universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Predicting Neurofeedback Learning Success - A Machine Learning Analysis

verfasst von | submitted by

Daniel Reiter BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Frank Scharnowski MSc
Mitbetreut von | Co-Supervisor:	Dipl.-Ing. Dr.techn. David Steyrl

English abstract

Neurofeedback is a technique that enables individuals to learn to regulate specific brain activities through real-time feedback from their own brain activities. It holds potential for improving cognitive functions and alleviating symptoms in various clinical conditions. However, a major challenge is the non-responder problem - substantial variability in individual learning success - which limits the efficacy and reliability of neurofeedback-based interventions. This thesis investigated what subject specific and study specific factors influence neurofeedback performance and improvement. A comprehensive dataset was compiled from 35 neurofeedback studies comprising 732 participants across clinical and healthy populations. Machine learning models were used to analyse the data and identify influential predictive features. Results showed that both neurofeedback performance [mean balanced accuracy = 61.67 %, SD = 3.90, per subject grouping] and improvement [56.28 %, SD = 3.51, per subject grouping] could be predicted significantly above chance. The most impactful features included age, sex, health status, length of neurofeedback run, and the existence of a pre-training transfer run. In an exploratory analysis incorporating study ID as an additional discrete feature, it emerged as the most impactful feature for neurofeedback improvement - suggesting that study specific factors, including potentially unmeasured ones, play a critical role. Although these features were significant in the analysis, the impact they had on the prediction was small. These findings underscore the complex interplay of subject specific and study specific factors in shaping neurofeedback performance and improvement. While predictive accuracy remains limited, this work highlights possible directions for optimizing future neurofeedback studies.

Deutscher Abstract

Neurofeedback ist eine Technik, mit der Menschen lernen können, bestimmte Gehirnaktivitäten durch Echtzeit-Feedback ihrer eigenen Gehirnaktivitäten zu regulieren. Es hat das Potenzial, kognitive Funktionen zu verbessern und Symptome verschiedener klinischer Erkrankungen zu lindern. Eine große Herausforderung ist jedoch das Non-Responder-Problem - eine erhebliche Variabilität im individuellen Lernerfolg -, das die Wirksamkeit und Zuverlässigkeit von Neurofeedback einschränkt. In dieser Arbeit wurde untersucht, welche subjekt- und studienspezifischen Faktoren Neurofeedbackleistung und -verbesserung beeinflussen. Es wurde ein Datensatz aus 35 Neurofeedback-Studien mit 732 Teilnehmenden aus klinischen und gesunden Populationen zusammengestellt. Machine Learning Modelle wurden eingesetzt, um die Daten zu analysieren und einflussreiche prädiktive Merkmale zu identifizieren. Die Ergebnisse zeigten, dass sowohl die Neurofeedback-Leistung [mean balanced accuracy = 61.67 %, SD = 3.90, per subject grouping] als auch die Verbesserung [56.28 %, SD = 3.51, per subject grouping] signifikant vorhergesagt werden konnten. Zu den einflussreichsten Merkmalen gehörten Alter, Geschlecht, Gesundheitszustand, Dauer des Neurofeedbacklaufs und das Vorhandensein eines Transferlaufs vor dem Training. In einer explorativen Analyse, in der die Studien-ID als zusätzliches diskretes Merkmal einbezogen wurde, erwies sie sich als das Merkmal mit dem größten Einfluss auf die Neurofeedbackverbesserung - was darauf hindeutet, dass studienspezifische Faktoren, einschließlich potenziell nicht gemessener Faktoren, eine entscheidende Rolle spielen. Obwohl die Merkmale in der Analyse signifikant waren, war ihr Einfluss auf die Vorhersage gering. Diese Ergebnisse unterstreichen das komplexe Zusammenspiel von subjekt- und studienspezifischen Faktoren bei der Gestaltung der Neurofeedback-Leistung und -Verbesserung. Obwohl die Vorhersagegenauigkeit begrenzt bleibt, zeigt diese Arbeit mögliche Wege zur Optimierung zukünftiger Neurofeedback-Studien auf.

Acknowledgements

Writing this master's thesis was a challenge I would not have been able to tackle alone – and I didn't. My friends, my family, and my colleagues at the university helped me in countless ways, from professional advice to calming my nerves when everything felt overwhelming. As I don't want to waste anyone's time reading a long list of names, I'll mention only a few who played a particularly important role in this journey.

First of all, I would like to thank Frank Scharnowski and David Steyrl, both of whom provided me with excellent advice and invaluable supervision throughout this thesis. Thank you for our conversations on equal footing, and for your patience in explaining complex concepts and methods. I am deeply grateful for the opportunity to write my thesis under your guidance.

The next person I would like to thank is my wonderful girlfriend, Kathi. Thank you for listening to me rant about topics you didn't know – always with a smile on your face. Your support was unwavering, and without you, I would not be who I am today.

Lastly, I would like to thank my parents and my siblings, who supported me every step of the way, even from afar.

Thank you to all of you – and to so many more!

Table of content

Introduction

Neurofeedback

In neurofeedback, an individual's brain activity is monitored and displayed to the individual in real time. This technique aims to promote conscious control over neural processes, which are linked with certain behaviours or disorders, offering a non-invasive approach to modulate brain function in a precise way (Hampson et al., 2020; Haugg et al., 2021; Sitaram et al., 2017). Behavioural changes observed as a result of individuals learning to modulate their own brain activity suggest that neurofeedback can act as a form of internal neural stimulation (Shibata et al., 2011). However, the rapid attempts to use neurofeedback for treatment in a clinical settings has raised many concerns, as the neural mechanisms and neuroplastic changes underlying neurofeedback have not been fully understood yet (Sitaram et al., 2017; Sulzer et al., 2013). Furthermore, the failure of some clinical trials to show treatment effects after promising preliminary studies shows how important it is to analyse and understand the underlying mechanisms (Lansbergen et al., 2011; Sitaram et al., 2017; Sulzer et al., 2013).

Experimental design

Neurofeedback studies usually follow the same design principles, even if they pursue different objectives. Sulzer et al. (2013) summarised the typical experimental framework in the five following steps:

1. Defining a physiological target and response either anatomically or through a functional localiser. This determines what region, network and/or physiological response should be trained.
2. Next, participants receive real-time feedback based on their brain activity in the targeted region. They are encouraged to employ mental strategies to influence this

activity. Training sessions can last from a few minutes to several hours and may extend over multiple days.

3. After a successful training follows the transfer to different settings. When individuals have achieved successful regulation of the activity of the physiological target, they will be tested by demonstrating whether they are still able to regulate the activity in a different setting or task with or without feedback or not.
4. To rule out alternative explanations such as placebo effects or general learning, studies include either control groups or within-subject controls.
5. Finally, researchers examine whether changes in neural regulation are accompanied by behavioural shifts, typically using a pre- and post-training comparison.

As an example for the neurofeedback of the physiological response and assessment of the individual's performance the illustration in *Figure 1* from Weiskopf et al. (2004) depicts the information flow in a neurofeedback experiment. Neurofeedback experiments follow the so called closed-loop design. In a closed-loop design, participants are instructed to use cognitive strategies in combination with feedback from the region of interest (ROI) with the goal of regulating their neural signal. While the participants try to regulate the neural signal, the corresponding brain activity is being recorded using a functional magnetic resonance imaging (fMRI) machine. The signal is then processed in real time, allowing for the computation of differences between regulation and baseline conditions. This processed data is used to generate feedback (can be visual or auditory), which is presented to the participant, thus completing the feedback loop.

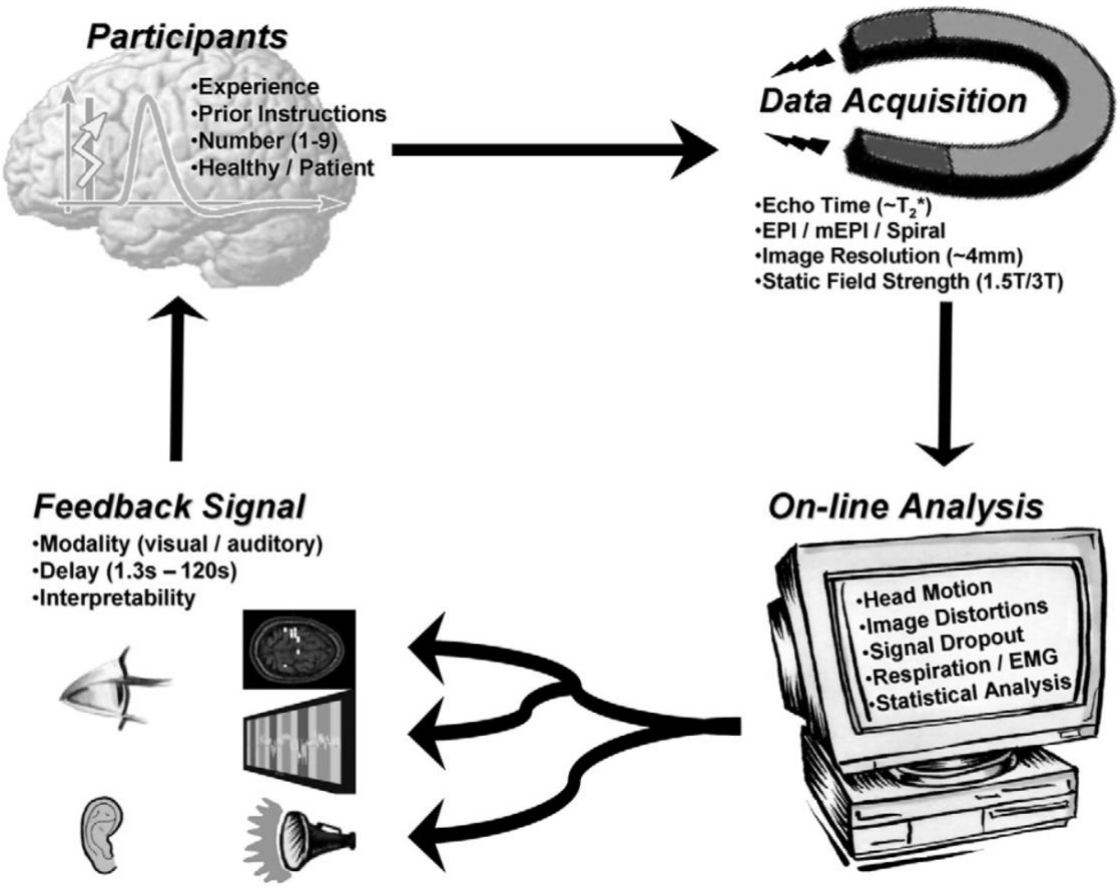

Figure 1 Closed-loop neurofeedback experimental design from Weiskopf et al. (2004).

A short history of neurofeedback

The first neurofeedback studies on humans were published in the 1960s and utilized EEG (electroencephalography) for the provided feedback (for a detailed history of the beginnings of neurofeedback please refer to Kamiya (2011)). These studies demonstrated that humans could control EEG signals themselves in real time. This line of development eventually gave rise to brain-machine interfaces – also referred to as brain-computer interfaces – where the focus shifts from modulating brain activity to enabling users to control external devices (Shanechi, 2019; Wolpaw, 2013). In recent years many neurofeedback studies utilizing fMRI have emerged – indicating a trend away from EEG-neurofeedback-studies, caused by the many advantages provided by fMRI. Although EEG has good temporal resolution, it cannot provide

the fMRIs whole brain coverage and millimetric spatial resolution (Sulzer et al., 2013; Thibault et al., 2018).

Functional Magnetic Resonance Imaging

Over the past 30 years, fMRI has revolutionized the study of the human brain by offering excellent spatial resolution, enabling whole-brain coverage and achieving a resolution of up to 0.6 mm^3 (Berman et al., 2021; Ogawa, Lee, Kay, et al., 1990; Ogawa, Lee, Nayak, et al., 1990; Weiskopf, 2012). The fMRI achieves this by detecting changes in the magnetic properties of oxygenated blood, which influence the intensity of the measured signal and allow researchers to infer neural activity in specific brain regions. This phenomenon is known as the blood-oxygen-level-dependent (BOLD) contrast, which serves as an indirect measure of neural activation (S.-G. Kim & Ogawa, 2012; Logothetis et al., 2001; Logothetis, 2008; Ogawa, Lee, Kay, et al., 1990; Ogawa, Lee, Nayak, et al., 1990). The BOLD response is based on the observation that increased neural activity leads to a localized increase in cerebral blood flow. While active brain regions use more oxygen, the accompanying vascular response supplies an excess of oxygenated blood, leading to a net increase in local blood oxygen levels. However, this vascular response is not immediate, it typically takes about five seconds to develop, a delay referred to as the haemodynamic response function. The fMRI signal captures this delayed change as the BOLD signal.

The BOLD signal differentiates neural activity levels by relying on the differing magnetic characteristics of oxygenated and deoxygenated haemoglobin (Hb) found in red blood cells. Oxygenated haemoglobin is diamagnetic, meaning it has weak negative magnetic susceptibility, which slows the dephasing of proton spins and contributes to an increase in the BOLD signal – an effect that can be linked indirectly to heightened neural activity in the corresponding brain region. On the other hand, deoxygenated haemoglobin is paramagnetic, adding positively to the local magnetic field and accelerating proton spin dephasing, which

leads to a decrease in the BOLD signal (S.-G. Kim & Ogawa, 2012; Mark et al., 2015; Shmuel et al., 2007).

Real-time functional Magnetic Resonance Imaging neurofeedback

The first study using real-time functional magnetic resonance imaging (rt-fMRI) was published in 1995 and defined rt-fMRI as a process of near simultaneous acquisition, processing, analysis and displaying of functional information from a fMRI machine (Cox et al., 1995). Current use-cases for rt-fMRI are surgical guidance (Hirsch et al., 2000), brain-computer interfaces (Shanechi, 2019; Sorger et al., 2012; Yoo et al., 2004) and neurofeedback. Similar to other neurofeedback methods, real-time fMRI neurofeedback is a non-invasive approach that allows individuals to intentionally modulate their brain activity, while benefiting from the high spatial resolution and other advantages of rt-fMRI (Haugg et al., 2021). Furthermore, it incorporates stricter control conditions (e.g., sham feedback from an unrelated brain signal) and measures not only the advancement in learning to control the BOLD signal, but also behavioural responses. It is important to distinguish these two outcome categories, as the assumption, that a change in the BOLD signal would inevitably cause a change in behaviour, would be false, as transfer effects do not happen reliably (Thibault et al., 2018). Additionally, rt-fMRI is not real time, like the name suggests, but shows a delay in its feedback. This delay arises from two main factors. First, the delay stems from haemodynamic coupling, which causes a temporal gap between neural activation and the resulting changes in the BOLD signal (S.-G. Kim & Bandettini, 2010). Typically, the signal begins to rise around 3 seconds after activation, with the peak occurring approximately 6 seconds later, though this can vary depending on the region of interest and the nature of the task (Sulzer et al., 2013; Weiskopf et al., 2004). Second, the time required for data preprocessing and statistical analysis contributes to the lag.

Use cases of real-time functional Magnetic Resonance Imaging neurofeedback

Many rt-fMRI neurofeedback studies have provided evidence for behavioural changes in healthy individuals and clinical improvement in the clinical population. More specifically rt-fMRI neurofeedback has been linked to improvements in visual perception (Scharnowski et al., 2012), speech performance (Rota et al., 2009), motor performance (Scharnowski et al., 2015), motivation (Zhi et al., 2018), memory (Scharnowski et al., 2015), emotion regulation (Koush et al., 2015; Paret & Hendler, 2020; Zich et al., 2020), attention (deBettencourt et al., 2015; Pamplona et al., 2020a), visual sensitivity (Shibata et al., 2011), colour perception (Amano et al., 2016), fear memory (Koizumi et al., 2016; Taschereau-Dumouchel et al., 2018), perceptual confidence (Cortese et al., 2016, 2017) and facial preference (Shibata et al., 2016) in healthy individuals. In the clinical population, researchers were able to provide evidence for the normalization of pathological neural characteristics and improvement of clinical measures in a multitude of disorders, including, but not limited to anxiety (Morgenroth et al., 2020), borderline personality disorder (Paret et al., 2016), depression (Young et al., 2017), post-traumatic stress disorder (Nicholson et al., 2017a), schizophrenia (Bauer et al., 2020), phobia (Zilverstand et al., 2015), nicotine addiction (Hanlon et al., 2013), Tourette syndrome (Sukhodolsky et al., 2020), chronic pain (deCharms et al., 2005), Huntington's disease (Papoutsi et al., 2019), obesity (Kohl et al., 2019), Parkinson's disease (Buyukturkoglu et al., 2013), stroke rehabilitation (Mehler et al., 2020), tinnitus (Emmert et al., 2017) and visuo-spatial neglect (Robineau et al., 2019).

The non-responder problem

Not all individuals – including both healthy, as well as individuals from the clinical population – show improvements in the previously mentioned fields (Haugg et al., 2020, 2021; Kadosh & Staunton, 2019; Sitaram et al., 2017; Sulzer et al., 2013; Thibault et al., 2018). The individuals who were not successful in regulating their brain signals were termed as non-

responders. However, there is no unified definition of a non-responder, and the definition varies between studies. In a study from Haugg et al. (2020), the authors tried to predict the neurofeedback learning success utilizing pretraining brain activity across 24 independent studies and estimated the non-responder rate to be about 38%. Additionally, even among participants who succeeded in modulating their brain signals, there was considerable variability in how effectively they regulated their neural activity (Haugg et al., 2020). This phenomenon reflects what is commonly referred to as the 'neurofeedback inefficacy problem', a term used to describe the significant differences in learning outcomes frequently observed in EEG-based neurofeedback research (Alkoby et al., 2018).

Improving the neurofeedback learning success rate

The design of neurofeedback studies might be a key factor for improving the rate of learning success of participants (Haugg et al., 2021). For example, two studies found that an intermittent presentation of feedback to the participants exhibited a better regulation performance of the brain signal compared to continuous presentation (Hellrung, Dietrich, et al., 2018; Johnson et al., 2012). However, a third study observed the same effect following a single neurofeedback session, but this effect did not persist across multiple sessions (Emmert et al., 2017). Another interesting aspect of neurofeedback studies is the influence of activity-based and connectivity-based feedback, although no significant difference in the performances of participants between the two approaches has been found (Papoutsi et al., 2019). The difference between these two approaches is that activity-based feedback focuses on a single brain area for training, whereas connectivity-based feedback focuses on changes in the connectivity between brain areas (Koush et al., 2013). A combination of both approaches appears to be promising, as one study reported enhanced effectiveness when connectivity-based feedback was integrated with activity-based neurofeedback (Kim et al., 2015).

In recent years, a third approach has emerged - the so-called decoded neurofeedback (DecNef). This approach combines fMRI neurofeedback with machine learning, which achieves a more fine-grained rendering of the manipulation of dynamics and representations in the brain (LaConte et al., 2007; Shibata et al., 2011). In contrast to univariate approaches, which assess overall activity within an ROI by analysing each voxel independently, DecNef employs multivoxel pattern analysis (MVPA), using machine learning to decode the information embedded in the distributed patterns of voxel activation. MVPA is utilised to achieve a high target specificity, referring to the accurate detection and modulation of the intended brain signal while avoiding non-target brain signals (Lubianiker et al., 2019; Shibata et al., 2019; Watanabe et al., 2017). This is possible because MVPA is able to detect and learn to decode information distribution in patterns of activity (Kamitani & Tong, 2005; Norman et al., 2006).

Furthermore, target neural representations can be inferred indirectly from different participants using a technique called hyperalignment (Haxby et al., 2011; Taschereau-Dumouchel et al., 2021). Hyperalignment is a functional alignment approach, which creates a high-dimensional common space by transferring patterns of neural activity across participants with a set of linear transformations (Haxby et al., 2011). These transformation can then be used to bring any new data to/from a participant's brain coordinate system and the common space (Haxby et al., 2011). Despite all these achievements the underlying neural mechanisms of neurofeedback remain unclear (Shibata et al., 2019).

On a subject level, the attention and motivation of the participants was suggested as an important factor for the success of neurofeedback (Kadosh & Staunton, 2019). Furthermore, factors like monetary rewards (Sepulveda et al., 2016) and social reward (Mathiak et al., 2010) might be important reinforcers increasing neurofeedback learning success. One of the most debated aspects of the design of neurofeedback studies is if whether precise regulation strategies should be provided to participants or not and which strategies improve the

neurofeedback learning success rate the most (Sitaram et al., 2017). Haugg et al. (2021) brought up two further factors that might influence the neurofeedback performance of participants. Firstly, the authors mentioned the number of training runs, as this is particularly relevant for fMRI studies as every run is connected to a high monetary cost since scanning hours are expensive. Secondly, whether neurofeedback training should be performed within a single day or across multiple days, as neurofeedback training might get facilitated through sleep consolidation. One still must keep in mind that these and many more factors can possibly influence the efficacy of neurofeedback.

To summarise, there are numerous factors and interactions between these factors – subject and study specific – that might influence neurofeedback learning success. In order to improve the efficacy of these studies an optimization of their design would be of great benefit. Unfortunately, this is hardly feasible in empirical studies. Therefore, a "big data" approach is suitable to investigate this problem. Additionally, approaching the problem with a big data methodology allows one to account for possible interaction between variables, and typically yields more generalizable findings.

Methods

Study objectives

This thesis seeks to advance the understanding of the mechanisms behind neurofeedback or more specifically neurofeedback learning success. As a first step, neurofeedback data was collected from various sources, including data from Haugg et al. (2021), the five studies of the decoded neurofeedback dataset (Amano et al., 2016; Cortese et al., 2016, 2021; Koizumi et al., 2016; Shibata et al., 2016; Taschereau-Dumouchel et al., 2018) and unpublished data from Haugg and Frei. The objective was to identify features of studies and the participants that improved neurofeedback learning success. This could help resolving

the current issue of low effectiveness in training participants to gain control over their neural signals. Furthermore, knowledge about optimized methods of neurofeedback training could improve the development and testing of it (Hampson et al., 2020).

Study Design

Eligible to the analysis were all participants of neurofeedback studies whose data was available for analysis on a per subject level. The aim was to obtain a heterogenous sample, as greater variability in subject and study characteristics may enhance the generalizability of the results and may improve the prediction accuracy of machine learning models by capturing a broader spectrum of relevant patterns.

Sample Description

Overall, there were 732 participants across 35 studies included in the analysis and all data was available on a per subject basis. This sample is an expansion of the 28 studies (608 participants) (Auer et al., 2015; Emmert et al., 2017; Hellrung, Borchardt, et al., 2018; Hellrung, Dietrich, et al., 2018; Hellrung et al., 2022; Keynan et al., 2019; D.-Y. Kim et al., 2015; Kirschner, 2017; Kirschner et al., 2018; Kohl et al., 2019; MacInnes et al., 2016; Marins et al., 2015; Marxen et al., 2016; McDonald et al., 2017; Megumi et al., 2015; Nicholson et al., 2017b; Pamplona et al., 2020b; Papoutsi et al., 2018, 2019; Scharnowski et al., 2012, 2015; Sorger et al., 2018; Spetter et al., 2017; Yao et al., 2016; Young et al., 2017; Zich et al., 2020) used in the original paper from Haugg et al. (2021) and incorporated the decoded neurofeedback dataset (Cortese et al., 2021), which consists of 66 participants across 5 studies (Amano et al., 2016; Cortese et al., 2016; Koizumi et al., 2016; Shibata et al., 2016; Taschereau-Dumouchel et al., 2018). Unfortunately, 3 participants had to be excluded due to inconsistencies between the number of trials and the number of runs in the provided data. Furthermore, 60 participants were added to the dataset from unpublished studies from Haugg

and Frei. The high heterogeneity of the data should ensure generalizability of the results, as a high number of diverse variables, that capture subject specific and study specific factors that might influence neurofeedback learning success, was included in the dataset (for further information on the variables please refer to the chapter *Measured Variables*).

Measured Variables

Overall, 21 features were used for the prediction of the neurofeedback learning success measures. All of these features were available on a per subject level in the collected data. As this thesis included much of the same data as Haugg et al. (2021) the features used for the prediction stayed the same, except for some minor changes.

- Three subject-specific features: (1) age of the subject in years, (2) sex of the subject, (3) health status of the subject (healthy subject or patient);
- Six ROI-based features: (4), ROI(s) is/are cortical or subcortical, (5) ROI(s) is/are a sensory brain region. (6) ROI(s) is/are part of the default mode network (DMN), (7) ROI(s) is/are part of the salience network, (8) ROI(s) is/are part of the motor network, (9) ROI(s) is/are a single or multiple clusters;
- Eleven experimental design-specific features: (10) use of connectivity- vs activity- vs decoded neurofeedback, (11) use vs no use of a functional localiser for defining the ROI, (12) up- vs down-regulation vs decoded neurofeedback, (13) use of precise- vs open- vs no strategy suggestions provided to the participants, (14) use vs no use monetary rewards, (15) use vs no use of a pre-training no-feedback run, (16) length of a single neurofeedback training run in seconds, (17) length of a single neurofeedback regulation block in seconds, (18) length of training time in minutes, (19) neurofeedback training on one day vs across multiple days, (20) use vs no use of percent signal change and (21) continuous vs no continuous presentation of feedback.

Table 1 Overview of studies: Data of 732 participants across 35 studies (484 healthy participants and 248 patients; 596 participants in acitivity-based studies, 72 participants in 72 studies and 64 participants in DecNef studies)

Author	Participants	NFB Type
Auer	Healthy (N = 16)	Activity
Emmert	Patients (N = 14)	Activity
Megumi	Healthy (N = 12)	Connectivity
Keynan	Healthy (N = 33)	Activity
Kim (act)	Patients (N = 7)	Activity
Kim (con)	Patients (N = 7)	Connectivity
Kirschner	Healthy (N = 27), Patients (N = 24)	Activity
Kirschner	Patients (N = 14)	Activity
Liew	Healthy (N = 10)	Connectivity
MacInnes	Healthy (N = 19)	Activity
Marins	Healthy (N = 14)	Activity
McDonald	Healthy (N = 68), Patients (N = 72)	Activity
Pamplona	Healthy (N = 15)	Activity
Papoutsi	Patients (N = 10)	Activity
Scharnowski	Healthy (N = 7)	Activity
Scharnowski	Healthy (N = 10)	Activity
Shuxia	Healthy (N = 18)	Activity
Sorger	Healthy (N = 10)	Activity
Spetter	Patients (N = 8)	Connectivity
Young	Patients (N = 18)	Activity
Zich	Healthy (N = 27)	Connectivity
Hellrung	Healthy (N = 34)	Activity
Hellrung	Healthy (N = 16)	Activity
Hellrung	Healthy (N = 11)	Activity
Papoutsi (act)	Patients (N = 8)	Activity
Papoutsi (con)	Patients (N = 8)	Connectivity
Marxen	Healthy (N = 32)	Activity
Nicholson	Patients (N = 14)	Activity
Kohl	Patients (N = 16)	Activity
Kohl	Patients (N = 9)	Activity
Shibata	Healthy (N = 23)	Decoded
Amano	Healthy (N = 12)	Decoded
Koizumi	Healthy (N = 9)	Decoded
Cortese	Healthy (N = 9)	Decoded
Taschereau-Dumouchel	Healthy (N = 11)	Decoded
Haugg	Healthy (N = 41), Patients (N = 3)	Activity
Frei	Patients (N = 16)	Activity

Success in neurofeedback is typically evaluated using different metrics such as percent signal change values, beta values, Bayes factors, correlation values, etc. These feedback values are used to assess each subject's performance and are also presented to the participant during the neurofeedback training. However, defining a common measure of neurofeedback learning success is important yet challenging, as it enables comparison across different subjects and studies, and allows for the pooling of participants. To address this, the same two targets as in Haugg et al. (2021) were used: neurofeedback performance and neurofeedback improvement. Two neurofeedback learning success measures were defined as follows:

- Neurofeedback performance: This is the ratio of successful neurofeedback runs compared to unsuccessful neurofeedback runs per subject. A successful run was defined as a run with positive feedback values for neurofeedback up-regulation studies, negative feedback values for neurofeedback down-regulation studies For example, if a subject is instructed to increase activity in a specific brain region (up-regulation) and achieves a positive percent signal change in that region during the run, this would count as a successful up-regulation run. Similarly, a negative signal change during a down-regulation task would indicate success. In DecNef studies a run with an induction likelihood of $P>0.5$ was defined as successful. For the purposes of the classification, a subject who showed more than 50% of successful runs was labelled as successful, otherwise the subject was labelled as unsuccessful.
- Neurofeedback improvement: This is the slope of the regression line over all the neurofeedback training values of each run of each subject. Successful subjects were defined as subjects with a slope value greater than zero, unsuccessful subjects were defined as subjects with a slope value equal to or smaller than zero.

Additionally, a third target was defined for analysing the data in a regression task:

- Neurofeedback improvement (regression): This is the slope of the regression line over all the neurofeedback training values of each run of each subject.

Research questions

To improve the efficacy of neurofeedback learning and to gain knowledge about the optimal study design, this thesis aims to explore what subject and study specific factors influence neurofeedback learning success:

1. To what extent is neurofeedback learning success influenced subject specific or study specific factors?
2. To what extent is neurofeedback performance influenced by subject specific or study specific factors?
3. To what extent is neurofeedback improvement influenced by subject specific or study specific factors?
4. If an influence is found, which subject specific and study specific factors influenced neurofeedback performance the most?
5. If an influence is found, which subject specific and study specific factors influenced neurofeedback improvement the most?

If an influence is found further factors not explicitly represented in the measured variables will be investigated in an exploratory fashion.

Procedure

All the analyses were performed using Python v3.12 (van Rossum, 1995).

Temporal dynamics and variability of neurofeedback measures

First, the mean standard deviation and baseline corrected slopes of the neurofeedback values over the training time were plotted to assess both the variability in performance (standard deviation) and the direction and magnitude of learning-related changes (baseline

corrected slopes) – once for all subjects, once for (1) all subjects of activity based neurofeedback studies, once for (2) all subjects of connectivity based neurofeedback studies and once for (3) all subjects of DecNef studies.

Interpretable machine learning

The interpretable machine learning analysis was performed in the machine-learning framework scikit-learn library v1.6.0 (Pedregosa et al., 2011). Machine learning was chosen over traditional statistical approaches due to its ability to flexibly model complex, potentially non-linear interactions between multiple features without relying on strong parametric assumptions. Furthermore, it enables the identification and quantification of feature relevance, thereby providing not only predictive performance but also insight into the underlying mechanisms associated with neurofeedback learning (Bzdok et al., 2018). By leveraging cross-validation and model interpretability techniques, the analysis ensures both generalizability and transparency in the derived findings. In order to assess the impact of the subject and study specific factors on neurofeedback learning success a total of 6 different analyses were performed: For answering the first, the second and the fourth research question (1) a classification task of the neurofeedback performance (successful vs. unsuccessful) with a per subject grouping and (2) a classification task of the neurofeedback performance (successful vs. unsuccessful) with a per study grouping were performed. For answering the first, the third and the fifth research question (3) a classification task of the neurofeedback improvement (successful vs. unsuccessful) with a per subject grouping, (4) a classification task of the neurofeedback improvement (successful vs. unsuccessful) with a per study grouping, (5) a regression of the neurofeedback improvement with a per subject grouping and (6) a regression task of the neurofeedback improvement with a per study grouping were performed. In this context, grouping refers to the level at which datapoints were considered as belonging to the same unit for the purpose of model construction and evaluation. When grouping was done at

the subject level, all data points from a single subject were treated as belonging to one group. Conversely, when grouping was done at the study level, datapoints from all subjects within the same study were treated as one group. Each of these analyses consisted of three parts: (1) prediction model training, (2) generalizability testing and (3) model analysis.

Exploratory analysis

After the completion of the six planned analyses described above, two additional exploratory analyses were conducted to further investigate factors not explicitly represented in the measured variables. To this end, the variable "study ID" was added as an additional discrete feature. As a perfect study indicator, "study ID" summarizes all study specific factors, including those captured in the feature set and those not explicitly represented. The classification tasks for (7) neurofeedback performance and (8) neurofeedback improvement were repeated using the same machine learning procedures as in the main analysis. *Table 2* provides an overview over all the analyses performed in the scope of this thesis.

Table 2 Analyses overview

	Neurofeedback Performance	Neurofeedback Improvement
Per Subject Grouping	Classification Task	Classification Task and Regression Task
Per Study Grouping	Classification Task	Classification Task and Regression Task
Exploratory Analysis	Classification Task with study ID and per subject grouping	Classification Task with study ID and per subject grouping

Prediction model training

Light gradient boosted machine models were used for both the classification tasks and the regression tasks as they are computationally efficient, fast, capable of handling large-scale data and highly accurate (Geurts et al., 2006; Microsoft Coperation, 2024a). Additionally, these

machine learning models are robust to multicollinearity and outliers in the data. Furthermore, they are capable of modelling non-linear relationships and interactions between features and target variables (Hastie et al., 2009; Lundberg & Lee, 2017). The features used in the models are the 21 previously described in the chapter *Sample Description*.

Generalizability testing

To assess the models' generalizability and evaluate the performance of the models on previously unseen data, a nested cross-validation (CV) was utilized (Cawley & Talbot, 2010). A nested CV consists of two major parts: an outer CV and an inner CV, both being a repeated k-fold CV. In a repeated k-fold CV the data is split into a certain number (k) of folds. One of these folds is used for testing the model, the remaining (k-1) get utilized for training. This process gets repeated until each fold was used for testing once. Thereafter, the data's association to the folds gets changed, and the splitting process gets repeated j times. Therefore, overall k x j models are generated, over which the final value for the predictive power gets averaged upon (Coutanche & Hallion, 2020). Due to the process of CV, issues like sampling bias – where certain groups or characteristics are over- or under-represented in the data splits – and overfitting to a specific subset – where the model learns patterns that are too specific to a particular partition and fail to generalize – should be countered (Cawley & Talbot, 2010). For the classification and the regression tasks a 5 fold outer CV with repeats till 20000 predictions were achieved (*outer CV*, 5 folds, *minimum predictions* = 20000) was used. Furthermore, for both task a 5 fold inner CV with repeats till 1000 predictions were achieved (*inner CV*, 5 folds, *minimum predictions* = 1000) was used for hyperparameter tuning. In this analysis, the hyperparameters are model settings that define boundaries of the structure and complexity of the decision trees (Cawley & Talbot, 2010). Random search was used to get hyperparameter candidates. The search-space was defined as follows: *'estimator__colsample_bytree'*: uniform(0.1, 0.9), *'estimator__extra_trees'*: [True, False], *'estimator__reg_lambda'*:

loguniform(0.1, 100). The following hyperparameters fixed for the regression tasks: boosting_type='gbdt', num_leaves=100, max_depth=-1, learning_rate=0.01, n_estimators=1000, subsample_for_bin=100000, objective='huber', min_split_gain=0.0001, min_child_weight=0.0001, min_child_samples=10, subsample=1.0, subsample_freq=0, colsample_bytree=1.0, reg_alpha=0.0, reg_lambda=0.0, random_state=None, n_jobs=1, importance_type='gain', 'data_random_seed': None, 'data_sample_strategy': 'bagging', 'extra_seed': None, 'feature_fraction_seed': None, 'feature_pre_filter': False, 'force_col_wise': True, 'min_data_in_bin': 1, 'use_quantized_grad': False, 'verbosity': -1. For the classification tasks the following hyperparameters were fixed: boosting_type='gbdt', num_leaves=100, max_depth=-1, learning_rate=0.01, n_estimators=1000, subsample_for_bin=100000, objective='multiclass', class_weight='balanced', min_split_gain=0.0001, min_child_weight=0.0001, min_child_samples=10, subsample=1.0, subsample_freq=0, colsample_bytree=1.0, reg_alpha=0.0, reg_lambda=0.0, random_state=None, n_jobs=1, importance_type='gain', 'data_random_seed': None, 'data_sample_strategy': 'bagging', 'extra_seed': None, 'feature_fraction_seed': None, 'feature_pre_filter': False, 'force_col_wise': True, 'min_data_in_bin': 1, 'num_class': task['n_classes'], 'use_quantized_grad': False, 'verbosity': -1. The remainder of the hyperparameters were left to the default (Microsoft Coperation, 2024b). The best parameters found were used for training the model in the outer CV loop. Classification performance was evaluated using balanced accuracy (García et al., 2009). Regression performance was evaluated using two metrics: (1) the prediction coefficient of determination (R^2), and (2) the mean absolute error (MAE). The MAE represents the average magnitude of prediction errors, expressed in the same units as the dependent variable (Hastie et al., 2009). For both tasks models trained on the same data, but with shuffled labels were calculated. For the comparison between the original and the shuffled model – both for classification and regression tasks – a for the lack of independency in the sample corrected

version of the paired sample t-test was utilized. This correction was needed because of the repeated use of the sample for training the models. If the models trained on the original data perform significantly better than the models trained on the data with shuffled labels, this suggests that the independent variables have real predictive value. In such cases, the average predictive power of the model can be interpreted as either its average prediction accuracy or as the proportion of variance in the dependent variable explained by the independent variables dependent of the task (Bouckaert & Frank, 2004; Nadeau & Bengio, 2003).

Model analysis

For assessing the feature importances and the importance of interactions, SHAP (SHapley Additive exPlanations) was used. SHAP is a method from interpretable machine learning that calculates SHAP values based on Shapley values, a method from cooperative game theory. Shapley values indicate the average deviation of the target variable from the expected value explained by the individual feature. SHAP measures the contributions of each predictor to the model's performance. Pooling those contributions over many predictions allows for a comprehensive analysis of the importance of individual features for the prediction task (Campbell et al., 2022; Lundberg & Lee, 2017; Molnar, 2019).

Results

The following section presents a summary of the findings from the conducted analyses. First, an analysis of the temporal dynamics and variability of the data will be presented. Subsequently, the performance of the machine learning models and their results, including prediction results and feature importances, will be described in detail. Finally, the results of the exploratory analysis will be described.

Plotted temporal dynamics and variability of neurofeedback measures

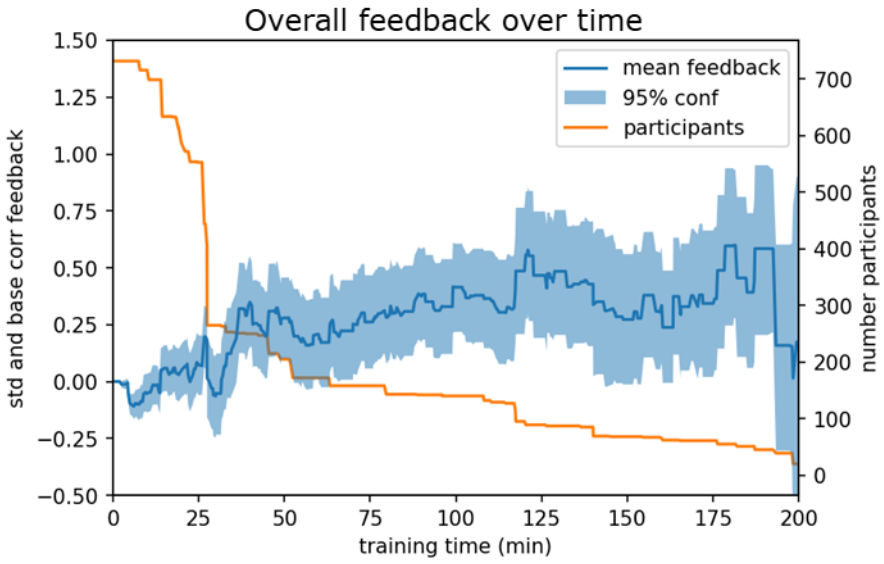

Figure 2 Mean standard deviation and baseline corrected feedback (neurofeedback improvement) over training time: The blue line depicts the average feedback value across all participants as a function of training time (in minutes). The light blue shaded region represents the 95% CI (confidence interval) of the mean. The information about the standard deviation and baseline corrected feedback can be gathered from the left-hand y-axis. The orange line indicates the number of participants contributing data at each time point. The information about the number of participants can be gathered from the right-hand y-axis.

Figure 2-5 depict the mean standard deviation and baseline corrected neurofeedback improvement over the training time (in minutes). In each figure, the blue line represents the mean feedback, with the shaded region indicating the 95% confidence interval (CI), and the orange line (right-hand y-axis) depicting the number of participants at each time point. The 95% CI was calculated using bootstrapping: 10000 bootstrap samples were generated by resampling participants with replacement and computing the mean feedback for each sample at each time point. The lower and upper bounds of the shaded region correspond to the 2.5th and 97.5th percentiles of these bootstrapped means.

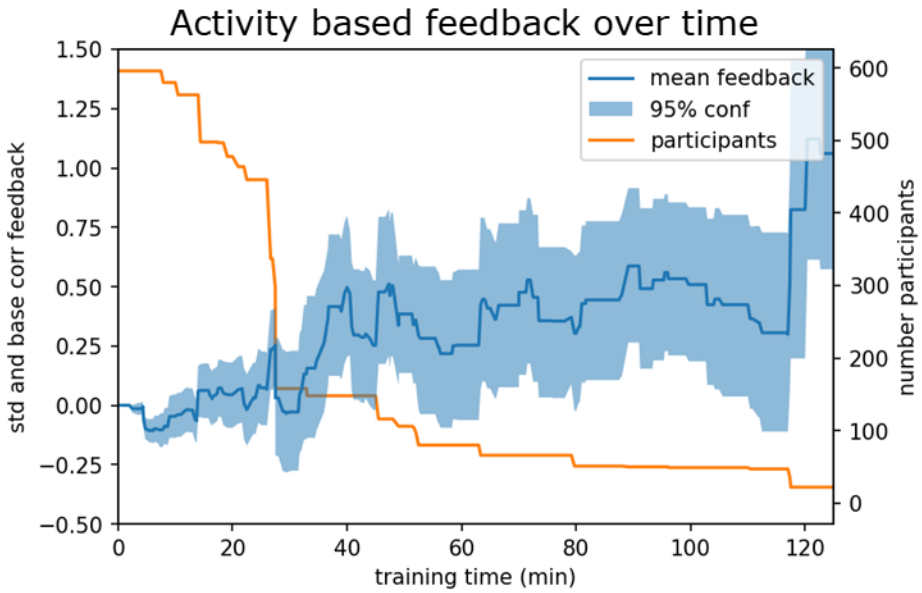

Figure 3 Mean standard deviation and baseline corrected feedback (neurofeedback improvement) over training time of activity based neurofeedback data: The blue line depicts the average feedback value across all participants as a function of training time (in minutes). The light blue shaded region represents the 95% CI (confidence interval) of the mean. The information about the standard deviation and baseline corrected feedback can be gathered from the left-hand y-axis. The orange line indicates the number of participants contributing data at each time point. The information about the number of participants can be gathered from the right-hand y-axis.

Figure 2 depicts the feedback over the training time for all 732 participants, *Figure 3* for participants of activity based neurofeedback studies, *Figure 4* for participants of connectivity based neurofeedback studies and *Figure 5* for DecNef studies.

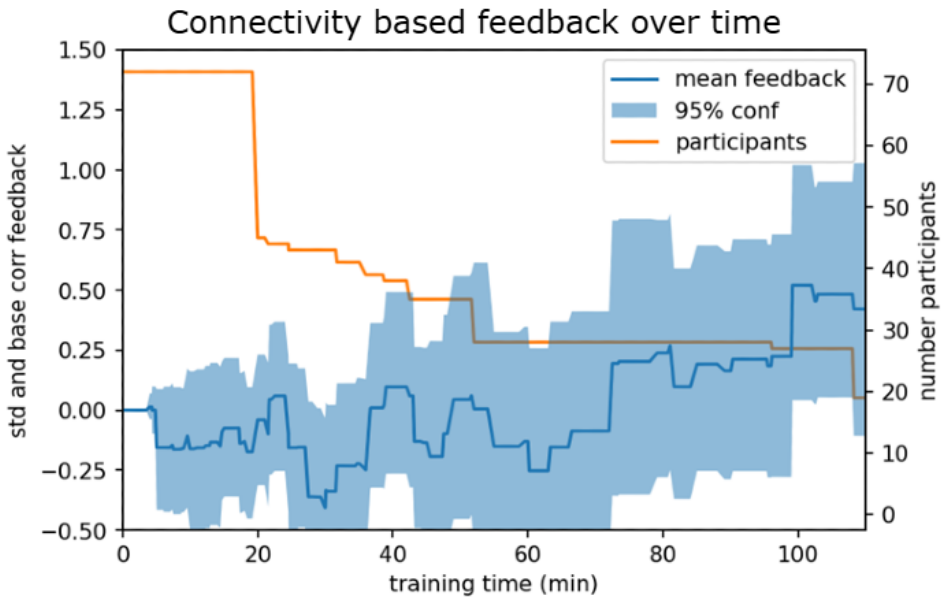

Figure 4 Mean standard deviation and baseline corrected feedback (neurofeedback improvement over training time of connectivity based neurofeedback data: The blue line depicts the average feedback value across all participants as a function of training time (in minutes). The light blue shaded region represents the 95% CI (confidence interval) of the mean. The information about the standard deviation and baseline corrected feedback can be gathered from the left-hand y-axis. The orange line indicates the number of participants contributing data at each time point. The information about the number of participants can be gathered from the right-hand y-axis.

In studies utilizing activity based neurofeedback or connectivity based neurofeedback a drop-off in participants happened sooner (approximately 25 minutes and 20 minutes of training time respectively) compared to the DecNef studies, which experienced a drop-off after approximately 150 minutes.

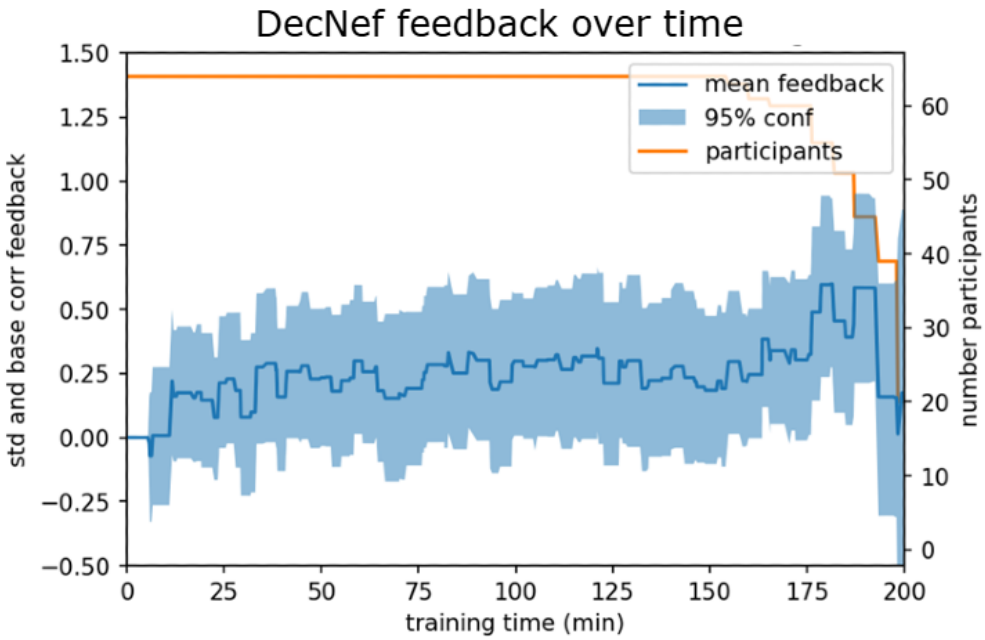

Figure 5 Mean standard deviation and baseline corrected feedback (neurofeedback improvement) over training time of DecNef-data: The blue line depicts the average feedback value across all participants as a function of training time (in minutes). The light blue shaded region represents the 95% CI (confidence interval) of the mean. The information about the standard deviation and baseline corrected feedback can be gathered from the left-hand y-axis. The orange line indicates the number of participants contributing data at each time point. The information about the number of participants can be gathered from the right-hand y-axis.

Machine learning analysis results

In the following section, the model fit and the feature importances of the different models are described. Both tasks – classification and regression – were performed with a per subject and a per study grouping.

Neurofeedback performance classification task results

Per subject grouping classification performance

To evaluate the ability of the classifier to predict the neurofeedback performance (neurofeedback_performance) with a per subject grouping, the balanced accuracy of the model using cross-validation was assessed.

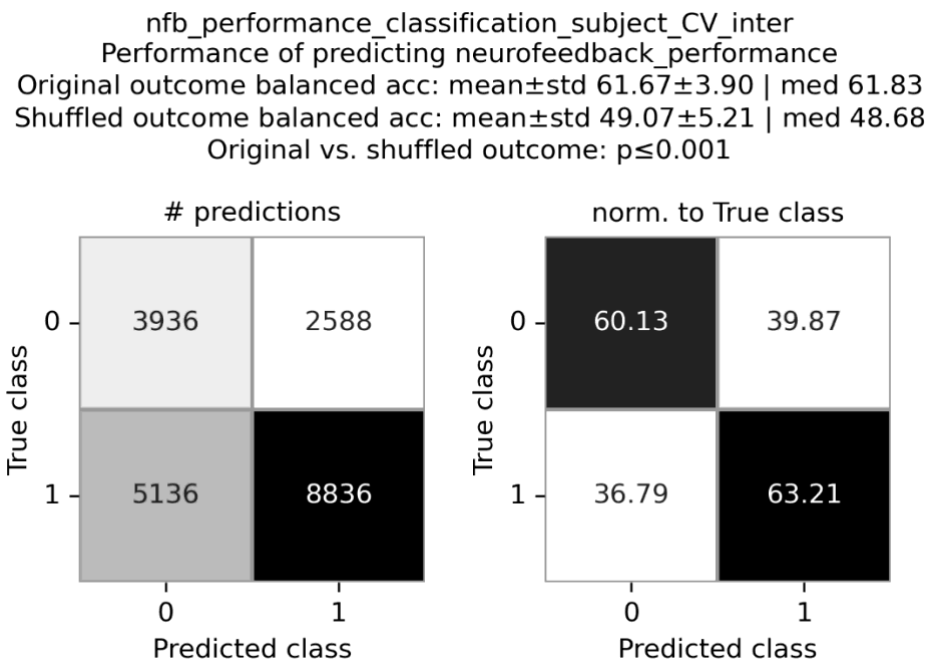

Figure 6 Classification results with per subject grouping: The outcome on the original dataset was a mean balamced accurcacy of 61.67 % [SD = 3.90, median = 61.83]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 49.07 % [SD = 5.21, median = 48.68]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in $p \leq 0.001$, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback performance (neurofeedback_performance). The left panel shows the number of predictions for each class: 3936 true negatives, 2588 false positives, 5136 false negatives, and 8836 true positives. The right panel presents the normalized values based on the true class distribution, with 60.13 % true negatives, 39.87 % false positives, 36.79 % false negatives, and 63.21 % true positives.

The results are summarized in *Figure 6*. The classifier trained on the original dataset achieved a mean balanced accuracy of 61.67 % [SD = 3.90, median = 61.83]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 49.07 % [SD = 5.21, median = 48.68]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in $p \leq 0.001$, suggesting that the classifier's performance was significantly better than chance.

The confusion matrices in *Figure 6* depict the classifier's performance in absolute counts (left matrix) and the normalized to the true class distributions (right matrix). The absolute number is bigger than the number of subjects due to CV. The classifier exhibited a slight bias toward predicting neurofeedback performance (class 1), as indicated by a higher true positive rate (63.21 %) compared to the true negative rate (60.13 %).

Per subject grouping feature importance analysis

To gain deeper insight into how each individual feature influenced the classification outcome, SHAP was performed for the model's predictions.

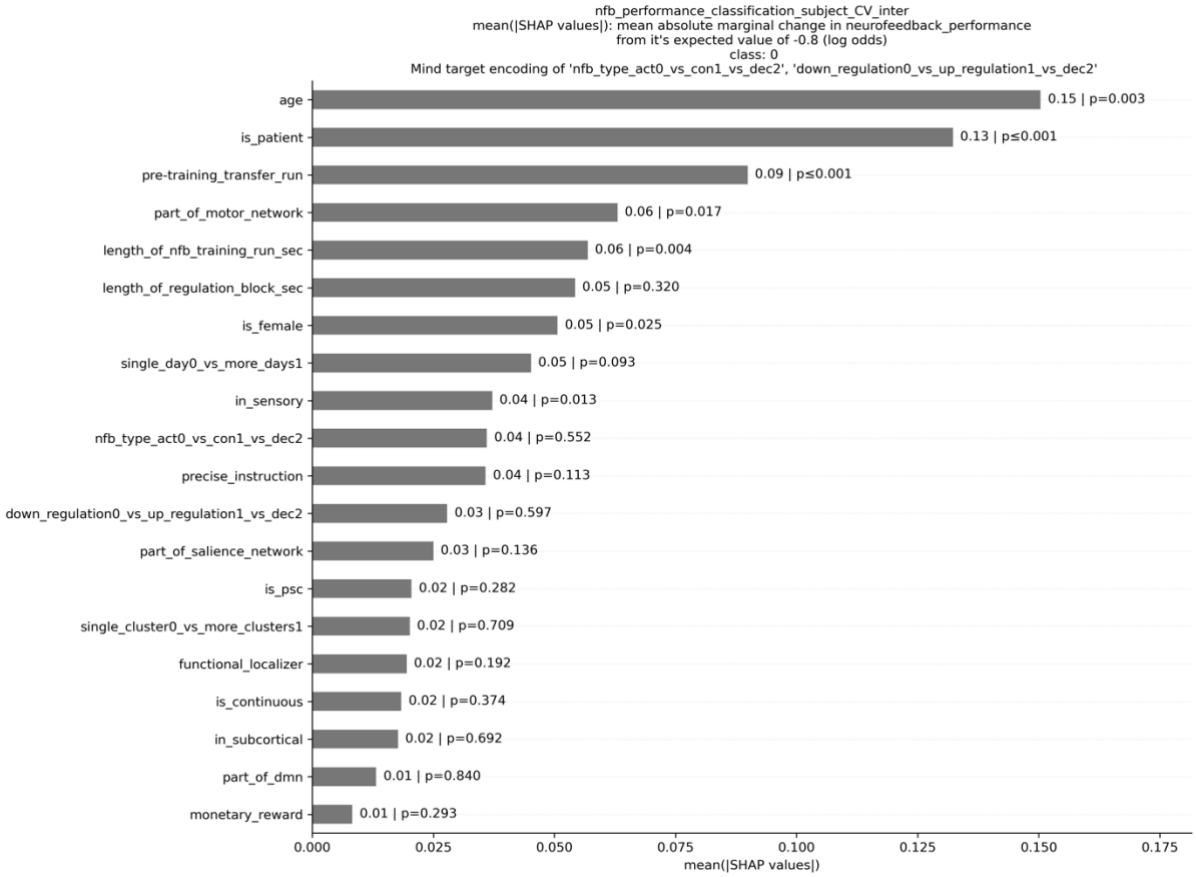

Figure 7 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting a non-successful neurofeedback performance. Bar plot displaying the mean absolute SHAP values for the most important features in predicting a non-successful neurofeedback performance (class 0). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include age (SHAP ~0.15, p = 0.003), is_patient (SHAP ~0.13, p ≤ 0.001), and pre-training_transfer_run (SHAP ~0.09, p ≤ 0.001).

The mean absolute SHAP values for the most influential features are depicted in *Figure 7* and *Figure 8* for predicting no successful neurofeedback performance (class:0) and *Figure 9* and *Figure 10* for predicting successful neurofeedback performance (class:1). The feature with the highest importance for the prediction of a non-successful neurofeedback performance (class 0) was age [mean absolute SHAP value = 0.15, p = 0.003], which indicates that on average, the feature age moved the neurofeedback performance (class 0) by SHAP = 0.15 from the expected value -0.8 (log odds). Further important features were the health status of the subject

(is_patient) [mean absolute SHAP value = 0.13, p ≤ 0.001], and the presence of a pre-training transfer run (pre-training_transfer_run) [mean absolute SHAP value = 0.09, p ≤ 0.001]. For a successful neurofeedback performance (class 1), the most impactful features include the presence of pre-training transfer runs (pre-training_transfer_run) [mean absolute SHAP value = 0.13, p ≤ 0.001], the length of the neurofeedback runs (length_of_nfb_training_run_sec) [mean absolute SHAP value = 0.11, p = 0.002], the health status of the subject (is_patient) [mean absolute SHAP value = 0.09, p = 0.005], sex (is_female) [mean absolute SHAP value = 0.09, p = 0.014], and age (mean absolute SHAP value = 0.09, p = 0.015).

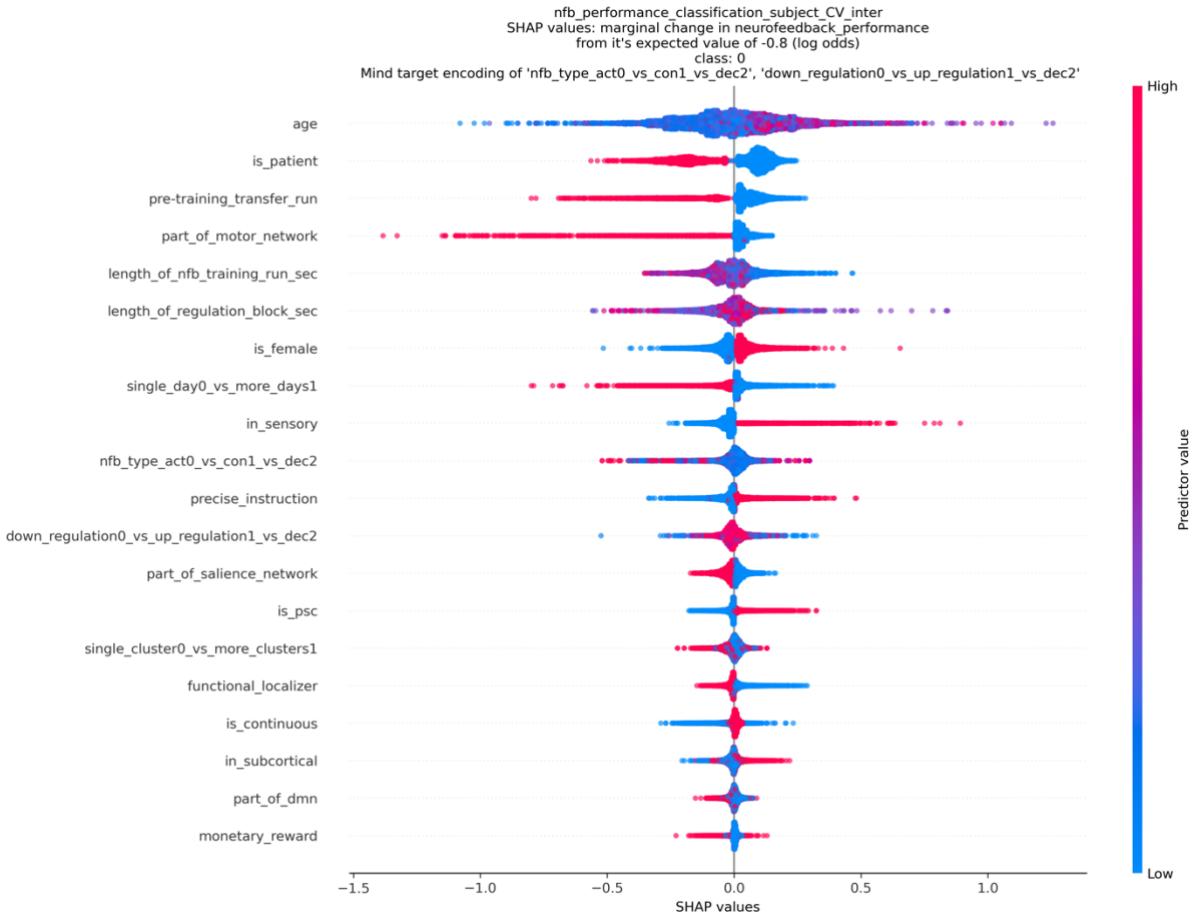

Figure 8 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (non-successful neurofeedback performance with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

Overall, some features showed a significant predictive value and the classifier's performance was significantly different from chance. Neurofeedback performance was successfully predicted above chance level with the current features, grouping and modeling approach.

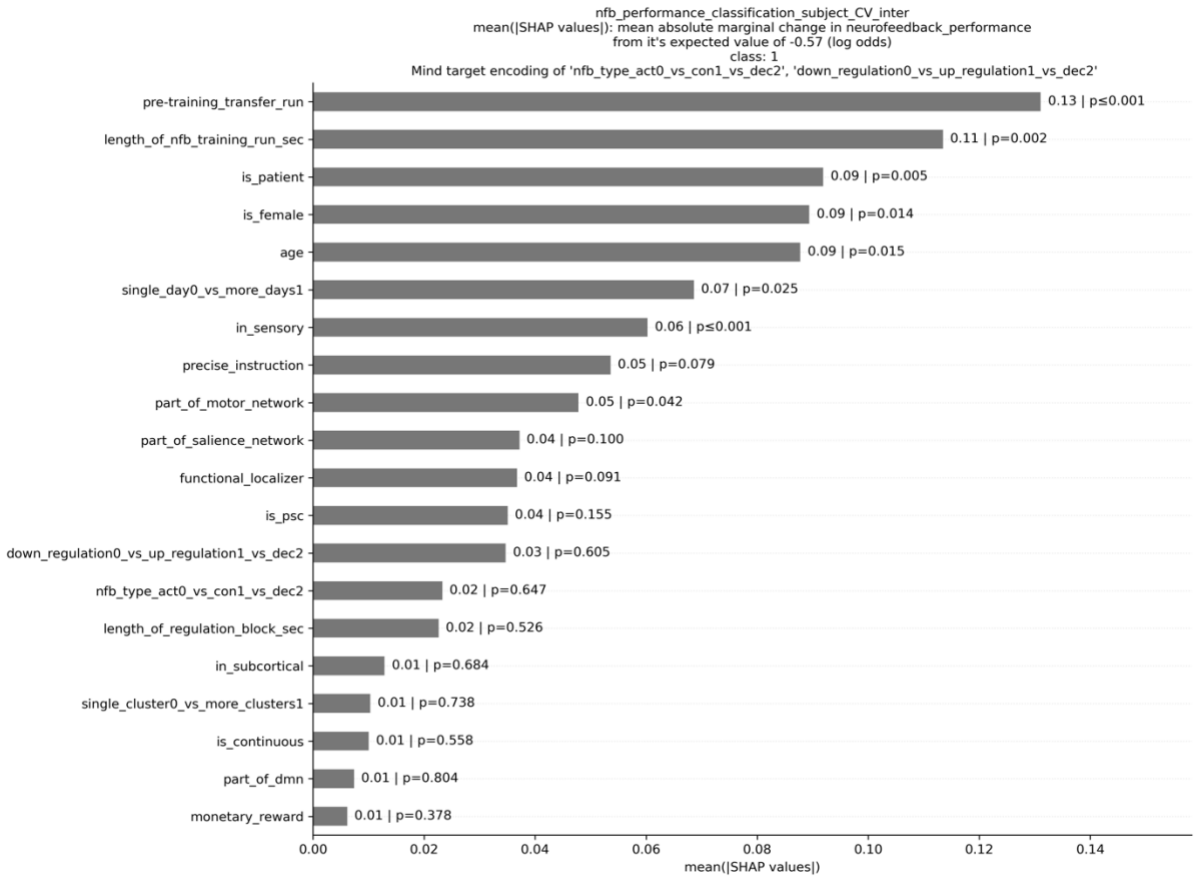

Figure 9 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting a successful neurofeedback performance. Bar plot displaying the mean absolute SHAP values for the most important features in predicting a successful neurofeedback performance (class 1). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include pre-training_transfer_run (SHAP ~0.13, $p \leq 0.001$), length_of_nfb_training_run_sec (SHAP ~0.11, $p = 0.002$), is_patient (SHAP ~0.09, $p = 0.005$), is_female (SHAP ~0.09, $p = 0.014$), and age (SHAP ~0.09, $p = 0.015$).

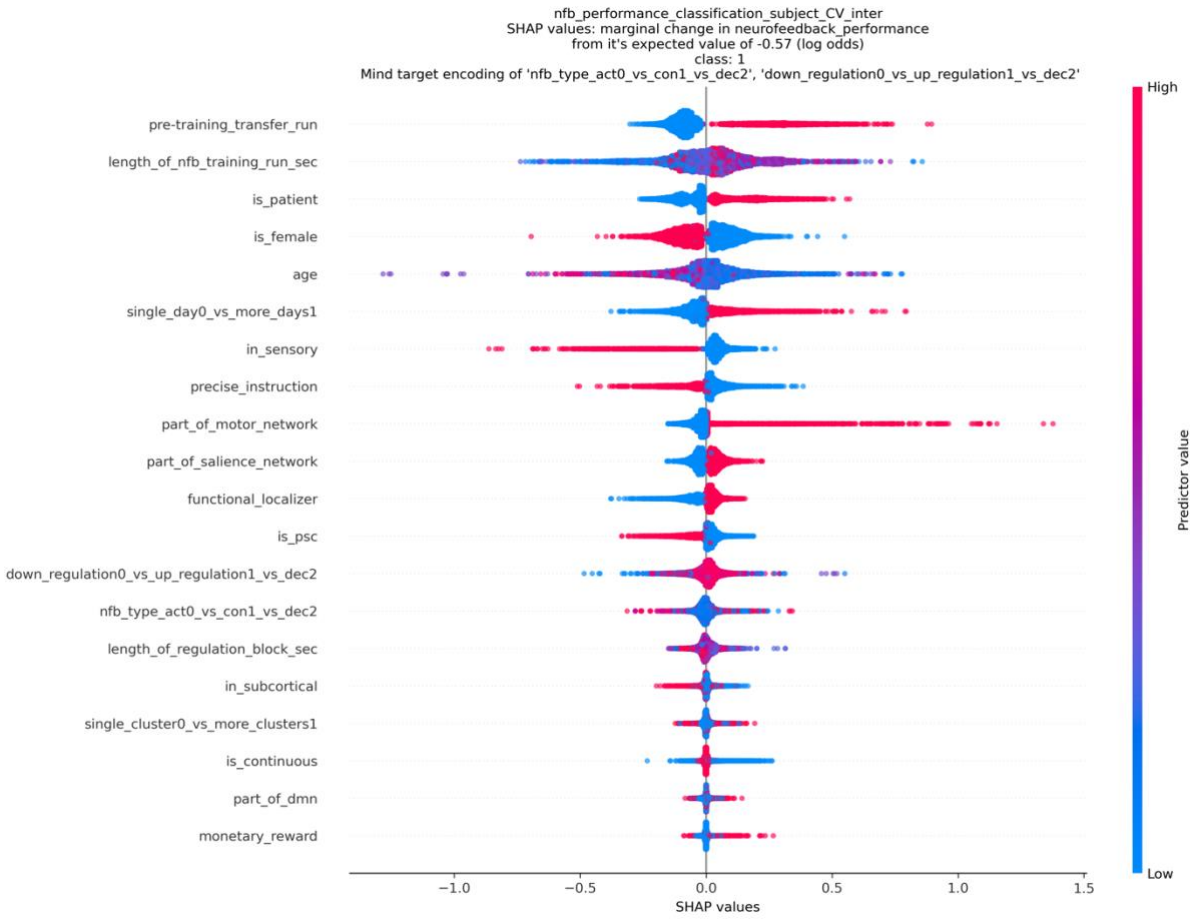

Figure 10 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (successful neurofeedback performance with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

Per study grouping classification performance

To assess the model's ability to predict the neurofeedback performance (neurofeedback_performance) with a per study grouping, the balanced accuracy was evaluated using CV.

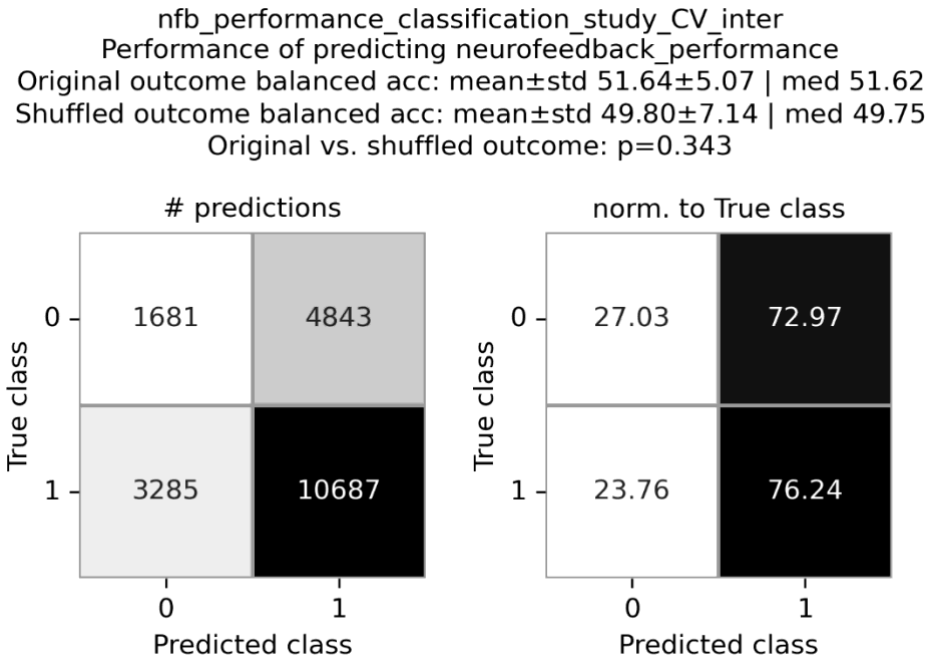

Figure 11 Classification results with per study grouping: The outcome on the original dataset was a mean balamced accurcacy of 51.64 % [SD = 5.07, median = 51.62]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 49.80 % [SD = 7.14, median = 49.75]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in p = 0.343, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback performance (neurofeedback_performance). The left panel shows the number of predictions for each class: 1681 true negatives, 4843 false positives, 3285 false negatives, and 10687 true positives. The right panel presents the normalized values based on the true class distribution, with 27.03 % true negatives, 72.97 % false positives, 23.76 % false negatives, and 76.24 % true positives.

The results are summarized in *Figure 11*. The classifier on the original dataset showed a mean balanced accuracy of 51.64 % [SD = 5.07, median = 51.62], which declined to 49.80 % [SD = 7.14, median = 49.75] with shuffled labels. The corrected paired sample t-test yielded p = 0.343, suggesting that the classifier's predictive ability did not exceed chance level.

Neurofeedback improvement classification task results

Machine learning was applied to predict the neurofeedback improvement. The target was binary – whether a participant showed neurofeedback improvement (1) or not (0).

Per subject grouping classification performance

To evaluate the ability of the classifier to predict neurofeedback improvement (slope_successful) with a per subject grouping, the balanced accuracy of the model using CV was assessed.

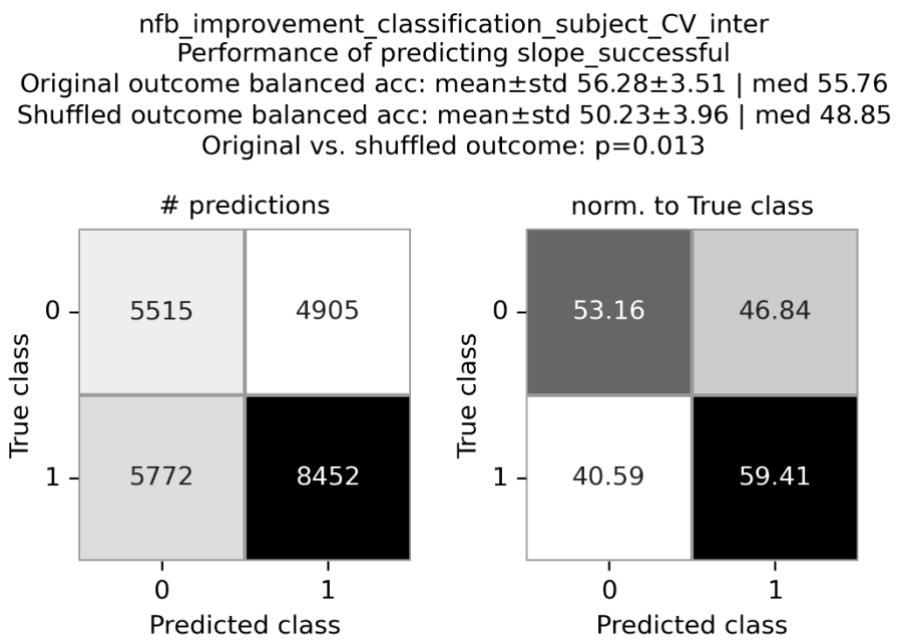

Figure 12 Classification results with per subject grouping: The outcome on the original dataset was a mean balamced accurcacy of 56.28 % [SD = 3.51, median = 55.76]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 50.23 % [SD = 3.96, median = 48.85]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in p = 0.013, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback improvement (slope_successful). The left panel shows the number of predictions for each class: 5515 true negatives, 4905 false positives, 5772 false negatives, and 8452 true positives. The right panel presents the normalized values based on the true class distribution, with 53.16 % true negatives, 46.84 % false positives, 40.59 % false negatives, and 59.41 % true positives.

The results are summarized in *Figure 12*. The classifier trained on the original dataset achieved a mean balanced accuracy of 56.28 % [SD = 3.51, median = 55.76]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 50.23 % [SD = 3.96, median = 48.85]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in p = 0.013, suggesting that the classifier's performance was not significantly better than chance.

The confusion matrices in *Figure 12* depict the classifier's performance in absolute counts (left matrix) and the normalized to the true class distributions (right matrix). The classifier exhibited a slight bias toward predicting neurofeedback improvement (class 1), as indicated by a higher true positive rate (59.41 %) compared to the true negative rate (53.16 %).

Per subject grouping feature importance analysis

To achieve a better understanding of the contributions of each single feature to the classification decision, SHAP was computed for the classifier's predictions.

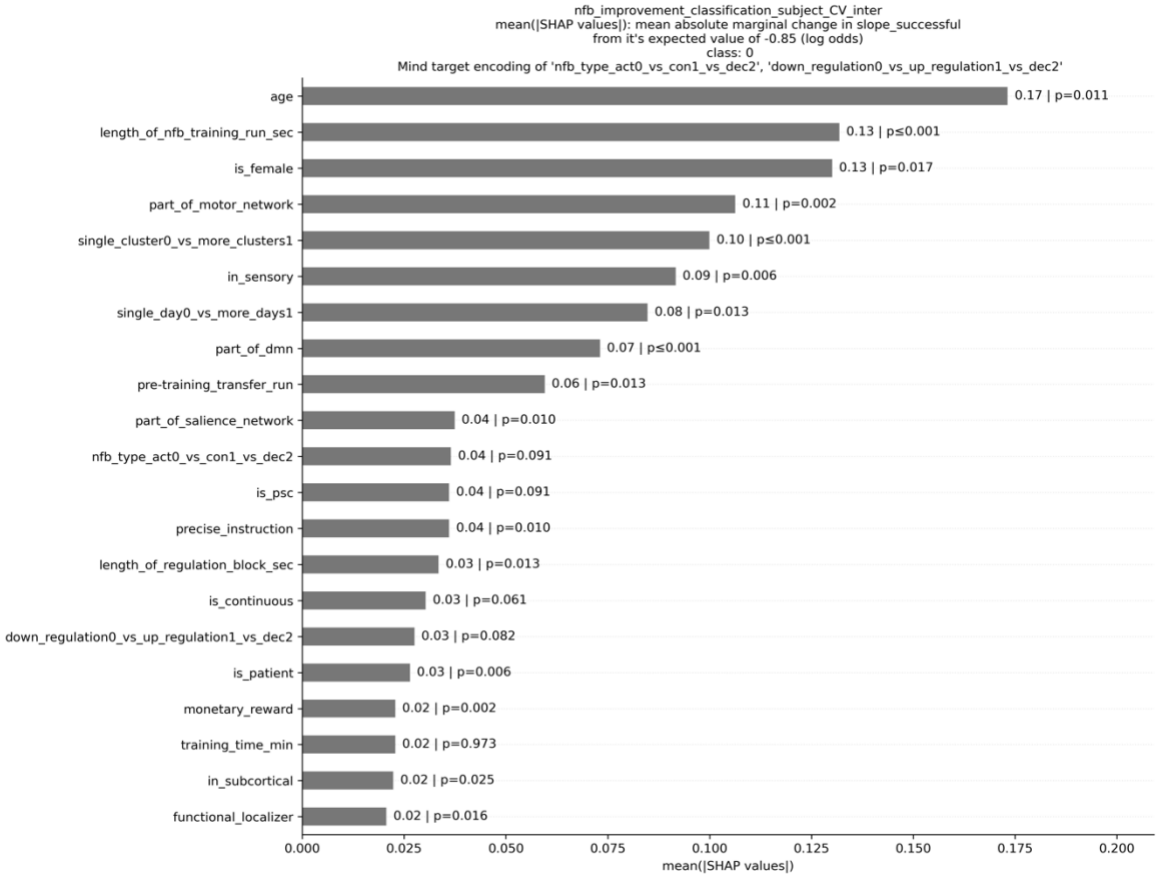

Figure 13 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting no learning success. Bar plot displaying the mean absolute SHAP values for the most important features in predicting no neurofeed improvemnet (class 0). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include age (SHAP ~0.17, p = 0.011), length_of_nfb_training_run_sec (SHAP ~0.13, p ≤ 0.001), and is_female (SHAP ~0.13, p = 0.017).

The mean absolute SHAP values for the most influential features are depicted in *Figure 13* and *Figure 14* for predicting no neurofeedback improvement (class 0) and *Figure 15* and *Figure 16* for predicting neurofeedback improvement (class 1).

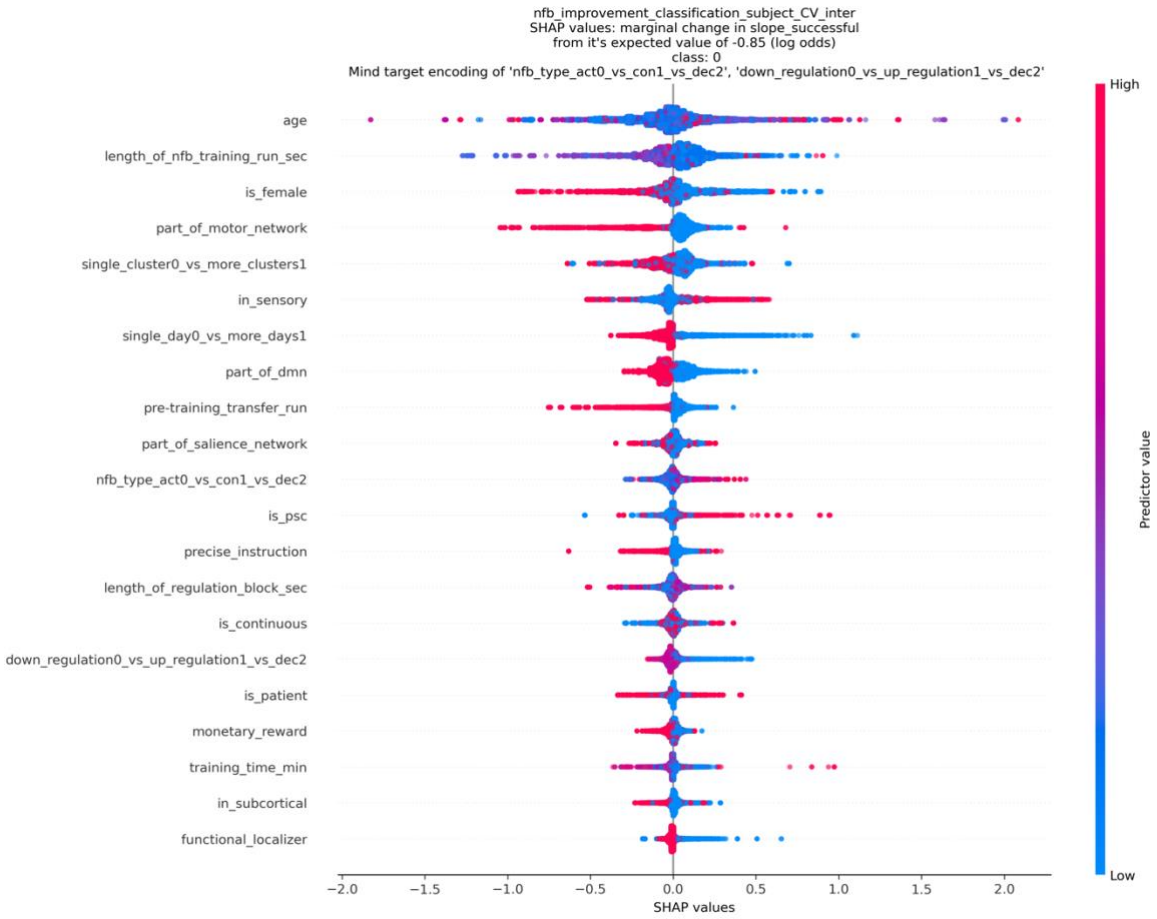

Figure 14 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (no learning success with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

The features with the highest importance for both prediction targets include age, with a mean absolute SHAP value of 0.17 for class 0 and 0.21 for class 1 [class 0: $p = 0.011$; class 1: $p = 0.011$], the length of the neurofeedback run (length_of_nfb_training_run_sec) [class 0: mean absolute SHAP value = 0.13, $p \leq 0.001$; class 1: mean absolute SHAP value = 0.13, $p = 0.013$] and the sex of the subject (is_female) [class 0: mean absolute SHAP value = 0.13, $p = 0.017$; class 1: mean absolute SHAP value = 0.12, $p \leq 0.001$].

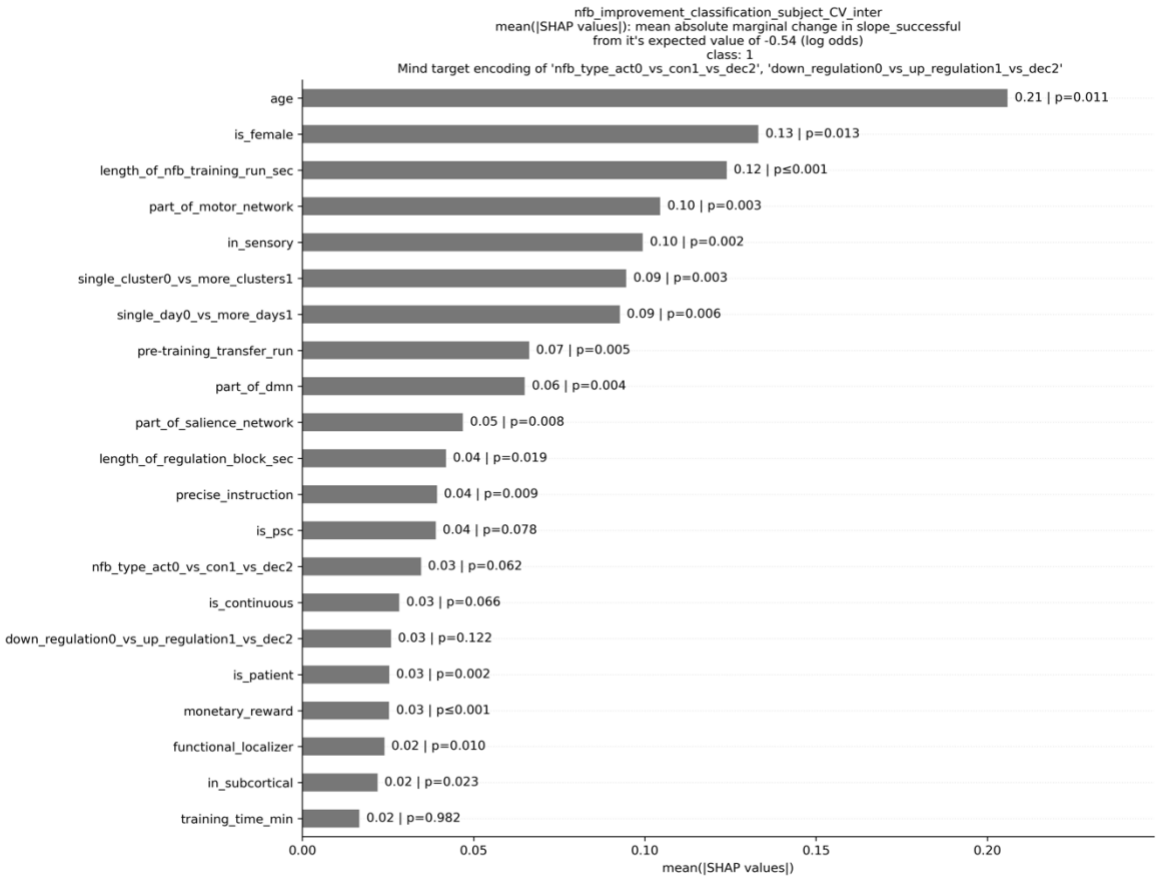

Figure 15 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting learning success. Bar plot displaying the mean absolute SHAP values for the most important features in predicting neurofeed improvemnet (class 1). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include age (SHAP ~0.21, p = 0.011), is_female (SHAP ~0.13, p = 0.013), and length_of_nfb_training_run_sec (SHAP ~0.12, p ≤ 0.001).

Overall, the classifier's performance was significantly different from chance.

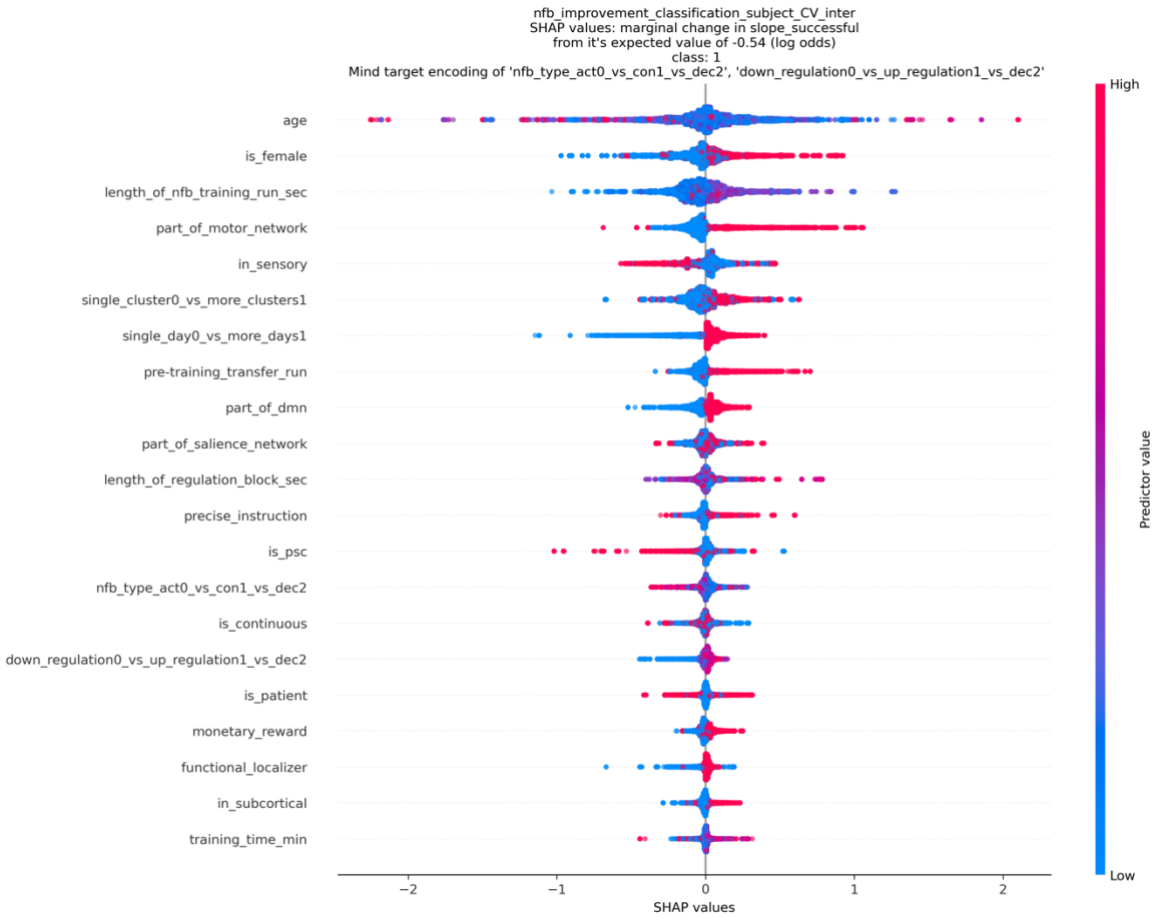

Figure 16 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (learning success with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

Per study grouping classification task results

To assess the classifier's ability to predict the neurofeedback improvement (slope_successful) with a per study grouping, the model's balanced accuracy was evaluated using CV. The results are presented in *Figure 17*. The classifier trained on the original dataset achieved a mean balanced accuracy of 52.71 % [SD = 6.74, median = 54.88]. When the outcome labels were shuffled, the mean balanced accuracy decreased to 49.18 % [SD = 4.94, median = 50.15]. A for the lack of independency in the sample corrected version of the paired sample t-test was used to compare the original and shuffled outcomes, resulting in p = 0.254, indicating that the classifier's performance was not significantly better than chance.

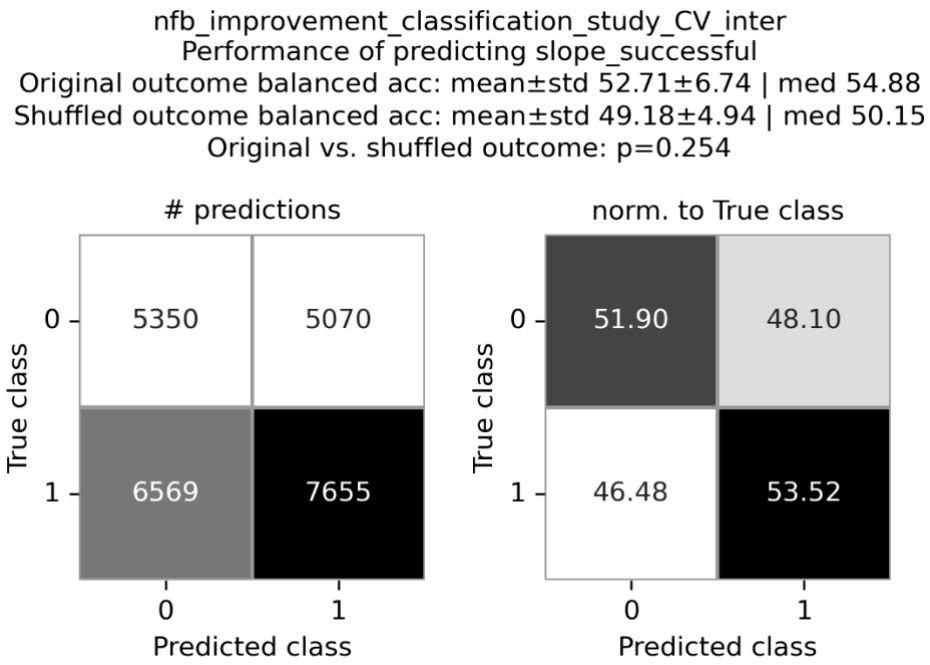

Figure 17 Classification results with per study grouping. The outcome on the original dataset was a mean balamced accurcacy of 52.71 % [SD = 6.74, median = 54.88]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 49.18 % [SD = 4.94, median = 50.15]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in p = 0.254, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback improvement (slope_successful). The left panel shows the number of predictions for each class: 5350 true negatives, 5070 false positives, 6569 false negatives, and 7655 true positives. The right panel presents the normalized values based on the true class distribution, with 51.90 % true negatives, 48.10 % false positives, 46.48 % false negatives, and 53.52 % true positives.

Neurofeedback improvement regression task results

Furthermore, machine learning was also applied in a regression task. The goal was to explain the variance in the slopes of the regression line over all the neurofeedback training values of each run of each participant, i.e. neurofeedback improvement.

Per subject grouping regression task performance

To evaluate the model's ability to predict the slopes with a per subject grouping, its performance using R^2 and Mean Absolute Error (MAE) across CV folds was assessed. The results are summarized in *Figure 18*. The model trained on the original dataset achieved a mean R^2 of -0.008 [SD = 0.403, median = 0.151]. In contrast, when the outcome labels were shuffled, the model's performance dropped to a mean R^2 of -0.021 [SD = 0.015, median = -0.020]. The

statistical comparison between the original and shuffled datasets yielded p = 0.477, indicating a non-significant difference in model performance.

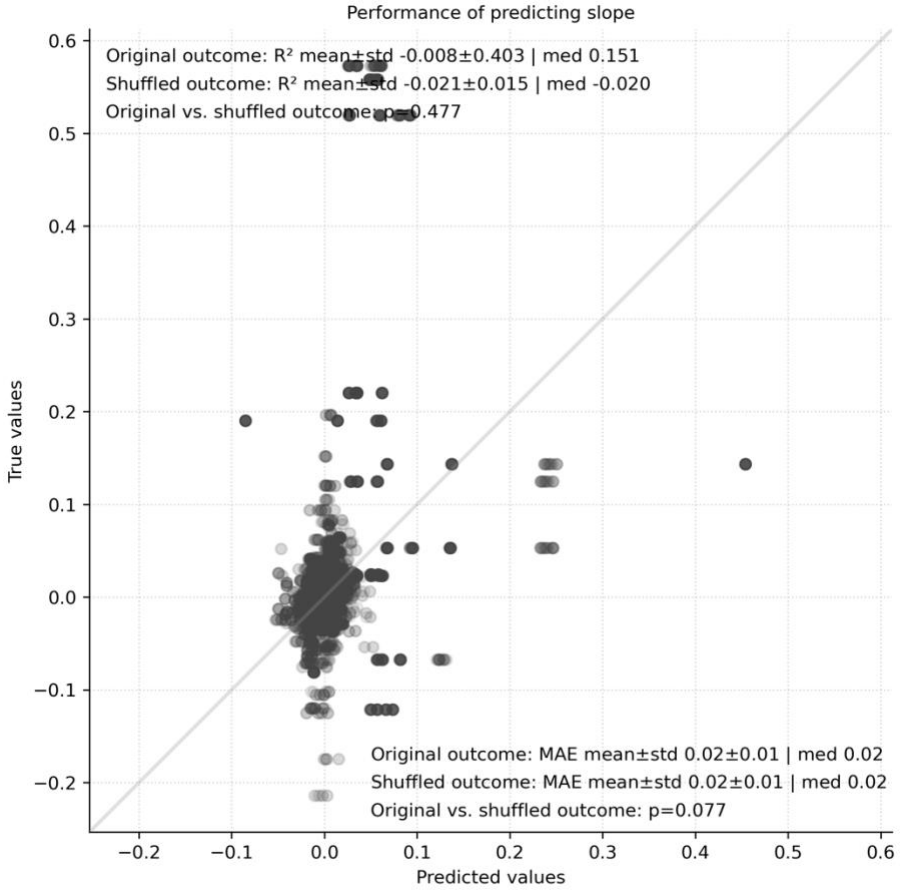

Figure 18 Regression results per subject grouping: Scatter plot displaying predicted vs. true values for slopes. The plot includes performance metrics (R^2 and MAE) for both the original [mean R^2 = -0.008, R^2 SD = 0.403, median R^2 = 0.151, mean MAE = 0.02, MAE SD = 0.01, median MAE = 0.02] and shuffled models [mean R^2 = -0.021, R^2 SD = 0.015, median R^2 = -0.020, mean MAE = 0.02, MAE SD = 0.01, median MAE = 0.02], highlighting a non significant difference in predictive power [p = 0.077].

The absolute prediction errors were also assessed using MAE. The original model resulted in a mean MAE of 0.02 [SD = 0.01, median = 0.02], whereas the shuffled model had a higher MAE of 0.02 [SD = 0.01, median = 0.02]. The results did not significantly differ (p = 0.077).

Per study grouping regression task performance

The model's effectiveness in predicting the slopes with a per study grouping was evaluated in the same way as the per subject grouping. *Figure 19* summarizes these results. The model trained on the original dataset achieved an average R^2 of -0.061 [SD = 0.175, median

= -0.064]. When the outcome labels were shuffled, the model's R^2 decreased to -0.146 [SD = 0.222, median = -0.088]. The statistical comparison between the original and shuffled datasets yielded p = 0.334, indicating a non-significant difference between the models.

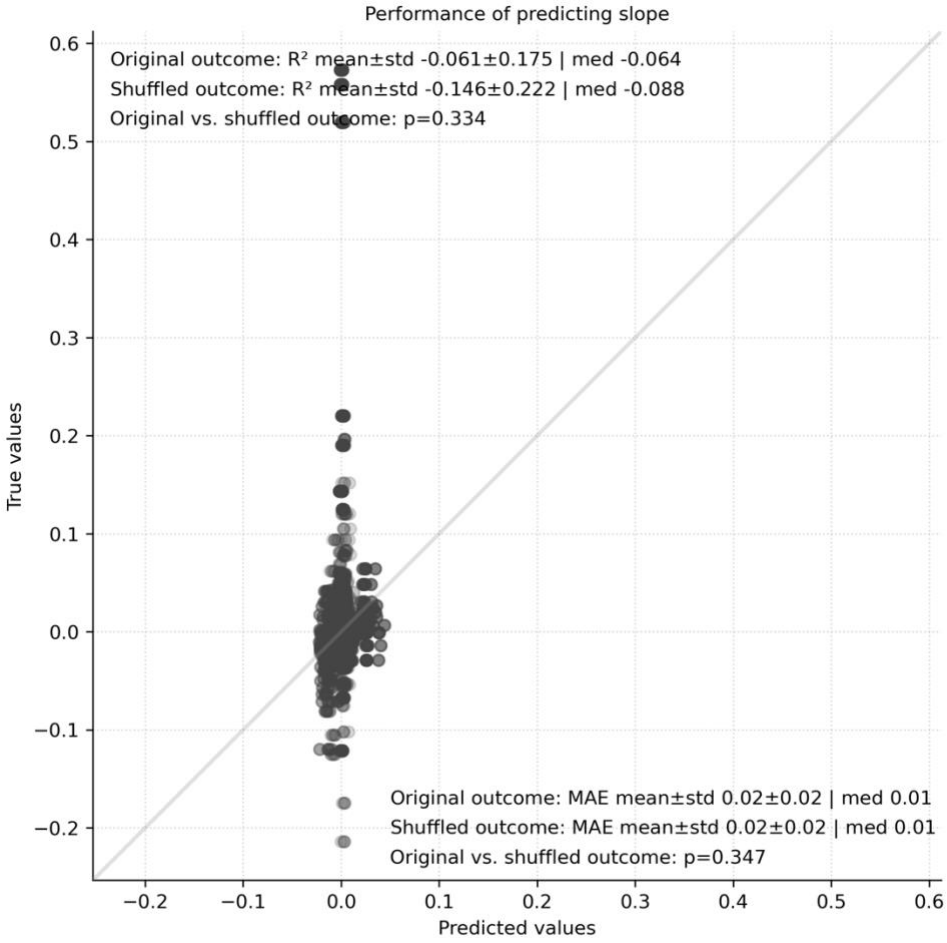

Figure 19 Regression results per study grouping: Scatter plot displaying predicted vs. true values for slopes. The plot includes performance metrics (R^2 and MAE) for both the original [mean R^2 = -0.061, R^2 SD = 0.175, median R^2 = -0.064, mean MAE = 0.02, MAE SD = 0.01, median MAE = 0.02] and shuffled models [mean R^2 = -0.146, R^2 SD = 0.222, median R^2 = -0.088, mean MAE = 0.02, MAE SD = 0.01, median MAE = 0.02], highlighting a non significant difference in predictive power [p = 0.334].

Additionally, prediction errors were assessed using MAE. The original model recorded a mean MAE of 0.02 [SD = 0.02, median = 0.01], whereas the shuffled model exhibited a higher MAE of 0.02 [SD = 0.02, median = 0.01]. These results also did not significantly differ from each other (p = 0.347).

Explorative Analysis: Per subject grouping classification task with study ID as discrete predictor

Neurofeedback performance classification with study ID as discrete predictor

To evaluate the impact of including the study ID on the prediction of neurofeedback performance (successful vs. unsuccessful), a classification model with per subject grouping was trained using nested CV.

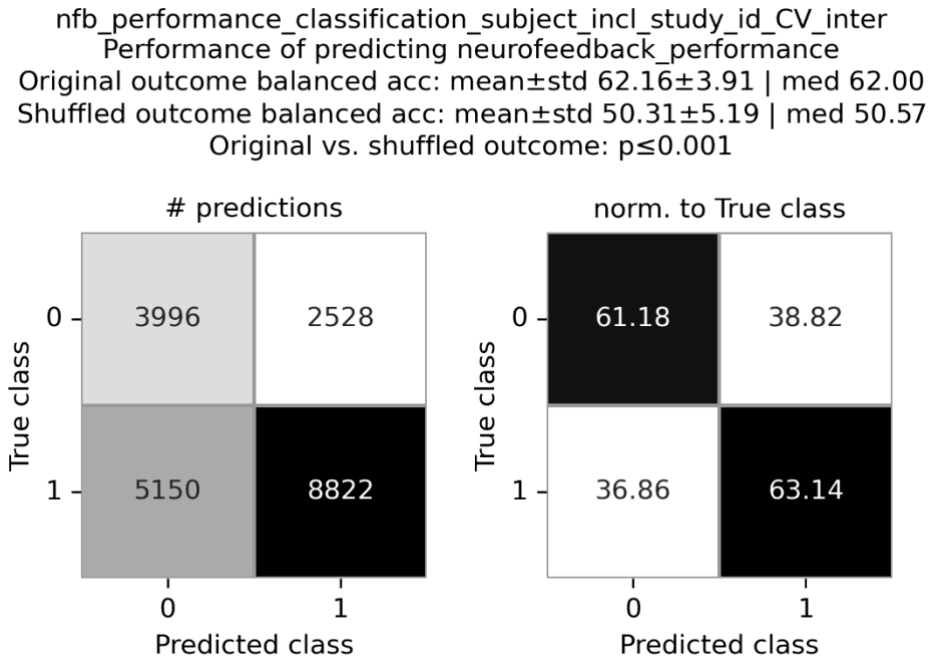

Figure 20 Classification results with per subject grouping and study ID as discrete predictor: The outcome on the original dataset was a mean balamced accurcacy of 62.16 % [SD = 3.91, median = 62.00]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 50.31 % [SD = 5.19, median = 50.57]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in $p \leq 0.001$, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback performance (neurofeedback_performance). The left panel shows the number of predictions for each class: 3936 true negatives, 2528 false positives, 5150 false negatives, and 8822 true positives. The right panel presents the normalized values based on the true class distribution, with 61.18 % true negatives, 38.82 % false positives, 36.86 % false negatives, and 63.14 % true positives.

As illustrated in *Figure 20*, the classifier achieved a mean balanced accuracy of 62.16 % [SD = 3.91, median = 62.00] on the original dataset. When outcome labels were shuffled, the balanced accuracy dropped to 50.31 % [SD = 5.19, median = 50.57]. A corrected version of the paired sample t-test was applied to account for dependencies due to repeated use of the

same data, resulting in a statistically significant difference [$p \leq 0.001$]. These results suggest that the model performed significantly better than chance when study ID was included.

Per subject grouping feature importance analysis

To achieve a better understanding of the contributions of each single feature to the classification decision, SHAP was computed for the classifier's predictions.

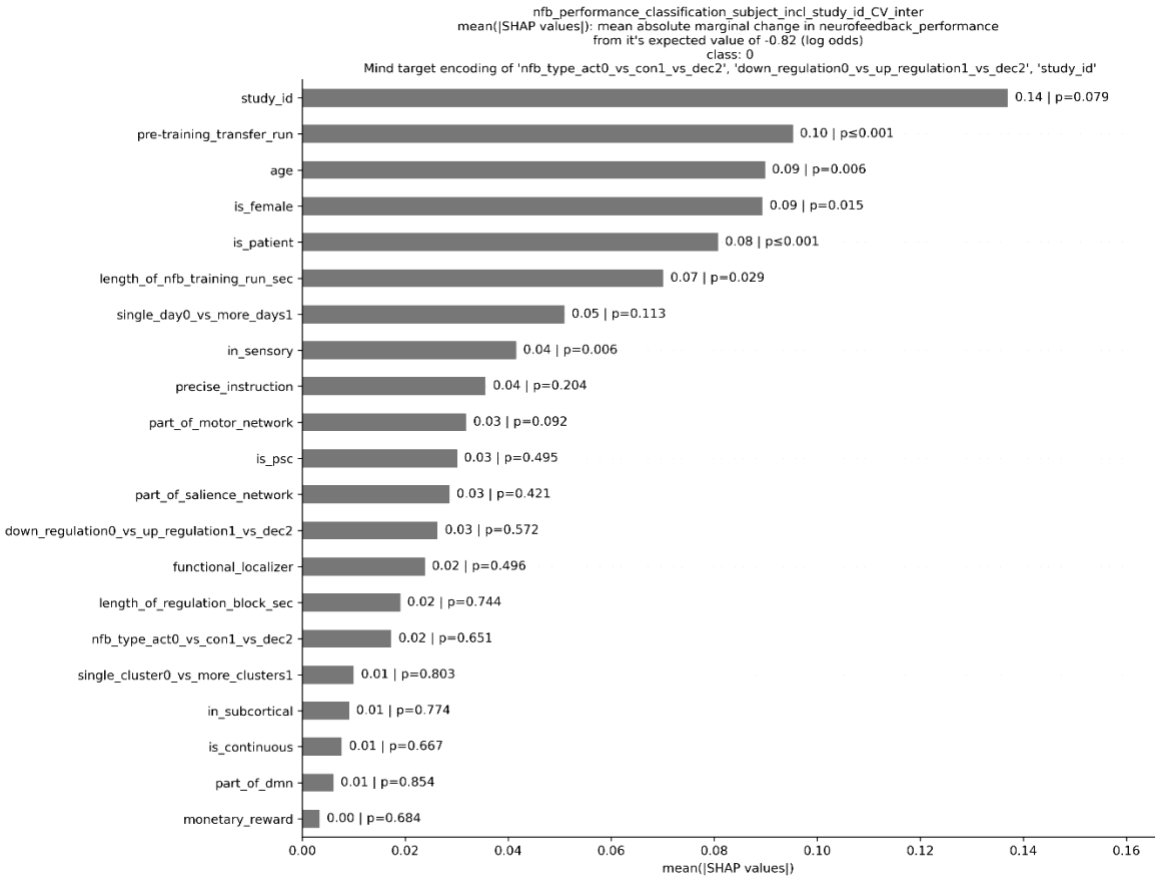

Figure 21 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting a non-successful neurofeedback performance including study ID as discrete predictor. Bar plot displaying the mean absolute SHAP values for the most important features in predicting a non-successful neurofeedback performance (class 0). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include study_id (SHAP ~0.14, p = 0.079), pre_training_transfer_run (SHAP ~0.10, p ≤ 0.001), age (SHAP ~0.09, p = 0.006), and is_female (SHAP ~0.09, p = 0.015).

For the classification of participants as non-successful in neurofeedback performance (class 0), the most influential predictor was the study ID, with a mean absolute SHAP value of 0.14 [$p = 0.079$]. Although the significance threshold was not met, the result suggests that variation across studies was an important factor in the model's decisions. Further important features were the presence or absence of a pre-training transfer run [mean absolute SHAP value

of 0.10, p ≤ 0.001], age [mean absolute SHAP value = 0.09, p = 0.006] and sex [mean absolute SHAP value = 0.09, p = 0.015].

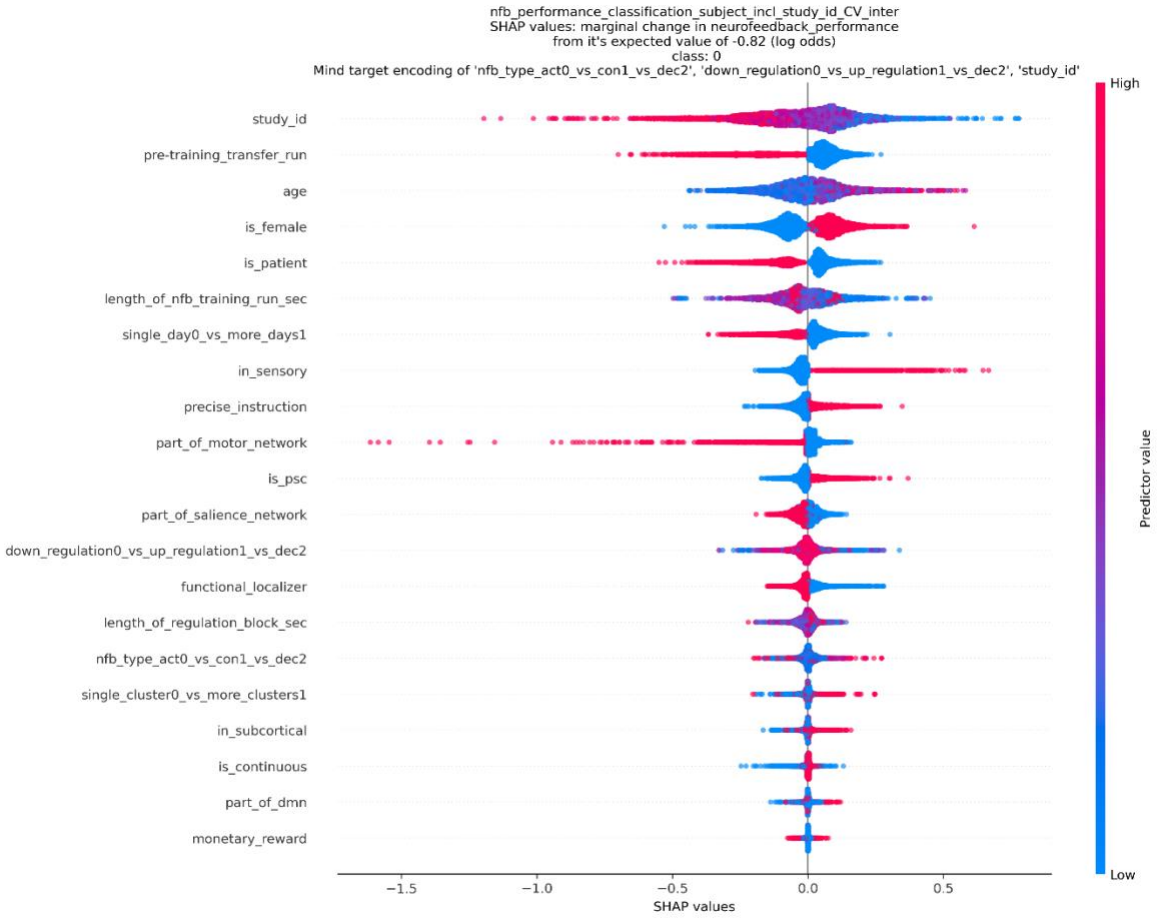

Figure 22 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (non-successful neurofeedback performance with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

When predicting successful neurofeedback performance (class 1), similar but not identical features emerged as important. Once again, study ID showed an influence on the model's decisions [mean absolute SHAP value = 0.13, p = 0.095]. Further important features were age [mean absolute SHAP value = 0.12, p = 0.004], participant health status (is_patient) [mean absolute SHAP value = 0.09, p ≤ 0.001] and the presence of a pre-training transfer run [mean absolute SHAP value = 0.09, p = 0.004].

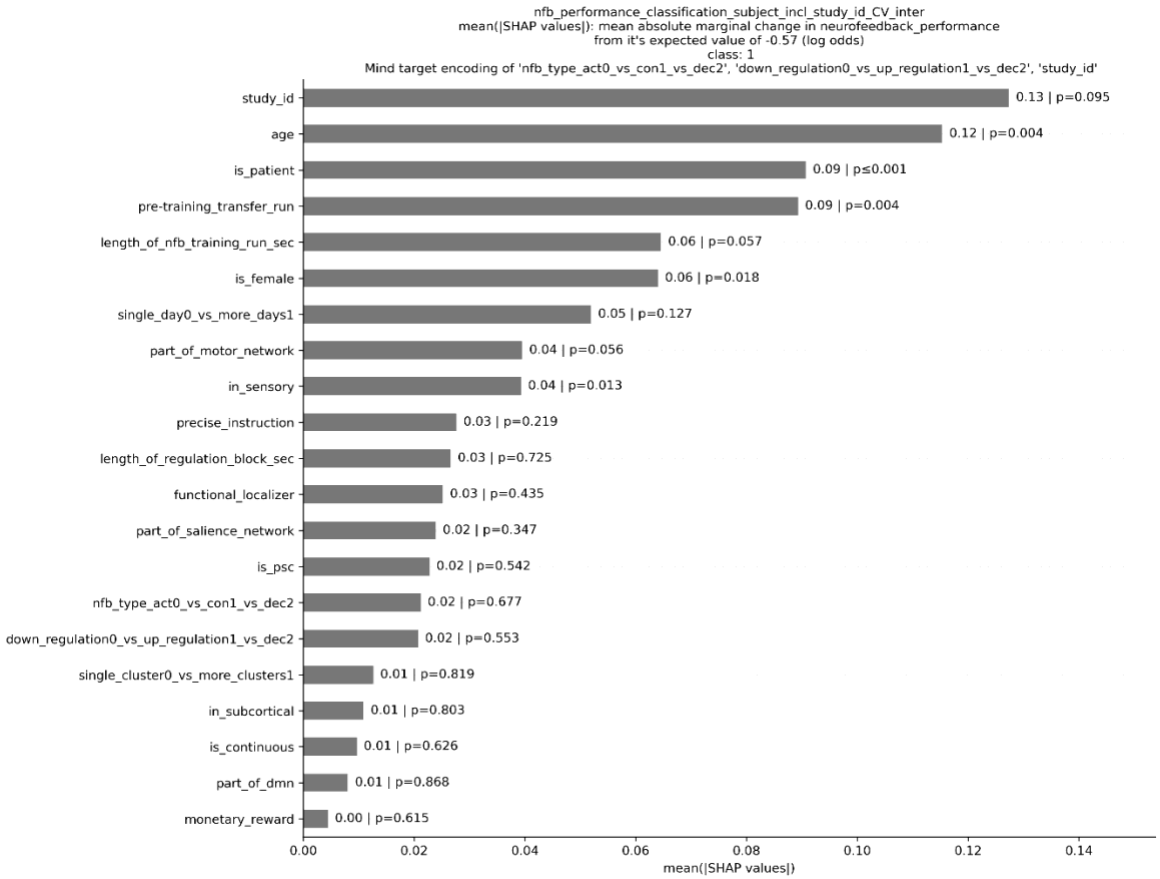

Figure 23 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting a successful neurofeedback performance including study ID as discrete predictor. Bar plot displaying the mean absolute SHAP values for the most important features in predicting a successful neurofeedback performance (class 1). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include study_id (SHAP ~0.13, p = 0.095), age (SHAP ~0.12, p = 0.004), is_patient (SHAP ~0.09, p ≤ 0.001), and pre_training_transfer_run (SHAP ~0.09, p = 0.004).

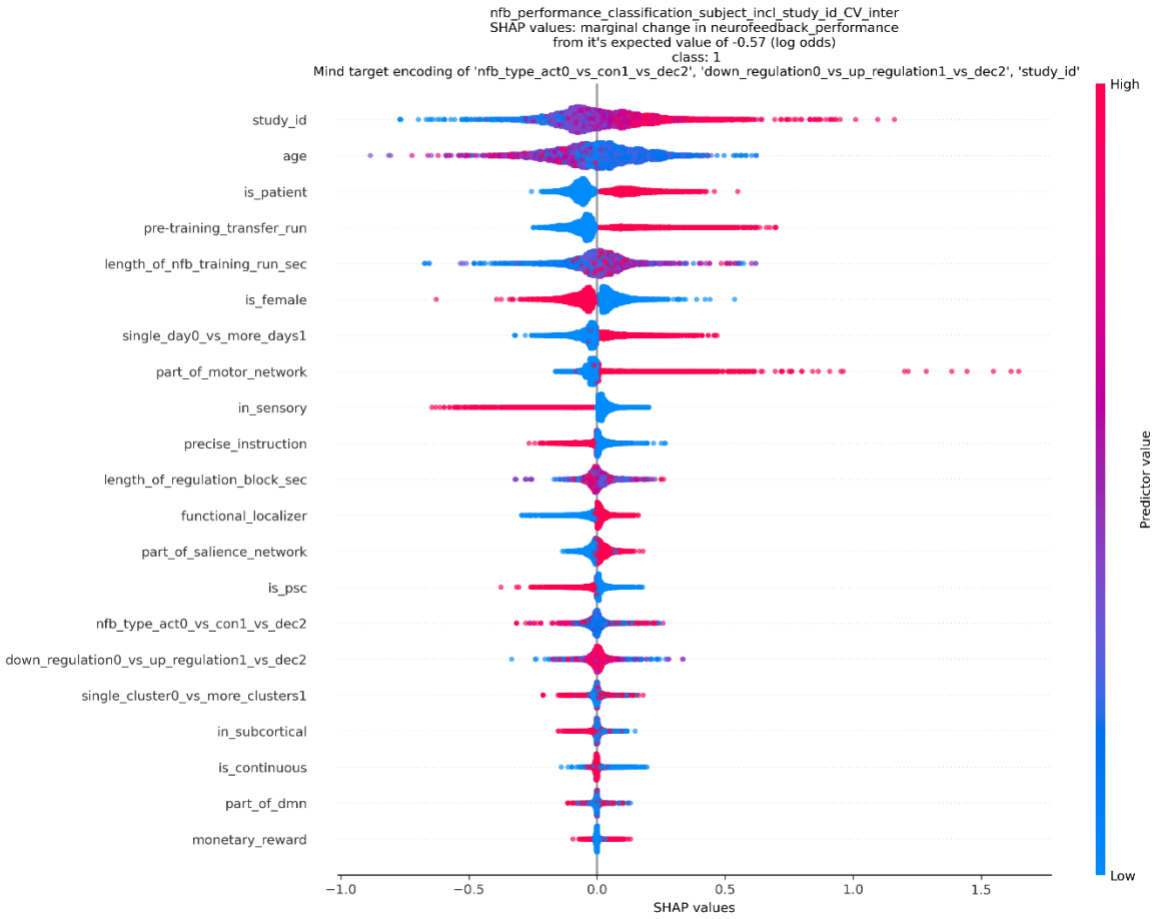

Figure 24 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (successful neurofeedback performance with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

Neurofeedback improvement classification with study ID as discrete predictor

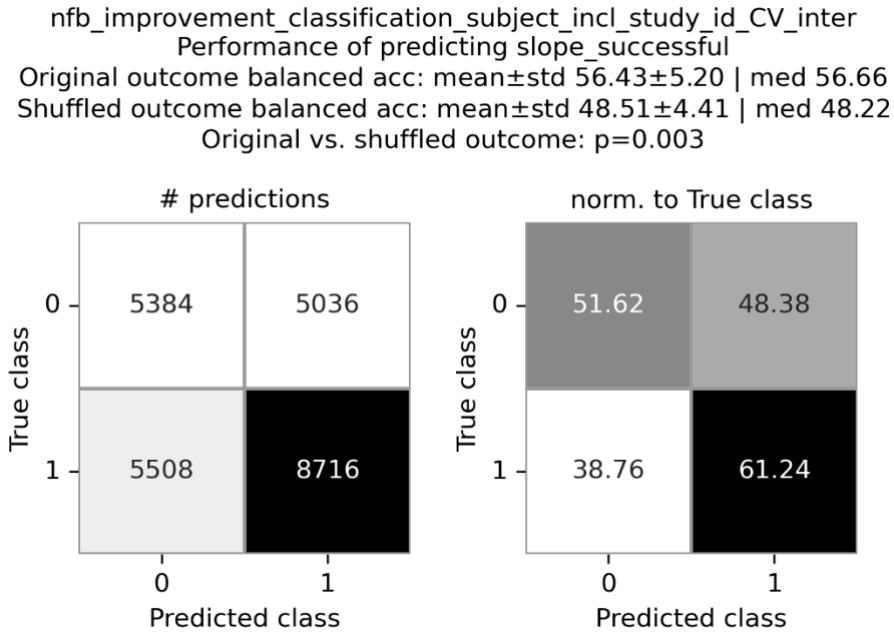

Figure 25 Classification results with per subject grouping and study ID as discrete predictor: The outcome on the original dataset was a mean balamced accurcacy of 56.43 % [SD = 5.20, median = 56.66]. When the outcome labels were shuffled, the mean balanced accuracy dropped to 48.51 % [SD = 4.41, median = 48.22]. A for the lack of independency in the sample corrected version of the paired sample t-test was utilized to compare the original and shuffled outcomes and resulted in p = 0.003, suggesting that the classifier's performance was not significantly better than chance. Confusion matrices displaying classification performance for predicting neurofeedback improvement (slope_successful). The left panel shows the number of predictions for each class: 5384 true negatives, 5036 false positives, 5508 false negatives, and 8716 true positives. The right panel presents the normalized values based on the true class distribution, with 51.62 % true negatives, 48.38 % false positives, 38.76 % false negatives, and 61.24 % true positives.

As shown in *Figure 25*, the model that incorporated study ID as an additional predictor achieved a mean balanced accuracy of 56.43 % [SD = 5.20, median = 56.66] on the original dataset. When outcome labels were shuffled, accuracy decreased to 48.51 % [SD = 4.41, median = 48.22]. A for the lack of independency in the sample corrected version of the paired sample t-test yielded a statistically significant difference [p = 0.003], confirming that the classifier performed significantly better than chance when including the study ID.

The confusion matrices indicate a relatively balanced model performance, with a slight bias toward positive classifications. Specifically, the model correctly identified 51.62 % of true negatives and 61.24 % of true positives.

Per subject grouping feature importance analysis

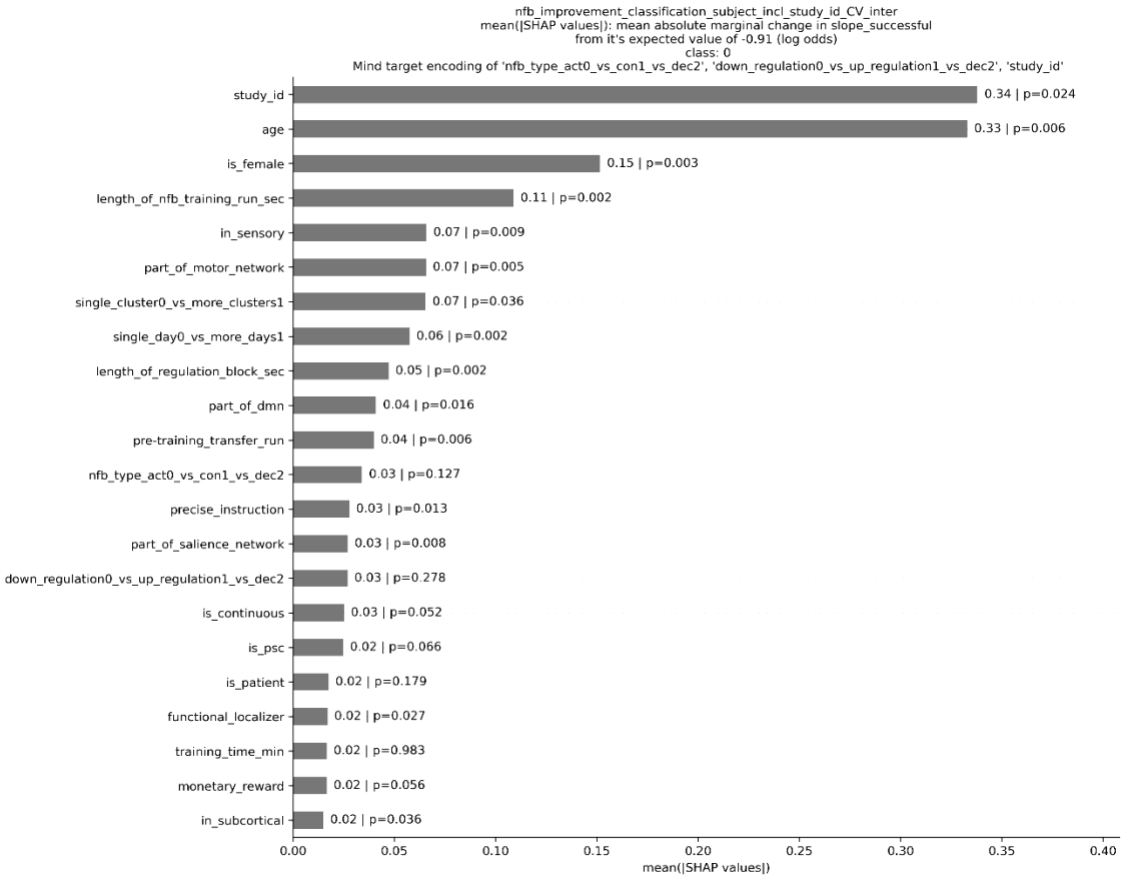

Figure 26 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting no learning success including study ID as discrete predictor. Bar plot displaying the mean absolute SHAP values for the most important features in predicting no neurofeed improvemnet (class 0). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include study_id (SHAP ~0.34, p = 0.024), age (SHAP ~0.33, p = 0.006), and is_female (SHAP ~0.15, p = 0.003).

To interpret how the features contributed to classification decisions, SHAP was computed. When predicting no neurofeedback improvement (class 0), the most impactful feature was the study ID itself, with a mean absolute SHAP value of 0.34 [p = 0.024]. Further important features were age [mean absolute SHAP value = 0.33, p = 0.006] and sex [mean absolute SHAP value = 0.15, p = 0.003].

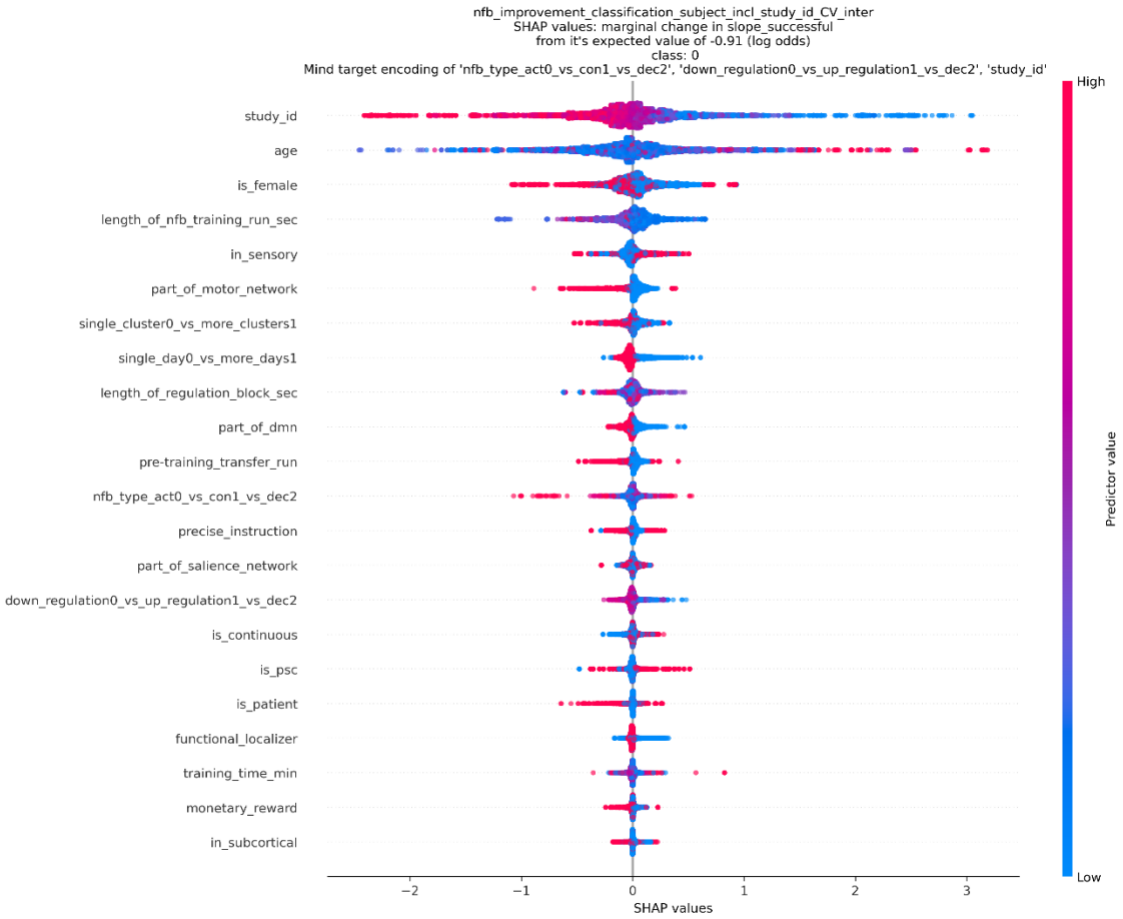

Figure 27 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (no learning success with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

For the successful group (class 1) study ID was the most important feature, with a mean absolute SHAP value of 0.35 [p = 0.020], followed by age [mean absolute SHAP value = 0.30, p = 0.016] and sex [mean absolute SHAP value = 0.15, p = 0.008].

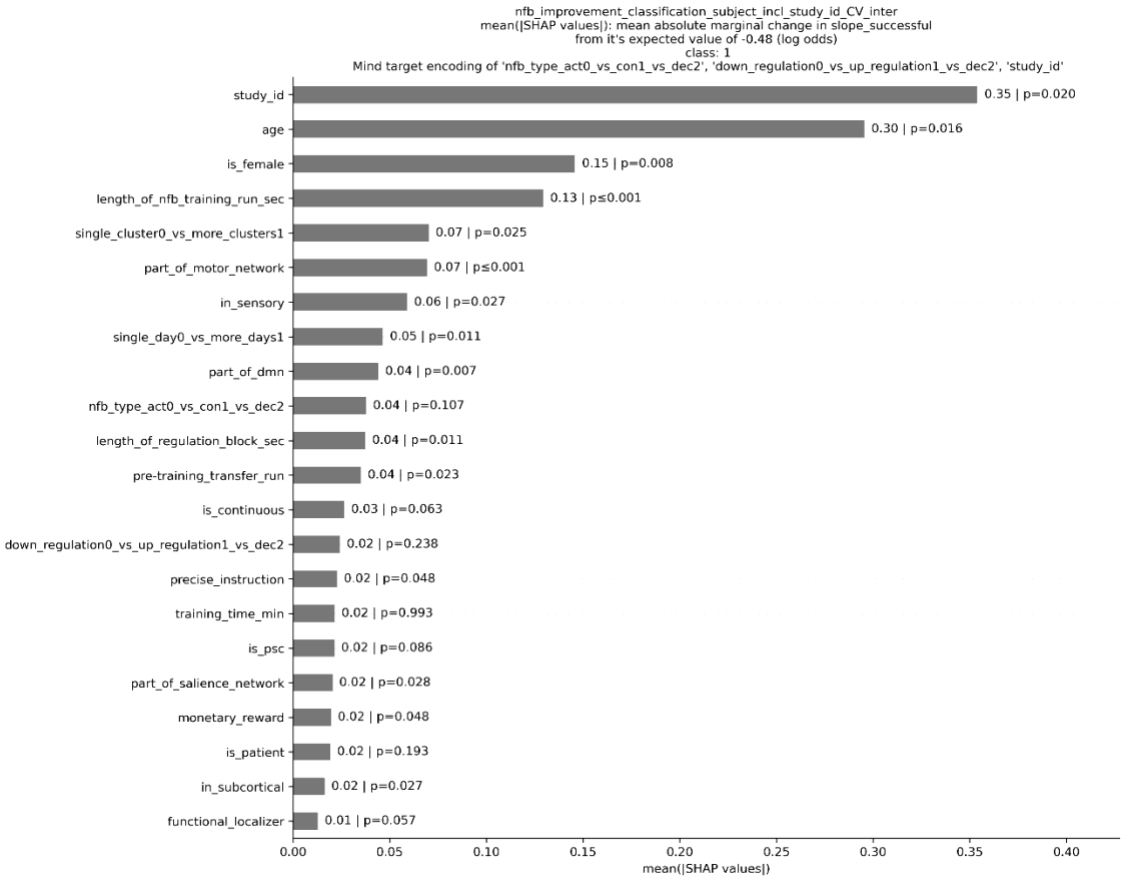

Figure 28 Mean absolute SHAP (SHapley Additive exPlainations) values for the most influential features in the model with a per subject grouping predicting learning success including study ID as discrete predictor. Bar plot displaying the mean absolute SHAP values for the most important features in predicting neurofeed improvemnet (class 1). The p-values indicate the statistical significance of each feature's contribution. The most impactful features include study_id (SHAP ~0.35, p = 0.020), age (SHAP ~0.30, p = 0.016), and is_female (SHAP ~0.15, p = 0.008).

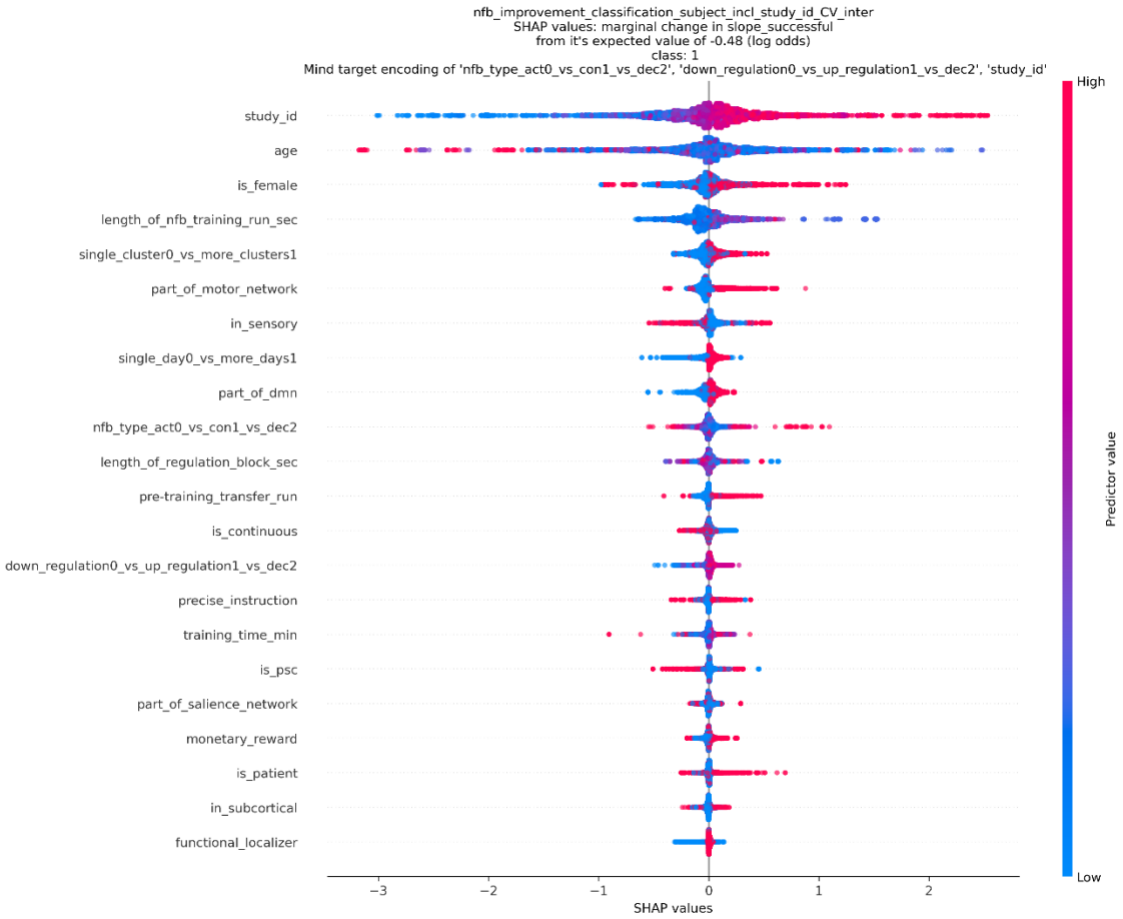

Figure 29 SHAP values (color coded: low = blue, red = high, and displaying non-linearity in associations), depicting the individual impact of features on a specific prediction (learning success with per subject grouping). Visual representation of the impact of the features on the prediction. Both, the directionality of the impact (positive or negative SHAP values) and the degree of influence, depicted by the colour gradient (blue = low influence, red = high influence).

Discussion

This thesis investigated neurofeedback learning success from multiple perspectives. First, it analysed the temporal dynamics and variability of neurofeedback improvement. Next, it examined the influence of 21 different factors on neurofeedback performance and improvement using machine learning, employing both per-subject and per-study groupings. For neurofeedback performance, a classification task was conducted, while for neurofeedback improvement, both a classification task and a regression task were performed. Finally, the thesis explored the impact of including study ID as a discrete feature on predicting neurofeedback performance and improvement using per subject grouping in a classification

task. Study ID was included as it serves as a perfect study indicator, summarising both the individual factors explicitly included in the feature set and those not captured within it.

The classification tasks for both neurofeedback performance and improvement yielded significant results when using per subject grouping. Additionally, both exploratory analyses demonstrated significant predictive performance. In contrast, neither the classification tasks with per study grouping nor the regression tasks achieved predictions above chance level.

These findings align with previous results by Haugg et al. (2021), who reported an average weighted accuracy of 60.3 % [SD = 12.3] when predicting neurofeedback performance, with the most impactful features being the presence of a pre-training transfer run and participants' health status. This reinforces the view that both subject specific and study design factors influence neurofeedback learning. A key distinction between the current work and that of Haugg et al. (2021) is the successful prediction of neurofeedback improvement, which may be attributed to the larger sample size and differences in the applied machine learning methodology.

Temporal dynamics and variability of neurofeedback improvement

The examination of temporal dynamics in this thesis revealed that neurofeedback improvement tended to remain relatively stable across training time when averaged. When comparing different neurofeedback paradigms – activity based neurofeedback, connectivity based neurofeedback and DecNef –, it became evident that training structure and study design influenced data availability throughout sessions. Data drop-off occurred earlier in activity based and connectivity based neurofeedback compared to DecNef, likely reflecting the more standardized protocol in the DecNef studies. Furthermore, it is to note, that there seems to be an upward trend in the mean feedback in activity based and connectivity based neurofeedback, although to a lesser extent in the latter. However, in DecNef no change over time was revealed. This could be due to many reasons. Firstly, it could be due to the fact that participants are

unaware of the content and purpose of the training, which might hinder learning effects. However, this is done with the goal of minimizing noise caused by cognitive processes or knowledge about the manipulated dimensions (Lubianiker et al., 2019; Muñoz-Moldes & Cleeremans, 2020). Another contributing factor may be the greater standardization of DecNef protocols. This standardization likely results in a more homogeneous set of variables, which may in turn reduce the likelihood of incidental factors that could enhance neurofeedback learning, as is more often the case in activity- and connectivity-based paradigms. Importantly, although DecNef does not show improvement over time, it consistently achieves a relatively high level of neurofeedback performance compared to the other paradigms. Nevertheless, the lack of temporal change in DecNef outcomes warrants further investigation in future research.

Features that influenced neurofeedback learning success

The most impactful features for predicting neurofeedback performance were age, health status, length of the neurofeedback run, and the presence of a pre-training transfer run. In contrast, for neurofeedback improvement, the most influential features included age, length of the neurofeedback run, and sex. In the exploratory analysis, the key features for predicting neurofeedback performance were age, sex, health status and the presence of a pre-training transfer run. For neurofeedback improvement, the features with the highest predictive impact were age, sex, length of the neurofeedback training, and study ID. These findings partially replicate previous results by Haugg et al. (2021), who identified the pre-training transfer run and health status as significant predictors of neurofeedback performance. The mentioned features were the most impactful features across multiple models, however other features were significant as well, but their impact on the predictions was very low and will therefore not be discussed in detail.

Pre-training transfer run

Pre-training transfer runs involve attempts to regulate brain activity prior to the actual neurofeedback training, without any feedback (D.-Y. Kim et al., 2015; Kirschner et al., 2018; MacInnes et al., 2016; Young et al., 2017). These runs can fulfil various functions. First, they can familiarize the participants with the scanning environment and regulation task. Secondly, these measurements can provide a baseline for subsequent neurofeedback training (Auer et al., 2015; MacInnes et al., 2016). The results of this thesis suggest that the presence of a pre-training transfer run is associated with a higher likelihood of successful neurofeedback learning. These findings align with results from Haugg et al. (2021), who suggested that prior familiarisation with the task and additional practice during these runs may enhance participants' neurofeedback learning success.

Health status

The results of the analyses suggest that patients tend to perform better in neurofeedback learning compared to healthy participants, which is consistent with the findings of Haugg et al. (2021). One possible explanation would be, that new paradigms are often piloted on healthy participants before clinical trials. Another potential explanation is a difference in motivation: patients typically undergo neurofeedback training with the goal of alleviating clinical symptoms, whereas healthy participants often engage in neurofeedback studies out of general interest or for monetary compensation. Finally, patients may exhibit greater neurofeedback learning due to their often more dysfunctional brain patterns, which reduces the likelihood of ceiling effects compared to healthy participants. A more detailed examination of the differences between different clinical populations would be valuable, as cognitive impairments affecting attention and self-regulation abilities may vary across different disorders (Heeren et al., 2014; Li et al., 2010; Lussier & Stip, 2001).

Age

The results of age are more complex: in neurofeedback performance younger participants seem to perform better compared to older participants. In neurofeedback performance age has a high impact as well, the direction of the effect however is mixed. In general the effects of age on neurofeedback learning have not been widely investigated. One study found differences in the effect of neurofeedback training among groups of adolescence of different ages. They reported positive effects on brain activation, attention and stress resistance which were linked to neurodevelopmental characteristics in adolescence (Byun, 2011). However, a systematic review on the effects of neurofeedback on aging related cognitive decline found no significant positive effect on attention, but did report improvements across other examined cognitive domains, including memory, executive functions and fluid intelligence (Laborda-Sánchez & Cansino, 2021). Another study compared a younger group of participants (mean age of 23) with an older group (mean age of 69) using near-infrared spectroscopy neurofeedback and found differences in the within-session trainability of neural responses with younger participants showing a superior ability to modulate their brain signals (Kober et al., 2019). Further research on the effects of age and aging on fMRI neurofeedback learning should focus on differences between groups with large age gape and their performance over longer time periods.

Sex

The results of the analyses suggest that in neurofeedback performance, men tend to perform better than women. However, the opposite pattern was observed for neurofeedback improvement, with women showing a higher likelihood of success, although this effect was less pronounced. To date, sex differences in neurofeedback learning success have been investigated to a very limited extent. A study from 2010 found significant differences in training effects on the prefrontal lobe between male and female adolescents (Byun & Park,

2011). Due to the low number of studies, additional research is needed to better understand the role of sex in neurofeedback learning.

Length of neurofeedback training run

The results of this thesis suggest that longer neurofeedback training runs tend to be associated with a higher likelihood of neurofeedback learning success, although the directionality of this effect is not strongly pronounced. To date, the specific impact of the length of neurofeedback training runs has not been a major focus of research. However, a systematic review and meta-analysis on the effects of neurofeedback on posttraumatic stress disorder symptoms found that more frequent and longer training sessions resulted in stronger positive outcomes compared to fewer, more condensed sessions, which is consistent with the findings of this thesis (Choi et al., 2023). Optimizing the length of neurofeedback training could improve its overall efficacy while also helping to minimise scanning-related costs.

Study ID

The prominent role of study ID in predicting both neurofeedback performance and neurofeedback improvement underscores the substantial impact that study-specific factors might have on neurofeedback learning. It was the most important feature when included in the analysis; however, it is important to note that although study ID had the highest SHAP value of all the features when predicting neurofeedback performance and improvement. However, it is important to note that despite its high SHAP values, study ID was not a significant predictor of neurofeedback performance. This highlights the need for further investigation into the role of specific study-related factors, as study ID serves as a perfect study indicator, summarising all individual factors of a study, including those not explicitly captured in the feature set. Moreover, the study-specific factors explicitly included in the analyses of this thesis were also encompassed within study ID, which likely explains why it emerged as the most prominent feature in the exploratory analysis.

Together, these findings suggest that both individual and experimental design variables are critical for neurofeedback learning success. Nonetheless, some researchers have argued that neurofeedback effects may be attributable to placebo responses or physiological artefacts (Thibault et al., 2016, 2017, 2018). However, the results of this thesis offer some evidence, that neurofeedback training may have genuine effects beyond these alternative explanations.

Limitations

There are many limitations this thesis and the collected dataset faced.

Impact of significant features

Although many features were statistically significant, their impact on the predictions was limited. This applied not only to variables excluded from the discussion but also to those included. For example, in the exploratory analysis, the most impactful feature for predicting neurofeedback performance success was the study ID, with a mean absolute SHAP value of 0.35. The expected value for predicting neurofeedback success was -0.48 in log odds, indicating that, on average, the study ID shifted the log odds by 0.35 from the expected value to -0.13 log odds. In terms of probability, this corresponded to a shift from 0.38 to 0.47 in the probability of predicting neurofeedback performance success.

Publication bias and data sharing

As with any meta-level research, this thesis faced several limitations. A major challenge was the potential bias in the included studies. Studies with non-significant results face the publication bias – i.e. they are less likely to get published in a scientific journal –, and their authors may also be less inclined to share their data with other researchers. Since this thesis was heavily reliant on the results of multiple studies and the authors' willingness to share their collected per subject data, such biases could have affected the analyses. However, increasing adoption of open science practices may mitigate this in future research, enabling broader

assessment of neurofeedback learning. Even with access to the raw imaging data and a re-analysis of all studies with an identical pipeline, discrepancies would remain, as the reanalysed data may not reflect the actual feedback presented to the participants during the original studies.

Heterogeneity in feedback metrics and study design

Heterogeneity in feedback metrics (e.g., prevent signal change, beta values, etc.) further complicates cross-study comparisons. Participant attrition and dropout may also introduce bias, as individuals who do not succeed in neurofeedback training may be more likely to discontinue participation.

Non recorded potentially relevant factors

Another limitation concerns the feature set, which was limited to 21 features (excluding study ID) and unknown confounding variables may exist. For example, real time cognitive strategies employed by the participants were not captured, though they could have significant predictive value. Other potentially relevant subject specific variables include biological indicators (e.g., heart rate, stress level, skin conductance), personality traits and cognitive characteristics, such as intelligence, mental imagery, attention and motivation (Kadosh & Staunton, 2019). Additionally, brain-based measures such as eigenvector centrality, grey matter volume and the connectivity between target regions and higher order brain areas may influence neurofeedback learning (Bassett & Khambhati, 2017; Skouras & Scharnowski, 2019; Zhao et al., 2021).

Neural activity due to cognitive tasks

Another limitation relates to the definition of neurofeedback performance. This metric did not control for the potential activation of target brain regions resulting from the performance of a cognitive task. Few studies in the field of neurofeedback include control conditions in which participants perform the cognitive task without receiving feedback, while

undergoing a fMRI scan. However, Haugg et al. (2020) found that neurofeedback learning appears to go beyond simple task-induced activation, as brain activity during pre-training transfer runs was not correlated with neurofeedback success. This suggests that the feedback provided during the training might have a higher impact on neurofeedback learning than the cognitive task performed in the MR scanner itself.

Definition of neurofeedback success

The outcomes presented in this thesis may be influenced not only by the selected features but also by the definition and measurement of neurofeedback success. Currently, there is no universally accepted metric for assessing neurofeedback learning success, and measures of success vary considerably across studies (Haugg et al., 2020; Paret et al., 2019; Thibault et al., 2018). Neurofeedback values can take many forms (e.g. percent signal change, beta values, and connectivity measures, etc.), which complicates the definition of a common target for machine learning analyses. One possible solution for pooling these heterogeneous feedback values and therefore increasing statistical power and generalizability was proposed by Haugg et al. (2021) and also applied in this thesis – neurofeedback performance and neurofeedback improvement. Further improvements in the heterogeneity of neurofeedback success measures may be achieved through the development and adoption of a universally accepted model of neurofeedback learning. However, the underlying mechanisms of neurofeedback are not yet fully understood (Kadosh & Staunton, 2019; Sitaram et al., 2017), making it challenging to identify the key components of neurofeedback learning necessary to establish a comprehensive and standardized success measure.

Conclusion

This thesis provided a comprehensive examination of neurofeedback learning success, analysing temporal dynamics, subject specific and study specific factors influencing

neurofeedback performance and improvement across a large multi-study dataset. It demonstrated that while the prediction of neurofeedback performance and improvement is feasible at levels above chance, the effect individual features have on the prediction remain modest. The findings suggest that both subject specific (e.g., age, sex, health status) and study specific factors (e.g., length of neurofeedback run, presence of a pre-training transfer run) possibly shape neurofeedback learning success. Notably, study ID was the feature with the highest SHAP value when predicting neurofeedback improvement in the exploratory analysis, underscoring the impact of study specific factors – including factors not captured within the feature set of this thesis. Collectively, these results picture some of the complicated mechanisms behind neurofeedback learning and show, that further research is needed to gain a better understanding of these processes.

Outlook

Future research should explore the integration of real-time cognitive strategy data, enabling the assessment of how participants' mental approaches during training influence neurofeedback learning success. Additionally, utilising larger and more diverse datasets will improve the generalisability of findings across different populations and neurofeedback paradigms. Furthermore, the application of advanced machine learning methods, including deep learning and transformer-based models, may uncover complex, non-linear patterns in the data, potentially enhancing predictive performance beyond the current models. To further improve the understanding of neurofeedback learning, it will be essential to systematically investigate a broader range of study specific factors, clarifying how variations in protocol design, feedback modalities, and training structures affect learning outcomes. Together, these advancements could contribute to the development of more effective and personalised neurofeedback training protocols in both clinical and non-clinical settings.

References

Alkoby, O., Abu-Rmileh, A., Shriki, O., & Todder, D. (2018). Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning. *Neuroscience*, *378*, 155–164. https://doi.org/10.1016/j.neuroscience.2016.12.050

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., & Watanabe, T. (2016). Learning to Associate Orientation with Color in Early Visual Areas by Associative Decoded fMRI Neurofeedback. *Current Biology*, *26*(14), 1861–1866. https://doi.org/10.1016/j.cub.2016.05.014

Auer, T., Schweizer, R., & Frahm, J. (2015). Training Efficiency and Transfer Success in an Extended Real-Time Functional MRI Neurofeedback Training of the Somatomotor Cortex of Healthy Subjects. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00547

Bassett, D. S., & Khambhati, A. N. (2017). A network engineering perspective on probing and perturbing cognition with neurofeedback. *Annals of the New York Academy of Sciences*, *1396*(1), 126–143. https://doi.org/10.1111/nyas.13338

Bauer, C. C. C., Okano, K., Ghosh, S. S., Lee, Y. J., Melero, H., Angeles, C. D. L., Nestor, P. G., Del Re, E. C., Northoff, G., Niznikiewicz, M. A., & Whitfield-Gabrieli, S. (2020). Real-time fMRI neurofeedback reduces auditory hallucinations and modulates resting state connectivity of involved brain regions: Part 2: Default mode network -preliminary evidence. *Psychiatry Research*, *284*, 112770. https://doi.org/10.1016/j.psychres.2020.112770

Berman, A. J. L., Grissom, W. A., Witzel, T., Nasr, S., Park, D. J., Setsompop, K., & Polimeni, J. R. (2021). Ultra-high spatial resolution BOLD fMRI in humans using combined segmented-accelerated VFA-FLEET with a recursive RF pulse design. *Magnetic Resonance in Medicine*, *85*(1), 120–139. https://doi.org/10.1002/mrm.28415

Bouckaert, R. R., & Frank, E. (2004). Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 3–12). Springer. https://doi.org/10.1007/978-3-540-24775-3_3

Buyukturkoglu, K., Rana, M., Ruiz, S., Hackley, S. A., Soekadar, S. R., Birbaumer, N., & Sitaram, R. (2013). Volitional regulation of the supplementary motor area with fMRI-BCI neurofeedback in Parkinson's disease: A pilot study. *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 677–681. https://doi.org/10.1109/NER.2013.6696025

Byun, Y.-E. (2011). The Effect of Neurofeedback Training on Age differences groups in Adolescence. *Journal of the Korea Academia-Industrial cooperation Society*, *12*(6), 2561–2566. https://doi.org/10.5762/KAIS.2011.12.6.2561

Byun, Y.-E., & Park, P.-W. (2011). The Effect of Neurofeedback Training on Sex differences groups in Adolescence. *Journal of the Korea Academia-Industrial cooperation Society*, *12*(3), 1171–1177. https://doi.org/10.5762/KAIS.2011.12.3.1171

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Campbell, T. W., Roder, H., Georgantas III, R. W., & Roder, J. (2022). Exact Shapley values for local and model-true explanations of decision tree ensembles. *Machine Learning with Applications*, *9*, 100345. https://doi.org/10.1016/j.mlwa.2022.100345

Cawley, G., & Talbot, N. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107.

Choi, Y.-J., Choi, E.-J., & Ko, E. (2023). Neurofeedback Effect on Symptoms of Posttraumatic Stress Disorder: A Systematic Review and Meta-Analysis. *Applied Psychophysiology and Biofeedback*, *48*(3), 259–274. https://doi.org/10.1007/s10484-023-09593-3

Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, *7*(1), 13669. https://doi.org/10.1038/ncomms13669

Cortese, A., Amano, K., Koizumi, A., Lau, H., & Kawato, M. (2017). Decoded fMRI neurofeedback can induce bidirectional confidence changes within single participants. *NeuroImage*, *149*, 323–337. https://doi.org/10.1016/j.neuroimage.2017.01.069

Cortese, A., Tanaka, S. C., Amano, K., Koizumi, A., Lau, H., Sasaki, Y., Shibata, K., Taschereau-Dumouchel, V., Watanabe, T., & Kawato, M. (2021). The DecNef collection, fMRI data from closed-loop decoded neurofeedback experiments. *Scientific Data*, *8*(1), 65. https://doi.org/10.1038/s41597-021-00845-7

Coutanche, M. N., & Hallion, L. S. (2020). Machine Learning for Clinical Psychology and Clinical Neuroscience. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology* (pp. 467–482). Cambridge University Press. https://doi.org/10.1017/9781316995808.041

Cox, R. W., Jesmanowicz, A., & Hyde, J. S. (1995). Real-Time Functional Magnetic Resonance Imaging. *Magnetic Resonance in Medicine*, *33*(2), 230–236. https://doi.org/10.1002/mrm.1910330213

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*(3), 470–475. https://doi.org/10.1038/nn.3940

deCharms, R. C., Maeda, F., Glover, G. H., Ludlow, D., Pauly, J. M., Soneji, D., Gabrieli, J. D. E., & Mackey, S. C. (2005). Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences*, *102*(51), 18626–18631. https://doi.org/10.1073/pnas.0505210102

Emmert, K., Kopel, R., Koush, Y., Maire, R., Senn, P., Van De Ville, D., & Haller, S. (2017). Continuous vs. Intermittent neurofeedback to regulate auditory cortex activity of tinnitus patients using real-time fMRI - A pilot study. *NeuroImage: Clinical*, *14*, 97–104. https://doi.org/10.1016/j.nicl.2016.12.023

García, V., Mollineda, R. A., & Sánchez, J. S. (2009). Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In H. Araujo, A. M. Mendonça, A. J. Pinho, & M. I. Torres (Eds.), *Pattern Recognition and Image Analysis* (pp. 441–448). Springer. https://doi.org/10.1007/978-3-642-02172-5_57

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Hampson, M., Ruiz, S., & Ushiba, J. (2020). Neurofeedback. *NeuroImage*, *218*, 116473. https://doi.org/10.1016/j.neuroimage.2019.116473

Hanlon, C. A., Hartwell, K. J., Canterberry, M., Li, X., Owens, M., LeMatty, T., Prisciandaro, J. J., Borckardt, J., Brady, K. T., & George, M. S. (2013). Reduction of cue-induced craving through realtime neurofeedback in nicotine users: The role of region of interest selection and multiple visits. *Psychiatry Research: Neuroimaging*, *213*(1), 79–81. https://doi.org/10.1016/j.pscychresns.2013.03.003

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Haugg, A., Renz, F. M., Nicholson, A. A., Lor, C., Götzendorfer, S. J., Sladky, R., Skouras, S., McDonald, A., Craddock, C., Hellrung, L., Kirschner, M., Herdener, M., Koush, Y., Papoutsi, M., Keynan, J., Hendler, T., Cohen Kadosh, K., Zich, C., Kohl, S. H., … Steyrl, D. (2021). Predictors of real-time fMRI neurofeedback performance and improvement – A machine learning mega-analysis. *NeuroImage*, *237*, 118207. https://doi.org/10.1016/j.neuroimage.2021.118207

Haugg, A., Sladky, R., Skouras, S., McDonald, A., Craddock, C., Kirschner, M., Herdener, M., Koush, Y., Papoutsi, M., Keynan, J. N., Hendler, T., Cohen Kadosh, K., Zich, C., MacInnes, J., Adcock, R. A., Dickerson, K., Chen, N., Young, K., Bodurka, J., … Scharnowski, F. (2020). Can we predict real-time FMRI neurofeedback learning success from pretraining brain activity? *Human Brain Mapping*, *41*(14), 3839–3854. https://doi.org/10.1002/hbm.25089

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Heeren, A., Maurage, P., Perrot, H., De Volder, A., Renier, L., Araneda, R., Lacroix, E., Decat, M., Deggouj, N., & Philippot, P. (2014). Tinnitus specifically alters the top-down executive control sub-component of attention: Evidence from the Attention Network Task. *Behavioural Brain Research*, *269*, 147–154. https://doi.org/10.1016/j.bbr.2014.04.043

Hellrung, L., Borchardt, V., Götting, F. N., Stadler, J., Tempelmann, C., Tobler, P. N., Walter, M., & Meer, J. N. van der. (2018). *Motion and physiological noise effects on amygdala real-time fMRI neurofeedback learning* (p. 366138). bioRxiv. https://doi.org/10.1101/366138

Hellrung, L., Dietrich, A., Hollmann, M., Pleger, B., Kalberlah, C., Roggenhofer, E., Villringer, A., & Horstmann, A. (2018). Intermittent compared to continuous real-time fMRI neurofeedback boosts control over amygdala activation. *NeuroImage*, *166*, 198–208. https://doi.org/10.1016/j.neuroimage.2017.10.031

Hellrung, L., Kirschner, M., Sulzer, J., Sladky, R., Scharnowski, F., Herdener, M., & Tobler, P. N. (2022). Analysis of individual differences in neurofeedback training illuminates successful self-regulation of the dopaminergic midbrain. *Communications Biology*, *5*(1), 1–13. https://doi.org/10.1038/s42003-022-03756-4

Hirsch, J., Ruge, M. I., Kim, K. H. S., Correa, D. D., Victor, J. D., Relkin, N. R., Labar, D. R., Krol, G., Bilsky, M. H., Souweidane, M. M., DeAngelis, L. M., & Gutin, P. H. (2000). An Integrated Functional Magnetic Resonance Imaging Procedure for Preoperative Mapping of Cortical Areas Associated with Tactile, Motor, Language, and Visual Functions. *Neurosurgery*, *47*(3), 711.

Johnson, K. A., Hartwell, K., LeMatty, T., Borckardt, J., Morgan, P. S., Govindarajan, K., Brady, K., & George, M. S. (2012). Intermittent "Real-time" fMRI Feedback Is Superior to Continuous Presentation for a Motor Imagery Task: A Pilot Study. *Journal of Neuroimaging*, *22*(1), 58–66. https://doi.org/10.1111/j.1552-6569.2010.00529.x

Kadosh, K. C., & Staunton, G. (2019). A systematic review of the psychological factors that influence neurofeedback learning outcomes. *NeuroImage*, *185*, 545–555. https://doi.org/10.1016/j.neuroimage.2018.10.021

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kamiya, J. (2011). The First Communications About Operant Conditioning of the EEG. *Journal of Neurotherapy*, *15*(1), 65–73. https://doi.org/10.1080/10874208.2011.545764

Keynan, J. N., Cohen, A., Jackont, G., Green, N., Goldway, N., Davidov, A., Meir-Hasson, Y., Raz, G., Intrator, N., Fruchter, E., Ginat, K., Laska, E., Cavazza, M., & Hendler, T. (2019). Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature Human Behaviour*, *3*(1), 63–73. https://doi.org/10.1038/s41562-018-0484-3

Kim, D.-Y., Yoo, S.-S., Tegethoff, M., Meinlschmidt, G., & Lee, J.-H. (2015). The Inclusion of Functional Connectivity Information into fMRI-based Neurofeedback Improves Its Efficacy in the Reduction of Cigarette Cravings. *Journal of Cognitive Neuroscience*, *27*(8), 1552–1572. https://doi.org/10.1162/jocn_a_00802

Kim, S.-G., & Bandettini, P. A. (2010). Principles of Functional MRI. In S. H. Faro & F. B. Mohamed (Eds.), *BOLD fMRI: A Guide to Functional Imaging for Neuroscientists* (pp. 3–22). Springer. https://doi.org/10.1007/978-1-4419-1329-6_1

Kim, S.-G., & Ogawa, S. (2012). Biophysical and Physiological Origins of Blood Oxygenation Level-Dependent fMRI Signals. *Journal of Cerebral Blood Flow & Metabolism*, *32*(7), 1188–1206. https://doi.org/10.1038/jcbfm.2012.23

Kirschner, M. (2017). Self-regulation of the Dopaminergic Reward System Via Real Time fmri Neurofeedback in Schizophrenia. *European Psychiatry*, *41*(S1), S58–S58. https://doi.org/10.1016/j.eurpsy.2017.01.041

Kirschner, M., Sladky, R., Haugg, A., Stämpfli, P., Jehli, E., Hodel, M., Engeli, E., Hösli, S., Baumgartner, M. R., Sulzer, J., Huys, Q. J. M., Seifritz, E., Quednow, B. B., Scharnowski, F., & Herdener, M. (2018). Self-regulation of the dopaminergic reward circuit in cocaine users with mental imagery and neurofeedback. *eBioMedicine*, *37*, 489–498. https://doi.org/10.1016/j.ebiom.2018.10.052

Kober, S. E., Spörk, R., Bauernfeind, G., & Wood, G. (2019). Age-related differences in the within-session trainability of hemodynamic parameters: A near-infrared spectroscopy–based neurofeedback study. *Neurobiology of Aging*, *81*, 127–137. https://doi.org/10.1016/j.neurobiolaging.2019.05.022

Kohl, S. H., Veit, R., Spetter, M. S., Günther, A., Rina, A., Lührs, M., Birbaumer, N., Preissl, H., & Hallschmid, M. (2019). Real-time fMRI neurofeedback training to improve eating behavior by self-regulation of the dorsolateral prefrontal cortex: A randomized controlled trial in overweight and obese subjects. *NeuroImage*, *191*, 596–609. https://doi.org/10.1016/j.neuroimage.2019.02.033

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., & Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses

conscious exposure. *Nature Human Behaviour*, *1*(1), 0006. https://doi.org/10.1038/s41562-016-0006

Koush, Y., Meskaldji, D.-E., Pichon, S., Rey, G., Rieger, S. W., Linden, D. E. J., Van De Ville, D., Vuilleumier, P., & Scharnowski, F. (2015). Learning Control Over Emotion Networks Through Connectivity-Based Neurofeedback. *Cerebral Cortex*, bhv311. https://doi.org/10.1093/cercor/bhv311

Koush, Y., Rosa, M. J., Robineau, F., Heinen, K., W. Rieger, S., Weiskopf, N., Vuilleumier, P., Van De Ville, D., & Scharnowski, F. (2013). Connectivity-based neurofeedback: Dynamic causal modeling for real-time fMRI. *NeuroImage*, *81*, 422–430. https://doi.org/10.1016/j.neuroimage.2013.05.010

Laborda-Sánchez, F., & Cansino, S. (2021). The Effects of Neurofeedback on Aging-Associated Cognitive Decline: A Systematic Review. *Applied Psychophysiology and Biofeedback*, *46*(1), 1–10. https://doi.org/10.1007/s10484-020-09497-6

LaConte, S. M., Peltier, S. J., & Hu, X. P. (2007). Real-time fMRI using brain-state classification. *Human Brain Mapping*, *28*(10), 1033–1044. https://doi.org/10.1002/hbm.20326

Lansbergen, M. M., Arns, M., Van Dongen-Boomsma, M., Spronk, D., & Buitelaar, J. K. (2011). The increase in theta/beta ratio on resting-state EEG in boys with attention-deficit/hyperactivity disorder is mediated by slow alpha peak frequency. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(1), 47–52. https://doi.org/10.1016/j.pnpbp.2010.08.004

Li, C.-T., Lin, C.-P., Chou, K.-H., Chen, I.-Y., Hsieh, J.-C., Wu, C.-L., Lin, W.-C., & Su, T.-P. (2010). Structural and cognitive deficits in remitting and non-remitting recurrent depression: A voxel-based morphometric study. *NeuroImage*, *50*(1), 347–356. https://doi.org/10.1016/j.neuroimage.2009.11.021

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. https://doi.org/10.1038/nature06976

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157. https://doi.org/10.1038/35084005

Lubianiker, N., Goldway, N., Fruchtman-Steinbok, T., Paret, C., Keynan, J. N., Singer, N., Cohen, A., Kadosh, K. C., Linden, D. E. J., & Hendler, T. (2019). Process-based framework for precise neuromodulation. *Nature Human Behaviour*, *3*(5), 436–445. https://doi.org/10.1038/s41562-019-0573-y

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions* (No. arXiv:1705.07874). arXiv. https://doi.org/10.48550/arXiv.1705.07874

Lussier, I., & Stip, E. (2001). Memory and attention deficits in drug naive patients with schizophrenia. *Schizophrenia Research*, *48*(1), 45–55. https://doi.org/10.1016/S0920-9964(00)00102-X

MacInnes, J. J., Dickerson, K. C., Chen, N., & Adcock, R. A. (2016). Cognitive Neurostimulation: Learning to Volitionally Sustain Ventral Tegmental Area Activation. *Neuron*, *89*(6), 1331–1342. https://doi.org/10.1016/j.neuron.2016.02.002

Marins, T. F., Rodrigues, E. C., Engel, A., Hoefle, S., Basílio, R., Lent, R., Moll, J., & Tovar-Moll, F. (2015). Enhancing Motor Network Activity Using Real-Time Functional MRI Neurofeedback of Left Premotor Cortex. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00341

Mark, C. I., Mazerolle, E. L., & Chen, J. J. (2015). Metabolic and vascular origins of the BOLD effect: Implications for imaging pathology and resting-state brain function. *Journal of Magnetic Resonance Imaging*, *42*(2), 231–246. https://doi.org/10.1002/jmri.24786

Marxen, M., Jacob, M. J., Müller, D. K., Posse, S., Ackley, E., Hellrung, L., Riedel, P., Bender, S., Epple, R., & Smolka, M. N. (2016). Amygdala Regulation Following fMRI-Neurofeedback without Instructed Strategies. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00183

Mathiak, K. A., Koush, Y., Dyck, M., Gaber, T. J., Alawi, E., Zepf, F. D., Zvyagintsev, M., & Mathiak, K. (2010). Social reinforcement can regulate localized brain activity. *European Archives of Psychiatry and Clinical Neuroscience*, *260*(2), 132–136. https://doi.org/10.1007/s00406-010-0135-9

McDonald, A. R., Muraskin, J., Dam, N. T. V., Froehlich, C., Puccio, B., Pellman, J., Bauer, C. C. C., Akeyson, A., Breland, M. M., Calhoun, V. D., Carter, S., Chang, T. P., Gessner, C., Gianonne, A., Giavasis, S., Glass, J., Homann, S., King, M., Kramer, M., … Craddock, R. C. (2017). The real-time fMRI neurofeedback based stratification of Default Network Regulation Neuroimaging data repository. *NeuroImage*, *146*, 157–170. https://doi.org/10.1016/j.neuroimage.2016.10.048

Megumi, F., Yamashita, A., Kawato, M., & Imamizu, H. (2015). Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00160

Mehler, D. M. A., Williams, A. N., Whittaker, J. R., Krause, F., Lührs, M., Kunas, S., Wise, R. G., Shetty, H. G. M., Turner, D. L., & Linden, D. E. J. (2020). Graded fMRI Neurofeedback Training of Motor Imagery in Middle Cerebral Artery Stroke Patients: A Preregistered Proof-of-Concept Study. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00226

Microsoft Coperation. (2024a). *LightGBM 4.0.0 documentation*. https://lightgbm.readthedocs.io/en/v4.0.0/

Microsoft Coperation. (2024b). *LightGBM 4.5.0 documentation*. https://lightgbm.readthedocs.io/en/stable/Parameters.html

Molnar, C. (2019). *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/

Morgenroth, E., Saviola, F., Gilleen, J., Allen, B., Lührs, M., W. Eysenck, M., & Allen, P. (2020). Using connectivity-based real-time fMRI neurofeedback to modulate attentional and resting state networks in people with high trait anxiety. *NeuroImage: Clinical*, *25*, 102191. https://doi.org/10.1016/j.nicl.2020.102191

Muñoz-Moldes, S., & Cleeremans, A. (2020). Delineating implicit and explicit processes in neurofeedback learning. *Neuroscience & Biobehavioral Reviews*, *118*, 681–688. https://doi.org/10.1016/j.neubiorev.2020.09.003

Nadeau, C., & Bengio, Y. (2003). Inference for the Generalization Error. *Machine Learning*, *52*(3), 239–281. https://doi.org/10.1023/A:1024068626366

Nicholson, A. A., Rabellino, D., Densmore, M., Frewen, P. A., Paret, C., Kluetsch, R., Schmahl, C., Théberge, J., Neufeld, R. W. J., McKinnon, M. C., Reiss, J. P., Jetly, R., & Lanius, R. A. (2017a). The neurobiology of emotion regulation in posttraumatic stress disorder: Amygdala downregulation via real-time fMRI neurofeedback. *Human Brain Mapping*, *38*(1), 541–560. https://doi.org/10.1002/hbm.23402

Nicholson, A. A., Rabellino, D., Densmore, M., Frewen, P. A., Paret, C., Kluetsch, R., Schmahl, C., Théberge, J., Neufeld, R. W. J., McKinnon, M. C., Reiss, J. P., Jetly, R., & Lanius, R. A. (2017b). The neurobiology of emotion regulation in posttraumatic stress disorder: Amygdala downregulation via real-time fMRI neurofeedback. *Human Brain Mapping*, *38*(1), 541–560. https://doi.org/10.1002/hbm.23402

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Ogawa, S., Lee, T.-M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, *14*(1), 68–78. https://doi.org/10.1002/mrm.1910140108

Pamplona, G. S. P., Heldner, J., Langner, R., Koush, Y., Michels, L., Ionta, S., Scharnowski, F., & Salmon, C. E. G. (2020a). Network-based fMRI-neurofeedback training of sustained attention. *NeuroImage*, *221*, 117194. https://doi.org/10.1016/j.neuroimage.2020.117194

Pamplona, G. S. P., Heldner, J., Langner, R., Koush, Y., Michels, L., Ionta, S., Scharnowski, F., & Salmon, C. E. G. (2020b). Network-based fMRI-neurofeedback training of sustained attention. *NeuroImage*, *221*, 117194. https://doi.org/10.1016/j.neuroimage.2020.117194

Papoutsi, M., Magerkurth, J., Josephs, O., Pépés, S. E., Ibitoye, T., Reilmann, R., Hunt, N., Payne, E., Weiskopf, N., Langbehn, D., Rees, G., & Tabrizi, S. J. (2019). *Activity or Connectivity? Evaluating neurofeedback training in Huntington's disease* (p. 481903). bioRxiv. https://doi.org/10.1101/481903

Papoutsi, M., Weiskopf, N., Langbehn, D., Reilmann, R., Rees, G., & Tabrizi, S. J. (2018). Stimulating neural plasticity with real-time fMRI neurofeedback in Huntington's disease: A proof of concept study. *Human Brain Mapping*, *39*(3), 1339–1353. https://doi.org/10.1002/hbm.23921

Paret, C., Goldway, N., Zich, C., Keynan, J. N., Hendler, T., Linden, D., & Cohen Kadosh, K. (2019). Current progress in real-time functional magnetic resonance-based neurofeedback:

Methodological challenges and achievements. *NeuroImage*, *202*, 116107. https://doi.org/10.1016/j.neuroimage.2019.116107

Paret, C., & Hendler, T. (2020). Live from the "regulating brain": Harnessing the brain to change emotion. *Emotion*, *20*(1), 126–131. https://doi.org/10.1037/emo0000674

Paret, C., Kluetsch, R., Zaehringer, J., Ruf, M., Demirakca, T., Bohus, M., Ende, G., & Schmahl, C. (2016). Alterations of amygdala-prefrontal connectivity with real-time fMRI neurofeedback in BPD patients. *Social Cognitive and Affective Neuroscience*, *11*(6), 952–960. https://doi.org/10.1093/scan/nsw016

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Robineau, F., Saj, A., Neveu, R., Van De Ville, D., Scharnowski, F., & Vuilleumier, P. (2019). Using real-time fMRI neurofeedback to restore right occipital cortex activity in patients with left visuo-spatial neglect: Proof-of-principle and preliminary results. *Neuropsychological Rehabilitation*, *29*(3), 339–360. https://doi.org/10.1080/09602011.2017.1301262

Rota, G., Sitaram, R., Veit, R., Erb, M., Weiskopf, N., Dogil, G., & Birbaumer, N. (2009). Self-regulation of regional cortical activity using real-time fMRI: The right inferior frontal gyrus and linguistic processing. *Human Brain Mapping*, *30*(5), 1605–1614. https://doi.org/10.1002/hbm.20621

Scharnowski, F., Hutton, C., Josephs, O., Weiskopf, N., & Rees, G. (2012). Improving Visual Perception through Neurofeedback. *The Journal of Neuroscience*, *32*(49), 17830–17841. https://doi.org/10.1523/JNEUROSCI.6334-11.2012

Scharnowski, F., Veit, R., Zopf, R., Studer, P., Bock, S., Diedrichsen, J., Goebel, R., Mathiak, K., Birbaumer, N., & Weiskopf, N. (2015). Manipulating motor performance and memory

through real-time fMRI neurofeedback. *Biological Psychology*, *108*, 85–97. https://doi.org/10.1016/j.biopsycho.2015.03.009

Sepulveda, P., Sitaram, R., Rana, M., Montalba, C., Tejos, C., & Ruiz, S. (2016). How feedback, motor imagery, and reward influence brain self-regulation using real-time fMRI. *Human Brain Mapping*, *37*(9), 3153–3171. https://doi.org/10.1002/hbm.23228

Shanechi, M. M. (2019). Brain–machine interfaces from motor to mood. *Nature Neuroscience*, *22*(10), 1554–1564. https://doi.org/10.1038/s41593-019-0488-y

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., & Kawato, M. (2019). Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. *NeuroImage*, *188*, 539–556. https://doi.org/10.1016/j.neuroimage.2018.12.022

Shibata, K., Watanabe, T., Kawato, M., & Sasaki, Y. (2016). Differential Activation Patterns in the Same Brain Region Led to Opposite Emotional States. *PLOS Biology*, *14*(9), e1002546. https://doi.org/10.1371/journal.pbio.1002546

Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. *Science*, *334*(6061), 1413–1415. https://doi.org/10.1126/science.1212003

Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N. K., & Ugurbil, K. (2007). Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage*, *35*(2), 539–552. https://doi.org/10.1016/j.neuroimage.2006.12.030

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M. L., Rana, M., Oblak, E., Birbaumer, N., & Sulzer, J. (2017). Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, *18*(2), 86–100. https://doi.org/10.1038/nrn.2016.164

Skouras, S., & Scharnowski, F. (2019). The effects of psychiatric history and age on self-regulation of the default mode network. *NeuroImage*, *198*, 150–159. https://doi.org/10.1016/j.neuroimage.2019.05.008

Sorger, B., Kamp, T., Weiskopf, N., Peters, J. C., & Goebel, R. (2018). When the Brain Takes 'BOLD' Steps: Real-Time fMRI Neurofeedback Can Further Enhance the Ability to Gradually Self-regulate Regional Brain Activation. *Neuroscience*, *378*, 71–88. https://doi.org/10.1016/j.neuroscience.2016.09.026

Sorger, B., Reithler, J., Dahmen, B., & Goebel, R. (2012). A Real-Time fMRI-Based Spelling Device Immediately Enabling Robust Motor-Independent Communication. *Current Biology*, *22*(14), 1333–1338. https://doi.org/10.1016/j.cub.2012.05.022

Spetter, M. S., Malekshahi, R., Birbaumer, N., Lührs, M., van der Veer, A. H., Scheffler, K., Spuckti, S., Preissl, H., Veit, R., & Hallschmid, M. (2017). Volitional regulation of brain responses to food stimuli in overweight and obese subjects: A real-time fMRI feedback study. *Appetite*, *112*, 188–195. https://doi.org/10.1016/j.appet.2017.01.032

Sukhodolsky, D. G., Walsh, C., Koller, W. N., Eilbott, J., Rance, M., Fulbright, R. K., Zhao, Z., Bloch, M. H., King, R., Leckman, J. F., Scheinost, D., Pittman, B., & Hampson, M. (2020). Randomized, Sham-Controlled Trial of Real-Time Functional Magnetic Resonance Imaging Neurofeedback for Tics in Adolescents With Tourette Syndrome. *Biological Psychiatry*, *87*(12), 1063–1070. https://doi.org/10.1016/j.biopsych.2019.07.035

Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M. L., Bruehl, A. B., Cohen, L. G., deCharms, R. C., Gassert, R., Goebel, R., Herwig, U., LaConte, S., Linden, D., Luft, A., Seifritz, E., & Sitaram, R. (2013). Real-time fMRI neurofeedback: Progress and challenges. *NeuroImage*, *76*, 386–399. https://doi.org/10.1016/j.neuroimage.2013.03.033

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., Kawato, M., & Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings*

of the National Academy of Sciences, *115*(13), 3470–3475. https://doi.org/10.1073/pnas.1721572115

Taschereau-Dumouchel, V., Cortese, A., Lau, H., & Kawato, M. (2021). Conducting decoded neurofeedback studies. *Social Cognitive and Affective Neuroscience*, *16*(8), 838–848. https://doi.org/10.1093/scan/nsaa063

Thibault, R. T., Lifshitz, M., & Raz, A. (2016). The self-regulating brain and neurofeedback: Experimental science and clinical promise. *Cortex*, *74*, 247–261. https://doi.org/10.1016/j.cortex.2015.10.024

Thibault, R. T., Lifshitz, M., & Raz, A. (2017). Neurofeedback or neuroplacebo? *Brain*, *140*(4), 862–864. https://doi.org/10.1093/brain/awx033

Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., & Raz, A. (2018). Neurofeedback with fMRI: A critical systematic review. *NeuroImage*, *172*, 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071

van Rossum, G. (1995). *Python tutorial* (Technical Report No. CS-R9526). Centrum voor Wiskunde en Informatica (CWI). https://ir.cwi.nl/pub/5007/05007D.pdf

Watanabe, T., Sasaki, Y., Shibata, K., & Kawato, M. (2017). Advances in fMRI Real-Time Neurofeedback. *Trends in Cognitive Sciences*, *21*(12), 997–1010. https://doi.org/10.1016/j.tics.2017.09.010

Weiskopf, N. (2012). Real-time fMRI and its application to neurofeedback. *NeuroImage*, *62*(2), 682–692. https://doi.org/10.1016/j.neuroimage.2011.10.009

Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., & Mathiak, K. (2004). Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *Journal of Physiology-Paris*, *98*(4), 357–373. https://doi.org/10.1016/j.jphysparis.2005.09.019

Wolpaw, J. R. (2013). Chapter 6—Brain–computer interfaces. In M. P. Barnes & D. C. Good (Eds.), *Handbook of Clinical Neurology* (Vol. 110, pp. 67–74). Elsevier. https://doi.org/10.1016/B978-0-444-52901-5.00006-X

Yao, S., Becker, B., Geng, Y., Zhao, Z., Xu, X., Zhao, W., Ren, P., & Kendrick, K. M. (2016). Voluntary control of anterior insula and its functional connections is feedback-independent and increases pain empathy. *NeuroImage*, *130*, 230–240. https://doi.org/10.1016/j.neuroimage.2016.02.035

Yoo, S.-S., Fairneny, T., Chen, N.-K., Choo, S.-E., Panych, L. P., Park, H., Lee, S.-Y., & Jolesz, F. A. (2004). Brain–computer interface using fMRI: Spatial navigation by thoughts. *NeuroReport*, *15*(10), 1591. https://doi.org/10.1097/01.wnr.0000133296.39160.fe

Young, K. D., Siegle, G. J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W. C., & Bodurka, J. (2017). Randomized Clinical Trial of Real-Time fMRI Amygdala Neurofeedback for Major Depressive Disorder: Effects on Symptoms and Autobiographical Memory Recall. *American Journal of Psychiatry*, *174*(8), 748–755. https://doi.org/10.1176/appi.ajp.2017.16060637

Zhao, Z., Yao, S., Zweerings, J., Zhou, X., Zhou, F., Kendrick, K. M., Chen, H., Mathiak, K., & Becker, B. (2021). Putamen volume predicts real-time FMRI neurofeedback learning success across paradigms and neurofeedback target regions. *Human Brain Mapping*, *42*(6), 1879–1887. https://doi.org/10.1002/hbm.25336

Zhi, L., Zhang, C., Huang, J., Wang, Y., Yan, C., Li, K., Zeng, Y., Jin, Z., Cheung, E. F. C., Su, L., & Chan, R. C. K. (2018). Improving motivation through real-time fMRI-based self-regulation of the nucleus accumbens. *Neuropsychology*, *32*(6), 764–776. https://doi.org/10.1037/neu0000425

Zich, C., Johnstone, N., Lührs, M., Lisk, S., Haller, S. Pw., Lipp, A., Lau, J. Yf., & Kadosh, K. C. (2020). Modulatory effects of dynamic fMRI-based neurofeedback on emotion regulation

networks in adolescent females. *NeuroImage*, *220*, 117053. https://doi.org/10.1016/j.neuroimage.2020.117053

Zilverstand, A., Sorger, B., Sarkheil, P., & Goebel, R. (2015). fMRI neurofeedback facilitates anxiety regulation in females with spider phobia. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00148

Figures legends

Table legends

Abbreviations

Abbreviation	*Meaning*
ROI	Region of Interest
fMRI	Functional Magnetic Resonance Imaging
EEG	Electroencephalography
BOLD	Blood-Oxygen-Level-Dependent
Hb	Haemoglobin
Rt-fMRI	Real-time Functional Magnetic Resonance Imaging
DecNef	Decoded Neurofeedback
MVPA	Multivoxel Pattern Analysis
CV	Cross-Validation
SHAP	Shapley Additive exPlanations
SD	Standard Deviation

Appendix

All supplementary material and analyses code can be found in the open science framework: https://osf.io/wd845/?view_only=6d59586659c84d298d8dac0be7981cd1.

Use of large language models

Large language model GPT-4o was used for providing assistance while coding and text polishing.

Example prompts:

- Please help me solve the following error.
- Please improve the English and structure of the following section. Please don't change any of the content.