universität innsbruck

Functional Connectivity in Internet Addiction:
A Predictive Modeling Approach

Luis **Glenzer**, 11726404
Luis.Glenzer@student.uibk.ac.at

Innsbruck, April 2024

Master Thesis

submitted to the University of Innsbruck, Faculty of Psychology and Sports Science to obtain the academic degree

Master of Science (MSc)

Master's program in Psychology

Supervised by:
Univ.-Prof. Dr. Carmen Morawetz
Institute of Psychology
Faculty of Psychology and Sports Science

Contributing Supervisor:
Ass. Prof. Dr. Matthias Schurz
Institute of Psychology
Faculty of Psychology and Sports Science

Abstract

Due to increasing omnipresence of digital services, numbers of people frequently using the Internet grow steadily and so does the prevalence of Internet Addiction (IA), a non-substance dependency. While the Internet may be a supportive tool for many, others suffer from excessive and addictive use accompanied by depression, anxiety and loneliness, however to date the underlying neural mechanisms are not fully understood. This work aims to fill the gap by applying a gradient boosted decision tree machine learning regressor trying to predict IA out of resting-state functional connectivity (rsFC). Successful prediction would contribute to the framework of digital phenotyping helping to identify relevant networks and robust markers of psychologic dysfunctionality, more valid than self-report questionnaires. Therefore, a previously collected sample of N=186 participants, each providing 3*15min resting state scans was analyzed hypothesizing that three networks are relevant for predicting IA, namely the reward-network, self-referential-network and attention-network. In a nested cross-validation scheme, both an original and a null model were fitted. The original model showed significantly better performance than the shuffled model (MAE: $M = 6.76 \pm 0.697$ vs. $M = 7.497 \pm 0.708$, $p < .001$; R^2: $M = .086 \pm .104$ vs. $M = -.105 \pm .103$, $p < .001$) however original feature importance for the respective networks did not significantly exceed that of the shuffled model, leading to the rejection of all hypotheses. As prediction mainly failed due to the distribution of the outcome scale, this work underlines the need for data examination in modeling tasks.

Words: 248

Keywords: Predictive Modeling, Functional Connectomes, Internet Addiction, Social-Media, rs-fMRI

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ACC	**Anterior Cingulate Cortex**
ADHD	**Attention Deficit Hyperactivity Disorder**
CbT	**Colsample by Tree**
CIAS	**Chen Internet Addiction Scale**
CV	**Cross-Validation**
DS	**Dorsal Striatum**
FC	**Functional Connectivity**
GBDT	**Gradient Boosted Decision Trees**
GIA	**General Internet Addiction**
IA	**Internet Addiction**
IAT	**Internet Addiction Test**
KS	**Kolmogorov Smirnov Test**
MAE	**Mean Absolute Error**
ML	**Machine Learning**
MRI	**Magnet Resonance Imaging**
MSE	**Mean Squared Error**
MW-U	**Mann-Whitney U Test**
NAcc	**Nucleus Accumbens**
OFC	**Orbitofrontal Cortex**
PCC	**Posterior Cingulate Cortex**
PFC	**Prefrontal Cortex**
PS	**Path Smooth**
ROI	**Region of Interest**

rs	**Resting State**
SRA	**Substance Related Addiction**
SNS	**Social Networking Sites**
SVM	**Support Vector Machines**
SW	**Shapiro Wilk Test**
VS	**Ventral Striatum**
VTA	**Ventral Tegmental Area**
YDQ	**Young Diagnostic Questionnaire**

1 Introduction

Since digital transformation in the 1990s, the digital world has grown, and its user numbers have risen continuously. This development can be observed in the ubiquity of Internet applications and the way they are changing human behavior. Until 2019, half of the world population, speaking of 3.5 billion people, used the Internet and one in three people engaged in social media. Not only are the absolute numbers remarkable, but so are the rates of growth. For instance, between 2016 and 2018, TikTok, a social media platform for sharing short videos, added 20 million new users on average per month (Ortiz-Ospina & Roser, 2023). The Internet offers a variety of opportunities that improve the daily lives of many people and increases flexibility of fulfilling several needs such as socializing, accessing a variety of information, and ordering goods via online-shopping in a time-efficient manner. Even social networks, despite the widespread belief that they negatively impact mental health, are neither inherently good nor bad. The effects and valence of use is dependent on the dosage and usage pattern, and users may even have favorable psychological effects (Akram & Kumar, 2017; Verduyn et al., 2017). However, Internet services and applications as well as smartphones have characteristics that lead to extensive use, which further can cause addiction. As with other addictions, Internet addiction (IA) is characterized as a strong urge to use the Internet, despite that the current usage pattern leads to detrimental situations, especially socially (Young, 1998). Besides the urge to use, also the need to reduce usage is present but not achievable. It could be shown that the relation between mental health issues and IA is reciprocal, such that mentally unstable individuals tend to develop IA, which in turn worsens their condition. Therefore, IA can be seen as both a consequence and antecedent of mental health problems, e.g. depression (Ostovar et al., 2016; Yang et al., 2022). In addition to having a large user base, social media is also used extensively. According to data from 2018, Americans used social media for an average of six hours a day, mostly on mobile devices (>50%) (Ortiz-Ospina & Roser, 2023). These findings highlight the need for additional research on the effects of social media and the Internet. As a result, the relevant literature examines the connection between addictive Internet use and the brain, however neurological studies are underrepresented, and the results are found to be contradictory. Various neuronal correlates of this disorder have been identified, however most of the findings, particularly for resting-state (rs) data, cannot be replicated, which contradicts the understanding of the underlying mechanisms (J.-T. Sun et al., 2023). Furthermore, digital behavior is constantly evolving, new sites, games and services appear, which highlights the need to move forward in the examination process and researchers call for a better understanding of the neurological underpinnings of social media and Internet use (Montag et al., 2023). This study addresses the issue of various contradictory findings in a conclusive attempt under the framework of systems neuroscience by incorporating and combining the brain regions found to be associated with IA into three functional networks (reward,

self-reference, attention) in IA based on rs functional magnetic resonance imaging (fMRI) functional connectivity (FC). Those networks are tested on their predictive power of IA. Considering the complex interplay of brain regions in behavior, hypothesis testing is to be achieved by predictive modeling of the Internet Addiction Test (IAT) (Young, 1998) using multivariate machine learning regression, aiming to identify underlying mechanisms, relevant neurological features and biomarkers of IA.

1.1 Internet Addiction

The evolution of the digitalized era, especially with the invention of the Internet has brought many improvements. Information and goods have become widely accessible, social connections around the globe more convenient and entertainment more individually consumable. Particularly the need for connectedness drives people to use the Internet on a variety of devices and keeps their activity on several social networks. That the fulfilment of these needs has become easier to meet, may have positive effects, however the valence of a new technology is moderated by the form of its use. Like with every mean of need fulfilment, there is a threshold marking a functional vs. dysfunctional use. Considering possible detrimental effects, Young (1998) was one of the first to highlight the addictive nature of the Internet and rising prevalence rates of dependent users on an international level prove those claims (Lozano-Blasco et al., 2022; Pan et al., 2020). A meta-analytic finding revealed the average prevalence for general Internet addiction (GIA) to be 7.02% with an average growing rate of 0.6% per year (Pan et al., 2020).

1.1.1 Addictive Properties

Explaining the reasons for this development, Greenfield (2012) worked out five main properties of the Internet, shaping its desirability: (1) Content: the information, media (e.g., music, videos) and means of social connection available on the Internet are inherently pleasurable and due to its user numbers and the heterogeneity of individuals, a variety of content can be found online, so that almost every preference can be served. (2) Process and access/availability factors: The Internet is a steady accessible world, enabling the user to fulfil a wide range of desires anonymously without time and space restrictions. Most services are ‘free, so there is no need to spend money on. Therefore, the threshold from desire to gratification is almost absent. (3) Reinforcement and reward factors: The services provided by the Internet are based on variable-ratio reinforcement schedules, known to be particularly addictive. Content is presented apparently random but well-trained algorithms place rewards (particularly pleasurable content) at the most effective positions. As users will never know, when to expect a reward, the constant anticipation of the next reward keeps them engaged. Besides

those primary gains, the Internet provides secondary a tool to conceal and divert themselves from negative real-life events such as negative feelings, performance pressure and social issues. (4) Social factors: The Internet is unique in enabling social connections with a high degree of control over the distance preferred by users. For example, Facebook and Instagram only implemented like buttons while there is no form of disliking except of commentaries or messaging, other users can be approached easily with no requirement of spontaneous actions and are rather impersonal. If some contacts become uncomfortable and bring issues with them, they can be blocked. In addition, the way to present oneself in social media can be well planned and made up so that characteristics which are favorable will be displayed while others will be held back. Thus, social interaction can be organized individually, maximizing comfort (e.g. reputation) while minimizing distress (e.g. social anxiety). (5) Gen-D factors: Generation-Digital users are persons which have been raised with this technology, creating feelings of naturality in use. This means that children learn from early on that the possession and the usage of smartphones or other digital devices is normal and that a large part of daily life takes place online. This, in addition leads to perceived power over elderly people without Internet or digital experiences.

Besides those inherent properties, providers of Internet services design their products to increase the attractiveness and binding effects of those, resulting in excessive use. Montag et al. (2019) worked out several mechanisms which are implemented in both social media sites as well as online gaming. Neyman (2017) highlighted these characteristics for social media and the findings of both will be combined and presented here. Note that online gaming is an umbrella term for every kind of digital games. Hence online does not necessarily mean the connection between multiple players via the Internet. However, the most popular and addictive games take place online and rely on the connection between different players.

Conclusively, social media and online games: (1) Evoke the need for social reciprocity and create social pressure. In social media this refers to giving likes back or answering received messages. For example, this mechanism can be found on WhatsApp with its blue ticks. Knowing, that the recipient of a message has received and read it, evokes feelings of social pressure in the recipient to reply. As receiving and reading messages is rewarding and mostly pleasurable, the ticks turn blue fast, and responses are expected immediately. Even though this feature can be turned off, it is activated by default and arguing with the ‘power of defaults’ the feature is widely used. In online games users often join groups and are asked to interact with each other. In some cases, the formation of groups is necessary to fulfil tasks or solve quests which results in pressure for people to be online and give support when their group tries to progress. (2) Lead to infinite scrolling or gaming by evoking a feeling of flow which is achieved by keeping tasks as complex that it is still interesting while easy enough to go on using. This flow is known to be capable of disturbing feelings of time and space. Newsfeeds, where every post appears after the other without the requirement of switching

sites, or games with the combination of short levels, low difficulties, and high rewards, result in infinite usage. (3) Create user investment by leaving users to build up their profiles in social media or to invest in games, which enables faster progress. These behaviors result in high value attribution to the own profile or the game which further strengthens the attraction effect. (4) Maintain an illusion of choice by providing a variety of content while in fact they are individually tailored and one-sided. Huge datasets collected by gathering user data through the services in question build the basis for machine learning algorithms which tie certain user types to specific content. This enables the providers to predict the expected pleasure of the specific content and arrange them accordingly. Users mostly do not know about the decision criteria of the algorithm so the feeling of choosing out of a wide variety of content is maintained. In turn the attractiveness leading to increased usage makes the algorithms better and reinforces it further. (5) Implement gamification, which refers to the provision of small rewards in response to certain actions. This gives users the feeling of achievement which often is suggested to be shared with the own social network. As users start so sense the possibility to achieve rewards for certain actions on the website or app, it becomes more attractive. In addition, providers of these services can shape behavior in the sense of operational conditioning. These mechanisms can particularly be observed in fitness apps, where various statistics are formed and based on the performance, rewards like badges or the placement in a leaderboard are awarded. Furthermore, the performance can be shared with the own networks which in turn results in further likes and prestige and required reciprocity.

Those properties explain the success of Internet applications and services but further provide a risk of dysfunctional behavior, namely IA. Many different names and definitions came up in the ongoing study of this disease. The different definitions agree on several properties like an overuse of the Internet, the strong desire to use it, the failed willingness to reduce the usage and detrimental outcomes mainly on a social level (Zou et al., 2017). Combining these characteristics, IA will be defined in the current study as an obsessive overuse of the Internet, caused by mental or behavioral disorder, and accompanied by an unrealizable desire to both re-use and stop or reduce to use, resulting in individual and social problems and eventually somatic symptoms.

1.1.2 Operationalization

One of the first instruments to capture IA was the Internet Addiction Test (IAT) by Young (1998). Within the study to record the characteristics of IA, it could be shown that highly Internet addicted individuals, diagnosed by an early version of this questionnaire, spent time worth a full-time employment on the Internet. Back to the time of the study, most used services were chatrooms, compared to social media services today and newsgroups. Except of the physical, every other domain in life was impaired moderately to severely in addicted individuals. It was hypothesized, that this

addiction works almost the same way as 'traditional' addictions like that of alcohol. Today the diversity and omnipresence of the Internet has drastically increased and as research on IA has progressed, it has been pointed out that more specific categorizations are needed, arguing that it is not solely the Internet leading to addiction, but the content consumed. The Internet serves as a platform where several needs can be satisfied like social desires, materialistic needs, competitive needs and sexual needs (Brand et al., 2016). Following this kind of logic, Montag et al. (2015) provides data supporting the distinguishable character of IA as some specific forms of IA did not overlap significantly with GIA as correlations ranged from $p = .13$ to .76, so that a division in GIA and specific IA is suggested.

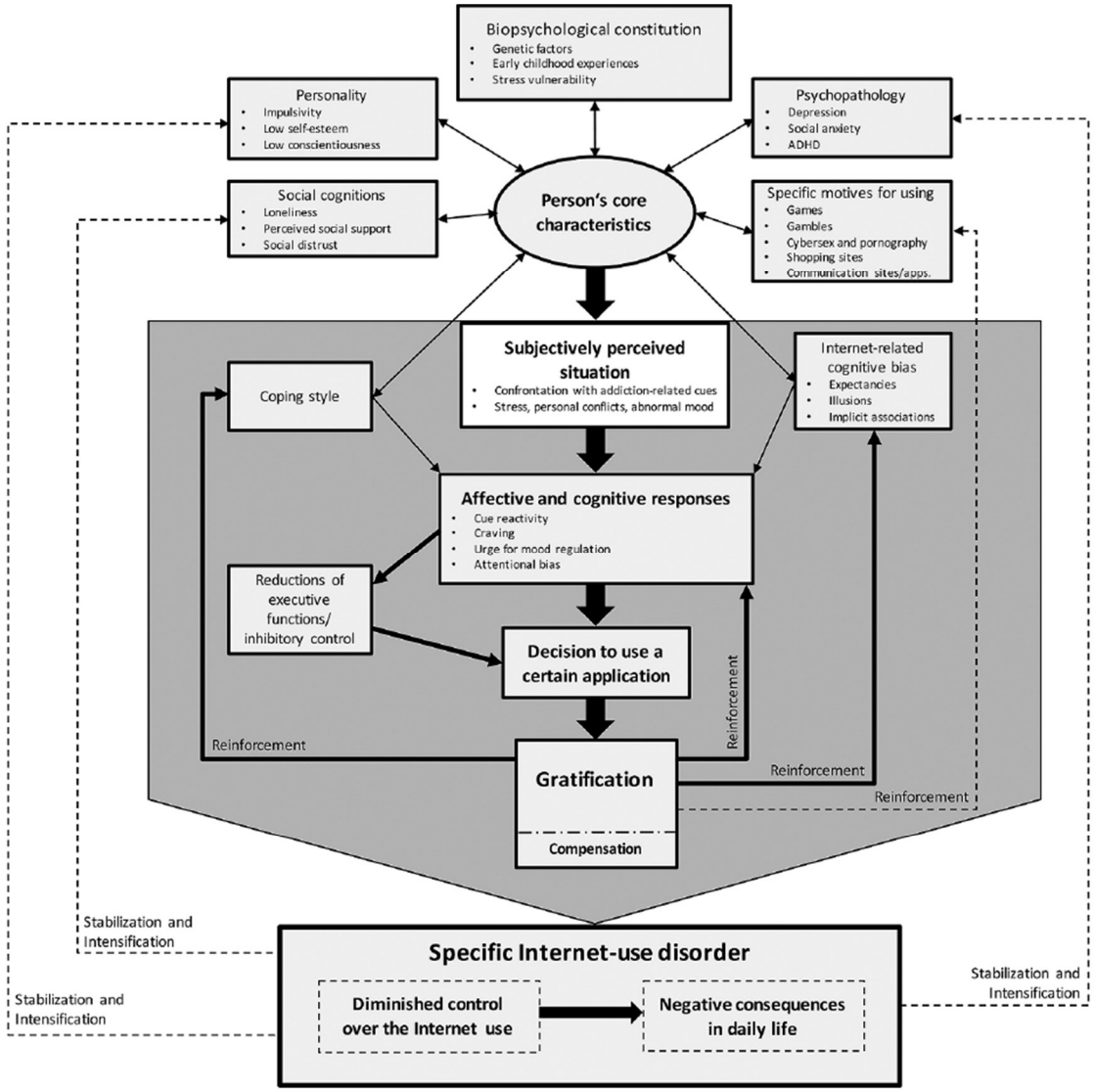
Based on this assumptions and evidence towards different underlying mechanisms, multiple scales have evolved (e.g. Young diagnostic Questionnaire (YDQ), Chen Internet addiction scale (CIAS)), however the IAT (Young, 1998), measuring GIA still is the most prominent scale and widely used. In a review it was found, that between 1996 and 2018 the IAT was used in almost 44% of all scientific publications concerning IA, followed by the YDQ (29%), CIAS (11%) and others (16%). It must be stated, that compared to other instruments such as the YDQ, the IAT measures IA more broadly, including more peripheral symptoms and therefore demonstrating higher prevalence rates (Pan et al., 2020). In addition, it has been shown that it is strongly associated to social media use (SMU) and mainly captures the social aspect of need satisfaction, as Montag et al. (2015) found that even though some specific forms are not captured by GIA, online social network addiction shows the highest correlation to GIA. In addition Yadav et al. (2013) could show, that the use of social networking sites (SNS) was 50% more frequent among participants with present IA and social networking and using chat rooms significantly predicted IA. This leads to the conclusion, that SMU is captured by the operationalization of IA. Hence, the main concern of this work lies on GIA, but the specific focus will be set on SMU.

1.1.3 The I-PACE model

When attempting to understand the mechanisms behind the development of IA, the Interaction of Person-Affect-Cognition-Execution (I-PACE) model by Brand et al. (2016) must be considered. This model, which is derived by theoretical and empirical assumptions of substance related addictions (SRAs), incorporates all the many expressions of IA, and offers a theoretical framework for explaining the onset and maintenance of this non-SRA (nSRA) disorder. Its methodology is to identify factors for specific types of IA. However, as GIA is the subject of the current study, the key elements essential to the general form of this disease will be addressed. This section is structured based on the I-PACE model enumerating the different stages (Figure 1).

Figure 1.

I-PACE Model (Brand et al. (2016)



Note. Schematic model of development and maintenance of IA influenced by personal characteristics, affective and cognitive responses, and executive function.

(P) The first part of the model refers to the personal characteristics of individuals developing IA. Predisposing variables can be found in different domains such as genetics, psychopathology, and personality. IA and several mental disorders like depression, (social) anxiety and attention deficit hyperactivity disorder (ADHD) often occur as comorbidities. Furthermore, among personality traits, high impulsivity, low self-esteem, low conscientiousness, high shyness, high neuroticism, a tendency to procrastinate and low self-directedness are observed among individuals with IA. They also report a perceived lack of social support, feelings of isolation and loneliness. The motivation of use can at least be partially derived from these characteristics. For example, people with high social anxiety

and shyness feel more confident in rather impersonal social contact enabled by social media, while people with ADHD could offset their lack of dopamine through the various rewards provided by Internet services. Possibly a high number of Facebook friends, followers and likes stands against feelings of isolation and loneliness. These motivations are caused by strong desires which cannot be fulfilled through other domains, but the Internet. Therefore, usage increases and heightened perception of social pressure makes them even more prone to fulfil expectancies (sharing one's life, commit to a gaming group), resulting in excessive use.

(A & C) The letters A and C refer to the affective and cognitive responses to external or internal stimuli. In line with the findings of depression and anxiety as comorbidities, the subjective perception of stress becomes a relevant factor. It seems to influence decisions to focusing short-term rewards and therefore to use the easily, infinitely accessible Internet as a dysfunctional coping mechanism. It is stated that the cognitive belief of the Internet as a distracting and helpful (avoidance expectancy) as well as pleasurable (positive expectancy) space, which provides the possibility to escape from stressors make people more prone to use it. While the perceived positive effects may be true for short-term, they can turn out detrimental in the long run, such that the Internet is seen as a (dysfunctional) coping mechanism. Once this affective and cognitive connection has been built, feelings of *craving* to use the Internet arise in the need for coping. Craving is defined as an intense urge to satisfy a specific need, which can be observed in all kinds of addiction and is accompanied by diminished behavioral control. It is seen as a key mechanism, increasing the difficulty to stop the dysfunctional consumption, and can be triggered by heightened *cue-reactivity* in interned addicted individuals. Cue-reactivity means the physiological and psychological response on an Internet related cue and those cues can appear external (e.g. receiving notifications, advertisements for new games) or internal (e.g. boredom, loneliness, stress, etc.). They pose a risk for abstinent individuals for relapse and the craving which arises out of the cue-reactivity is accompanied by *attentional biases*. Attentional biases in IA refer to the tendency to exhibit heightened attention and increased focus towards Internet related cues, such that not only the reaction on these is increased but also, they are recognized and processed faster. This results in a stronger occupation with the Internet and impaired cognitive function when attention for Internet cues is activated. Observations of this process go in hand with the previously outlined positive and negative expectancies, such that it can be described in the following two paths: (1) Individuals with IA are more likely to perceive Internet related cues, react stronger towards them and therefore feel a craving to use it. This results in positive expectation towards the usage and with high probability in turn to usage itself. (2) Beginning with negative mood, the individual searches for coping strategies and as feelings of relief are closely linked to Internet used, the attention and motivation could be focused on the Internet, which would lead to the beginning of process 1. The combination of cue-reactivity, craving, attentional bias, and expectancies results in repeated Internet engagement.

(E) The E part refers to executive function, inhibitory control and decision making with respect to usage. It is important to consider, that Internet addicted individuals potentially are aware of the detrimental effects of their behavior. If this holds true it is to be clarified why these individuals are not able to stop the dysfunctional behavior. This can be explained by a lack of *salience attribution* and *inhibitory control*, which mainly has been examined by neuroscientific findings of reduced functioning of the prefrontal cortex (PFC) and limbic structures (1.2.1 *Relevant Structures and their Roles*) and can also be found in SRAs (Volkow & Boyle, 2018). Salience attribution refers to the perceived importance of a cue and the attentional resources used to keep it salient. This mechanism is closely related to the interplay of attentional biases and expectations from the previous section but is proposed here as a cause of diminished inhibitory control, which describes the individual's potential to control impulses derived from short-term desires and motivations to use so that long-term goals such as finding functional coping mechanisms can be pursued. As stated in the definition, IA is associated with the desire to stop or reduce use, but beyond a certain level of severity of addiction, this change becomes almost impossible. Results in this direction are inconsistent but it appears that decision-making and executive functioning is impaired in Internet addicted individuals to a comparable degree as in alcohol-dependent individuals. Studies concerning inhibitory control did not find strong effects, however this was dependent in part on whether the cues were related to the specific type of addiction or rather general characteristics of the Internet. It is therefore hypothesized, that effects will be stronger for specific cues paired with specific Internet use disorders. This supports both the quest for a more specific investigation of IA and the assumption that motivational factors in combination with susceptible prior characteristics are strong enough to keep individuals engaged despite negative outcomes.

1.1.4 Clinical Relevance

The clinical relevance of IA becomes evident when the pathological relations are considered. It has been demonstrated that IA is strongly associated with depression, anxiety (Bisen & Deshpande, 2020; Ko et al., 2012; Moon et al., 2018; Stănculescu & Griffiths, 2022; Yadav et al., 2013), loneliness, low self-control (Ostovar et al., 2016; Özdemir et al., 2014) and ADHD (Evren et al., 2018; Marin et al., 2021; Sariyska et al., 2015; Yoo et al., 2004). Furthermore, constant occupation with mobile phones can lead to stress and lower well-being (Montag, 2019; Vahedi & Saiphoo, 2018). In line with the definition and the I-PACE model (Brand et al. (2016), see *1.1.3 The I-PACE Model*) that it is mainly mental health issues, which cause the development of IA, people suffering from mental health problems were found to be more prone to develop IA (Tóth-Király et al., 2021). However, it can be assumed that this relation is bidirectional as a longitudinal study with a reasonable sample size was able to show, that pathological Internet use could lead to the development of

depression (Tian et al., 2021). Based on these findings, the interrelation between those clinical states forms a self-reinforcing cycle, however, until now it is not clear which properties are antecedents and which consequences and this applies also to the neurological underpinnings of IA. However, one longitudinal study reported, that a four-week excessive SMU intervention for people low in SMU could change individuals FC, so that they exhibited patterns likewise found in people high in SMU (Hu et al., 2022).

Not just GIA but also SMU shows this relation to mental health issues, explainable mainly by social comparison and following negative affect. Facebook use was found to be related to depression and negative affectivity if social comparison and envy are present and as one of the properties of social media is the urge of positively biased presentation of ones' self, these feelings are commonly observed in users (Appel et al., 2016). In addition, the time spent on Facebook is positively associated with social comparison and depression without any influence of gender, but social comparison mediated the relation (Steers et al., 2014). In line with that, technology-based social comparison and feedback-seeking are associated with depressive symptoms (Nesi & Prinstein, 2015). Those insights are supported and extended by a review stating that passive use of social-networking-sites evokes feelings of social comparison and envy and therefore is related to depression while active use can have positive effects for mental well-being. However, this relation is less robust than the negative relation (Verduyn et al., 2017). Active use here is defined as active exchange with other participants in either targeted way such as texting personally or nontargeted such as posting status updates or sharing photos, while passive use refers to only monitoring the content shared by other users. It is important to note that this review focused on studies with healthy participants and that individuals with previous mental health issues and personality traits such as low self-esteem, high shyness and high neuroticism are more prone to develop IA (Brand et al., 2016). It therefore can be concluded, that individuals with these characteristics use Internet services such as social media excessively but tend to use social media passively, so that they will not benefit from possible positive effects and rather suffer even more from the upward comparison they are facing (Nesi & Prinstein, 2015).

Besides the most studied affective disorders like depression and anxiety, another issue related to SMU is body dissatisfaction. A longitudinal study found that SMU predicted body dissatisfaction and this effect is the same for boys and girls (Schønning et al., 2020; Vries et al., 2016). Highlighting the comparative character and possible detrimental effects on self-identity, social media elicits body dissatisfaction and is followed by increased internalizing symptoms which were operationalized as depressive and anxiety symptoms. This effect was significantly greater for highly visual social media (such as Instagram or Snapchat) and it could further been found that body dissatisfaction is a mediator of the relation between SMU and internalizing symptoms (Marengo et al., 2018). These findings are supported by a meta-analysis, stating a positive relation between social media use and body image

disturbing with a moderating role of the type of social media such that appearance-focused social media obtained stronger effects (Saiphoo & Vahedi, 2019).

Thus, the Internet and social media combine addictive properties with negative health impacts. As individuals with prior mental health issues spend more time on social media or online gaming such that they are more prone to develop IA, the risk of decreasing in well-being and mental health is high and presents a major problem. The relations to mental health disorders are particularly concerning in the light of emerging prevalences of IA over the years (Lozano-Blasco et al., 2022; Pan et al., 2020) due to the rise of digitalization, which corroborate the necessity of further investigation.

1.2 Neuronal Basis of Addiction

Early in addiction research, the dependence of a substance or behavior has been seen as a lack of willpower and "moral deficiency", resulting in high stigmatization. Even compared to other mental illnesses, addicted individuals face stronger rejections and prejudices (Barry et al., 2014). Today, it is argued, that highlighting addictive behaviors from a neuroscientific perspective can reduce those stigmata and provide a more objective investigation of this disease. This can be achieved by explanative relations between the mechanisms behind dependencies and the brain functioning, which are difficult to be influenced by pure willpower (Buchman et al., 2011). Besides changing the perspective on the responsibility of addictive behavior, neuroscientific findings contributed to the general understanding and therefore provided reliable opportunities for prevention and treatment, not only for SRAs (Volkow et al., 2016). Hence, neurological underpinnings of addiction are of particular importance and have been studied extensively resulting in the detection of antecedents and outcomes.

1.2.1 Relevant Structures and their Roles

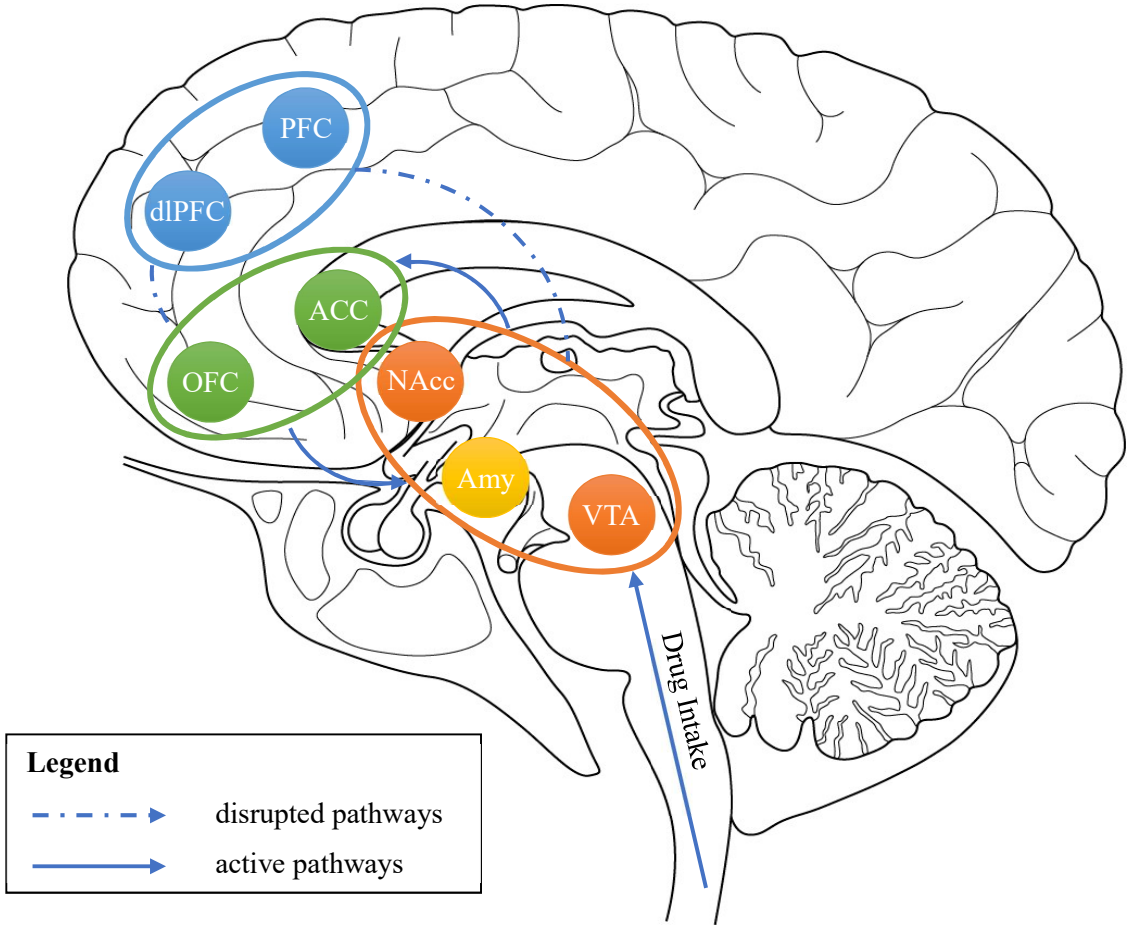
In the following paragraph (*1.2.2 Process Model of Addiction in the Brain*) a schematic model of the development and maintenance of addiction based on its neurological characteristics will be presented (Volkow & Boyle, 2018). To increase the comprehensibility of the following mechanisms, the relevant regions and their role in behavior will be outlined previously. The order of presentation is similar to the appearance in the model, so that the process can be understood more easily. To gain understanding of the localization of these structures, their allocation to specific networks and connectivity are depicted in Figure 2.

Ventral Striatum (VS) and *Nucleus Accumbens* (NAcc): The NAcc is located at the bottom part of the ventral striatum and a key region in the reward network. It regulates motivation for drug seeking in addiction. Given its location in the basal ganglia region in the forebrain there are

connections to PFC. Those connections are assumed to be a mediator of cognitive control over the craving to further engage in drugs. As the VS and NAcc gets activated strongly when it is exposed to rewards, a dysfunctional VS and NAcc may be able to override the executive functions of the PFC (Wadsley & Ihssen, 2023). *Ventral Tegmental Area (VTA):* The VTA is located near the top of the brain stem and is highly intertwined with the VS and NAcc as they belong to the 'feels better'

Figure 2.

Process Model of Addiction



Note. Mode of action of drug intake in the brain. Substances and intake effects enter through the brainstem and affect reward circuits (orange). Dopamine is released and the link is reinforced by the learning and motivational circuit (green). In the withdrawal phase the amygdala (Amy; emotion area; yellow) gets active. The control circuits (blue) cannot prevent this interplay. Graphic created by Gill Brown [Link] and adapted with permission for individual purpose.

pathway, processing the positive (pleasurable) and negative (stress reduction) mechanisms of usage (Brand, 2022). *Amygdala*: The amygdala is part of the limbic system and known for its role in processing emotions and motivation. The amygdala shows activation when drug related cues are

presented, and it further is associated with reward expectancy. The amygdala is underlying changes with respect to addictive behavior and it is assumed that this causes problems in emotion regulation, which in turn complicates the resistance against craving (Wadsley & Ihssen, 2023). *Orbitofrontal cortex* (OFC): The OFC is part of the PFC and therefore responsible for executive functioning, decision-making and regulation of impulsivity. Connections to striatal regions involves the OFC in reward functions and with both roles it can be seen as a mediator between the memory of positive effects of 'drug' engagement and the decision to engage (J.-T. Sun et al., 2023). *Anterior cingulate cortex* (ACC): The ACC is part of the cingulate gyrus which in turn belongs to the limbic system. In this system, emotional control, action outcome learning and memory are processed and therefore it is seen as part of the cognitive control network (Wadsley & Ihssen, 2023). It receives information from the OFC about expected rewards from certain actions and behaviors. Those responsibilities and connections make it a crucial part in remembering the link between drug consumption and feelings of pleasure and following the action to engage in the drug even more (J.-T. Sun et al., 2023). *Prefrontal Cortex* (PFC): Besides the OFC, lateral and medial regions belong to the PFC. While the lateral regions are responsible for cognitive control and executive functions, the medial regions are associated with motivation, decision making and self-referential processing. Both outlined parts can be seen as the controlling instance of the brain, responsible for action-outcome decision. If it's working becomes dysbalanced, the 'requests' from other structures such as VS, NAcc, OFC and ACC cannot be handled properly anymore (Wadsley & Ihssen, 2023).

1.2.2 Process Model of Addiction in the Brain

In their review, Volkow and Boyle (2018) provide a process model describing the alterations in brain functioning in addicted individuals. The characteristics of an addiction, namely the perceived rewards, conditioning to the drug and its cues, negative mood, increased stress reactivity and craving could be mapped to several neuronal circuits. In the beginning of an addiction, the 'drug' causes bursts of dopamine in the reward system, consisting of the nucleus accumbens (NAcc), dorsal striatum (DS) and VTA (Brand, 2022; Volkow et al., 2019), linking the drug with feelings of pleasure and therefore reinforcing the consumption. Repeated drug intake makes the individuals reaction towards rewards attenuated, which causes even stronger dependency to get at least some pleasurable effect. When this effect wears off, the so-called withdrawal phase begins, where the individual is confronted with feelings of guilt, anxiety, negative mood, increased sensitivity to stress and significant dysphoria. It is mostly observed in individuals with longer drug exposure history and depends on the pharmacological and societal characteristics of the drugs. However, it can be assumed, that every dependently use creates this pattern of feelings. This phase is neurologically characterized by basal forebrain areas like the extended amygdala and habenula and is associated

with the development of feelings of craving. At the same time, the reward and motivation system are hypofunctional, contributing to anhedonia and the aversive state during withdrawal. This unpleasant experience leads to the desire to counteract with additional drug intake. When the craving phase starts, the conditioned link between drug intake and pleasure response becomes activated and triggers dopamine release in the striatum, which drives the motivation to seek for the drug. In this phase, prefrontal circuits including the OFC and ACC, which are related to salience, as well as circuits in the hippocampus and amygdala are involved. Additional contributing factors are (1) the hypofunctionality of the prefrontal circuits responsible for executive functions and emotion regulation, causing impulsiveness and decreased inhibitory control over consumption behavior, and (2) the hyperfunctionality of limbic circuits increasing stress reactivity and decreasing sensitivity to nondrug rewards. Both alterations are consistently found in repeated drug exposure and represent the lack of ability to balance short-term reward vs. long-term goals. Furthermore, these alterations are assumed to be favored by drug use, such that again a self-reinforcing cycle increases the difficulty to stop usage. The validity of these findings has been confirmed by successful intervention studies targeting enhanced dopamine signaling in prefrontal areas, counteraction on the enhanced reactivity to cues in the PFC, amygdala and VTA, and counteraction of the negative mood during the withdrawal state on neurotransmitters.

1.2.3 Substance Related Addictions and non-Substance Related Addictions

These mechanisms are not unique to SRAs, as it is argued, that addictive behaviors, both non-substance- and substance-related, share overlapping features such as ‘loss of control’ and ‘despite adverse consequences’ and therefore symptoms and even neurological characteristics are similar to some extent (Zou et al., 2017). Several studies have demonstrated the involvement of specific brain regions in both forms of addiction and common mechanisms like cue reactivity, inhibitory control and craving were revealed (Brand et al., 2020). Increased cue reactivity occurs not only in substance related addictions but also in gambling, gaming and shopping disorders and can be observed in the brain in the caudate nucleus, inferior frontal gyrus, angular gyrus, precuneus, and inferior network with results consistent with meta-analyzes on cue reactivity in substance-use disorders (Starcke et al., 2018). Reduced inhibitory control in IA could be observed with alterations in the PFC over a wide range of studies (Brand et al., 2016) and cognitive control in general, processed by the PFC has been associated with addictive behavior in SRAs, but also with IA (Brand et al., 2014). A systematic review found that the neuronal correlates of risk-related decision making in individuals with SRA and non-SRA were similar in the OFC and striatum as well in ventral ACC and dACC while differences were found in PCC and precuneus, although these regions seem to be affected in both forms of addiction (Hüpen et al., 2023). It is also noted that the similar neuronal

underpinnings to SRAs become apparent in the reward system. Relevant structures here are VS and DS, which are often associated with dopamine release, the OFC, involved in craving, and the dlPFC, responsible for behavior control (Brand, 2022; Brand et al., 2016). While many studies provide insights about the neuronal features of this disorder, these findings need to be revisited.

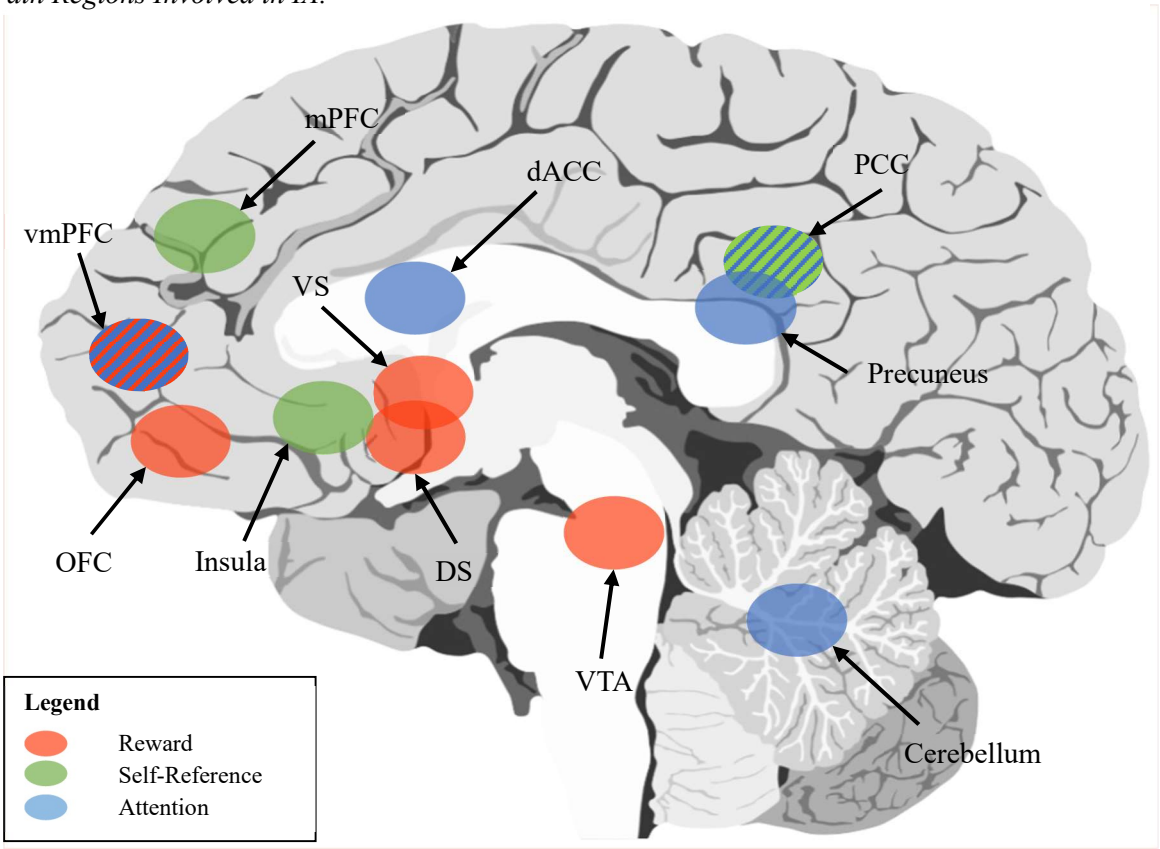
The current findings on the neuronal basis addiction and particularly IA, with respect to a functional connectomes approach (see *1.4. Functional Connectomes*) give rise to the research question if variations in the FC of specific networks can predict the level of IA. Which networks are assumed to be predictive of IA will be assessed in the following paragraph.

1.3 Neuronal Basis of IA

This section concludes current theoretical assumptions as well as evidential findings on the neuronal underpinnings of IA with respect to its associated networks to contribute to the framework of functional connectomes. As previously outlined, studies found many regions associated with both

Figure 3.

Brain Regions Involved in IA.



Note. Brain regions associated with IA and assigned to their respective networks. Graphic created by Gill Brown [Link] and adapted with permission for individual purpose.

SRA and nSRA. However, this variety could not be concluded to coherent networks, particularly with respect to IA, as out of a meta-analysis of 31 rs-fMRI studies the FC of only two structures (PCC, insula; whole-brain) were consistently associated with IA in an *activation likelihood estimation* but none were found in the *seed-based d mapping with permutation* of subject images (J.-T. Sun et al., 2023). Therefore, the current state of research will be outlined more broadly, to discover functional networks (Figure 3), which play a role as factors and effects of IA. Findings will be incorporated in the formulation of the hypotheses to test within this study (Figure 4).

1.3.1 Reward

Like previously elaborated, using the Internet provides many rewards and most research investigating the neurological underpinnings of addictions, both SRAs and nSRAs, reveals the involvement of reward circuits. The findings supporting the association between addictions and the reward network and among others, dopamine pathways, are consistent over multiple studies (Brand, 2022; Volkow & Boyle, 2018). On the Internet, which is mainly an information platform, searching for information via search engines such as Google is rewarding, as it satisfies the mechanism of seek and find in providing instant answers to almost any question (Parsons, 2017). Other services such as social media or online gaming are by design even more rewarding and in turn have addictive potential. Variable-ratio schedules, known from gambling and operant conditioning are widespread, as it is unpredictable, which results a search engine will return, what articles will be displayed on news sites and whose friend post you will see on a social networking site, thus maintaining surprise and curiosity. In particular social networks as platforms for sharing and mapping one's life are rewarding as they offer the opportunity to gain social prestige, which triggers the dopamine system in the brain (Parsons, 2017). A meta-analysis could reveal that individuals with IA have steeper delay discounting rates, which means, that short term rewards with lower value are preferred before long term rewards. This in turn indicates the lack of patience, inhibitory control and impulsivity when waiting for rewards (Cheng et al., 2021). In addition, higher activation in regions of the reward-area has been observed when comparing photos with many likes to photos with few likes in a simulated Instagram paradigm (Sherman et al., 2016) and reward-related activities accessed through reputation gains predicts Facebook use (Meshi et al., 2013). Additionally, a higher daily frequency of visiting Facebook is associated with a lower volume of grey matter of the NAcc, which belongs to the VS (Montag et al., 2017), which is known to be associated with cue-reactivity and craving in Internet addicted individuals (Brand et al., 2016). To capture the reward processes associated with IA, the reward network proposed by Meshi et al. (2015), consisting of the ventromedial prefrontal cortex (vmPFC), VS and the VTA, is adapted.

However, other regions of interest (ROIs) have been linked to addiction and are part of the reward network. The OFC is thought to play a role in the reward system as it is responsible for decision making and regulation of impulsivity (Chun et al., 2018; J.-T. Sun et al., 2023). It is involved in addictive behavior when relying on the operant conditioning of learning framework (Y. Sun & Zhang, 2021). The OFC is important for the representation and expectation of rewards and for ranking and selecting the value of rewards based on their identity (Howard & Kahnt, 2021). In addition, alterations in the FC between the OFC and midcingulate cortex (MCC) and NAcc have been associated with excessive smartphone use, and the FC including the OFC are considered to be important for cognitive control of emotional stimuli, including reward (Chun et al., 2018).

Further, the DS, including the putamen and caudate nucleus, is involved in choice impulsivity in relation to rewarding behavior. In particular, communication between the VS as the learner of the tie between behavior and reward and the DS as the actor to decide about actions, is considered to be the main mechanism (Balleine et al., 2007; Kim & Im, 2019). The putamen was found to be the most prominent brain region in the network found to be altered in relation to IA (Hong et al., 2013) and FC between DS, VS and middle frontal gyrus was altered between individuals with Internet gaming disorders and those with recreational game use (Dong et al., 2021). Given this evidence, the OFC and DS will be included in the reward network. The reward network is strongly associated with IA and therefore its within rsFC is supposed to be predictive of IA forming the first hypothesis of this work.

1.3.2 Self-Reference

In addition, the self-referential cognition network is assumed to be altered by IA. With respect to the SMU characteristics of IA, it is hypothesized that the constant sharing and monitoring of personal information highlights the importance of 'valuable' characteristics. This is a source of self-evaluation and leads to comparisons and an attention shift towards self-identity, which may be disturbed by unrealistic expectancies (Meshi et al., 2015). In line with this assumption, it was found that IA was associated with self-identity confusion which consists of three common types: (1) Disturbed identity, describing the tendency to acquire thoughts and attitudes of significant others hindering the development of an own identity. (2) Unconsolidated identity, which describes the failure, to live in self-defined roles and to develop stable beliefs, attitudes, and values. As a result, individuals perceive a (3) lack of identity which manifests in sudden and dramatic shifts in self-image and respectively own convictions (Hsieh et al., 2019). While the authors examined this relation in the direction that self-identity confusion would be the antecedent of IA, it is hypothesized in this work, that the relation is bidirectional, as for example bodily sensations appear to be affected, as SMU disrupts self-images regarding the body. The assumption that this is caused by social media

focusing idealistic appearance was confirmed and the effect sizes were larger for appearance-focused sites, compared to general social media (Saiphoo & Vahedi, 2019). Furthermore IA goes in hand with lower self-esteem (Aydm & San, 2011; Brand et al., 2016) and is negatively associated with self-kindness and positively with self-judgment (Iskender & Akin, 2011). Not only the properties of social media are detrimental to self-referential processes, as Gaming-addicted teens appear to integrate their game characters into their own self-identity, accompanied by activation of mPFC and ACC (Choi et al., 2018). Hence, the self-referential cognition network is assumed to be associated with IA.

Based on theoretical assumptions, the self-reference network associated with SMU includes the medial prefrontal cortex (mPFC) and the PCC (Meshi et al., 2015). These propositions relate to SMU but can also be applied to IA due to the overlap between the two constructs. The role of the self-referential network is empirically undermined as the PCC is activated during both social and self-referential processes and the mPFC shows stronger activation in self-referential processes contrasted over social processes (Herold et al., 2016). The PCC is further a part of the default mode network (DMN), which is known to process social and self-referential information. The PCC has found to be activated when individuals were shown personalized content (Wadsley & Ihssen, 2023), indicating a role in social comparisons and therefore self-referential processes. Findings support the assumption of involvement of mPFC and precuneus in SMU based on altered FC between mPFC and dorsolateral prefrontal cortex (dlPFC) and between precuneus and dlPFC in relation to sharing self-related information (Meshi et al., 2016). Individuals with high SMU showed lower differences in the evaluation on statements about skills and qualities in a self-concept vs. reflected-peer paradigm, being related to the Disturbed Identity construct. The observed evaluation similarity was related to less mPFC-activity for reflected-peer-judgments and higher for self-judgements (Peters et al., 2021). Further reviews and meta-analyses corroborated these considerations (J.-T. Sun et al., 2023; Wadsley & Ihssen, 2023) but the ROIs suggested by Meshi et al. (2015) will be expanded.

The insula is associated with self-awareness and interoception, the perception and integration of the body and its states (Namkung et al., 2017; Naqvi & Bechara, 2010; Turel et al., 2018). This becomes even more relevant considering the altered body image, that results from SMU. In addition to its contribution to self-referential processes, the insula also shows alterations in grey matter volume in relation to SMU mediated by delay discounting, which provides an explanation of addictive tendencies through poor foresight and impulsivity shown by the insula (Turel et al., 2018). It further seems to play a role in the maintenance of addictive behavior, as in one study its damage led to disruption of smoking addiction (Naqvi et al., 2007).

Given the theoretical and empirical findings the self-reference network is assumed to play a role in IA and therefore it is supposed that its rsFC is predictive of IA, forming hypothesis two of this study.

1.3.3 Attention

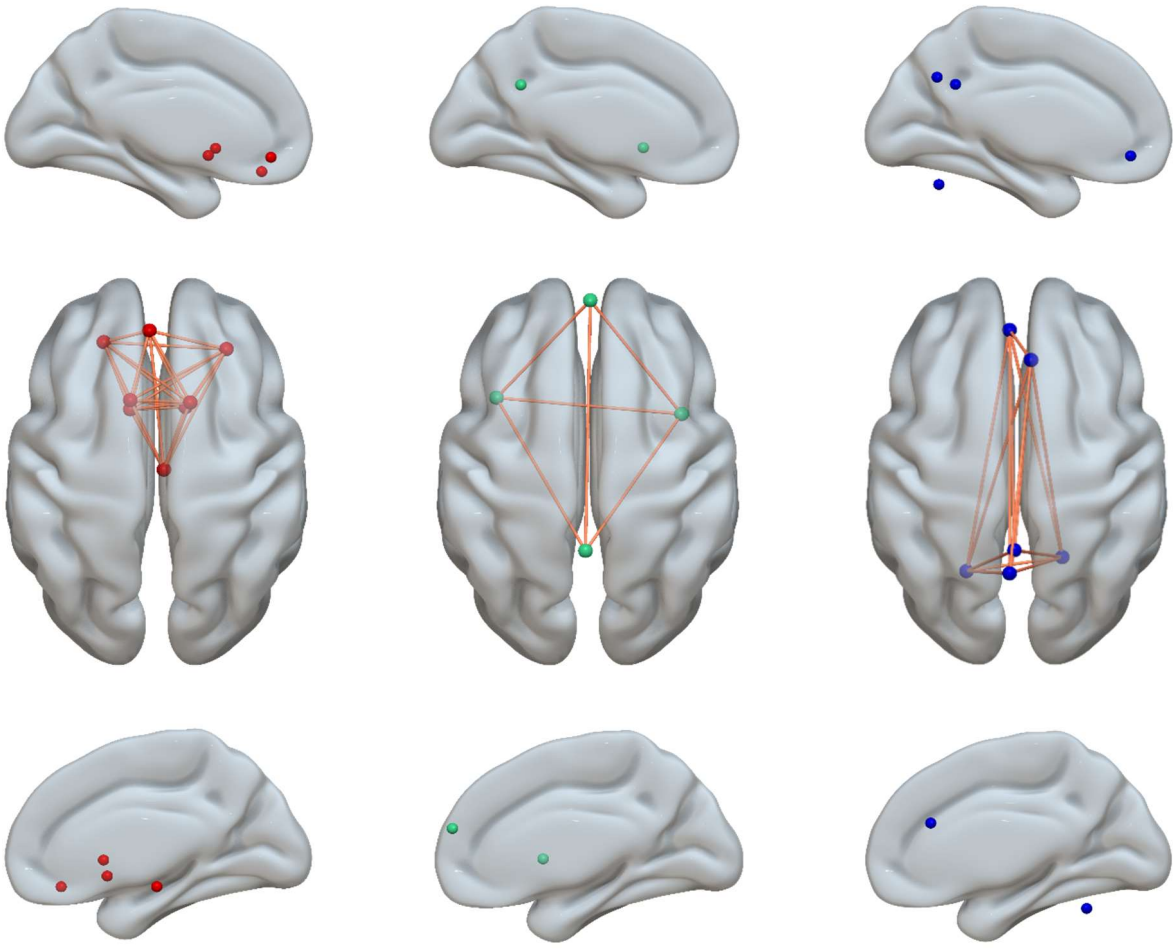
As previously mentioned, there is a relation between IA and ADHD, as both constructs capture attention deficits, hyperactivity (B. Wang et al., 2017; Yen et al., 2009) and impulsiveness (Konrad & Eickhoff, 2010). Those characteristics are further generally found in (n)SRAs (Brand et al., 2016; Volkow & Boyle, 2018). Particularly the association of IA with attentional changes in processing Internet related cues makes both conditions overlapping (Brand et al., 2016). Therefore, several studies found IA to be correlated with ADHD (Evren et al., 2018; Tateno et al., 2016; Weinstein et al., 2015; Yoo et al., 2004) and the association remained even when it was controlled for confounding factors such as predisposing social variables and overall health (Y.-L. Chen et al., 2015). Both diseases can be considered as comorbidities, while the preexisting neuropsychological characteristics of ADHD are assumed to make individuals more vulnerable to develop IA. As patients with ADHD have increased sensitivity to short-term, immediate rewards and difficulties to resist to them due to a lack of dopamine, the Internet with its rewarding effects could be used as a tool to compensate (Yen et al., 2007). This assumption has been supported, as ADHD has been found to be a predictor of problematic Internet use and that this association was even stronger as this for depression. Furthermore, compared between the different types of ADHD (predominantly inattentive, predominantly hyperactive, combined type) the scale for attention deficits was the strongest predictor (Sariyska et al., 2015). Most studies concerning the association between IA and ADHD focus on the attentional bias (1.1.3 *The I-PACE Model*) and a neurological similarity between both diseases has been found in smaller volume decrease in regions related attention and other executive functions, among others (Marin et al., 2021).

Neurologically the findings of similarities between both diseases can partially be explained by a mechanism in which striatal dopamine compensates for the usual attention failure while using the Internet or other preferential activities, so that ADHD patients find themselves concentrated and in control, while this feeling is lacking in real life (Y. Wang et al., 2019; Yen et al., 2007). Individuals with IA exhibit increased FC in the DMN which is typically deactivated during attention-demanding tasks and hypo-connectivity in the visual attention network (VAN), which plays a major role in visual information integration and attention processing. The regions affected in the DMN are associated with object-context associations and contextual predictions which becomes relevant given the conditioned reward response towards drug/Internet-cues. The regions affected in the VAN were also found to be affected in ADHD and are associated with impaired concentration (Y. Wang et al., 2019).

The neuronal basis of ADHD is mainly found in the DMN, which includes regions such as the vmPFC, precuneus and PCC. In addition, the dACC and cerebellum play a role in attention deficits as they exhibited lower FC for ADHD patients in working memory tasks (Konrad & Eickhoff, 2010). These regions have also been found to be altered structurally and functionally in Internet addicted individuals. Smaller grey matter volume of ACC and decreased FC between ACC and precuneus were found in individuals with higher multimedia multitasking. The ACC is known for higher cognitive and emotional/motivational processes and activated in tasks, that require decision making under conflict and cognitive control. Therefore, lower grey matter of this structure indicates

Figure 4.

Networks forming Hypotheses.



Note. Networks hypothesized to play a role in IA with its nodes and edges. Edges are used as features in the modeling algorithm. Red: Reward-Network (vmPFC, VTA, OFC, VS & DS); Green: Self-Reference-Network (mPFC, PCC & insula); Blue: Attention-Network (vmPFC, precuneus, PCC, dACC & cerebellum).

lower cognitive and inhibitory control in these individuals. Furthermore, higher MMI is related to higher failure in attention (Loh & Kanai, 2014). Besides, a meta-analysis found stronger resting state FC from PCC to the whole brain in subjects with IA and findings are attributed to the role of the PCC in visuo-spatial and sensorimotor areas. The abnormal functionality is linked to decreased attention and poor executive ability found in Internet addicted individuals (J.-T. Sun et al., 2023). Altered FC due to IA has been found in the PCC to precuneus, NAcc and cerebellum. While the PCC is known to be relevant for short-term memory and attention regulation, the cerebellum is associated with cognitive regulation and signal processing and the precuneus plays a role in visual processing (Ding et al., 2013). Those properties make both structures relevant for the addiction characteristics attentional bias, cue reactivity and impulsivity. Their malfunctioning assigns Internet relevant cues more importance, increases the salience and processing speed and therefore impedes inhibitory control. Furthermore, the cerebellum has been linked to both ADHD and addiction as a mediator between reward, motivation and cognitive control systems and is involved in reinforcement learning of motivational behavior, modulating emotions as in the withdrawal phase and cue-reactivity in the form of craving (Moulton et al., 2014). This structure was found to underlie changes between Internet addicted individuals and healthy controls in fMRI FC and regional homogeneity (ReHo) (Liu et al., 2010). Hence, the attention-network associated with IA and adapted for this work consists of vmPFC, precuneus, PCC, dACC and cerebellum.

Given the evidence of associated ADHD and IA as well as overlapping findings of the regions being involved, the rsFC within the attention network is supposed to be predictive of IA and thus forms the third hypothesis in this work. All hypothesized networks and their respective ROIs are shown in Fig. 4.

The current literature in this field trying to examine the neuronal mechanisms of IA is broad but therefore also inconclusive. Most studies focus on single regions or specific interactions, but do not incorporate a comprehensive prospective on the subject so that the current study tries to counteract by applying the functional connectomes framework.

1.4 Functional Connectomes

Before introducing functional connectomes, it is essential to explain the subject of its parent field, systems neuroscience. Neuroimaging studies, which investigate neuronal correlates of behavior mostly adopt two distinct approaches based on neurological properties. Many studies focus primarily on correlating the activation of specific ROIs with behavior, an alternative perspective is obtained by the dynamic interplay among these regions, which form networks associated with behavioral and cognitive states. Systems neurosciences goal is to understand these complex interactions among different brain regions and their role in shaping behavior and cognition. The emergence of systems

neuroscience stemmed from the observation that resting-state functional magnetic resonance imaging (rs-fMRI) time series exhibited systematic correlations within networks, but distinct patterns between networks. This foundational observation has caused extensive investigation into the organizational principles governing neuronal networks and their functional significance (Smith et al., 2013). The advantage of this approach lies in the investigation of more complex interplays between regions and possible alterations based on differing individuals' characteristics.

One part of systems neuroscience is functional connectomes. Connectomes refers to the human connectome, which is a map of the human brain's circuitry, consisting of different areas and its structural and functional connections. Connections which are represented anatomically are called structural while co-(de-)activation (e.g. measured by blood oxygenation) is called functional. Therefore, functional connectomes focus on the communication and coordination of different regions and networks. Findings are often visualized in graphs, where nodes represent the ROIs and edges represent the connectivity (e.g. Figure 4), which are given weights for the strength of connection. Since the application of rs-fMRI and FC in functional connectomes is commonly used to investigate clinical disorders, this approach—aiming to identify biomarkers of disease state, severity, and prognosis—suits the objectives of this work and serves as the applied framework.

The most commonly used neuroimaging method in functional connectomes is (f)MRI due to its near-ubiquity, non-invasiveness and high spatial resolution (Craddock et al., 2015). In fMRI studies there are two main approaches to map behavior and cognition to brain functioning, namely task-activation and resting state. In task-activation studies participants are confronted with tasks or different stimuli (e.g. visual, auditory, tactile, etc.) while brain activation is captured. Following the data acquisition, the brain activation in the rest and task condition are compared. Activation differences are then explained by the neuronal reaction to the task or stimuli. In resting-state fMRI participants are told to stay still in the scanner and should not think about something in specific. It is hypothesized, that peoples' characteristics or frequent behaviors (e.g. addictions) are represented in the brain functioning at rest, such that only one condition is required. The obtained connectivity across different networks is then compared between subjects. It has been shown that previously identified networks and connections can be observed even if participants are in non-conscious states (e.g. sleep, anesthesia) (Smith et al., 2013). This rs-approach has its advantage in less complex data acquisition and analysis and interpretation is more convenient. Therefore observed functioning is not biased by erroneous experimental design and it is the most common fMRI-method in functional connectomes (Craddock et al., 2015).

In fMRI, the previously mentioned time series of specific nodes in the brain (single voxels or whole regions representing the average signal of all included voxels) is obtained by measuring the oxygenation of blood flowing through these nodes over time. A voxel is a three-dimensional pixel representing a cube (mostly of the size $3mm^3$) within the space of the brain. This means that the brain

in fMRI data consists of up to millions of single voxels which are localized through a three dimensional coordinate system (x,y,z) and each voxel has its own time series. The measure is called BOLD-time series referring to *Blood Oxygenation Level Dependent*, which reflects the changes in the concentration of oxygenated and deoxygenated hemoglobin in the blood vessels of the brain. It is hypothesized that the level of oxygenated blood gives insight about the activation or deactivation of the specific nodes as activated cells need more oxygen to work. The correlation of the oxygenation's signals (time series) refers to simultaneous in-/activity indicating correspondence between those nodes, forming the edges and weights.

It is important to note that simple correlation, in FC mostly computed as Pearson's correlation coefficient is assumed to be an indication of connection but there is no evidence of causality and /or if the connection is direct versus indirect. Those questions are addressed by effective connectivity (EC), which incorporates confirmatory approaches such as structural equation modeling (SEM) and dynamic causal modeling (DCM) (Smith et al., 2013). However, in this study FC will be analyzed as EC is computationally expensive and performs better on task-fMRI (Smith et al., 2013) while the current data is rs-fMRI which in turn is mostly used in functional connectomes (Craddock et al., 2015).

One key consideration which has to be made when functional connectivity is analyzed for the comparison of different individuals is the parcellation method (Craddock et al., 2015). Parcellation refers to the division of the brain into different networks and their respective nodes, which are assumed to play a role in the experience or behavior in question. The previous section (1.3 *Neurological Basis of IA*) outlined relevant functional networks in the context of IA which form the hypotheses to be tested.

1.5 Outlook

Summarized, prevalence of IA is rising constantly, which provides a risk for increasing numbers of mental health disorders, and in turn increases the vulnerability for those individuals even more, to get addicted. Simultaneously, the (neurological) mechanisms behind this addictive behavior remain unclear as to the best knowledge of the author, no study has been conducted incorporating a functional connectomes framework combined with a predictive modeling approach. This work tries to shed light on the neurological underpinnings of IA to validate and combine previous findings on possible biomarkers and further provide a model which would be able to make diagnosis based on fMRI data. This goal is to be achieved by predictive modeling of IA through FC of three networks, namely reward- self-reference and attention-network with gradient boosted decision trees. The hypotheses will be tested by assessing the feature importance of each network for the success of prediction. Therefore, two models will be fit by FC matrices of each participant and the respective

IAT score. One with correct labels the other one with shuffled labels (null model), such that they can be tested against each other. It is expected that the true model will perform better and that feature importance for the networks outperform the ones in the shuffled model.

2 Methods

Within this section the current dataset (secondary data) will be described on the data acquisition and preprocessing procedures and the inclusion criteria the authors chose (Mendes et al., 2019). The variables of interest will be presented based on its role in the modeling procedure, split between features and outcome variable. The outlier strategy, statistical analyses as well as their assumptions and prerequisites will be explained, and the final modeling procedure outlined. Finally, the used software and its additional packages are presented so that the results are reproducible.

2.1 Dataset

The Max Planck Institute Leipzig acquired the data, which was peer reviewed, published, and is publicly available as the MPI-Leipzig-Mind-Brain-Body dataset (Mendes et al., 2019). The data includes rs-fMRI and several behavioral measurements. MRI data (structural scan, four 15 minutes rs-fMRI scans, two gradient echo field maps and two pairs of spin echo images with reversed phase encoding direction) was obtained by a whole-body 3 Tesla magnetic resonance imaging scanner (Siemens Magnetom Verio, Siemens Healthcare, Erlangen, Germany) at the Day Clinic for Cognitive Neurology, University of Leipzig.

Healthy participants were screened for past and present psychiatric disorders using the Structured Clinical Interview for DSM-IV (SCID-I; First et al., 1997). Data acquisition took place five days in a row over two years and on-site appointments were held at Max Planck Institute for Human Cognitive and Brain Sciences (MPI-CBS). For the questionnaires and tasks, the order of presentation was randomized across participants to remove positional effects. The authors aimed for a healthy sample, with exclusion criteria comprising a history of psychiatric diseases accompanied by hospitality within the last 10 years, history of neurological disorders, history of malignant diseases, intake of certain medication, a positive drug anamnesis, extensive testing experience in academic institutions and past or present study of psychology. Furthermore, MRI testing poses some prerequisites, so that participants with metallic implants, tattoos, pregnancy, claustrophobia, tinnitus, and a surgical operation in the last 3 months were excluded.

Structural and functional MRI-data has already been preprocessed and denoised. The structural scan T1 was transformed to the MNI152 1mm standard space. To ensure non-identifiability, a defacing mask was applied to all scans. Preprocessing of the functional data included the exclusion of initial volumes for the stabilization of the signal, motion correction, masking of the functional images onto the structural scans, removing motion parameters and outliers, temporal filtering for the removal of physiological artifacts and the focus on low-frequency fluctuations of the BOLD-signal, normalization to ensure comparability across subjects and projection onto MNI152 space. Details

and processing python code are reported in Mendes et al. (2019). The data was first accessed in April 2023 and analyzed in October 2023 after publishing preregistration in the OSF registry (https://osf.io/5v7my).

Data was accessible via different domains noted within the article. MRI data was retrieved from the following link without any restrictions: https://ftp.gwdg.de/pub/misc/MPI-Leipzig_Mind-Brain-Body/. Behavioral data was downloaded from the following link after the declaration that (1) I will not attempt to establish the identity or attempt to contact any of the included human subjects. (2) I understand that under no circumstances will the code that would link these data to Protected Health Information be given to me, nor will any additional information about individual human subjects be released to me under these Open Access Behavioral Data Use Terms. (3) Failure to abide by these guidelines will result in termination of my privileges to access MPI-Leipzig Mind-Brain-Body behavioral data: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/-DVN/VMJ6NV.

The whole dataset consisted of 194 fMRI and 214 IAT datapoints (Mendes et al., 2019). Unzipping of resting state .gz-files led to problems in some participants, such that the last session was not available. As most resting state analysis is performed on 8-15 minutes (Smith et al., 2013), it was decided that the first three sessions with a total of 45min of resting state data would be sufficient for subsequent analysis. As a result, eight participants could not be included in further analyses, leaving 186 participants with valid data. The authors stated that the data of participants who failed to complete questionnaires was excluded from analyses, therefore this sample was considered to consist of only complete cases, and no strategy for the handling of missing data has been developed.

2.2 Variables

Following the framework of predictive modeling, variables of interest must be split into an outcome vector (y), consisting of IAT-values, the set of features, consisting of FC matrices (X_{FC}), and control variables ($X_{Contr.}$), consisting of age and sex, which are outlined in this section.

2.2.1 Outcome

For the current study, IA is quantified by the IAT (Young, 1998), a 20-item self-report questionnaire. Statements can be answered on a six point Likert scale, indicating the frequency of certain behaviors linked to excessive, uncontrollable Internet use (e.g. “How often do you find that you stay online longer than you intended?; 0 = ‘does not apply’ – 5 = ‘always’). All items are summed

up obtaining a global score with a possible range between 0 and 100. Based on the summed scores four categories in IA severity (normal, mild, moderate, severe) can be specified. People scoring 50 or higher are meant to be Internet addicted. In the current dataset a German translated version was used, and item 3 (i.e., "How often do you prefer the excitement of the Internet to intimacy with your partner?") has been removed due to different scaling, such that the questionnaire consisted of 19 items with a possible continuum of [0:95]. Internal consistency was good with $\alpha = 0.91$ (Mendes et al., 2019).

2.2.2 Features

To predict IA rs-fMRI FC matrices for each subject were used. Networks as the basis of FC-matrices were decided to be created by selecting ROIs, assuming a better fit to the data than previously created atlases (Khosla et al., 2019). This procedure enabled flexible creation of individual networks. The FC matrices then were obtained by correlation between time-series for each preselected ROI, namely (1) reward network (intercorrelation values for single nodes for vmPFC and VTA and for bilateral nodes for OFC, VS and DS); (2) self-referential network (intercorrelation values for single nodes for mPFC and PCC and for bilateral nodes for insula); (3) attention network (intercorrelation values for single nodes for vmPFC, precuneus, PCC and dACC and for bilateral nodes for cerebellum). This led to $N_{within} = 49$ continuous predictor variables ([-1:1]) divided by each network (reward: 28; self-reference: 6; attention: 15). It was decided to keep feature-sample ratio as low as possible, so that only within-network connectivity was used as features. Therefore, variables representing between-network connectivity were disregarded ($N_{between} = 72$). To calculate FC between those regions, firstly the MNI152 spaces for the respective ROIs were retrieved from neurosynth.org (Yarkoni et al., 2011) and extended by a 6mm sphere like follows: (1) brain activation masks were downloaded for each ROI using a respective search term (Table S1), (2) masks were loaded in MRIcroGL (v. 1.2.20220720; https://www.mccauslandcenter.sc.edu/mricrogl/). thresholded to a certain activation level and activation clusters <32mm^3 were removed to get peak coordinates which were used to (3) create own ROIs in the setup-section of the CONN-toolbox (v. 22.a; Nieto-Castanon & Whitfield-Gabrieli, 2022), details can be found in Table 1.

Computation of the FC-matrices was further conducted in the CONN-toolbox using three times 15 minutes of already preprocessed (Mendes et al., 2019) rs-fMRI with a repetition time of 1.4 seconds and continuous acquisition time. The preprocessing pipeline conducted by the authors of the dataset contained denoising, hence this step has been skipped in the toolbox by running the analysis without any parameter selected. First-level analyses were set to ROI-to-ROI connectivity (RRC), computed by Fisher-transformed bivariate correlations between a pair of ROIs BOLD timeseries via a weighted GLM (for computational details see Nieto-Castanon (2020)). This resulted in a 16x16

connectivity matrix per subject (for schematic display see Figure 8). The current computation method is suggested for the simultaneous study of entire networks of connections (Nieto-Castanon, 2020). As the following cross-validation strategy (*2.3.2.6 Procedure*) requires independency between the train and test set in cross-validation (CV) schemes (Scheinost et al., 2019; Shen et al., 2017), such that no between-subject computation could lead to data leakage, those matrices were calculated on single-subject basis.

Table 1.

MNI-Coordinates of ROIs

ROI	*z*-Threshold	Removal Size	Peak Coordinates			Final Coordinates		
			x	y	z	x	y	z
vmPFC	11	32	-2.8	43.4	-9.1	-3	43	-9
VS (L)	15	32	-10.9	8.7	-7.6	-11	9	-8
VS (R)	15	32	12	10.2	-7.6	12	10	-8
VTA	15	32	3.1	-17.1	-13.5	3	-17	-14
OFC (L)	10	32	-21.9	38.2	-16.5	-22	38	-17
OFC (R)	10	32	29	34.5	-13.5	29	35	-14
DS (L)	7	32	-10.9	13.1	-3.9	-11	13	-4
DS (R)	7	32	14.2	12.4	1.2	14	12	1
mPFC	12	32	0.9	55.9	20.4	1	56	20
PCC	8	32	-1.3	-51.8	30.7	-1	-52	31
Insula (L)	12	32	-38.2	13.9	-3.9	-38	14	-4
Insula (R)	12	32	39.3	7.2	2.7	39	7	3
Precuneus	12	32	-2.8	-62.1	35.2	-3	-62	35
Cerebellum (L)	15	32	-21.2	-60.6	-24.6	-21	-61	-25
Cerebellum (R)	15	32	19.4	-54.7	-24.6	19	-55	-25
dACC	7	32	6.1	30.1	23.4	6	30	23

Note. Activation maps derived from neurosynth.org were reduced at given z-value thresholds and activation clusters below the size of 32mm^3 were removed. Final coordinates were derived from rounded peak coordinates.

2.2.3 Control Variables

Besides networks FC, age and sex are also used as predictors, as several studies identified both as important markers for FC and effects of technology use. Age is associated with decreased FC (N.-K. Chen et al., 2009; Schulz et al., 2022) and can be assumed to be negatively correlated with IA severity, considering higher prevalences for younger individuals (Lozano-Blasco et al., 2022). Furthermore, sex mediates and moderates technology use and IA, such that female users tend to show higher SMU engagement, while online gaming would be more prevalent among male individuals (Nesi & Prinstein, 2015; Ostovar et al., 2016). In addition interactions on the effect of SMU on various self-concepts are expected as e.g. physical positivity decreases for females based on SMU while males are not affected (Peters et al., 2021). It is assumed that the role of age and sex will

become relevant in the interaction with FC predictors as differences in both variables should be associated with different types of Internet usage. Their predictive performance on IA does not form a hypothesis but the influence will be tested exploratory. Therefore, besides FC-matrices, age in 5-year bins as a continuous variable ([20:80]; [1:12]) and binary sex as categorial variable (male: 0, female: 1) are included as features in the modeling process.

2.3 Machine Learning

Before outlining the analysis strategy, the concept of predictive modeling via machine learning, its advantages, techniques and decisions on the actual strategy have to be presented.

Machine Learning (ML) methods in neuroscience have gained importance and application in recent years, as relationships between neurological activation and behavior are often complex and non-linear and therefore require advanced modeling approaches (Glaser et al., 2019; Simpson et al., 2013). The complexity of neuronal data mainly results from the high number of possible predictors (ML jargon: features) that interact and create high-dimensional datasets. The most common template for fMRI data, the MNI152 space consists of hundreds of thousands of voxels. Even when dimensionality is reduced by forming ROIs (averaging the time series for respective voxels), the number of features exceed the threshold, where simple models are performing well (Coutanche & Hallion, 2019).

In addition to the improved ability to model complex relationships, ML techniques are mostly conceptualized to predict a particular variable y from a feature set X, while the model is required to generalize on unseen data. This is achieved by conducting two phases of the modeling process. (1) In the training phase, a portion of the dataset is used for training, in supervised ML X and y (labels) are provided together, so that the algorithm can learn the relationship between X and y. Depending on the specific modeling objective (classification vs. regression) the model is optimized to maximize accuracy or minimize prediction error, respectively. (2) In the testing phase, the obtained model is tested on unseen data to verify the performance of prediction. This makes ML methods superior to traditional statistical approaches as they do not overfit on the entire dataset and estimation of predictive power is more realistic (Coutanche & Hallion, 2019). In line with that, it is argued, that most models developed in psychology have not been tested for their true predictive performance, resulting in a lack of validation about their generalizability, contributing to the replication crisis (Yarkoni et al., 2011). The current work agrees on those claims and therefore follows the framework of predictive modeling.

The application of ML in neuroscience is manifold. Glaser et al. (2019) worked out four roles of ML in neuroscience out of which two are relevant for this work. (1) Solving engineering problems refers to quantify a certain variable y out of a set of variables X. This is relevant for diagnosis of

diseases so that practitioners could scan patients to determine their state of disease. This is one of the strived for goals of this work. (2) Identifying predictive variables is closely related as relevant predictors also contribute to diagnosis accuracy but here the focus is set on the feature set *X*. Revealing the interplay of features responsible for increased model performance gives insight about underlying mechanisms and therefore contributes to neuroscientific theories. Hence, modeling *y* out of *X* supported by methods to detect feature importance is another goal of this work.

In fMRI, ML is mostly used to model second-level analyses like e.g. predicting behavioral outcomes based on neurological features. This framework is widely used because of its suitable characteristics to discover biomarkers and the principled means for multivariate models without the urge to correct for multiple comparisons in mass univariate techniques which especially in high dimensional datasets can cause problems (Craddock et al., 2015). Many studies using this approach showed high predictive performance with (classification) accuracies mostly >80% for a wide range of mental disorders like schizophrenia, major depression, bipolar disorder, autism spectrum disorder, ADHD, obsessive–compulsive disorder, social anxiety disorder, posttraumatic stress disorder, and substance dependence (Du et al., 2018; Rashid & Calhoun, 2020). However, the numbers of studies focusing dependencies is low and they do not incorporate nSRAs. Therefore, this gap must be filled and the increasing interest in those methods in revealing biomarkers and the evidence of successful application undermine the promising character.

2.3.1 Classifier and Regression Models

Once the decision has been made to incorporate machine learning methods to answer the research question, it must be chosen from a variety of different models, which fit best to the problem. Models differ mainly in two categories, the form of prediction they make (classification vs. regression) and the degree of linearity in the model structure. In classification the samples in question are predicted in different classes (categories), which is mostly done with sets of healthy vs. ill individuals. In regression a continuous value is predicted, which allows a finer separation between the samples and will be applied within this work.

Linearity refers to a straight proportional nature of relations, which means, that the steady increase/decrease of one variable x_i (out of *X*) predicts the increase/decrease of another variable *y* without fluctuations and can be depicted as a straight line. As there might be linear relations in general, those are rare in psychology and especially brain functioning is assumed to be non-linear. Non-linearity, in contrast means that the relation can fluctuate at given points, so that the strength and direction of the correlation may change. An example is the relation of age (*x*) and fluid intelligence (*y*). From $x = 0$ on, intelligence increases until it reaches its climax at $x = 25$, followed by a decline in *y* for $x > 25$. In this section the most frequently used ML models in rs-fMRI analysis

will be presented (Khosla et al., 2019). Based on their advantages and disadvantages an argumentation for the decision in favor of one of them follows.

Regularized linear models: Those models provide linear predictions between the features *X* and the target variable *y*. Their optimization process, the attempt to increase accuracy or to decrease prediction error, relies on either a conditional likelihood estimation which tries to maximize the conditional probability of the target variable given the feature set or a maximum a posteriori estimation (MAP), where the goal is to find the parameters that maximize the posterior probability of *y* given *X*. Regularization is a technique which prevents the model from overfitting, trying to keep models less complex and penalizing too heavy feature weights. L1 (Lasso) regularization penalizes a model for incorporating too many features, L2 (Ridge) regularization penalizes a model for too strong features and Elastic Net regularization combines both approaches. Linear models can be used for both classification and regression problems.

Support Vector Machine (SVM): SVMs try to find the best possible separation between different classes or cases. This is done by constructing a hyperplane in an *n*-dimensional space where *n* = the number of features in *X*. This hyperplane, based on the features, must be constructed under the prerequisite that it divides the space in the most selective form, such that the different samples of *y* can be differentiated with a maximal distance to the hyperplane. The construction of the hyperplane is based on the samples of *y* which are closest to the separating plane, which are called the support vectors. Mostly, SVMs are linear but the hyperplane can be constructed in a curvy way so that even non-linear data can be modeled. They can be used for both classification and regression tasks.

Decision Trees and Random Forests: The prediction is obtained by a sequence of splits in the input feature space *X*. A decision tree consists of nodes, which represent decision points for certain features and lead to following edges representing the outcome. Dependent on the number of nodes, several variables are used to differentiate the sample. The end of a tree is formed by 'leaves', representing the final decision for the current sample *y*. Single decision trees mostly do not perform well compared to other models but given the accuracy-interpretability trade-off they are well suited for investigating the relevance of specific features. To further improve accuracy or prediction error, ensemble learning techniques such as random forests or boosted trees can be used (2.3.2.1 *Decision Trees* & 2.3.2.2 *Gradient Boosted Trees*). They can be used for both classification and regression tasks.

Deep Neural Networks: Those methods perform particularly well on highly complex datasets. Commonly used for supervised tasks, they map *X* on *y* via multiple interconnected nodes and layers which are called neurons. The connections are associated with weights determining the strength of connection and each node consists of a mostly non-linear activation function. Based on the 'reaction' of this function on its input and associated bias term further information is passed to the next layer of neurons. Weights and biases are tuned using a gradient descent to minimize a loss

function to provide the best results. Gradient descent is an iterative process adapting the parameters in question based on the loss function representing the prediction error. Based on its accuracy this method is the most promising but poses issues in other domains which are discussed below.

2.3.2 Decisions on Current Model

Given this variety of potent models, their strengths and weaknesses need to be examined regarding the research question at hand to decide which model to use in this study. First the requirements for the current research question must be outlined.

(1) Non-Linearity: As mentioned above, linear models are assumed to be insufficient for the current problem. Evidence for the superiority of non-linear models is sparse, nevertheless there are indications that these models better fit the requirements of neuroimaging based modeling (Coutanche & Hallion, 2019). (2) Interpretability of the Model: Since a main goal of this work is the identification of biomarkers, the features which are most responsible for prediction success must be extractable and interpretable. Therefore, the model is required to both maximize complexity and interpretability. (3) Required Sample-Size: One of the key issues of predictive modelling of fMRI parameters is the time and financial effort to create the datasets. Therefore, sample sizes are often small, which in turn limits the maximum complexity of the model as complex models require larger datasets to learn from. Even though the current dataset is comparably large, it is not sufficient for certain approaches.

Based on these considerations, regularized linear models and SVMs, considered as linear models, will not be applied, although they are the most widely used models. In addition, SVMs can lack interpretability. An argument in favor of deep neural networks would be the non-linearity of the activation functions and their high prediction accuracy, however this accuracy is based on the complexity of the model and thus, a large amount of data is required. While there are current attempts to combine multiple neurological datasets to obtain a sufficient size for neural networks, this study relies on a single dataset and therefore a less complex model will be the better fit. In addition, model complexity of neural networks poses another limitation, as it makes the model almost uninterpretable. As decision trees combine interpretability, non-linearity and sufficient accuracy when used as an ensemble, this will be the method to use within the current study (Kingsford & Salzberg, 2008).

After choosing the specific model the next decision is about classification vs. regression. This is due to two major reasons: (1) The diagnosis of mental disorders in classical categories is debated because of the remaining uncertainty of cut-off thresholds (Craddock et al., 2015). (2) Most predictive modeling in the field is conducted via classification techniques but the call for continuous predictions is growing. Many classification studies lack real samples as the common approach to classify healthy and ill individuals mostly relies on data on both extremes, which in turn decreases specificity (Craddock et al., 2015). Furthermore, important insights about underlying mechanisms

can only be achieved by continuous measurements enabling diagnosis on finer levels including sub-clinical manifestation, observation of development and course of the disease (Rashid & Calhoun, 2020). Therefore, and because the current dataset consists of mostly healthy participants, this work strives for a continuous prediction.

2.4 Analysis

The analysis consisted of three major steps: (1) Descriptive analysis to get an oversight about distributions of the relevant variables and their interrelations, (2a) Modeling of the IAT through FC and control features in an original and shuffled version to obtain a null model. Those models were compared to assess model performance and (2b) Hypothesis testing to evaluate the predictive importance of the outlined networks, (3) Exploratory analysis, to deepen insights about causes and implications of results.

The measure for central tendency in this study is the mean-value (*M*) in combination with the standard deviation (*SD*) which will be presented with ± after the mean. In most cases the mean is a good estimator for the expected values although in some distributions this assumption does not hold true as it is highly influenced by outliers. In those cases, the median represents a better indicator, therefore in critical cases it is reported as well and its computation represents the value, which distributes the sample in two halves of equivalent size. An indicator, if the median (*Med.*) should be preferred over the mean is the presence of outliers.

Therefore, it is important to spot the outliers and handle them in a way so that no datapoints are lost but the skewness of the distribution is minimized, resulting in a more robust mean value. In this work, the outliers are identified by Tukey's boxplots, which use the *1.5xIQR rule* (Rousseeuw & Hubert, 2011). Every data point which is lower/higher than Q1–1.5xIQR/Q3 + 1.5xIQR is labeled as an outlier, representing a more robust metric than the commonly used *Z* scores method (Bakker & Wicherts, 2014). IQR refers to the Interquartile-Range, the range of values between the first and third quartiles (25^{th} and 75^{th} percentiles). The factor 1.5 leads to a threshold even narrower as three *SD*s of the normal distribution ($\mathcal{N}$; in $\mathcal{N}$ 99.7% of all datapoints fall into the range of the *1.5xIQR rule*). The strategy to deal with those outliers was chosen to be *Winsorizing*. This method provides a robust mean but can also be detrimental to conservative testing. However, it is argued, that the removal of outliers would be even worse for the inflation of Type I error (Bakker & Wicherts, 2014) such that the current strategy is to find the compromise. The data points identified to be outliers are assigned to the threshold values obtained by the *1.5xIQR rule*. Outlier handling like *Winsorizing* have their limitations in adding a bias to the original data, however its advantage lies in not losing samples while keeping the variance as original as possible. Therefore, the beneficial effects for a valid mean value and the bias and risk for Error I inflation are balanced. The significance tests which this

technique is used for assessing the performance differences between both models. As the variation in those performance and importance scores can be assumed to follow a random distribution, the creation of outliers is assumed to be random and therefore no important information is lost in transforming. The data to be analyzed is plotted before transformation to ensure that the changes to the data can be traced. Even though *Winsorizing* can rearrange the distribution it is not guaranteed, that it results in $\mathcal{N}$. However, the current testing strategy is robust in this case and as the sample size is large ($N = 100$, for details see *2.3.2.4 Procedure*), it is argued by the Central Limit Theorem (CLT), that the probability distribution of the mean will closely approximate $\mathcal{N}$ (Kwak & Kim, 2017). Furthermore, the differences between the *p*-values for the original and transformed mean will be reported to prevent the intransparency of questionable research practices (Bakker & Wicherts, 2014).

2.4.1 Descriptive Analysis

The descriptive analysis was conducted for all used variables, to capture the distributions, test for the representativeness of the sample and estimate possible interconnections between those variables. To gain insights about the average strengths of connectivity, the distribution of the correlational matrices was computed and further stratified by network affiliation.

2.4.1.1 Mann-Whitney-U-Test

Besides describing the overall distribution of sex and age, it had to be tested, whether sex was distributed equally among the different age bins, to estimate if the following modeling and prediction would be representative for both sexes. For this purpose, a *Mann-Whitney-U-Test* (MW-U) was conducted, which is a non-parametric test (needed because of non-interval scaling of sex) to compare distributions of two independent samples. The prerequisites are independence of measurements, nominal scaling of the independent measurement and at least ordinal scaling of the dependent measurement, which all are met. For the requirement of equal forms of distributions in both groups, a *Kolmogorov-Smirnov-Test* (KS) had to be conducted. If the *p*-value of the MW-U-Test does reach the threshold of significance ($p < .05$), the two samples are distributed differentially.

2.4.1.2 Linear Model

The next step was assessing relations between the control variables age and sex on the outcome variable IAT. For this purpose, a linear model was fitted. Even though it is not suitable to model the IAT through FC via a linear model (see *2.3.1 Classifier and Regression Models*; *2.3.2*

Decisions on Current Model), the relationship between age and sex and IAT is theoretically assumed to be linear. This linear model gives insight about the nature and strength of relations, which are represented in an equation of the form $y = ß_0 + ß_1 x_1 + ß_2 x_2 + e$. y is the dependent variable (IAT), $ß_0$ is the intercept, representing the starting point (default mean value without influence of the two predictors), $ß_1$ is the coefficient representing the strength of the relation between x_1 (age) and y and $ß_2$ the coefficient representing the strength of relation between x_2 (sex), and y and e represents the random error term. The model estimates the coefficients which best fit to the goal of minimizing the sum of squared differences between the observed values of the dependent variable and predicted values. The requirements for a linear model are a linear relationship between the variables, no multicollinearity, homoscedasticity, and normal distribution of the residuals. As previously stated, a linear relationship between predictors and the outcome is assumed but will be tested via linearity plot. Multicollinearity can be ruled out if the MW-U test reveals a non-significant result, normality of the residuals will be evaluated with the *Shapiro-Wilk-Test* (SW) and QQ-Plots. Dependent on the result, homoscedasticity will be evaluated by either the *Breusch-Pagan-Test* (BG; under normality) or the *White-Test* (under non-normality) (Long & Ervin, 2012).

2.4.2 Modeling

For model building, multivariable regression analysis was performed by fitting Gradient Boosted Decision Tree (GBTD) regression models using the LightGBM framework (Ke et al., 2017). This algorithm has proven to achieve high accuracy compared to other state-of-the-art algorithms while saving computational power (Ke et al., 2017). Two models (original vs. shuffled) were fitted with the features age, sex, and single predictors of the aforementioned networks (FC values for each predictor group) and scores of the IAT served as the outcome.

2.4.2.1 Decision Trees

Decision trees are a widely used, supervised modeling algorithm, because they have a simple structure and can handle high-dimensional data very well. The argumentation for the application of decision trees can be found in (*2.3.2 Decisions on Current Model*). A decision tree consists of several nodes, which are represented by the input variables. Those nodes can consist of both continuous as well as categorial data. Depending on certain values that these variables have, the node is split in a way to maximize separability. This results in subordinate nodes that represent a different variable and according to the depth of tree are split further themselves. The characteristic that a decision tree consists of multiple nodes and therefore multiple decision boundaries, poses an advantage of this

method. At the end of the tree, the so-called leaves form the answer to the classification or regression problem. By following all the nodes and their decision conditions, instances (data points/ subjects) can be classified or predicted in the expression of the outcome variable. This procedure also makes decision trees interpretable, as the value of a feature for the prediction performance can be determined easily. To avoid overfitting, it is important to limit the complexity of the tree by limiting the number of its nodes or assess generalizability via Cross-Validation (CV; *2.4.2.4 Procedure*) to ensure, that the model complexity reaches the threshold for the best bias-variance trade-off (Kingsford & Salzberg, 2008). Prediction performance can be increased by using ensemble learning which refers to the combination of multiple weak learners (single decision trees) to stronger ensembles, which can cut errors made by single instances out (Polikar, 2012). Decision trees are often combined to random forests or boosted (*2.4.2.2 Gradient Boosted Models*). In random forests each tree will be fitted with a randomly chosen certain feature subset so that each tree consists of different nodes and the results of all trees are combined to build an ensemble (Kingsford & Salzberg, 2008).

2.4.2.2 Gradient Boosted Trees

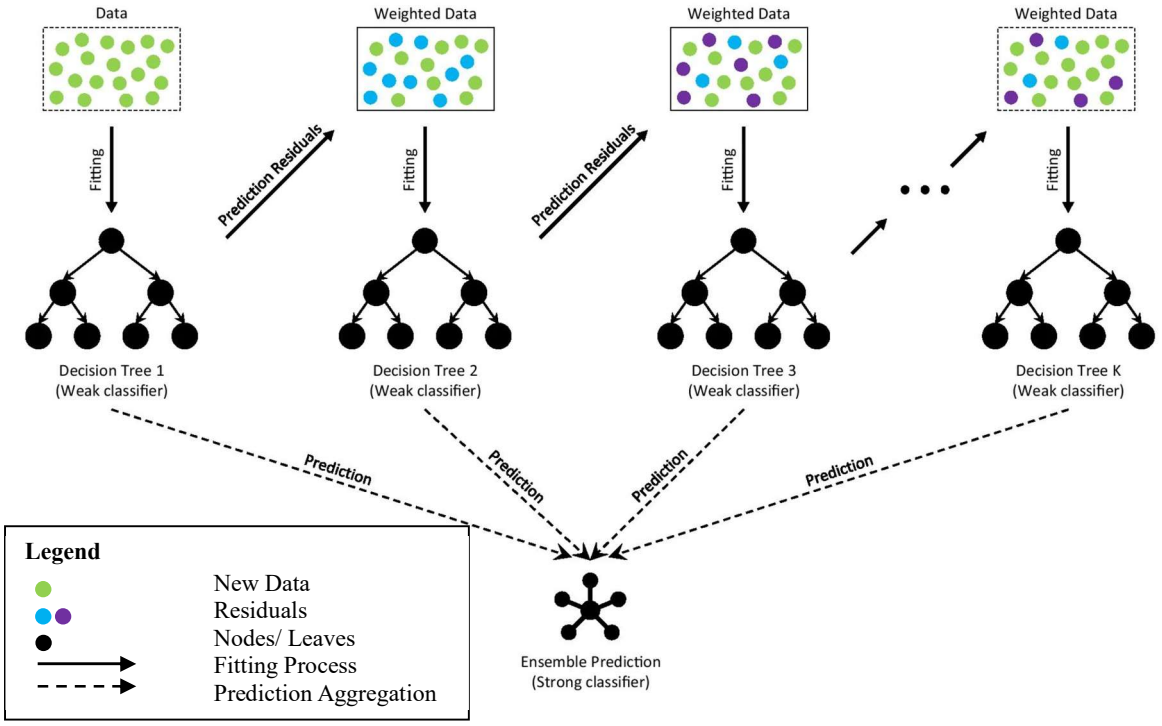
Boosting, like random forests, is a method for combining multiple decision trees, but there are several differences. In random forests, the decision trees are fitted independently, whereas in boosting, trees are fit iteratively, with the subsequent trees focusing on the prediction errors by previously generated models.

The prediction error, also called prediction residuals are used to assign weights to instances, which indicates to the algorithm to put more effort into these. This means that multiple models are fitted to the training set and the subsequent models focus on the instances with higher weights to increase accuracy on those instances which failed in the first place. With this technique the algorithm focuses more on explaining variations in complex features than those which provide clear decision boundaries. At the end of this procedure the predictions of the different models are combined (Polikar, 2012). In addition, the combination process differs between the two approaches, as the combination in random forests is usually computed as the average of all trees while in boosting single trees are weighted based on their performance, highlighting the trees with relevant variables and best splitting decisions. Since this procedure can become computationally expensive, new methods needed to be evaluated resulting for example in LightGBM, which reduces the computational demand in two ways: (1) by keeping instances with large gradients (undertrained instances with large errors), which contribute strongly to information gain and randomly drop the instances with small gradients (Gradient-based One-Side Sampling (GOSS)) and (2) reducing the number of features through selecting and bundling the effective features of the set (Exclusive Feature Bundling (EFB)). In the current modeling strategy, only decision trees with a certain gradient size are kept and the

decision tree splits each node at the most informative feature (Ke et al., 2017). A schematic depiction of the boosted trees algorithm can be found in Figure 5.

Figure 5.

Schematic Graph of Boosted Trees (Deng et al., 2021)



Note. Trees are fit sequentially from left to right (1,2,3 … k), while prediction residuals (data points ≠ green) are assigned to stronger weights in each subsequent tree. Each tree contributes to the ensemble prediction but those are weighted based on tree performance. Circles in single decision trees depict nodes if their respective layer < n. Otherwise these are respective decision leaves.

2.4.2.3 Hyperparameters

Every ML Model needs to be set up with special hyperparameters, which control the behavior of the algorithm. A selection of the used hyperparameters (Ke et al., 2017) will be introduced in this section, as their setting can have a high influence on model performance and three of them will further be outlined in 3.4 *Exploratory Analysis* and 4.1 *Key Findings* section.

As previously described the tree leaves represent the final decisions, such that *number of tree leaves* indicates the number of all possible outcomes the tree can predict. In a regression tree each leaf node has a certain y-value, which, if this leaf was chosen, becomes the prediction for the instance

leading to the decision. For the current problem the *number of tree leaves* should not fall below 100, as the range of IAT is from 0 to 100 and every possible score should be able to be modeled. *Number of boosting rounds* refers to the number of iterations in the boosting algorithm and specifies how many models are fitted to reduce the residuals. The choice of this number is important to avoid under- and overfitting. *Maximum number of bins* refers to the division of continuous features in certain value-bins. This is important for the current data, as FC is a real continuous metric unlike data obtained by Likert-scales and therefore could lead to infinite splitting points (decimal places $\rightarrow \infty$). Controlling this hyperparameter is important for the tree's interpretability and generalization performance. It is further controlled by another hyperparameter *minimum data in bin*. This parameter specifies the minimum number of instances that must be present in each bin and thereby makes bins important for the model and prevents the maximum number of bins from being too high. *Learning rate* determines the scale of the contribution of each tree in the ensemble, such that a low learning rate keeps the contribution of each tree rather low. This means a slower learning process but on the other hand prevents overfitting as the model does not rely too much on single trees.

In addition, three hyperparameters which were tuned were evaluated on their role in model performance. *Column sample ratio per tree* is especially used in ensemble learning and specifies the proportion of features to be randomly selected for splitting at each node. This randomness has an impact on the diversity and correlation of the trees and number of features refers to model complexity such that its tuning is an attempt to avoid under- and overfitting. *Extra trees* is a method to increase the randomness of the feature handling. Instead of the usage of greedy algorithms searching for the best split points, in extra trees the split points and resulting child notes are set randomly. This randomness is helpful to decrease the likelihood of overfitting but oftentimes requires more trees as similar methods. However, compared to methods searching for the optimum split, *Extra Trees* saves computational power. *Path smoothing* refers to a concept of regularization strategies used to deal with complex models, which can end up having a little number of instances in the leaves. Those leaves tend to overfit on the data, which is corrected by path smoothing, where high parameters stand for stronger regularization.

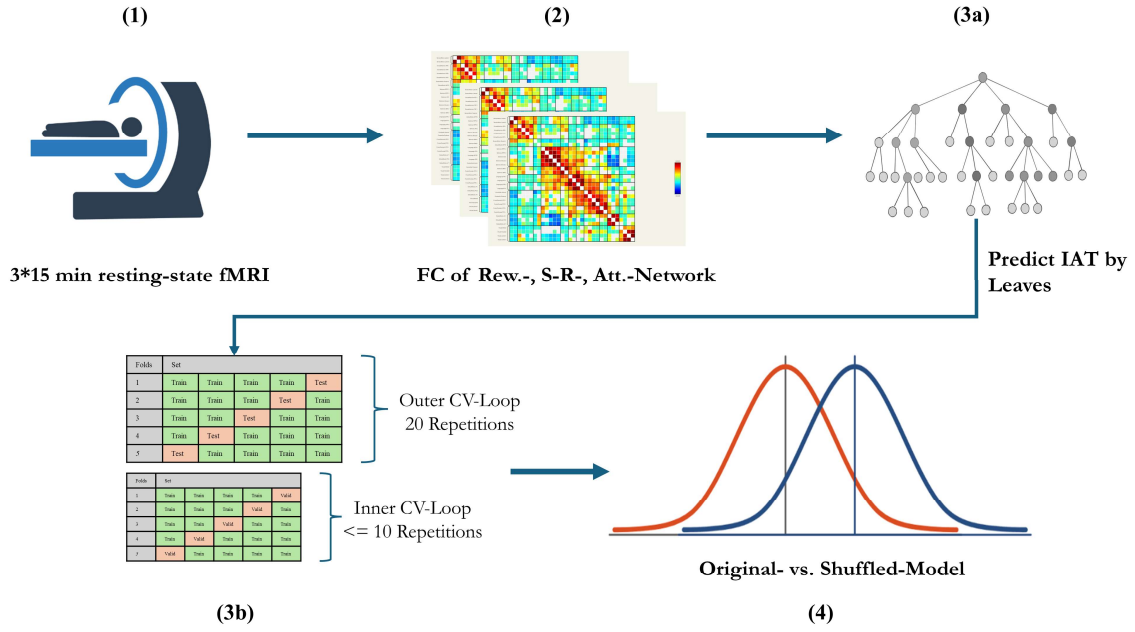
2.4.2.4 Procedure

To assess the generalizability of the model and the overall performance of the regression, a nested CV procedure was used (Cawley & Talbot, 2010). CV has proven as the gold standard in generalizability testing as the dataset is randomly split into independent train and test sets, to test the model on unseen data (Rosenberg & Finn, 2022). It is particularly important for smaller datasets, as variance becomes high with lower sample size, such that CV is a good way to control for randomly large effects. A widely used technique is *k*-fold CV, where the dataset is split into *k* equally sized,

non-overlapping subsets. *k* -1 subsets are assigned to the train set and the leftover represents the test set. This whole procedure is conducted *k*-times leading to *k* performance measurements which can be analyzed (Scheinost et al., 2019). Another crucial step is hyperparameter tuning, which means finding the model settings which are best for predictive performance (*2.4.2.3 Hyperparameters*). Similarly to the goal of the outer CV loop, namely preventing overfitting and therefore maintaining generalizability, the tuning of hyperparameters should be conducted on a train- and tested on a validation-set (Cwiek et al., 2022). Therefore, a nested cross-validation scheme was applied. A schematic overview about the CV procedure can be found in Figure 6.3b.

Figure 6.

Analysis Process



Note. Steps involved in analysis strategy: (1) fMRI data acquisition, (2) preprocessing and FC computation for each subject and the three networks, (3a&b) modeling via regression trees and nested-CV-scheme to predict IAT (CV sets: green, train; red, test/validation), (4) hypothesis testing via paired t-tests between predictor importance between original and shuffled model. Abbreviations: Rew., Reward; S-R, Self-Reference; Att., Attention.

In the main (i.e., outer) CV loop, a 5-fold CV scheme with 20 iterations and an 80/20 split ratio was applied and within each training set of the outer CV loop, another 5-fold CV with at least 10,000 predictions or 10 repetitions and the same split ratio was conducted. Hyperparameters were tuned using a random search procedure (100 random hyperparameter combinations) within the inner CV loop, tuning parameters include: Column sample ratio per tree, using extra trees or not, and path smoothing for controlling tree complexity. Some hyperparameters were not tuned but set to different

values than default: the number of tree leaves was set to 100, the number of boosting rounds was set to 1,000, the maximum number of bins was set to 1,000, the minimum data in bin was set to 1, top rate for gradient-based one-side sampling was set to 0.5, learning rate was 0.01. All other parameters were left in default setting. The best parameter combination was selected based on lowest mean squared error (MSE) and used for training in the respective outer CV repetition.

Furthermore, predictor importance was evaluated using SHAP (Lundberg & Lee, 2017), making even complex models explainable and interpretable. This method acts like a 'virtual lesion' as it removes the respective feature in every possible feature-subset, measuring the changing prediction performance, resulting in a distribution of the predictive value of that single feature (Wu et al., 2023). SHAP values were calculated for each predictor on randomly subsampled groups of n = 37/38 (186 = 4*37+1*38) instances per test set, repeated 100 times (5*20) within the main cross-validation (CV) loop, generating a total of 3,720 SHAP values per predictor.

Two approaches were used to determine global predictor importance: (1) Predictor-wise importance was assessed for descriptive purposes by computing the mean absolute SHAP values across participants within each repetition, resulting in 100 absolute mean SHAP values per individual feature. The features were then ranked based on the overall mean of these 100 absolute mean SHAP values in descending order. (2) Group-wise importance was evaluated for addressing the hypotheses. SHAP values for groups of predictors were calculated like follows: Initially, participant-wise group values were obtained by summing SHAP values of individual predictors belonging to a predictor group. Subsequently, absolute mean values were computed for each predictor group across instances within each repetition, yielding 100 SHAP values per predictor group. To understand the process, the calculation for group-wise importance is shown in Equation 1, divided in three steps: (I) Obtaining SHAP-sums for each predictor-group G, (II) averaging the sums firstly over all participants j in each iteration i and then (III) averaging over all iterations.

Equation 1.

Calculation of Averaged, Grouped SHAP-Values

$$G_{ijg} = \sum_{p=1}^{|p|} S_{ijpg}$$

$$M_{\mathrm{ig}} = \frac{1}{|j_i|} \sum_{j=1}^{|j_i|} G_{ij}$$

$$m_g = \frac{1}{|i|}\sum_{i=1}^{|i|} M_{ig}$$

Note. G_{ijg} is the matrix for the sum of SHAP-values for each group g (reward, self-reference, attention), where p iterates over all predictors in g, S_{ijpg} is the SHAP value of predictor p in group g for prediction j in iteration i. M_{ig} is the matrix of absolute mean SHAP-values of g over all predictions j in iteration i. m_g is the mean for the absolute mean SHAP-values of each g. Values denoted such as $|p|$ represent the length of the vector of values.

The 100 absolute mean SHAP-values per predictor group M_{ig} were used to address hypotheses. To conduct significance testing, SHAP values were computed for the same 37/38 participants per test set (across all iterations *i*) based on models with shuffled target values (values for null-model). This means that predictors were randomly shuffled before the model was fit. This creates by logic the perfect null-model (representing the null-hypothesis), that features will not have any meaningful relation to the outcome measurement. Therefore, the calculation process described in Equation 1 has been conducted both for the original and the shuffled model.

For hypothesis testing the resulting distributions M_{ig}, consisting of $N = i = 100$ mean absolute SHAP values per group, were compared using corrected one-sided paired *t*-test for all predictor groups. A *t*-test determines if there is a significant difference between the means of two groups. Two choices for this method have been made: (1) The test was one-sided as there is an expected direction of difference. It is assumed that M_{ig} is higher for the original than for the shuffled model. (2) The test must be paired because of the characteristic of the CV-scheme. The variability of prediction performance based on the random split into train- and test set is assumed to have a large impact on model performance, therefore it is necessary to counter this influence by pairing M_{ig}, such that the values for each model are compared within the respective iteration (Bouckaert & Frank, 2004; Nadeau & Bengio, 1999).

2.4.2.5 Inference Criteria

Sample sizes for one sided paired *t*-test were derived from the number of repetitions within the outer CV loop ($N = i = 100$). Yet, guidelines for significant effect strength thresholds in machine learning-based analyses, particularly comparing original vs. shuffled distributions, are lacking. Therefore, established benchmarks like $d = 0.50$ standardized mean difference representing moderate effect strength (Cohen, 1988) were adopted for global model performance assessment (mean absolute

error (*MAE*) and R^2). Criteria for model interpretation was original vs. shuffled *MAE* distributions' standardized mean difference ≥ .50 (lower means for original models) with a significance threshold at $p < .05$ by corrected one sided paired *t*-test. For comparison of the difference for R^2 distributions the assumption was vice versa, as higher values indicate better prediction. Considerations for the use of paired-*t*-test were the same as for hypothesis testing (*2.4.2.4 Procedure*). Calculation of *d* was derived by dividing *MAE* score mean difference by the shuffled model's standard deviation (not pooled standard deviation). Based on traditional formulas (Seabold & Perktold, 2010), the *t*-test had 97% power for detecting $d = .50$ ($N = 100$ per group; paired-sample; one-sided). For both significance tests of model scores (*MAE*, R^2) and hypothesis testing (SHAP values), inferences were based on $p < .05$ (paired; one-sided), whereas in hypothesis testing, SHAP values in the original model were expected to be larger.

2.4.3 Exploratory Analysis

To gain further insights in the modeling results, two additional exploratory analyses were conducted. The first focused on the hyperparameters of the model and the second strives for an explanation of prediction performance of age and sex as features in the model. To investigate the underlying mechanisms of the modelling algorithm a linear model with hyperparameters as predictors and *MAE* as the outcome were fitted. Characteristics and prerequisites for a linear model are described in *2.3.1.2 Linear Model* and will be applied for this analysis as well. Further, feature importance was analyzed for the control variables age and sex in the same way as hypothesis testing (*2.4.2.4 Procedure*), such that one-sided paired *t*-tests were conducted to see if these features performed better in the original than in the shuffled model.

2.5 Software

FC computation was conducted in MatLab v. R2023b (https://matlab.mathworks.com) with additional toolboxes CONN-toolbox (v. 22.a; Nieto-Castanon & Whitfield-Gabrieli, 2022) and MRIcroGL (v. 1.2.20220720; https://www.mccauslandcenter.sc.edu/mricrogl/). All statistical analyses subsequent to FC computation were performed with Python (v. 3.11.6) with Spyder IDE (v. 5.4.3), using the following libraries: numpy (v. 1.24.3), pandas (v. 2.1.1), lightgbm (v. 4.0.0; Ke et al., 2017), scipy (v. 1.11.3), shap (0.42.1; Lundberg & Lee, 2017), pingouin (v. 0.5.4) and scikit-learn (1.3.0).

3 Results

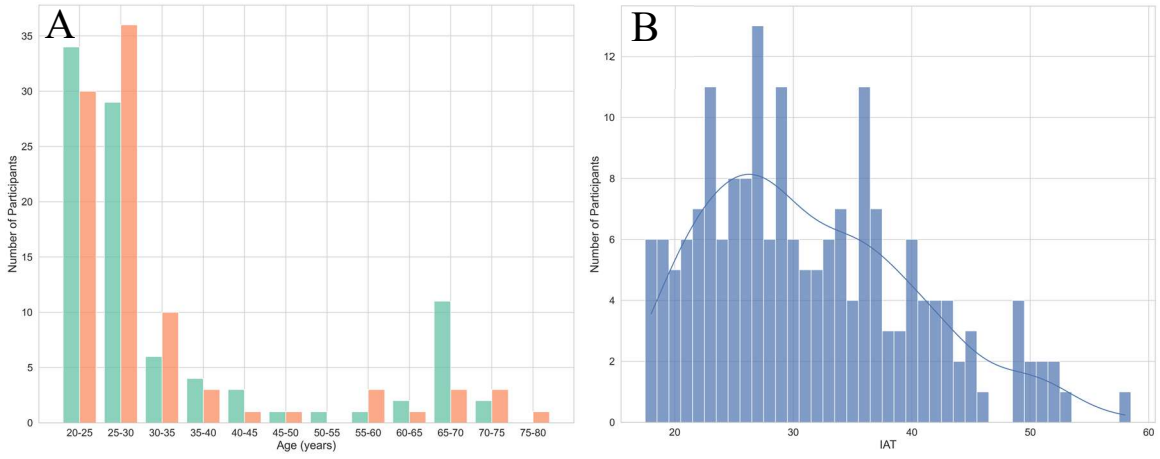
The Results section is divided into four steps: Descriptive, Model Performance, Hypothesis Testing and Exploratory Analysis and presented in respective order.

3.1 Descriptive Analysis

All selected participants provided IAT measurements, leaving a sample size of 186 participants. Participants' sex was almost equally distributed, 50.538% categorized themselves as female. Age ranged from 20 to 80 while the most frequent age bins were 20-25 (34.409%) and 25-30 (34.946%). Mean IAT-scores were 31.333 ± 8.841 (min.:18; max.:58). Distribution of participants in IAT-severity scores is like follows: N_{normal} (0-30) = 99 (53.226%), N_{mild} (31-49) = 79 (42.473%), $N_{moderate}$ (50-79) = 12 (4.301%) and does not yield normality based on a SW ($W(184) = .958$, $p < .001$). While this does not affect the choice of model, it must be discussed in the following section. No participant displayed severe IA, hence in total the proportion of Internet addicted individuals in this sample was 4.301%. Details in demographics and IAT values are presented in Figure 7 A&B.

Figure 7.

Distributions for Descriptive Analysis



Note. (A) Distribution of age stratified by sex (green: female; red: male) and binned into 5-year bins. (B) Distribution and density of IAT (Internet Addiction Test) values.

Before the MW-U test, a KS test has been conducted to test for equal age distributions between both sexes. The distributions were found to be equal ($D(186) = 0.092$, $p = .776$) and the subsequent MW-U test revealed no significant difference in age between men and women (U =

4228.0, $p = .786$). For descriptive explanation of the relation of age and sex on IAT a linear regression model was fitted after testing for its assumptions. The assumption of linearity between age and IAT was given (Figure S1). A SW test revealed non-normality for the residuals ($W(186) = .968$, $p < .001$). This does not influence the choice of the regression based on the assumptions of the CLT, that parametric tests still can be interpreted if the sample size exceeds 30. This holds true in this case, particularly with incorporation of the QQ-plot (Figure S2), which indicates, that the relation between theoretical quantiles and sample quantiles follows approximately a straight line and therefore the assumption of normality does not have to be rejected. However, this result influences the choice of the test for homoscedasticity towards the *White-Test*. The residuals were homoscedastic ($W(2) = 9.294$, $p = .054$) and a significant regression was found ($F(2,183) = 19.0$, $p = < .001$, adj$R^2 = .163$). Age negatively predicted the severity of IAT ($B = -1.198$, $t = -6.039$, $p < .001$) while sex (0 = male, 1 = female) did not significantly contribute to the model ($B = -.918$, $t = -0.772$, $p = .441$). Results and details of the linear model are displayed in Table S2, and the model equation looks as follows:

Equation 2.

Linear Model of IAT

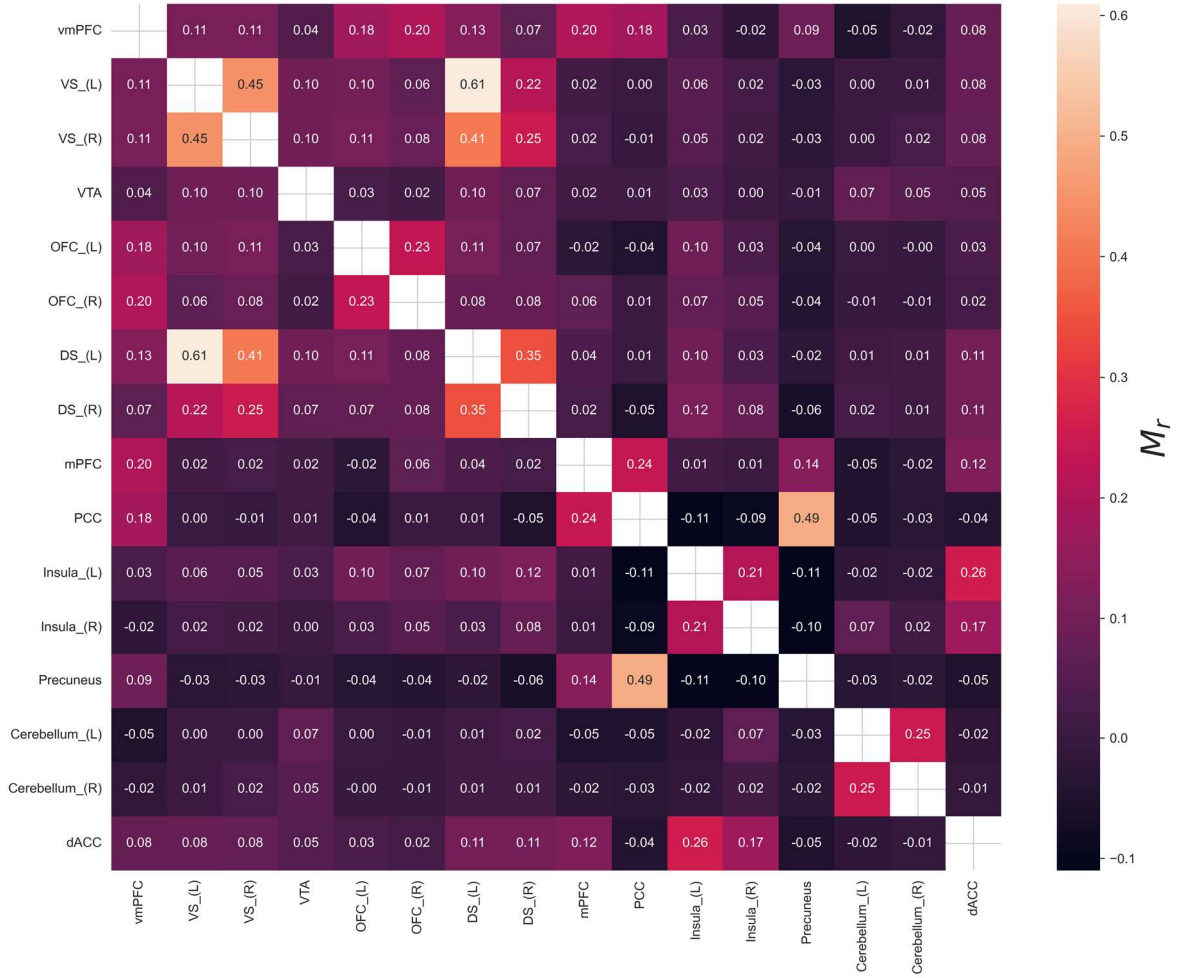
$$IAT = 35.502 - 1.198 * age - .918 * sex$$

For descriptive purposes the FC-matrices for each participant were averaged, such that the average correlation for each connectivity could be observed. The resulting correlation matrix can be seen in Figure 8. Most edges center around zero with few peaks mostly in positive direction. The average correlation strength over all edges is $M_r = .062$ with a SD_r of .115 (*min* = -.01, *max* = .61). Furthermore, the average correlation for edges between networks was $M_r = .028$ with a SD_r of .124. As expected, those values were lower as the average connectivity strength within the networks (Reward: $M_r = .16 \pm .195$; Self-Reference: $M_r = .045 \pm .197$; Attention: $M_r = .052 \pm .196$). Highest connectivity strength was observed for regionally and functionally overlapping regions such as DS_(L)_VS_(L) ($M_r = .61$), DS_(L)_DS_(R) ($M_r = .41$), VS_(L)_VS_(R) ($M_r = .45$) or PCC_Precuneus ($M_r = .49$). However, functional networks are not only determined by positive correlations and the correlation coefficient has the special characteristic, that 0 is the lowest connectivity strength while $r < 0$ still represents connectivity. Therefore, average FC was additionally computed with absolute values means across participants ($absM_r = .083 \pm .102$). Connectivity

strength between networks was still lower than within networks (Between: $absM_r$= .048 ± .049, Reward: $absM_r$= .133 ± .132 Self-Reference: $absM_r$= .112 ± .089, Attention: $absM_r$= .094 ± .124).

Figure 8.

Averaged FC-Matrix



Note. All FCs were averaged over participants (N = 186). Lowest connectivity values are coloured black, whle highest connectivities tend towards white. Theoretical range is [-1:1] but color scale only ranges from [-0.1:-0.6] due to low average connectivity. Abbreviations: vmPFC, ventral medial prefrontal cortex; VS, ventral striatum; VTA, ventral tegmental area; OFC, orbitofrontal cortex; DS, dorsal striatum; mPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; dACC, dorsal anterior cingulate cortex.

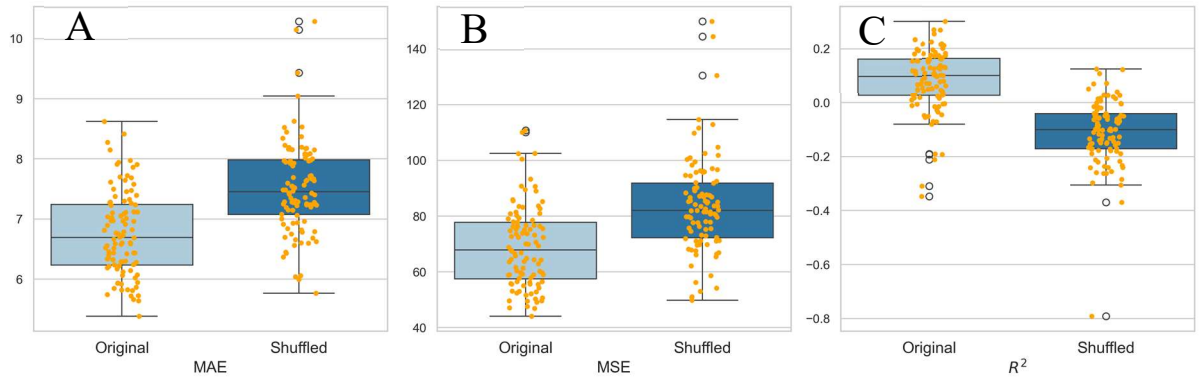
3.2 Model Performance

Model evaluation with paired, one-sided t-tests showed for all three metrics a significantly better fit of the original versus shuffled model (*MAE*: t = -12.971, p <.001, d = 1.035, *95% CI* = [-

∞:-0.64]; *MSE*: t = -13.647, p <.001, d = 0.912, *95% CI* = [-∞:-12.19]; R^2: t = 14.917, p < .001, d = 1.855, *95% CI* = [0.17:∞]) (see Figure 9A/B/C respectively), all tests with df = 99, meeting the criteria to interpret the model based on SHAP values. As can be seen in Figure 9, the raw data contained outliers in almost every distribution. Therefore, the tests were conducted with *winsorized* means. Winsorized *MAE* (M = 6.76 ± 0.697 vs. M = 7.497 ± 0.708) and *MSE* (M = 69.03 ± 14.436 vs. M = 82.912 ± 15.148) of the original model are lower than the shuffled version. In the case of R^2 the original model revealed higher scores than in the shuffled model (M = 0.086 ± 0.104 vs. M = -0.105 ± 0.103). All test statistics including the pre-*Winsorizing* metrics are also reported in Tables S3/S4.

Figure 9.

Model Performance Comparison



Note. Boxplots to compare model performance between the original model (light blue) and the shuffled model (dark blue). Scores are represented by *MAE* (mean absolute error), *MSE* (mean squared error) and R^2 (r-squared).

For descriptive purposes the predictors with their respective SHAP-values are reported in descending order, Table 2 displays the ten predictors with the highest values. The first two predictors (age = 0.069, sex = 0.024) in the original model yielded much higher SHAP-values than the first two predictors in the shuffled model (sex = 0.034, VS_(R)_VTA = 0.015). But the subsequent predictors values are converging, indicating a very low explanatory value in the original model. The complete listing of all predictors can be found in Table S1. A Beeswarm-plot (Fig. S5) shows that SHAP-values for age increased with reduced age, indicating a more informative variance of IA for individuals with lower age. Further it reveals that the FC of OFC_(L)_OFC_(R) becomes more important with higher values but nothing can be said about the direction of the influence, and as overall model-performance and the mean SHAP-values for this predictor is almost zero (-.008) this observation is not meaningful for any discussion.

Table 2.

Sorted SHAP-Values per Model Version

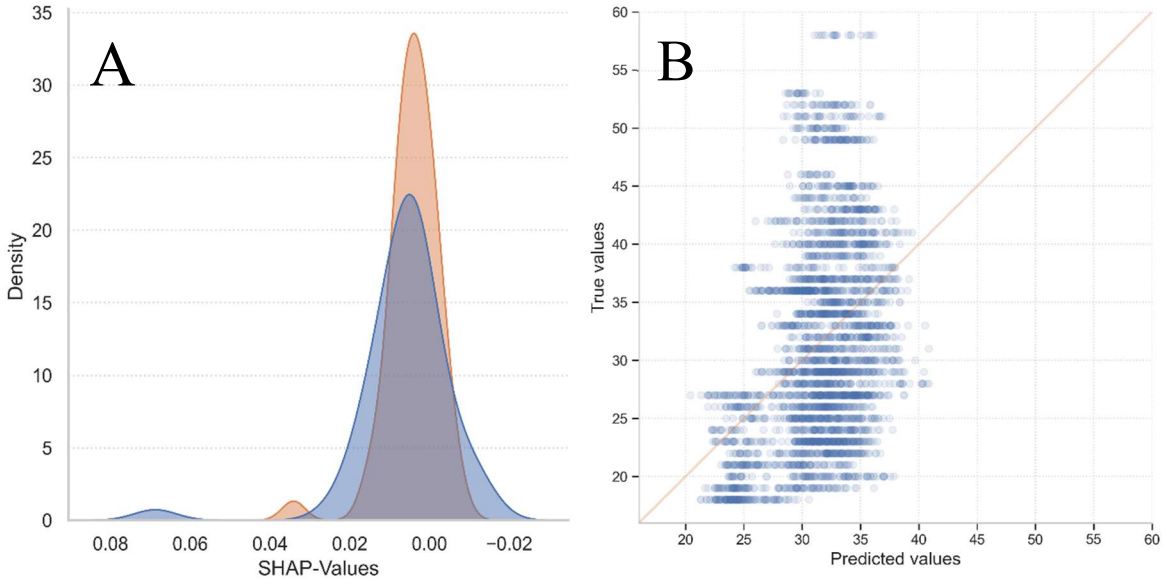
Original			Shuffled		
Feature	*M*	*SD*	Feature	*M*	*SD*
age	0.069	0.569	sex	0.034	0.097
sex	0.024	0.063	VS_(R)_VTA	0.015	0.065
VS_(L)_VTA	0.018	0.049	PCC_Insula_(R)	0.014	0.066
Insula_(L)_Insula_(R)	0.017	0.035	VTA_DS_(R)	0.013	0.05
vmPFC_dACC	0.017	0.027	OFC_(R)_DS_(L)	0.010	0.041
vmPFC_PCC	0.016	0.032	VS_(L)_DS_(R)	0.009	0.046
VS_(L)_OFC_(R)	0.016	0.025	vmPFC_PCC	0.009	0.055
VTA_OFC_(R)	0.013	0.06	Insula_(L)_Insula_(R)	0.009	0.052
vmPFC_VTA	0.012	0.05	vmPFC_VTA	0.009	0.057
vmPFC_Cerebellum_(R)	0.012	0.06	PCC_Cerebellum_(L)	0.009	0.051

Note. Listing of the ten strongest features by SHAP-values for each model. *M* = mean, *SD* = standard deviation.

That SHAP-values between model versions do not differ to an important extent can also be observed in Fig. 10A, which represents the density of SHAP-values for both models and indicating that both distributions are centered around zero, indicating no explanatory value. For the shuffled curve values range from 0.034 to -0.007 ($M = 0.004 \pm .006$), while the original curve shows a higher

Figure 10.

Distribution of SHAP-Values and Predicted Labels



Note. (A) Density of SHAP-values for original model (blue) and shuffled model (orange). (B) Relation between true and predicted values, orange line represents perfect fit.

spread and ranges from 0.069 to -0.013 ($M = 0.006 \pm .012$). It is to be highlighted, that original values contain more negative values, indicating learning effects which could not be transferred from the train to the test set. Fig. 10B displays the relation between real and predicted values. Smaller true values could be at least partially predicted while the model failed completely for values >41 (min: 20.405, max: 40.878).

As the variance in SHAP values of different iterations is assumed to be high, deeper insights about the distributions of the SHAP-values for single predictors are necessary. Therefore, they were listed based on the following three criteria: lowest minimum, highest maximum and lowest standard deviation over all 3,720 predictions. This listing can show the variations of feature importance over different CV-iterations and sample splits. It is noticeable, that those values with the highest maximum SHAP-values also show low minimum values and therefore a high standard deviation and variance. This listing is displayed in Table 3.

Table 3.

Predictors Sorted after SHAP-Stats

$Feature_{min}$	*min*	*max*	*SD*	*M*
vmPFC_Cerebellum_(L)	-0.372	0.975	0.132	-0.003
vmPFC_OFC_(L)	-0.451	0.979	0.114	0.01
PCC_dACC	-0.458	0.447	0.101	0.005
Cerebellum_(R)_dACC	-0.48	0.588	0.126	0.008
vmPFC_VS_(R)	-0.501	0.673	0.106	0.001
Cerebellum_(L)_Cerebellum_(R)	-0.511	0.659	0.140	0.002
VTA_DS_(L)	-0.534	0.582	0.131	0.003
VS_(R)_DS_(L)	-0.534	0.74	0.122	0.004
VS_(R)_VTA	-0.576	0.909	0.152	0.002
PCC_Precuneus	-0.593	0.64	0.112	0.004
$Feature_{max}$				
OFC_(L)_OFC_(R)	-1.596	2.983	0.763	-0.008
age	-0.967	2.612	2.81	0.069
VS_(R)_OFC_(L)	-0.69	2.153	0.227	0.005
VS_(R)_DS_(R)	-1.292	1.855	0.489	-0.013
OFC_(L)_DS_(R)	-0.864	1.775	0.256	0.002
vmPFC_Cerebellum_(R)	-0.931	1.713	0.353	0.012
sex	-1.289	1.583	0.348	0.024
mPFC_PCC	-1.032	1.514	0.322	-0.01
Precuneus_Cerebellum_(R)	-1.247	1.482	0.371	-0.013
vmPFC_DS_(L)	-1.322	1.378	0.294	< 0.001
$Feature_{sd}$				
PCC_dACC	-0.458	0.447	0.101	.005
Precuneus_Cerebellum_(L)	-0.634	0.618	0.103	0.004
vmPFC_VS_(R)	-0.501	0.673	0.106	0.001
Precuneus_dACC	-0.681	0.652	0.11	0.004
PCC_Precuneus	-0.593	0.64	0.11	0.004

vmPFC_OFC_(L)	-0.451	0.979	0.114	0.01
mPFC_Insula_(L)	-0.703	0.515	0.115	0.008
PCC_Cerebellum_(R)	-0.661	1.136	0.120	0.002
VS_(R)_DS_(L)	-0.534	0.74	0.122	0.004
Cerebellum_(R)_dACC	-0.48	0.588	0.126	0.008

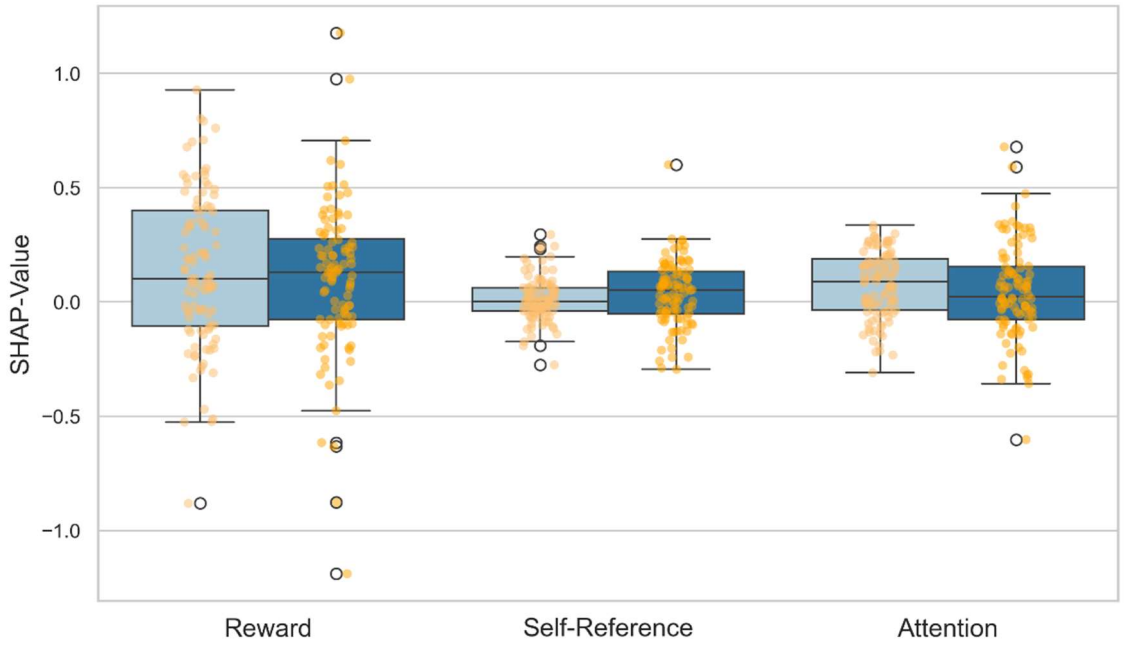
Note. The first ten features of each sorting scheme are listed. Underscore at feature indicates the metric the list is sorted after. *min* = minimum, *max* = maximum, *SD* = standard deviation, *M* = mean

3.3 Hypothesis Testing

For hypothesis testing SHAP-values in each CV-iteration have been summed up by networks affiliation and averaged over all participants in the test set (*Equation 1*). Again, original values were tested against shuffled values in three one-sided, paired *t*-tests, one for each network. A boxplot (Figure 11) revealed multiple outliers which were removed with *Winsorizing*. Results indicated no significant difference in the expected direction (higher values for original vs. shuffled datasets). Beyond that, reward ($t = 0.713$, $p = .239$, $d = 0.11$, *CI95%* = [-0.05: ∞]) and attention ($t = 1.225$, $p =$

Figure 11.

Hypothesis Testing



Note. Boxplots indicating SHAP-values for each network stratified by model version (light blue: original, dark blue: shuffled)

.13, $d = 0.169$, 95% CI = [-0.01: ∞]) networks showed differences in the hypothesized direction, but they did not yield significance. Furthermore, Figure 11 revealed, that the Median in the comparison for the reward-network differed opposite to the expected direction. This holds true even for the winsorized dataset (*Med* = 0.102 vs. *Med*. = 0.128; original vs. shuffled). An almost similar pattern was revealed for the self-reference network, such that original values showed even lower values in the original vs. shuffled dataset ($t = -1.495$, $p = .931$, $d = 0.193$, 95% CI = [-0.05: ∞]). Representative for original vs. shuffled means are Reward: $M = 0.137 \pm 0.336$ vs. $M = 0.103 \pm 0.29$; Self-Reference: $M = 0.012 \pm 0.09$ vs. $M = 0.035 \pm 0.136$ and Attention: $M = 0.075 \pm 0.146$ vs. $M = 0.046 \pm 0.197$ respectively. Those results do not meet the inference criteria, leading to a rejection of every hypothesis. Results are presented in detail also for pre-*Winsorizing* values in Tables S6/S7.

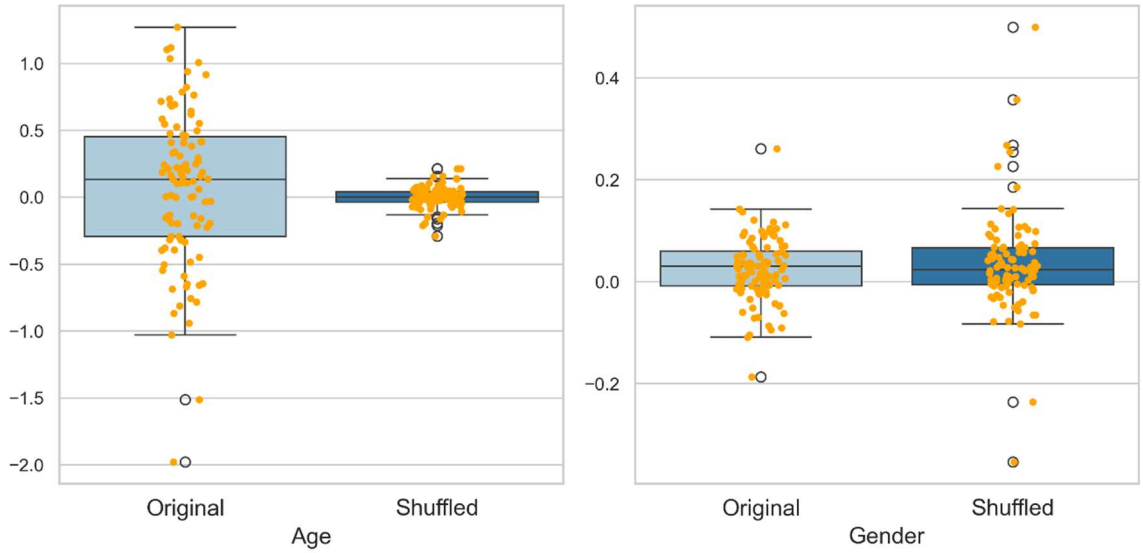
3.4 Exploratory Analysis

For explanatory purposes the distribution of the best hyperparameters retrieved from the inner CV-loop were computed and their influence on *MAE* were analyzed fitting a linear model. *Colsample By Tree*, a hyperparameter taking continuous values from 0 to 1, was high over most iterations ($M = .926 \pm 0.819$, *min* = .445, *max* = .999, *Med.* = .95), *Extra Trees*, a categorial hyperparameter, indicated, that this method was used in every iteration (1 (used) = 100%) and *Path Smooth*, a continuous value, was rather low in average, exhibited a high *SD*, *M* and *Med*. differed strongly ($M = 15.583 \pm 60.161$, *min* = 0.101, *max* = 337.033, *Med.* = 0.464), so that it was assumed, that the distribution was highly skewed, which was confirmed by SW ($W(100) = .27$, $p < .001$). Therefore, the median should be interpreted for the estimation of the central tendency, which explains, that most values were close to zero. However, as a linear model does not assume normality for its predictors, the analysis could be conducted as planned. In the following linear model, *Extra Trees* was not included, as it just took one value and therefore did not show any variance which could influence MAE. Linearity was given for the relationship between both hyperparameters and *MAE*, however it could further be observed in Figure S3 & S4, that there is no significant relationship. The SW-test revealed normality for the residuals of the model ($W(100) = .979$, $p = .11$) and the BP-test revealed homoscedasticity for the residuals ($X^2(3) = 3.44$, $p = .179$). Therefore, the linear model can be interpreted. As assumed by linearity plots, it did not reach significance ($F(3,96) = 0.004$, $p = .996$, adj$R^2 = .0$). While the hyperparameters *Path Smooth* ($B < 0.001$, $t = -0.018$, $p < .986$) and *Colsample By Tree* ($B = -0.077$, $t = -0.089$, $p = .930$) did not show any relation to MAE. Further details can be seen in Table S8.

To analyze the relevance of the non-neurological features age and sex the same comparison analyses like in hypothesis testing were performed. Neither of the two variables were predictive of IA given the non-significant differences between (winsorized) original and shuffled SHAP values (age: $t = 1.312$, $p = .096$, $d = 0.186$, *95% CI* = [-0.02: ∞]; sex: $t = -0.773$, $p = .779$, $d = 0.108$, *95% CI* = [-0.02: ∞]). Means for age differed in the expected direction for original vs. shuffled models ($M = 0.069 \pm .569$ vs. $M = 0.001 \pm 0.077$), while SHAP-means for sex were higher in the shuffled models ($M = 0.024 \pm 0.063$ vs. $M = 0.034 \pm 0.097$). Details of data before transformation can be found in Figure 12 and Tables S9/S10.

Figure 12.

Model Comparison Control Variables



Note. Boxplots to compare SHAP-values of control variables between the original model (light blue) and the shuffled model (dark blue).

To estimate the effect of *Winsorizing* on the inference criteria, the mean deviations of *p*-values were calculated and revealed a low negative difference ($M = -.009 \pm .014$), indicating that in average *p*-values became lower. However, in eight *t*-tests both approaches led to three significant and five non-significant results, such that the outlier strategy did not influence the rejection of hypotheses.

4 Discussion

The discussion section starts with summarizing the key findings of the study, followed by examining the differences to the existing body of knowledge and discussing the reasons for divergence. Finally, the methodological limitations are addressed.

4.1 Key Findings

The current sample showed an equal frequency of men and women, and sex was distributed equally over differing age bins. General IA in this sample was low with a proportion of 4.3% addicted individuals, particularly compared to the global prevalence of 7.2% (Pan et al., 2020). As expected, age was negatively associated with IA, being in line with the assumption that younger people are socialized with digital technology making them more prone to become addicted (Greenfield, 2012; Lozano-Blasco et al., 2022). Sex was not significantly associated with IA, which can be explained by the general focus of the IAT, such that different specific forms of Internet use are combined within this scale. Therefore, the differing characteristics of Internet usage could not be observed with the instrument at hand (Montag et al., 2015). For example, while it can be assumed, that women are more active in social media use, men use the Internet more for gaming purposes. Those using patterns could have equaled out, so that no difference was observed.

Average FC over all features was low and in combination with a comparably high *SD* indicates the high individual variability, such that most connections could be either positive or negative without a clear trend. Connectivity strength was higher within regions of the same network than distinct regions, indicating, that networks were specified correctly as FC-strength within networks is assumed to be higher than between (Smith et al., 2013). Although it must be noted that only the reward network showed a considerable within-FC strength for non-absolute mean.

Modeling was moderately successful as all three scores in question (*MAE*, *MSE* and R^2) showed significant differences in the expected direction and inference criteria. However, the low R^2 indicated model failure, which was confirmed by sorting the features based on their SHAP-values in descending order. Unexpectedly, the original values were slightly better, however the difference was primarily due to the control variables age and sex. Age was almost three times as important than the following feature but seemed to only be relevant among younger individuals (Figure S5), sex was surprisingly important in the shuffled model, however this may be coincidental, as there cannot be any meaningful relation in the shuffled model. Contrary, neuronal features did not differ significantly between model versions (Table 2/S5 and Figure 10A), such that it can be assumed that the differences were caused by random variation. The predicted values ranged from 20.405 to 40.878 and exhibited a much narrower distribution than the original IAT values. With a maximum value in the current

sample of 58, many instances could not be modeled. Further investigation revealed that in the original model, the features associated with the highest positive SHAP-values also had the highest negative SHAP-values, resulting in a large standard deviation and variation over CV-iterations. This could mean two things: either those features were important on some subsets but detrimental on others, or their high importance is due to the model learning noise.

In line with the fact of poor performance metrics, all hypothesis tests revealed no significant difference in the expected direction, leading to the conclusion that the research question must be answered in the negative, IA cannot be predicted by rsFC of the reward, self-reference and attention network. The reward and attention network showed differences in the expected direction, but these were not statistically significant. It rather should be emphasized, that comparing the median between model versions for the reward network revealed a difference in the opposite direction, highlighting the need to reject the hypothesis. Furthermore, the means of the model versions for the self-reference network revealed a difference in the direction opposite to the expected one.

The linear model, estimating the MAE by the tuned hyperparameters *Colsample by Tree* (CbT) and *Path Smooth* (PS) as predictors, did not reveal any significant relationship, implying, that the combination of best hyperparameters varied across the CV-iterations and that no inferences about ideal model settings for this type of predictive modeling can be drawn. Values for CbT were generally high, that the model decided to use almost the whole feature-space for tree construction. The result can be interpreted with respect to the poor predictive power of individual features. Even though the feature to sample ratio was rather high, only a little number of features has been dropped because there were no features with high explanative value, the model could rely on. For some instances, this led to overfitting, as almost every feature was included but turned out to be detrimental for generalization, such that in some iterations the model scored R^2 down to -.348. Somewhat surprising is the combination with generally low PS values, as PS is a hyperparameter controlling for overfitting (Ke et al., 2017). When the feature-sample ratio is high and CbT does not control sufficiently for using too many features, the probability for a little number of instances in every leaf increases such that PS should regularize stronger. However, considering the distribution of the IAT, most values were assigned frequently, and only high values were sparse. This could explain the few but strong outliers in PS, such that in four iterations the parameter must be set to > 200 while mostly the parameter ranged between 0 and 1. That *Extra Trees* was used in every iteration indicates that increased randomness in the learning process was advantageous for predictive performance. Concluding, the data at hand was particularly complex or noisy, such that the search for best splitting points should be conducted rather randomly than gradually optimizing. The finding is unsurprising as complexity and noise are typical of neurological data.

Testing the predictive importance of the control variables revealed no significant differences in the expected direction. This counters the finding, that age became a significant predictor of IAT in

the linear model fitted for descriptive analysis, as well as the strongest feature in SHAP-values for the original model. The discrepancy could be explained by the different approaches of the linear model and machine learning modeling as the former was neither tested on generalizability nor compared to a null model. For the non-significance of sex in the comparison the same reasoning as for non-significance in the linear model can be applied. Despite these observations, it is noteworthy that sex was among the most important predictors for the original and shuffled model. As high feature importance in the shuffled model must represent learning from noise, the result that the SHAP-value for sex in the original model is lower, also indicates learning from noise to create a conclusive picture of those observations.

The presentation of the average *p*-value difference between original and winsorized *t*-statistics indicates overall decreased *p*-values and thus the risk of Error I inflation. However, the deviation was rather small and did not affect the resulting significances. *Winsorizing* rather resulted in a more accurate estimation of means for the tested parameters, such that the application of this method can be concluded as successful.

4.2 Why did the Model Fail?

Even though the original model performed significantly better than the shuffled model, the performance of the original model was not sufficient. This makes the interpretation of the feature importance almost impossible and further analysis showed, that the predictor groups (networks) cannot be said to be significant contributors to the model. Therefore, this section does not address the role of single neurological features but rather makes assumptions about the causes of failed replication of previous findings outlined in the theoretical background. The section starts with reasoning about the causes for the non-contribution of the networks in predicting IAT and further outlines methodological deviations which could explain model failure.

4.2.1 Reward

Most studies' findings point to the involvement of reward circuits in IA and SMU (Montag, 2019). Therefore, the actual finding, that the reward network is not predictive of the IAT is surprising, particularly given that the argument of the IAT as a too generalized measurement does not hold true, as every aspect of Internet usage and thus every form of Internet use disorder should be associated with rewarding experiences (Brand et al., 2016). This finding could be attributed to the operationalization of the reward network in this study. Based on previous studies and theoretical assumptions the ROIs of the reward network have been collected and added together. This could

have resulted in an inflated network, such that ROIs which are less involved in reward circuits have been included and interfered the relevance of other structures. While Meshi et al. (2015) only proposed VTA, vmPFC and VS for the reward system, the OFC and DS were also included. This issue may also be related to the findings, that rather the whole brain connection of single ROIs is associated with IA than the FC within entire closed and specified networks (J.-T. Sun et al., 2023). For example, the OFC is the only region thought to be part of the reward network, which is also consistently reported to be affected by IA, but only in voxel-based morphometry studies (J.-T. Sun et al., 2023). However, if this finding is reliable, the descriptive comparison of single feature SHAP-values for both models would yield a different picture. The connections of specific ROIs (e.g. OFC) to all ROIs in the same network should have become more frequent but as this requirement is not fulfilled (Table 2 & 3), no such pattern has been observed. This could be due to the high sample size, in comparison to other neuroscientific studies, such that possible effects of alteration could have equaled out.

Besides, the frequently mentioned dopamine pathways mainly found between the striatum (VS, DS) and PFC (Brand, 2022; Volkow et al., 2019) could not be found to be altered in this study. This result may be caused by fMRI as imagining method as it is already stated, that the functionality of neurotransmitters cannot be observed by neuroimaging and even if correlates are found, estimated relations are based on reversed inference (Montag et al., 2023). One may argue, that increased exchange of neurotransmitters should cause inhibitory or excitatory effects which could be seen in fMRI data, but the findings of this study cannot confirm that, at least not for rs-fMRI. The assumption, that the linkage of regions based on neurotransmitters is not observable in rs-fMRI is in line with the findings, that no connectivity in the dopamine-pathways is consistently associated with IA. Among FC studies, abnormal FC was only consistent between middle occipital gyrus and dlPFC and the dorsal attention network for SMU studies (Wadsley & Ihssen, 2023) and whole brain FC of the PCC and insula were found for general IA (J.-T. Sun et al., 2023). All those regions are no part of the reward system outlined in this study. Findings about reward circuits could only be observed for reduced GMV (Wadsley & Ihssen, 2023). Nevertheless, FC in the reward network showed the highest mean SHAP-values and listing the FC based on the highest max values indicated higher frequency for those of the reward network. Even though the comparison between both models did not yield significance, the idea of rewarding factors of the Internet and social media shaping the risk for addictive behavior should not be rejected.

4.2.2 Self-Reference

In contrast the considerations regarding the self-reference network were rather theoretical and there is less empirical evidence for an involvement of this network in IA. The reasoning for the

investigation of its involvement relied mainly on the SMU properties of IA, which seem to play a major role, even in the general IAT (Montag et al., 2015). However, self-referential cognition is possibly only significantly affected by SMU, which has not been captured to a sufficient degree by the scale at hand. Furthermore, although the listed studies in the theoretical background were able to reveal connections to self-relevant constructs, these may still be very different with terms such as identity-confusion, self-esteem, self-kindness, and body-image disturbance. Possibly all those constructs are depicted differently in brain activation patterns so that the operationalization for the self-reference network was not able to capture them properly. For example, there must be a differentiation between cognitive and physical representation of the self which would be the most obvious in the comparison of identity-confusion and body-image disturbance. While identity-confusion refers to the building of stable beliefs and roles, body-image disturbance may be stronger related to the perception of the own attractiveness and appearance. Although both constructs are related to self-esteem and self-kindness (Hsieh et al., 2019; Saiphoo & Vahedi, 2019), the underlying neuronal properties could vary. For example the ROIs of the self-refence network proposed by Meshi et al. (2015) were expanded by the insula with the argument, that the insula is responsible for interoception and therefore for a sense of self. But this process is rather of physical nature (Namkung et al., 2017), which is supported by the finding of disruption of smoking in individuals with damaged insula (Naqvi et al., 2007). However, the conclusion, that the insula is rather relevant in SRAs, where drug intake triggers physiological changes cannot be correct, as the insula was one of the few regions, which in a meta-analysis was revealed to show stronger rsFC to the whole brain in individuals with IA (J.-T. Sun et al., 2023). At this point, the functional connectomes approach of this work can be questioned, as like outlined for the reward network, consistent findings are more common for single ROIs and their connectivity to the whole brain.

Another consideration must be the high individuality of identity constructs. It may be possible to recognize differences in self-esteem within the brain, however it will probably be more difficult to map attitudes, values, and beliefs via neuroimaging. As these are highly complex constructs and therefore extremely variable it will be difficult to find recognizable patterns for the model to decide on the severity of IA in these individuals. For this purpose, more complex models as well as larger sample sizes would be needed. An example for the difficulty of discovering such patterns could be the fact, that significant others (parents, partners) can be incorporated into one's self concept and that this can also be observed in excessive online gamers who start to identify with their game character (Choi et al., 2018). Even if this effect is reliable, it is not possible to differentiate if the person to identify with is real or part of an online game. The authors found differences between the addicted group and non-addicted group but did not investigate, if the same neuronal correlates would be activated if the non-addicted individuals are thinking of their real significant others. Furthermore, this study has been conducted with task-activation, such that the participants saw their

game characters while lying in the scanner and the argument of differing methodologies mentioned above comes to bear.

The same conclusion should also be drawn with regard to the other studies cited, with respect to the activation of regions belonging to the self-reference network (Herold et al., 2016; Meshi et al., 2016). The sparsity of empirical evidence was reflected in the result as the comparison between the models for the SHAP-values of the self-reference network revealed the lowest *t*-statistics and the mean difference pointed into the direction opposite to the expected direction. The current study was not able to reveal the assumed patterns based on the complexity of the disease and self-related concepts. Given the sparse theoretical background and highly individual expressions of self-identity, self-perception, and self-esteem, those constructs and their association with addiction must be investigated more in detail before any attempt to map them neurologically. Based on these considerations, the idea of disturbed self-identity and altered self-perception in Internet addicted individuals should not be rejected, rather refined.

4.2.3 Attention

The incorporation of the attention network into the investigation was mainly due to the strong relationship between IA and ADHD. Because the dataset's authors excluded participants with a history of psychiatric diseases requiring inpatient treatment and the intake of centrally acting drugs, which include standard ADHD treatment such as Ritalin (Mendes et al., 2019), it is reasonable to assume that a large proportion of individuals with ADHD were excluded from the sample. Therefore, in the first line Internet addicted individuals were excluded and secondly those which were included would be mostly free of ADHD (for further limitations based on the sampling method see 4.3 *Limitations*). This could imply, that the Internet addicted individuals in the current sample did not belong to the subsample which developed the disease due to ADHD-like neuronal alterations, making it more difficult for the model to find the expected patterns in the sample. For those instances, which exhibited this pattern of relation, the assumption of lacking striatal dopamine as a consequence of ADHD and an antecedent of IA (Y. Wang et al., 2019; Yen et al., 2009) should hold true but likewise in the reward network this consideration could not be found in this study.

In another study it was found that individuals with IA showed increased FC within the DMN (Y. Wang et al., 2019). According to this finding, one FC out of DMN-ROIs was identified to be one of the most important features for the original model (vmPFC_PCC, $M = .016$). However, the same FC was listed as one of the most important in the shuffled model too with a nearly identical average importance ($M = .009$). Therefore, it is reasonable to concluded that this scoring was created randomly and has no explanation value. Unlike in the study of Y. Wang et al. (2019), no altered FC

between ROIs of the DMN was observed and the ROIs which were incorporated into their visual attention network were not investigated in this study.

Surprisingly, the FC with the cerebellum did not show any importance. This structure is mentioned frequently in the addiction literature based on its role for motor- and emotion-processing (Moulton et al., 2014). However, this structure, which plays a nearly ubiquitous role in addiction, has relevant connections also with structures which were not included in the attention network (e.g. insula, PFC and reward structures). The same applies to the PCC, whose altered FC associated with IA reaches to different areas in the whole brain (Ding et al., 2013; J.-T. Sun et al., 2023). Of the proposed connections in those studies, only PCC_precuneus and PCC_cerebellum were tested as well but still did not reach any significant importance. This leads to questioning the exclusive focus on within-network connections. Furthermore, the cerebellum is a relatively large brain region with multiple sub-regions, associated with various tasks (Moulton et al., 2014). Possibly, the decision in the current regions should have been conducted on a higher granularity. This consideration can be backed with the findings by (Ding et al., 2013), who discovered that the connection between PCC and cerebellum to correlate with IA, however in their case they investigated the anterior lobe which has different peak MNI coordinates compared to the ones used here. Another important aspect is that the decision to investigate the cerebellum has been made based on its role in SRAs. Given the relevant differences between IA and SRAs, this could have led to incorrect conclusions as it is also stated, that yet it is unclear, if the structural and functional alterations in the cerebellum are an antecedent of the addiction or a result of pharmacological effects (Moulton et al., 2014). Those, logically cannot be found in IA.

Further the cited studies in the theoretical background all applied different methods such as structural analysis (Loh & Kanai, 2014) and task-activation procedures (Konrad & Eickhoff, 2010). While ADHD and IA are related and attentional processes such as cue-reactivity are found in both forms of addiction, those properties could not be observed within this study, which could be due to the parcellation or the invalid transmission of SRA findings on IA. Based on the strong background of the theory of the involvement of attention in IA, these ideas should not be rejected but a more thorough investigation is required.

4.2.4 Methods

Model failure can further be interpreted in the context of the diverse field of research on the neuronal underpinnings of IA. Besides a wide range of different scales for IA, the neuronal methods vary significantly. Given these disparities in methods, it is not surprising that findings are also diverging, with reviews and meta-analyses only revealing sparse commonalities between studies (J.-T. Sun et al., 2023; Wadsley & Ihssen, 2023). Furthermore, the field focusing on the neuronal

underpinnings of IA and SMU is narrow in terms of publication numbers, and the findings of those published studies are limited by small sample sizes and a lack of longitudinal data, and as previously stated, interpreting the results is difficult due to different methodologies (Montag, 2019). Most research focuses on single ROIs in structural MRI or rather task-based fMRI settings, whereas this study was conducted with rsFC. That in this study previous findings with different foci were concluded, relied on the mechanism, that functional alterations can be observed before structural alterations. However, the transmission of these findings to the assumption that connectivity between the observed ROIs varies to the same extent depending on IA severity may be incorrect, and thus no neuronal feature derived from the literature could be identified as relevant in this study. As this study did not investigate single activation but rather the co-(de)activation of ROIs, the comparative information value is limited.

While most predictive modeling studies use rs-fMRI variables as features there is a call for the investigation of task-based MRI as activation and effect strengths of ROI activation are assumed be larger when participants are faced with actual stimuli (Greene et al., 2018). Given the low overall FC average, task-based fMRI would increase distinguishable patterns, so it may be effective in detecting IA-related alterations. This method could be used to observe alterations in reactions of participants to Internet relevant cues such as message notifications, follower numbers, likes, newsfeeds etc. This consideration is supported as machine learning analysis with task-based fMRI led to better accuracy than structural MRI, such that possibly it would exceed the performance of rsFC-analyses as well (Rashid & Calhoun, 2020).

Another transmission which could have led to the incorrect conclusions concerns the findings for SRAs and nSRAs. As previously stated, there are key characteristics, which are shared by both forms and the argumentation for the involved networks has been derived from literature mapping specific regions to SRAs. Now, that the proposed hypotheses have been rejected, it is necessary to consider to which extent the findings for both forms can be attributed interchangeably. For example it was found that in risk-taking there are commonalities as well as distinctions when it comes to neuronal correlates for SRAs and nSRAs (Hüpen et al., 2023). These findings highlight the necessity for more fine-grained analyses about the extent of similarity between those constructs such that the transfer of findings from one form to the other can be better informed. One major difference between both diseases is that in SRAs structural and functional changes in the brain can be attributed to pharmacological effects. As those substances are capable of passing the blood-brain barrier, they have a more direct influence on the brain (Volkow & Boyle, 2018). While the behavioral characteristics are overlapping, this difference is particularly important for the current study. Furthermore, even though ML for biomarker investigation is a fast-growing field, dependencies are understudied diseases within this framework. In addition, the current studies mainly focus on SRAs

(Rashid & Calhoun, 2020), such that there are to the best knowledge of the author no reliable results for nSRAs yet.

To this end, the results of this study are partially in line with the literature, highlighting the sparsity of evidence, the variety of findings and the difficulty of revealing overlaps and drawing a coherent picture. However, evidence for the involvement for every present region was outlined in the theoretical background. Therefore, it must be noted that the failed investigation of any relevant feature could be a sign of publication bias, as it is common sense, that almost no null-results are published. Therefore, it could be that similar studies have been conducted with the same results, but they were not published due to the null findings. If those would be publicly available, a more valid picture about the true relationships could be drawn and the current results would be less surprising. However, methodological limitations biasing the deviating results are present and will be outlined in the following paragraph.

4.3 Limitations

Besides the explanation for failed prediction based on the theoretical background also methodological limitations could have caused the unexpected current results. As those are only as reliable as the chosen methodology, considerations about current pitfalls must be made to understand possibly limited interpretability of findings.

The hypothesis testing strategy may have prevented the networks from becoming significant contributors to the model. It was not focused on single features, as networks SHAP-values were computed as the average sum of single features, including also negative SHAP-values, which may have removed the effect of relevant features. This means that some individual features, that performed particularly bad may have cancelled out the power of stronger features. In line with this, Figure 10A shows that the spread of SHAP-values for the original model was higher with an *SD* twice as large as for the shuffled version, resulting in a higher density for stronger absolute values on each side. This is explainable given the logic of the null-model as features in the shuffled condition could by logic not have any meaningful influence on the model and thus the SHAP-scores did not spread as much as in the original model. However, a descriptive comparison of SHAP-scores for each model revealed no significant differences in the values of the individual FC-features and the scores for the original model are only interpretable in comparison with shuffled scores as it needs to be estimated, how much the model is learning from random noise. The fact that the original model revealed more, and stronger negative SHAP-values indicates learning errors of the model, as learned relations in the train set were contradictory in the test set, deteriorating successful prediction. These learning errors could be caused by two factors: (1) It is possible that this observation is due to highly individual differences in the representation of IA in the brain. A predictor that has been identified as

relevant in the train set but failed completely in the test set indicates relevance for only a subset of participants and is likely to be mediated through other relevant properties that the current model was unable to incorporate. If this assumption is correct, further investigation and integration on possible interacting variables may increase the value of features that failed completely and yielded negative SHAP-values. However, interpreting the hyperparameter settings of high CbT and low PS, the model was already overfit on the training data, so adding more features would exacerbate the problem. (2) Another possible explanation for the learning errors could be, that the model learned 'noise' in the data and overfit on the training set, which refers to patterns that are not meaningful. Arguing for this reason, average feature importance for the different networks changes for both model versions correspondingly. This means, that the network with the highest feature importance for the shuffled model was the same for the original model. Therefore, it is to be concluded, that in this case a large proportion of average SHAP-values must be seen as noise. To account for both possibilities, the sample size must have been larger as adding more features would necessitate a larger sample size due to the importance of the feature-sample ratio. Alternatively, noise could be cancelled out by more data instances because it should vary randomly across different samples, so that the model would not detect irrelevant patterns.

In contrast, based on reviews of predictive modeling of brain and mental disorders, the sample size of this study is larger than that of most studies claiming prediction based on fMRI data. The majority of studies (>50%) had sample sizes lower than 50 and >90% had a maximum sample size of 100 (Poldrack et al., 2020). The numbers were slightly higher in another review, however given an observed mean of 126.7 and a median of 77 subjects in the datasets (Cwiek et al., 2022), the current study exceeds this sample size. Predictive performance suffers with larger sample sizes, indicating uninflated predictive performance in the current study, which partially explains the high *MAE*- and low R^2-values due to the sample size. This mechanism can be understood as certain characteristics which are found in small samples, may not be present in a larger population, so they cannot be generalized to the entire population (Yarkoni & Westfall, 2017). Because the goal of this study was to obtain a generalizable model, it is critical that the current sample size is sufficient, even though this meant identifying no significant features. The mechanism at work in the current model can be seen in the number of predictors with negative SHAP-values, which, as previously stated, indicate patterns found in a subset but were not predictive for the rest of the sample. Predictive accuracy has been reported to range between .20 and .40 for larger sample sizes and a correlation coefficient of r =-.265 between sample size and prediction accuracy (Wu et al., 2023). The conclusions drawn from these reports indicate that brain features have generally low real predictive power for behavioral variables in sufficiently large samples. However, it is possible that increasing the current sample size could lead to even poorer results in this study, if the distribution of the IAT exhibits the observed properties of skewness, representing unequal frequencies for higher IAT-

values. This is particularly important when comparing the current sample size to studies using similar methods. Lower sample sizes would have sufficed if they had obtained better sample characteristics (equal distribution, sufficient number of ill participants). Combining the considerations of the two previous paragraphs, careful consideration on the sample size particularly in relation to the number of features is inevitable. However, if the goal is to draw conclusions about entire populations, large sample sizes are generally necessary to obtain valid effect sizes.

Another methodological limitation that may have contributed to the failure of the model is the attempt to regress IAT on brain features. As stated in *2.3.2 Decisions on Current Model*, predicting continuous variables using regression is more difficult to achieve than classifying disease states (Shen et al., 2017). Continuous measures have a higher variability, particularly the IAT (scores can range from 0 to 100; Young, 1998) necessitating more detailed models as the differences in scores are much more subtle. Even though the scores in this sample range from 18 to 58, there are fewer observations for each IAT-score. Therefore, it is assumed that particularly high IAT values could not be learned because there were fewer examples to learn from for observations with an IAT >40. Most individuals in the current dataset did not use the Internet addictively (95.699%), whereas moderate IA, which is used to determine whether IA is present, was only observed in eight participants (4.301%) and severe IA was entirely absent. This issue becomes even more apparent when considering the CV strategy.

In general, the performance of a predictive model is measured by the ability to generalize on unseen data. However, it should be noted that this approach does not take the overall prevalence of the disorder in the population into account but only the patterns which it learned from in the training set (Craddock et al., 2015). As a result, representative datasets are required and both positive and negative predicted values must be calculated after considering general prevalence. This refers to specificity and sensitivity, both of which were not incorporated into the current strategy. The folds in this work were generated randomly without any attempt to make the test set representative of the training set (i.e. stratifying the sets by age and sex, checking for similar distributions, or weighting cases by frequency of their IAT-score), and the model predicted high frequency values too often without considering single other cases. Thinking of possible splits where most participants with high IAT values were allocated in the test set, it is not surprising that the model failed to predict these values. Furthermore, even if the majority of highly Internet addicted participants were included in the training set, the proportion would not have exceeded 5%, which is insufficient to influence the model significantly. Especially for the investigation of diseases which are uncommon in general population (e.g. GIA 7.02%, Pan et al., 2020), the sampling must be conducted carefully. Therefore, one important rule of predictive modeling has been disregarded (“Be mindful of sample characteristics”; Scheinost et al., 2019). Furthermore, given the current procedure, the distribution of IAT in this study, which is best described as a Poisson, is a limitation. In contrast, a uniform

distribution for the IAT would have been required to accurately model even high IAT cases properly. This limitation is due to the sampling of Mendes et al. (2019) who only included healthy participants removing the possibility of modeling actual diseases but only the approximation to a specific disease threshold. It has already been discussed that predicting mental health states from healthy samples will not be successful (Wu et al., 2023).

The train-test-split used in the current CV-scheme should also be discussed in terms of its applicability and rigidity. To avoid overfitting on the train data, the current study used a conservative testing strategy that included predictive modeling and CV. Predictions were made to independent test-sets, which is considered a valid approach to assess generalizability and it was successful in that regard, as age was found to be a significant predictor of IAT in the descriptive linear model, but not in the predictive modeling approach. This observation demonstrates how important out-of-sample predictions are for estimating true model performance and generalizability. However, the current approach could have been even stricter. Out of sample prediction is considered to be the gold-standard but for fMRI data out of site prediction provides even stronger evidence of successful or rather failed prediction, as the process of data acquisition and processing in MRI contains multiple steps that can affect the model, resulting in overfitting to these settings (Rosenberg & Finn, 2022). Furthermore, in addition to the need for multi-site data, there is a call for multimodal studies in which multiple neuroimaging techniques are used, such as EEG, structural MRI and fMRI (Rashid & Calhoun, 2020). This eliminates the bias of each method and yields more informative and reliable results. A model that aims to make diagnoses must perform consistently across different data collections. Given that the current model is already overfit on the training data, it is reasonable to expect that a stricter testing strategy would result in even worse prediction performance, which must be offset by stricter feature selection, larger sample size, and stronger overfitting regularization. However, it must be noted, that the current study applied a reliable and conservative generalizability approach in the application of CV and the prevention of data leakage. To date it is still common to confound train and test sets by applying feature reduction techniques on the whole sample oftentimes valid train-test-splits are not conducted (Poldrack et al., 2020).

Regarding the goal of diagnosis making based on FC properties in individuals, the quality of the dependent measurement is a significant limitation of this study. Even though prediction failed it should be mentioned that successful prediction only could have provided an indication for future studies. Furthermore, for clinical utility making diagnoses for a specific disease is insufficient; comorbidities and disease history must also be differentiated (Rashid & Calhoun, 2020). The issue with the IAT, a GIA measure, comes to play, as there is no possibility of informative distinction. The IAT provides information about current conditions but does not consider the history of Internet usage. Although items referring to failed attempts at rehabilitation indicate persistent problematic Internet use, the current scale cannot estimate the duration of dependency. If the neuronal underpinnings of

IA are identified as outcomes of this disease, this information is critical for predicting functional or structural changes in the brain because of the disease. Therefore, the interpretability of the current findings is limited as this study did not follow a longitudinal approach. Furthermore, it is commonly stated that self-report questionnaires lack reliability and validity, which can be applied to the IAT as well. While Cronbach's α has been found to be .889 in a large scale meta-analysis (Frangos et al., 2012) and .91 in the current sample (Mendes et al., 2019), its validity after more than 20 years of digital evolution needs to be proven. Modeling behavioral measures requires high reliability and sensitivity, while other variables, such as performance-based ones are more likely to be replicated (Rosenberg & Finn, 2022). To capture the characteristics of IA with greater quality, measures such as passive monitoring or digital tracking (e.g., number of social media contacts and messages, time spent in front of a screen, money spent in online shopping/ gaming etc.) may be more appropriate for the current approach. Furthermore, it has been widely investigated and demonstrated that the IAT should be divided into subscales such as social networking, online video gaming, pornography, and online shopping addiction (Montag et al., 2015). Thus, this specific scale is overly broad and implies difficulties in mapping brain specific parameters onto a scale that captures a wide range of behaviors, resulting in a lack of specificity. This issue is compounded by the fact that the theoretical deduction of networks in this study focused primarily on characteristics of social media addiction, implying that the investigation of networks role should be considered on a more specific scale.

Another limitation concerning the hypotheses of this study is feature selection. Three networks were chosen and tested for predictive performance. To keep the feature-sample ratio small, only within-network connections were used as features. As a result, predictive power of inter-network connections has not been assessed but it may be relevant for IA. Future studies using a similar approach should investigate potential relevance of inter-network connections while maintaining an appropriate feature-sample ratio. Furthermore, all networks were derived from theoretical rather than empirical suggestions, and expanding those networks with additional ROIs may have resulted in spurious correlation in the FC matrices, as those values do not provide insight into the reason for apparent connectivity. The existence of these networks containing the additional ROIs must be empirically proven as otherwise, neurobiological interpretability is almost impossible (Smith et al., 2013). This represents a significant limitation because no sanity check of the actual existence of the mentioned networks has been performed and the found difference in FC strength may not be sufficient to indicate, if the outlined regions all belong to the same networks. Particularly, since no predefined networks from established atlases were used to better fit the data at hand, an assessment of network validity would have been necessary. This goal might have been missed.

4.4 Conclusion

In conclusion this study provides strong evidence for the non-predictability of the IAT across the specified networks. The main reasons for predictive failure were the transmission of findings from different methodologies onto the current approach of rs-FC functional connectomes, the operationalization and extension of the tested networks mainly by theoretical assumptions, poor quality of the outcome variable, particularly in terms of lacking cases with present IA, an overestimation of the similarity between characteristics of SRAs and nSRAs and inconsistent empirical evidence for underlying neuronal features. Even though specific FCs reported in previous findings were incorporated in this study, they could not be replicated, most likely due to the conservative analysis strategy and the comparatively large sample size, combined with the relatively good health of the current sample. Particularly the quality of the dependent variable was insufficient for the current approach, and it is assumed that replicating it with different metrics to capture IA or, more specifically, SMU would still be worthwhile. The applied modeling methodology was robust, and the findings should be interpreted as a strong indication of the lack of consistent theoretical and empirical evidence for reliable and generalizable neuronal correlates of IA. Concerning the highly variable previous findings, it was demonstrated, that a conservative approach such as predictive modeling incorporating CV was required to validate these, which failed in this approach. Rejection of all hypotheses of this study proves a non-significant contribution of the reward, self-reference and attention network in IA. However, given the strong evidence from previous studies, the assumption of the role of the reward network in IA should not be rejected, and its prediction failure may be attributed to the methodological limitations. These should be considered into future research, with the goal to shed light on the underlying brain-based mechanisms of this concerning disease and development.

5 References

Akram, W., & Kumar, R. (2017). A Study on Positive and Negative Effects of Social Media on Society. *International Journal of Computer Sciences and Engineering*, *5*(10), 351–354.

Appel, H., Gerlach, A. L., & Crusius, J. (2016). The interplay between Facebook use, social comparison, envy, and depression. *Current Opinion in Psychology*, *9*, 44–49. https://doi.org/10.1016/j.copsyc.2015.10.006

Aydm, B., & San, S. V. (2011). Internet addiction among adolescents: The role of self-esteem. *Procedia - Social and Behavioral Sciences*, *15*, 3500–3505. https://doi.org/10.1016/j.sbspro.2011.04.325

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the Type I error rate in independent samples t tests: the power of alternatives and recommendations. *Psychological Methods*, *19*(3), 409–427. https://doi.org/10.1037/met0000014

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27*(31), 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Barry, C. L., McGinty, E. E., Pescosolido, B. A., & Goldman, H. H. (2014). Stigma, discrimination, treatment effectiveness, and policy: Public views about drug addiction and mental illness. *Psychiatric Services*, *65*(10), 1269–1272. https://doi.org/10.1176/appi.ps.201400140

Bisen, S., & Deshpande, Y. (2020). Prevalence, predictors, psychological correlates of internet addiction among college students in India: a comprehensive study. *Anatolian Journal of Psychiatry*, *21*(2), 117–123. https://doi.org/10.5455/apd.47328

Bouckaert, R. R., & Frank, E. (2004). Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In H. Dai, Srikant, Ramakrishnan, Srikant, & C. Zhan (Eds.), *Lecture Notes in Computer Science. Advances in Knowledge Discovery and Data Mining* (Vol. 3056, pp. 3–12). Springer-Verlag. https://doi.org/10.1007/978-3-540-24775-3_3

Brand, M. (2022). Can internet use become addictive? *Science*, *376*(6595), 798–799. https://doi.org/10.1126/science.abn4189

Brand, M., Rumpf, H.-J., Demetrovics, Z., Müller, A., Stark, R., King, D. L., Goudriaan, A. E., Mann, K., Trotzke, P., Fineberg, N. A., Chamberlain, S. R., Kraus, S. W., Wegmann, E., Billieux, J., & Potenza, M. N. (2020). Which conditions should be considered as disorders in the International Classification of Diseases (ICD-11) designation of "other specified disorders due to addictive behaviors"? *Journal of Behavioral Addictions*, *11*(2), 150–159. https://doi.org/10.1556/2006.2020.00035

Brand, M., Young, K. S., & Laier, C. (2014). Prefrontal control and internet addiction: A theoretical model and review of neuropsychological and neuroimaging findings. *Frontiers in Human Neuroscience*, *8*, 375. https://doi.org/10.3389/fnhum.2014.00375

Brand, M., Young, K. S., Laier, C., Wölfling, K., & Potenza, M. N. (2016). Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. *Neuroscience and Biobehavioral Reviews*, *71*, 252–266. https://doi.org/10.1016/j.neubiorev.2016.08.033

Buchman, D. Z., Illes, J., & Reiner, P. B. (2011). The Paradox of Addiction Neuroscience. *Neuroethics*, *4*(2), 65–77. https://doi.org/10.1007/s12152-010-9079-z

Cawley, G. C., & Talbot, N. L. C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *The Journal of Machine Learning Research*, *11*, 2079–2107. https://doi.org/10.5555/1756006.1859921

Chen, N.-K., Chou, Y., Song, A. W., & Madden, D. J. (2009). Measurement of spontaneous signal fluctuations in fMRI: Adult age differences in intrinsic functional connectivity. *Brain Structure and Function*, *213*(6), 571–585. https://doi.org/10.1007/s00429-009-0218-4

Chen, Y.-L., Chen, S.-H., & Gau, S. S.-F. (2015). Adhd and autistic traits, family function, parenting style, and social adjustment for Internet addiction among children and adolescents in Taiwan: A longitudinal study. *Research in Developmental Disabilities*, *39*, 20–31. https://doi.org/10.1016/j.ridd.2014.12.025

Cheng, Y.-S., Ko, H.-C., Sun, C.-K., & Yeh, P.-Y. (2021). The relationship between delay discounting and Internet addiction: A systematic review and meta-analysis. *Addictive Behaviors*, *114*, 106751. https://doi.org/10.1016/j.addbeh.2020.106751

Choi, E. J., Taylor, M. J., Hong, S.-B., Kim, C., Kim, J.-W., McIntyre, R. S., & Yi, S.-H. (2018). Gaming-addicted teens identify more with their cyber-self than their own self: Neural evidence. *Psychiatry Research: Neuroimaging*, *279*, 51–59. https://doi.org/10.1016/j.pscychresns.2018.05.012

Chun, J.-W., Choi, J., Cho, H., Choi, M.-R., Ahn, K.-J., Choi, J.-S., & Kim, D.-J. (2018). Role of Frontostriatal Connectivity in Adolescents With Excessive Smartphone Use. *Frontiers in Psychiatry*, *9*, Article 437. https://doi.org/10.3389/fpsyt.2018.00437

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Coutanche, M. N., & Hallion, L. S. (2019). *Machine Learning for Clinical Psychology and Clinical Neuroscience*. Center for Open Science. https://doi.org/10.31234/osf.io/7zswh

Craddock, R. C., Tungaraza, R. L., & Milham, M. P. (2015). Connectomics and new approaches for analyzing human brain functional connectivity. *GigaScience*, *4*(1), 1–12. https://doi.org/10.1186/s13742-015-0045-x

Cwiek, A., Rajtmajer, S. M., Wyble, B., Honavar, V., Grossner, E., & Hillary, F. G. (2022). Feeding the machine: Challenges to reproducible predictive modeling in resting-state connectomics. *Network Neuroscience*, *6*(1), 29–48. https://doi.org/10.1162/netn_a_00212

Deng, H., Zhou, Y., Wang, L., & Zhang, C. (2021). Ensemble learning for the early prediction of neonatal jaundice with genetic features. BMC Medical Informatics and Decision Making, 21(1), 338. https://doi.org/10.1186/s12911-021-01701-9

Ding, W., Sun, J., Sun, Y [Ya-wen], Zhou, Y., Li, L., Xu, J [Jian-rong], & Du, Y [Ya-song] (2013). Altered default network resting-state functional connectivity in adolescents with Internet gaming addiction. *PloS One*, *8*(3), e59902. https://doi.org/10.1371/journal.pone.0059902

Dong, G.-H., Dong, H., Wang, M., Zhang, J., Zhou, W., Du, X [Xiaoxia], & Potenza, M. N. (2021). Dorsal and ventral striatal functional connectivity shifts play a potential role in internet gaming disorder. *Communications Biology*, *4*(1), Article 866. https://doi.org/10.1038/s42003-021-02395-5

Du, Y [Yuhui], Fu, Z., & Calhoun, V. D. (2018). Classification and Prediction of Brain Disorders Using Functional Connectivity: Promising but Challenging. *Frontiers in Neuroscience*, *12*, 525. https://doi.org/10.3389/fnins.2018.00525

Evren, B., Evren, C., Dalbudak, E., Topcu, M., & Kutlu, N. (2018). Relationship of internet addiction severity with probable ADHD and difficulties in emotion regulation among young adults. *Psychiatry Research*, *269*, 494–500. https://doi.org/10.1016/j.psychres.2018.08.112

First, M. B., Spitzer, R. L., Gibbon Miriam, W., & Janet, B. W. (1997). Structured Clinical Interview for DSM-IV Axis I disorders (SCID I). *New York: Biometric Research Department*. https://cir.nii.ac.jp/crid/1130282271541670912

Frangos, C. C [Constantinos C.], Frangos, C. C [Christos C.], & Sotiropoulos, I. (2012). A meta-analysis of the reliabilty of young's internet addiction test. In L. Gelman, A. Hunter, S. I. Ao, D. W. L. Hukins, & am Korsunsky (Eds.), *World Congress on Engineering: Vol. 2012, Lecture Notes in Engineering and Computer Science* (pp. 368–371). Newswood Limited. https://discovery.ucl.ac.uk/id/eprint/10129833/

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137. https://doi.org/10.1016/j.pneurobio.2019.01.008

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), 2807. https://doi.org/10.1038/s41467-018-04920-3

Greenfield, D. (2012). The Addictive Properties of Internet Usage. In K. S. Young & C. N. de Abreu (Eds.), *Internet Addiction: A handbook and guide to evaluation and treatment* (pp. 133–153). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118013991.ch8

Herold, D., Spengler, S., Sajonz, B., Usnich, T., & Bermpohl, F. (2016). Common and distinct networks for self-referential and social stimulus processing in the human brain. *Brain Structure and Function*, *221*(7), 3475–3485. https://doi.org/10.1007/s00429-015-1113-9

Hong, S.-B., Zalesky, A., Cocchi, L., Fornito, A., Choi, E.-J., Kim, H.-H., Suh, J.-E., Kim, C.-D., Kim, J.-W., & Yi, S.-H. (2013). Decreased functional brain connectivity in adolescents with internet addiction. *PLOS ONE*, *8*(2), e57831. https://doi.org/10.1371/journal.pone.0057831

Howard, J. D., & Kahnt, T. (2021). To be specific: The role of orbitofrontal cortex in signaling reward identity. *Behavioral Neuroscience*, *135*(2), 210–217. https://doi.org/10.1037/bne0000455

Hsieh, K.-Y., Hsiao, R. C., Yang, Y.-H., Lee, K.-H., & Yen, C.-F. (2019). Relationship between Self-Identity Confusion and Internet Addiction among College Students: The Mediating Effects of Psychological Inflexibility and Experiential Avoidance. *International Journal of Environmental Research and Public Health*, *16*(17), 3225. https://doi.org/10.3390/ijerph16173225

Hu, B., Yu, Y., Yan, L.-F., Qi, G.-Q., Wu, D., Li, Y.-T., Shi, A.-P., Liu, C.-X., Shang, Y.-X., Li, Z.-Y., Cui, G.-B., & Wang, W. (2022). Intersubject correlation analysis reveals the plasticity of cerebral functional connectivity in the long-term use of social media. *Human Brain Mapping*, *43*(7), 2262–2275. https://doi.org/10.1002/hbm.25786

Hüpen, P., Habel, U., Votinov, M., Kable, J. W., & Wagels, L. (2023). A Systematic Review on Common and Distinct Neural Correlates of Risk-taking in Substance-related and Non-substance Related Addictions. *Neuropsychology Review*, *33*(2), 492–513. https://doi.org/10.1007/s11065-022-09552-5

Iskender, M., & Akin, A. (2011). Self-Compassion and Internet Addiction. *Turkish Online Journal of Educational Technology - TOJET*, *10*(3), 215–221. https://eric.ed.gov/?id=EJ944970

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*.

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging*, *64*, 101–121. https://doi.org/10.1016/j.mri.2019.05.031

Kim, B., & Im, H.-I. (2019). The role of the dorsal striatum in choice impulsivity. *Annals of the New York Academy of Sciences*, *1451*(1), 92–111. https://doi.org/10.1111/nyas.13961

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, *26*(9), 1011–1013. https://doi.org/10.1038/nbt0908-1011

Ko, C. H., Yen, J. Y., Yen, C. F., Chen, C. S., & Chen, C. C. (2012). The association between Internet addiction and psychiatric disorder: A review of the literature. *European Psychiatry : The Journal of the Association of European Psychiatrists*, *27*(1), 1–8. https://doi.org/10.1016/j.eurpsy.2010.04.011

Konrad, K., & Eickhoff, S. B. (2010). Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human Brain Mapping*, *31*(6), 904–916. https://doi.org/10.1002/hbm.21058

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: The cornerstone of modern statistics. *Korean Journal of Anesthesiology*, *70*(2), 144–156. https://doi.org/10.4097/kjae.2017.70.2.144

Liu, J [Jun], Gao, X.-P., Osunde, I., Li, X [Xin], Zhou, S.-K., Zheng, H.-R., & Li, L.-J. (2010). Increased regional homogeneity in internet addiction disorder: A resting state functional magnetic resonance imaging study. *Chinese Medical Journal*, *123*(14), 1904–1908. https://pubmed.ncbi.nlm.nih.gov/20819576/

Loh, K. K., & Kanai, R. (2014). Higher media multi-tasking activity is associated with smaller gray-matter density in the anterior cingulate cortex. *PloS One*, *9*(9), e106698. https://doi.org/10.1371/journal.pone.0106698

Long, J. S., & Ervin, L. H. (2012). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, *54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

Lozano-Blasco, R., Robres, A. Q., & Sánchez, A. S. (2022). Internet addiction in young adults: A meta-analysis and systematic review. *Computers in Human Behavior*, *130*, 107201. https://doi.org/10.1016/j.chb.2022.107201

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions.* https://doi.org/10.48550/ARXIV.1705.07874

Marengo, D., Longobardi, C., Fabris, M. A., & Settanni, M. (2018). Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. *Computers in Human Behavior*, *82*, 63–69. https://doi.org/10.1016/j.chb.2018.01.003

Marin, M. G., Nuñez, X., & Almeida, R. M. M. de (2021). Internet Addiction and Attention in Adolescents: A Systematic Review. *Cyberpsychology, Behavior and Social Networking*, *24*(4), 237–249. https://doi.org/10.1089/cyber.2019.0698

*Mendes, N., Oligschläger, S., Lauckner, M. E., Golchert, J., Huntenburg, J. M., Falkiewicz, M., Ellamil, M., Krause, S., Baczkowski, B. M., Cozatl, R., Osoianu, A., Kumral, D., Pool, J., Golz, L., Dreyer, M., Haueis, P., Jost, R., Kramarenko, Y., Engen, H., . . . Margulies, D. S.

(2019). A functional connectome phenotyping dataset including cognitive state and personality measures. *Scientific Data*, *6*(1), 1–19. https://doi.org/10.1038/sdata.2018.307

Meshi, D., Mamerow, L., Kirilina, E., Morawetz, C., Margulies, D. S., & Heekeren, H. R. (2016). Sharing self-related information is associated with intrinsic functional connectivity of cortical midline brain regions. *Scientific Reports*, *6*(1), Article 22491. https://doi.org/10.1038/srep22491

Meshi, D., Morawetz, C., & Heekeren, H. R. (2013). Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use. *Frontiers in Human Neuroscience*, *7*, 439. https://doi.org/10.3389/fnhum.2013.00439

*Meshi, D., Tamir, D. I., & Heekeren, H. R. (2015). The Emerging Neuroscience of Social Media. *Trends in Cognitive Sciences*, *19*(12), 771–782. https://doi.org/10.1016/j.tics.2015.09.004

Montag, C. (2019). The Neuroscience of Smartphone/Social Media Usage and the Growing Need to Include Methods from 'Psychoinformatics'. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience* (pp. 275–283). Springer, Cham. https://doi.org/10.1007/978-3-030-01087-4_32

Montag, C., Bey, K., Sha, P., Li, M [Mei], Chen, Y.-F., Liu, W.-Y., Zhu, Y.-K., Li, C.-B., Markett, S., Keiper, J., & Reuter, M. (2015). Is it meaningful to distinguish between generalized and specific Internet addiction? Evidence from a cross-cultural study from Germany, Sweden, Taiwan and China. *Asia-Pacific Psychiatry*, *7*(1), 20–26. https://doi.org/10.1111/appy.12122

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories. *International Journal of Environmental Research and Public Health*, *16*(14), 2612. https://doi.org/10.3390/ijerph16142612

Montag, C., Marciano, L., Schulz, P. J., & Becker, B. (2023). Unlocking the brain secrets of social media through neuroscience. *Trends in Cognitive Sciences*, *27*(12), 1102–1104. https://doi.org/10.1016/j.tics.2023.09.005

Montag, C., Markowetz, A., Blaszkiewicz, K., Andone, I., Lachmann, B., Sariyska, R., Trendafilov, B., Eibes, M., Kolb, J., Reuter, M., Weber, B., & Markett, S. (2017). Facebook usage on smartphones and gray matter volume of the nucleus accumbens. *Behavioural Brain Research*, *329*, 221–228. https://doi.org/10.1016/j.bbr.2017.04.035

Moon, S. J., Hwang, J. S., Kim, J. Y., Shin, A. L., Bae, S. M., & Kim, J. W. (2018). Psychometric Properties of the Internet Addiction Test: A Systematic Review and Meta-Analysis. *Cyberpsychology, Behavior and Social Networking*, *21*(8), 473–484. https://doi.org/10.1089/cyber.2018.0154

Moulton, E. A., Elman, I., Becerra, L. R., Goldstein, R. Z., & Borsook, D. (2014). The cerebellum and addiction: Insights gained from neuroimaging research. *Addiction Biology*, *19*(3), 317–331. https://doi.org/10.1111/adb.12101

Nadeau, C., & Bengio, Y. (1999). Inference for the generalization error. *Advances in Neural Information Processing Systems*(12).

Namkung, H., Kim, S.-H., & Sawa, A. (2017). The Insula: An Underestimated Brain Area in Clinical Neuroscience, Psychiatry, and Neurology. *Trends in Neurosciences*, *40*(4), 200–207. https://doi.org/10.1016/j.tins.2017.02.002

Naqvi, N. H., & Bechara, A. (2010). The insula and drug addiction: An interoceptive view of pleasure, urges, and decision-making. *Brain Structure and Function*, *214*(5-6), 435–450. https://doi.org/10.1007/s00429-010-0268-7

Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. *Science*, *315*(5811), 531–534. https://doi.org/10.1126/science.1135926

Nesi, J., & Prinstein, M. J. (2015). Using Social Media for Social Comparison and Feedback-Seeking: Gender and Popularity Moderate Associations with Depressive Symptoms. *Journal of Abnormal Child Psychology*, *43*(8), 1427–1438. https://doi.org/10.1007/s10802-015-0020-0

Neyman, C. (2017). A Survey of Addictive Software Design. *Computer Science and Software Engineering*. https://digitalcommons.calpoly.edu/cscsp/111

Nieto-Castanon, A. (2020). *Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*. Hilbert Press. https://doi.org/10.56441/hilbertpress.2207.6598

Nieto-Castanon, A., & Whitfield-Gabrieli, S. (2022). *CONN functional connectivity toolbox: RRID SCR_009550, release 22*. Hilbert Press. https://doi.org/10.56441/hilbertpress.2246.5840

Ortiz-Ospina, E., & Roser, M. (2023). The rise of social media. *Our World in Data*. https://ourworldindata.org/rise-of-social-media?ref=tms

Ostovar, S., Allahyar, N., Aminpoor, H., Moafian, F., Nor, M. B. M., & Griffiths, M. D. (2016). Internet Addiction and its Psychosocial Risks (Depression, Anxiety, Stress and Loneliness) among Iranian Adolescents and Young Adults: A Structural Equation Model in a Cross-Sectional Study. *International Journal of Mental Health and Addiction*, *14*(3), 257–267. https://doi.org/10.1007/s11469-015-9628-0

Özdemir, Y., Kuzucu, Y., & Ak, Ş. (2014). Depression, loneliness and Internet addiction: How important is low self-control? *Computers in Human Behavior*, *34*, 284–290. https://doi.org/10.1016/j.chb.2014.02.009

Pan, Y.-C., Chiu, Y.-C., & Lin, Y.-H. (2020). Systematic review and meta-analysis of epidemiology of internet addiction. *Neuroscience and Biobehavioral Reviews*, *118*, 612–622. https://doi.org/10.1016/j.neubiorev.2020.08.013

*Parsons, T. D. (2017). *Cyberpsychology and the brain: The interaction of neuroscience and affective computing* (First published.). Cambridge University Press.

Peters, S., van der Cruijsen, R., van der Aar, L. P. E., Spaans, J. P., Becht, A. I., & Crone, E. A. (2021). Social media use and the not-so-imaginary audience: Behavioral and neural mechanisms underlying the influence on self-concept. *Developmental Cognitive Neuroscience*, *48*, 100921. https://doi.org/10.1016/j.dcn.2021.100921

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, *77*(5), 534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671

Polikar, R. (2012). Ensemble Learning. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning: Methods and Applications* (pp. 1–34). Springer US. https://doi.org/10.1007/978-1-4419-9326-7_1

Rashid, B., & Calhoun, V. (2020). Towards a brain-based predictome of mental illness. *Human Brain Mapping*, *41*(12), 3468–3535. https://doi.org/10.1002/hbm.25013

Rosenberg, M. D., & Finn, E. S. (2022). How to establish robust brain-behavior relationships without thousands of individuals. *Nature Neuroscience*, *25*(7), 835–837. https://doi.org/10.1038/s41593-022-01110-9

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79. https://doi.org/10.1002/widm.2

Saiphoo, A. N., & Vahedi, Z. (2019). A meta-analytic review of the relationship between social media use and body image disturbance. *Computers in Human Behavior*, *101*, 259–275. https://doi.org/10.1016/j.chb.2019.07.028

Sariyska, R., Reuter, M., Lachmann, B., & Montag, C. (2015). Attention Deficit/Hyperactivity Disorder is a Better Predictor for Problematic Internet use than Depression: Evidence from Germany. *Journal of Addiction Research & Therapy*, *6*(1). https://doi.org/10.4172/2155-6105.1000209

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E., Salehi, M., Gao, S., Shen, X., O'Connor, D., Barron, D. S., Yip, S. W., Rosenberg, M. D., & Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, *193*, 35–45. https://doi.org/10.1016/j.neuroimage.2019.02.057

Schønning, V., Hjetland, G. J., Aarø, L. E., & Skogen, J. C. (2020). Social Media Use and Mental Health and Well-Being Among Adolescents - A Scoping Review. *Frontiers in Psychology*, *11*, 1949. https://doi.org/10.3389/fpsyg.2020.01949

Schulz, M., Mayer, C., Schlemm, E., Frey, B. M., Malherbe, C., Petersen, M., Gallinat, J., Kühn, S., Fiehler, J., Hanning, U., Twerenbold, R., Gerloff, C., Cheng, B., & Thomalla, G. (2022). Association of Age and Structural Brain Changes With Functional Connectivity and Executive Function in a Middle-Aged to Older Population-Based Cohort. *Frontiers in Aging Neuroscience*, *14*, 782738. https://doi.org/10.3389/fnagi.2022.782738

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The Power of the Like in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media. *Psychological Science*, *27*(7), 1027–1035. https://doi.org/10.1177/0956797616645673

Simpson, S. L., Bowman, F. D., & Laurienti, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain*†. *Statistics Surveys*, *7*, 1–36. https://doi.org/10.1214/13-SS103

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., Nichols, T. E., Robinson, E. C., Salimi-Khorshidi, G., Woolrich, M. W., Barch, D. M., Uğurbil, K., & van Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, *17*(12), 666–682. https://doi.org/10.1016/j.tics.2013.09.016

Stănculescu, E., & Griffiths, M. D. (2022). Social media addiction profiles and their antecedents using latent profile analysis: The contribution of social anxiety, gender, and age. *Telematics and Informatics*, *74*, 101879. https://doi.org/10.1016/j.tele.2022.101879

Starcke, K., Antons, S., Trotzke, P., & Brand, M. (2018). Cue-reactivity in behavioral addictions: A meta-analysis and methodological considerations. *Journal of Behavioral Addictions*, *7*(2), 227–238. https://doi.org/10.1556/2006.7.2018.39

Steers, M.-L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing Everyone Else's Highlight Reels: How Facebook Usage is Linked to Depressive Symptoms. *Journal of Social and Clinical Psychology*, *33*(8), 701–731. https://doi.org/10.1521/jscp.2014.33.8.701

Sun, J.-T., Hu, B., Chen, T.-Q., Chen, Z.-H., Shang, Y.-X., Li, Y.-T., Wang, R., & Wang, W. (2023). Internet addiction-induced brain structure and function alterations: A systematic review and

meta-analysis of voxel-based morphometry and resting-state functional connectivity studies. *Brain Imaging and Behavior*, *17*(3), 329–342. https://doi.org/10.1007/s11682-023-00762-w

Sun, Y [Yalin], & Zhang, Y [Yan] (2021). A review of theories and models applied in studies of social media addiction and implications for future research. *Addictive Behaviors*, *114*, 106699. https://doi.org/10.1016/j.addbeh.2020.106699

Tateno, M., Teo, A. R., Shirasaka, T., Tayama, M., Watabe, M., & Kato, T. A. (2016). Internet addiction and self-evaluated attention-deficit hyperactivity disorder traits among Japanese college students. *Psychiatry and Clinical Neurosciences*, *70*(12), 567–572. https://doi.org/10.1111/pcn.12454

Tian, Y., Qin, N., Cao, S., & Gao, F. (2021). Reciprocal associations between shyness, self-esteem, loneliness, depression and Internet addiction in Chinese adolescents. *Addiction Research & Theory*, *29*(2), 98–110. https://doi.org/10.1080/16066359.2020.1755657

Tóth-Király, I., Morin, A. J. S., Hietajärvi, L., & Salmela-Aro, K. (2021). Longitudinal Trajectories, Social and Individual Antecedents, and Outcomes of Problematic Internet Use Among Late Adolescents. *Child Development*, *92*(4), e653-e673. https://doi.org/10.1111/cdev.13525

Turel, O., He, Q., Brevers, D., & Bechara, A. (2018). Delay discounting mediates the association between posterior insular cortex volume and social media addiction symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(4), 694–704. https://doi.org/10.3758/s13415-018-0597-1

Vahedi, Z., & Saiphoo, A. (2018). The association between smartphone use, stress, and anxiety: A meta-analytic review. *Stress and Health*, *34*(3), 347–358. https://doi.org/10.1002/smi.2805

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review. *Social Issues and Policy Review*, *11*(1), 274–302. https://doi.org/10.1111/sipr.12033

Volkow, N. D., & Boyle, M. (2018). Neuroscience of Addiction: Relevance to Prevention and Treatment. *The American Journal of Psychiatry*, *175*(8), 729–740. https://doi.org/10.1176/appi.ajp.2018.17101174

Volkow, N. D., Koob, G. F., & McLellan, A. T. (2016). Neurobiologic Advances from the Brain Disease Model of Addiction. *The New England Journal of Medicine*, *374*(4), 363–371. https://doi.org/10.1056/NEJMra1511480

Volkow, N. D., Michaelides, M., & Baler, R. (2019). The Neuroscience of Drug Reward and Addiction. *Physiological Reviews*, *99*(4), 2115–2140. https://doi.org/10.1152/physrev.00014.2018

Vries, D. A. de, Peter, J., Graaf, H. de, & Nikken, P. (2016). Adolescents' Social Network Site Use, Peer Appearance-Related Feedback, and Body Dissatisfaction: Testing a Mediation Model.

Journal of Youth and Adolescence, *45*(1), 211–224. https://doi.org/10.1007/s10964-015-0266-4

Wadsley, M., & Ihssen, N. (2023). A Systematic Review of Structural and Functional MRI Studies Investigating Social Networking Site Use. *Brain Sciences*, *13*(5), 787. https://doi.org/10.3390/brainsci13050787

Wang, B., Yao, N., Zhou, X., Liu, J [Jian], & Lv, Z. (2017). The association between attention deficit/hyperactivity disorder and internet addiction: A systematic review and meta-analysis. *BMC Psychiatry*, *17*(1), 260. https://doi.org/10.1186/s12888-017-1408-x

Wang, Y., Qin, Y., Li, H., Yao, D., Sun, B., Li, Z., Li, X [Xin], Dai, Y., Wen, C., Zhang, L., Zhang, C [Chenchen], Zhu, T., & Luo, C. (2019). Abnormal Functional Connectivity in Cognitive Control Network, Default Mode Network, and Visual Attention Network in Internet Addiction: A Resting-State fMRI Study. *Frontiers in Neurology*, *10*, 1006. https://doi.org/10.3389/fneur.2019.01006

Weinstein, A., Yaacov, Y., Manning, M., Danon, P., & Weizman, A. (2015). Internet Addiction and Attention Deficit Hyperactivity Disorder Among Schoolchildren. *The Israel Medical Association Journal : IMAJ*, *17*(12), 731–734.

Wu, J., Li, J [Jingwei], Eickhoff, S. B., Scheinost, D., & Genon, S. (2023). The challenges and prospects of brain-based prediction of behaviour. *Nature Human Behaviour*, *7*(8), 1255–1264. https://doi.org/10.1038/s41562-023-01670-1

Yadav, P., Banwari, G., Parmar, C., & Maniar, R. (2013). Internet addiction and its correlates among high school students: A preliminary study from Ahmedabad, India. *Asian Journal of Psychiatry*, *6*(6), 500–505. https://doi.org/10.1016/j.ajp.2013.06.004

Yang, X., Guo, W., Tao, Y., Meng, Y., Wang, H [Hui-yao], Li, X [Xiao-jing], Zhang, Y [Ya-min], Zeng, J., Tang, W., Wang, Q., Deng, W., Zhao, L., Ma, X., Li, M [Ming-li], Xu, J [Jia-jun], Li, J [Jing], Liu, Y., Tang, Z., Du, X [Xiang-dong], . . . Li, T. (2022). A bidirectional association between internet addiction and depression: A large-sample longitudinal study among Chinese university students. *Journal of Affective Disorders*, *299*, 416–424. https://doi.org/10.1016/j.jad.2021.12.013

Yarkoni, T., Poldrack, R. A., Nichols, T. E., van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yen, J.-Y., Ko, C.-H., Yen, C.-F., Wu, H.-Y., & Yang, M.-J. (2007). The comorbid psychiatric symptoms of Internet addiction: Attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, *41*(1), 93–98. https://doi.org/10.1016/j.jadohealth.2007.02.002

Yen, J.-Y., Yen, C.-F., Chen, C.-S., Tang, T.-C., & Ko, C.-H. (2009). The association between adult ADHD symptoms and internet addiction among college students: The gender difference. *CyberPsychology & Behavior*, *12*(2), 187–191. https://doi.org/10.1089/cpb.2008.0113

Yoo, H. J., Cho, S. C., Ha, J., Yune, S. K., Kim, S. J., Hwang, J., Chung, A., Sung, Y. H., & Lyoo, I. K. (2004). Attention deficit hyperactivity symptoms and internet addiction. *Psychiatry and Clinical Neurosciences*, *58*(5), 487–494. https://doi.org/10.1111/j.1440-1819.2004.01290.x

Young, K. S. (1998). Internet Addiction: The Emergence of a New Clinical Disorder. *CyberPsychology & Behavior*, *1*(3), 237–244. https://doi.org/10.1089/cpb.1998.1.237

Zou, Z., Wang, H [Huijun], d'Oleire Uquillas, F., Wang, X., Ding, J., & Chen, H. (2017). Definition of Substance and Non-substance Addiction. In X. Zhang, J. Shi, & R. Tao (Eds.), *Substance and Non-substance Addiction* (Vol. 1010, pp. 21–41). Springer. https://doi.org/10.1007/978-981-10-5562-1_2

6 Supplementary Material

Table S 1.

ROIs and Respective Neurosynth Search-Terms

ROI	Search-Term
vmPFC	ventromedial prefrontal
VS	ventral striatum
VTA	ventral tegmental
OFC	orbitofrontal cortex
DS	dorsal striatum
mPFC	medial prefrontal
PCC	posterior cingulate
insula	insula
precuneus	precuneus
cerebellum	cerebellum
dACC	dacc

Note. Presentation of search-terms used on neurosynth.org for retrieval of coordinates for ROIs

Table S 2.

Linear Regression for Descriptive Results

			95% CI			
Variable	*B*	*SE*	*LL*	*UL*	*t*	*p*
Intercept	35.502	1.016	33.497	37.507	34.932	<.001
Age	-1.1984	0.198	-1.590	-0.807	6.039	<.001
Sex	-0.9179	1.190	-3.265	1.429	-0.772	.441

Note. $F(2,183)=19.0$, $p<.001$, $adjR^2=.163$; *B* = unstandardized coefficient, SE = standard error, *LL* = lower limit, *UL* = upper limit

Table S 3.

Paired One-Sample t-tests of Model Performance - Pre Winsorizing

	Original		Shuffled				*95% CI*		
Score	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
MAE	6.76	0.697	7.515	0.764	-12.913	<.001	-∞	-0.66	0.983
MSE	69.074	14.556	83.52	17.033	-13.416	<.001	-∞	-12.66	0.844
R^2	0.082	0.115	-0.109	0.121	12.943	<.001	.17	∞	1.583

Note. $N = 100$, $df = 99$, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared, *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Table S 4.

Paired One-Sample t-tests of Model Performance - Winsorized

	Original		Shuffled				*95% CI*		
Score	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
MAE	6.76	0.697	7.497	0.708	-12.972	<.001	-∞	-0.64	1.035
MSE	69.03	14.436	82.912	15.148	-13.647	<.001	-∞	-12.19	0.912
R^2	.086	0.104	-0.105	0.103	14.917	<.001	0.17	∞	1.855

Note. $N = 100$, $df = 99$, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared, *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Table S 5.

All Sorted Predictors by SHAP-Values per Model-Version

Original		Shuffled	
Feature	SHAP	Feature	SHAP
age	.069	sex	.034
sex	.024	VS_(R)_VTA	.015
VS_(L)_VTA	.018	PCC_Insula_(R)	.014
Insula_(L)_Insula_(R)	.017	VTA_DS_(R)	.013
vmPFC_dACC	.017	OFC_(R)_DS_(L)	.01
vmPFC_PCC	.016	VS_(L)_DS_(R)	.009
VS_(L)_OFC_(R)	.016	vmPFC_PCC	.009
VTA_OFC_(R)	.013	Insula_(L)_Insula_(R)	.009
vmPFC_VTA	.013	vmPFC_VTA	.009
vmPFC_Cerebellum_(R)	.012	PCC_Cerebellum_(L)	.008
PCC_Cerebellum_(L)	.011	PCC_Cerebellum_(R)	.008
vmPFC_Precuneus	.011	vmPFC_VS_(L)	.008
vmPFC_DS_(R)	.011	OFC_(L)_DS_(L)	.007
vmPFC_OFC_(L)	.01	vmPFC_Precuneus	.007
OFC_(L)_DS_(L)	.008	mPFC_Insula_(R)	.006
vmPFC_OFC_(R)	.009	DS_(L)_DS_(R)	.006
OFC_(R)_DS_(L)	.008	VTA_DS_(L)	.006
mPFC_Insula_(L)	.008	Cerebellum_(L)_Cerebellum_(R)	.006
Cerebellum_(R)_dACC	.008	PCC_Insula_(L)	.006
VTA_DS_(R)	.008	VTA_OFC_(R)	.006
VS_(L)_OFC_(L)	.006	Cerebellum_(R)_dACC	.005
PCC_dACC	.005	VS_(R)_OFC_(L)	.005
mPFC_Insula_(R)	.005	PCC_Precuneus	.005
VS_(R)_OFC_(L)	.005	vmPFC_Cerebellum_(L)	.004
VS_(L)_DS_(R)	.005	vmPFC_DS_(L)	.004
PCC_Precuneus	.004	VS_(R)_DS_(R)	.004
Precuneus_dACC	.004	OFC_(L)_DS_(R)	.003
VS_(R)_DS_(L)	.004	VS_(R)_DS_(L)	.003
Precuneus_Cerebellum_(L)	.004	vmPFC_DS_(R)	.003
VS_(R)_OFC_(R)	.004	mPFC_PCC	.003
VS_(L)_VS_(R)	.003	VS_(L)_OFC_(R)	.002
OFC_(R)_DS_(R)	.003	VS_(L)_OFC_(L)	.002
DS_(L)_DS_(R)	.003	Precuneus_Cerebellum_(R)	.002
VTA_DS_(L)	.003	vmPFC_dACC	.002

vmPFC_VS_(L)	.003	VS_(R)_OFC_(R)	.001
VS_(L)_DS_(L)	.003	age	.001
VS_(R)_VTA	.003	VTA_OFC_(L)	.001
OFC_(L)_DS_(R)	.002	VS_(L)_VTA	.000
Cerebellum_(L)_Cerebellum_(R)	.002	VS_(L)_VS_(R)	.000
PCC_Cerebellum_(R)	.002	PCC_dACC	-.000
vmPFC_VS_(R)	.001	vmPFC_OFC_(L)	-.000
vmPFC_DS_(L)	.001	mPFC_Insula_(L)	-.001
PCC_Insula_(L)	-.001	vmPFC_Cerebellum_(R)	-.001
vmPFC_Cerebellum_(L)	-.003	VS_(L)_DS_(L)	-.002
VTA_OFC_(L)	-.004	Cerebellum_(L)_dACC	-.002
Cerebellum_(L)_dACC	-.005	Precuneus_Cerebellum_(L)	-.002
PCC_Insula_(R)	-.006	OFC_(L)_OFC_(R)	-.004
OFC_(L)_OFC_(R)	-.008	vmPFC_OFC_(R)	-.004
mPFC_PCC	-.01	vmPFC_VS_(R)	-.004
Precuneus_Cerebellum_(R)	-.013	Precuneus_dACC	-.005

Note. Predictors sorted after their predictive performance in both models from high to low

Table S 6.

Paired One-Sample t-tests of Predictive Performance of Networks - Pre-Winsorizing

	Original		Shuffled				*95% CI*		
Network	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
Reward	0.137	0.337	0.10	0.332	0.727	.234	-0.05	∞	0.113
Self-Reference	0.013	0.096	0.037	0.142	-1.483	.929	-0.05	∞	0.192
Attention	0.075	0.146	0.047	0.209	1.135	.13	-0.01	∞	0.157

Note. N = 100, *df* = 99, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared, *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Table S 7.

Paired One-Sample t-tests of Predictive Performance of Networks - Winsorized

	Original		Shuffled				*95% CI*		
Network	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
Reward	0.137	0.337	0.103	0.289	0.713	.239	-.05	∞	0.11
Self-Reference	0.012	0.09	0.035	0.136	-1.495	.931	-.05	∞	0.193
Attention	0.075	0.146	0.047	0.209	1.225	.112	-.01	∞	0.169

Note. N = 100, *df* = 99, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared, *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Table S 8.

Linear Regression of Hyperparameters on MAE

			95% CI			
Variable	*B*	*SE*	*LL*	*UL*	*t*	*p*
Intercept	6.832	0.813	5.218	8.447	8.399	<.001
Cols.Bytree	<0.001	0.001	-0.002	0.002	-0.018	.986
PathSmooth	-0.077	0.873	-1.809	1.655	-0.089	.93

Note. $F(2,97)=.004$, $p=.996$, $adjR^2=-.021$; *B* = unstandardized coefficient, *SE* = standard error, *LL* = lower limit, *UL* = upper limit

Table S 9.

Paired One-Sample t-tests of Predictive Performance of Control Variables – Pre-Winsorizing

	Original		Shuffled				*95% CI*		
Network	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
Age	0.069	0.569	0.001	0.077	1.172	.122	-0.03	∞	0.166
Self-Reference	0.024	0.063	0.034	0.097	-0.895	.813	-0.03	∞	0.124

Note. *N* = 100, *df* = 99, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Table S 10.

Paired One-Sample t-tests of Predictive Performance of Control Variables – Winsorized

	Original		Shuffled				*95% CI*		
Network	*M*	*SD*	*M*	*SD*	*t*	*p*	*LL*	*UL*	*d*
Age	0.069	0.569	0.001	0.077	1.312	.096	-0.02	∞	0.186
Self-Reference	0.024	0.063	0.034	0.097	-0.773	.779	-0.02	∞	0.108

Note. *N* = 100, *df* = 99, MAE = Mean Absolute Error, MSE = Mean Squared Error, R^2 = r-squared, *M* = mean, *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit, *d* = Cohens' d

Figure S 1.

Linearity Check Age-IAT

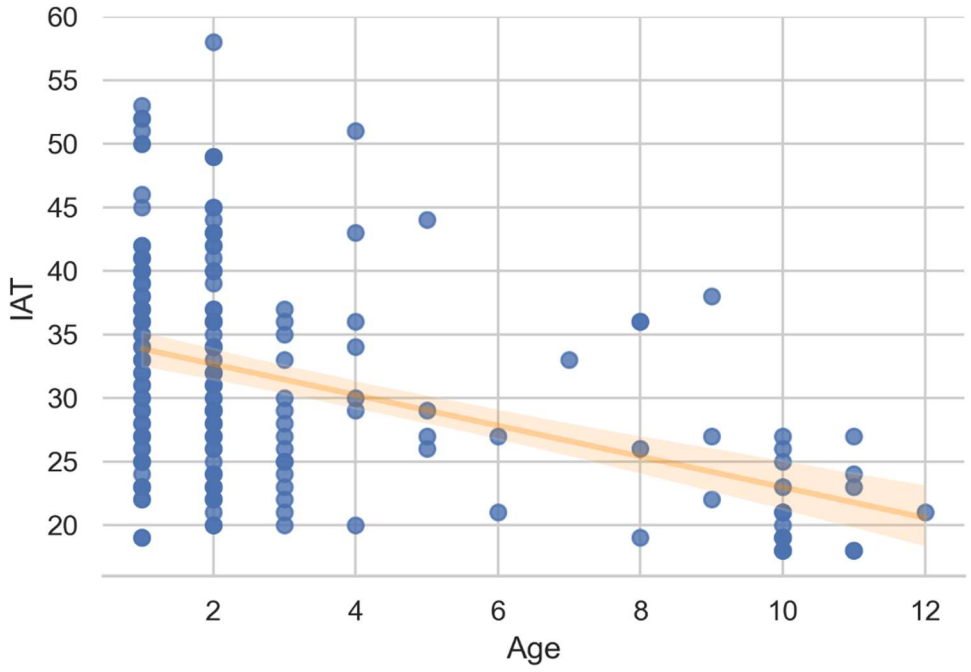


Figure S 2.

QQ-Plot Descriptive Linear Model

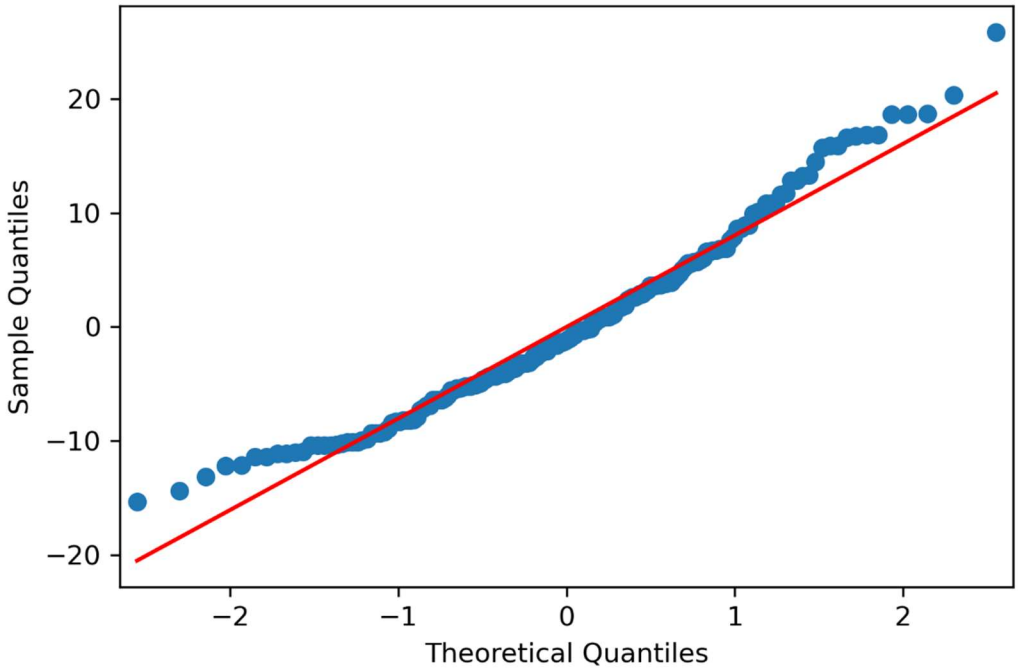


Figure S 4.

Linearity Check Cols.-MAE

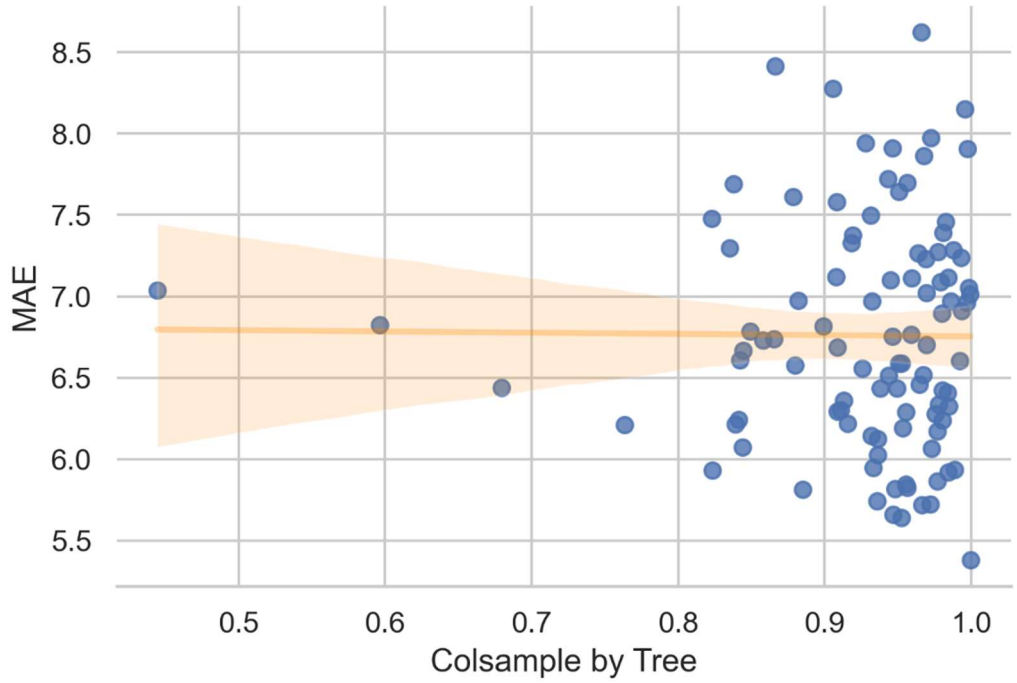


Figure S 3.

Linearity Check PathSmooth-MAE

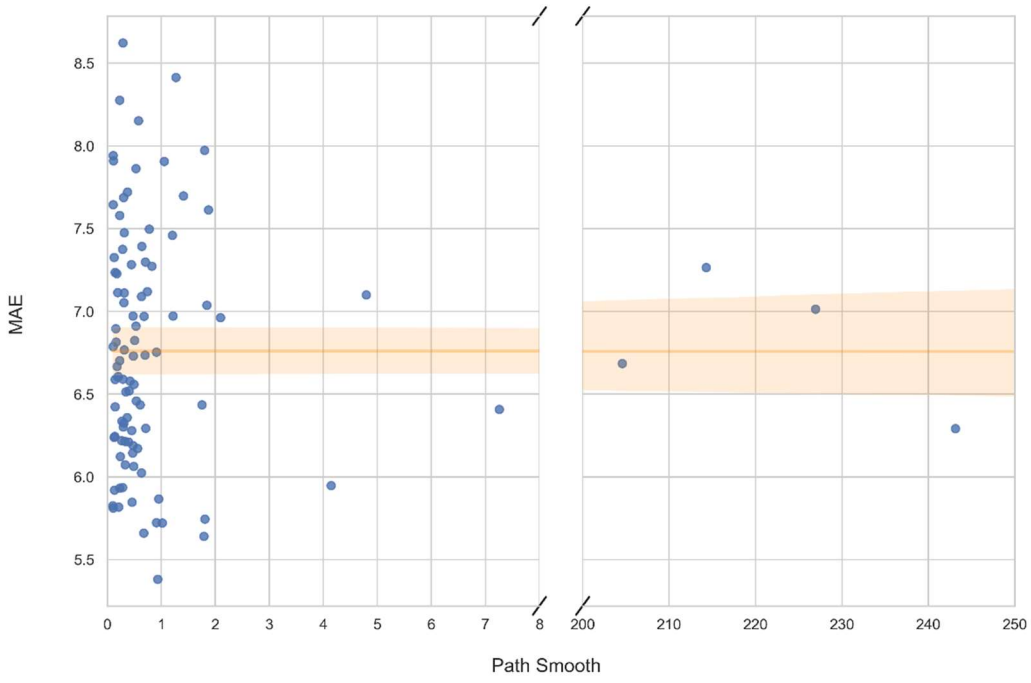
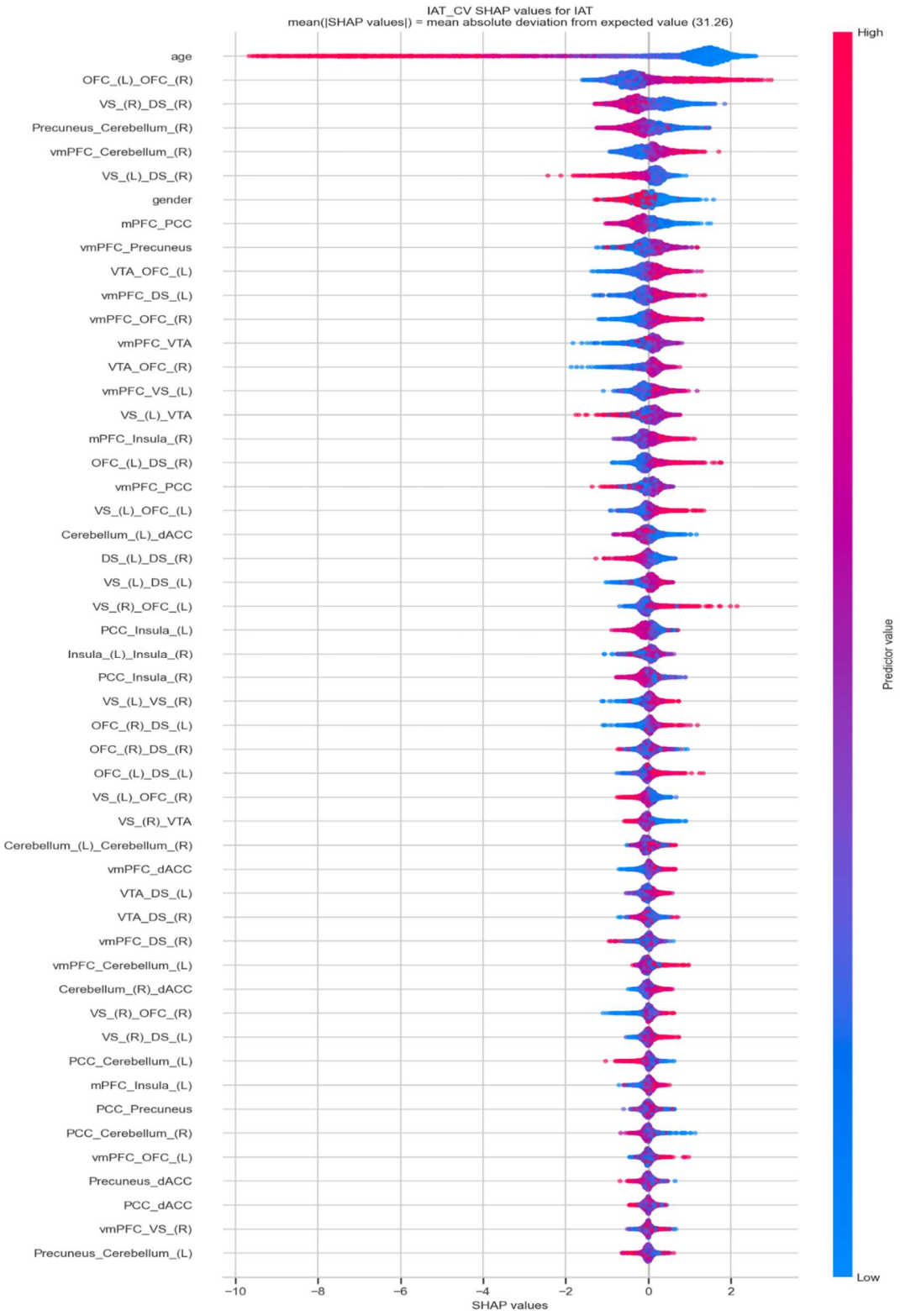


Figure S 5.

Beeswarm Plot



Note. Indicating relationship between feature value and and feature importance. High SHAP-values = high importance.

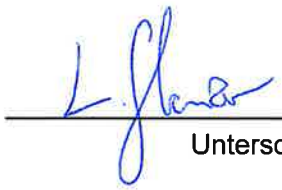
universität
innsbruck

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Magister-/Master-/Diplomarbeit/Dissertation eingereicht.

29.04.2024
Datum



Unterschrift