels (see Methods). This achieved an accuracy of 86.61% (m = 15) with Fourier transform (Table 17). Using only 18 channels and the sliding-window DA technique (m = 15), we achieved an accuracy of 83.42%. Hence, using DA, we achieved higher accuracy with 64 channels than with 18 channels. Interestingly, without DA, the situation was flipped: the 64-channel data had lower accuracy

71.47(±1.9) than the 18-channel data 81.73(±2.5) (). This suggests that, if one dataset has lower accuracy than another without DA, it does not necessarily mean that the first dataset would also have lower accuracy than the second after DA.

As noted above, our accuracies were higher than those of Dai et al. (2020) (Table 15)—which was the top result in the field. Besides higher accuracies on average, our accuracies for individual participants were 90.54% or higher (Table 15), while Dai et al. (2020) achieved this accuracy or higher for just for 5 of the 9 participants. Further, we were interested in the effect of DA on the accuracy of their results. But they did not report that for BCI 2a. And we were unable to obtain their code. What is more, they did not specify the details of their DA techniques. We therefore reimplemented their architecture from their paper, as per the details in their methods, without DA, to compare it with our architecture without DA. The accuracy of our proposed CNN without DA—at 89.11% (±3.8; SE here and below) and 86.28% (±7.4)— outperformed the NN reproduced from Dai et al. (2020)—at 75.61% (±14.6) and 78.88% (±11.4)—for BCI 2a and 2b datasets, respectively.

Following the above, an exciting potential use of DA is to replace lengthy, multi-session data-acquisition efforts [228, 229]. For brain-imaging studies, it would decrease the time and funds that researchers need to spend on data collection and reduce the inconvenience of participants. This is especially pertinent for situations where gathering additional data is financially, ethically, or otherwise difficult. Though DA would of course come at the expense of additional training time for the statistical models. We tested this by training on only some of the training set—25, 50, 75, or 100% (see Table 10)—while testing different DA techniques on the remaining data.

We therefore tested the extent to which data augmentation could replace gathering more data, at least for the dataset that we collected (Table 18). More specifically, we collected 400 trials from each participant (see Methods) and used different proportions of the MI dataset (100%, 75%, 50% and 25%) to train the model. We then augmented those different proportions of the dataset with various DA techniques that have different magnification factors. Our aim was to test the effects of those DA parameters on classification accuracy (Table 10). With 100%, 75% and 50% of the data, m =15, 5, and 10, using Fourier transform achieved the highest accuracies, that were overall similar, at 86.61%, 88.26%, and 86.18% accuracy. Yet, classification based on just 25% of the data, m=15 and GAN (conditional left vs. right and channels) resulted in a lower accuracy, 82.18%. It might be that the smaller dataset required a more sophisticated DA technique that for the other proportions was needed to achieve its best accuracy. Though this accuracy was clearly lower than for the other proportions of data. This hints at the limits of DA for EEG.

It is well known that there is general anatomical similarity as well as structure-function correspondence among humans. But the anatomy of different brains also differs, at least to some extent, as does the structure-function correspondence. So, brain science typically operates at the aggregate level [216]. In particular, Smith et al. delineated structural differences, suggesting that the number of folds and thickness of the cortex could be associated with whole-brain functional networks [217]. Furthermore, inter-participant variability in topography occurs due to participant-specific cognitive style and strategy to perform a task over time [233], which could augment the underlying learning processes, e.g., motor and perceptual learning [219].

This question has clear implications for the analysis of EEG over groups of participants. We therefore wanted to investigate to what extent the number of participants over which we trained and tested our machine-learning model reduced the classification accuracy of the statistical model over that group. We thus trained and tested our model on all individual participants, on all pairs of participants, all triplets, quadruplets, and so on. It appears that, for all 3 datasets, the accuracy dropped most markedly between training and testing on individual participants to training and testing on pairs. Then there were diminishing decreases going from pairs to triplets, triplets to quadruplets, and so on, leading to roughly a plateau from groups of 6 participants and on. This suggests that the costs associated with inter-individual differences in brain structure and activity outweigh the benefits of the additional data when training over a group of participants. Though the decoding accuracy appeared to stop decreasing and reached somewhat of a plateau after around 6 participants. Future work, with a larger number of participants, could test the hypothesis that the accuracy would begin to rise again when training and testing over enough additional participants. One reason that this could happen is that the introduction of an ever-increasing number of additional participants might end up more than compensating for the neural variability between different brains. In other words, the advantages of the increasingly larger data available to train the model would outweigh the disadvantages of the variability across additional brains. Testing this hypothesis is left for future studies.

Following the discussion of inter-participant brain variability above, another key question in EEG analysis and especially for classification using DL is the extent to which a machine-learning model that was training on one group of participants could be generalized to new participants [234]. Put differently, we were wondering to what extent transfer learning, which has been increasingly used in the machine-learning literature, especially of late [235-237], would be useful for EEG classification using DL. We tested this by directly comparing two analyses. In the first, we trained a model on all but one participant and then tested it on that remaining participant (i.e., leave-one-participant-out classification). The second analysis comprised of again training on all but one participant, but then using transfer learning and finetuning the model on one part of the left-out participant. Finally, we tested the model on independent data from that participant (see Results 3.7). Our results clearly indicated that transfer learning led to higher accuracy than leave one out (Figure 9)—an increase in accuracy of 16.66%, 11.35%, and 18.6% for BCI 2a, BCI 2b, and for our dataset, respectively. This demonstrates the clear advantages of transfer learning for EEG analysis using DL. With DL models getting increasingly complex, the ability to finetune them for new participants rather than retrain them from scratch becomes increasingly important. In addition, our results suggest that the BCI community could use transfer learning with EEG to train a model on an existing dataset and then improve its performance for a new participant using only finetuning of the model [235, 238]. According to our results, this could markedly improve the performance of BCI classifiers.

Due to the good classification performance of our proposed neural-network architecture and the relatively simple data processing, without prior manual feature extraction, our method holds promise for online, real-time, EEG-based classification of MI. It is left to future work to test how well the system will work in real time. Further, based on our results, it seems useful to use transfer learning between participants in a real-time paradigm. Furthermore, our neural-network architecture uses an attentional mechanism that helps identify the most salient brain regions that drive the network's classification ability. However, we leave the analysis of these brain regions for future work.

5 A systematic review on Restricted Environmental Stimulation Therapy (REST) Floatation

5.1 Introduction

Early sensory deprivation research attempted to elucidate the effects of reduced environmental stimulation on a range of psychophysiological, motoric, perceptual, cognitive, and emotional measures by placing subjects in sensory deprivation chambers for extended periods of time [239]. During the 1950s, a new line of research, spearheaded by John C. Lilly at The National Institute of Mental Health, shifted away from this original focus and instead sought to investigate consciousness in the absence of external stimuli [240]. Notably, in 1954, Lilly would go on to create the first sensory deprivation tank.

Today, sensory deprivation tanks are known as Floatation-Restricted Environmental Stimulation Technique (REST) tanks. Additionally, the phrase “sensory deprivation tank” is often used interchangeably with “float tank”, “floatation tank”, “float pod”, and “isolation tank”. During a typical float session, participants rest in a supine position in a large, lightproof chamber filled with body temperature (35-37 degrees Celsius) saltwater contaning approximately five hundred kilograms of dissolved magnesium sulphate. This solution creates complete buoyancy for the user. As modern-day use of floatation tanks has shifted away from scientific inquiry towards commercial use, music is now played at the beginning and end of float sessions. These sessions typically last around an hour [241].

While Lilly was researching the effects of sensory deprivation though water immersion, psychologist Donald Hebb concurrently began to investigate the effects of Dry REST, commonly known as Chamber-REST [239]. Chamber-REST occurs in a soundproof and lightproof room, in which the participant lies on a bed and is encouraged to keep as still as possible. Sessions last for prolonged periods of time; while participants are allowed to leave at any point, most stay for a complete 24-hour period [239].

The initial response to sensory deprivation from both the general and scientific communities was largely negative. In fact, these unfavorable views were supported by some of the leading researchers themselves; for instance, Hebb hypothesized that people's susceptibility to external influences increased after being in the sensory deprivation chamber. Consequently, it become common procedure to have participants sign a psychological damage liability waiver before entering the chamber as well as include a "panic button" within the chamber (Suedfeld, 2012). The view that sensory deprivation led to adverse effects on cognition and consciousness persisted into the 1970s [243].

In 1968, Peter Suedfeld reignited a second wave of sensory deprivation research and, in 1980, coined the term 'Restricted Environmental Stimulation Therapy'. From Suedfeld's work, extensive academic research began globally, the majority in Sweden by Annette Kjellgren (2001 – 2017) from Karlstads Universitet [244-251]. In more recent years, American neuropsychologist Justin Feinstein (2016) and his team at the Laureate Institute of Brain Research (LIBR) conducted the first functional Magnetic Resonance Imaging (fMRI) study on the effects of sensory deprivation while in float tanks [252].

A number of review studies about floatation-REST have been published, generally focusing on subtopics such as the effects on stress [253, 254] and sleep [255]. However, to the best of our knowledge, no review exists which surveys recent trends of floatation-REST applications. Thus, a review of floatation-REST is warranted, with a particular focus on studies published after 1960. Here, we attempt to meet this need through a systematic review.

5.2 Methods

5.2.1 Search method for identification of related studies

The search was conducted on 3rd June 2020 within the Google Scholar and PubMed databases using the following group of keywords: "Restricted Environmental Stimulation Therapy" OR "Float pod" OR "Floatation REST" OR "Floatation pod" OR "Floatation Tank" OR "Sensory isolation".

Only studies that met the inclusion criteria (see Figure 31) are included below. Furthermore, duplicates among these databases were removed from the search results. The full texts of the remaining studies were then screened.

Inclusion criteria	Exclusion criteria
• Written in English • Journal and conference papers • Electronic preprints • Related to Floatation-REST	• Review papers • Lying on a bed in a dark, soundproof room (studies with both chamber-REST and Floatation-REST are examined to highlight the information from the floatation-REST segment)

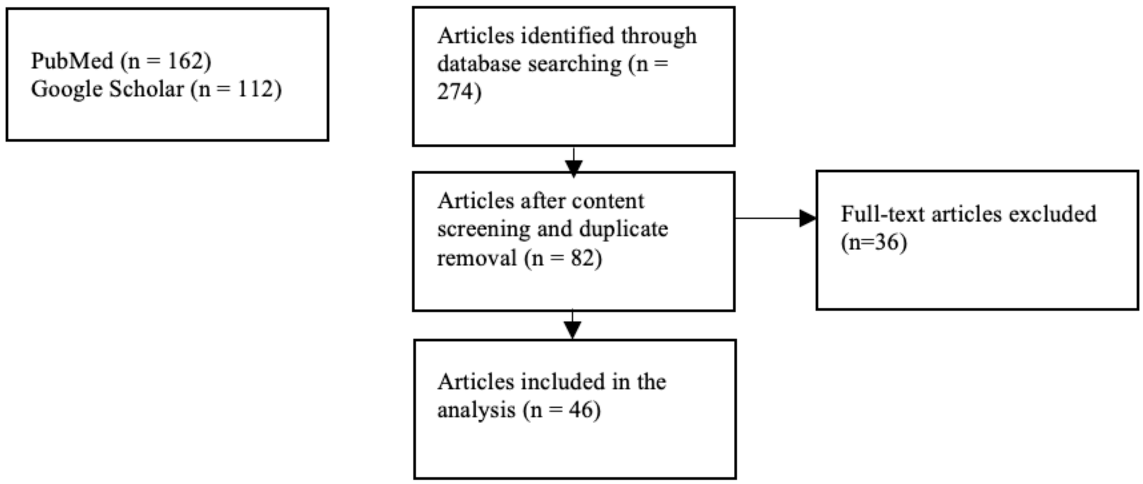

Figure 31. Selection process for the papers

The database queries yielded 274 matching results. Of those, 101 were duplicated. Manually screening the remaining 156 papers suggested that 91 of them were not relevant for this review (e.g., the keywords were included in the references rather than in the paper itself). We thus ended up with 82 papers, which we read carefully to make sure they met all our inclusion criteria. We found 36 that did not meet the inclusion criteria following closer inspection. Hence, based on our inclusion and exclusion criteria, 46 papers were selected for inclusion in this analysis (Figure 31).

Regarding our inclusion criteria, we should also mention more specifically that the water immersion and floatation research began in late 1950s as an attempt to elucidate the effects of monotonous or reduced environmental stimulation on a range of psychophysiological, motoric, cognitive, emotional and other measures. We found 2 review papers in 1983 and 2005 most of these studies [253, 254]. By looking to all papers, we can get more sense about the growing importance of Float pod after 2000.

5.2.2 Data extraction and presentation

For each selected paper, around 30 features were extracted covering seven categories: Article origin, data sample, application of float pod, experiment design, treatment delivery, questionnaire and tools, and results (Table 20).

Table 20. Data items extracted from each selected article

Category	Data item
Article origin	Type of publication (Journal article, conference article, or in an electronic preprint repository)
Sample	Quantity of data, # subjects, age, selection method, conditions
Application of float pod	Category of treatment using float pod
Experiment design	Randomization description or pre-post
Treatment delivery	Number and length of sessions
Questionnaire and tools	Outcome measures
Results	Effects of float pod on subjects

5.3 Results

5.3.1 Origin of the selected studies and different applications of float pod

Our search returned 46 journal papers, 19 of which were based in Sweden, 15 in the United States, 7 in Canada, and one in Australia, Japan, Switzerland, United Kingdom, and New Zealand each. The reviewed studies were published in 11 different journal groups: psychology (28%), medicine (17%), health research (11%), psychiatry (7%), cognition (7%), pain research (7%), perceptual and motor skills (7%), behavioral research (7%), addictive behaviors (4%), biofeedback and self-regulation, and stress management (2%) (Figure 32). Half of these studies focus on either the psychology or clinical applications of float pods.

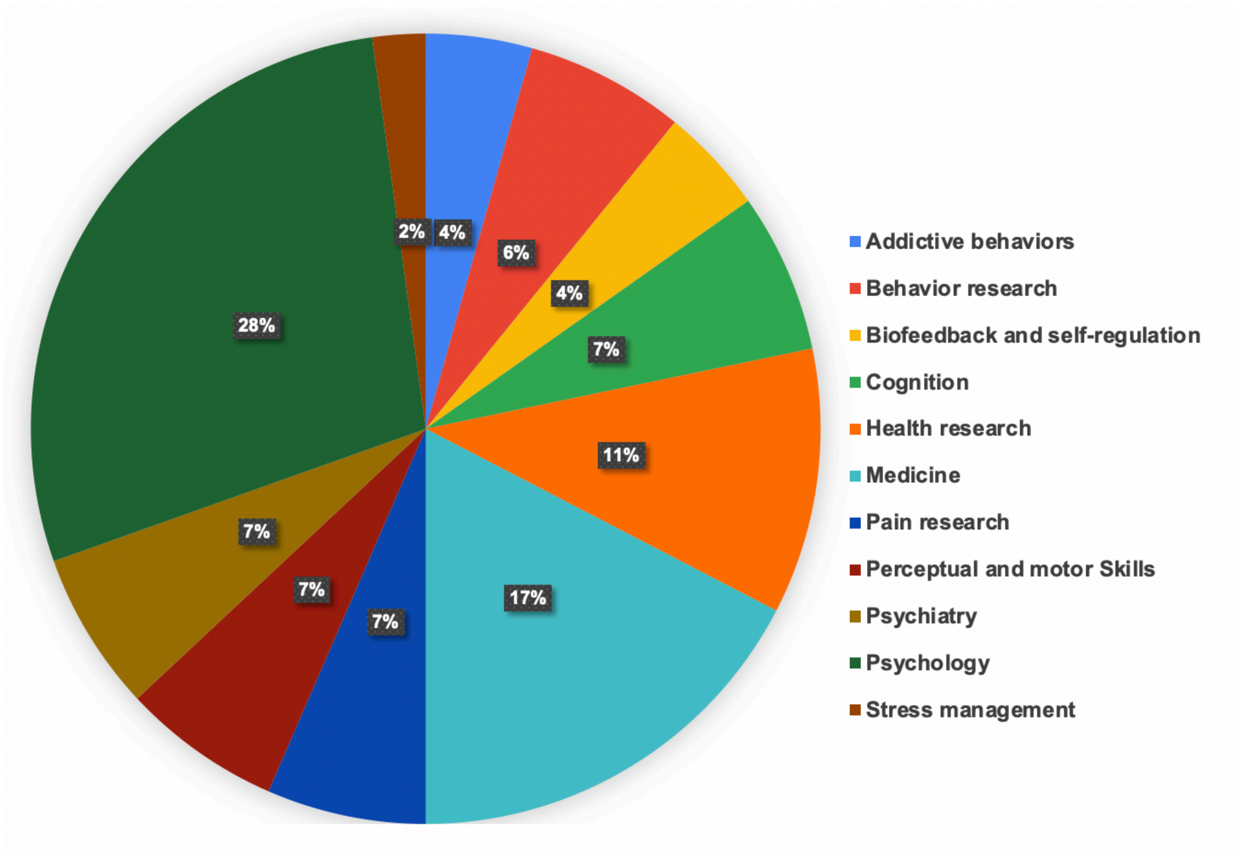

Figure 32. The percentage of different journals categories across all reviewed studies

Interestingly, we found that between 1960 to 2010, much of the research on float pods focused on their positive effects on pain relief. However, after 2010, there was a shift towards examining the effects of float pods on general physiology, sleep, and anxiety (Figure 33).

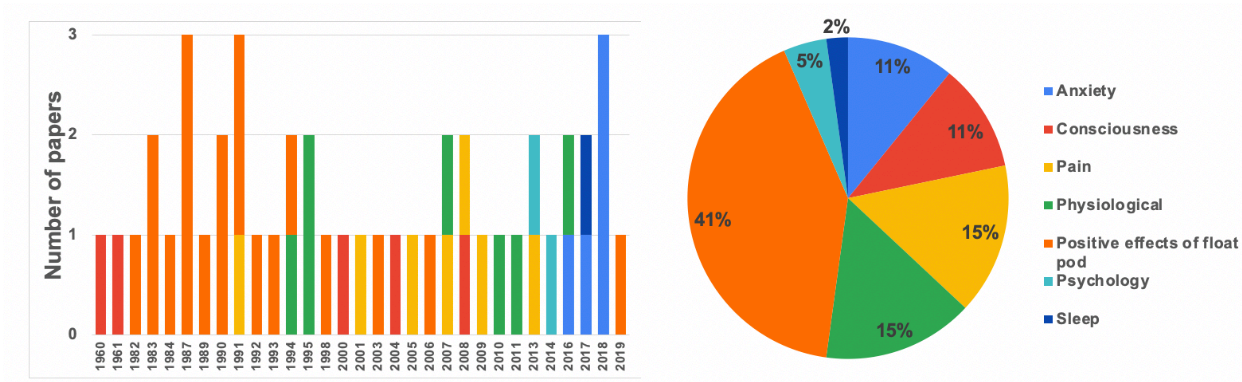

Figure 33. Different applications of float pods. (Left) Number of publications per domain for different applications of float pods. (Right) Percentage of different applications of float pods across all studies

5.3.2 Sample characteristics

The number of participants in the 46 selected studies ranged from one to 94, with an average of 29.9 subjects per study. The total number of participants across all studies was 1349, with a majority of these participants being women. Figure 34 shows the number of participants in each category of float pod application. Based on the figure, 609 subjects were in positive effects of FP, 233 in pain, 200 in anxiety, 169 in physiology, 71 in consciousness, 66 in psychology, and one in sleep.

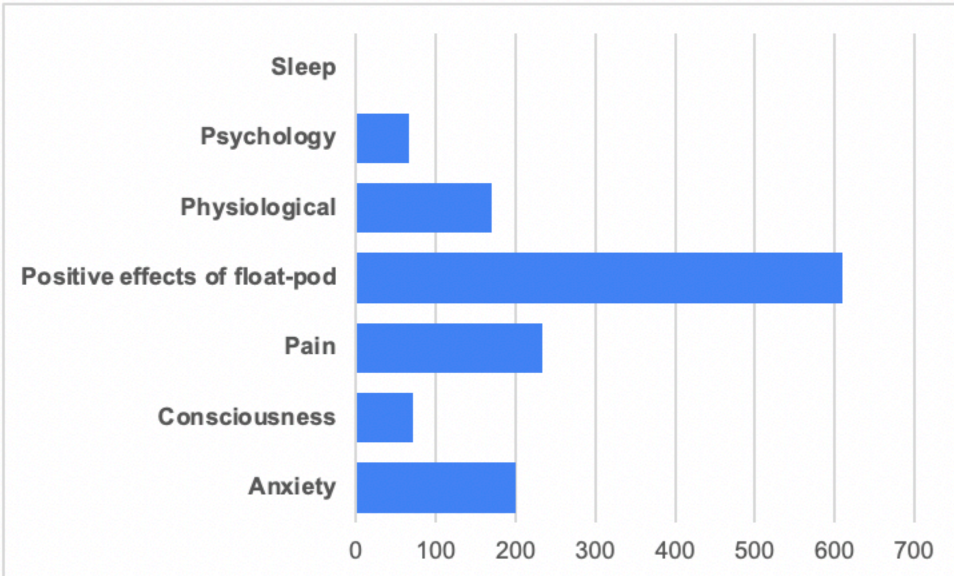

Figure 34. Total number of participants in each application of float pod across 46 studies

5.3.3 Experiment design

We found three distinct categories across the 46 studies concerning experiment design (see Figure 35). 31 studies used randomization (either a within-group or randomized controlled trial design), 3 studies used pre-/post-test measurements, and 3 studies used a single-subject design. Additionally, there were a number of that did not fit into any of these categories (listed as "Other" in Figure 35).

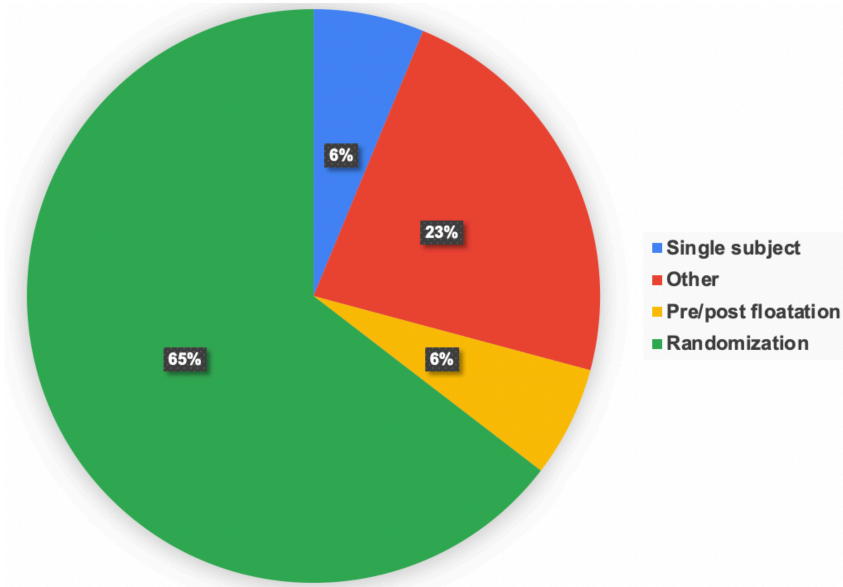

Figure 35. Experiment design for 46 studies

5.3.4 Treatment delivery

Floatation-REST sessions varied from 1 to 114 (mean: 13.33). Durations ranged from 35 to 240 minutes (mean: 63.28). Table 21 represents different treatment deliveries for each float pod application.

Table 21. Treatment delivery for different application of float pod

	Positive effects	Physiology	Pain	Anxiety	Consciousness	Psychology	Sleep
Possible combination	1session (60 min)	1 session (45 min)	9 sessions (45 min)	1 session (60 min)	1 session (60 min)	75 sessions (45 min)	22 sessions (60 min)
	2 sessions	1 session	12 sessions	1 session	6 sessions		

	(45 min) 2 sessions (60 min) 3 sessions (45 min) 3 sessions (60 min) 4 sessions (35 min) 5 sessions (60 min) 5 sessions (150 min) 6 sessions (50 min) 6 sessions (60 min)	(60 min) 1 session (90 min) 10 sessions (45 min) 35 sessions (45 min)	(45 min) 15 sessions (45 min) 114 sessions (90 min)	(90 min) 4 sessions (45 min) 12 sessions (45 min)	(45 min) NA sessions (up to 240 min)		

	8 sessions (40 min) 10 sessions (45 min) 12 sessions (45 min) 12 sessions (90 min) 20 sessions (40 min)							

5.3.5 Questionnaire

Self-rated questionnaires were used to assess subjects' experiences during the floatation sessions. Table 22 shows a list of all of the questionnaires used across the 46 studies.

Table 22. Different questionnaires used in this review paper

Test name	Description
PSWQ	Penn state worry questionnaire
GAD-Q-IV	the generalized anxiety disorder questionnaire 4th edition
MADRS-S	Montgomery-Asberg depression rating scale
PSQI	the Pittsburgh sleep quality index
DERS	the dysfunctional emotional regulation scale

MAAS	the mindful attention awareness scale
EDN	the experienced deviation from normal state scale
ASI-3	Anxiety sensitivity Index
OASIS	Overall anxiety severity and impairment scale
PHQ-9	Patient health questionnaire
PSS	Perceived stress scale
SDS	Sheehan disability scale
HM	Happiness Measure
STAI-Y state form	State trait anxiety inventory
PANAS-X	Positive and negative effect schedule-expanded form
KSS	Karolinska Sleepiness scale
VAS	Visual Analogue Scales
HAD	Hospital Anxiety and Depression scales
PAI	Pain Area Inventory
SE	Stress and energy
LOT	Life orientation test
DTs	sensory detection thresholds
PE	pain endurance
Syllogisms I-II	A test presented in two versions [36] that measures the ability of logical and deductive thinking
Beer Can/Brick	A test of divergent that measures the number of relevant responses for how many different ways one may use a beer can or a brick, respectively.
FS	change stability
Composition Test	Participants were instructed to write an essay based on four words: ambition, choice, ring, and disappointment
SQ	Sleep quality
GHQ-12	The General Health Questionnaire
MDMQ	multidimensional mood-state questionnaire (for pleasant-unpleasant, awake-sleepy, calm-restless)
Muscle Soreness	0 (not sore - 10 (max soreness)

Figure 36 visualizes the aggregated information about the questionnaires of the reviewed studies. This figure helps to reveal the trends in the types of questionnaires used for the different types of float pod applications.

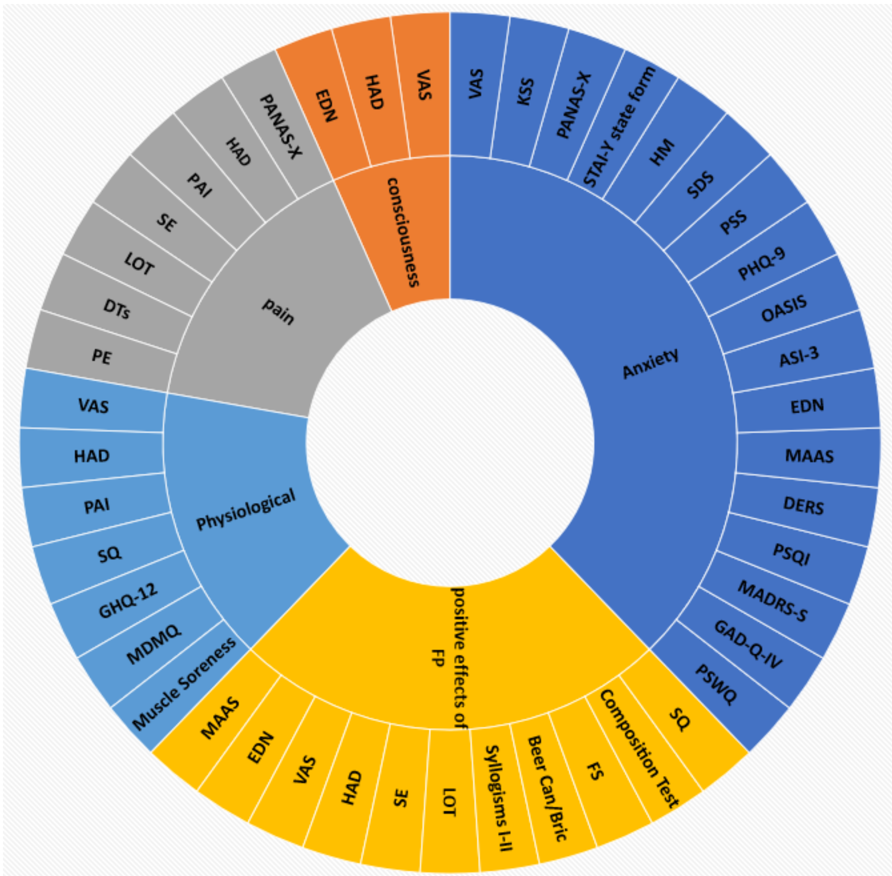

Figure 36. Aggregated information of different questionnaires across different types of float pod applications

5.4 Float pod application

The applications of float pod in these studies fell into 7 groups: positive effects of FP (28%), pain (20%), anxiety (20%), physiological (16%), consciousness (8%), sleep (4%) and psychology (4%).

5.4.1 Positive effects of float pod

Most of the studies in this section are related to the first studies ever conducted on float pods which mainly focus on their general effects.

Fine and Turner (1982) piloted a study in which they investigated the effects of controlled frequent brief REST relaxation session on blood pressure in subjects with borderline essential hypertension [256]. Three subjects were involved in the study, identified as subjects A, B and C. Subject A participated in 20 float sessions (40 minutes each) over a 2-month period for treatment of essential hypertension. Subjects B and C participated in 20 float sessions (40 minutes each) over a 10-week period. Each subject had their blood pressure taken before every floatation session. The Spielberger STAl X-l was also administered pre-test and post-test to subject A. Post-treatment follow-ups were conducted in 3-month, 6-month, and 10-month increments. Results of this experiment indicated reductions in both systolic and diastolic blood pressure occurring during the treatment period and persisting throughout the follow-up sessions for all subjects.

In 1983, Turner and Fine continued their line of inquiry by investigating the effects of brief repeated floatation REST assisted relaxation on plasma cortisol, ACTH, and luteinizing hormone (LH) in normal healthy subjects [257]. 12 male volunteers from a medical school class were recruited as subjects and randomly assigned to two equally-sized conditions: a REST-assisted relaxation procedure and a control relaxation procedure without REST. For the REST-assisted relaxation procedure, subjects were placed in a float tank with dim lighting. In contrast, control subjects in the non-REST relaxation procedure rested in a reclining chair with the same dim lighting. Throughout both procedures, the lights would dim until dark, and a relaxing audio message would play for 90 seconds. Lights would be then be brought back to the original dim lighting for 60 seconds before the cycle would repeat for a total of 10 times. Both conditions took baseline measures from subjects as they quietly reclined in a chair for 30 minutes. Each subject experienced two baseline sessions (1 and 2), four 35-minute REST (or non-REST) relaxation sessions (3, 4, 5, 6), and two follow-up sessions (7 and 8). Blood samples were drawn before and after sessions 1, 2, 5, and 8. At the end of each follow-up session, subjects filled out a brief questionnaire surveying their overall relaxation experience. Turner and Fine found that the repeated REST-assisted relaxation sessions were associated with a significant decrease in plasma cortisol and ACTH throughout all eight sessions. However, there were no significant changes in LH across each session [257].

Suedfeld et al. (1983) attempted to establish the general effects of floatation by conducting an experiment on healthy normal subjects independent from any goal- or research-oriented program [253]. 27 subjects were recruited through a commercial tank facility in Vancouver, Canada to participate in a single 1-hour long REST session. A total of 5 measures were administered: The Arousal Seeking Tendency test, which was administered pre-float [258]; the Subjective Stress Scale, administered both pre-float and post-float [259]; and the Body Consciousness Scale [260], the Russell Mood Scale (Person) [261], and the Russell Mood Scale (Place) [262], which were all administered post-float. The data indicated that an hour of floatation was generally a relaxing and pleasant experience for most users.

In 1984, Jacobs et al. attempted to compare the effect of floatation-REST on relaxation to that elicited in a normal sensory environment [263]. 28 subjects were recruited and randomly assigned to either an experimental group or a control group. All subjects participated in a simple relaxation program consisting of guided point-to-point relaxation, breathing techniques, and visual imagery techniques. Physiological measures such as EMG, blood pressure, and skin temperature were taken pre-test and post-test. Psychological measures like the Subjective Relaxation Questionnaire were also administered. The experimental group completed 10 floatation-REST sessions, each lasting for 45 minutes, while the control group also completed a total of 10 45-minute sessions, during which subjects relaxed in a supine position in a small room. The results of this study supported the hypothesis that, given adequate preparation, subjects who practiced relaxations techniques in a floatation-REST environment achieved greater levels of relaxation than subjects who practiced the same techniques in a normal sensory environment [263].

In 1987, researchers tested whether variations in REST type, session duration or message presentation schedule could increase the success rate and/or reduce financial and temporal costs associated with smoking [264]. Their study included 83 subjects who had smoked at least one pack of cigarettes per day for the past 5 years. The study compared 12-hour chamber-REST sessions with 24-hour sessions. Another group of subjects also engaged in 5 one-hour floatation-REST sessions. Messages used in the study promoted things like protecting one's body from cigarettes, teaching deep-breathing relaxation techniques for cravings, describing imagery to deal with emotions, preparing for relapse, and congratulating the subject for becoming a non-smoker [264]. There were 4 different presentation schedules for the messages: 1) Distributed: one message an in the last 45 minutes, one message an hour after, and others at irregular intervals. Massed: two messages within the first hour and at the midpoint and one 45 minutes before the end. Self-demand: a message after the first hour and whenever the subject pressed a button. Floatation: a one-hour float without a message, then one message around the midpoint of each of the four subsequent floats, and one message in the last 45 minutes in the final float session. Results were measured in reduction in smoking and abstinence from smoking over a 3 month and 12 months follow up. Their results showed that chamber REST was indeed effective as a smoking intervention method, although parametric variation did not identify an optimal combination. In contrast, their floatation results did not yield as promising data in favor for reduction or abstinence from smoking, leading them to suggest that the application of floatation-REST requires further investigation [264].

Suedfeld et al. (1987) sought to examine the effects of floatation-REST on creative performance [265] using a within-subjects experimental design. Subjects included 7 full-time faculty members from the Department of Psychology at the University of British Columbia. Each subject completed 8-12 60-minute float sessions followed by 30 minutes of dictating ideas into an audio recorder and an additional 8-12 sessions of sitting in an office and completing the same dictation task. Dependent variables included the Profile of Mood States (POMS) Questionnaire and a list of self-reported measures including: (1) The number of times subjects changed from one idea or topic to the next; (2) the number of times a novel idea or topic came to mind; (3) the level of "quality" (i.e., creativeness) of each idea or topic rated on a scale from 1-10, with these ratings being made 1-3 months after the completion of each session; and (4) the number of ideas that eventually led to new research publications, grant proposals, and other similar accomplishments, with this measure being completed 12-15 months after the completion of all sesssions [265]. Suedfeld et al. found that new ideas generated during floatation-REST were rated as more creative than those originating in office settings, confirming their hypothesis [265].

In 1989, Turner et al. examined the effect of light on relaxation during floatation REST by measuring plasma cortisol levels, mean arterial pressure, and psychometric parameters [266]. 21

subjects were recruited and then separated into 2 groups: one group completed float-REST in the presence of light, while the other group completed float-REST in complete darkness. The study took place over the course of 6 weeks, during which participants experienced a total of 8 float sessions (40 minutes each) and had their blood drawn before and after each session for a total of 16 blood drawings. Psychological measures such as the Taylor Manifest Anxiety Scale, the Marlowe-Crown Social Desirability Scale, and the Profile of Mood States (POMS) were also administered. Results indicated that, across all sessions, repeated REST was associated with decreases in both plasma cortisol levels and mean arterial blood pressure, regardless of the presence or absence of light [266].

In 1990, Suedfeld and Bruno carried out a study designed to measure the effect of REST, coupled with a guided mental imagery task, on athletic performance [267]. The study consisted of 30 subjects, all of which had little to no experience playing basketball, and divided them into 3 treatment groups: REST, alpha chair, and a control group. Those in the REST group floated in a dark, soundproofed tank; those in the alpha chair condition rested in a shell-like chair which enclosed the subject and was specifically "designed to induce relaxation and concentration"; and those in the control group sat in a regular comfortable armchair situated within an office. Before each treatment session across all three conditions, subjects were asked to shoot 20 regular free throws. Then, while in their assigned environments, subjects listened to an hour-long tape recording that directed them through a multisensory imagery task, in which subjects were asked to listen to a guide on how to make basketball free throws. Finally, after finishing the recording, subjects completed an additional 20 free throws. Questionnaires were administered to assess previous experiences with playing basketball as well as participating in guided imagery tasks;

subjects' confidence in their athletic ability was also measured using self-reported predictions of their own free-throw success [267]. The results demonstrated significant improvement in athletic skill after REST. Additionally, those in the REST group reported higher post-treatment levels of confidence in athletic performance compared to the alpha chair group, although no significant difference existed compared to the control group [267].

McAleney et al. (1990) investigated the joint effects of floatation-REST and a visual imagery task on the competitive performance of intercollegiate tennis players [268]. The study consisted of 20 university varsity tennis players (10 men and 10 women) and involved 2 conditions: floatation-REST with the visual imagery task, and a visual imagery-only condition in which participants were exposed to imagery and messages in a room without restrictions. The visual-imagery message was eight minutes in length and included images of five or six alternative skills where the subjects visualized themselves making optimal tennis shots. All subjects participated in six 50-minute treatment sessions across a 3-week period. To measure improvement in athletic performance, tennis matches from both before and after completion of the treatments sessions for each participant were videotaped. Performance was then assessed based on three key aspects: first service, key shot, and points won/lost. Their results indicated that floatation-REST, in combination with a visual imagery task, enhanced the performance of one previously well-learned key athletic skill amongst varsity tennis players [268].

Turner and Fine (1991) conducted a study on the effects of repeated sessions of floatation-REST on overall plasma cortisol levels as well as plasma cortisol level variability [269]. The study recruited 27 subjects who were pair-matched based on initial values of plasma cortisol and were then split into two groups: floatation-REST and non-REST. For their treatment sessions, those in the floatation-REST group rested in a supine position within a dark floatation chamber, while those in the non-REST group rested in a fully reclined cushioned chair. All subjects participated in eight 40-minute sessions across 3 weeks. Blood samples were taken before and after sessions 5-8. The results showed that both the average plamsa cortisol concentration and variability decreased across sessions in the REST group, whereas no significant change was detected in the non-REST group [269].

In 1991, Jeffery D. Wagaman investigated the joint effects of REST and an imagery training task on subjective and objective measures of athletic performance while playing basketball [270]. 22 male basketball players were randomly assigned to participate in either a REST-imagery or an imagery-only treatment. During each 60-minute session, those assigned to the REST-imagery treatment rested in a supine position within a float tank, while those assigned to the imagery-only treatment sat in a comfortable chair situated within an office where they were free to do as they pleased once a tape relaying the imagery training task had finished playing. The imagery task used in this study was adapted from that originally created by Lee and Hewitt (1987) which was designed to enhance athletic performance [271]. The 20-minute long tape guided subjects to relax and visualize the steps of an optimal basketball game performance. An objective measure of athletic performance was made using a point-based system assessing subjects' individual performances during real basketball games. Successful shots or passes resulted in subjects gaining a point, whereas unsuccessful shots or passes resulted in subjects losing a point. Subjective measures of athletic perfromance were also collected using the Performance Evaluation Questionnaire (AAHPERD, 1984), which was a short survey that measured partcipants' subjective perceptions on their own performance. Wagaman found that floatation REST, when accompanied with an imagery training task, led to better athletic performance in comparison to completion of just the imagery task.

Forgays et al. (1992) evaluated the potential ability of floatation to enhance creativity, as measured by the Guilford Fluency Test and similar indices [243]. 30 subjects (13 male, 17 female), half of whom spent one hour in a float pod and the other half in a dark room, completed a number of pre-test and post-test measures assessing creativity and personality/affect. Subjects who completed the float pod treatment showed significant increases on the Guilford Fluency Test pre- to post-float along with meaningful increases on other thinking measures as compared to non-floating control subjects. Floating was also associated with a decrease in anxiety/tension, depression, hostility, and fatigue, and an increase in vigor. Forgays et al. speculated that the increased creativity brought about by floatation may result from some of these cognitive and behavioral changes.

In 1993, Suedfeld et al. examined the effects of REST plus imagery training on perceptual-motor accuracy in comparison to both REST and imagery training alone [272]. 40 subjects who had played a game of darts at least twice during the previous year were recruited and randomly assigned to one of four groups: imagery only, REST only, REST plus imagery, and a control group. Subjects in the imagery-only group sat in a small room. After 40 minutes, participants were asked to listen to a 13-minute tape recording that started with a brief exercise during which subjects were asked to describe the “feel” of throwing a perfect bull’s eye. Subjects in the REST-only group rested in a supine position within a floatation tank for 1 hour without any imagery training. Subjects in the REST plus imagery group also rested in a supine position within a floatation tank, but. after 40 minutes had elapsed, also coompleted the imagery training. Lastly, subjects in the control group were placed within a small room and were told they could do as they pleased for the entire 1-hour session. Suedfeld used a pretest-posttest design, and participants were given 10 practice throws before they completed 20 recorded dart throws at a standard dart board. Subjects were scored based on the distance of their dart from the bullseye on the dart board. The study results indicated a marked improvement in dart-throwing accuracy after floatation-REST irrespective of wheter imagery training had been completed during the float. Imagery training alone did not increase dart-throwing accuracy, and no synergistic interaction between floatation-REST and imagery training was found. From this, Suedfeld et al. concluded that floatation REST improves accuracy and precision in athletic performance for sports that do not require speed or strength [272].

A study in Switzerland investigated the neuroendocrine changes that occur after REST using a crossover design comparing floatation-REST to a control treatment [273]. Five healthy male subjects were recruited. For the floatation-REST treatment, the subjects rested in a supine position within a float tank; for the control treatment, subjects rested in a supine position in a dimly lit room. Each treatment session lasted for 60 minutes. Subjects completed three pre-test measures: the Minnesota Multiphasic Personality Inventory [274], the Sensation Seeking Scale Form IV [275], and von Zerssen's Mood Questionnaire [277]. Before the experiment, subjects also engaged in two habituation floatation-REST sessions over a period of 2-3 weeks. Three subjects started the experiment with the floatation-REST treatment, while the other two subjects started with the control treatment. Three blood samples were drawn 30 minutes before the start of each treatment. Blood samples were then drawn every 15 minutes for 2 hours after the completion of both treatments. Subjects completed two post-test measures: the Stanford Sleepiness Scale [276] and the von Zerssen's Mood Questionnaire [277]. Biochemical analyses of the blood samples assessed the following: cortisol, TSH, prolactin, LH, GH, melatonin, ß-endorphin, ADH, T4, GABA, HVA, magnesium, urinary VMA, osmolality, and lactate. The results demonstrated no significant neuroendocrine changes that could be attributed to floatation-REST across all five subjects, apart from a 37% average reduction in urinary VMA. Psychological results, however, indicated an increase in sedation, relaxation, and euphoria resulting from floatation-REST [273].

In Sweden, a study was conducted to investigate the effects of REST on creative problem-solving and originality [278]. The study consisted of 2 experiments. The first experiment recruited 40 subjects and separated them into a control group and a floatation group. Participants' pulses were measured every 60 seconds one week before the experiment for all subjects. At the beginning of the experiment, subjects also completed the FS-Change and Stability Test [279], which measured subjects' attitudes towards change and stability. The control group engaged in a creative problem known as the "Cheap Necklace Problem" (Silveira, 1971; Best, 1995) for 5 minutes before sitting in a reclined armchair for 45 minutes. After the 45 minutes, subjects could return to working on the puzzle to completion or until 25 minutes had elapsed. The floatation group worked with the same cheap necklace problem for the first 5 minutes but then floated in a supine position within a float tank for the following 45 minutes. After the 45 minutes had elapsed, subjects could continue to work on the problem to completion or until 25 minutes had elapsed in the same fashion as the control subjects.

The second experiment was comprised of 54 subjects randomly assigned to one of three conditions: floatation-REST, dry-REST, and non-REST. Like the first experiment, participants' pulses were measured every 60 seconds one week before the experiment for all conditions. The floatation-REST group rested in a supine position within a floatation tank for 45 minutes before taking three psychological pen-and-paper tests: the Syllogisms I Test [280], the FREGO Test

[281], and a self-constructed assessment in which subjects were tasked with coming up with as many consequences as possible to six dramatic events (e.g. “What would happen if there suddenly was a new ice age in Northeurope?”) [278]. The dry-REST group completed the same tests, except they rested on a couch for 45 minutes in a dark, sound-proofed chamber prior to completing them. Similarly, the non-REST group followed the same procedure but sat in an armchair for the first 45 minutes, where they were allowed to either read magazines or do nothing.

The study found three main results: 1) In the first experiment, subjects in the floatation group took longer to complete the problem than those in the non-REST group; 2) the quicker subjects in the first experiment solved the problem, the greater their average heart rate variability tended to be; and 3) in the second experiment, subjects in the float-REST group scored higher in originality in comparison with the dry-REST and non-REST groups [278].

Forgays et al. ran a study attempting to validate float-REST, coupled with motivational auditory messages, as an effective treatment for smoking cessation and/or reduction in heavy smokers motivated to quit [282]. Participants completed follow-up measures for up to 12 months post-intervention. Smoking reductions after 12 months compared favorably with other interventional techniques, and floats of longer durations appeared to be more effective at reducing smoking than shorter floats. The motivational auditory messages were not found to affect the strength of the float-REST treatment. Control subjects reduced their smoking more than the experimental subjects, suggesting that the procedures used on them were actually more effective interventions.

In 2003, Norlander et al. examined the effects of float-REST on stress and cognition by completing a two-part study [283]. In the first study, 38 participants were recruited and randomly assigned to either a group that completed a single 45-minute float session or a group that completed three 45-minute float sessions. The study used a pretest-posttest design and administered the LOT, Syllogisms (either version I or II), and beer can/brick (either beer can or brick) measures. When subjects completed the Syllogisms and beer can/brick tests post-intervention, they were administered the versions of the measures that they had not received pre-intervention.

In the second study, which used a 2x2 design, 32 participants were recruited and randomly assigned to either the floatation-REST or chamber-REST group as well as to the stress or non-stress group [283]. Following each treatment session, subjects were tasked with writing an essay which was later assessed for elaboration, liveliness, originality, and realism.

The results indicated that: 1) Float-REST resulted in no differences concerning divergent or logical thinking, regardless of whether one or three float sessions had transpired; 2) float-REST led to higher scores in orginality but lower scores in deductive thinking; 3) chamber-REST led to higher scores in realistic and elaborated thinking; 4) subjects in the floatation-REST/stress group were more lively compared to their floatation-REST/non-stress group counterparts; 5) subjects in the chamber-REST/stress group demonstrated greater realism compraed to their chamber-REST/non-stress group counterparts; 6) both floatation-REST and chamber-REST were equally effective in reducing stress, i.e., showed comparable efficacy as relaxation techniques, although floatation-REST altered consciousness to a greater extent compared to chamber-REST; and 7) correlational analyses indicated that the more adaptable, optimistic, and receptive to change one was, the more originality one exhibited in his/her essay; additionally, those who scored higher in liveliness tended to score lower in realism.

Bood et al. (2006) aimed to replicate earlier findings that demonstrated increased wellness after float-REST as well as investigate its long-term effects [284]. 70 subjects were recruited, 54 of which were women and 16 men, participated. All subjects had been diagnosed as having stress-related pain; an additional 26 subjects also had been diagnosed with burnout depression. Participants were randomly assigned in equal numbers to either a control group or a floatation–REST group; each participant completed a total of 12 sessions. Results indicated that, for those in the float-REST group, levels of pain, stress, anxiety, and depression decreased, whereas sleep quality, optimism, and prolactin increased. These positive effects also generally persisted 4 months after completing the intervention. Bood et al. concluded that float-REST is an effective method for the treatment of stress-related pain.

Broderick et al. (2019) examined the effects of float-REST on recovery from exercise [285]. They hypothesized that float-REST would reduce muscle soreness, improve sleep quality, and enhance subsequent performance recovery in trained athletes. Following pre-exercise testing and warm-up, participants performed an exercise circuit known as the Basketball Exercise Simulation Test. Upon completion of the test, post-exercise measures (performance test, saliva collection, perceptual measures, and algometer) were administered to determine the level of fatigue experienced by the participants. Once all post-exercise tests were completed, participants were assigned to one of two recovery intervention: float-REST or a control intervention. Participants then went home to sleep before returning the next morning to perform the same Basketball Exercise Simulation Test. Finally, perceptual measures were recorded once more 24 hours after completing the test. Float-REST was found to significantly enhance CMJ ($p = 0.05$), 10 m sprint ($p = 0.01$) and 15 m sprint performance ($p = 0.05$) with small to moderate effects ($d = 0.21–0.68$) for all performance measures, except CMJ (unclear), compared to the control group. The results also show significantly higher pressure-to-pain thresholds across all muscle sites (p's < 0.01) and lower MS and PF 12 h following float-REST ($p < 0.05$). All sleep measures resulted in small to large effects ($d = 0.20–0.87$) with a significantly greater perceived sleep quality ($p = 0.001$) for subjects in the float-REST group compared to thosein the control group. There were no significant differences and a trivial effect size between trials for changes in cortisol concentration. The use of float-REST following exercise in the late afternoon/early evening may be an effective strategy to enhance relaxation and subsequent sleep. Furthermore, this is the first study to our knowledge to show the benefits of floast-REST on next-day performance recovery, specifically in measures of power and speed. Future research should attempt to control for the possible placebo effect of such a treatment or include the comparison of other post exercise recovery strategies. Additionally, future research may give insight into whether habitual or regular floaters have different cortisol responses than those relatively new to the technique.

5.4.2 Pain

In 1991, Wallbaum et al. investigated the treatment potential of float-REST for individuals with chronic tension headaches [286]. 31 subjects were recruited and assigned to one of four treatments: 1) Chamber/control: subjects rested in a supine position on a bed situated within a small, dimly lit room; 2) chamber/tank: subjects rested in a supine position in both a float tank and, later, in the same dimly lit room used in the control treatment, 3) chamber/relaxation: subjects completed a series of muscle relaxation exercises while resting in a supine position on a bed in the same room used in the control treatment; and 4) tank/relaxation: subjects completed a series of muscle relaxation exercises while resting in a supine position in a float tank. All subjects completed a total of 8 90-minute sessions across 4 weeks. A Headache Diary [287] was used by subjects to track the frequency, duration, and intensity of their headaches. Subjects maintained entries in these diaries from at least two weeks before the first treatment until six months after treatment completion. These entries were used to calculate a headache index for each subject along with an index of self-reported use of headache-related medications. The results indicated that floatation-REST may serve as an effective treatment for chronic tension headaches. Additionally, the combination of float-REST and muscle relaxation exercises may increase the duration of positive treatment effects [286].

Kjellgren et al. (2001) investigated whether floatation-REST may serve as an effective and reliable method for treating chronic pain as well as psychosomatic symptoms resulting from long term muscle tension and/or stress-related headaches [288]. 37 patients were recruited and randomly assigned to either complete a nine-session float-REST treatment or no treatment at all. For those in the float-REST treatment group, each float session lasted 45 minutes, and subjects were asked to complete a total of 9 sessions across three weeks. The study used a pretest-posttest design; the administered measures asked participants to estimate self-assessed pain severity, duration, onset, and treatment efficacy, as well as report any other health-related experiences or symptoms. Blood samples were also taken from each participant. Information regarding sleep, dreams, and tobacco and alcohol habits of each subject was also collected. The results indicated that subjects in the float-REST treatment group who reported the most severe perceived pain intensity experienced significant reductions in pain intensity, whereas subjects who reported lower perceived pain intensity were not affected by float-REST. Furthermore, circulating levels of the noradrenaline metabolite 3-methoxy-4hydroxyphenylethyleneglycol were reduced significantly in the float-

REST treatment group but not in the control group, whereas endorphin levels were not affected by floatation. Floatation-REST treatment also elevated the participants' optimism while reducing anxiety or depression. Finally, subjects in the float-REST treatment group reported being able to fall asleep more easily compared to those in the control group. The present findings tentatively suggest that floatation-REST may effectively alleviate low to moderately severe pain induced by muscle tension. Further investigations should aim to extend the scope of the measured subjective changes and neurochemical markers (e.g. with the serotonin metabolite 5-hydroxyindole acetic acid, oxytocin and cortisol).

In 2005, Bood et al. examined the effect of attention on the therapeutic effects of float-REST using a single-blind design [289]. 32 subjects were recruited and randomly assigned to either receive special attention throughout the experiment for 12 weeks (high attention group) or receive normal attention for only six weeks (normal attention group). The study used a pretest-posttest design; subjects completed a number of measures assessing pain (intensity, areas and types, frequency, duration, onset and treatment efficacy). Subjects also completed the PANAS, LOT, PAI, SE and HAD questionnaires. The results demonstrated that participants exhibited lowered blood pressure; reduced pain, anxiety, depression, stress and negative affect; and increased optimism, energy and positive affect. However the results were largely unaffected by the degree of attention-placebo or diagnosis. Also, analysis showed that the participants reported reducing their alcohol intake per month during treatment period. Also analysis showed that the participants reported reducing the number of types of pain medication during the treatment period.

This group in 2007 investigated whether 33 float-REST sessions were more effective in treating stress-related ailments compared to 12 sessions [290]. The results after 12 sessions were similar to those found in other float-REST studies (Bood et al., 2005; 2006); surprisingly, however, there were little to no differences in treatment effects between the subjects in the 33-session group versus those in the 12-session group. Although the number of comprehensive pain areas was significantly lower after 33 floatation sessions, for PAI, most severe pain intensity, normal pain intensity, and pain frequency, 12 sessions were enough to get considerable improvements across these measures, and no further improvements were found after 33 sessions. A similar pattern was observed among the stress-related psychological variables. After 12 floatation sessions, 25% of participants experienced decreased stress, 26% experienced decreased anxiety, 11% experienced decreased negative affect, and 32% experienced lower levels of depression, while dispositional optimism increased among 10% of participants and sleep quality among 18%; there were no further improvements after 33 floatation sessions. Additionally, no effects in blood pressure were observed after 12 floatation sessions, while there was a significant effect for diastolic blood pressure after 33 sessions. Finally, as expected, subjects diagnosed with burn-out depression measured higher for depression and negative affect than did patients without a diagnosis. Consequently, it was these diagnosed subjects who saw significant improvements with regards to depression and negative affect.

In 2008, Edebol et al. investigated whether floatation-REST might be able to help treat chronic whiplash associated disorders (WAD) [291]. Six women and one man, all diagnosed by licensed physicians as having chronic whiplash associated disorder, were recruited for this study. Two of the participants were beginners in floatation-REST (i.e., previously participated in 2 to 3 float sessions), while the remaning five were more experienced (i.e., previously participated in 7 to 15 float sessions). The floatation-REST treatment involves the creation of new elements in the world of experience. The spiral illustrates how the participants with chronic WAD experience the short-term effects of the floatation-REST treatment. After completing the floatation-REST treatment, subjects reported improvements in pain reduction and stress management, changed attitudes towards their pain, renewed coping strategies, openness to perceptions, and the sense of a centered self. Floatation-REST treatment favors the patient in a physiological sense as well as mentally and cognitively. The environment in the floatation-REST tank supports reflection and relaxation to a great extent. In comparison to other common treatments of chronic WAD, such as hydrotherapy, electromagnetic field therapy, radio wave neurotonomy, cognitive behavior therapy, and physiotherapy, floatation-REST treatment may offer the unique ability to explicitly benefit the mental and cognitive health of WAD patients. However, this study represents only an initial evaluation of the effects of float-REST on WAD, so future research must be conducted before floast-REST can be established as a reliable and effective treatment intervention.

Bood et al. (2009) explored how sex differences may alter the effects of float-REST among patients diagnosed with stress-related pain [292]. The results indicated that float-REST had beneficial effects on stress, anxiety, depression, sleep quality and pain, and that these effects were generally consistent regardless of sex differences. The only notable difference between the sexes was that men were exhibited to show greater endurance both before and after the float-REST treatment. Although women were found to score higher in depression than men before the float-REST treatment, this difference was not found post-treatment. The results also demonstrated, for the first time, that both sexes improved their ability to endure experimentally-induced pain following successful completion of the floatation-REST treatment.

5.4.3 Anxiety

In 2016, Jonsson et al. evaluated the effects of float-REST on self-diagnosed GAD [244]. The study implemented a 12-session floast-REST treatment program designed to target problems related to GAD such as pathological worry, emotional regulation difficulties, low levels of mindfulness, depression, fatigue, sleep difficulties, and muscle tension.

Jonsonn et al. (2017) evaluated the effects of a 7-week float-REST program on subjects diagnosed with GAD. A total of 9 subjects were recruited; each subject completed 12 45-minute flotation sessions over the seven-week period. Interviews were conducted with each subject after each float session along with the administration of a host of measures. The data was analyzed and interpreted by applying the Empirical Phenomenological Psychological (EPP) method developed by Karlsson. The results highlight that flotation-REST treatment of GAD was experienced as a comprehensive process that were both challenging and pleasant. The results indicated that the float-REST program reduced overall GAD symptomology. The present study also generated some initial understanding regarding potential mechanism that might mediate and maintain positive treatment effects when flotation-REST is applied as an intervention of GAD. Few things measured in the study were obstacles in treatment, a relaxed and safe vantage point, non-ordinary states of consciousness, connecting with oneself, new attitudes and coping strategies, and enhanced life quality.

In 2018, Khalsa et al. investigated the safety and tolerability of float-REST on 21 subjects diagnosed with anorexia nervosa [293]. The study found that there was no evidence of systolic or diastolic orthostatic hypotension after each float in any participant. Additionally, none of the subjects reported any adverse events throughout the duration of the float-REST program.

Significant improvements in anxiety and negative affect along with significant reductions in body dissatisfaction ratings were observed.

Feinstein et al. (2018) examined whether floatation-REST might attenuate symptoms of anxiety, stress, and depression in a sample of 50 subjects diagnosed with either depression, anxiety, or a stress-related disorder (e.g., PTSD, GAD, panic disorder) [294]. This open-label study found that a single one-hour session of floatation-REST was capable of inducing a strong reduction in state anxiety along with a substantial improvement in mood. Participants also reported significant reductions in stress, muscle tension, pain, depression and negative affect, accompanied by an increase in feelings of serenity, relaxation, happiness and overall well-being. Further analysis revealed that the most severely anxious participants consequently reported the largest effects. These findings suggest that floatation-REST may be a promising technique for acutely reducing symptoms of anxiety and depression, although the persistence of these effects is presently unknown.

This group in 2018 investigated the affective and physiological changes induced by floatation-REST and assessed whether individuals with high AS experienced any alterations in interoceptive sensation while immersed in an environment lacking exteroceptive sensation [295]. 31 subjects, all diagnosed with either anxiety or depression, were randomly assigned to complete either a 90-minute session of floatation-REST or to complete a 90-minute session of an exteroceptive comparator that entailed watching a nature documentary from the BBC Planet Earth series (i.e., control group). There were two main findings in this study: 1) Subjects experienced a robust relaxation response during and after floatation-REST that was decisively anxiolytic in nature and 2) the float environment enhanced interoceptive awareness and attention to cardiorespiratory sensations. For both findings, the effects during floatation-REST were significantly greater than those during the exteroceptive comparison condition.

Edebol et al. (2013) examined the individual experience of a long-time user of float-REST diagnosed with Whiplash Associated Disorder (WAD), grade IV [296]. The subject of interest was a middle-aged, native-born Caucasian male from Sweden who had been diagnosed with chronic WAD-IV by a licensed physician. The subject performed regular floatation for one and a half years and wrote about his experiences in a diary; additionally, a semi-structured interview was conducted at the end of his float-REST treatment. Edebol et al. provide a model describing the rehabilitative circuit brought about by float-REST that is in line with the potential role of a stress response system for the development and management of chronic whiplash. The study provides qualitative insights into the use of float-REST as a pain- and stress-management system for chronic whiplash. Thus, the study findings suggest that float-REST can be used to relieve chronic pain and enhance the quality of life for a more comprehensive group of patients with whiplash-associated disorders.

5.4.4 Physiological

In 1994, Raab and Gruzelier conducted a study in which they examined subjects before and after floatation-REST with neuropsychological tests chosen because they disclosed changes before and after hypnosis [297]. They predicted that like hypnosis, floatation would produce an improvement in right, relative to left, hemispheric processing. 32 subjects with no knowledge on hemispheric specialization and no experience in floating were randomly divided into two conditions: The experimental/float group and the control/non-float group. 2 different measures were used during the study for pre-test and post-test measurements: The Haptic Processing test and the Warrington

Recognition Memory test. The Haptic Processing test consisted of sorting letters and numbers while blindfolded after a brief familiarization period with the items. In order to obtain hemispheric processing times, scoring on this test was calculated by subtracting the mean movement times from the mean sorting times for each hand. The Warrington Recognition Memory test involved presenting the subject with 50 stimulus items in which subjects decided whether they liked or disliked the word or face. Retention was tested by two choice recognition after the presentation of the test stimuli and the subject was instructed to distinguish the stimulus item from the distracter item [297]. Additionally, a floatation questionnaire was administered post-test which covered responses on possible occurrences during floatation like the loss of conception of time, sudden insights, and unusual memories. After the administration of the Haptic Sorting Test and the Warrington Recognition memory test, participants floated supine in a floatation tank for 90 minutes. After the float, participants were given both tests again and the Floatation Questionnaire. Their results supported the hypothesized right hemispheric processing enhancement after floatation REST across both neuropsychological tasks. They also found that floatation did not produce significant detrimental effects on the left hemisphere like hypnosis, as there was only a slight reduction in right-hand sorting times with repetition in comparison with the control group [297].

A study conducted in Japan tested the hypothesis that floatation-REST would facilitate deep relaxation and a hypnagogic state, which may enhance the generation of random sequences [298]. 7 participants (4 men and 3 women) were recruited, and all subjects completed two types of treatment sessions in a counterbalanced order: 1) subjects rested in a supine position within a floatation tank (float-REST), and 2) subjects rested in a supine position on a bed within a dark, quiet room (bed-REST). Subjects participated in a total of two float-REST sessions and one bed-REST session (each session lasting 40 minutes), with a one-week interval between float-REST session and a month-long interval between the two intervention types. Prior to each treatment session, electrodes were attached to the subject in order to measure EEG, EOG, heart rate, and respiration. Before each session, subjects were instructed to complete a task in which they were asked to orally generate a list of numbers while attempting to be as random as possible. Subjects completed this same task at the 40-minute mark of their treatment session as well as at the end of the session. Results indicated that random number generation was enhanced by floatation-REST in comparison to bed-REST. Randomization indices scores were lower during both the 40-minute mark and at the end of floatation-REST, indicating an increase in randomness. Additionally, another interesting result showed that although no significant difference was observed between the spectral power from the band activities of theta to beta in both conditions, the delta band power in the bed-REST condition was 1.7 times larger than that in the floatation-REST condition in the latter half of the REST period [298].

In 1995, Suedfeld and Eich conducted a two-part study in which they examined the effect of floatation-REST on self-rated mood and arousal along with its effects on mood change and autobiographical memory [299]. For both parts of the study, the Profile of Mood States (POMS) Questionnaire was administered once before each float-REST session and twice afterward. The first study recruited 32 participants and randomly assigned them to one of two treatment groups: the float-REST group, in which subjects rested in a supine position within a float tank for 1 hour; or the control group, in which subjects simply waited in the university's psychology building and returned to the laboratory after 1 hour. POMS result for this study showed post-float REST subjects to be significantly lower than controls on both scales [299].

The second study recruited 24 participants who were sorted into equally sized float-REST and control groups. Subjects were first asked to complete two measures assessing their current levels of mood and arousal. Then, during their assigned treatment, subjects were asked to complete the same measures, either over an intercom system for subjects in the float-REST group or on pen-and-paper for those in the control group. Subsequently, subjects completed a new task in which they had 120 seconds to retrieve a specific autobiographical memory in response to a list of neutral probe words. Subjects were then asked to date their rerieved event and rate it based on seven scales adapted from prior research on moon congruence in autobiographical memory (Eich et al., 1990; Eich et al., 1994).

Suedfeld and Eich found that systematic self-ratings in both studies confirmed common reports of serenity and pleasant relaxation during float-REST. Findings from the second study support the hypothesis that neither an explicit mood induction nor the recognition that such an induction was being attempted is necessary for the occurrence of mood congruence in autobiographical memory [299]. Additionally, the results showed that the more relaxed a float-REST subject was during the autobiographical memory retrieval task, the more vivid and emotionally intense the memories experienced by the subject were.

In 2007, Asenlof et al. examined how float-REST might be used in tandem with standard therapy to treat patients with severe stress problems [300]. Two women on long-term sick-leave, aged 55 and 58, participated in the study, which was carried out over a period of one year. One subject was diagnosed with burnout depression and the other with fibromyalgia. Asenlof et al. implemented a treatment program with several components: floatation-REST, group therapy, conversational therapy, and picture production. Throughout the study, the subjects were asked to track their progress in journals and participated in long-form interviews on two separate occasions. The Empirical Phenomenological Psychological Method (Karlsson, 1995) was used in the study to generate four overarching themes: 1) the therapeutic work model, 2) transformation of feelings, 3) self-insight, and 4) meaning. The method entails an analysis in several stages including techniques for dividing the texts into smaller so-called "meaning units" (MU). This division is not based on grammatical rules but entirely on the content the researcher discovers and where there is a suitable shift of meaning. These together constituted a "therapeutic circle" which after a while transformed in to a "therapeutic spiral" of increased meaning and enhanced wellbeing.

Kjellgren et al. (2010) examined the efficacy of float-REST, coupled with psychotherapy, for the treatment of persons suffering from diverse ailments as chronic fatigue, depression, pain and/or anxiety [247]. Four women and two men, all on disability leave, between the ages of 33 and 57 years old took part in the study. Kjellgren et al. designed a 10-week treatment program consisting of twice-weekly 45-minute floatation-REST sessions and weekly psychotherapy with a psychologist. Participants were interviewed twice during the ten weeks about their experiences of the treatments and its effects in their daily lives. The first interview (about 30 minutes) was conducted after four weeks, and the last interview (about 60 minutes) was conducted after ten weeks once the subject had completed the entire treatment program. The findings of this study suggested that float-REST induced deep relaxation and altered states of consciousness, with experiences like feelings of flying, entering a state of "nothingness", and feelings of distinguishing the mind from bodily limitations commonly being felt. Additionally, some subjects reported that the feeling that the mind and body were separate entities gave rise to insights concerning their ailments. Subjects also reported a heightened awareness of physical sensations, breathing patterns, bodily responses, body image, and body processes in general. By the end of treatment program, all the participants were reported as being so full of energy and strength that neither they themselves nor their physicians assessed that any disability leave was needed.

This group in 2011 designed a 10-week treatment program combining float-REST with psychotherapy to examine its efficacy as a treatment for persons suffering from high stress-load and burnout syndrome [248]. Five subjects were recruited, all being diagnosed as on the brink of taking a sick leave and suffering from burnout syndrome, experiencing symptoms of fatigue and problems keeping up with daily life. Interestingly, all five subjects experienced a complete recovery in their mental health within the relatively short span of 10 weeks. While promising, the results should be further evaluated in a randomized control trial.

In 2016, Driller et al. examined the physical and psychological effects of float-REST in 60 elite, international-level athletes (28 males, 32 females) across a range of 9 sports, with the goal of determining if float-REST was a viable strategy for athlete recovery regarding both mood state and muscle soreness. Following exercise training for their sport, each subject completed a 45 minute float-REST. Pre- and post-float, subjects filled out the Multidimensional Mood-state Questionnaire (MDMQ) as well as a questionanire on perceived muscle soreness. Subjects also reported whether they had napped during the float-REST session or whether they had remained awake. A single float-REST session was found to significantly enhance 15 of the 16 MDMQ mood-state variables and also lowe perceived muscle soreness. Small to moderate effect sizes in favor of napping for 9 of the 16 MDMQ mood-state variables were also found when compared to the no-napping group, suggesting that napping in combination with float-REST may provide additional benefits to enhance certain mood-state variables. In summary, subjects saw an improvement in both mood state and muscle soreness, indicating that float-REST may be an effective tool for both physical and psychological recovery following training in elite athletes.

5.4.5 Consciousness

In 1960, Jay T. Shurley designed a study focused on refining the procedural methods used in early floatation-REST [302]. Shurley attempted to refine three fundamental aspects of floatation-REST: the physical aspect, physiological aspect, and psycho-social aspect. On the physical level, Shurley noted that it was important to create an environment that obtained the maximum achievable reduction of ambient physical stimuli along with a dynamic maintenance of ambient temperature. To achieve these conditions, Shurley created a special two-room laboratory that provided a significant reduction in light, sound, vibration, odor, and taste. Next, on the physiological level, there was a priority to eliminate all sources of pain and discomfort stemming from body position, pressure ischemia, and hollow viscus distention. Thus, a tank was designed that allowed subjects to float upright in order to reduce body discomfort. Additionally, a stimulus-restricting oxygen mask was used by subjects when fully immersed in the water to maintain neutral buoyancy via the use of weights. Lastly, on the psycho-social level, Shurley noted that certain types of persons were optimal subjects for float-REST studies. Such subjects were skilled in self-observation, memory, and attention to detail as well as able to communicate their experience fully and freely with minimal distortion. Based on this criteria, Shurley recruited several subjects, the first of which floated for over 4 hours and displayed emotions that shifted randomly from calm, contemplation, anxiety, elation, and depression. The subject also reported experiencing mild to severe audible and visual hallucinations throughout the session. Other subjects were observed to display similar non-consistent mood swings and hallucinations, with additional reportings of hyperawareness of body function, such as being able to hear the sounds of their own hearts. Overall, this study presents a feasible and effective method for studying a wide range of psychophysiological phenomena under the optimal conditions for float-REST [302].

In 1961, Lilly and Shurley designed a study to observe the ego-altering effects of float-REST as well as establish methods of self-observation by finding, defining, and setting limits for subject's psychological sets [303]. The floatation environment used in this study was similar to that of Shurley (1960) such that participants floated neutrally buoyant (via the use of weights) while upright and submerged in a tank inside of a stimulus-isolating room. Participants also wore a stimulus-restricting mask that allowed for normal breathing while underwater while still limiting the subject's senses. During this study, participants rotated among three different roles. Participants would first act as a regular subject floating within the tank and were instructed that they were free to explore whatever internal processes they wished to while floating in isolation for as long as they decided, with the goal of attenuating to their egos. Then, after exiting the tank, participants acted as "safety man" in which they would sit outside of the isolation area and observe another subject floating, operate the floatation equipment, and be on standby should any issues arise. Lastly, the participant became the "self-observer", in which they were allowed to float without a safety man in the room, allowing for maximum isolation and ego freedom. Subjects became self-observers only after great consideration and agreement between the subject and the safety man. Notable results from this experiment were the idea that freedom from external exchanges and transactions allowed the isolation-constrained ego to develop sources new information from within, such that sources can be experienced as if they are outside with greater or lesser degrees of awareness. Drives such as needing to leave the tank or needing to interact with other persons increased during floatation but decreased after multiple exposures to floatation. Subjects were also very susceptible to both negative and positive suggestions during the session, such that assessments of performance would induce heightened negative ideations and experiences. Positive experiences were found only in times where inner and outer relational stimuli was minimized.

Norlander et al. (2000), ran a study to investigate if experiences induced by floatation-REST might be affected by either settings and/or subjects' earlier experiences of altered states of consciousness (ASC) [304]. The comparison of subjects' dreams with their float-REST experiences indicated there was no significant difference between Group or Setting with regard to precision, participation, familiarity or reality. Float-REST was shown to significantly reduce pain and enhance mood. Overall, subjects reported that floatation-REST was a pleasurable experience. Subjects also reported that they experienced different kinds of visual and acoustic effects which altered their time perception, along with a sense of weightlessness. Additionally, reports of deep transpersonal experiences were quite common and could be distinguished into three types: experiences of one's own childbirth/delivery, feeling of cosmic unity, and experiences of losing contact with the body or out-of-body experiences. In summary, float-REST appears to be an effective consciousness-altering method with promising potential for clinical and therapeutic use.

In 2004, Kjellgren et al. investigated whether or not the degree or level of altered state of consciousness could be of importance for the subjective experience of experimental pain induced when the participant was already in a mild altered state of consciousness [250]. In order to practically achieve this altered state of consciousness; sensory deprivation was used in a floatation-REST tank and on a couch in a dark, silent room (chamber-REST), respectively. 23 participants were exposed to one 45-minute exposure to floatation-REST and one exposure to chamber-REST on two occasions, incorporating random assignment to either floatation-REST followed by chamber-REST or vice versa. On each occasion, the Restricted Environmental Stimulation Technique (REST) procedure was followed immediately by testing experimentally induced pain to one arm using a blood pressure cuff. The questionnaires and tools used to asses these differences were: VAS, HAD, Pulse Oximeter, Sphygmomanometer and EDN. It was found that floatation-REST induced a significantly higher degree of altered states of consciousness (ASC) than did chamber-REST. Participants experiencing High Experienced deviation from normal state in the floatation-REST condition reported higher levels of both "experienced pain" and "experienced stress" than did those experiencing Low EDN.

In 2008, Kjellgren et al. ran a qualitative analysis of interviews conducted with eight subjects, diagnosed with depresion, burntout syndrome, or chronic pain, to assess the subjective effects of float-REST during and following the float session. Kjellgren et al. used the Empirical Phenomenological Psychological method to analyze the interview transcripts, finding four common themes throughout the interviews: 1) experiences during floatation, 2) perceived effects afterwards, 3) technical details, and 4) the participant's background, motivation, and expectations. Overall, subjects reported float-REST to be a pleasant experience. Subjects reported experiencing altered states of consciousness, varying from milder states involving profound relaxation and altered time perception to more powerful states with notable perceptual changes and profound sensations such as out-of-body experiences and perinatal experiences. Further research may attempt to combine float-REST with a psychotherapeutic intervention by psychologist or psychotherapist in order to investigate if such a combination might generate even more beneficial effects.

5.4.6 Sleep

In 2017, Dunham et al. determined whether Bispectral Index values obtained during float-REST have a similar profile in a single observation compared to literature-derived results found during sleep and other relaxation-induction interventions [306]. One single subject experienced 22, 1-hour floatation-REST sessions and during sessions 14 and 16, BIS monitoring was performed where BIS values were recorded on the BIS-X hard drive every minute during floatation. The results indicated that Pre-floatation mood scores progressively increased from 5 at sessions 1–7 to 8 at sessions 16–22. Similarly, post-floatation mood scores increased during later sessions. The mean pre-floatation and post-floatation difference for the 22 float sessions was 3.5 ± 0.5. The objective BIS electrophysiological signature implies that relaxation-induction, stage I sleep, and floatation REST may be comparable conditions of consciousness.

We should mention that, in 2020, Kjellgren et al. published a review study titled, "Does floatation-rest (restricted environmental stimulation technique) have an effect on sleep?" [255].

5.4.7 Psychology

In 2013, Kjellgren et al. reported the subjective experiences of an individual who had completed two and a half years of float-REST treatment (totaling 75 sessions) in order to treat a number of neuropsychiatric and mental health disorders [307]. The subject was a 24-year-old woman diagnosed with attention deficit hyperactivity disorder, atypical autism, post-traumatic stress disorder, anxiety, and depression. Interviews regarding her experiences were analyzed. The subject reported experiencing an overall improved quality of life, wellbeing, and healthy behavior due to the float-REST treatment. The subject also reported no negative effects from treatment. These results suggest that float-REST may have beneficial therapeutic effects on mental health.

Kjellgren et al. (2014) evaluated the psychological effects of floatation-REST in healthy participants [251]. Based on previous studies, Kjellgren et al. hypothesized that float-REST would have beneficial effects on levels of pain, depression, anxiety, stress, energy, optimism and sleep quality. 65 subjects were recruited, all of which were a part of a cooperative-health project initiated by their individual companies, and were randomly assigned into the float-REST group or the control group. The results showed significant decreases in experienced stress, pain, anxiety, and depression, as well as significant increases in sleep quality and optimism for the floatation-REST group compared to the control group. In addition, it was found that the dimensions mindfulness and altered states of consciousness, at least to some extent, seemed to be overlapping constructs.

5.5 Discussion

Here we review the most important findings from our systematic review and discuss the significance and impact of various trends highlighted in the review.

Lilly's (1962) initial hypothesis regarding sensory deprivation was that the brain would turn off in the absence of sensory input. Anecdotally, Lilly was reported as having believed that the sensory deprivation tank was a portal one could use to go "inside yourself" (Lilly, 1977). While no valid research project has supported such claims, many of Lilly's early anecdotal claims catalyzed further research into sensory deprivation (Suedfeld, 1980). In fact, the meditative state brought about by sensory deprivation that Lilly first researched in the 1960s is still what attracts many to float tanks today (Feinsten, 2015). Another of Lilly's early hypothesis (1962), positing that the minimization of external stimuli in sensory deprivation tanks forces individuals to refocus on internal stimuli, was supported by the stimulus-hunger hypothesis proposed by Borrie (1990). After 2000, research on sensory deprivation began to shift towards examining its potential as a form of mental health therapy. There is a growing body of empirical research that has identified the benefits of both alternative medicine and therapy for mental health, when used both in isolation and in combination with traditional psychotherapy. In 2013, 40% of Americans in the United States used alternative treatment (Suedfeld, 2014). Different modalities of alternative therapy and medicine are included in medical insurance schemes reflecting the growth in its popularity. Further, therapies and practices such as yoga, tai chi and meditation although once considered "alternative" are now considered mainstream as evidence by the inclusion as reimbursable procedures by medical insurance companies. The demands of daily living increase and products based on stress management are at the forefront of mainstream culture (Feinsten, 2016). Products and services that aim to increase the human "relaxation response" (Benson, 1975) have been found to be protective factors against autoimmune disorders, cardiovascular diseases, neurodegenerative as well as behavioral disorders (Kerr, 2000). We found that between 1960-2010, the focus of float pod research was primarily on the general positive effects of float-REST and its effects on pain. After 2010, however, research attention switched to the physiological effects of float-REST along with its effects on sleep and anxiety. We should mention that Anette Kjellgren et al. had a good contribution the application of float pod. Most studies used a randomized design and implemented float-REST sessions that varied between 35 to 240 minutes in duration, with a majority of studies using 45-minute sessions.

A myraid of positive effects associated with float-REST have been documented. Many of these effects have been found from placebo-controlled studies, suggesting that these effects occur as a product of the floating experience. Increased creativity, faster exercise recovery, relaxation, mental coping, enhanced energy, muscular and motor improvement, and increased subjective well-being have all been reported to be positive effects resulting from float-REST. Furthermore, float-REST has also shown promise in the clinical setting, being able to aid with smoking reduction, pain management, and diminishing symptoms of depression and anxiety.

Research on the effects of floating and the commercialization of the float pod industry are growing simultaneously. Although the floatation industry typically advocates the use of floating for causes that have been verified through research, no systematic study has looked qualitatively at the different applications of float pod in recent years. Additionally, although a plethora of anecdotal accounts on float pod use exists on the Internet, no formal research study has examined the trend of its application for non-specialized populations. In general, there is limited information available in the form of quantitative studies of floatation therapies and even less in the form of qualitative studies. Considering the popularity and adoption of floatation as an intervention to treat a number of health disorders, it seems imperative to fill these gaps in the scientific literature. Understanding our perception of experience, treatment delivery, experiment design, samples, questionnaire and tools for each application could provide insight supporting established research findings. Currently, the subjective perceived experience inside the sensory deprivation tank has not been thoroughly researched. Understanding how one perceives the environment from inside the floatation tank may help us understand the mechanisms by which float tanks engender relaxation, which could have great clinical implications. Such a research project might aim to understand where the mind “goes” when in a maximum state of relaxation. Achieveing this understanding might allow us to recreate the float pod experience outside the tank, perhaps in the therapy room or in a meditation setting.

5.6 Limitations and strengths

Some limitations of the current review should be noted. First, the qualitative ratings of the 46 studies were remarkably low, so the conclusion that floatation-REST is effective for some application such as sleep is tentative and awaiting new studies with stronger methodology.

Second, although all studies showed float-REST to have a generally positive effect on measures such as pain, anxiety, and sleep, further research is needed to determine whether floatation-REST is effective in specific types of clinical problems (i.e., certain types of disorders).

Third, in some studies, all measurements were self-reported [307]. However, it is not standard practice to use objective measurements in float-REST studies.

Fourth, all studies failed to mention any explicitly negative effects resulting from sensory deprivation. Lastly, a number of studies used self-referred samples, which reduces the external validity of these studies [306].

In addition, although the search methodology we used to identify relevant studies is well-founded, it undeniably did not capture all of the existing literature on the topic. Since the number of publications available at the time of writing this manuscript was limited, we decided to include all the papers we could find (note that some of the newer trends are more visible in repositories such as arXiv and bioRxiv, as those manuscripts may be going through the publication process). They have been adopted by the psychological community to quickly disseminate results and encourage a fast research-iteration cycle. Our goal was to provide a transparent and objective analysis of the trends in float pod application.

Also, it is important to note that researchers submit predominantly positive results for publication due to a highly competitive publishing landscape. If researchers withhold negative results from publication — i.e., publication bias — this could result in major changes to our review.

Despite these limitations, this review has several strengths. First, six of the nine studies were RCTs, which increase their internal validity and the reliability of their conclusions. Second, our quality assessment employed established criteria. Third, our literature searches covered several large databases, combined with complementary searches by researchers who are knowledgeable about floatation (AK and KJ), making it unlikely that relevant studies have been overlooked.

5.7 Future research and potential implications

Float pods offer non-invasive and safe treatment. Other comparable methods, such as meditation, yoga, or qigong typically require regular practice and a dedicated focus before benefits become apparent. In contrast, the beneficial effects of float pods are apparent often after just a single float session. Additionally, float pods are shallow, so they pose little risk of drowning; many tank models have a water depth of just 10 to 12 inches. There are no typical floatation tub users. The time of day for which the float session was scheduled was not controlled. Some experienced floaters shared differences in effects based on the time of day, and it is quite possible that floating at different times of the day may have some unknown measurable effect.

People of all ages and walks of life enjoy the experience of floating, including children [308]. There is a possibility to use float pods in schools and incorporate floating sessions to as part of an academic curriculum. Though most of the research that exists is older, there is some evidence that sensory deprivation may improve focus and concentration, leading to clearer and more precise thinking. These effects have been linked to improved learning and enhanced performance in school and career groups. According to an article published in 2014 in the European Journal of Integrative Medicine, floating in a sensory deprivation tank has also been found to increase originality, imagination, and intuition, which can all lead to enhanced creativity. Spending time in floating can also help children fit in better at school. Floating sessions allow for changes in learning habits and social behavior. If a child is struggling to make friends or having a difficult time getting good grades, a few hours in a sensory deprivation tank each month can give them a chance to unlock their mind and learn how to overcome those obstacles [309]. Finally, further research is being conducted on the positive effects of floatation tubs on diabetic and autistic children. Current research suggests that the sensory deprivation experienced within floatation tubs stimulates positive changes in learning, social behavior, and cognitive function in autistic children. Research also shows that floatation tubs improve the body’s ability to use insulin, which reduces the incidence and severity of diabetes.

While some parents worry that their children may experience fear or claustrophobia within the floatation tub, there are no reports of such adverse effects. In fact, children often report that the darkness in the tub is calm and peaceful as opposed to the scary darkness of their bedrooms at night. Additionally, floating is physically safe for children and no swimming ability is required as anyone will effortlessly float due to the buoyancy effect of the epsom salt solution. Given that childhood is the most formative time for our bodies and minds, the benefits of floating may offer positive, long-lasting effects on children.

Other physical benefits brought about by float-REST include boosting the immune system, increasing endorphin production, preventing sports injuries, speeding up athletic recovery, improving the absorption of nutrients and even improving the formation of joint proteins, brain tissue and mucin proteins. Mental benefits of floating include accelerated learning, heightened mental clarity and alertness, increased creativity and problem solving, and drastic reductions in anxiety. Floating even facilitates freedom from phobias, addictions, and destructive habits. Lastly, floatation tubs can improve one's quality of sleep and thereby help with sleep disorders.

Float pods may be implemented into gyms in the future. The various beneficial effects of float-REST on athletic performance are well documented. For example, in a study of 24 college students, float-REST was found to speed up recovery after strenuous physical training by decreasing blood lactate. A 2016 study of 60 elite athletes also found float-REST to improve both psychological and physical recovery following intense training and competition.

Floatation therapy is a simple method of holistic healing that has proven to help millions of people from all different walks of life around the world. It may not be the end-all, cure-all for all ailments, but there's proof in the pudding that it can serve as a potent treatment for a wide variety of conditions . The most recent research suggests that floatation therapy can be an effective supplement to treatment programs for substance abuse and addiction. For instance, the National

Center for Biotechnology Information (NCBS) show that floatation therapy may be a viable resource for people struggling with addiction. One study explains that “restricted environmental stimulation therapy (REST) [can be] offered as a useful, flexible tool that can facilitate change in addictive variables at each level of complexity, from habitual acts through attitudes to self-concept and spirituality.”

Your own home sensory deprivation tank can cost between $10,000 and $30,000. The cost for a one-hour float session at a floatation center or float spa ranges from about $50 to $100, depending on the location. Having mobile float pods in the future would provide convenient access to a greater population of users as well as allow for deployment and relocation to multiple sites.

In sum, we hope this review will constitute a good entry point for those looking to use float pods in their work and will assist the field to produce high-quality, reproducible results.

Table 23. Details of all the papers that we found for our review paper.

# Ref	Title	R	Sample: Size, age, selection method	Treatment delivery	Questionnaire & tools	Category	Results
[244]	Promising effects of treatment with floatation-REST (restricted environmental stimulation technique) as an intervention for generalized anxiety disorder (GAD): a randomized controlled pilot trial.	2	59 participants in age range of 18-65 years and who suffer prolonged anxiety problems	Treatment group received a 7-week treatment program of float pod which consisted of 12 sessions,45 minutes each. Float pod temp of 35 celsius and epsom salt saturated water.	PSWQ GAD-Q-IV MADRS-S PSQI DERS MAAS EDN	anxiety	A significant Time x Group interaction effect for GAD-symptomatology [$F(2,88) = 2.93$, $p < .001$, $\eta 2p = .062$] was found. Further analyses showed that the GAD-symptomatology was significantly reduced for the treatment group ($t(23) = 4.47$, $p < .001$), but not for the waiting list

								control group ($t(21) = 0.98$, $p > .05$), when comparing baseline to post-treatment scoring. Regarding clinical significant change, 37 % in the treatment group reached full remission at post-treatment. Significant beneficial effects were also found for sleep difficulties, difficulties in emotional regulation, and depression, while the treatment had

							ambiguous or non-existent effects on pathological worry and mindfulness. All improved outcome variables at post-treatment, except for depression, were maintained at 6-months follow. No negative effects were found.
[294]	Examining the short-term anxiolytic effect of floatation-REST.	0	50 participants were recruited across a spectrum of anxiety and stress related disorders. Also 30 participants without any anxiety and stress were recruited for reference. recruiting subjects from the	60 minutes in float pod for subjects and 90 minutes for control group. Anxious group had the choice of having lights on or off whereas	ASI-3 OASIS PHQ-9 PSS SDS HM STAI-Y state PANAS-X KSS Wong Baker	anxiety	participants reported significant reductions in stress, muscle tension, pain, depression, and negative affect, accompanied by a significant

			T1000 database. Targeting participants with anxiety and stress disorders, many with comorbid unipolar depression. Participants with very high levels of anxiety sensitivity, anxiety sensitivity index (ASI-3)total score>= 30. also included cut off score of 8 or greater for their overall anxiety severity and impairment scale(OASIS). Also selected participants with no float-pod experience but some swimming experience before	control group were instructed to have it off.	Pain scale VAS		improvement in mood characterized by increases in serenity, relaxation, happiness and overall well-being ($p < .0001$ for all variables). In reference to a group of 30 non-anxious participants, the effects were found to be more robust in the anxious sample and approaching non-anxious levels during the post-float period. Further analysis revealed that the most severely anxious

							participants reported the largest effects. There were 2 side effects reported, dry mouth and itchiness
[250]	Altered Consciousness In Floatation-Rest And Chamber-Rest: Experience Of Experimental Pain And Subjective Stress	2	23 participants were recruited. Their mean age was 29.48 years (SD = 4.97, range = 21 to 41), and 13 individuals were students whereas 10 had professions. They were recruited through association with sports-active groups in the province of Värmland	one 45-minute exposure to floatation-REST and one exposure to chamber-REST on two occasions, incorporating random assignment to either floatation-REST followed by chamber-REST or vice versa. The	VAS HAD: Blood pressure EDN	conscio usness	It was found that floatation-REST induced a significantly higher degree of altered states of consciousness (ASC), as measured with an instrument assessing experienced deviation from normal state (EDN), than did chamber-REST. Participants experiencing

				study was carried out for two weeks, with a six-week interval separating each week. On each occasion, the Restricted Environmental Stimulation Technique (REST) procedure was followed immediately by testing experimentall y induced pain to one arm using a blood pressure cuff			High EDN in the floatation-REST condition reported higher levels of both “experienced pain” and “experienced stress” than did those experiencing Low EDN. These results suggest that the particular distinguishing features of floatation-REST and chamber-REST may cause selective deviations from normal levels of consciousness, under experimental

							conditions, that may underlie the subjective experience of pain and stress thresholds.
[290]	Effects of floatation Rest (Restricted Environmental stimulation technique) on stress Related muscle Pain: Are 33 floatation sessions more Effective THAN 12 sessions?	2	Participants were 37 patients, 29 women and 8 men, all diagnosed as having stress-related pain of a muscle tension type. Subjects recruited from the waiting list at the Human Performance Laboratory at Karlstad University. They had been diagnosed by a physician as having stress-related pain, of a muscle tension type. They reported having had pain for	All participants, irrespective of condition, were treated with floatation-REST during two periods consisting of two treatments per week for three weeks, separated by a week without treatment. The group with 12 treatments visited the	PAI Pain Matcher SE HAD LOT PANAS	pain	analyses for subjective pain typically indicated that 12 sessions were enough to get considerable improvements and no further improvements were noticed after 33 sessions. A similar pattern was observed concerning the stress-related psychological variables: experienced

			an average of 11.14 years (SD = 8.41) and 23 of the patients took analgesics on a regular basis. The average age of the patients was 49.54 years (SD = 8.67). Among the patients, 14 had also received the diagnosis of burn-out depression including symptoms such as fatigue, less energy, loss of self-esteem, problems with organizing daily life, problems with memory and processing new information, problems with sleep, finding that the ailments are not relieved by rest, and	laboratory twice a week during 6 weeks (over a total of 7 weeks). The other group with 33 treatments visited the laboratory after the initial seven weeks for seven more three-week periods of treatment, with only one session a week (over a total of 35 weeks).			stress, anxiety, depression, negative affectivity, dispositional optimism, and sleep quality. For blood pressure no effects were observed after 12 sessions, but there was a significant lower level for diastolic blood pressure after 33 sessions.

			feelings of low-spiritedness.				
[292]	Treating stress-related pain with the floatation restricted environmental stimulation technique: Are there differences between women and men?	0	88 patients (69 women, 19 men).The mean (± SD) age of the patients was 49.28±9.24 years. all recruited from the waiting list at the Human Performance Laboratory at Karlstad University. subjects were selected because they had been diagnosed by a physician as having chronic stress-	Participants were randomly assigned to one of three different experimental studies that included floatation-REST groups with assessments before and after treatment	DTs PE PAI Pain Matcher: SE HAD	pain	The analyses indicated that the floatationREST treatment had beneficial effects on stress, anxiety, depression, sleep quality and pain and that there were few sex differences. Women were more depressed than men before

			related muscle tension pain	during 12 or more sessions – study 1 (13), study 2 (20) and study 3 (31). The participants were treated with floatation-REST during two three-week periods consisting of two treatments (45 min each) per week for three weeks, followed by a week without treatment.			treatment, but after treatment there was no difference between sexes. However, there was a sex difference in the ability to endure experimentally induced pain, suggesting that men exhibited greater endurance both before and after the floatation-REST treatment. The results also showed, for the first time, that both sexes improved their ability to endure experimentally induced pain

							(higher scores for upper pain threshold) following the successful floatation-REST pain treatment.
[289]	Effects of floatation-restricted environmental stimulation technique on stress-related muscle pain: What makes the difference in therapy attention-placebo or the relaxation response?	2	Thirty-two patients (25 women and seven men). The average age of the patients was 48.46±9.51 years. Subjects recruited from the waiting list at the Human Performance Laboratory at Karlstad University, Karlstad, Sweden. They had been diagnosed by a physician as having stress-related pain of a muscle tension type	All participants, regardless of condition, were treated with floatation-REST for a seven-week period. The period consisted of two treatments per week for three weeks, followed by a week without treatment, then another	PAI SE HAD:LOT PANAS	pain	The participants exhibited lowered blood pressure, reduced pain, anxiety, depression, stress and negative affectivity, as well as increased optimism, energy and positive affectivity. The results were largely unaffected by the degree of

				three weeks of treatments. Thus, the participants received a total of 12 floatations during two periods of three weeks each.			attention-placebo or diagnosis.
[307]	Quality of Life with Floatation Therapy for a Person Diagnosed with Attention Deficit Disorder, Atypical Autism, PTSD, Anxiety and Depression	1(single subject)	a 24-year-old female subject with psychiatric and neuropsychiatric disorders. This subject was chosen because she was thoroughly assessed and diagnosed with Attention Deficit Hyperactivity Disorder predominantly inattentive subtype (DSMIV; 314.00)	At week one, floatation was performed for 3 × 45 minutes, at week two and three for 2 × 45 minutes, at week four to six for 1 × 45 minutes, and week seven and forward included one or two	Subjective floatation experience	psychology	From this qualitative single-subject study we learn that floating was associated with beneficial therapeutic effects in terms of quality of life, subjective wellbeing, and healthy behavior. The respondent: “feel good well

			and with atypical autism (DSM-IV; 299:80) by a neuropsychiatric specialist-team. When she initiated floating, she suffered from PTSD (due to the earlier episode of assault), high stress load, fatigue, social phobia, anxiety, recurring episodes of depression, muscle tension pain and general stiffness.	sessions per month. At the time of the first interview she had performed floatation for one and half year including approximately 50 sessions in total. At the one-year follow-up, she had floated approximately 75 sessions. The first interview was conducted at the respondents floatation center, it prolonged for 74 minutes			like a new person and so it has made a great difference… it really has, and I really want to continue with this because I really need it.”

				and was recorded on a mini-disc. The interview was semi-structured with questions like: how come you started floating, how do you experience floating, has your life somehow been affected by floating, has your experience of floating changed over time? A one-year follow-up was performed to understand			

				more about the experiences from long-lasting floatation and handwritten notes were taken			
[306]	Comparison of Bispectral Index [TM] values during the floatation restricted environmental stimulation technique and results for stage I sleep: a prospective pilot investigation	1(single subject)	1 subject (the author)	22 1-hour floatation-REST sessions. During sessions 14 and 16, BIS monitoring was performed where BIS values were recorded on the BIS-X hard drive every minute during floatation.	BIS Mood score	sleep	Pre-floatation mood scores progressively increased from 5 at sessions 1–7 to 8 at sessions 16–22. Similarly, post-floatation mood scores increased during later sessions. The mean pre-floatation and post-floatation difference for the 22 float sessions was 3.5 ± 0.5.

[288]	Effects of floatation-REST on muscle tension pain	2	Thirty-seven patients (14 men and 23 women). The mean age of the participants was 31.63 years. A portion of the participants were recruited through a 'remission' procedure (a procedure that allows patients access to specialized treatment on recommendation from their physician) from each one's general practitioner. A portion of the participants responded to announcements by the Karlstad University, Sweden, for individuals suffering from	nine treatments (three times per week for three weeks). Each floatation treatment lasted 45 min, resulting in a total of 300 h of treatment.	HAD Subjective floatation experience APZ OAVAV Analysis of blood samples	pain	The results indicated that the most severe perceived pain intensity was significantly reduced, whereas low perceived pain intensity was not influenced by the floating technique. Further, the results indicated that circulating levels of the noradrenaline metabolite 3-methoxy-4hydroxyphenyl ethyleneglycol were reduced significantly in the experimental group but not in

			localized muscle tension pain in the neck and shoulder area, with or without temporal headache				the control group following treatment, whereas endorphin levels were not affected by floatation. Floatation-REST treatment also elevated the participants' optimism and reduced the degree of anxiety or depression; at nighttime, patients who underwent floatation fell asleep more easily. The present findings describe possible changes, for the better, in patients presenting with

							chronic pain complaints.
[247]	Psychotherapeutic Treatment in Combination with Relaxation in a Floatation Tank: Effects on "Burn-Out Syndrome"	2	Four women and two men between the ages of 33 and 57 years old took part in the study. The average age of the clients was 42.7 years. They were all diagnosed as suffering from burnout syndrome with symptoms of fatigue, listlessness, and problems organizing daily life.	Weekly for 45 minutes over a ten-week period consisting of floatation-REST treatments and psychotherapy with a psychologist	EPP-method MU	physiological	Deep relaxation and altered states of consciousness were induced, with experiences like feelings of flying, entering a state of "nothingness" and feelings of distinguishing the mind from bodily limitations. Experiencing how the mind

							and body are separate entities gave rise to insights concerning their close connectedness. A heightened awareness of physical sensations of breathing patterns and bodily responses were noticed, as was an augmented awareness of body image and body processes in general. Also a deep physical relaxation, as well as mental relaxation with fewer thought

								processes were achieved and were greatly appreciated. By the end of the course of treatment ten weeks later, all the participants were so full of energy and strength that neither they themselves nor their physicians' assessed that any disability leave was needed. They all returned to work full-time.

[300]	Case Studies on Fibromyalgia and Burn-Out Depression Using Psychotherapy in Combination with Floatation-Rest: Personality Development and Increased Well-Being	1(single subject)	Two women on long-term sick-leave, aged 55 and 58. They had indicated their interest in participating in the experiment with floatation-REST treatment at the Human Performance Laboratory at Karlstad University but, after a medical examination, had been excluded from participating in a planned experiment because their anxiety levels were adjudged to be too high; they were subsequently referred to this study (the two women).	(a) floatation-REST treatment at least once every other week for 45 minutes on a total of 35 occasions, (b) group therapy on eight occasions, (c) conversational therapy on eight occasions and (d) picture Floatation-REST and therapy on eight occasions. The group, conversational and pictures therapy	MU	physiological	In only a total of 28 hours of group, conversational and picture therapy spread over a year (which represents approximately 15 minutes' therapy per client per week) the two therapists achieved a dramatic change in the lives of the two women, thanks to the combination with floatation-REST. A follow-up 18 months after the

				sessions were coordinated in a total of eight meetings.			completion of the therapy revealed that the spiral of increased meaning and enhanced wellbeing was still in operation.
[283]	EFFECTS OF FLOATATION-VERSUS CHAMBER-RESTRICTED ENVIRONMENTAL STIMULATION TECHNIQUE (REST) ON CREATIVITY AND REALISM UNDER STRESS AND NON-STRESS CONDITIONS	2	first study:38 subjects, 21 were female subjects and 17 were male subjects. The mean age of the participants was 34.31 second study: Thirty-two subjects were recruited,13 men and 19 woman (Subjects were recruited from announcements placed on announcement boards	1 or 3,45 minute sessions depending on the assigned group	Syllogisms I-II HAD LOT FS Composition Test EDN Stress test	positive effects of FP	1. No differences concerning divergent or logical production were obtained whether floatation occurred once or on three occasions. 2. Floatation-REST induced more originality, yet less deductive thinking. 3. Chamber-

			throughout the university.)				REST induced more realistic and elaborated thinking. 4. Subjects that were stressed in the Floatation-REST condition were more lively compared with the non-stressed subjects in that condition. 5. Stressed subjects in the Chamber-REST condition showed more realism than their non-stressed counterparts. 6. Both Floatation-REST and Chamber-REST were

							equally effective in reducing stress, i.e., showed comparable efficacy as relaxation techniques. However, Floatation-REST altered consciousness to a greater extent than Chamber-REST. 7. Correlational analysis indicated that the more adaptable and receptive to change and the more optimistic one was, the more originality on essay writing one exhibited.

							Further, there was a relationship between more lively accompanied by less realistic.
[293]	A Clinical Trial Investigating the Safety and Tolerability of Floatation-Rest in Anorexia Nervosa	0	21 Anorexia Nervosa Patients	Four float sessions for each participant.	Blood pressure Subjective floatation experience BMI	anxiety	Twenty-one patients completed the study (average EDE-Q: 2.3+/-1.4, average BMI 22+/-2.7. Primary outcome: there was no evidence of systolic or diastolic orthostatic hypotension after each float in any participant, and no adverse events. We

							significantly observed improvements in anxiety ($p<0.001$, Cohen’s $d>1$), negative affect ($p<0.01$, Cohen’s $d>0.5$), heightened interoceptive aware- ness for cardiorespiratory ($p<0.01$, Cohen’s d 0.2-0.5) but not gastrointestinal sensations, and reduced body dissatisfaction ratings ($p<0.001$, Cohen’s $d>0.5$) following floating.

[284]	Eliciting the Relaxation Response with the Help of Floatation–REST (Restricted Environmental Stimulation Technique) in Patients With Stress-Related Ailments	2	54 women and 16 men. Subjects recruited from the waiting list at the Human Performance Laboratory at Karlstad University, participated in the study. They had been diagnosed by a physician as having stress-related pain of a muscle tension type.	12 session/45 min. Participants were randomly assigned in equal numbers (35 participants) to one of two experimental groups: a control group and a floatation–REST group (see the sections “Design” and “Procedure”).	Questionnaire 1Before the treatment (floating in the tank), a questionnaire was provided that estimated each subject’s self-assessed pain: intensity, frequency, duration, onset, sleep quality, treatment as well as experiences/symptoms of other types of complaints. Each subject’s own descriptions of “sleep quality” were estimated	positive effects of FP	awareness of body image and body processes in general. Also, a deep physical relaxation, as well as mental relaxation with fewer thought processes were achieved and were greatly appreciated.

					on visual analog scales (0 –100). Questionnaire 2At a final meeting directly after the 7 weeks of the experimental floatation procedure, the same questions were presented as in questionnaire 1.		
[282]	Floatation REST as a Smoking Intervention	2	33 females between the ages of 20 and 64. All have made efforts to stop smoking. The subjects on average smoke 30 cigarettes per day for ~20 years. Subjects volunteered for the experiment.	The subjects were each placed in the tank. The participants were divided into groups. Half of these groupings	health and personality questionnaires	positive effects of FP	The percent reduction data at twelve months was subjected to analysis of variance. The between effects were messages, float/no float

			The experiment was advertised via radio, magazines, etc.	receive brief taped messages on an intercom system early in floats two through five; the others receive no messages. The first float was up to 150 minutes. Thereafter half receive four more floats of up to 60 minutes duration on consecutive days; the other half receive four more floats, each a week apart, and for up to			(60-minute groups, 150-minute groups, and control groups), and the interaction of messages and float/no float conditions. Only the main effect for float/no float conditions was found to be significant (p = < .05 for 2,27 df). Fisher's multiple comparisons tests reveal that the short float groups (Groups 1 and 2) are reliably lower than the control groups (Groups 5 and 6) at the p = < .05 (1, 27

				150 minutes duration. Five floats, each a week apart. Then, follow-up phone calls for up to 12 months after the last float.			df), and that Group 1 and Group 2 are both reliably lower than Group 6 (both at the p = < .05, for 1, 27 df). Groups 3 and 4 do not differ reliably from Groups 1 and 2 nor from Groups 5 and 6. Thus, the control groups reduce smoking reliably more than the short float groups and the control group without messages reduces smoking reliably more than each of the

							short float groups.
[251]	Beneficial effects of treatment with sensory isolation in floatation-tank as a preventive health-care intervention – a randomized controlled pilot trial	2	Sixty-five participants (14 men, 51 women) from three different companies with a mean age of 47.95 years. The participants were all part of a cooperative-health project initiated by their individual companies. There was a wide range of occupational groups varying from managers, employers and employees all in the retail industry.	7-week period with a total of 12 floatation-REST sessions (45 min each).	SE HADS LOT SQ MAAS VAS EDN	positive effects of FP	The main findings were significant decreased experienced stress, worst pain, anxiety, and depression - as well as significant increased sleep quality and optimism for the floatation-REST group compared to the control group. In addition, it was found that the dimensions mindfulness and altered states of

							consciousness, at least to some extent, seemed to be overlapping constructs.
[245]	Characterizing the experiences of flotation-REST (Restricted Environmental Stimulation Technique) treatment for generalized anxiety disorder (GAD): A phenomenological study	0	9 individuals (two men and seven women) with a mean age of 45 years (age range 24–61). They were selected from a sample comprising 24 individuals participating in a research project evaluating flotation-REST as an intervention for GAD	flotation treatment consisted of 12 flotation session (á 45 min) over a seven-week period with two sessions a week, and with the fourth week treatment free	health and personality questionnaires	anxiety	The result highlights that flotation-REST treatment of GAD was experienced as a comprehensive process that were both challenging and pleasant. The results indicate that the method positively affected symptoms and the core issue associated with GAD on an experiential level. The

							present study also generated some initial understanding regarding potential mechanism that might mediate and maintain positive treatment effects when flotation-REST is applied as an intervention of GAD
[248]	Preventing Sick-leave for Sufferers of High Stress-load and Burnout Syndrome: A Pilot Study Combining	0	Four women and two men between the ages of 33 and 57 years old. They were all diagnosed as on the brink for sick leave and suffering from ‘burn-out syndrome’ with symptoms of	twice weekly for 45 minutes during a 10-week period. For each person in this study, there will be a total of three hours in the	GHQ-12 HAD PAI VAS SQ	physiol ogical	The results revealed a significant decrease in degree of depression and anxiety and an increase in positive outlook on life. There

			fatigue and problems organizing daily life	floatation-laboratory and one hour psychotherapy each week; this multiplied with the scheduled ten weeks treatment time generates a total of 40 hours per included person.			was also a significant decrease in extent of painful areas and a significant decrease in their experienced worst pain-intensity. After the treatment period, they all continued to work, and there was no need for sick leave.
[291]	Chronic Whiplash-Associated Disorders and Their Treatment Using Floatation-REST (Restricted Environmental	0	7 subjects (6 women and 1 man). All being diagnosed as having chronic WAD by licensed physicians. Six participants either had WAD grade II (neck complaints and musculoskeletal signs) or WAD grade	Number of sessions varied from 2 to 15, 45 minute each treatment	Subjective floatation experience	pain	The results therefore contain two models; the first model covers the participants' experiences of the crises that took place in times prior to the

	Stimulation Technique)		III (neck complaints and neurological signs), and one participant had WAD grade IV (neck complaints and evidence of fracture or dislocation).				treatment, and the second model describes the short-term effects of the floatation-REST treatment in terms of floatation phases. A linear story about the experienced effects of the floatation-REST treatment in participants with chronic WAD appears as the background and adds depth, the foreground adds light, and the participants' quotations add illustrations to the story. short-

							term effects of the treatment in terms of five phases: (a) intensification, (b) vitalization, (c) transcreation, (d) deflocculation, and (e) reorientation. Results indicated that floatation REST is a meaningful alternative for treating chronic whiplash-associated disorder
[301]	Floatation restricted environmental stimulation therapy and napping on	0	60 athletes, 28 male and 32 female across 9 sports (athletics = 8; basketball = 8; boxing = 2; cycling = 10; football = 11;	Subjects were told to arrive within 1-3 hours after finishing their training.	MDMQ Muscle Soreness	physiological	A significant reduction in perceived muscle soreness in was seen in pre to post FLOAT. A

	mood state and muscle soreness in elite athletes: A novel recovery strategy?		netball = 15; rowing = 2; rugby = 2; swimming = 2). Found volunteers of elite Australina athletes that have not taken part in a floatation tank session. All athletes represent their country at an international level from summer and winter sports	Subjects would fill out a questionnaire pre and 10-minutes post FLOAT pod. Subjects were also told to come in a hydrated state. FLOAT was done for ~45 (48 ± 15 min) after training for 6 months.			moderate correlation (r = -0.35) was ween between the pre to post FLOAT muscle soreness. This indicated that a higher pre-FLOAT muscle soreness was associated with greater reductions in muscle soreness for post FLOAT. 15 of the 16 mood-states was significantly enhanced following FLOAT. Alert was the mood-state that seemed to not change significantly. The greatest

								change was "relaxed" as a mean change of 1.5 ± 1.3 was seen. 28/60 athletes reported to nap during FLOAT. A significance difference between nap and no-nap was seen in pre to post FLOAT for 5 of the 16 mood-state variables (“worn-out”, “at-ease”, “tense”, “fresh” and “exhausted”). Small and moderate effects were seen in 9 mood states. No significance was found between

							nap and no-nap for muscle soreness.
[295]	The Elicitation of Relaxation and Interoceptive Awareness Using Floatation Therapy in Individuals with High Anxiety Sensitivity	2	31 participants with high AS. Subjects all had high anxiety sensitivity	one 90-minute float session	ASI-3 OASIS PHQ-9 SDS	anxiety	Relative to the comparison condition, Floatation-REST generated a significant anxiolytic effect characterized by reductions in state anxiety and muscle tension and increases in feelings of relaxation and serenity (p= 0.001 for all variables). Significant blood pressure reductions were

							evident throughout the float session and reached the lowest point during the diastole phase (average reduction .12 mm Hg). The float environment also significantly enhanced awareness and attention for cardiorespiratory sensations.
[304]	THE EXPERIENCE OF FLOATATION-	2	28 total (The participants' mean age was 32.04 years (SD = 8.29))/24 men	1 session/60m minutes. Fourteen former drug-	health and personality questionnaires Subjective	conscio usness	The findings of this study indicate that floatation-REST

	REST AS A FUNCTION OF SETTING AND PREVIOUS EXPERIENCE OF ALTERED STATE OF CONSCIOUSN ESS		and 4 women. Participants were randomly assigned to either a “strict” setting (strict- condition) or to a “fantasy” setting (fantasy- condition)	users, they were matched against 14 participants without drug experience of comparable age, gender, and occupation.	floatation experience		offers a technique with notable potential for clinical and therapeutic application. Previous investigations have shown the technique to be used in association with the treatment of drug abuse problems. In the present study half of the subjects had a background of illegal substance abuse. There were no indications that floatation-REST was unsuitable for persons with

							a background of abuse. Besides the more ‘classical’ applications (pain alleviation, stress-reduction, mental training in sports), the technique ought to provide a useful adjunct within psychotherapy and personal development.
[305]	Sensory Isolation in Floatation Tanks: Altered States of Consciousness and Effects on Well-being	0	Eight persons, six females and two males, aged 35 to 69 years old (M = 49.5, SD = 12.4). Subjects were part of a floatation project at the stress clinic of the Human Performance Laboratory, where	N/A. patients with earlier experience of floatation tank therapy (at least eight times) were included	EPP-method NCT	positive effects of FP	Following the sorting and analyses of the material, 471 MU was created providing insight into the research question as to how floating affects the

			they had previously made an appointment because of mental or physical difficulties. They were selected from a group of patients who had floated at least eight times and visited the clinic within the previous month. The reason for this procedure was to include participants who were familiar with the technique, and whose experiences were recent. patients with earlier experience of floatation tank therapy (at least eight times) were included					individual and the circumstances surrounding the floatation. The MU: s then generated 21 categories which can be summarized into 4 themes: Experiences during floatation-REST, Effects of floatation-REST, Technical details and the target group for floatation. Floating was perceived as total relaxation, hardly comparable to anything else.

								The relaxation provides a sense of total calm and rest from everyday life. An altered state of consciousness (ASC) was induced during floating. It could vary from a mild ASC, e.g., like meditative daydreaming to a more powerful ASC with more profound, cognitive, perceptual, or transpersonal experiences. Floating was shown to produce pain relief and profound

							relaxation in the present study as well as in several others, effects that were much valued. A few participants noted that they might feel some discomfort leaving the tank
[285]	Floatation-restricted environmental stimulation therapy improves sleep and performance recovery in athletes	2	Nineteen trained, male team-sport athletes (age: 21 ± 2 years). They were chosen based on their athletic background. The team-sports from which the participants partook in were basketball (n = 4), football (n = 11), and rugby (n = 4).	two 45-minute float session. they either sat in a dim light room after the exercise or in a float pod for 45 minutes	VAS cortisol levels in saliva isometric mid-thigh pull dynamometer wrist actigraphy	positive effects of FP	FLOAT was found to significantly enhance CMJ (p = 0.05), 10 m sprint (p = 0.01) and 15 m sprint performance (p = 0.05) with small to moderate effects (d = 0.21–0.68) for all performance measures, except CMJ (unclear),

								compared to CON. The results also show significantly higher pressure-to-pain thresholds across all muscle sites (p's < 0.01) and lower MS and PF 12 h following FLOAT (p < 0.05). All sleep measures resulted in small to large effects (d = 0.20–0.87) with a significantly greater perceived sleep quality (p = 0.001) for the FLOAT trial compared to CON. There

							were no significant differences and a trivial effect size between trials for changes in cortisol concentration.
[251]	Beneficial effects of treatment with sensory isolation in floatation-tank as a preventive health-care intervention – a randomized controlled pilot trial	2	Sixty-five participants (14 men and 51 women) were all part of a cooperative-health project initiated by their individual companies, they were all health care workers.	12 awaiting floatation sessions (around two per week for a period of seven weeks) where each session was of 45 minutes duration and 30 minutes to shower and relax	SE HADS LOT SQ MAAS VAS EDN	positive effects of FP	Stress, depression, anxiety, and worst pain were significantly decreased whereas optimism and sleep quality significantly increased for the floatation-REST group. No significant results for the control group were seen. There was also a

							significant correlation between mindfulness in daily life and degree of altered states of consciousness during the relaxation in the floatation tank

6 Short-term Floatation-REST (Restricted Environmental Stimulation Technique) reduces stress: Analyses based on neural and cardiac components

6.1 Introduction

(Float pod) Sensory deprivation research, one of the first systematic laboratory explorations of environmental psychology, began in the late 1950s [310]. Its goal was to reduce environmental effects on a range of psycho-physiological, perceptual, cognitive, emotional, and other measures. Floatation-REST (Reduced Environmental Stimulation Therapy), an intervention that attenuates exteroceptive sensory input to the nervous system, has been self-reported to produce deep relaxation [244]. During Floatation-REST, a person is lying horizontally, face up, inside a quiet and dark tank, filled with heated salt-saturated water. Earlier research has documented deep relaxation and beneficial effects on fatigue [284], muscle tension [249], stress [254], sleep difficulties [311, 312], anxiety [251, 295] and depression [284, 290]. Floatation-REST is cost effective and secure, with minimal or a complete absence of adverse effects [313, 314].

EEG is the most important tool to study brain behavior because it enables researchers to select localized information from the composite inner mechanisms of the brain [315]. The recorded waveforms reflect the activity of the brain structures underneath the cortex. The frontal cortex plays an important role in both emotional and motivational processes [316]. More specifically, the left frontal region is involved in the management of arousal and regulation of the stress response [317, 318]. Stress assessment based on EEG spectral analysis is discussed in [319-321]. A busy brain decreases alpha power (8-13 Hz) and increases beta power (13-30 Hz). Focused mental processes are closely associated with a high-frequency EEG rhythm – the gamma rhythm (30 Hz and above) [322]. According to this model, asymmetry in frontal alpha activity reflects emotions and measures feelings. In fact, statistical differences of the frontal EEG alpha asymmetry have been observed under depression [323], examination stress [324] and sleep deprivation [325].

Electrocardiography (ECG) is another tool of stress assessment that shows heart rate variability (HRV), the fluctuation in time intervals between adjacent heartbeats [326]. HRV is an emergent property of interdependent regulatory systems which operate on different time scales to help us adapt to environmental and psychological challenges. HRV indexes neurocardiac function and is generated by heart-brain interactions and dynamic non-linear autonomic nervous system (ANS) processes. It reflects the regulation of autonomic balance, blood pressure (BP), gas exchange, gut, heart, and vascular tone, referring to the diameter of blood vessels that regulate BP, and possibly facial muscles [327, 328]. Two overlapping processes generate short-term HRV [329]: The first is the dynamic and complex relationship between the sympathetic and parasympathetic branches; the second includes the regulatory mechanisms that control HR via respiratory sinus arrhythmia (RSA), the baroreceptor reflex (negative-feedback control of BP), and rhythmic changes in vascular tone [330].

In HRV, frequency-domain measurements estimate the distribution of absolute or relative power into four frequency bands. The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) divided heart rate (HR) oscillations into ultra-low-frequency (ULF, (≤0.003 Hz)), very-low-frequency (VLF, 0.0033–0.04 Hz), low-frequency (LF, (0.04–0.15 Hz)), and high-frequency (HF, (0.15–0.40 Hz)) bands. The ratio of LF to HF power (LF/HF ratio) may estimate the ratio between sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity under controlled conditions [328].

Our laboratory has recently started to measure neural time series or electroencephalography (EEG) and electrocardiography (ECG) data during REST therapy [331]. The 6-channel EEG was recorded from the frontal lobe (FP1, FP2, AF3, AF4, AF7, AF8). As a result, after adjusting pre-float procedure and applying artifact removal algorithm, stable EEG with high signal-to-noise ratio (SNR) could be recorded for 45 minutes without the presence of ECG artifacts. In previous studies, authors mentioned that more active comparator, rather than self-reports, and blood pressure measures are better for assessing the efficacy of floatation-REST on anxiety [295]. Also, because the non-floatation state consisted of participants sitting upright in a chair, this posture likely magnified the differences between measures of interoceptive awareness and muscle tension.

In this study, we analyzed simultaneous central activity via EEG and autonomic heartbeat-to-heartbeat (RR intervals) from ECG during floating and non-floatation conditions. We focused on the existence of and variation in different EEG and ECG frequencies in 17 subjects during floating and lying-in bed. Furthermore, we tracked the temporal dynamics of power spectral density in two different conditions for different frequency bands of 6-channel EEG signals (alpha, beta, and gamma). We design the non-floatation situation, lying in the bed, to decrease dissimilarities between conditions.

Discovering a relation between autonomic nervous and brain dynamics will give us more insight into the effects of float pod on stress reduction. We hypothesized that floatation-REST -despite the pressure, body posture, temperature, humidity, and other conditions- can lower stress levels in a single session better than lying in the bed.

6.2 Methods and Materials

6.2.1 Participants Recruitment and Randomization

We used a within-subject crossover design, in which 17 participants who met specific inclusion and exclusion criteria were assigned to complete both a 21-minute session of Floatation-REST and a 21-minute session of an exteroceptive comparator (referred to as the bed condition), in random order. After completing one condition, participants crossed over to the other condition (average time between conditions was 30 minutes), with both conditions scheduled at the same time of day for each participant. We pre-determined the randomization sequence using a 1:1 allocation ratio, and the study used an open-label design with no blinding or concealed allocation.

6.2.2 Data acquisition and pre-processing

We recorded EEG and ECG during both conditions, i.e., floatation and lying-in bed, in 17 healthy subjects. Each recording lasts about 21 minutes, including 2 breaks, creating in effect 3 intervals of approximately 5.8 minutes each. A customized headband made from neoprene material secured the EEG electrodes using an elastic and adjustable cord. Tegaderm, medical-use tape, and a swim cap placed over the headband created a good seal, which helped to eliminate ECG artifacts. The EEG signals were recorded from frontal (AF3, AF4, AF7, and AF8) and frontopolar (FP1 and FP2) electrodes, with a customized 6-channel LiveAmp mobile EEG amplifier from Brain Products GmbH (Figure 37). The scalp EEG electrodes were referenced to AFz, and the ground was set at Fpz. Bipolar ECG signals were acquired using two passive electrodes placed below the collarbone and recorded from one of the auxiliary inputs on the LiveAmp sensor and trigger extension box. All data were digitized at 1000 Hz. After adjusting for pre-float procedures and removing artifacts, a stable EEG signal with high signal-to-noise ratio was recorded without the presence of ECG artifacts [331].

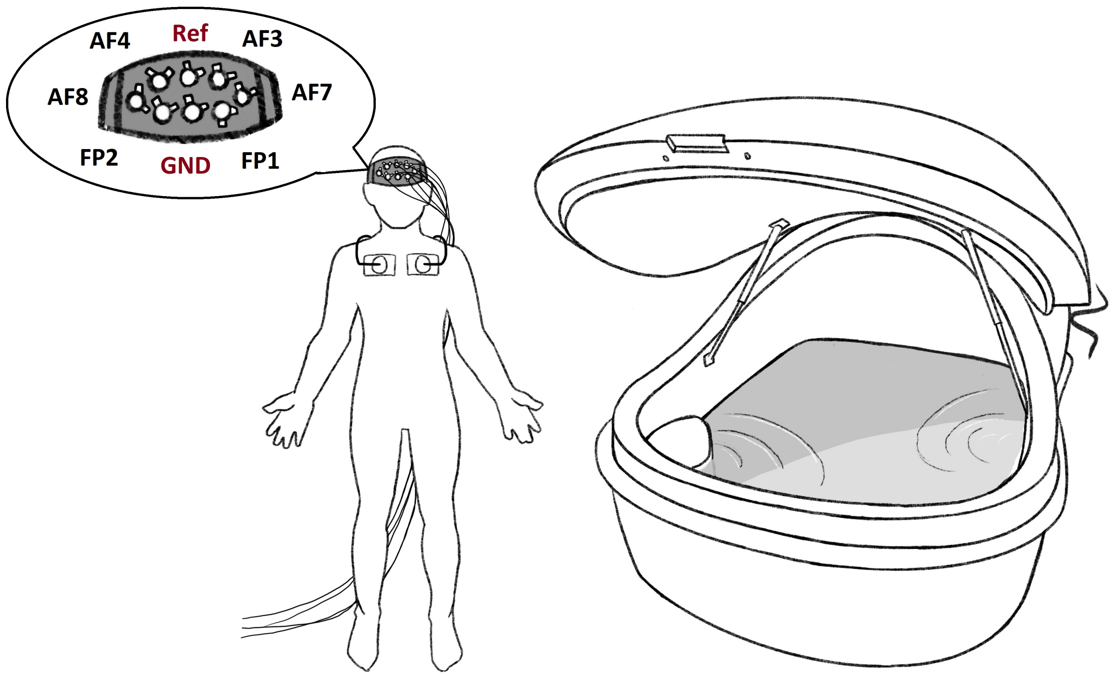

Figure 37. (Left) A customized 6-channel LiveAmp mobile EEG amplifier from Brain Products GmbH was used to collect EEG signal from the frontal lobe and bipolar ECG signal. (Right) Sensory deprivation tank.

6.2.3 Heart-beat detection and the RR time-series

We filtered the ECG signal with a 6th order Butterworth band-pass filter 0.5–100Hz, identified the R peaks in the ECG using the Pan-Tompkins method [332-334], and confirmed the results visually. We intentionally analyzed all artifact-free RR intervals without rejecting any RR intervals that were too short or too long [335]. To extract the continuous RR tachograms, the RR intervals were resampled (at 4Hz for power spectrum estimation) and interpolated by piecewise cubic spline.

Zero-phase Butterworth filters were applied to the interpolated RR time-series to extract the low frequency (LF; 0.04-0.15Hz) component of RR_{LF} and high frequency (HF; 0.15-0.4) component RR_{HF}[335]. The RR_{LF} reflects sympathetic nervous activity (although this is not universally accepted [336]) while the RR_{HF} components reflect parasympathetic (vagal) activity [337]. Figure 38 shows an example of the raw ECG, RR tachograms, RR_{LF} and RR_{HF}components for 40 seconds of a sample recording.

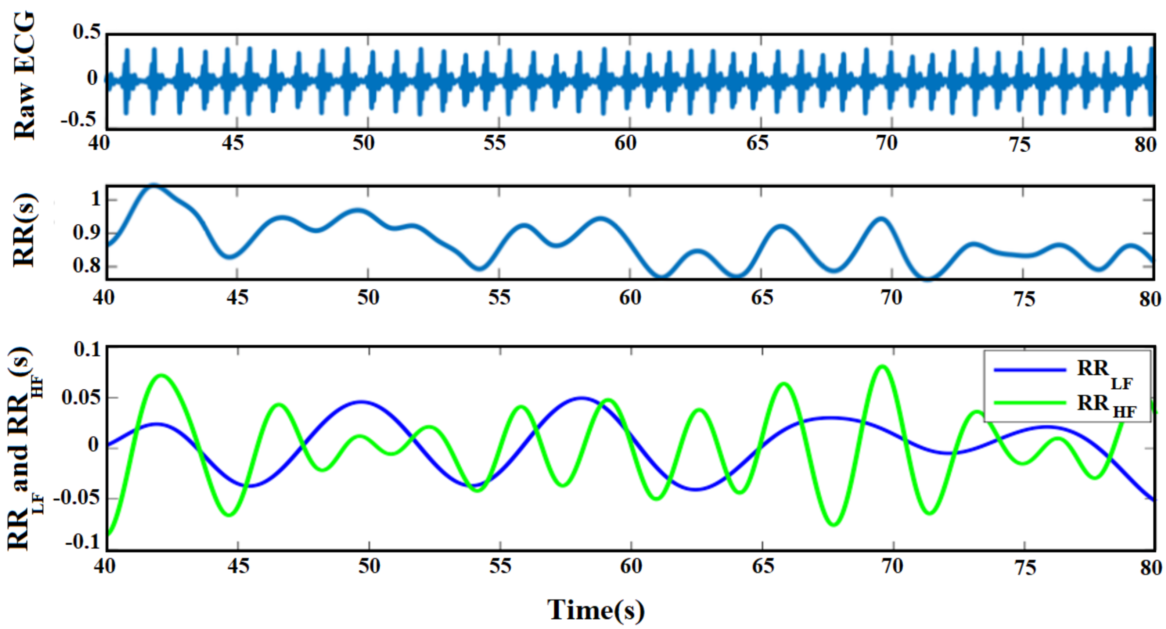

Figure 38. (Top) Raw ECG signal, (Middle) RR tachograms, (Bottom) RR_{LF} and RR_{HF} components for 40 seconds RR time series.

6.2.4 Extraction of the root mean of successive heart beats differences

The root mean square of successive differences between normal heartbeats (RMSSD) represents the beat-to-beat variance in the HR and is the primary time-domain measure used to estimate the vagally mediated changes reflected in HRV [338]. We obtained the RMSSD by first calculating each successive time differences between the heartbeats in ms. Then, we squared each of the values and averaged the result before obtaining the square root of the total. The RMSSD is positively correlated with the HF power and reportedly more influenced by the PNS activity [339]. Hence, higher RMSSD indicates greater parasympathetic influence. In this study, we found the RMSSD for three 5-minutes intervals during both conditions.

6.2.5 Power Spectrum of RR time-series

We calculated the RR power spectra for LF/HF ratio. Figure 39 represents the RR power spectrum for one subject (subject 4) during interval 3 for bed and pod conditions. For the purposes of statistical analyses, we calculated the LF/HF ratio as an index of the sympatho-vagal balance.

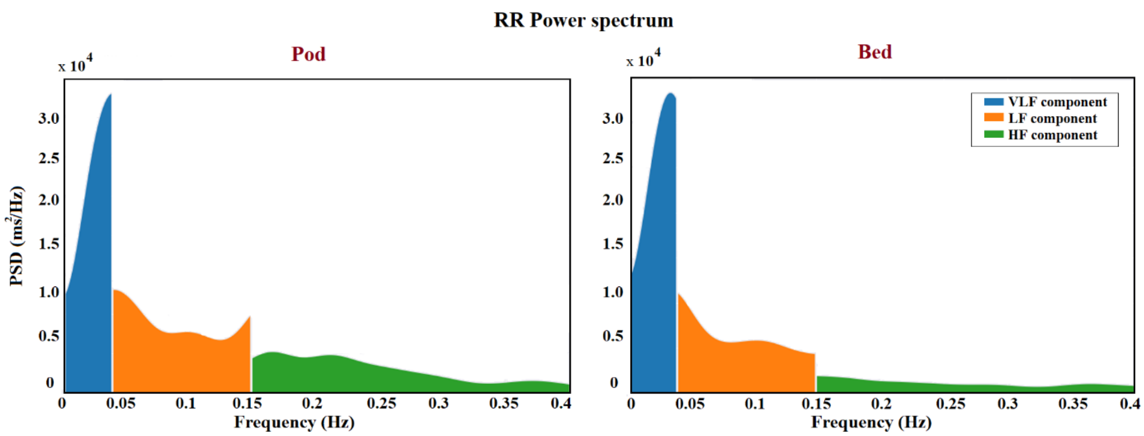

Figure 39. Power spectrum of RR time series for subject 4, interval 3 in pod (left) and bed (right) for 3 frequency bands (VLF: 0.0033–0.04 Hz, LF: 0.04–0.15 Hz, and HF: 0.15–0.40 Hz)

6.2.6 EEG signal processing

Preprocessing is an essential procedure for raw EEG data analysis. We used FieldTrip toolbox in this study. FieldTrip is open-source software available under the GNU General Public License

(GPL). The EEG signals were recorded from AF3, AF4, AF7, and AF8, FP1 and FP2 electrodes with a sampling frequency of 1000 Hz. The scalp EEG electrodes were referenced to AFz. The experiment has break times. We extracted the intervals before and after these break times.

We also removed linear trends from the data. Then, we applied base correlations to these intervals.

6.2.7 Statistical Analysis

We analyzed the cardiac data to explore whether there is a difference in autonomous nervous system activity between the bed and the pod conditions and/or between the three-time intervals. To do so, we analyzed the RR intervals, RMSSD, and the LF/HF ratio in three separate, repeated-measure ANOVAs with two factors (*condition* and *interval*) and their interaction (*condition*interval*). For this analysis, we used JASP 11.1.0.

Also, we were interested in several factors potentially influencing the EEG power spectral density: *condition*, *interval*, *channel* (which electrode produced the given datapoint), and frequency *band* (alpha, beta, gamma). To analyze the effects of these factors and all their interactions (including their three-wise and four-wise interactions), we used a linear mixed-effect model. In general, the mixed-effect models quantify the relationships between the independent and dependent variables by breaking down the regression into both fixed effects and random effects. Fixed effects represent the global relationship between independent and dependent variables, while random effects represent the deviations from the global relationships within each group of datapoints belonging to the same group [340]. Specifically, we collected several datapoints from each subject, hence these datapoints are associated with each other by being collected from the same subject; hence, these datapoints were associated with each other. We therefore added the random effect of *subject* into our analysis of the four above-mentioned fixed effects (*condition*, *interval*, *channel*, *band*). For this analysis, we used TIBCO Statistica 13.

6.2.8 Sliding-window Analysis

To provide more insight into the temporal dynamics of the EEG power spectral density, we averaged the PSD for each interval, channel, and band using a 60 s sliding window with a step size of 10 s. To compute the average band power, we first needed to compute an estimate of the power spectral density for each band. The most widely used method to do so is the Welch's periodogram, a method of averaging consecutive Fourier transforms of small windows in the temporal signal domain. We used the Welch's method to construct a sliding average of the EEG power spectral density for each combination of condition, interval, and channel.

6.3 Results

We examined the cardiac RR (inter-beat) intervals from the ECG, and the electrical brain activity from the frontal and frontopolar EEG during a period of floating in the pod and resting on the bed in 17 healthy subjects. In all analyses, we segmented each session into three, approximately five-minutes-long intervals to analyze potential changes of the effects over time. First, we analyzed the cardiac activity by examining whether *condition* and *interval* have effect over (1) the RR intervals, (2) the RMSSD, and (3) LF/HF ratio. Second, we analyzed the brain activity by examining whether *condition*, *interval*, *channel*, and frequency *band* had an effect over the EEG power spectral density. Finally, we plotted the EEG power spectral density in a graph, showing how it changes over time during the three intervals for each channel and between the two conditions.

6.3.1 Temporal and spectral analysis of RR intervals

We used JASP 11.1.0 for all statistical analyses of the RR intervals and the heart rate variability.

Temporal analysis: We performed a repeated-measures ANOVA to examine the effect of *condition* and *interval*, as well as their interaction, on the **RR intervals**. We found that the ANOVA's assumption of sphericity was violated for *interval* (Mauchly's W= 0.416, approx. $X^2(2) = 13.152$, $p = 0.001$) and the interaction *interval*condition* (Mauchly's W= 0.382, approx. $X^2(2) = 14.439$, $p < 0.001$). Because both the Greenhouse-Geisser ε and Huynh-Feldt ε were lower than 0.75, we applied the Greenhouse-Geisser correction to the p-values in both *interval* and *interval*condition*. We did not find any significant main effect of *interval* ($F(1.263, 16) = 0.372$, $p = 0.598$, $\eta^2 = 0.002$). However, we found a weak but significant main effect of *condition* ($F(1, 16) = 4.650$, $p = 0.047$, $\eta^2 = 0.178$); specifically, the RR intervals were shorter in the pod than in the bed, indicating faster heart rate in the pod. We also found a weak but significant effect of the interaction *interval*condition* ($F(1.263, 16) = 4.518$, $p = 0.039$, $\eta^2 = 0.022$); nevertheless, the Holm's post-hoc test did not identify any significant pairwise comparisons, possibly due to the weakness of this effect. For the points estimates of the mean RR intervals along different intervals in both conditions, see Figure 40.

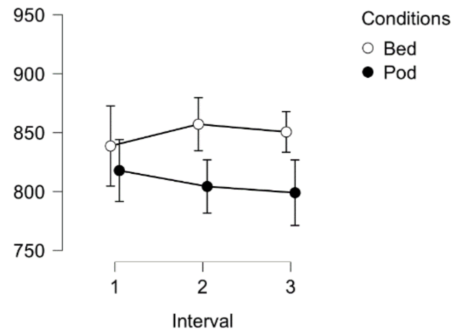

Figure 40. Point estimates of the RR intervals in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals.

Furthermore, we performed similarly designed repeated-measures ANOVA to examine the effect of *condition*, *interval*, and *interval*condition* on the cardiac **RMSSD**. We detected a violation of the sphericity assumption for *interval* (Mauchly's W= 0.620, approx. $X^2(2) = 7.178$, $p = 0.028$). Because the Greenhouse-Geisser ε was lower than 0.75, we applied the Greenhouse-Geisser p-value correction for the main effect of *interval*, which was not significant ($F(1.449, 16) = 1.212$, $p = 0.311$, $\eta^2 = 0.019$). Similarly, we did not find any significant effect for *condition* ($F(1, 16) = 0.389$, $p = 0.541$, $\eta^2 = 0.013$) nor for *interval*condition* ($F(2, 16) = 2.040$, $p = 0.147$, $\eta^2 = 0.021$). For the points estimates of the mean RMSSD, see Figure 41.

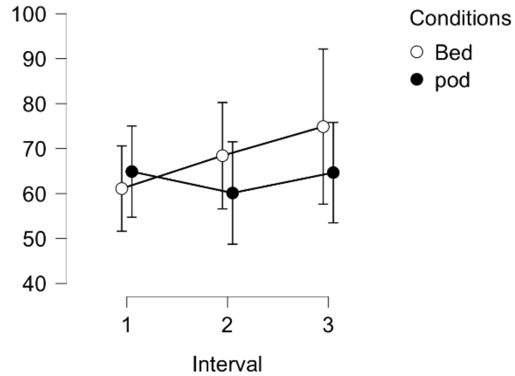

Figure 41. Point estimates of the RMSSD in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals.

Power spectrum analysis: We ran a repeated measures ANOVA to analyze the effect of *interval*, *condition*, and the interaction *interval*condition* on the LF/HF ratio. The analysis revealed a weak effect of *interval* ($F(2, 16) = 3.439$, $p = 0.044$, $\eta^2 = 0.042$). Nevertheless, due to the weakness of this effect, the Holm's post-hoc test did not identify any significant pairwise comparisons between the three intervals. As for the factor *condition* and the interaction *condition*interval*, the analysis revealed no significant effects ($F(1, 16) = 0.076$, $p = 0.786$, $\eta^2 = 0.003$, and $F(2, 16) = 0.355$, $p = 0.704$, $\eta^2 = 0.005$, respectively). For the point estimates of the mean LF/HF ratio, see Figure 42.

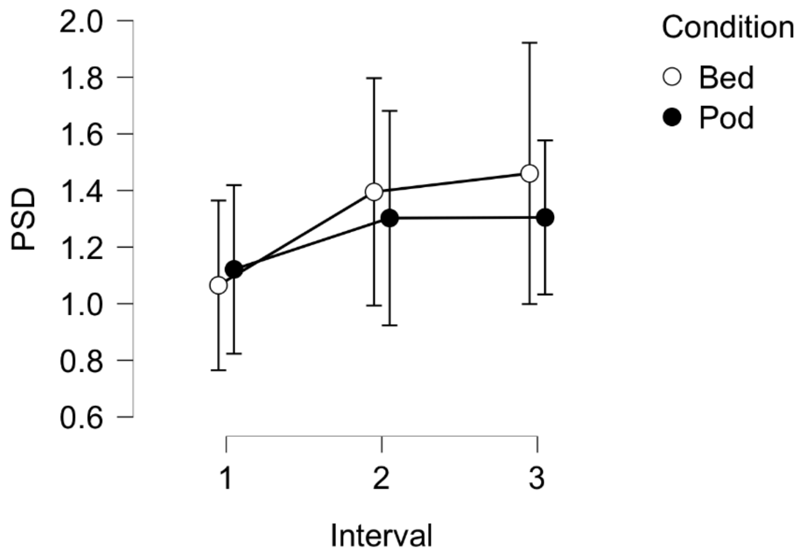

Figure 42. Point estimates of the LF/HF ratio in the three intervals in both the bed and the pod; the vertical lines denote 95% confidence intervals.

6.3.2 Spectral Analysis of the EEG signal

We analyzed the EEG results with TIBCO Statistica 13 using a mixed-effect model, i.e., a linear model including both fixed and random effects. In this analysis, we predicted the EEG power spectral density according to the random effects of *subject* and fixed effects of *condition, interval, channel*, and *band*. In addition, we also included all two-, three-, and four-way interactions between the fixed factors in the model. Although the input dependent variable is extremely positively skewed (see Figure 43-a), the model residuals are distributed symmetrically (see Figure 43-b). Despite that, our results from this analysis should be accepted cautiously because the residuals also exhibit non-negligible heteroscedasticity, probably due to the heavily skewed distribution of the input dependent variable (see Figure 43-c).

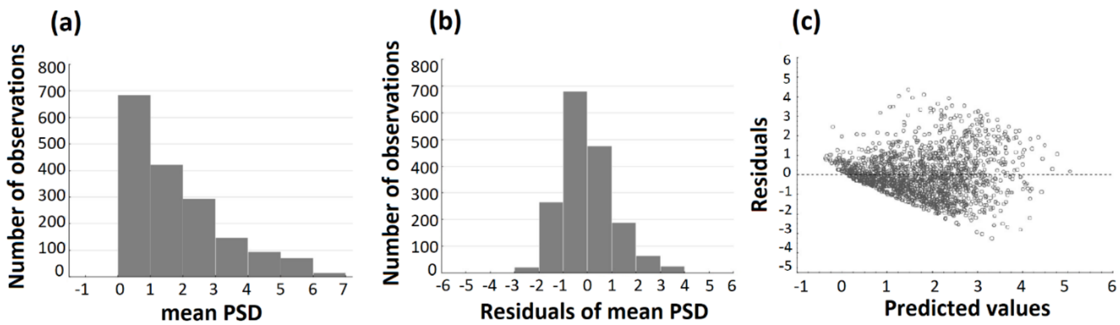

Figure 43. (a) Distribution of the dependent variable. (b) Distribution of the model's residuals. (c) Residual distribution vs. predicted values suggesting residuals' heteroscedasticity.

With appropriate caution, we proceeded to analyze the factors' effects. The main effects of factors *condition* ($F(1, 1602) = 41.918$, $p < 0.001$, $\eta^2 = 0.192$), *channel* ($F(5, 1602) = 148.423$, $p < 0.001$, $\eta^2 = 0.317$), and *band* ($F(2, 1602) = 35.632$, $p < 0.001$, $\eta^2 = 0.043$) were significant, while the main effect of *interval* remained insignificant ($F(2, 1602) = 1.746$, $p = 0.175$, $\eta^2 = 0.002$). The significant main effect of the condition is due to higher total PSD while lying in bed compared to floating (see Figure 44-a). Regarding to the significant effect of channel, the Scheffé post-hoc test revealed that it is caused by each pair of homologous channels (i.e., FP1 & FP2, AF3 & AF4, AF7 & AF8) exhibiting similar total PSD while at the same time differing significantly from the other pairs (all p's < 0.01; see Figure 44-b). The Scheffé post-hoc test on *band* revealed that the beta frequency band overall reaches significantly higher spectral power than the alpha and gamma bands (p's < 0.001), while the latter two do not differ in their total PSD ($p = 0.168$; see Figure 44-c).

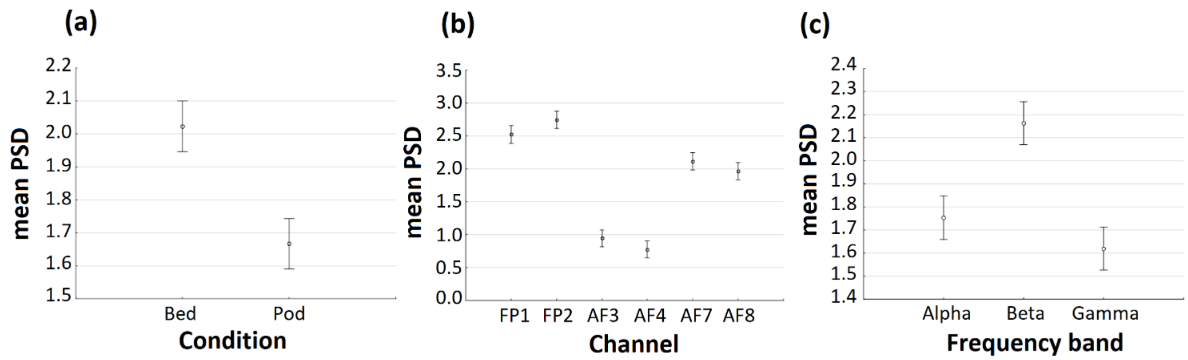

Figure 44. (a) Differences in total PSD between the bed and pod. (b) Differences in total PSD across the six channels. (c) Differences in total PSD between the three frequency bands. The vertical lines denote 95% confidence intervals.

The two-way interaction between *condition* and *channel* was significant ($F(5, 1602) = 2.402$, $p = 0.035$, $\eta^2 = 0.007$; see Figure 7). The Scheffé post-hoc test revealed one significant difference, suggesting that electrode AF8 exhibit lower spectral power from the bed to the float pod ($p = 0.007$). As can be seen from the Figure 45-a, a similar trend can also be observed for AF4 and AF7, although the Scheffé test did not identify these as significant ($p = 0.524$ and $p = 0.667$, respectively). All other pairwise comparisons were either irrelevant to our analysis, insignificant, or reflected the familiar, main effect of channel.

Crucially to our hypothesis, we found significant interaction between *condition* and *band* ($F(2, 1602) = 3.397$, $p = 0.034$, $\eta^2 = 0.004$). The Scheffé test did not reveal a significant difference between the two conditions in the alpha PSD ($p = 0.893$), but it did show a significant decrease of PSD in both the beta and gamma bands from the bed to the pod ($p < 0.001$ and $p = 0.0012$, respectively; see Figure 45-b). The remaining two-way and all the three- and four-way interactions remained insignificant.

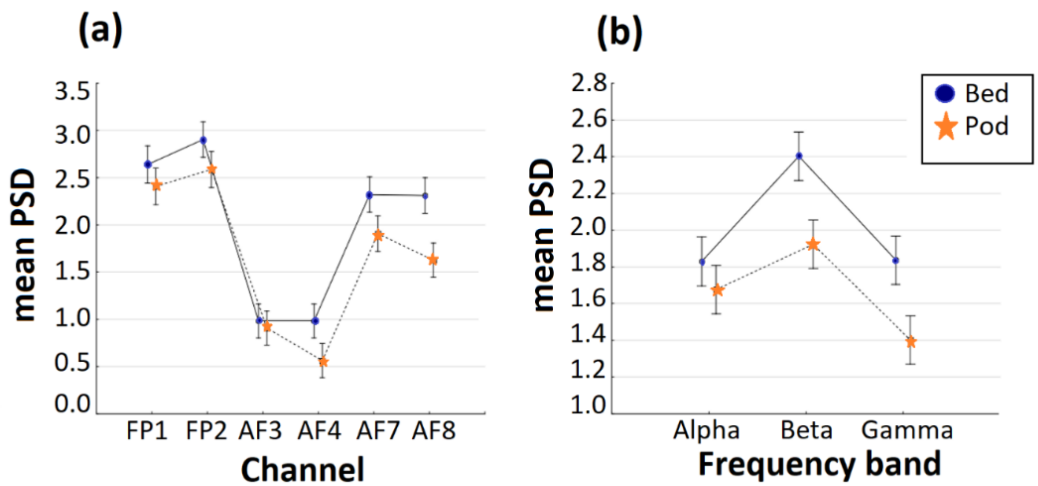

Figure 45. (a) Differences in total PSD between the six channels and the condition. (b) Differences in the PSD between alpha, beta, and gamma bands in the bed and pod conditions. The vertical lines denote 95% confidence intervals.

6.3.3 Sliding-window analysis

To investigate the PSD of EEG signals in more detail, we used a 60 s sliding window with a step size of 10 s on the PSD values of each frequency band for each of the 6 EEG channels. Figure 10 represents the analysis for the sliding average of the PSD over 17 subjects during the three intervals when the subjects are in float-pod or bed for 2 channels (AF7 and AF8). The shaded area shows the 95% confidence interval. From Figure 46-a we cannot see any significant differences between the conditions. However, during the second interval in beta and gamma frequency, there are some intervals for significant differences and at the end of experiments (Figure 46-c). Roughly 17 minutes after the start of the experiment, there is a significant difference between the conditions for 2 frequency bands, beta, and gamma on AF7 and AF8. The higher beta power in the bed condition suggests an increase in the subjects' alertness, and the higher gamma power in the bed condition suggests a prevalence of higher cognitive processes, such as attention and perception.

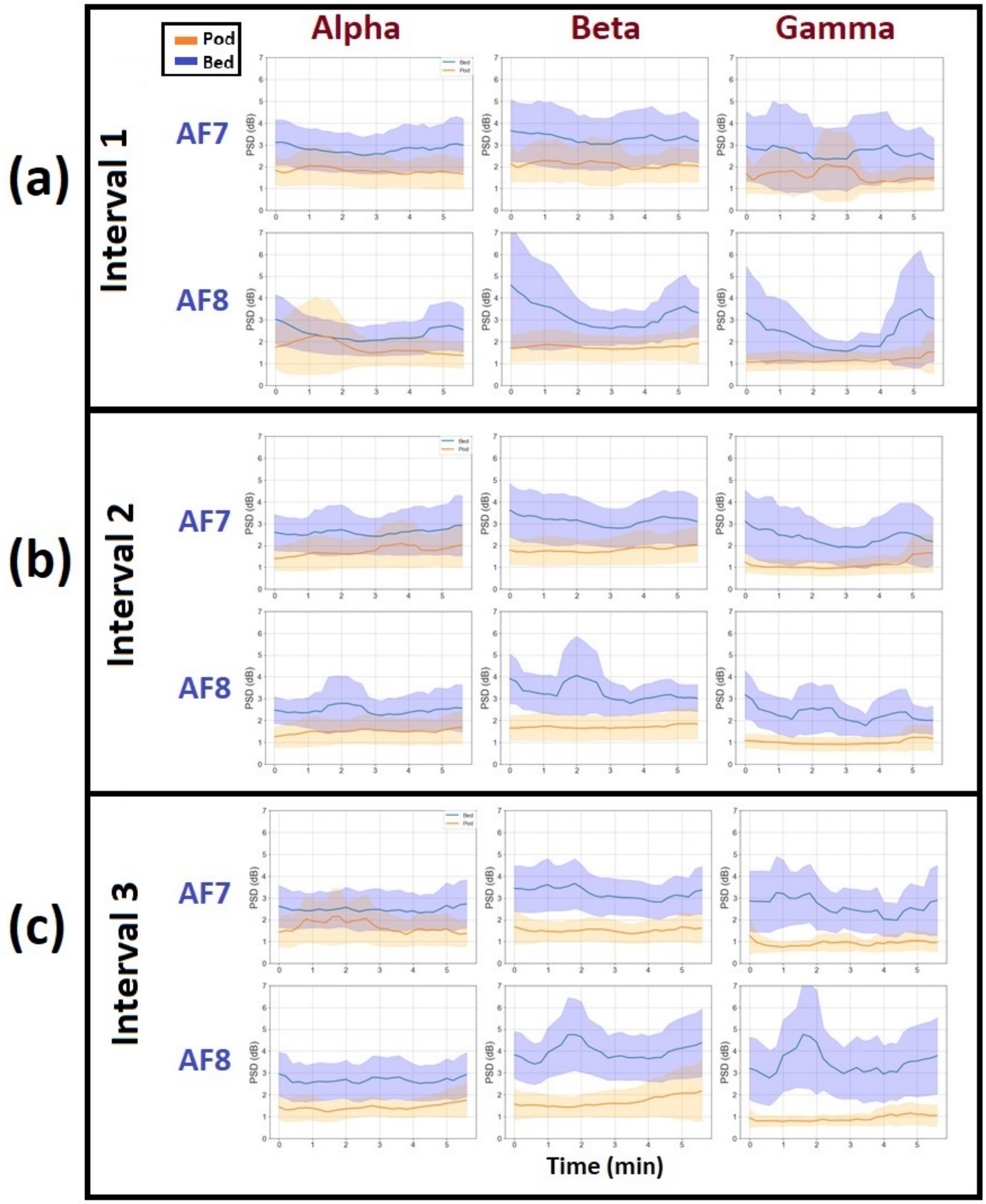

Figure 46. Sliding window with length 60 s and step size of 10 s on the EEG signals of 17 subjects. The plot represents three time intervals and two EEG channels (AF7 and AF8). Each column represents specific frequency band (left to right: alpha, beta and gamma). The shaded area depicts the 95% confidence band.

6.4 Discussion

Few studies on stress have combined EEG and ECG signals in their analyses [341]. But attending to both the brain and heart signals promises to increase the accuracy and reliability of an experiment because of the signals' synergy. Furthermore, such studies acknowledge the interrelation between neural and cardio functions during mental stress. Previous experiments on REST often reported positive effects of floating, such as improved relaxation [295], reduced anxiety [244, 245], stress reduction [295], pain and muscle tension reduction etc. [250, 290]. In 2008, Feinstein et al. examined whether Floatation-REST would attenuate symptoms of anxiety, stress, and depression in a clinical sample [294]. Their study found that a single, one-hour session of Floatation-REST was capable of inducing a strong reduction in state anxiety and a substantial improvement in mood in a group of 50 anxious and depressed participants with a range of different anxiety and stress-related disorders. The findings from their open-label study suggested that Floatation-REST was a promising technique for acutely reducing symptoms of anxiety and depression, although the persistence of these effects is presently unknown. They also suggested that a more active comparator than self-reported and blood pressure measuring would be needed to access the efficacy of floatation-REST on anxiety and stress [295]. Participants in that experiment reported significant reductions in stress, muscle tension, pain, depression and negative affect, accompanied by a significant improvement in mood characterized by increases in serenity, relaxation, happiness and overall well-being ($p < .0001$ for all variables). When compared to the results of a group of 30 non-anxious participants, the effects were found to be more robust in the anxious sample, approaching non-anxious levels during the post-float period. However, their non-floatation state was sitting upright in a chair, a feature that likely magnified the differences between conditions on measures of interoceptive awareness and muscle tension.

In this study, for the first time we recorded ECG and EEG in the float pod to analyze the effect of float pod based on neural and cardiac components [331]. There are some studies which recorded EEG, ECG, EOG and breathing pre-floatation REST and after, but none of them recorded these signal during floatation[298]. Also, our proposed non-floatation state was lying in bed to minimize the differences between conditions on measures of interoceptive awareness and muscle tension. However, we should consider that, even despite minimizing the posture differences in different condition, there may be two additional physiological mechanisms that likely contribute to the influence of posture on electrical scalp activity [342-344]: 1) alterations in cerebrospinal fluid (CSF) thickness and 2) changes in noradrenergic output. First, because CSF is highly conductive, minute shifts in CSF concentration can cause substantial alterations in EEG signals. Many studies revealed a main effect of posture in the beta and gamma bands. Beta and gamma activities increased from lying supine to inkling at 45 degree and increased further when subjects sat upright. These changes manifested regardless of whether participants engaged in a cognitive task and irrespective of whether their eyes were open or closed. However, in this study, by considering EEG and ECG signals, we examined the role of float pod as a relaxation or meditation technique and its effect on reducing stress for the first time. The analyses divided to: (1) temporal and spectral analyzing of RR intervals (2) EEG power spectral density.

We analyzed the temporal and spectral of RR intervals. We did not find any significant main effect of interval ($F(1.263, 16) = 0.372$, $p = 0.598$, $\eta^2 = 0.002$). However, we found a weak but significant main effect of condition ($F(1, 16) = 4.650$, $p = 0.047$, $\eta^2 = 0.178$); specifically, the RR intervals were shorter in the pod than in the bed, indicating a faster heart rate in the pod. We believe the faster heart rate in pod is the result of humidity. Furthermore, we found a weak but significant effect of the interaction interval*condition on the heart rate ($F(1.263, 16) = 4.518$, $p = 0.039$, $\eta^2 = 0.022$), which suggests the differences in heart rate between bed and pod appeared after 20 minutes. To capture fast changes in instantaneous heart rate, we measured RMSSD as well. However, we did not find any significant effect for condition ($F(1, 16) = 0.389$, $p = 0.541$, $\eta^2 = 0.013$) nor for interval*condition ($F(2, 16) = 2.040$, $p = 0.147$, $\eta^2 = 0.021$).

For spectral RR intervals, we calculated LF/HF ratio. Our assumptions underlying the LF/HF ratio was that LF power may be generated by the SNS while HF power is produced by the PNS. In this model, a low LF/HF ratio reflects parasympathetic dominance. The analysis revealed a significant difference of interval ($F(2, 16) = 3.439$, $p = 0.044$, $\eta^2 = 0.042$). Lower LF/HF ratio in pod rather than bed shows a more relaxing experience for participants in the pod [345].

We analyzed the EEG power spectral density with a mixed-effect model. We saw the significant differences in condition ($F(1, 1602) = 41.918$, $p < 0.001$, $\eta^2 = 0.192$), channel ($F(5, 1602) = 148.423$, $p < 0.001$, $\eta^2 = 0.317$), and band ($F(2, 1602) = 35.632$, $p < 0.001$, $\eta^2 = 0.043$). These findings on their own do not have any interpretation. Therefore, we investigated more to find differences in total PSD between conditions (channel and frequency band).

The two-way interaction between condition and channel was significant ($F(5, 1602) = 2.402$, $p = 0.035$, $\eta^2 = 0.007$; see Figure 45). The Scheffé post-hoc test revealed one relevant significant difference, suggesting that electrode AF8 exhibited lower spectral power in the bed than in the float pod ($p = 0.007$). This result was unexpected for us and needs more investigation.

We found significant interactions between condition and band ($F(2, 1602) = 3.397$, $p = 0.034$, $\eta^2 = 0.004$). Specifically, a lower PSD in beta and gamma in the pod compared to in bed. Stress due to tension, excitement, and anxiety has high power of beta band [346, 347]. We saw significantly less beta power in pod, implying less tension, excitement and anxiety in pod vs. in bed. Dedovic et al. found these difference in beta waves are highest in frontal part [346]. Furthermore, to the best of our knowledge, gamma power has been previously used to assess meditation [348, 349]. The cited meditation-related works found contrasting results regarding the positive or negative correlation between the gamma power and the meditation level. However, Minguillon et al. showed a positive correlation between gamma power and stress level, in particular, with the expected stress level on healthy subjects [350]. Our results support these findings by showing a significant decrease in gamma power in the float pod rather than in bed, revealing that floatation decreases stress more than lying in the bed.

To investigate more details of PSD over time, we used the sliding-window analysis with 60s length and step size 10s (Figure 46). In Figure 46, PSD for AF7 and AF8 were shown with 95% confidence interval. We can see during interval 3 or after 20 minutes in float-pod that there was a significant difference between conditions. This analysis is consistent with what we found in Figure 45.

Future research should explore these preliminary findings to determine whether Floatation-REST facilitates the practice of mindfulness and whether the combination of floating with specific mindfulness instructions can lead to even greater anxiolytic effects. Also, there should be more analysis on the relationship and connection between neural and cardiac signals such as phase-amplitude coupling (PAC) analysis [335]. It would also be interesting to compare neural and cardiac signals during floatation with those present during sleep.

Reference

1. Reaz, M.B.I., M. Hussain, and F. Mohd-Yasin, *Techniques of EMG signal analysis: detection, processing, classification and applications.* Biological procedures online, 2006. **8**(1): p. 11.
2. Chowdhury, R., et al., *Surface electromyography signal processing and classification techniques.* Sensors, 2013. **13**(9): p. 12431-12466.
3. Oskoei, M.A. and H. Hu, *Myoelectric control systems—A survey.* Biomedical signal processing and control, 2007. **2**(4): p. 275-294.
4. Lashgari, E. and E. Demircan, *Electromyography Pattern Classification with Laplacian Eigenmaps in Human Running.* World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 2017. **11**(4): p. 399-404.
5. Tsoli, A. and O.C. Jenkins. *Neighborhood denoising for learning high-dimensional grasping manifolds*. in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2008. IEEE.
6. Tsoli, A. and O.C. Jenkins. *2d subspaces for user-driven robot grasping*. in *Robotics, Science and Systems Conference: Workshop on Robot Manipulation*. 2007.
7. Totah, D., et al., *Low-back electromyography (EMG) data-driven load classification for dynamic lifting tasks.* PloS one, 2018. **13**(2).
8. Batzianoulis, I., et al., *EMG-based decoding of grasp gestures in reaching-to-grasping motions.* Robotics and Autonomous Systems, 2017. **91**: p. 59-70.
9. Na, Y., et al., *Ranking hand movements for myoelectric pattern recognition considering forearm muscle structure.* Medical & biological engineering & computing, 2017. **55**(8): p. 1507-1518.
10. Gailey, A., P. Artemiadis, and M. Santello, *Proof of concept of an online EMG-based decoding of hand postures and individual digit forces for prosthetic hand control.* Frontiers in neurology, 2017. **8**: p. 7.
11. Wu, Y., et al., *Exploration of Feature Extraction Methods and Dimension for sEMG Signal Classification.* Applied Sciences, 2019. **9**(24): p. 5343.
12. Boostani, R. and M.H. Moradi, *Evaluation of the forearm EMG signal features for the control of a prosthetic hand.* Physiological measurement, 2003. **24**(2): p. 309.
13. Lashgari, E., A. Pouya, and U. Maoz, *Decoding object weight from electromyography during human grasping.* BioRxiv, 2021.
14. Kauppi, J.-P., et al., *Three-way analysis of spectrospatial electromyography data: Classification and interpretation.* PloS one, 2015. **10**(6).
15. Phinyomark, A., P. Phukpattaranont, and C. Limsakul, *Feature reduction and selection for EMG signal classification.* Expert systems with applications, 2012. **39**(8): p. 7420-7431.

16. Mordohai, P. and G. Medioni, *Dimensionality estimation, manifold learning and function approximation using tensor voting.* Journal of Machine Learning Research, 2010. **11**(Jan): p. 411-450.
17. Lashghari, E., S. Rosenberg, and F. Qu, *Atrial flutter detection utilizing nonlinear dimension reduction*. 2020, Google Patents.
18. Lashgari, E., S. Rosenberg, and F. Qu, *Atrial flutter detection utilizing nonlinear dimension reduction*. 2021, Google Patents.
19. Jolliffe, I., *Principal component analysis.* Technometrics, 2003. **45**(3): p. 276.
20. Cox, M.A. and T.F. Cox, *Multidimensional scaling*, in *Handbook of data visualization*. 2008, Springer. p. 315-347.
21. Rabin, N., et al., *Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques.* Expert Systems with Applications, 2020. **149**: p. 113281.
22. Roweis, S.T. and L.K. Saul, *Nonlinear dimensionality reduction by locally linear embedding.* science, 2000. **290**(5500): p. 2323-2326.
23. Saul, L.K. and S.T. Roweis, *Think globally, fit locally: unsupervised learning of low dimensional manifolds.* Journal of machine learning research, 2003. **4**(Jun): p. 119-155.
24. Tenenbaum, J.B., V. De Silva, and J.C. Langford, *A global geometric framework for nonlinear dimensionality reduction.* science, 2000. **290**(5500): p. 2319-2323.
25. Belkin, M. and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation.* Neural computation, 2003. **15**(6): p. 1373-1396.
26. Lashgari, E. and E. Demircan, *Electromyography Pattern Classification with Laplacian Eigenmaps in Human Running.* International Journal of Electronics and Communication Engineering, 2017. **11**(4): p. 412-417.
27. Pham, H., M. Kawanishi, and T. Narikiyo. *Recognition of walking movement from EMG using a framework combining LLE and HMM*. in *2014 IEEE/SICE International Symposium on System Integration*. 2014. IEEE.
28. Lashgari, E., D. Liang, and U. Maoz, *Data Augmentation for Deep-Learning-Based Electroencephalography.* Journal of Neuroscience Methods, 2020: p. 108885.
29. Xia, P., J. Hu, and Y. Peng, *EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks.* Artificial organs, 2018. **42**(5): p. E67-E77.
30. Phinyomark, A., R. N Khushaba, and E. Scheme, *Feature extraction and selection for myoelectric control based on wearable EMG sensors.* Sensors, 2018. **18**(5): p. 1615.
31. Zhai, X., et al., *Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network.* Frontiers in neuroscience, 2017. **11**: p. 379.
32. Luciw, M.D., E. Jarocka, and B.B. Edin, *Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction.* Scientific data, 2014. **1**: p. 140047.
33. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.
34. Yu, H. and J. Yang, *A direct LDA algorithm for high-dimensional data—with application to face recognition.* Pattern recognition, 2001. **34**(10): p. 2067-2070.

35. Pai, G., et al. *Dimal: Deep isometric manifold learning using sparse geodesic sampling*. in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2019. IEEE.
36. Maaten, L.v.d. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(Nov): p. 2579-2605.
37. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, 2011. **12**: p. 2825-2830.
38. Van Der Maaten, L., E. Postma, and J. Van den Herik, *Dimensionality reduction: a comparative.* J Mach Learn Res, 2009. **10**(66-71): p. 13.
39. Van der Maaten, L., E.O. Postma, and H.J. van den Herik, *Matlab toolbox for dimensionality reduction.* MICC, Maastricht University, 2007.
40. Von Luxburg, U., *A tutorial on spectral clustering.* Statistics and computing, 2007. **17**(4): p. 395-416.
41. Kim, K.S., et al., *Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions.* Current applied physics, 2011. **11**(3): p. 740-745.
42. Verikas, A., et al., *Electromyographic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness.* Sensors, 2016. **16**(4): p. 592.
43. Atzori, M., M. Cognolato, and H. Müller, *Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands.* Frontiers in neurorobotics, 2016. **10**: p. 9.
44. He, H. and E.A. Garcia, *Learning from imbalanced data.* IEEE Transactions on knowledge and data engineering, 2009. **21**(9): p. 1263-1284.
45. He, H. and Y. Ma, *Imbalanced learning: foundations, algorithms, and applications*. 2013: John Wiley & Sons.
46. Luciw, M.D., E. Jarocka, and B.B. Edin, *Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction.* Scientific data, 2014. **1**(1): p. 1-11.
47. An, J. and S. Cho. *Hand motion identification of grasp-and-lift task from electroencephalography recordings using recurrent neural networks*. in *2016 International Conference on Big Data and Smart Computing (BigComp)*. 2016. IEEE.
48. Várszegi, K. *Comparison of algorithms for detecting hand movement from EEG signals*. in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2016. IEEE.
49. Eilbeigi, E. and S.K. Setarehdan, *Detecting intention to execute the next movement while performing current movement from EEG using global optimal constrained ICA.* Computers in Biology and Medicine, 2018. **99**: p. 63-75.
50. Shen, F., J. Liu, and K. Wu, *Multivariate Time Series Forecasting based on Elastic Net and High-Order Fuzzy Cognitive Maps: A Case Study on Human Action Prediction through EEG Signals.* IEEE Transactions on Fuzzy Systems, 2020.
51. Cisotto, G., et al. *Classification of grasping tasks based on EEG-EMG coherence*. in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. 2018. IEEE.
52. Côté-Allard, U., et al., *Interpreting deep learning features for myoelectric control: A comparison with handcrafted features.* Frontiers in Bioengineering and Biotechnology, 2020. **8**: p. 158.

53. Maoz, U., et al. *Predicting action content on-line and in real time before action onset–an intracranial human study*. in *Advances in Neural Information Processing Systems*. 2012.
54. Pless, R. and R. Souvenir, *A survey of manifold learning for images.* IPSJ Transactions on Computer Vision and Applications, 2009. **1**: p. 83-94.
55. Souvenir, R. and R. Pless, *Image distance functions for manifold learning.* Image and Vision Computing, 2007. **25**(3): p. 365-373.
56. Yannick, R., et al., *Deep learning-based electroencephalography analysis: a systematic review.* arXiv preprint arXiv:1901.05498, 2019.
57. Cohen, M.X., *Analyzing neural time series data: theory and practice*. 2014: MIT press.
58. Bigdely-Shamlo, N., et al., *The PREP pipeline: standardized preprocessing for large-scale EEG analysis.* Frontiers in neuroinformatics, 2015. **9**: p. 16.
59. Jas, M., et al., *Autoreject: automated artifact rejection for MEG and EEG data.* NeuroImage, 2017. **159**: p. 417-429.
60. Cole, S.R. and B. Voytek, *Cycle-by-cycle analysis of neural oscillations.* bioRxiv, 2018: p. 302000.
61. Gramfort, A., et al., *Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations.* NeuroImage, 2013. **70**: p. 410-422.
62. Hively, L., V. Protopopescu, and P. Gailey, *Timely detection of dynamical change in scalp EEG signals.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 2000. **10**(4): p. 864-875.
63. Subasi, A. and M.I. Gursoy, *EEG signal classification using PCA, ICA, LDA and support vector machines.* Expert systems with applications, 2010. **37**(12): p. 8659-8666.
64. Sanz-García, A., et al., *Potential EEG biomarkers of sedation doses in intensive care patients unveiled by using a machine learning approach.* Journal of neural engineering, 2019. **16**(2): p. 026031.
65. Lotte, F., et al., *A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update.* Journal of neural engineering, 2018. **15**(3): p. 031005.
66. Chollet, F., *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. 2018: MITP-Verlags GmbH & Co. KG.
67. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
68. Farabet, C., et al., *Learning hierarchical features for scene labeling.* IEEE transactions on pattern analysis and machine intelligence, 2013. **35**(8): p. 1915-1929.
69. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
70. Tompson, J.J., et al. *Joint training of a convolutional network and a graphical model for human pose estimation*. in *Advances in neural information processing systems*. 2014.
71. Hinton, G., et al., *Deep neural networks for acoustic modeling in speech recognition.* IEEE Signal processing magazine, 2012. **29**.
72. Mikolov, T., et al. *Strategies for training large scale neural network language models*. in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*. 2011. IEEE.
73. Sainath, T.N., et al. *Deep convolutional neural networks for LVCSR*. in *2013 IEEE international conference on acoustics, speech and signal processing*. 2013. IEEE.
74. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436.

75. Craik, A., Y. He, and J.L. Contreras-Vidal, *Deep learning for electroencephalogram (EEG) classification tasks: a review.* Journal of neural engineering, 2019. **16**(3): p. 031001.
76. Page, A., C. Shea, and T. Mohsenin. *Wearable seizure detection using convolutional neural networks with transfer learning*. in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2016. IEEE.
77. Perez, L. and J. Wang, *The effectiveness of data augmentation in image classification using deep learning.* arXiv preprint arXiv:1712.04621, 2017.
78. Wang, F., et al. *Data augmentation for eeg-based emotion recognition with deep convolutional neural networks*. in *International Conference on Multimedia Modeling*. 2018. Springer.
79. Zhang, C., et al., *Understanding deep learning requires rethinking generalization.* arXiv preprint arXiv:1611.03530, 2016.
80. Truong, N.D., et al., *Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram.* Neural Networks, 2018. **105**: p. 104-111.
81. Dümpelmann, M., *Early seizure detection for closed loop direct neurostimulation devices in epilepsy.* Journal of neural engineering, 2019. **16**(4): p. 041001.
82. Zhang, Z., et al., *A novel deep learning approach with data augmentation to classify motor imagery signals.* IEEE Access, 2019. **7**: p. 15945-15954.
83. Malhotra, R.K. and A.Y. Avidan, *Sleep stages and scoring technique.* Atlas of sleep medicine, 2013: p. 77-99.
84. Moser, D., et al., *Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters.* Sleep, 2009. **32**(2): p. 139-149.
85. Sors, A., et al., *A convolutional neural network for sleep stage scoring from raw single-channel EEG.* Biomedical Signal Processing and Control, 2018. **42**: p. 107-114.
86. Chambon, S., et al., *A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018. **26**(4): p. 758-769.
87. Kuanar, S., et al. *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)*. 2018. IEEE.
88. Sakai, A., Y. Minoda, and K. Morikawa. *Data augmentation methods for machine-learning-based classification of bio-signals*. in *2017 10th Biomedical Engineering International Conference (BMEiCON)*. IEEE.
89. Kwak, N.-S., K.-R. Müller, and S.-W. Lee, *A convolutional neural network for steady state visual evoked potential classification under ambulatory environment.* PloS one, 2017. **12**(2): p. e0172578.
90. Bashivan, P., et al., *Learning representations from EEG with deep recurrent-convolutional neural networks.* arXiv preprint arXiv:1511.06448, 2015.
91. Yin, Z. and J. Zhang, *Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights.* Neurocomputing, 2017. **260**: p. 349-366.
92. Yin, Z. and J. Zhang, *Cross-session classification of mental workload levels using EEG and an adaptive deep learning model.* Biomedical Signal Processing and Control, 2017. **33**: p. 30-47.

93. Hussein, R., et al., *Epileptic seizure detection: A deep learning approach.* arXiv preprint arXiv:1803.09848, 2018.
94. Salama, E.S., et al., *EEG-based emotion recognition using 3D convolutional neural networks.* Int. J. Adv. Comput. Sci. Appl, 2018. **9**(8): p. 329-337.
95. Parvan, M., et al. *Transfer Learning based Motor Imagery Classification using Convolutional Neural Networks*. in *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. 2019. IEEE.
96. Li, Y., et al., *A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019.
97. Goodfellow, I., et al. *Generative adversarial nets*. in *Advances in neural information processing systems*. 2014.
98. Zhang, H., et al. *Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
99. Bousmalis, K., et al. *Unsupervised pixel-level domain adaptation with generative adversarial networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
100. Antoniou, A., A. Storkey, and H. Edwards, *Data augmentation generative adversarial networks.* arXiv preprint arXiv:1711.04340, 2017.
101. Zhang, Q. and Y. Liu, *Improving brain computer interface performance by data augmentation with conditional Deep Convolutional Generative Adversarial Networks.* arXiv preprint arXiv:1806.07108, 2018.
102. Piplani, T., N. Merill, and J. Chuang, *Faking it, Making it: Fooling and Improving Brain-Based Authentication with Generative Adversarial Networks.*
103. Zhang, X., et al., *DADA: Deep adversarial data augmentation for extremely low data regime classification.* arXiv preprint arXiv:1809.00981, 2018.
104. Schlögl, A., *Outcome of the BCI-competition 2003 on the Graz data set.* Berlin, Germany: Graz University of Technology, 2003.
105. Hartmann, K.G., R.T. Schirrmeister, and T. Ball, *EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals.* arXiv preprint arXiv:1806.01875, 2018.
106. Salimans, T., et al. *Improved techniques for training gans*. in *Advances in neural information processing systems*. 2016.
107. Heusel, M., et al. *Gans trained by a two time-scale update rule converge to a local nash equilibrium*. in *Advances in Neural Information Processing Systems*. 2017.
108. Rabin, J., et al. *Wasserstein barycenter and its application to texture mixing*. in *International Conference on Scale Space and Variational Methods in Computer Vision*. 2011. Springer.
109. Luo, Y. and B.-L. Lu. *EEG data augmentation for emotion recognition using a conditional Wasserstein GAN*. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2018. IEEE.
110. Zheng, W.-L. and B.-L. Lu, *Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks.* IEEE Transactions on Autonomous Mental Development, 2015. **7**(3): p. 162-175.

111. Koelstra, S., et al., *Deap: A database for emotion analysis; using physiological signals.* IEEE transactions on affective computing, 2012. **3**(1): p. 18-31.
112. Luo, Y., L.-Z. Zhu, and B.-L. Lu. *A GAN-Based Data Augmentation Method for Multimodal Emotion Recognition*. in *International Symposium on Neural Networks*. 2019. Springer.
113. Wei, Z., et al., *Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain.* Biomedical Signal Processing and Control, 2019. **53**: p. 101551.
114. Chang, S. and H. Jun, *Hybrid deep-learning model to recognise emotional responses of users towards architectural design alternatives.* Journal of Asian Architecture and Building Engineering, 2019. **18**(5): p. 381-391.
115. Yang, B., et al. *A Framework on Optimization Strategy for EEG Motor Imagery Recognition*. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019. IEEE.
116. Panwar, S., et al. *Generating EEG signals of an RSVP Experiment by a Class Conditioned Wasserstein Generative Adversarial Network*. in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 2019. IEEE.
117. Gulrajani, I., et al. *Improved training of wasserstein gans*. in *Advances in neural information processing systems*. 2017.
118. Touryan, J., et al., *Estimating endogenous changes in task performance from EEG.* Frontiers in neuroscience, 2014. **8**: p. 155.
119. Radford, A., L. Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks.* arXiv preprint arXiv:1511.06434, 2015.
120. Kingma, D.P. and M. Welling, *Auto-encoding variational bayes.* arXiv preprint arXiv:1312.6114, 2013.
121. Aznan, N.K.N., et al. *Using variable natural environment brain-computer interface stimuli for real-time humanoid robot navigation*. in *2019 International Conference on Robotics and Automation (ICRA)*. 2019. IEEE.
122. O'Shea, A., et al. *Neonatal seizure detection using convolutional neural networks*. in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. 2017. IEEE.
123. Schirrmeister, R.T., et al., *Deep learning with convolutional neural networks for EEG decoding and visualization.* Human brain mapping, 2017. **38**(11): p. 5391-5420.
124. Ullah, I., M. Hussain, and H. Aboalsamh, *An automated system for epilepsy detection using EEG brain signals based on deep learning approach.* Expert Systems with Applications, 2018. **107**: p. 61-71.
125. Truong, N.D., et al., *Semi-supervised Seizure Prediction with Generative Adversarial Networks.* arXiv preprint arXiv:1806.08235, 2018.
126. Majidov, I. and T. Whangbo, *Efficient Classification of Motor Imagery Electroencephalography Signals Using Deep Learning Methods.* Sensors, 2019. **19**(7): p. 1736.
127. Mousavi, Z., et al., *Deep convolutional neural network for classification of sleep stages from single-channel EEG signals.* Journal of neuroscience methods, 2019: p. 108312.

128. Avcu, M.T., Z. Zhang, and D.W.S. Chan. *Seizure Detection Using Least Eeg Channels by Deep Convolutional Neural Network*. in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019. IEEE.
129. Tayeb, Z., et al., *Validating deep neural networks for online decoding of motor imagery movements from EEG signals.* Sensors, 2019. **19**(1): p. 210.
130. Tsiouris, K.M., et al., *A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals.* Computers in biology and medicine, 2018. **99**: p. 24-37.
131. Tang, Y., S. Wada, and K. Yoshihara. *Failure Prediction with Adaptive Multi-scale Sampling and Activation Pattern Regularization*. in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. 2017. IEEE.
132. Manor, R. and A.B. Geva, *Convolutional neural network for multi-category rapid serial visual presentation BCI.* Frontiers in computational neuroscience, 2015. **9**: p. 146.
133. Drouin-Picaro, A. and T.H. Falk. *Using deep neural networks for natural saccade classification from electroencephalograms*. in *2016 IEEE EMBS International Student Conference (ISC)*. 2016. IEEE.
134. Supratak, A., et al., *DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017. **25**(11): p. 1998-2008.
135. Dong, H., et al., *Mixed neural network approach for temporal sleep stage classification.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017. **26**(2): p. 324-333.
136. Ruffini, G., et al., *Deep learning using EEG spectrograms for prognosis in idiopathic rapid eye movement behavior disorder (RBD).* bioRxiv, 2018: p. 240267.
137. Sun, C., et al., *A Two-Stage Neural Network for Sleep Stage Classification Based on Feature Learning, Sequence Learning, and Data Augmentation.* IEEE Access, 2019. **7**: p. 109386-109397.
138. Thodoroff, P., J. Pineau, and A. Lim. *Learning robust features using deep learning for automatic seizure detection*. in *Machine learning for healthcare conference*. 2016.
139. Sengur, A., et al., *Neutrosophic similarity score-based entropy measure for focal and nonfocal electroencephalogram signal classification*, in *Neutrosophic Set in Medical Image Analysis*. 2019, Elsevier. p. 247-268.
140. Schwabedal, J.T., et al., *Addressing Class Imbalance in Classification Problems of Noisy Signals by using Fourier Transform Surrogates.* arXiv preprint arXiv:1806.08675, 2018.
141. Ruffini, G., et al., *Deep learning with EEG spectrograms in rapid eye movement behavior disorder.* bioRxiv, 2018: p. 240267.
142. Shamwell, J., et al. *Single-trial EEG RSVP classification using convolutional neural networks*. in *Micro-and Nanotechnology Sensors, Systems, and Applications VIII*. 2016. International Society for Optics and Photonics.
143. Said, A.B., et al. *Multimodal deep learning approach for joint EEG-EMG data compression and classification*. in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. 2017. IEEE.
144. Zhang, Y., et al., *Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network.* IEEE Journal of Biomedical and Health Informatics, 2019.

145. Dai, G., et al., *HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification.* Journal of neural engineering, 2020. **17**(1): p. 016025.
146. Frydenlund, A. and F. Rudzicz. *Emotional affect estimation using video and EEG data in deep neural networks*. in *Canadian Conference on Artificial Intelligence*. 2015. Springer.
147. Deiss, O., et al., *HAMLET: Interpretable Human And Machine co-LEarning Technique.* arXiv preprint arXiv:1803.09702, 2018.
148. Shovon, S.T.H., et al. *Classification of Motor Imagery EEG Signals with multi-input Convolutional Neural Network by augmenting STFT*. in *5th International Conference on Advances in Electrical Engineering (ICAEE). IEEE*. 2019.
149. Freer, D. and G.-Z. Yang, *Data augmentation for self-paced motor imagery classification with C-LSTM.* Journal of neural engineering, 2019.
150. Mokatren, L.S., et al., *Improved EEG Classification by factoring in sensor topography.* arXiv preprint arXiv:1905.09472, 2019.
151. Abiri, R., et al., *A comprehensive review of EEG-based brain–computer interface paradigms.* Journal of neural engineering, 2019. **16**(1): p. 011001.
152. Nicolas-Alonso, L.F. and J. Gomez-Gil, *Brain computer interfaces, a review.* sensors, 2012. **12**(2): p. 1211-1279.
153. Salvaris, M. and P. Haggard, *Decoding intention at sensorimotor timescales.* PloS one, 2014. **9**(2): p. e85100.
154. Schneider, L., et al., *What we think before a voluntary movement.* Journal of cognitive neuroscience, 2013. **25**(6): p. 822-829.
155. Schultze-Kraft, M., et al., *The point of no return in vetoing self-initiated movements.* Proceedings of the National Academy of Sciences, 2016. **113**(4): p. 1080-1085.
156. Lashgari, E. and U. Maoz, *Electromyography Classification during Reach-to-Grasp Motion using Manifold Learning.* bioRxiv, 2020.
157. Wolpaw, J.R., et al., *Brain–computer interfaces for communication and control.* Clinical neurophysiology, 2002. **113**(6): p. 767-791.
158. Daly, J.J. and J.R. Wolpaw, *Brain–computer interfaces in neurological rehabilitation.* The Lancet Neurology, 2008. **7**(11): p. 1032-1043.
159. Machado, S., L.F. Almada, and R.N. Annavarapu, *Progress and prospects in EEG-based brain-computer interface: clinical applications in neurorehabilitation.* Journal of Rehabilitation Robotics, 2013. **1**(1): p. 28-41.
160. Mulder, T., *Motor imagery and action observation: cognitive tools for rehabilitation.* Journal of neural transmission, 2007. **114**(10): p. 1265-1278.
161. Contreras-Vidal, J.L., et al., *Restoration of whole body movement: toward a noninvasive brain-machine interface system.* IEEE pulse, 2012. **3**(1): p. 34-37.
162. Pfurtscheller, G. and C. Neuper, *Motor imagery activates primary sensorimotor area in humans.* Neuroscience letters, 1997. **239**(2-3): p. 65-68.
163. He, B., et al., *Noninvasive brain-computer interfaces based on sensorimotor rhythms.* Proceedings of the IEEE, 2015. **103**(6): p. 907-925.
164. Yuan, H. and B. He, *Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives.* IEEE Transactions on Biomedical Engineering, 2014. **61**(5): p. 1425-1435.

165. Morash, V., et al., *Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries.* Clinical neurophysiology, 2008. **119**(11): p. 2570-2578.
166. Pfurtscheller, G. and F.L. Da Silva, *Event-related EEG/MEG synchronization and desynchronization: basic principles.* Clinical neurophysiology, 1999. **110**(11): p. 1842-1857.
167. Tabar, Y.R. and U. Halici, *A novel deep learning approach for classification of EEG motor imagery signals.* Journal of neural engineering, 2016. **14**(1): p. 016003.
168. Zabidi, A., et al. *Short-time Fourier Transform analysis of EEG signal generated during imagined writing*. in *2012 International Conference on System Engineering and Technology (ICSET)*. 2012. IEEE.
169. Amin, H.U., et al., *Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques.* Australasian physical & engineering sciences in medicine, 2015. **38**(1): p. 139-149.
170. Edelman, B.J., B. Baxter, and B. He, *EEG source imaging enhances the decoding of complex right-hand motor imagery tasks.* IEEE Transactions on Biomedical Engineering, 2015. **63**(1): p. 4-14.
171. Ang, K.K., et al. *Filter bank common spatial pattern (FBCSP) in brain-computer interface*. in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. 2008. IEEE.
172. Meng, J., et al., *Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks.* Scientific Reports, 2016. **6**: p. 38565.
173. Gandhi, T., B.K. Panigrahi, and S. Anand, *A comparative study of wavelet families for EEG signal classification.* Neurocomputing, 2011. **74**(17): p. 3051-3057.
174. Yang, Y., et al. *Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG*. in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012. IEEE.
175. Baldi, P., *Deep Learning in Science: Theory, Algorithms, and Applications*. 2021, Cambridge University Press, Cambridge, UK.
176. Lawhern, V.J., et al., *EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces.* Journal of neural engineering, 2018. **15**(5): p. 056013.
177. Lu, N., et al., *A deep learning scheme for motor imagery classification based on restricted boltzmann machines.* IEEE transactions on neural systems and rehabilitation engineering, 2016. **25**(6): p. 566-576.
178. Tortora, S., et al., *Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network.* Journal of Neural Engineering, 2020. **17**(4): p. 046011.
179. Zhang, H., et al., *Motor imagery recognition with automatic EEG channel selection and deep learning.* Journal of Neural Engineering, 2021. **18**(1): p. 016004.
180. Zhang, R., et al., *A novel hybrid deep learning scheme for four-class motor imagery classification.* Journal of neural engineering, 2019. **16**(6): p. 066004.
181. Zhang, G., et al., *Classification of hand movements from EEG using a deep attention-based LSTM network.* IEEE Sensors Journal, 2019. **20**(6): p. 3113-3122.
182. Wang, S., et al. *Training deep neural networks on imbalanced data sets*. in *2016 international joint conference on neural networks (IJCNN)*. 2016. IEEE.
183. Vaswani, A., et al., *Attention is all you need.* arXiv preprint arXiv:1706.03762, 2017.

184. Shaw, P., J. Uszkoreit, and A. Vaswani, *Self-attention with relative position representations.* arXiv preprint arXiv:1803.02155, 2018.
185. Cisotto, G., et al., *Comparison of Attention-based Deep Learning Models for EEG Classification.* arXiv preprint arXiv:2012.01074, 2020.
186. Mrini, K., et al., *Rethinking self-attention: An interpretable self-attentive encoder-decoder parser.* 2019.
187. Hertel, L., et al., *Sherpa: Robust Hyperparameter Optimization for Machine Learning.* arXiv preprint arXiv:2005.04048, 2020.
188. Li, Y., et al., *A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019. **27**(6): p. 1170-1180.
189. Zhang, X., et al. *Dada: Deep adversarial data augmentation for extremely low data regime classification*. in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019. IEEE.
190. Mirza, M. and S. Osindero, *Conditional generative adversarial nets.* arXiv preprint arXiv:1411.1784, 2014.
191. Brunner, C., et al., *BCI Competition 2008–Graz data set A.* Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, 2008. **16**: p. 1-6.
192. Leeb, R., et al., *BCI Competition 2008–Graz data set B.* Graz University of Technology, Austria, 2008: p. 1-6.
193. Gaur, P., et al. *An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface*. in *2015 International Joint Conference on Neural Networks (IJCNN)*. 2015. IEEE.
194. Luo, J., et al., *Dynamic frequency feature selection based approach for classification of motor imageries.* Computers in biology and medicine, 2016. **75**: p. 45-53.
195. Jurcak, V., D. Tsuzuki, and I. Dan, *10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems.* Neuroimage, 2007. **34**(4): p. 1600-1611.
196. Kevric, J. and A. Subasi, *Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system.* Biomedical Signal Processing and Control, 2017. **31**: p. 398-406.
197. Sadiq, M.T., et al., *Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform.* IEEE Access, 2019. **7**: p. 127678-127692.
198. Sadiq, M.T., et al., *Motor Imagery EEG Signals Decoding by Multivariate Empirical Wavelet Transform-Based Framework for Robust Brain–Computer Interfaces.* IEEE Access, 2019. **7**: p. 171431-171451.
199. Shahid, S. and G. Prasad, *Bispectrum-based feature extraction technique for devising a practical brain–computer interface.* Journal of neural engineering, 2011. **8**(2): p. 025014.
200. Ang, K.K., et al., *Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b.* Frontiers in neuroscience, 2012. **6**: p. 39.
201. Saa, J.F.D. and M. Çetin, *A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data.* Journal of neural engineering, 2012. **9**(2): p. 026020.

202. Li, M.-a., et al., *Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-isomap.* Applied Sciences, 2017. **7**(4): p. 390.
203. Zheng, Q., F. Zhu, and P.-A. Heng, *Robust support matrix machine for single trial EEG classification.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018. **26**(3): p. 551-562.
204. Lotte, F. and C. Guan, *Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms.* IEEE Transactions on biomedical Engineering, 2010. **58**(2): p. 355-362.
205. Raza, H., et al., *Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface.* Soft Computing, 2016. **20**(8): p. 3085-3096.
206. Gaur, P., et al., *A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry.* Expert Systems with Applications, 2018. **95**: p. 201-211.
207. Blankertz, B., et al., *The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects.* NeuroImage, 2007. **37**(2): p. 539-550.
208. Willett, F.R., et al., *High-performance brain-to-text communication via handwriting.* Nature, 2021. **593**(7858): p. 249-254.
209. Jeannerod, M., *Mental imagery in the motor context.* Neuropsychologia, 1995. **33**(11): p. 1419-1432.
210. Zich, C., et al., *Real-time EEG feedback during simultaneous EEG–fMRI identifies the cortical signature of motor imagery.* Neuroimage, 2015. **114**: p. 438-447.
211. Lotze, M., et al., *Phantom movements and pain An fMRI study in upper limb amputees.* Brain, 2001. **124**(11): p. 2268-2277.
212. Ruffino, C., C. Papaxanthis, and F. Lebon, *Neural plasticity during motor learning with motor imagery practice: Review and perspectives.* Neuroscience, 2017. **341**: p. 61-78.
213. Niazi, I.K., et al., *Detection of movement-related cortical potentials based on subject-independent training.* Medical & biological engineering & computing, 2013. **51**(5): p. 507-512.
214. de Lange, F.P., *Neural mechanisms of motor imagery*. 2008.
215. Mueller, S., et al., *Individual variability in functional connectivity architecture of the human brain.* Neuron, 2013. **77**(3): p. 586-595.
216. Honey, C.J., J.-P. Thivierge, and O. Sporns, *Can structure predict function in the human brain?* Neuroimage, 2010. **52**(3): p. 766-776.
217. Smith, S., et al., *Structural variability in the human brain reflects fine-grained functional architecture at the population level.* Journal of Neuroscience, 2019. **39**(31): p. 6136-6149.
218. Quinn, A.J., et al., *Task-evoked dynamic network analysis through hidden markov modeling.* Frontiers in neuroscience, 2018. **12**: p. 603.
219. Herzfeld, D.J. and R. Shadmehr, *Motor variability is not noise, but grist for the learning mill.* nature neuroscience, 2014. **17**(2): p. 149-150.
220. Betzel, R.F., et al., *The community structure of functional brain networks exhibits scale-specific patterns of inter-and intra-subject variability.* Neuroimage, 2019. **202**: p. 115990.
221. Edelman, B.J., et al., *Noninvasive neuroimaging enhances continuous neural tracking for robotic device control.* Science robotics, 2019. **4**(31).

222. Faller, J., et al., *Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task.* Proceedings of the National Academy of Sciences, 2019. **116**(13): p. 6482-6490.
223. Saha, S. and M. Baumert, *Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review.* Frontiers in Computational Neuroscience, 2019. **13**: p. 87.
224. Luck, S.J., *An introduction to the event-related potential technique*. 2014: MIT press.
225. Ostry, D.J. and P.L. Gribble, *Sensory plasticity in human motor learning.* Trends in neurosciences, 2016. **39**(2): p. 114-123.
226. Radford, A., et al., *Improving language understanding by generative pre-training.* 2018.
227. Bai, O., et al., *Prediction of human voluntary movement before it occurs.* Clinical Neurophysiology, 2011. **122**(2): p. 364-372.
228. McFarland, D.J., W.A. Sarnacki, and J.R. Wolpaw, *Electroencephalographic (EEG) control of three-dimensional movement.* Journal of neural engineering, 2010. **7**(3): p. 036007.
229. Suwannarat, A., S. Pan-ngum, and P. Israsena, *Comparison of EEG measurement of upper limb movement in motor imagery training system.* Biomedical engineering online, 2018. **17**(1): p. 103.
230. Ang, K.K., et al., *A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke.* Clinical EEG and neuroscience, 2015. **46**(4): p. 310-320.
231. Koo, B., et al., *A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery.* Journal of neuroscience methods, 2015. **244**: p. 26-32.
232. Tam, W.-K., et al., *A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2011. **19**(6): p. 617-627.
233. Seghier, M.L. and C.J. Price, *Interpreting and utilising intersubject variability in brain function.* Trends in Cognitive Sciences, 2018. **22**(6): p. 517-530.
234. Saha, S. and M. Baumert, *Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review.* Frontiers in computational neuroscience, 2020. **13**: p. 87.
235. Lee, D.-Y., et al. *Decoding movement imagination and execution from eeg signals using bci-transfer learning method based on relation network*. in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020. IEEE.
236. Wu, D., Y. Xu, and B.-L. Lu, *Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016.* IEEE Transactions on Cognitive and Developmental Systems, 2020.
237. Zhang, K., et al., *Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network.* Neural Networks, 2021. **136**: p. 1-10.
238. Fahimi, F., et al., *Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI.* Journal of neural engineering, 2019. **16**(2): p. 026007.

239. Suedfeld, P. and R.A. Bow, *Health and thxrapeutic applications of chamber and floatation restricted environmental stimulation thxrapy (REST).* Psychology & Health, 1999. **14**(3): p. 545-566.
240. Lilly, J.C., *The deep self: Profound relaxation and the tank isolation technique*. 1977: Simon & Schuster.
241. Borrie, R.A., *The use of restricted environmental stimulation therapy in treating addictive behaviors.* International journal of the addictions, 1991. **25**(sup7): p. 995-1015.
242. Jackson, C.W., *Sensory deprivation: Fifteen years of research*. 1969: Ardent Media.
243. Forgays, D.G. and D.K. Forgays, *Creativity enhancement through floatation isolation.* Journal of Environmental Psychology, 1992. **12**(4): p. 329-335.
244. Jonsson, K. and A. Kjellgren, *Promising effects of treatment with floatation-REST (restricted environmental stimulation technique) as an intervention for generalized anxiety disorder (GAD): a randomized controlled pilot trial.* BMC complementary and alternative medicine, 2016. **16**(1): p. 108.
245. Jonsson, K. and A. Kjellgren, *Characterizing the experiences of floatation-REST (Restricted Environmental Stimulation Technique) treatment for generalized anxiety disorder (GAD): A phenomenological study.* European Journal of Integrative Medicine, 2017. **12**: p. 53-59.
246. Kjellgren, A., et al., *Wellness through a comprehensive Yogic breathing program–A controlled pilot trial.* BMC Complementary and Alternative Medicine, 2007. **7**(1): p. 43.
247. Kjellgren, A., H. Buhrkall, and T. Norlander, *Psychotherapeutic Treatment in Combination with Relaxation in a Floatation Tank: Effects on" Burn-Out Syndrome".* Qualitative Report, 2010. **15**(5): p. 1243-1269.
248. Kjellgren, A., H. Buhrkall, and T. Norlander, *Preventing sick-leave for sufferers of high stress-load and burnout syndrome: A pilot study combining psychotherapy and the floatation tank.* International Journal of Psychology and Psychological Therapy, 2011. **11**(2): p. 297-306.
249. Kjellgren, A., et al., *Effects of floatation-REST on muscle tension pain.* Pain Research and Management, 2001. **6**(4): p. 181-189.
250. Kjellgren, A., et al., *Altered consciousness in floatation-REST and chamber-REST: Experience of experimental pain and subjective stress.* Social Behavior and Personality: an international journal, 2004. **32**(2): p. 103-115.
251. Kjellgren, A. and J. Westman, *Beneficial effects of treatment with sensory isolation in floatation-tank as a preventive health-care intervention–a randomized controlled pilot trial.* BMC complementary and alternative medicine, 2014. **14**(1): p. 417.
252. Carder, K.J.S., *What is the Lived Experience of Floatation-REST Inside a Tank?* 2018, Alliant International University.
253. Suedfeld, P., E.J. Ballard, and M. Murphy, *Water immersion and floatation: From stress experiment to stress treatment.* Journal of Environmental Psychology, 1983. **3**(2): p. 147-155.
254. Van Dierendonck, D. and J. Te Nijenhuis, *Floatation restricted environmental stimulation therapy (REST) as a stress-management tool: A meta-analysis.* Psychology & Health, 2005. **20**(3): p. 405-412.
255. Kjellgren, A., et al., *Does floatation-rest (restricted environmental stimulation technique) have an effect on sleep?* European Journal of Integrative Medicine, 2020. **33**: p. 101047.

256. Fine, T.H. and J.W. Turner Jr, *The effect of brief restricted environmental stimulation therapy in the treatment of essential hypertension.* Behaviour research and therapy, 1982. **20**(6): p. 567-570.
257. Turner, J.W. and T.H. Fine, *Effects of relaxation associated with brief restricted environmental stimulation therapy (REST) on plasma cortisol, ACTH, and LH.* Biofeedback and Self-regulation, 1983. **8**(1): p. 115-126.
258. Mehrabian, A. and J.A. Russell, *A measure of arousal seeking tendency.* Environment and Behavior, 1973. **5**(3): p. 315.
259. Berkun, M.M., et al., *Experimental studies of psychological stress in man.* Psychological Monographs: General and Applied, 1962. **76**(15): p. 1.
260. Miller, L.C., R. Murphy, and A.H. Buss, *Consciousness of body: Private and public.* Journal of personality and social psychology, 1981. **41**(2): p. 397.
261. Mehrabian, A. and J.A. Russell, *The basic emotional impact of environments.* Perceptual and motor skills, 1974. **38**(1): p. 283-301.
262. Russell, J.A. and G. Pratt, *A description of the affective quality attributed to environments.* Journal of personality and social psychology, 1980. **38**(2): p. 311.
263. Jacobs, G.D., R.L. Heilbronner, and J.M. Stanley, *The effects of short term floatation REST on relaxation: a controlled study.* Health Psychology, 1984. **3**(2): p. 99.
264. Suedfeld, P. and G. Baker-Brown, *Restricted environmental stimulation therapy of smoking: A parametric study.* Addictive behaviors, 1987. **12**(3): p. 263-267.
265. Suedfeld, P., J. Metcalfe, and S. Bluck, *Enhancement of scientific creativity by floatation REST (restricted environmental stimulation technique).* Journal of Environmental Psychology, 1987. **7**(3): p. 219-231.
266. Turner, J.W., et al., *The presence or absence of light during floatation restricted environmental stimulation: Effects on plasma cortisol, blood pressure, and mood.* Biofeedback and Self-regulation, 1989. **14**(4): p. 291-300.
267. Suedfeld, P. and T. Bruno, *Floatation REST and imagery in the improvement of athletic performance.* Journal of Sport and Exercise Psychology, 1990. **12**(1): p. 82-85.
268. McAleney, P.J., A. Barabasz, and M. Barabasz, *Effects of floatation restricted environmental stimulation on intercollegiate tennis performance.* Perceptual and Motor Skills, 1990. **71**(3): p. 1023-1028.
269. Turner Jr, J.W. and T.H. Fine, *Restricting environmental stimulation influences levels and variability of plasma cortisol.* Journal of Applied Physiology, 1991. **70**(5): p. 2010-2013.
270. Wagaman, J.D., A.F. Barabasz, and M. Barabasz, *Floatation REST and imagery in the improvement of collegiate basketball performance.* Perceptual and Motor Skills, 1991. **72**(1): p. 119-122.
271. Lee, A. and J. Hewitt, *USING VISUAL-IMAGERY IN A FLOATATION TANK TO IMPROVE GYMNASTIC PERFORMANCE AND REDUCE PHYSICAL SYMPTOMS.* International Journal of Sport Psychology, 1987. **18**(3): p. 223-230.
272. Suedfeld, P., D.E. Collier, and B.D. Hartnett, *Enhancing perceptual-motor accuracy through floatation REST.* The Sport Psychologist, 1993. **7**(2): p. 151-159.
273. Schulz, P. and C.-H. Kaspar, *Neuroendocrine and psychological effects of restricted environmental stimulation technique in a floatation tank.* Biological psychology, 1994. **37**(2): p. 161-175.

274. Hathaway, S.R. and J.C. McKinley, *Minnesota Multiphasic Personality Inventory; Manual, revised.* 1951.
275. Zuckerman, M. and M. Neeb, *Sensation seeking and psychopathology.* Psychiatry research, 1979. **1**(3): p. 255-264.
276. Schulz, P., et al., *Traduction française de la Stanford Sleepiness Scale (SSS) et utilisation de cette échelle de sédation après dose unique de midazolam ou d'amitriptylline.* Agressologie (Paris), 1983. **24**(8): p. 357-359.
277. von Zerssen, D., *Clinical self-rating scales (CSRS) of the Munich psychiatric information system (PSYCHIS München)*, in *Assessment of depression*. 1986, Springer. p. 270-303.
278. Norlander, T., H. Bergman, and T. Archer, *Effects of floatation REST on creative problem solving and originality.* Journal of Environmental Psychology, 1998. **18**(4): p. 399-408.
279. Holmquist, R., *Manual till FS: Förändring och stabilitet [Manual for the FS Test: Change and stability].* Stockholm: Psykologiförlaget AB, 1986.
280. Holmquist, R., *Testbeskrivningar för Syllogismer I och II.[Manual to the syllogisms I and II test]*. 1974, Stockholm: Psykologiförlaget.
281. Holmquist, R., *Testbeskrivning för FREGO [Manual for the FREGO test].* Stockholm: Psykologiförlaget AB, 1973.
282. Forgays, D.G., *Floatation REST as a smoking intervention.* Addictive behaviors, 1987. **12**(1): p. 85-90.
283. Norlander, T., A. Kjellgren, and T. Archer, *Effects of floatation-versus chamber-restricted environmental stimulation technique (REST) on creativity and realism under stress and non-stress conditions.* Imagination, Cognition and Personality, 2003. **22**(4): p. 343-359.
284. Bood, S.Å., et al., *Eliciting the relaxation response with the help of floatation-rest (restricted environmental stimulation technique) in patients with stress-related ailments.* International Journal of Stress Management, 2006. **13**(2): p. 154.
285. Broderick, V., L. Uiga, and M. Driller, *Floatation-restricted environmental stimulation therapy improves sleep and performance recovery in athletes.* Performance Enhancement & Health, 2019. **7**(1-2): p. 100149.
286. Wallbaum, A.B., et al., *Progressive muscle relaxation and restricted environmental stimulation therapy for chronic tension headache: a pilot study.* International Journal of Psychosomatics, 1991.
287. Blanchard, E.B., et al., *The efficacy and cost-effectiveness of minimal-therapist-contact, non-drug treatments of chronic migraine and tension headache.* Headache: The Journal of Head and Face Pain, 1985. **25**(4): p. 214-220.
288. Kjellgren, A., et al., *Effects of floatation-REST on muscle tension pain.* Pain Research and Management, 2001. **6**.
289. Bood, S.A., et al., *Effects of floatation-restricted environmental stimulation technique on stress-related muscle pain: What makes the difference in therapy-attention-placebo or the relaxation response?* Pain Research and Management, 2005. **10**.
290. Bood, S.Å., et al., *Effects of floatation rest (restricted environmental stimulation technique) on stress related muscle pain: are 33 floatation sessions more effective than 12 sessions?* Social Behavior and Personality: an international journal, 2007. **35**(2): p. 143-156.

291. Edebol, H., S. Åke Bood, and T. Norlander, *Chronic whiplash-associated disorders and their treatment using floatation-REST (Restricted Environmental Stimulation Technique).* Qualitative Health Research, 2008. **18**(4): p. 480-488.
292. Bood, S.Å., A. Kjellgren, and T. Norlander, *Treating stress-related pain with the floatation restricted environmental stimulation technique: Are there differences between women and men?* Pain Research and Management, 2009. **14**.
293. Khalsa, S., et al., *S66. A Clinical Trial Investigating the Safety and Tolerability of Floatation-Rest in Anorexia Nervosa.* Biological Psychiatry, 2018. **83**(9): p. S372.
294. Feinstein, J., et al., *F34. Examining the Short-Term Anxiolytic Effect of Floatation-REST.* Biological Psychiatry, 2018. **83**(9): p. S250-S251.
295. Feinstein, J.S., et al., *The elicitation of relaxation and interoceptive awareness using floatation therapy in individuals with high anxiety sensitivity.* Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2018. **3**(6): p. 555-562.
296. Edebol, H., T. Nordén, and T. Norlander, *Behavior change and pain relief in chronic whiplash associated disorder Grade IV using floatation restricted environmental stimulation technique: A case report.* Psychology and Behavioral Sciences, 2013. **2**: p. 206-2016.
297. Raab, J. and J. Gruzelier, *A controlled investigation of right hemispheric processing enhancement after restricted environmental stimulation (REST) with floatation.* Psychological Medicine, 1994. **24**(2): p. 457-462.
298. Sakata, S., et al., *Enhancement of randomness by floatation rest (restricted environmental stimulation technique).* Perceptual and motor skills, 1995. **80**(3): p. 999-1010.
299. Suedfeld, P. and E. Eich, *Autobiographical memory and affect under conditions of reduced environmental stimulation.* Journal of Environmental Psychology, 1995. **15**(4): p. 321-326.
300. Åsenlöf, K., et al., *Case studies on fibromyalgia and burn-out depression using psychotherapy in combination with floatation-REST: Personality development and increased well-being.* Imagination, Cognition and Personality, 2007. **26**(3): p. 259-271.
301. Driller, M.W. and C.K. Argus, *Floatation restricted environmental stimulation therapy and napping on mood state and muscle soreness in elite athletes: a novel recovery strategy?* Performance Enhancement & Health, 2016. **5**(2): p. 60-65.
302. Shurley, J.T., *Profound experimental sensory isolation.* American Journal of Psychiatry, 1960. **117**(6): p. 539-545.
303. Lilly, J.C. and J.T. Shurley, *Experiments in solitude, in maximum achievable physical isolation with water suspension, of intact healthy persons.* Psychophysiological aspects of space flight, 1961: p. 238-247.
304. Norlander, T., A. Kjellgren, and T. Archer, *The experience of floatation-REST as a function of setting and previous experience of altered state of consciousness.* Imagination, Cognition and Personality, 2000. **20**(2): p. 161-178.
305. Kjellgren, A., F. Lyden, and T. Norlander, *Sensory isolation in floatation tanks: Altered states of consciousness and effects on well-being.* The Qualitative Report, 2008. **13**(4): p. 636-656.
306. Dunham, C.M., J.V. McClain, and A. Burger, *Comparison of Bispectral Index™ values during the floatation restricted environmental stimulation technique and results for stage I sleep: a prospective pilot investigation.* BMC research notes, 2017. **10**(1): p. 1-6.

307. Kjellgren, A., et al., *Quality of life with floatation therapy for a person diagnosed with attention deficit disorder, atypical autism, PTSD, anxiety and depression.* Open Journal of Medical Psychology, 2013. **2**(3): p. 134-138.
308. Soulexdc.com. *Can children benefit from floating*. [cited 2021 7/11/2021]; Available from: https://soulexdc.com/blog/2017/3/15/can-children-benefit-from-floating.
309. Zenfloatco.com. *Chidren and floating*. 2013 [cited 2021 7/11/2021]; Available from: https://zenfloatco.com/blog-pages/2014/5/30/floating-and-children.
310. Lilly, J.C., *Mental effects of reduction of ordinary levels of physical stimuli on intact, healthy persons.* Psychiatric Research Reports, 1956.
311. Ballard, E., *REST in the treatment of persistent psychophysiological insomnia*, in *Clinical and experimental restricted environmental stimulation*. 1993, Springer. p. 187-203.
312. Bood, S.Å., A. Kjellgren, and T. Norlander, *Treating stress-related pain with the floatation restricted environmental stimulation technique: Are there differences between women and men?* Pain Research and Management, 2009. **14**(4): p. 293-298.
313. Suedfeld, P., *The Restricted Environmental Stimulation Technique in the modification of addictive behaviors: Through the centuries to frontiers for the Eighties.* Bulletin of the Society of Psychologists in addictive Behaviors, 1983.
314. Barabasz, A. and M. Barabasz, *Clinical and experimental restricted environmental stimulation*. 1993: Springer.
315. Jalaudin, N. and M.K.M. Amin, *Electroencephalography (EEG) analysis on human reflection towards relaxation of mind.* 2019.
316. Davidson, R.J., *Cerebral asymmetry and emotion: Conceptual and methodological conundrums.* Cognition & Emotion, 1993. **7**(1): p. 115-138.
317. Craig, A., *Forebrain emotional asymmetry: a neuroanatomical basis?* Trends in cognitive sciences, 2005. **9**(12): p. 566-571.
318. Goodman, R.N., et al., *Stress, emotion regulation and cognitive performance: The predictive contributions of trait and state relative frontal EEG alpha asymmetry.* International Journal of Psychophysiology, 2013. **87**(2): p. 115-123.
319. Alonso, J., et al., *Stress assessment based on EEG univariate features and functional connectivity measures.* Physiological measurement, 2015. **36**(7): p. 1351.
320. Kamei, T., et al., *Decrease in serum cortisol during yoga exercise is correlated with alpha wave activation.* Perceptual and motor skills, 2000. **90**(3): p. 1027-1032.
321. Yang, Q., et al. *Cortical synchrony change under mental stress due to time pressure*. in *2010 3rd International Conference on Biomedical Engineering and Informatics*. 2010. IEEE.
322. Strelets, V., Z.V. Garakh, and V.Y. Novototskii-Vlasov, *Comparative study of the gamma rhythm in normal conditions, during examination stress, and in patients with first depressive episode.* Neuroscience and behavioral physiology, 2007. **37**(4): p. 387-394.
323. Hinrikus, H., et al., *Electroencephalographic spectral asymmetry index for detection of depression.* Medical & biological engineering & computing, 2009. **47**(12): p. 1291.
324. Lewis, R.S., N.Y. Weekes, and T.H. Wang, *The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health.* Biological psychology, 2007. **75**(3): p. 239-247.
325. Ferreira, C., et al., *The relation between EEG prefrontal asymmetry and subjective feelings of mood following 24 hours of sleep deprivation.* Arquivos de Neuro-psiquiatria, 2006. **64**(2B): p. 382-387.

326. McCraty, R. and F. Shaffer, *Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk.* Global advances in health and medicine, 2015. **4**(1): p. 46-61.
327. Gevirtz, R., P. Lehrer, and M. Schwartz, *Cardiorespiratory biofeedback.* Biofeedback: A practitioner's guide, 2016: p. 196-213.
328. Shaffer, F. and J. Ginsberg, *An overview of heart rate variability metrics and norms.* Frontiers in public health, 2017. **5**: p. 258.
329. Nunan, D., G.R. Sandercock, and D.A. Brodie, *A quantitative systematic review of normal values for short-term heart rate variability in healthy adults.* Pacing and clinical electrophysiology, 2010. **33**(11): p. 1407-1417.
330. Karemaker, J.M., *Counterpoint: respiratory sinus arrhythmia is due to the baroreflex mechanism.* Journal of applied physiology, 2009. **106**(5): p. 1742-1743.
331. Gregory, J., et al., *Measuring neural time series data in a sensory deprivation tank.* 2019.
332. Pan, J. and W.J. Tompkins, *A real-time QRS detection algorithm.* IEEE transactions on biomedical engineering, 1985(3): p. 230-236.
333. Sedghamiz, H., *Matlab implementation of Pan Tompkins ECG QRS detector.* DataCite: Cambridge, UK, 2014.
334. Sedghamiz, H., *BioSigKit: A Matlab Toolbox and Interface for Analysis of BioSignals.* Journal of Open Source Software, 2018. **3**(30): p. 671.
335. Naji, M., et al., *Coupling of autonomic and central events during sleep benefits declarative memory consolidation.* Neurobiology of learning and memory, 2019. **157**: p. 139-150.
336. Billman, G.E., *The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance.* Frontiers in physiology, 2013. **4**: p. 26.
337. Camm, A.J., et al., *Heart rate variability. Standards of measurement, physiological interpretation, and clinical use.* 1996.
338. Shaffer, F., R. McCraty, and C.L. Zerr, *A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability.* Frontiers in psychology, 2014. **5**: p. 1040.
339. Kleiger, R.E., P.K. Stein, and J.T. Bigger Jr, *Heart rate variability: measurement and clinical utility.* Annals of Noninvasive Electrocardiology, 2005. **10**(1): p. 88-101.
340. Pinheiro, J. and D. Bates, *Mixed-effects models in S and S-PLUS*. 2006: Springer Science & Business Media.
341. Brown, R.P. and P.L. Gerbarg, *Sudarshan Kriya yogic breathing in the treatment of stress, anxiety, and depression: part I—neurophysiologic model.* Journal of Alternative & Complementary Medicine, 2005. **11**(1): p. 189-201.
342. Lifshitz, M., et al., *Source localization of brain states associated with canonical neuroimaging postures.* Journal of Cognitive Neuroscience, 2017. **29**(7): p. 1292-1301.
343. Raz, A., et al., *Ecological nuances in functional magnetic resonance imaging (fMRI): psychological stressors, posture, and hydrostatics.* Neuroimage, 2005. **25**(1): p. 1-7.
344. Thibault, R.T., et al., *Posture alters human resting-state.* cortex, 2014. **58**: p. 199-205.
345. Yamashita, K., et al., *The state anxiety inventory is useful for predicting the autonomic nervous system state of patients before the extraction of an impacted mandibular third molar.* Journal of Oral and Maxillofacial Surgery, 2020. **78**(4): p. 538-544.

346. Dedovic, K., et al., *The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain.* Journal of Psychiatry and Neuroscience, 2005. **30**(5): p. 319.
347. Sulaiman, N., et al. *Initial investigation of human physical stress level using brainwaves*. in *2009 IEEE Student Conference on Research and Development (SCOReD)*. 2009. IEEE.
348. Steinhubl, S.R., et al., *Cardiovascular and nervous system changes during meditation.* Frontiers in Human Neuroscience, 2015. **9**: p. 145.
349. Davidson, R.J., et al., *Alterations in brain and immune function produced by mindfulness meditation.* Psychosomatic medicine, 2003. **65**(4): p. 564-570.
350. Minguillon, J., M.A. Lopez-Gordo, and F. Pelayo, *Stress assessment by prefrontal relative gamma.* Frontiers in computational neuroscience, 2016. **10**: p. 101.
351. Yair, O., et al., *Spectral Discovery of Jointly Smooth Features for Multimodal Data.* arXiv preprint arXiv:2004.04386, 2020.
352. Li, B., Y.-R. Li, and X.-L. Zhang, *A survey on Laplacian eigenmaps based manifold learning methods.* Neurocomputing, 2019. **335**: p. 336-351.
353. Bengio, Y., et al. *Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering*. in *Advances in neural information processing systems*. 2004.
354. Quispe, A.M., C. Petitjean, and L. Heutte, *Extreme learning machine for out-of-sample extension in Laplacian eigenmaps.* Pattern Recognition Letters, 2016. **74**: p. 68-73.

Appendix A. Dimensionality Reduction for Classification of Object Weight from Electromyography Supplementary Material

In this supplementary material we give further details about the dimensionality-reduction methods we used. Much of this material follows Belkin et al. **[25]**.

A.1 Optimal embeddings

Given a data set, we construct a weighted graph G = (V, E) with edges connecting nearby points to each other (assuming the graph is connected). Consider the problem of mapping the weighted graph G to a line so that connected points stay as close together as possible. Let y = $(y_1, y_2, \dots, y_n)^T$ be such a map. A reasonable criterion for having good mapping is to minimize the following objective function:

$$\sum_{ij} (y_i - y_j)^2 W_{ij} \tag{1}$$

under appropriate constraints. The objective function with our choice of weights W_{ij} incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart. Therefore, minimizing it is an attempt to ensure that if x_i and x_j are "close," then y_i and y_j are close as well. It turns out that for any y, we have:

$$\frac{1}{2}\sum_{ij}\left(y_i - y_j\right)^2 W_{ij} = y^T L y \tag{2}$$

Where as $L = D - W$ and W_{ij} is symmetric and $D_{ii} = \sum_j W_{ji}$. Thus,

$$\sum_{ij}\left(y_i - y_j\right)^2 W_{ij} = \sum_{ij}({y_i}^2 + {y_j}^2 - 2y_i y_j)W_{ij} = \sum_i {y_i}^2 D_{ii} + \sum_j {y_j}^2 D_{jj} - 2\sum_{ij} y_i y_j W_{ij} = 2y^T L y \tag{3}$$

Note that this calculation also shows that L is positive semidefinite. Therefore, the minimization problem reduces to finding: $\operatorname{argmin} y^T L y \ \text{ s.t. } \ y^T D y = 1$

The constraint $y^T D y = 1$ removes an arbitrary scaling factor in the embedding. Matrix D provides a natural measure on the vertices of the graph. The bigger the value D_{ii} (corresponding to the ith vertex) is, the more important is that vertex. Because L is positive semidefinite, the vector y that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem: $Ly = \lambda D y$

Let **1** be the constant function taking 1 at each vertex. It is easy to see that **1** is an eigenvector with eigenvalue 0. If the graph is connected, **1** is the only eigenvector for λ = 0. To eliminate this trivial solution, which collapses all vertices of G onto the real number 1, we put an additional constraint of orthogonality and look for $\operatorname{argmin} y^T L y \ s.t. \ y^T D y = 1$ and $y^T D \mathbf{1} = 0$. Thus, finally, the solution is now given by the eigenvector with the smallest nonzero eigenvalue. The condition $y^T D \mathbf{1} = 0$ can be interpreted as removing a translation invariance in y.

Now consider the more general problem of embedding the graph into an m-dimensional Euclidean space. The embedding is given by the $k \times m$ matrix $\Upsilon = [y_1, y_2, \dots, y_m]$, where the i^{th} row provides the embedding coordinates of the ith vertex. Similarly, we need to minimize:

$$\sum_{ij} \left\| y^{(i)} - y^{(j)} \right\|^2 W_{ij} = tr(\Upsilon^T L \Upsilon) \tag{4}$$

where $y^{(i)} = [y_1(i), \dots, y_m(i)]^T$ is the m-dimensional representation of the i^{th} vertex. This reduces to finding:

$$\text{argmin } tr(\Upsilon^T L \Upsilon) \text{ s.t. } \Upsilon^T L \Upsilon = 1 \tag{5}$$

For the one-dimensional embedding problem, the constraint prevents collapse onto a point. For the m-dimensional embedding problem, the constraint presented above prevents collapse onto a subspace of dimension less than $m - 1$.

The above therefore suggest that the Laplacian-Eigenmap algorithm keeps samples from the original, higher-dimensional space close to each other also in the lower-dimensional embedding. The Laplacian graph is analogous to the Laplace-Beltrami operator on manifolds. The eigenfunctions of the Laplace Beltrami operator have properties desirable for embedding [25, 351]. Let $\mathcal{M}_x \subseteq \mathbb{R}^d$ be a smooth, compact manifold embedded in a d-dimensional Euclidean space. A function $f: \mathcal{M}_x \to \mathbb{R}^d$ is said to be smooth if $f \in C^\infty(\mathcal{M}_x, \mathbb{R})$, that is, the function f and all of its derivatives are continuous. Let us define a different notion of smoothness related to the Laplace-Beltrami operator. The Laplace-Beltrami operator L_x is a linear operator generalizing the Laplacian on Euclidean spaces to Riemannian manifolds. The eigenfunctions ψ_i of the Laplace-Beltrami operator span a dense subset of the function space $H^0 = L^2(\mathcal{M}_x, \mathbb{R})$. The eigenvalues of the Laplace-Beltrami are real (and non-negative), so we can sort the associated eigenfunctions ψ_i such that $\lambda_i \leq \lambda_j$ for $i < j$. We say that ψ_i is smoother than ψ_j if $\lambda_i \leq \lambda_j$. It was shown that the best representation basis, in terms of truncated representation of functions $f: \mathcal{M}_x \rightarrow \mathbb{R}^d$ such that $\| \nabla f \| \leq 1$, are in fact the eigenfunctions of the Laplace-Beltrami operator L_x. Thus, in that sense, we say that $f_i : \mathcal{M}_x \rightarrow \mathbb{R}$ with $\| f_i \| = 1$ is smoother than $f_j : \mathcal{M}_x \rightarrow \mathbb{R}$ with $\| f_j \| = 1$ if $\| L_x f_i \| \leq \| L_x f_j \|$.

Heat kernels and the choice of weight matrix: The Laplace Beltrami operator on differentiable functions on a manifold $\mathcal{M}$ is intimately related to heat flow. Let $f : M \rightarrow \mathbb{R}$ be the initial heat distribution and $u(x,t)$ be the heat distribution at time t $\left(u(x,0) = f(x)\right)$ (see [25, 351] for more details). The results show that we compute the graph Laplacian with the following weights:

$$W_{ij} = \begin{cases} e^{\frac{\|x_i - x_j\|^2}{4t}} & if\ \|x_i - x_j\| < \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

A.2 Properties of techniques for dimensionality reduction

In S1 Table 1, the dimensionality reduction techniques are listed by four general properties: (1) whether the mapping between the high-dimensional and the low-dimensional space is parametric, (2) the main free parameters to be optimized, (3) the computational complexity of the main computational part of the technique, and (4) the memory complexity of the technique [38, 39]. S1 Table 1 shows that most techniques for dimensionality reduction are non-parametric. This means that the technique does not specify a direct mapping from the high-dimensional to the low-dimensional space (or vice versa). The non-parametric nature of most techniques is a disadvantage for two main reasons: (1) it is not possible to generalize the mapping to a held-out or to a new test set (without carrying out the dimensionality-reduction technique again); (2) it is difficult to obtain insights into how much information of the high-dimensional data was preserved in the low-dimensional space by reconstructing the original data from its low-dimensional representation and measuring the error between the reconstructed and original data.

S1 Table 1. Properties of techniques for dimensionality reduction

Dimensionality reduction technique	Parametric	Free parameters	Computational complexity	Memory complexity
PCA	Yes	none	$O(D^3)$	$O(D^2)$
ISOMAP	No	k	$O(n^3)$	$O(n^2)$
LLE	No	k	$O(pn^2)$	$O(pn^2)$
Laplacian Eigenmaps	No	k, σ	$O(pn^2)$	$O(pn^2)$
t-SNE	No	perplexity	$O(n^2)$	$O(n^2)$

As for the free parameters, S1 Table 1 shows that the objective functions of most non-linear techniques for dimensionality reduction have free parameters that need to be optimized. In other words, there are parameters that directly influence the optimized cost function. Non-convex techniques for dimensionality reduction have additional free parameters, such as the learning rate and the permitted maximum number of iterations. Moreover, LLE uses a regularization parameter in the computation of the reconstruction weights.

The presence of free parameters has both advantages and disadvantages. The main advantage of free parameters is that they make the technique more flexible. However, they then need to be tuned to optimize performance. S1 Table 1 also provides more details on the computational and memory complexities of the techniques. The computational complexity of a dimensionality-reduction technique is important for its practical applicability. Algorithms grow increasingly infeasible as computational or memory demands rise. The computational complexity of a dimensionality reduction technique is determined by: (1) properties of the dataset, such as the number of datapoints n and their dimensionality D, and (2) by parameters of the techniques, such as the target dimensionality d, the number of nearest neighbors k, σ (for techniques based on neighborhood graphs). In S1 Table 1, p denotes the ratio of nonzero elements in a sparse matrix to the total number of elements.

In the next section, we show how we deal with out-of-sample extension where there is no explicit projection function between the original data and their low dimensional representations in the original LE algorithm.

A.3 Out-of-sample extension

An important requirement for dimensionality reduction techniques is the ability to embed new high-dimensional datapoints into an existing low-dimensional data representation. However, there is no explicit projection function between the original data and their low dimensional representations in the original LE algorithm, which makes out-of-sample extension difficult. To find projection of any additional samples, LE needs to be run on all the data together with the additional samples, resulting in considerable computational cost, especially when applying it to large scale data pattern recognition. Fortunately, various methods have been developed to mitigate the out-of-sample problem [352]: Linear approximation to LE, Kernel extensions to LE, Tensor representation of LE, incremental learning for LE, neural network approaches, and Extreme Learning Machine. The out-of-sample extension for spectral techniques has been presented in [353]. Nyström approximation supports out-of-sample extensions for spectral techniques such as

ISOMAP, LLE, and Laplacian Eigenmaps. In the next section, we explain Nyström approximation in greater detail.

A.3.1 Nyström extension

Let D denote the dimension of the initial set, N the number of samples (or points), x_i a sample in $X \in \mathbb{R}^D$ and the $D \times N$ training matrix containing the samples. Let y_i denote the coordinates in the embedded space, included in $\mathbb{R}^d$ where d is the reduced dimension that corresponds to x_i. Finally, let x_{N+1} denote a sample not belonging to the initial set of samples—i.e. an out-of-sample point. The goal is to estimate its reduced coordinates y_{N+1}. The Nyström method speeds up kernel-method computations by performing the eigen-decomposition on a subset of examples [354]. It was previously used to propose an out-of-sample extension to kernel-based spectral methods [353]. Let us recall the general framework in which spectral dimension-reduction techniques can be cast. Let W be a symmetric matrix of size $N \times N$, expressing the affinity between the N points of the training set. Let $K(\cdot,\cdot)$ denote a data-dependant kernel function giving rise to matrix W with $W_{ij} = K(x_i, x_j)$.

Let $(v_k,\ \lambda_k)$ denote the eigenvector and eigenvalue pairs such that $W v_k = \lambda_k\, v_k$. For dimensionality reduction, retain the d largest (or smallest, depending on the method) eigenvalues and their associated eigenvectors. The embedding (or reduced coordinates) of each training sample x_i is the ith row of a matrix U that contains the d eigenvectors in columns. The Nyström extension for an out-of-sample point is a weighted sum of the previously calculated eigenvectors and eigenvalues. More precisely, the kth reduced coordinate of the out-of- sample point is approximated as: $y_{N+1} = \frac{1}{\lambda_k} \sum_{i=1}^{N} v_{ki}\, K(x_{N+1}, x_i)$ for all $k = 1, \ldots, d$ or, in matrix form: $\hat{y}_{N+1} =$

$\frac{1}{\sqrt{\lambda}} U^T K_{N+1}$, where $\frac{1}{\sqrt{\lambda}} = diag\ (\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_d}})$. U is the matrix whose columns are the eigenvectors, and $K_{N+1} = [K(x_{N+1}, x_1) \dots K(x_{N+1}, X_N)]$. In [4], Bengio et al. have designed a formulation of $K\ (\cdot,\cdot)$ for Laplacian eigenmaps:

$$K(a,b) = \frac{1}{n} \frac{K(a,b)}{\sqrt{E_x[K(a,x)]E_{x'}[K(b,x')]}} \tag{7}$$

The Nyström extension is applicable to any technique that make use of a kernel function. This method requires some parameter choice for the kernel $K\ (\cdot,\cdot)$, usually made heuristically.

A1 Table 2. Statistical analysis Performance of different classifiers

Paired Samples T-Test

Classifiers			t	df	p	Mean Difference	SE Difference	Cohen's d
k-NN	-	RBF SVM	4.000	6	0.007	6.771	1.693	1.512
k-NN	-	Linear SVM	5.616	6	0.001	17.257	3.073	2.123
k-NN	-	Random Forest	2.928	6	0.026	5.671	1.937	1.107

Note. Student's t-test.

A1 Table 3. Statistical analysis Performance of different dimension reduction techniques

Paired Samples T-Test

			t	df	p	Cohen's d
LE (simple minded)	-	PCA	-5.613	11	< .001	-1.620
LE (simple minded)	-	ISOMAP	-3.721	11	0.003	-1.074
LE (simple minded)	-	LLE	2.214	11	0.049	0.639
LE (simple minded)	-	LE (rbf)	1.064	11	0.310	0.307

Paired Samples T-Test

			t	df	p	Cohen's d
LE (simple minded)	-	t-SNE	-7.844	11	< .001	-2.264
LE (rbf)	-	PCA	-5.158	11	< .001	-1.489
LE (rbf)	-	ISOMAP	-3.882	11	0.003	-1.121
LE (rbf)	-	LLE	1.161	11	0.270	0.335
LE (rbf)	-	t-SNE	-6.148	11	< .001	-1.775

Note. Student's t-test.

A1 Table 4. Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Classifiers	1087.127	3	362.376	15.490	< .001
Residual	421.096	18	23.394		

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Residual	125.647	6	20.941		

Note. Type III Sum of Squares

A1 Table 5. Post Hoc Comparisons - Classifiers

		Mean Difference	SE	t	Cohen's d	p_{holm}
Linear SVM	RBF SVM	-10.486	3.170	-3.307	-1.250	0.049
	Random Forest	-11.586	2.703	-4.286	-1.620	0.026
	k-NN	-17.257	3.073	-5.616	-2.123	0.008

A1 Table 5. Post Hoc Comparisons - Classifiers

		Mean Difference	SE	t	Cohen's d	p_{holm}
RBF SVM	Random Forest	-1.100	2.586	-0.425	-0.161	0.685
	k-NN	-6.771	1.693	-4.000	-1.512	0.028
Random Forest	k-NN	-5.671	1.937	-2.928	-1.107	0.053

Note. Cohen's d does not correct for multiple comparisons.

Note. Bonferroni adjusted confidence intervals.