From the Institute for Systems Neuroscience

at the Heinrich-Heine-University Düsseldorf

Interpretability and Reliability in Neuroimaging

Dissertation

to obtain the academic title of Doctor of Philosophy (Ph.D.) in Medical Sciences

from the Faculty of Medicine at Heinrich-Heine-University Düsseldorf

submitted by

Jean-Philippe Kröll

(2025)

As an inaugural dissertation printed with permission of the

Faculty of Medicine at Heinrich-Heine-University Düsseldorf

signed:

Dean: Prof. Dr. med. Nikolaj Klöcker

Examiners: PD Dr. Susanne Weis,

Prof. Dr. Julian Caspers

Parts of this work have been accepted and published:

Jean-Philippe Kröll, Simon B. Eickhoff, Felix Hoffstaedter, Kaustubh Patil, (2020). Evolving complex yet interpretable representations: application to Alzheimer's diagnosis and prognosis,*2020 IEEE Congress on Evolutionary Computation (CEC),* pp. 1-8. 10.1109/CEC48606.2020.9185843

Jean-Philippe Kröll, Patrick Friedrich, Xuan Li, Kaustubh R. Patil, Lisa Mochalski, Laura Waite, Xing Qian, Michael WL Chee, Juan Helen Zhou, Simon Eickhoff, Susanne Weis, (2023). Naturalistic viewing increases individual identifiability based on connectivity within functional brain networks. *NeuroImage*, *Vol*(273), https://doi.org/10.1016/j.neuroimage.2023.120083.

Mochalski, L. N., Friedrich, P., Li, X., Kröll, J.-P., Eickhoff, S. B., & Weis, S., I. (2024). Inter- and intra-subject similarity in network functional connectivity across a full narrative movie. *Human Brain Mapping, 45(11),* https://doi.org/10.1002/hbm.26802

Zusammenfassung

Die Entwicklung von Biomarkern auf der Grundlage der Magnetresonanztomographie (MRT) ist ein ständiges Bestreben auf dem Gebiet der klinischen Neurowissenschaften. Obwohl diese Biomarker ein großes Potenzial haben, wurden bisher nur wenige für den routinemäßigen klinischen Einsatz übernommen. Die größten Herausforderungen bei der Umsetzung in die klinische Anwendung sind die Genauigkeit, Zuverlässigkeit und Interpretierbarkeit eines Biomarkers. In dieser Dissertation wird daher ein neues maschinelles Lernverfahren (ML) vorgestellt, das die Genauigkeit der Diagnose und Prognose einer der häufigsten neurologischen Erkrankungen, der Alzheimer-Krankheit, durch die Konstruktion komplexer Darstellungen von Basis-Featuren verbessert. Durch die Verwendung einer kontextfreien Grammatik werden die konstruierten Repräsentationen gezwungen, menschlich interpretierbar zu bleiben, was die Validierung einer Beziehung zwischen dem Biomarker und dem vermuteten zugrunde liegenden pathologischen Korrelat ermöglicht. Darüber hinaus wird untersucht, ob *Naturalistic Viewing* (NV) Paradigmen geeignet sind, die für die Entwicklung von Biomarkern wichtigen Eigenschaften von MRT-Messungen zu verbessern, wie z. B. Reliabilität, geringere Variabilität innerhalb von Probanden und verbesserte Erkennung individueller Unterschiede im Vergleich zu Ruhemessungen (RS). Daher wird die Wirkung von NV-Stimuli mit unterschiedlichem sozialem Inhalt und unterschiedlicher Länge in 14 funktionellen Gehirnnetzwerken untersucht. Es wird gezeigt, dass NV-Stimuli, basierend auf der funktionellen Netzwerkkonnektivität (NFC), die Erkennung von individuellen Unterschieden in 10 von 14 Netzwerken verbessern, wobei die Stimuli mit dem höchsten Grad an sozialem Inhalt die größte Verbesserung erzielen. Eine anschließende Analyse bestätigt, dass Filmstimuli mit einem höheren Maß an sozialem Inhalt ähnliche NFC-Muster hervorrufen, die sich von RS und einem Stimulus ohne soziale Interaktionen unterscheiden. Darüber hinaus wird gezeigt, dass NV-Stimuli die Intra-Subjekt-Variabilität in meta-analytischen Netzwerken reduzieren können, die für die Wahrnehmung und Verarbeitung von Handlungen, Verhalten und Emotionen wichtig sind. Zusätzlich wird gezeigt, dass NV-Stimuli die Zuverlässigkeit von Graph-Metriken, die aus NFC extrahiert werden, gegenüber RS erhöhen können. Die Ergebnisse machen jedoch auch deutlich, dass NV-Stimuli die Metriken nicht uneingeschränkt über das gesamte Gehirn hinweg verbessern. Insbesondere für Netzwerke, die mit intrinsisch orientierten Funktionen verbunden sind, erweist sich RS als das zu bevorzugende Paradigma. Daher ist die Auswahl des geeigneten Stimulus und des funktionellen Netzwerks für die Beantwortung der jeweiligen Forschungsfrage von entscheidender Bedeutung. Schließlich stellt diese Dissertation einen neuen öffentlich zugänglichen NV-Datensatz zur Verfügung, um die Wirkung von NV-Stimuli weiter zu analysieren.

Summary

The development of magnetic resonance imaging (MRI) based biomarkers is a constant endeavor in the field of clinical neuroscience. Although these biomarkers hold great potential, only few have been adopted for routine clinical use. Primary challenges for the translation into clinical use are accuracy, reliability and interpretability of a given biomarker. Consequently, this dissertation presents a new machine learning (ML) framework that improves accuracy of diagnosis and prognosis of one of the most common neurological diseases, Alzheimer' Disease (AD), by constructing complex representations of base features. Further, by using a context-free grammar (CFG), the constructed representations are forced to remain humanly interpretable, thus enabling the validation of a relationship between the biomarker and the supposed underlying pathologic correlate. Additionally, it is investigated if naturalistic viewing (NV) paradigms are suited to improve characteristics of MRI measurements that are important for biomarker development, such as reliability, reduced intra-subject variability and enhanced detection of individual differences, in comparison with resting-state (RS). Therefore, the effect of NV stimuli with varying levels of social content and different lengths is investigated in 14 functional brain networks. It is shown that, based on network functional connectivity (NFC), NV stimuli improve the detection of individual differences in 10 out of 14 networks, with the stimuli with the highest level of social content achieving the most improvement. A subsequent analysis confirms that movie stimuli with higher levels of social content evoke similar NFC patterns that are distinct from RS and a stimulus lacking social interactions. Further, it is demonstrated that NV stimuli can reduce intra-subject variability in meta-analytic networks that are essential for perception and processing of action, behavior and emotions. In addition, it is shown that NV stimuli can increase the reliability of graph metrics extracted from NFC, over RS. However, the results also emphasize that NV stimuli do not unconditionally improve metrics of interest across the whole brain. In particular for networks that are related to intrinsically oriented functions, RS proves to be the more favorable paradigm. Therefore, selecting the appropriate stimulus and functional network is essential for addressing the specific research question at hand. Finally, this dissertation provides a new publicly available NV dataset to further analyze the effect of NV stimuli.

List of abbreviations

Abbreviation	Definition
MRI	magnetic resonance imaging
AD	Alzheimer's Disease
ML	machine learning
fMRI	functional magnetic resonance imaging
sMRI	structural magnetic resonance imaging
rs-fMRI	resting-state functional magnetic resonance imaging
NV	naturalistic viewing
AI	artificial intelligence
SVM	support vector machine
RBF	radial basis function
GE	grammatical evolution
FC	functional connectivity

RS	resting state
NFC	network functional connectivity
AM	autobiographical memory
ER	emotion regulation
SM	semantic memory
ToM	theory of mind
eSAD	extended socio-affective default
FNM	full narrative movie
ADHD	attention-deficit/hyperactivity disorder

Table of contents

1 Introduction

Since initial discoveries that MRI-measured structural brain differences between healthy individuals and patients can be used to monitor or even predict disease progression, many researchers have tried to develop biomarkers for use in clinical settings. Apart from conventional statistical approaches, the use of ML based methods has gained a lot of popularity. Especially with more availability of larger datasets, researchers have turned to ML as such methods better handle complex, high-dimensional data than conventional approaches. However, one of the primary challenges in the translational use of ML methods is the lack of explainability, particularly with non-linear techniques. Explainable (i.e., human-interpretable) methods on the other hand not only offer valuable insights into disease mechanisms but also foster clinician-patient trust, which is crucial for the broader social acceptance of ML approaches. Besides the need for interpretable ML models, the reliability of MRI measurements itself are an important factor for ensuring accurate and reproducible results. In functional MRI (fMRI) research, resting-state fMRI (rs-fMRI) has been the gold standard for the study of brain connectivity because it measures intrinsic functional organization independent of task constraints. However, rs-fMRI is not without limitations, including intra-subject variability, susceptibility to motion artifacts, and the influence of unconstrained mental processes. All these compromise the reliability of rs-fMRI. NV paradigms, where participants engage with dynamic, real-world stimuli, offer a promising alternative to traditional rs-fMRI for studying brain connectivity. However, the reliability of NV paradigms and their ability to capture individual differences not only across the brain, but also in functional networks, are yet to be assessed.

1.1 Alzheimer's Disease and Interpretability

One of the diseases which has extensively been studied with MRI measures is AD. AD is the most common form of dementia, affecting about 50 million people worldwide. It significantly impairs memory, language, and intellectual capabilities, making day-to-day tasks increasingly difficult for patients. With increasing life expectancy, AD has become an emerging public health problem (Nandi et al., 2022). It has therefore become a main research objective to develop accurate methods for early diagnosis of AD. One promising method of distinguishing AD from healthy controls is the application of machine learning to structural MRI (sMRI) data (Lahmiri and Shmuel, 2019; Zhu et al., 2021). Since brain atrophy is a feature of AD and can be observed in sMRI scans, most machine learning approaches have been successful in leveraging this information for classification. Since sMRI is already a staple in clinical practice, developing an accurate diagnostic tool based on such scans is of great clinical value. However, aside from accuracy, another critical factor is interpretability because it enables researchers to understand which areas of the brain are most affected by AD and how these regions interact. That kind of insight can open up knowledge on mechanisms of the disease and ultimately help improve treatment outcomes. Furthermore, building trust between clinicians using ML/artificial intelligence (AI) methods and patients is essential for the acceptance of these technologies. This can be achieved by clearly explaining the reasoning behind decisions and the uncertainties associated with different options. Therefore, the latest EU guidelines for trustworthy AI, make transparency one of the main requirements for the application of machine learning algorithms ("EU guidelines on ethics in artificial intelligence: Context and implementation," 2019). Accuracy and interpretability, however, usually come with a trade-off. While interpretable models like linear support vector machine (SVM) and logistic regression can not capture very complex feature interactions, more complex models like SVM with radial basis function (RBF) kernel or neural networks would be capable of recognizing such patterns, but they are less interpretable in terms of how they arrived at those decisions.

One approach that can be utilized to enhance both accuracy and interpretability is employing grammatical evolution (GE) for feature construction and selection. GE is an evolutionary algorithm which combines the concepts of genetic programming and formal grammar systems with the aim of evolving solutions to complex problems. Like all evolutionary algorithms, GE works by maintaining a population of solutions (i.e. newly constructed features), selecting the fittest individuals for reproduction, and applying genetic operators like crossover and mutation to create new candidates. Each individual is evaluated based on a fitness function, and over multiple generations, the population evolves toward better solutions. However, unlike other evolutionary algorithms where program structures are evolved explicitly, GE uses a user-defined formal grammar. By restricting this grammar to perform only basic arithmetic operations during feature generation, interpretability of the resulting features can be enforced.

1.2 Resting-State vs Naturalistic Viewing

Most research on functional connectivity (FC) has focused on connectivity patterns observed during task-free resting state (RS), where participants lie in a scanner without engaging in any specific task (Amft et al., 2015; Damoiseaux et al., 2006; Langner and

Eickhoff, 2013). RS is believed to reflect the brain's intrinsic organization and has also been shown to align well with findings derived from task-based studies (Smith et al., 2009). Further, the ease of implementing RS data allows for the rapid acquisition of large healthy and clinical samples due to minimal participant demands. While the RS paradigm has provided valuable insights into brain organization, it also has limitations: RS data can be heavily influenced by head movement and drowsiness due to its unconstrained nature (Tagliazucchi and Laufs, 2014; Van Dijk et al., 2012), as participants struggle to stay awake and still without a task or stimulus. Moreover, RS is susceptible to the influence of spontaneous thoughts (Christoff et al., 2004; Gonzalez-Castillo et al., 2021).

NV paradigms, where participants watch a story or film, have recently gained popularity as they offer a more ecologically valid approach to studying brain function. Compared to RS, NV offers several advantages. By providing a stimulus, NV reduces the variability caused by spontaneous thoughts. Furthermore, NV has been shown to reduce fatigue and head movement by increasing participant engagement, as compared to RS (Finn and Bandettini, 2020; Vanderwal et al., 2019). Finally, watching movies can make scanning more tolerable for participants who find it challenging to stay still (e.g., individuals with ADHD) or complete demanding tasks (e.g., those with cognitive impairments) (Eickhoff et al., 2020).

1.3 The two present Samples

Since the rise of NV, a plethora of samples have been made available. Researchers have implemented different stimuli, varying from short clips that last less than two minutes to full length movies (DuPre et al., 2020). More so, movie clips differ in their content. On the one hand, studies have used rather neutral clips e.g. depicting landscapes or even more abstract clips like the movie Inscapes as a baseline comparison to RS (Van Essen et al., 2012; Vanderwal et al., 2015). On the other hand, many authors have suggested that movie clips with social content are more likely to engage participants (Finn et al., 2018; Finn and Bandettini, 2020; Nguyen et al., 2019; Rikandi et al., 2017). Related, several studies have suggested that the cultural background of a person can influence the effect of NV. Cultural norms, values, and experiences shape one's interpretation of social interactions and narrative elements depicted in movie stimuli. Consequently, elucidating the interplay between cultural background and naturalistic viewing is crucial for understanding the variability in neural responses across populations (Eickhoff et al., 2020).

To address these points, two identical samples were acquired for this dissertation. Both samples employed the same three different movie stimuli with different levels of social content. The first movie was the movie Inscapes which depicts only abstract animations and lacks any form of social interaction. The second movie, The Circus (United Artists Digital Studios, 1928, directed by Charlie Chaplin) is a silent black-and-white film that shows the protagonist being chased through a circus by the police and unintentionally causing comic situations during his escape. Due to the lack of spoken words and the chaplin-typical pantomime-esque depiction, this movie is employed as a stimulus with moderate level of social interactions. The third movie, Indiana Jones and the Temple of Doom (Paramount Pictures, 1984, directed by Steven Spielberg) the protagonist is shown during an intense negotiation and afterwards has to fend off several hitmen who try to kill him. Due to the complex interactions between the characters during this scene, the movie is seen as the stimulus with the highest level of social content. To enable the comparison of cultural effects, one sample was acquired in Singapore and one sample in Jülich.

1.4 Meta-analytic Networks

The human brain is commonly seen as being organized into modules of spatially distinct areas that form functional networks (Sporns and Betzel, 2016). These networks correspond to particular cognitive domains, such as memory (Spreng et al., 2009), social cognition (Bzdok et al., 2012) and executive function (Rottschy et al., 2012). Since NV paradigms use complex, multimodal stimuli that elicit activation patterns across the whole brain, adopting a network perspective can explain the effect of movie stimuli on particular cognitive processes.

In the context of NV, one would expect that networks that relate to different functions, should also differ in their response to the same stimulus. For example, a functional network that processes emotions should be differently affected by a movie scene with strong emotional content, in comparison to the motor network. Therefore, investigating FC in networks that cover different cognitive domains under NV, extends the knowledge over traditional whole-brain studies. There are various methods to define functional networks (Power et al., 2011; Schaefer et al., 2018; Smith et al., 2009) one of which is the use of meta-analysis (Eickhoff et al., 2012). Meta-analytically defined networks integrate converging data from a multitude of studies and thus represent the most likely core nodes that are involved in a given function. Therefore, studying FC in meta-analytical networks could offer new and more detailed insights into the effects of naturalistic viewing, as compared to conventional whole-brain studies.

1.5 Individual Differences

While traditional neuroscience has mostly focused on group-level analysis, exploring variability between subjects is essential for a comprehensive understanding of individual brain function. Characterizing individual variations in FC offers additional insights into the relation of brain function, behavior and cognition. Furthermore, individual differences hold significant implications for personalized medicine. Understanding variations between individual brains will help to assess personal susceptibility to neurological disorders and response to interventions. However, the detection of individual differences in FC has been a challenging task. Due to motion artifacts and physiological fluctuations inherent to fMRI data, true individual differences are partly obscured and difficult to disentangle from noise (Dubois and Adolphs, 2016). In addition, the typically used RS paradigm is influenced by attention fluctuations and spontaneous thoughts of the participant (Christoff et al., 2004). Moreover, the passive nature of the RS paradigm might not fully capture the individual's cognitive abilities, thus limiting the sensitivity with which subtle variations across participants can be detected. Previous research has shown that certain tasks improve the sensitivity to individual differences in FC in comparison to RS (Finn et al., 2017). However, the authors themselves point out that alternative paradigms are worth exploring for the analysis of individual differences, as they might enhance the detection of individual differences over RS and task approaches. One of these paradigms is NV. In contrast to RS and task, NV employs rich stimuli that better reflect the complexity of real-life experiences. By exposing participants to a wide array of sensory, emotional and contextual input, NV stimuli probe the human brain under a condition that allows past experiences, cultural beliefs and cognitive strategies to shape the neuronal response. Thereby, NV imposes richer brain state dynamics and therefore more individual connectivity profiles, which might better reflect individual characteristics than RS (Vanderwal et al., 2017).

1.6 Reliability

fMRI has become an indispensable tool in neuroscience research and has granted substantial insight into the function of the human brain. As applications of fMRI expand to the prediction of clinical outcomes, the reliability of the measurement has become a major concern. In order to guide clinicians in the diagnosis and prognosis of brain disorders, a measure has to consistently give accurate results. However, the reported reliability of fMRI measures varies vastly across studies (Bennett and Miller, 2010), partly due to small test-retest samples, but also due to different analysis choices. Therefore, finding methods that increase reliability has become a priority. Traditionally, the field has relied either on task-free RS or on highly controlled task designs. Although both paradigms have their benefits, the former lacks specificity which makes it challenging to relate the observed neural activity to function, while the latter has limited generalizability because it uses highly artificial tasks to focus on one specific cognitive process. One of the methods with potential to increase reliability of fMRI is NV, because it engages the brain in a more structured, yet ecologically valid context. NV paradigms present participants with stimuli that mimic conditions under which the brain naturally operates, such as movies depicting dynamic social interactions. Thereby, participants might react in a manner that is more reflective of their typical cognitive processes, possibly leading to more consistent and reliable results.

1.7 Ethics protocols

The acquisition and use of the JUMAX dataset has been approved by the Heinrich-Heine-University Düsseldorf (Study-Nr. 2019-791). The IMAX dataset was acquired under protocols approved by the National University of Singapore (NUS-IRB REFERENCE CODE: B-14-045). Data collection and sharing for the ADNI project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904).

1.8 Aims of the thesis

This thesis aims to advance the development of biomarkers. Therefore, a new ML framework is provided that can accurately diagnose and prognose AD while retaining interpretability of the model. Interpretability is one of the major hurdles for the translation of ML based MRI research to real-world applications. Subsequently, the thesis will focus on another key challenge of biomarker development which is improving the reliability and the ability to detect individual differences. The thesis will explore the use of NV paradigms as an alternative to resting-state fMRI (rs-fMRI) for studying brain connectivity. It will assess how NV paradigms capture individual differences in functional networks and how different stimuli influence within- and between-subject similarity. In addition, the reliability of NV will be analyzed and compared to that of RS.

This dissertation pertains to four studies. Study 1 establishes a ML framework that maximizes predictive accuracy while retaining feature interpretability. The framework is applied to the diagnosis and prognosis of AD. The features of the final ML model are then examined in terms of their interpretability. Study 2 investigates the effect of NV on FC in fourteen meta-analytic networks. Particularly, the study focuses on the identifiability of individuals based on FC during NV and RS. In addition, individual variability in network FC (NFC) is assessed by comparing within- and between-subject similarity during NV and RS. These results are then compared between different NV stimuli and functional networks. Study 3 investigates the within- and between-subject similarity in NFC of the same 14 networks using a full narrative movie (FNM) and employing a linear mixed model to assess which factors explain inter- and intra-subject similarity. Study 4 focuses on the reliability of NV and compares it to that of RS. The influence of NV on reliability is characterized on the basis of graph metrics extracted from the same 14 functional networks.

2 Evolving complex yet interpretable representations: application to Alzheimer's diagnosis and prognosis, Jean-Philippe Kröll, Simon B. Eickhoff, Felix Hoffstaedter, Kaustubh Patil, *2020 IEEE Congress on Evolutionary Computation (CEC),* pp. 1-8, (2020)

3 Naturalistic viewing increases individual identifiability based on connectivity within functional brain networks, Jean-Philippe Kröll, Patrick Friedrich, Xuan Li, Kaustubh R. Patil, Lisa Mochalski, Laura Waite, Xing Qian, Michael WL Chee, Juan Helen Zhou, Simon Eickhoff, Susanne Weis, *NeuroImage, Vol*(273), (2023)

4 Inter- and intra-subject similarity in network functional connectivity across a full narrative movie, Lisa N. Mochalski, Patrick Friedrich, Xuan Li, Jean-Philippe Kröll, Simon B. Eickhoff, Susanne Weis, *Human Brain Mapping, 45*(11), e26802, (2024)

5 Test–Retest Reliability of Meta Analytic Networks During Naturalistic Viewing, Jean-Philippe Kröll, Patrick Friedrich, Xuan Li, Yulia Nurislamova, Nevena Kraljevic, Anna Geiger, Julia Mans, Laura Waite, Julian Caspers, Xing Qian, Michael WL Chee, Juan Helen Zhou, Simon Eickhoff, Susanne Weis, *bioRxiv, 2024.05.15.594266 doi: https://doi.org/10.1101/2024.05.15.594266*

5 Discussion

5.1 Interpretable ML frameworks

Despite the opportunities provided by large datasets and ever more advanced ML algorithms, only a fraction of proposed ML methods make it to clinical application. One of the major hurdles for clinical translation is the interpretability of a given method (Dinsdale et al., 2022; Thibeau-Sutre et al., 2023). With first disease-modifying treatments becoming available for AD (Mintun et al., 2021; Van Dyck et al., 2023), accurate diagnosis and prognosis of the disease have become even more urgent. Likewise, this places a high demand on the explainability of potential biomarkers as decisions based on these biomarkers will lead to drug-administration.Various examples have shown that “blackbox” models applied in clinical settings can produce seemingly accurate results, relying on confounders, but will fail to generalize on new data (DeGrave et al., 2021; Thibeau-Sutre et al., 2023; Winkler et al., 2019). This dissertation provides a framework that not only improves accurate diagnosis and prognosis of AD, by constructing complex features, but also maintains interpretability of the constructed features. For both applications, the features constructed by the GE framework improved performance across all four metrics used in the study, in comparison with models using base features (Study 1, Table 3, 4). The performance of the models were comparable to results reported in other studies that employed explainable ML frameworks for diagnosis and prognosis of AD (Bloch et al., 2021; Bogdanovic et al., 2022; Böhle et al., 2019; Pohl et al., 2022). The features that were constructed by the GE framework integrate information about the complex interactions between base features, such that the result is still interpretable. An analysis of the constructed features showed that they combined brain regions that are known to be affected in AD, such as the temporal pole (Scheltens et al., 1992), Amygdala (Poulin et al., 2011), Putamen (de Jong et al., 2008) and Thalamus (de Jong et al., 2008). More so, the constructed features were still interpretable as it was observable that they contained information about the atrophy or co-atrophy of AD-involved brain regions.

5.2 Individual differences during RS and NV

One of the primary goals of neuroscience is to relate differences in brain functions to differences in phenotypes. However, analyzing differences in individual FC patterns that occur during RS has led to unsatisfactory results. Therefore, study 2 investigates how a new paradigm, NV, enhances individual differences in comparison with RS. By calculating the identifiability of individual FC matrices extracted during three NV stimuli and RS, this dissertation provides clear evidence for improved detection of individual differences during movie-watching. Using identifiability or "fingerprinting" of FC matrices as a proxy for individual differences has been made popular by Finn et al (Finn et al., 2015) and was previously used by Vanderwal et al to show that NV can enhance individual differences on a whole brain basis (Vanderwal et al., 2017). Similarly, identifiability was used in this dissertation to compare the effects of NV and RS on individual differences in functional networks. In ten out of fourteen networks, NV improved identification accuracy over RS (Study 2, Tab.1). The improvement seen for NV was most prominent for the Indiana Jones stimulus, which led to the highest identifiability in eight of the networks. On the other hand, the movie Inscapes was generally similar or inferior to RS, while Circus showed improvement in only two networks. These results are in line with previous studies that have suggested that in order to maximally engage the participant, NV stimuli with more social content might be preferable over neutral/abstract stimuli (Dmochowski et al., 2014; Finn and Bandettini, 2020; Nummenmaa et al., 2014; Schmälzle et al., 2015). This notion was further supported by comparing patterns of inter-individual NFC between conditions. NFC patterns during Inscapes were mostly similar to those during RS, while Circus and Indiana Jones exhibit connectivity profiles that are distinct from RS across networks (Study 2, Fig.1, Fig. 2).

5.3 Variability across functional networks

The vast majority of studies that investigate the impact of NV have focused on whole-brain connectivity. This dissertation instead focuses on connectivity on a network level and provides evidence that the effect of NV deviates across networks covering different cognitive domains. Study 2 investigated three different NV stimuli and revealed differences in within- and between-subject correlations during RS and NV that were obscured on a whole-brain level (Study 2, Fig 4, 5, S1, S2). Based on the overall increased identifiability during NV in Study 2, one might expect that within-subject correlations (as a measure for stable individual patterns) are increased during NV as well. However, identifiability is always dependent on the ratio of within- and between-subject correlations, e.g. subjects that are too similar to each other will be harder to identify even though they might exhibit stable patterns across sessions (Finn et al., 2017). In Study 2, increased within-subject correlations were observed in meta-analytic networks that are essential for the perception and processing of action, behavior and emotions. With regards to the assumption that the social aspect of a movie stimulus induces stable individual connectivity patterns, it is only reasonable to expect that this effect is more pronounced in networks that deal with the processing of social interactions. On a whole-brain level, between-subject correlations are generally presumed to be increased by NV, given that all subjects are presented with the same stimulus (Hasson et al., 2004; Vanderwal et al., 2017). However, this dissertation importantly shows that this effect is not unambiguously true across functional networks. NV increased between-subject correlations in networks that are associated with executive functions and/or stimulus evaluation. On the other hand, networks that are more related to intrinsically oriented functions exhibited reduced between-subject correlations during NV. Presumably, the function of these networks is suppressed during the processing of complex stimuli, thus preventing coordinated activity in these networks which in turn reduces similarity between subjects.

This dissertation also investigated within- and between-subject correlations in functional networks during a full narrative movie (FNM), Forrest Gump, from studyforrest project (Hanke et al., 2016). Contrary to the shorter stimuli used in Study 2 (10 minutes), a FNM provides emotions embedded in a richer context and evolving over a longer time, allowing for a more comprehensive study of socio-affective processes. Study 3 showed that the effect of the FNM on changes in within- and between-subject correlations was dependent on the network (Study 3, Fig. 3), confirming results from Study 2. Furthermore, Study 3 implemented linear mixed models to analyze how the narrative of the movie and the portrayed valence and arousal affected within- and between-subject correlations across networks. Based on valence and arousal annotations from Labs et al (Labs et al., 2015), the analysis revealed that within- and between-subject correlations were best accounted for by network, movie segment, valence and a movie segment by valence interaction. Within-subject correlations were further explained by an interaction of movie segment, valence and arousal. Taken together, these findings show that within- and between-subject correlations during NV are sensitive to the progressing narrative and emotions portrayed in a stimulus and differ between networks. Lower within-subject correlations during the FNM were observed in the AM, ER, SM, ToM, and eSAD networks (Study 3, Fig. 3), which align with patterns observed in Study 2 that show a tendency for lower within-subject correlations during NV than during RS, in these networks (Study 2, Fig. 4). In addition, Study 3 could also demonstrate that within- and between-subject correlations across networks increased as the movie progressed, suggesting a general trend towards greater similarity in subjects' NFC over time (Study 3, Fig. 3). Related, previous work has demonstrated that certain cognitive and emotional processes develop only over extended time periods (Hasson et al., 2010). This dissertation shows that both, movie clips and FNMs, have different effects across functional networks, emphasizing that a network perspective grants more detailed insights into the full effect of NV paradigms than whole-brain analysis.

5.4 Reliability of NV stimuli

For any research question at hand, the crucial prerequisite is that the used measurement is reliable, such that differences across subjects and time points can be meaningfully interpreted. This dissertation investigates the reliability of NV paradigms and compares it to that of conventional RS. Study 4 shows that NV can improve reliability over RS across networks dealing with affective, social, executive, memory and motor functions, in two samples (Study 4, Fig. 1, Fig. 2). These results indicate that NV can increase engagement not only in sensory, but also in higher order networks. The observed reliability in study 4 matched results from previous studies investigating graph metrics extracted from RS and NV (Braun et al., 2012; Cao et al., 2014; Wang et al., 2017). Similar to Wang et al (Wang et al., 2017) and results from study 2 and study 3, effects of NV varied across networks. However, in contrast to results from Wang et al, where the majority of networks showed improved reliability during NV, NV was less reliable than RS in the majority of networks and graph metrics in study 4. Possibly, since sessions for both datasets were conducted within a week, participants might have been rather familiar with the movie stimuli during the second session. Multiple studies have shown that expected stimuli reduce the neuronal response (Alink et al., 2010; Koster-Hale and Saxe, 2013), which in turn might have led to the relative decrease in reliability for NV here. Still, NV at least partially increased reliability over RS. In these cases, it was again observable that the NV stimuli with more social content, Circus and Indiana Jones, improved reliability more often than Inscapes. In addition, Study 4 investigated if the effect of NV stimuli is different across cohorts with different cultural backgrounds. Previous studies have demonstrated cultural differences for the perception of faces (Adams et al., 2010; Goh et al., 2010; Harada et al., 2020) and rating of emotions when watching movie clips (Sneddon et al., 2011). Similarly, the results of the Asian and European cohort in Study 4 were mostly different across stimulus, network and graph metric (Study 4, Fig. 3). Therefore, future studies should consider the cultural background of a cohort when choosing a movie stimulus.

5.5 Conclusions

This dissertation addressed primary challenges for the translation of MRI based biomarkers into clinical use, such as accuracy, reliability and interpretability. Therefore, a simple GE based framework was provided that constructs complex feature representations while remaining interpretability. The GE framework was demonstrated to be applicable to the diagnosis and prognosis of AD, one of the most prevalent neurological diseases as of today, where it could significantly improve predictive performance. Subsequent inspection of the features uncovered humanly interpretable patterns of co-atrophy in brain regions typically impacted by AD. Further, this dissertation investigated if NV paradigms can improve key biomarker metrics such as reliability, reduced intra-subject variability and enhanced detection of individual differences, in comparison with RS. Therefore, different NV stimuli with varying levels of social content, as well as different lengths and their effect in functional brain networks were compared. The comparison of different NV stimuli revealed that certain stimuli, The Circus and Indiana Jones, are better suited to improve the detection of individual differences, possibly due to a higher level of social content. A clustering of the connectivity profiles during the different stimuli confirmed that these two stimuli were more distinct from RS than the movie Inscapes, which lacks social interaction. Further, an analysis of within- and between subject correlations demonstrated that shorter movie clips as well as a FNM can improve similarity within and between subjects, in comparison with RS. In addition, it was shown that NV stimuli can increase the reliability of fMRI, as measured by graph metrics. This dissertation extends the current knowledge about NV paradigms by examining their effect in functional networks. Contrary to previous studies that focused on whole-brain, it was demonstrated that NV stimuli do not unconditionally improve reliability, as well as within- and between-subject correlations across the brain, but rather that the effect varies between functional networks. Especially in networks that are related to intrinsically related functions, RS was shown to be preferable over NV.

Looking forward, the provided GE framework can be helpful in future biomarker studies where interpretability of a model is a must, by promoting both accuracy and interpretability. As drug development for neurological diseases advances, biomarkers that diagnose and monitor these diseases will become increasingly important, and methods like the proposed framework have the potential to play a crucial role in the development of such biomarkers. Further, the results here encourage the use of NV stimuli to improve signal properties of fMRI, that are important for biomarker research. However, the results highlight the importance to carefully chose the appropriate stimulus for the research question at hand. Generally, NV stimuli with social content should be preferred. Future biomarker studies might benefit from NV paradigms by selecting a stimulus that is specific to their research focus such as anxiety-inducing movie clips to probe patients with anxiety or a NV stimulus with distractors for patients with attention-deficit/hyperactivity disorder (ADHD). Finally, this dissertation provides a new NV dataset, which employs three NV stimuli with different levels of social content, that is publicly available to the neuroimaging community.

6 References

Adams, R.B., Rule, N.O., Franklin, R.G., Wang, E., Stevenson, M.T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., Ambady, N., 2010. Cross-cultural Reading the Mind in the Eyes: An fMRI Investigation. Journal of Cognitive Neuroscience 22, 97–108. https://doi.org/10.1162/jocn.2009.21187

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus Predictability Reduces Responses in Primary Visual Cortex. Journal of Neuroscience 30, 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Annals of the New York Academy of Sciences 1191, 133–155. https://doi.org/10.1111/j.1749-6632.2010.05446.x

Bloch, L., Friedrich, C.M., for the Alzheimer's Disease Neuroimaging Initiative, 2021. Data analysis with Shapley values for automatic subject selection in Alzheimer's disease data sets using interpretable machine learning. Alz Res Therapy 13, 155. https://doi.org/10.1186/s13195-021-00879-4

Bogdanovic, B., Eftimov, T., Simjanoska, M., 2022. In-depth insights into Alzheimer's disease by using explainable machine learning approach. Sci Rep 12, 6508. https://doi.org/10.1038/s41598-022-10202-2

Böhle, M., Eitel, F., Weygandt, M., Ritter, K., 2019. Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. Front. Aging Neurosci. 11, 194. https://doi.org/10.3389/fnagi.2019.00194

Braun, U., Plichta, M.M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., Walter, H., Seiferth, N., Kirsch, P., Meyer-Lindenberg, A., 2012. Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. NeuroImage 59, 1404–1412. https://doi.org/10.1016/j.neuroimage.2011.08.044

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct 217, 783–796. https://doi.org/10.1007/s00429-012-0380-y

Cao, H., Plichta, M.M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., Esslinger, C., Kirsch, P., Meyer-Lindenberg, A., Tost, H., 2014. Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. NeuroImage 84, 888–900. https://doi.org/10.1016/j.neuroimage.2013.09.013

Christoff, K., Ream, J.M., Gabrieli, J.D.E., 2004. Neural Basis of Spontaneous thought Processes.

Cortex 40, 623–630. https://doi.org/10.1016/S0010-9452(08)70158-8
de Jong, L.W., van der Hiele, K., Veer, I.M., Houwing, J.J., Westendorp, R.G.J., Bollen, E.L.E.M., de Bruin, P.W., Middelkoop, H.A.M., van Buchem, M.A., van der Grond, J., 2008. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain 131, 3277–3285. https://doi.org/10.1093/brain/awn278
DeGrave, A.J., Janizek, J.D., Lee, S.-I., 2021. AI for radiographic COVID-19 detection selects shortcuts over signal. Nat Mach Intell 3, 610–619. https://doi.org/10.1038/s42256-021-00338-7
Dinsdale, N.K., Bluemke, E., Sundaresan, V., Jenkinson, M., Smith, S.M., Namburete, A.I.L., 2022. Challenges for machine learning in clinical translation of big data imaging studies. Neuron 110, 3866–3881. https://doi.org/10.1016/j.neuron.2022.09.012
Dmochowski, J.P., Bezdek, M.A., Abelson, B.P., Johnson, J.S., Schumacher, E.H., Parra, L.C., 2014. Audience preferences are predicted by temporal reliability of neural processing. Nat Commun 5, 4567. https://doi.org/10.1038/ncomms5567
Dubois, J., Adolphs, R., 2016. Building a Science of Individual Differences from fMRI. Trends in Cognitive Sciences 20, 425–443. https://doi.org/10.1016/j.tics.2016.03.014
DuPre, E., Hanke, M., Poline, J.-B., 2020. Nature abhors a paywall: How open science can realize the potential of naturalistic stimuli. NeuroImage 216, 116330. https://doi.org/10.1016/j.neuroimage.2019.116330
Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. NeuroImage 59, 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017
Eickhoff, S.B., Milham, M., Vanderwal, T., 2020. Towards clinical applications of movie fMRI. NeuroImage 217, 116860. https://doi.org/10.1016/j.neuroimage.2020.116860
EU guidelines on ethics in artificial intelligence: Context and implementation, 2019.
Finn, E.S., Bandettini, P.A., 2020. Movie-watching outperforms rest for functional connectivity-based prediction of behavior (preprint). Neuroscience. https://doi.org/10.1101/2020.08.23.263723
Finn, E.S., Corlett, P.R., Chen, G., Bandettini, P.A., Constable, R.T., 2018. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. Nat Commun 9, 2043. https://doi.org/10.1038/s41467-018-04387-2
Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? NeuroImage 160, 140–151.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci 18, 1664–1671. https://doi.org/10.1038/nn.4135
Goh, J.O.S., Leshikar, E.D., Sutton, B.P., Tan, J.C., Sim, S.K.Y., Hebrank, A.C., Park, D.C., 2010.

Culture differences in neural processing of faces and houses in the ventral visual cortex. Soc Cogn Affect Neurosci 5, 227–235. https://doi.org/10.1093/scan/nsq060

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F.R., Nigbur, R., Waite, A.Q., Baumgartner, F., Stadler, J., 2016. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. Sci Data 3, 160092. https://doi.org/10.1038/sdata.2016.92

Harada, T., Mano, Y., Komeda, H., Hechtman, L.A., Pornpattananangkul, N., Parrish, T.B., Sadato, N., Iidaka, T., Chiao, J.Y., 2020. Cultural influences on neural systems of intergroup emotion perception: An fMRI study. Neuropsychologia 137, 107254. https://doi.org/10.1016/j.neuropsychologia.2019.107254

Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. Trends in Cognitive Sciences 14, 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject Synchronization of Cortical Activity During Natural Vision. Science 303, 1634–1640. https://doi.org/10.1126/science.1089506

Koster-Hale, J., Saxe, R., 2013. Theory of Mind: A Neural Prediction Problem. Neuron 79, 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., Golz, M., Hartigs, B., Hoffmann, N., Keil, S., Perlow, M., Peukmann, A.K., Rabe, L.N., Von Sobbe, F.-R., Hanke, M., 2015. Portrayed emotions in the movie “Forrest Gump.” F1000Res 4, 92. https://doi.org/10.12688/f1000research.6230.1

Lahmiri, S., Shmuel, A., 2019. Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer’s disease. Biomedical Signal Processing and Control 52, 414–419. https://doi.org/10.1016/j.bspc.2018.08.009

Mintun, M.A., Lo, A.C., Duggan Evans, C., Wessels, A.M., Ardayfio, P.A., Andersen, S.W., Shcherbinin, S., Sparks, J., Sims, J.R., Brys, M., Apostolova, L.G., Salloway, S.P., Skovronsky, D.M., 2021. Donanemab in Early Alzheimer’s Disease. N Engl J Med 384, 1691–1704. https://doi.org/10.1056/NEJMoa2100708

Nandi, A., Counts, N., Chen, S., Seligman, B., Tortorice, D., Vigo, D., Bloom, D.E., 2022. Global and regional projections of the economic burden of Alzheimer’s disease and related dementias from 2019 to 2050: A value of statistical life approach. eClinicalMedicine 51, 101580. https://doi.org/10.1016/j.eclinm.2022.101580

Nguyen, M., Vanderwal, T., Hasson, U., 2019. Shared understanding of narratives is correlated with shared neural responses. NeuroImage 184, 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jääskeläinen, I.P., Hari, R., Sams, M., 2014. Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. NeuroImage 102, 498–509. https://doi.org/10.1016/j.neuroimage.2014.07.063

Pohl, T., Jakab, M., Benesova, W., 2022. Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease. Int J Imaging Syst Tech 32, 673–686. https://doi.org/10.1002/ima.22657

Poulin, S.P., Dautoff, R., Morris, J.C., Barrett, L.F., Dickerson, B.C., 2011. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Research: Neuroimaging 194, 7–13. https://doi.org/10.1016/j.pscychresns.2011.06.014

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional Network Organization of the Human Brain. Neuron 72, 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Rikandi, E., Pamilo, S., Mäntylä, T., Suvisaari, J., Kieseppä, T., Hari, R., Seppä, M., Raij, T.T., 2017. Precuneus functioning differentiates first-episode psychosis patients during the fantasy movie Alice in Wonderland. Psychol. Med. 47, 495–506. https://doi.org/10.1017/S0033291716002609

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A.R., Schulz, J.B., Fox, P.T., Eickhoff, S.B., 2012. Modelling neural correlates of working memory: A coordinate-based meta-analysis. NeuroImage 60, 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral Cortex 28, 3095–3114. https://doi.org/10.1093/cercor/bhx179

Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H.C., Vermersch, P., Kuiper, M., Steinling, M., Wolters, E.C., Valk, J., 1992. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. Journal of Neurology, Neurosurgery & Psychiatry 55, 967–972. https://doi.org/10.1136/jnnp.55.10.967

Schmälzle, R., Häcker, F.E.K., Honey, C.J., Hasson, U., 2015. Engaged listeners: shared neural processing of powerful political speeches. Social Cognitive and Affective Neuroscience 10, 1137–1143. https://doi.org/10.1093/scan/nsu168

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. U.S.A. 106, 13040–13045. https://doi.org/10.1073/pnas.0905267106

Sneddon, I., McKeown, G., McRorie, M., Vukicevic, T., 2011. Cross-Cultural Patterns in Dynamic Ratings of Positive and Negative Natural Emotional Behaviour. PLoS ONE 6, e14679. https://doi.org/10.1371/journal.pone.0014679

Sporns, O., Betzel, R.F., 2016. Modular Brain Networks. Annu. Rev. Psychol. 67, 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Spreng, R.N., Mar, R.A., Kim, A.S.N., 2009. The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. Journal of Cognitive Neuroscience 21, 489–510. https://doi.org/10.1162/jocn.2008.21029

Thibeau-Sutre, E., Collin, S., Burgos, N., Colliot, O., 2023. Interpretability of Machine Learning Methods Applied to Neuroimaging, in: Colliot, O. (Ed.), Machine Learning for Brain Disorders, Neuromethods. Springer US, New York, NY, pp. 655–704. https://doi.org/10.1007/978-1-0716-3195-9_22

Van Dyck, C.H., Swanson, C.J., Aisen, P., Bateman, R.J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L.D., Iwatsubo, T., 2023. Lecanemab in Early Alzheimer's Disease. N Engl J Med 388, 9–21. https://doi.org/10.1056/NEJMoa2212948

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. The Human Connectome Project: A data acquisition perspective. NeuroImage 62, 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. NeuroImage 157, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes : A movie paradigm to improve compliance in functional magnetic resonance imaging. NeuroImage 122, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069

Wang, J., Ren, Y., Hu, X., Nguyen, V.T., Guo, L., Han, J., Guo, C.C., 2017. Test-retest reliability of functional connectivity networks during naturalistic fMRI paradigms: Test-Retest Reliability of Naturalistic fMRI. Hum. Brain Mapp. 38, 2226–2241. https://doi.org/10.1002/hbm.23517

Winkler, J.K., Fink, C., Toberer, F., Enk, A., Deinlein, T., Hofmann-Wellenhof, R., Thomas, L., Lallas, A., Blum, A., Stolz, W., Haenssle, H.A., 2019. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. JAMA Dermatol 155, 1135. https://doi.org/10.1001/jamadermatol.2019.1735

Zhu, W., Sun, L., Huang, J., Han, L., Zhang, D., 2021. Dual Attention Multi-Instance Deep Learning for Alzheimer's Disease Diagnosis With Structural MRI. IEEE Trans. Med. Imaging 40, 2354–2366. https://doi.org/10.1109/TMI.2021.3077079

Acknowledgements

I would first like to thank Prof. Dr. Simon Eickhoff for the opportunity to study in his institute, first as a master student and then for my PhD. I'm greatly appreciative of the support, guidance, and motivation throughout my studies. I would also like to thank Prof. Dr. Julian Caspers for his guidance and assistance as my secondary supervisor during my PhD project.
A particular thanks goes to PD Dr. Susanne Weis for assisting me in all aspects of my PhD. Thank you for your patience and for encouragement in times of need. Your supervision has helped me to become an independent researcher and prepared me for my future career in science. An additional thanks goes to all my colleagues from INM-7 who have always been very kind and made my PhD more enjoyable. I want to particularly thank the people from my working group Dr. Lisa Mochalski, Dr. Patrick Friedrich, Dr. Xuan Li, Dr. Julia Camilleri, Dr. Lisa Wiersch, Gianna Kuhles and Natalie Schlothauer for supporting me throughout my PhD and my Düsseldorf office colleagues, Kaustubh, Jürgen, Leo, Tobias, Georgios, Mamaka and Sam for the interesting lunch time discussions. I also want to thank Anna and Julia. Thank you both for bringing fun and joy to my years at the INM-7 and thanks for putting up with me.

A special thanks goes to:
Nevena, from bachelor's students to PhDs - what a ride! Thank you for making the PhD a fun (and almost enjoyable) experience. Although you are never on time, you are always there when I need you. Tobi, you never answer my texts - but I know you love me. Thank you for always supporting me, no matter what. Felix, you have enough on your plate as a BVB fan so I will **not** say "Ha, I made it before you!". Thank you for always listening to my problems, and sometimes even share personal information yourself, you're a great friend.

To my family. Thank you, Mama & Papa for all the trust and confidence you have in me. Thank you for all the support throughout my studies. I could not have done this without you. Thank you, Freddy, for making life fun even in the most challenging times. I know I can always count on you.

Most importantly, I want to thank my wife, Janina. Thank you for being my partner in crime. Your unwavering support has helped me through this PhD (and through life in general if I'm being honest). You and Merle make me the happiest man alive. I will forever be grateful for everything you have done for me.

Disclaimer: As of now, my heartfelt thanks are extended to the children I currently have. If, at any future point, additional children should exist, he/she is to be considered equally acknowledged and appreciated, and any expressions of heartfelt thanks should be understood to include him/her as well. Likewise, any friends who may not have been specifically mentioned are nonetheless sincerely valued and included in my gratitude.