COMENIUS UNIVERSITY BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

UNIVERZITA
KOMENSKÉHO
V BRATISLAVE

COMPUTATIONAL MODELS OF MENTAL HEALTH DISORDERS:

A REVIEW

Master Thesis

Study program:	Cognitive Science
Field of study:	2503 Cognitive Science
Supervising department:	Department of Applied Informatics
Supervisor:	prof. RNDr. Ľubica Beňušková, PhD.

Bratislava, 2025 **MA, Milica Kiš**

Declaration

I hereby declare that I elaborated this diploma thesis independently using cited literature.

Bratislava, 2025 **Signature**

....

Acknowledgements

I would like to express my gratitude to my supervisor, prof. RNDr. Ľubica Beňušková, PhD, for all her support, patience and invaluable guidance in the writing of this thesis.

Abstract

Computational models are becoming an indispensable tool for studying and understanding the underlying mechanisms of various scientific phenomena. Thus, in recent years, the complexity of diagnosing and treating mental health disorders has proven to be an exciting challenge for the interdisciplinary field of computational psychiatry. In this thesis, an extensive literature search and review of articles, books and lectures from the field of computational psychiatry was used to provide a systematic review of computational approaches applied to mental health disorders and identify possible research gaps.

This critical review has three main aims: (1) to provide a detailed and up-to-date review of computational models used in computational psychiatry (data-driven, theory-driven and combined approaches), as well as alternative approaches to traditional classification systems; (2) to offer a systematic overview of all data types used in computational psychiatry (i.e. neuroimaging, genetic, digital, and behavioral data) and highlight possible problems in their collection, implementation into computational models and validation; (3) to synthesize findings from the application of these approaches to depression and to identify possible research gaps and understudied phenomena. Research into the application of computational models to the study of depression have revealed so far that they are well-suited to investigating different facets and mechanisms of the disorder, while machine learning models show great promise for tackling practical issues in psychiatry (e.g., diagnosis, prognosis, and treatment selection).

In conclusion, the inherently interdisciplinary approach of the field of computational psychiatry should enable moving away from the traditional symptom-based categorization of mental health disorders and provide findings that are more useful for translational psychiatry, thus advancing a more individualized, efficient and optimal approach in clinical practice.

Keywords: computational psychiatry, mental health disorders, depression

Abstrakt

Výpočtové modely sa stávajú nenahraditeľným nástrojom na štúdium a pochopenie základných mechanizmov rôznych vedeckých javov. V posledných rokoch sa tak zložitosť diagnostiky a liečby duševných porúch ukázala ako vzrušujúca výzva pre interdisciplinárnu oblasť počítačovej psychiatrie. V tejto práci bola použitá rozsiahla rešerš literatúry a prehľad článkov, kníh a prednášok z oblasti počítačovej psychiatrie na poskytnutie systematického prehľadu prístupov aplikovaných na duševné poruchy a identifikáciu možných medzier vo výskume.

Tento kritický prehľad má tri hlavné ciele: (1) poskytnúť podrobný a aktuálny prehľad výpočtových modelov používaných vo výpočtovej psychiatrii (prístupy založené na dátach, prístupy založené na teórii a kombinované prístupy), ako aj alternatívne prístupy k tradičným klasifikačným systémom; (2) ponúknuť systematický prehľad všetkých typov dát používaných vo výpočtovej psychiatrii (neurozobrazovacie, genetické, digitálne a behaviorálne dáta) a poukázať na možné problémy pri ich zhromažďovaní, implementácii do výpočtových modelov a validácii; (3) syntetizovať zistenia z aplikácie týchto prístupov na depresiu a identifikovať možné medzery vo výskume a nedostatočne preskúmané javy. Výskum aplikácie výpočtových modelov na štúdium depresie odhalil, že sú vhodné na skúmanie rôznych aspektov a mechanizmov tejto poruchy, pričom modely strojového učenia sú veľmi sľubné na riešenie praktických otázok v psychiatrii (napr. diagnostika, prognóza, a výber liečby).

Na záver, interdisciplinárny prístup, ktorý je neodmysliteľnou súčasťou oblasti počítačovej psychiatrie, by mal umožniť odklon od tradičnej kategorizácie duševných porúch založenej na symptómoch a poskytnúť poznatky, ktoré sú užitočnejšie pre translačnú psychiatriu, čím sa posunie individualizovanejší, účinnejší a optimálnejší prístup v klinickej praxi.

Kľúčové slová: výpočtová psychiatria, poruchy duševného zdravia, depresia

Contents

List of Figures and Tables

Figures

Tables

Abbreviations

5-HT 5-hydroxytryptamine
ADHD attention deficit hyperactivity disorder
AN affective network
ANN artificial neural networks
ANS autonomic nervous system
APA American Psychiatric Association
APS attenuated psychosis syndrome
ASD autism spectrum disorders
ATR antidepressant treatment response
BDT Bayesian decision theory
CBT cognitive behavioral therapy
CCA canonical correlation analysis
CCN cognitive control network
CHR clinical high-risk
CP computational psychiatry
CSF cerebrospinal fluid
CST clinical support tools
CSTC cortico-striato-thalamo-cortical
DALY Disability Adjusted Life Years
DDM drift diffusion model
DL deep learning
DLPFC dorsolateral prefrontal cortex
DMN default mode network
DMPFC dorsomedial prefrontal cortex
DSM-5-TR Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, Text Revision
DTI diffusion tensor imaging
EHR electronic health record
EMA ecological momentary assessments
FEP free energy principle
FL federated learning
fMRI functional magnetic resonance imaging
GABA gamma-aminobutyric acid
GAD generalized anxiety disorder
GBD Global Burden of Disease
GWAS Genome-wide association studies
HAMD Hamilton Depression Rating Scale
HPA hypothalamic-pituitary-adrenal (axis)
ICA independent component analysis
ICD-11 International Classification of Diseases
IGT Iowa gambling task
IHME Institute of Health Metrics and Evaluation
IPT interpersonal therapy
LOO-CV leave one out-cross validation
MAO(A) monoamine oxidase (A)

MDD major depressive disorder
MEG magnetoencephalography
MHD mental health disorder
ML machine learning
NIMH National Institute of Mental Health (US)
NLP natural language processing
OCD obsessive-compulsive disorder
OFC orbitofrontal cortex
PCA principal component analysis
PCR principal component regression
PET positron emission tomography
PFC prefrontal cortex
PRS polygenic risk score
qEEG quantitative electroencephalography
RDoC Research Domain Criteria
rEEG referenced electroencephalography
RL reinforcement learning
RN reward network
RPE reward prediction error
ROD recent onset depression
ROI region of interest
(r)TMS (repetitive) transcranial magnetic stimulation
sgACC subgenual anterior cingulate cortex
sMRI structural magnetic resonance imaging
SNRI selective norepinephrine reuptake inhibitors
SSRI selective serotonin reuptake inhibitors
TAU treatment-as-usual
TD temporal difference
TCA tricyclic antidepressants
TRD treatment-resistant depression
VMPFC ventromedial prefrontal cortex
WHO World Health Organisation
XAI explainable artificial intelligence

1 Introduction

1.1 Overview of the prevalence and impact of mental health disorders worldwide

To grasp the influence of mental health disorders in general population, it is necessary to look at ***prevalence*** of particular mental health disorders (i.e. the share of people affected in a population), ***incidence*** (i.e. number of new cases) or some other statistical indicators, such as data from GBD (Global Burden of Diseases, Injuries and Risk Factors)[1] study (e.g., pertaining to the impact of the DALY metric[2]).

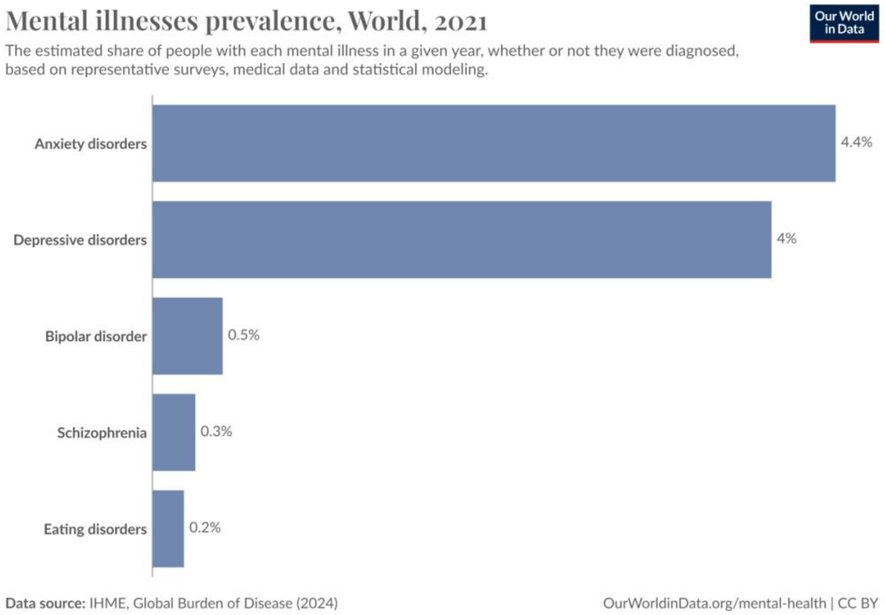
Although, ideally, we would like to rely solely on official medical records (i.e. clinical data), these estimations have to be made by taking into account some additional factors. Therefore, in some instances, the estimated figures might be even higher. For example, mental health disorders are usually underreported, due to social stigma or fear of discrimination. Also, the lack of awareness of the illness might prevent people from asking for professional help. In some cases, the access to proper healthcare facilities is simply lacking. Matters are further complicated in cases of undiagnosed or misdiagnosed disorders or multiple diagnoses (comorbidity).

To get a complete picture of the impact of mental health disorders in overall population, ideally, we should also have longitudinal data at our disposal, since mental health disorders are often characterized by their recurring nature, resistance to treatments or the early age at which they first appear, and thus may remain undetected over longer periods of time.

As we can see from Figure 1.1 and Figure 1.2, depressive and anxiety disorders, schizophrenia, bipolar and eating disorders are the five most common mental health disorders, out of which depressive and anxiety disorders (which are classified as mild) constitute the largest

[1] The latest source of data is from 2021, and it has been collected since 1990 by the Institute of Health Metrics and Evaluation (IHME).

[2] DALYs (Disability Adjusted Life Years) represent the sum of mortality and morbidity and are a metric by which researchers from GBD study measure the "burden of disease", which corresponds to one year of loss of good health either due to premature death or disease or disability (Roser et al., 2024). This is certainly a more informative metric than, for example, only taking into consideration the suicide rates (i.e. mortality) from a particular disorder. However, alarming rise in suicide rates in a particular population (e.g., teenagers and young adults) may point to a need for raising awareness or implementing certain prevention strategies.



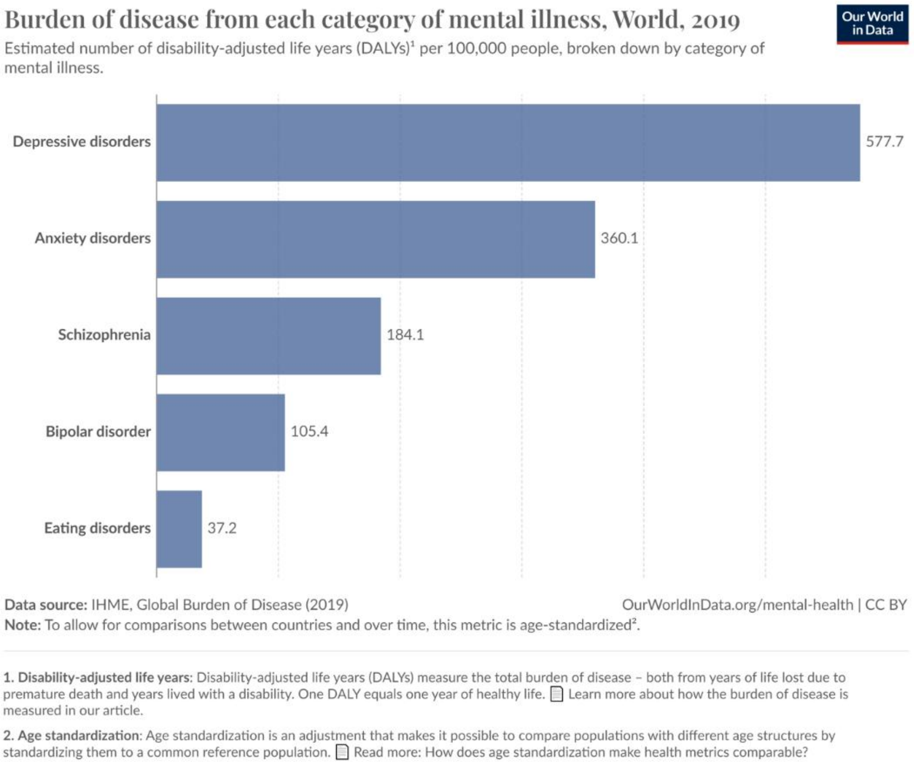
Mental illnesses prevalence, World, 2021

The estimated share of people with each mental illness in a given year, whether or not they were diagnosed, based on representative surveys, medical data and statistical modeling.

Data source: IHME, Global Burden of Disease (2024)

OurWorldinData.org/mental-health | CC BY

Figure 1.1 Prevalence of mental illnesses worldwide (Source: https://ourworldindata.org/grapher/mental-illnesses-prevalence).



Burden of disease from each category of mental illness, World, 2019

Estimated number of disability-adjusted life years (DALYs)[1] per 100,000 people, broken down by category of mental illness.

Data source: IHME, Global Burden of Disease (2019)

OurWorldInData.org/mental-health | CC BY

Note: To allow for comparisons between countries and over time, this metric is age-standardized[2].

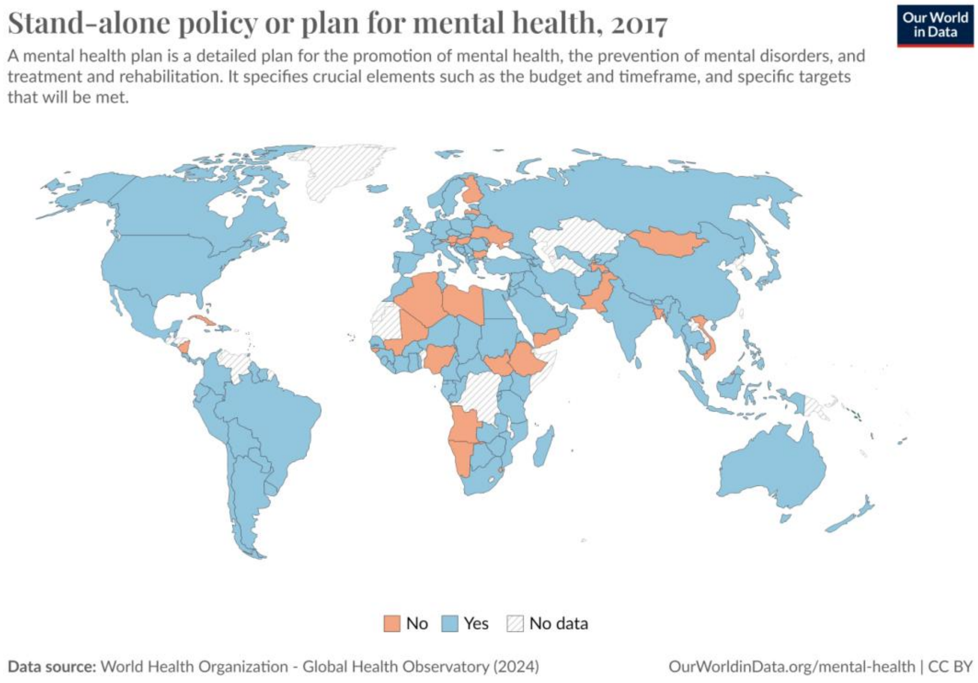
[1] **Disability-adjusted life years:** Disability-adjusted life years (DALYs) measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability. One DALY equals one year of healthy life. 🗎 Learn more about how the burden of disease is measured in our article.

[2] **Age standardization:** Age standardization is an adjustment that makes it possible to compare populations with different age structures by standardizing them to a common reference population. 🗎 Read more: How does age standardization make health metrics comparable?

Figure 1.2 Global burden of disease from each category of mental illness (Source: https://ourworldindata.org/grapher/burden-disease-from-each-mental-illness).

share, and which recorded a rise from 2019 onwards (with COVID-19 pandemic cited as the most likely contributing factor[3]). Although depression and anxiety are classified as mild, they also have high prevalence, thereby costing the global economy $1 trillion in lost productivity each year, with a cost projected to rise to $6 trillion by 2030, while the depression is cited as the leading cause of disability worldwide (The Lancet Global Health, 2020).

The WHO Mental Health Atlas initiative from 2017 requested that countries estimate their government's total spending on mental health. They found that, on average, mental health expenditure accounted for less than 2% of government budgets for health (The Lancet Global Health, 2020). According to the findings of Rajkumar (2022), government spending on mental health was below 1% of health expenditure in 24.4% of the 78 countries studied. However, Figure 1.3 shows that majority of countries have some kind of policy or plan in place to address the issue of mental health.



Stand-alone policy or plan for mental health, 2017

A mental health plan is a detailed plan for the promotion of mental health, the prevention of mental disorders, and treatment and rehabilitation. It specifies crucial elements such as the budget and timeframe, and specific targets that will be met.

Data source: World Health Organization - Global Health Observatory (2024)

OurWorldinData.org/mental-health | CC BY

Figure 1.3 Stand-alone policy or plan for mental health, 2017 (Source: https://ourworldindata.org/grapher/stand-alone-policy-or-plan-for-mental-health).

[3] WHO estimates that COVID-19 has directly or indirectly contributed to an additional 53.2 million cases of depression and 76.2 million cases of anxiety, an increase of 28% and 26% in prevalence, respectively, since the start of the pandemic (Kämpfen et al., 2020).

These plans generally focus on raising awareness and prevention of mental health disorders, encouraging population to seek adequate professional help, and outlining specific strategies with allocation of funds or time limits within which certain targets have to be met.

1.2 Importance of the computational modeling approach for mental health disorders

It may be argued that research and treatment of mental health disorders have been in a relationship of mutual dependency, since one informs the other and vice-versa, which was especially prominent in the early stages of the psychiatric study and practice. For example, cognitive behavioral therapy (CBT) grew from the early 20th century psychological tradition of behaviorism (Watson, 1913; Skinner, 1938) (Seriès, 2020). Therefore, methods such as exposure therapy to treat disorders such as phobias have been developed on the assumption that behavioral response can be “unlearned” by gradual exposure to observed triggers, but uncovering the cause or the neural basis of such maladaptive responses might have still remained unclear. Furthermore, since the treatment was used to eliminate the undesired observed behavior, it seemed that the main objective was achieved, without necessarily understanding the exact mechanism behind it.

The main problem in psychiatry compared to other branches of medicine is that the mechanism or the neural basis of many psychopathologies are still unknown or not completely understood, due to the absence of specific (neuro)biological bases or biomarkers. The advance of neuroimaging techniques and genetic studies marked a significant leap towards better understanding of etiology of the disorders, whether by offering insight into the neural mechanisms or into the influence of heritability on potential development of the disorder (risk factors). At the cellular or molecular level, the knowledge about neurotransmitters and neuromodulators (serotonin, dopamine, GABA) enabled targeting some of the imbalances that are characteristic for certain disorders, with the use of medication.

Apart from psychotherapeutic treatments (based on traditions of psychodynamic or behavioral theory), other options in psychiatry include pharmacology and more recently, brain stimulation techniques. Psychotherapeutic and pharmacological interventions[4] in use today

[4] Psychotherapeutic approaches include, for example, analytical psychotherapy, cognitive behavioral therapy (CBT) or interpersonal therapy (IPT), while pharmacological treatments include chlorpromazine and other

(the so-called “first line” treatments) have mostly been discovered over 50 years ago (Seriès, 2020). Although they have since undergone some modifications (in terms of better tolerability, but not significantly improved efficacy), recent meta-analyses suggest a ceiling effect in treatment research[5] (Leichsenring et al., 2022). Brain stimulation techniques, which offer alternative treatment in cases where conventional treatment options fail, may target specific brain structures implicated in the disorders, but different protocols used still yield varying degrees of success (with significant between- and within-subject variability).

Therefore, the central problem of psychiatry remains, but it is possible to approach it from different perspectives. The emergence of computational neuroscience, or more precisely, the first successful computational models such as Hodgkin-Huxley model of action potential generation and propagation (Hodgkin & Huxley, 1952) or Hebb’s rules of plasticity and learning (Hebb, 1949), changed the way basic neurobiological processes can be described, validated and simulated. However, only recently it has been proposed that computational models of cognitive function could be used to explain psychopathology. The novelty of the computational approach consists in formalizing the biological structures and mechanisms of the nervous system in terms of ***information processing*** (Seriès, 2020)[6].

Figure 1.4 shows how computational neuroscience, translational modeling and various areas of application thereof (with focus on computational psychiatry) are related to each other. However, as a still predominantly theoretical discipline, computational psychiatry does not seek to offer any novel treatments; it has the aim of ***optimizing*** current treatment options for better response in patients. Also, its ultimate goal is to translate its findings into meaningful interventions in psychiatric practice.

Apart from explaining the cause of the disorders, current challenges for psychiatric practice include better ***classification*** (in case of overlapping symptoms or comorbidity),

typical antipsychotics, lithium for bipolar disorder, tricyclic antidepressants, SSRI/SNRI for depressive disorders, with various modifications to improve their tolerability and alleviate side-effects (Seriès, 2020).

[5] A random effect meta-analytic evaluation of the effect sizes reported by the largest meta-analyses per disorder yielded a standardized mean difference (SMD) of **0.34** (95% CI: 0.26-0.42) for ***psychotherapies*** and **0.36** (95% CI: 0.32-0.41) for ***pharmacotherapies*** compared with treatment-as-usual (TAU) or placebo.

[6] However, the term “computational” in the context of psychiatry may also have another meaning, i.e. ***inferring*** physiological or cognitive processes from measurements of brain activity and behavioral responses, respectively (Stephan & Mathys, 2014).

treatment selection, ***prediction*** of treatment outcomes (e.g., by means of simulations) and tailoring the treatments for ***individual*** patients (e.g., in the domain of precision psychiatry[7]). Computational approach, which is inherently interdisciplinary in the case of computational psychiatry (CP), relies on the knowledge from branches of science such as machine learning (ML) which is a powerful tool for the classification problems. In addition, CP relies on statistical and modeling methods as a tool for prediction, and knowledge of translational medicine and precision psychiatry to find ways of individualizing the treatment options, and therefore has the necessary means to provide answers to some of those challenges.

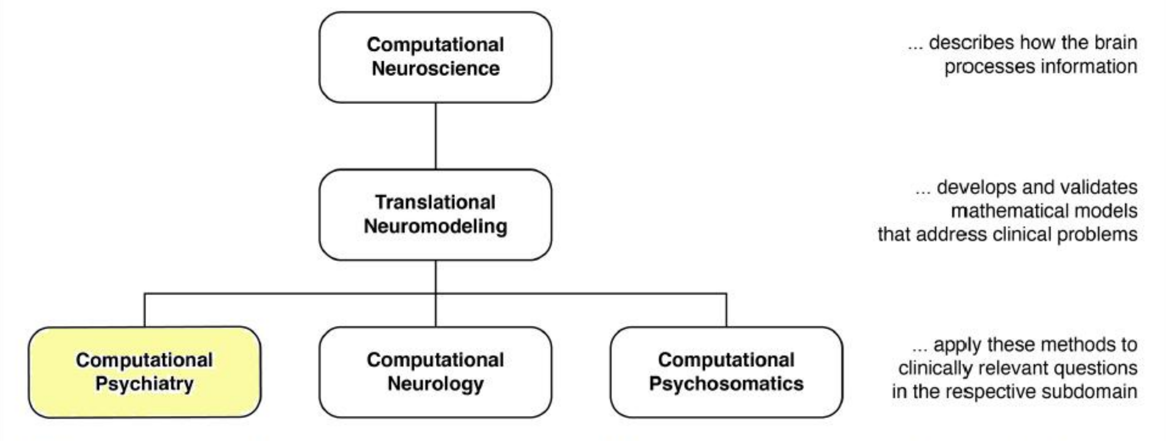

Computational Neuroscience
... describes how the brain processes information
Translational Neuromodeling
... develops and validates mathematical models that address clinical problems
Computational Psychiatry
Computational Neurology
Computational Psychosomatics
... apply these methods to clinically relevant questions in the respective subdomain

FIGURE 1 | Taxonomy for different disciplines in the computational neurosciences and their relation to clinical questions. Translational Neuromodeling (TN) develops and validates mathematical models for addressing clinical problems, whereas Computational Psychiatry (CP), Neurology (CN), and Psychosomatics (CPS) then apply these methods to clinically relevant questions. Reprinted with permission from Frässle et al. (13). Copyright 2018 Wiley.

Figure 1.4 Taxonomy of disciplines in computational neurosciences. Adapted from Frässle et al. (2018).

On the one hand, computational models constrain the number of parameters, processes and potential outcomes and thus provide a predominantly mechanistic representation of a certain disorder, but on the other, they fail to take into the account other dynamic influences, such as environmental or social factors.

[7] Precision medicine applied to psychiatry, or precision psychiatry, is a new, promising approach in psychiatric care, boosted by recent advances in neuroscience, that aims to tailor treatments and interventions to individual patients based on their unique characteristics, including genetic makeup, biomarkers, clinical symptoms, and personal preferences.

1.3 Scope and aim of the review

SCOPE

Since computational psychiatry is a relatively new field, the research will be focused on the articles and books published from 2007 onwards (PubMed, ResearchGate, Google Scholar), as well as on the material presented at the Computational Psychiatry Course Zurich 2023 & 2024[8] (which is keeping up-to-date with the latest research in the field). This year has been chosen as a tentative date since the first article containing the term "computational psychiatry" appeared in 2007 (Montague et al., 2007). However, many concepts and findings related to this research certainly predate this point in time, so related resources with earlier publication dates will also be included. Thematically, only the research concerning the most prevalent or the most known disorders is going to be presented.

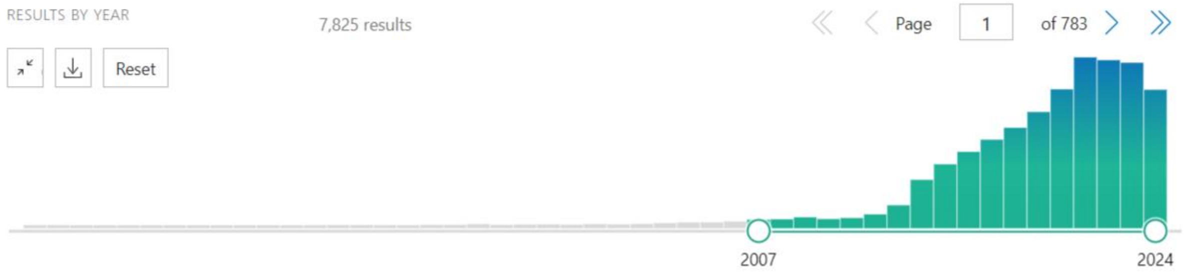


Figure 1.5 Computational psychiatry as a research trend. Search for "computational psychiatry" on PubMed yielded 7,825 results for the period between 2007 and 2024 and shows an increasing trend.

METHODS

Since this is a critical review (extensive, consistent overview of theoretical approaches and concepts in a new area of study), the method used is finding relevant literature (articles, books and lectures on CP), systemizing knowledge about computational approaches to psychiatry in general, and more concretely, applying it to the study of one particular disorder (depression). Another effort at systemizing findings from the literature refers to a more comprehensive presentation of the types of data used in CP models. Various tools (e.g., ConnectedPapers,

[8] This course has been organized by the Translational Neuromodeling Unit, University of Zurich & ETH Zurich since 2014. A list of recommended literature for the course can be found in the following link: https://www.tnu.ethz.ch/de/teaching/cpcourse/cpc2021readinglist

ResearchRabbit) are used to establish connections between articles and authors researching overlapping phenomena.

In the conclusion of the thesis, the author will try to provide a critical overview of the approaches in CP (possible limitations, ethical considerations, future directions), as well as presenting possible new directions in their own research.

AIMS

This thesis has three main aims: (1) to provide a detailed and up-to-date review of computational models used in computational psychiatry (data-driven, theory-driven and combined approaches) (Huys et al., 2016), as well as other alternative models (e.g., RDoC initiative by NIMH)[9]; (2) to synthesize findings from the application of these approaches to depression and to identify possible research gaps and understudied phenomena. These examples will try to illustrate how the necessity in clinical practice encourages novel solutions by leveraging theoretical knowledge and computational tools (3) to offer a systematic overview of all the types of data used in computational psychiatry (neuroimaging, genetic, digital, behavioral data) and highlight possible problems that may be encountered when attempting to include them in models and operationalize them as meaningful constructs, often at different levels of analysis (e.g., relating neuroimaging and behavioral data).

2 Definition and overview of the most common mental health disorders

2.1 Definition of mental health disorders

The question of ‘what is a mental health disorder?’ is a fundamental one for the philosophy of psychiatry, but it is also of great practical importance for both clinicians and patients. The first instance refers to attempts to delineate behavior that is considered unacceptable or harmful (e.g. crime, authoritarian behavior) in the broader context of society from the inherently dysfunctional behavior that significantly impairs individual’s everyday functioning. The latter is more concerned with coming up with a classification system of the disorders based on their

[9] Research Domain Criteria (RDoC) is a new conceptual model proposed by the US National Institute of Mental Health in 2010 (cf. 2.3. Alternative approaches (RDoC initiative by NIMH)).

descriptive or phenomenological characteristics (categorical approach) that enables comparability and validation by clinical practitioners, and is determined by its usefulness for clinical practice.

There are currently two widely established systems that classify mental disorders: *ICD-11 Chapter 06: Mental, behavioral or neurodevelopmental disorders*, part of the International Classification of Diseases produced by the WHO (in effect since 1 January 2022) and *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, Text Revision* (DSM-5-TR) produced by the American Psychiatric Association (APA) in 2022.

Table 2.1 contains complete definitions of mental disorders by both manuals (with common themes highlighted).

Table 2.1 Definitions of mental disorders by DSM-5-TR and ICD-11.

DSM-5-TR	**ICD-11**
A mental disorder is a **syndrome** characterized by **clinically significant disturbance** in an individual's **cognition, emotion regulation, or behavior** that reflects a **dysfunction** in the **psychological**, **biological**, or **developmental** processes underlying **mental functioning**. Mental disorders are usually associated with significant **distress** or disability in **social**, **occupational**, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above.	Mental, behavioural and neurodevelopmental disorders are **syndromes** characterised by **clinically significant disturbance** in an individual's **cognition, emotional regulation, or behaviour** that reflects a **dysfunction** in the **psychological**, **biological**, or **developmental** processes that underlie **mental** and behavioural **functioning**. These disturbances are usually associated with **distress** or impairment in personal, family, **social**, educational, **occupational**, or other important areas of functioning.

There are some notable formal differences between these two classification systems. Firstly, DSM-5-TR is primarily used in the US; while ICD-11 presents a broader, international system covering all aspects of health, including mental health (corresponding chapters include Chapter 06, 07 and 17). Secondly, ICD-11 uses alphanumeric codes that are integrated with global health information systems, whereas DSM-5-TR uses combination of its own codes and ICD-11 codes for compatibility.

These manuals have undergone multiple revisions, reflecting changes in societal attitudes, scientific knowledge, and clinical practices. This just highlights the fact that what constitutes a mental health disorder might sometimes be regarded as a social construct and that categorical classifications (due to highly overlapping symptoms between disorders) might at some point have to give way to a more nuanced approach, i.e. that these disorders should be observed as a continuum rather than a discrete category.

2.2 Brief overview of the most common classifications of mental health disorders (DSM-5-TR, ICD-11)

The following table contains broad groups of mental health disorders from the latest versions of DSM and ICD (Table 2.2).

Table 2.2 Groups of mental disorders from DSM-5-TR and ICD-11. Cells with a blue background indicate complete match of diagnostic groups and cells with the yellow background indicate a difference.

ICD-11	**DSM-5-TR**
Neurodevelopmental Disorders	Neurodevelopmental Disorders
Schizophrenia and Other Primary Psychotic Disorders	Schizophrenia Spectrum and Other Psychotic Disorders
Catatonia	
Mood Disorders	Bipolar and Related Disorders
	Depressive Disorders
Anxiety and Fear-Related Disorders	Anxiety Disorders
Obsessive-Compulsive and Related Disorders	Obsessive-Compulsive and Related Disorders
Disorders Specifically Associated with Stress	Trauma- and Stressor-Related Disorders
Dissociative Disorders	Dissociative Disorders
Feeding or Eating Disorders	Feeding and Eating Disorders
Elimination Disorders	Elimination Disorders
Disorders of Bodily Distress and Bodily Experience	Somatic Symptom and Related Disorders
Disorders Due to Substance Use and Addictive Behaviors	Substance Use and Addictive Disorders
Impulse Control Disorders	Disruptive, Impulse-Control, and Conduct Disorders
Disruptive Behavior or Dissocial Disorders	
Personality Disorders and Related Traits	Personality Disorders
Paraphilic Disorders	Paraphilic Disorders

Factitious Disorders	In Somatic Symptoms and Related Disorders
Neurocognitive Disorders	Neurocognitive Disorders
Mental or Behavioral Disorders Associated with Pregnancy, Childbirth and Puerperium	No separate grouping
Secondary Mental or Behavioral Syndromes Associated with Disorders or Diseases Classified Elsewhere	No separate grouping
Sleep-Wake Disorders (Ch. 7)	Sleep-Wake Disorders
Sexual Dysfunctions (Ch. 17 Conditions Related to Sexual Health)	Sexual Dysfunctions
Gender Incongruence (Ch. 17 Conditions Related to Sexual Health)	Gender Dysphoria

As we can see from Table 2.2, many groups of disorders have similar or identical names or they encompass similar disorders, which is indicative of the need for the convergence in classification and terminology between these two systems.

DSM and ICD, in general, are based on a theoretical, descriptive approach, whereby each disorder is characterized by a list of possible symptoms. A minimum number of those symptoms need to be present concurrently and during a certain time span in order to warrant a diagnosis.

A debate about the development and utility of these diagnostic manuals is ongoing and often contentious; however, the span of this thesis is not broad enough to discuss them here. Yet, just to illustrate how clinicians perceive utility of these classifications, we cite the following example from a global survey from 2018: both classifications (DSM and ICD) were rated to be most useful for assigning a diagnosis, communicating with other health care professionals and teaching, and least useful for treatment selection and determining prognosis (First et al., 2018).

2.3 Alternative approaches (RDoC initiative by NIMH)

Classifications based on current versions of DSM and ICD have facilitated reliable clinical diagnosis and research for decades. However, recently it has become apparent that diagnostic categories based on clinical consensus fail to align with findings emerging from clinical neuroscience and genetics (Insel et al., 2010). Namely, classifications based on descriptions of symptoms have not always been able to capture the underlying pathophysiology.

Therefore, in 2010, the National Institute of Mental Health (NIMH, USA) launched the Research Domain Criteria (RDoC) initiative to create a framework for research on pathophysiology, especially for genomics and neuroscience. Instead of having some *a priori* categories to which certain symptoms need to conform, RDoC, as a research framework, offers greater flexibility to include and explain away other symptoms that potentially contribute to the clinical picture of a particular disorder. However, RDoC is not intended to serve as a diagnostic guide nor does it attempt to replace current diagnostic systems (Sèries, 2020).

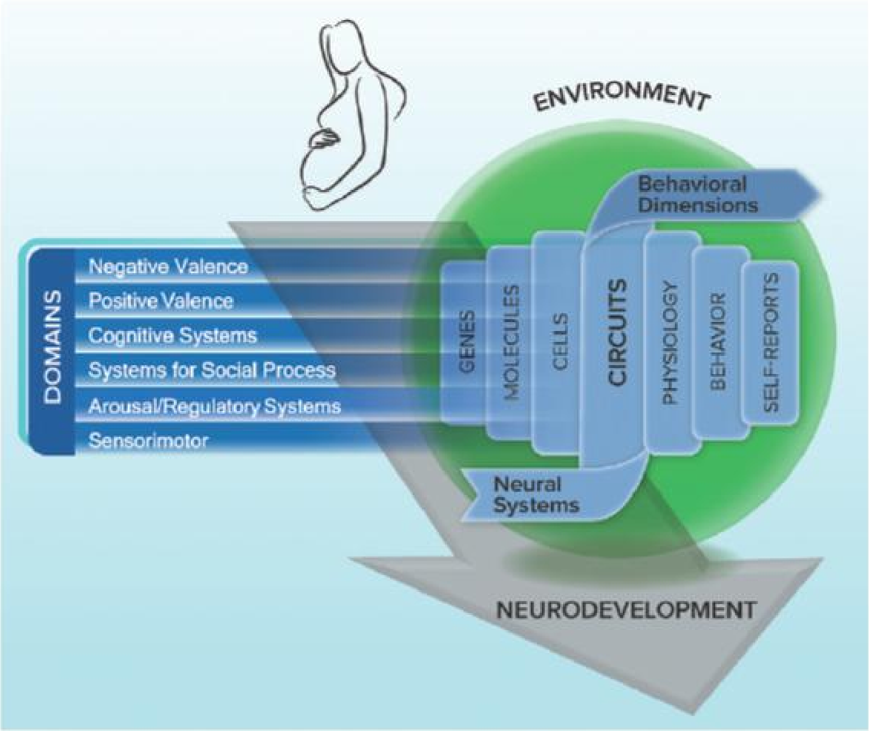


Figure 2.1 Illustration of the RDoC matrix. Adapted from Seriès (2020).

As it can be seen from Figure 2.1, RDoC operates on several levels and utilizes several types of qualitatively diverse data. It conceptualizes mental health as a continuum, i.e. the disorders are viewed on a spectrum from complete health to varying levels of dysfunction. RDoC proposes human behavior to be broken down into fundamental domains of function (like negative/positive valence, cognitive systems, systems for social processes, arousal and regulatory processes, and sensorimotor systems)[10]. These domains are further differentiated into psychological-level constructs, which should link the behavior to the function of specific neural circuits or (biological) systems[11]. The matrix uses different levels of analysis: from genes, molecules and cells, which correspond to neural systems, to physiology, behavior and self-reports, which constitute behavioral dimensions, with neural circuits as a crossing point[12]. The effects of environment and neurodevelopmental processes are also taken into account.

In order for us to validate these constructs and assess their clinical utility, they need to be in line with empirical findings, and to be able to withstand rigorous testing[13]. Although it is conceptualized as a research framework, the ultimate goal of the RDoC initiative is to be able to translate these findings into clinical practice (e.g., for early detection, better classification, treatment selection and targeting, prediction of prognosis and treatment outcomes). RDoC is also taking into consideration subjective feedback from the patients (by means of self-reports), which has possibly been undervalued in determining the complete clinical picture, i.e. subjective experiences need to conform to particular established symptom descriptions to be evaluated for diagnosis. This shift in conceptualization of the problems faced in psychiatry serves as a basis for the emerging field of computational psychiatry, because it provides a framework within which specific theories can be applied and models tested. The existence of domains and dimensions in RDoC encourages interdisciplinary cooperation, which is also one of the cornerstones of computational psychiatry research. Levels of analysis that transcend the individual (as a biological entity), such as environmental and neurodevelopmental factors can

[10] Negative valence systems involve responses to aversive situations or contexts, such as fear, anxiety, and loss. Positive valence systems relate to responses to positive motivational situations, such as reward seeking and habit learning. Cognitive systems operate with constructs such as attention, perception, declarative and working memory. Systems for social processes mediate the responses to various interpersonal settings. Arousal/Regulatory systems enable activation of appropriate neural responses for achieving homeostatic balance (arousal, circadian rhythms, and sleep-wake patterns). Sensorimotor systems are responsible for the control and execution of motor behaviors (NIMH» RDoC Matrix, n.d).

[11] For example, ***reward motivation*** is a construct that can be used to explain the following dimension of functioning: if it is excessive, it may contribute to substance abuse or gambling, if it is deficient, it may be a factor in anhedonia or anorexia (Kozak & Cuthbert, 2016). Biological underpinnings of reward seeking behavior have already been well-documented (e.g., dopaminergic system).

[12] Circuit-level is the focal element in the RDoC organisation (Insel et al., 2010).

[13] The lack of specific biomarkers in psychiatry was one of the reasons for stagnation in this area of medicine, compared to other areas, in which, once the right biomarker was identified, it was possible to establish the diagnosis and identify targets for treatment.

also be modeled this way. On the other hand, interoceptive processes have also gained importance as a contributing factor to the overall understanding of mental health disorders (Khalsa et al., 2018).

2.4 Diagnostic difficulties

As it was mentioned before, one of the main reasons for the difficulties in diagnosing mental health disorders is that in other branches of medicine, in contrast to psychiatry, it is possible to validate the diagnosis empirically by means of specific biomarkers or (neuro)biological bases. This is possible only for a small number of mental health disorders,[14] but is not routinely performed.

Mental health disorders are usually associated with a heterogeneous clinical representation, which often leads to diagnostic challenges, since multiple disorders have a lot of overlapping symptoms. In other words, it is very rare to have a set of symptoms that are unique for one particular disorder.

The use of current classification and diagnostic systems (DSM, ICD) has led to discussions on both theoretical and practical levels. On the theoretical level, the concept of ***comorbidity***[15] in psychiatry seems to require significant revisions in order to be clearly defined.

On the other hand, in clinical practice, it has been suggested that psychiatric comorbidity might be an artifact or a by-product of the DSM/ICD strategy to “split” categorical diagnoses. Another feature of the diagnostic manuals which contributes to comorbidity is that users are instructed to follow the general rule of recording as many diagnoses as are necessary to cover the clinical picture (First, 2005). However, this strategy should be reviewed in light of clinical utility and whether it ***adds*** or ***obscures*** important

[14] For example, genetic studies have yielded some significant results, especially about the higher heritability (the proportion of causation attributable to genetic factors) for the more severe and less common disorders such as autism, schizophrenia and bipolar disorder. Genome-wide association studies (GWAS) have identified more than a hundred genetic variants associated with severe mental illness (Uher & Zwicker, 2017). GWAS test hundreds of thousands of genetic variants across many genomes to find those statistically associated with a specific trait or disease.

[15] In simple terms, psychiatric comorbidity refers to the co-occurence of two or more mental disorders. It is implied, however, that these disorders are mutually independent. Some authors (First, 2005) argue that ***true comorbidity*** in psychiatry is rare (it needs to fulfill the criteria of either known etiology and/or circumscribed pathology, similarly to general medicine), and that most of the cases can be classified as ***artifactual comorbidity***, a by-product of the DSM/ICD strategy to “split“ categorical diagnoses.

clinical information. For example, allowing the diagnosis of panic disorder in the presence of the diagnosis of schizophrenia ***adds*** clinically useful information in terms of clinical management (e.g., choice of treatment, prognosis). However, the question remains whether representation of a panic disorder due to agoraphobia, for example, is the same as the picture of a panic disorder with comorbid schizophrenia. On the other hand, in DSM-IV, generalized anxiety disorder (GAD) was not diagnosed if it occurred only during major depressive disorder (MDD), since it was a commonly associated feature of MDD, thus ***obscuring*** the presence of anxiety, with various therapeutic and prognostic implications. However, recording of multiple disorders is important for indicating the complexity of the clinical picture, which is an obvious predictor of greater severity, disability and service utilization (Maj, 2005).

This is also related to the hierarchical structure of the diagnostic manuals, which is based on the order of diagnoses given by Kraepelin, so that disorders that are higher in the hierarchy (e.g., organic disorders) take precedence over disorders of the lower order. For example, if both are present in a patient, the diagnosis of schizophrenia will be ranked higher than MDD (First, 2005). Another characteristic of the diagnostic systems is that they are based on categorical principles. In such systems, diagnoses are established when the patient's symptoms exceed some (often arbitrarily set) threshold. In contrast, in dimensional systems, symptoms can be viewed on a continuum (according to the degree of severity). Although helpful in establishing diagnoses, categorical classifications are still artificial constructs to some extent, since they are rarely associated with objective measurements. A more pragmatic approach to psychiatric assessment, which would allow recognition of individual experiences of distress, instead of strictly relying on clinical categories, would probably enable a more holistic perspective (Maj, 2005). One of the dangers of allowing proliferation of diagnoses is polypharmacy (simultaneous use of multiple medicines), which might lead to various adverse effects and complications in treatment (Maj, 2005).

In psychiatric practice, there is also a risk of a misdiagnosis or a failure to detect the disorder. If a misdiagnosis happens due to inadequately following diagnostic guidelines, it is both unethical and harmful for the patient (Nordgaard et al., 2023). Apart from the heterogeneous clinical picture, mental health disorders often have unusual trajectories, from symptom onset (which may not always be recognized) to recurrent episodes or remission, over a period that might span for years or decades. Therefore, temporal dimension and progression of mental health disorders is a significant facet, which is why computational models are well-suited to provide answers to questions regarding treatment outcomes or disease prediction.

3 Definition and scope of computational psychiatry

Computational psychiatry (CP) is an emerging ***interdisciplinary field***[16] that aims to integrate computational modeling, empirical data, and theoretical insights from various fields, such as psychology, neuroscience, computer science, and mathematics, in order to better understand psychiatric disorders and their underlying mechanisms (Vasilchenko & Chumakov, 2023). Although it is difficult to give an all-encompassing and comprehensive definition of CP, since it does not have a uniform "manifesto" (possibly due to its interdisciplinarity and/or complexity of phenomena studied), it seems that it has emerged as a (mostly) theoretical framework in a Seriès of articles of like-minded authors (Montague et al., 2012, Friston et al., 2014, Huys et al., 2016)[17]. The hope is, however, that these findings will eventually be translated into clinical practice[18].

Computational models are becoming an indispensable tool for studying and explaining the underlying mechanisms of various scientific phenomena. The complexity of diagnosing and treating mental health disorders has proven to be an exciting challenge for the emerging field of computational psychiatry. Simply defined, this field represents the application of computational modeling and theoretical approaches to psychiatric questions (e.g., to explain the underlying mechanisms of psychopathologies). Despite the progress in brain imaging and success of computational neuroscience in explaining various phenomena related to the

[16] Interdisciplinarity in the context of computational psychiatry refers not only to cooperation between various branches of science and scientific disciplines, but also to the cooperation between theorists and clinical practitioners.

[17] The first international computational psychiatry meeting was held in 2013, and 2014 saw the inception of the Max Planck Society-University College London Initiative on Computational Psychiatry and Ageing Research (Friston et al., 2014).

[18] In general, translational research is a bidirectional concept in which the knowledge generated from the "benches" of laboratory science can be translated to "bedside" (or the population) and vice-versa. When applied to psychiatry, it involves the translation of (neuro)scientific discoveries into clinically meaningful interventions. Additionally, translational psychiatry involves translating observations and clinical insights from patients into hypotheses for basic research, facilitating a bidirectional flow of information between laboratory research and clinical practice (Weissman et al., 2011).

functioning of the brain, in the case of psychiatry, there is still an ***explanatory gap***. Due to insufficient understanding of human cognition (and ***cognitive phenotypes***)[19], the researchers have been unable to provide a bridge between the molecular/neural and phenomenological/behavioral levels and to explain how some changes on neural level give rise to the changes in behavior.

Apart from the most obvious, observable changes at the symptom level (e.g., changes in mood), some authors emphasize the importance of ***aberrant decision-making*** in a large number of psychiatric conditions. For example, patients suffering from depression choose not to explore, persons affected by obsessive-compulsive disorder (OCD) choose to repeat some behavior (despite the lack of some rational basis), etc. (Montague et al., 2012). But why patients "choose" to act along these lines remains unexplained and indicates that aberrant decision-making is not a primal cause, but rather a consequence or symptom of some underlying cause. Other authors (Friston et al., 2014) place importance on the ***production of false beliefs*** as the central problem in psychiatry (e.g., false beliefs about agency in schizophrenia, learned hopelessness or helplessness in depression). The question about different kinds of false beliefs in different kinds of mental disorders however points to the assumption that these false beliefs are rather a consequence than a cause of the underlying disorder.

Computational models have the advantage of formalizing operational concepts and explicitly stating experimental hypotheses. The models can be constrained in terms of number of parameters, processes and potential outcomes in order to test a particular hypothesis. Both manipulated (independent) and measured (dependent) variables can be incorporated into the model, so that the extent to which experimental results match model predictions can quantitatively and qualitatively inform our mechanistic understanding and guide future experiments. Computational models can also explicitly incorporate time, giving the possibility to understand temporal progression of mental health disorders (e.g., by means of treatment outcome simulations) (Seriès, 2020). If there is a discrepancy between the model predictions

[19] Cognitive phenotype can be defined as a measurable trait of some aspect of cognitive functioning. Similarly, a ***computational phenotype*** is a measurable behavioral or neural type defined in terms of some computational model. An individual's computational phenotype is defined as a set of mechanistically interpretable parameters obtained from fitting models to behavioral data (Schurr et al., 2024). Large-scale computational phenotyping in humans has not yet been carried out.

and empirical data, the model can be refined in order to include some latent/hidden variables that might be required to explain the phenomenon, but were not obvious at the time.

In order to fully understand the computational approach to mental health disorders, we need to look at some broader theoretical and methodological concepts that constitute the operational framework of computational psychiatry. Predictive coding, free energy principle and Bayesian inference are three interconnected concepts that are used to explain the transmission of neural messages in the brain, or more generally, information processing. The brain is constantly confronted with a wealth of sensory information that must be processed efficiently in order to facilitate appropriate reactions. One way of optimizing this processing effort is to ***predict*** incoming sensory information based on previous experience, so that resources can be allocated to novel or surprising stimuli. This idea emerged as a ***predictive coding framework*** (Friston, 2005; Rao & Ballard, 1999). Predictive coding states that brain is continually generating models of the world and that it is trying to predict sensory input. These predictions are then compared to actual sensory input, and the "mismatch" between the two, i.e. the ***prediction error***, is then used to ***update*** the brain's model of the world. Predictive coding assumes a hierarchical brain structure (Figure 3.1). A predictive model is created in higher cortical areas and passed through feedback connections (top-down) to lower sensory areas (directly receiving sensory stimuli), while certain feedforward connections project an error signal, i.e. prediction error (bottom-up). The predictive model is constantly updated according to this error signal (Rao & Ballard, 1999).

The free energy principle (FEP), introduced by Karl Friston, is a theoretical framework that generalizes the idea of predictive coding. It states that all biological systems (including the brain) are driven to minimize a quantity called "free energy", which also corresponds to discrepancy between the predicted and actual signal. The brain minimizes "free energy" either by updating its internal model or by acting on the environment so that it fits its prediction. ***Active inference*** is a corollary of the free energy principle that describes the process of inferring the causes of sensory data, which have to be actively chosen or sampled (Friston et al., 2014). ***Bayesian inference*** is a formal mechanism central to both predictive coding and FEP, and it presents a statistical method used to update beliefs or models based on new evidence. In simple terms, it entails updating our existing belief (***the prior distribution***) with new information (***the likelihood distribution***) to form our new belief (***the posterior distribution***).

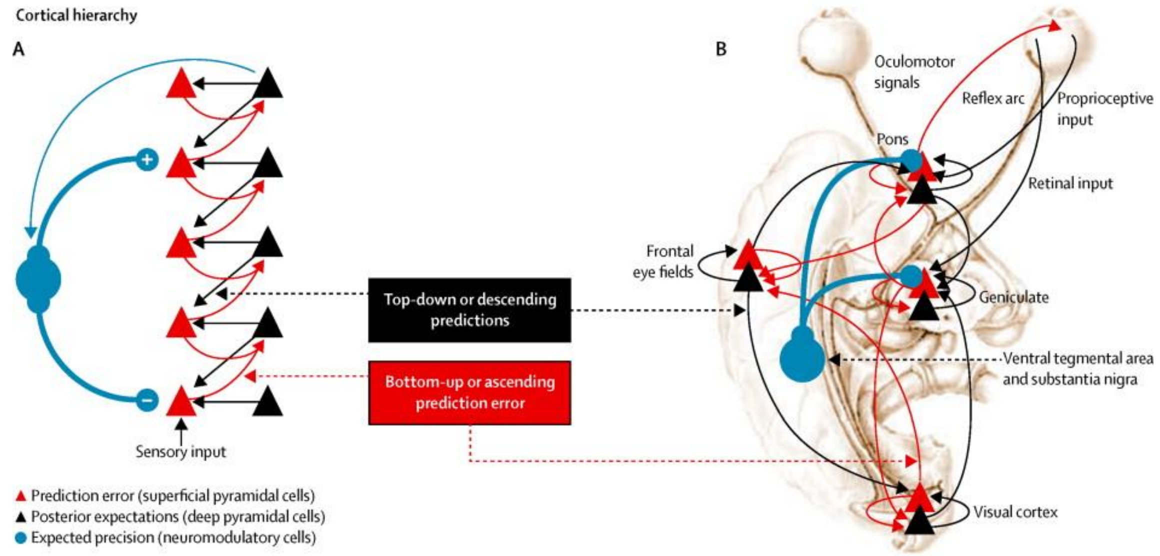

Cortical hierarchy
A
B
Sensory input
Top-down or descending predictions
Bottom-up or ascending prediction error
Oculomotor signals
Reflex arc
Proprioceptive input
Pons
Retinal input
Frontal eye fields
Geniculate
Ventral tegmental area and substantia nigra
Visual cortex
▲ Prediction error (superficial pyramidal cells)
▲ Posterior expectations (deep pyramidal cells)
● Expected precision (neuromodulatory cells)

Figure 3.1 Hierarchical representation of the neuronal message transmission in predictive coding. Adapted from Friston et al. (2014).

Taking all of the above into the consideration, computational psychiatry can be conceptualized as an attempt to characterize mental dysfunction in terms of aberrant computations over multiple scales (Montague et al., 2012). The idea is that computational models can provide possible explanations and hypotheses testing by comparing how these computations differ in healthy controls and patients suffering from mental health disorders. For example, a key issue in schizophrenia research is a false belief about agency, i.e. the fact that the patients suffering from schizophrenia attribute beliefs about their actions to external forces. The proof for this is the resistance of patients towards sensory attenuation compared to healthy subjects. The study by Shergill et al. (2005)[20] showed that self-generated forces were attenuated less in the patient group, suggesting a dysfunction in their ability to predict the sensory consequences of their actions. This can be conceptualized as a propagation of the prediction error of the proprioceptive signal in a computational model (Friston et al., 2014)

[20] This study used the force-matching task to assess the level of sensory attenuation, whereby participants are instructed to reproduce the sensation, i.e. applying pressure on a passive finger, by directly pressing it with a finger of the other hand. Healthy subjects tend to overestimate the required force needed to match the force of the stimulus.

(Figure 3.1). Therefore, in the case of schizophrenia, this error signal is not transmitted optimally. However, mental health disorders are complex phenomena, so computational models are usually able to capture only some facet(s) of the process(es) underlying particular disorders. Additionally, due to their reductionist nature, models need to constrain the number of variables required for the explanation.

The main aims or goals of computational psychiatry, apart from providing theoretical explanations, seem to be more practical and pragmatic, with the intention to facilitate processes in clinical practice. Therefore, main aims of computational psychiatry include, but are not limited to, improved ***classification*** of mental health disorders, ***predictions*** and ***simulation*** of treatment outcomes and longitudinal disease course, ***treatment selection*** etc. The ultimate goal of computational psychiatry, however, is to be able to translate these findings into useful interventions in clinical practice.

4 Types of data used in computational psychiatry

4.1 Main types of data used in computational psychiatry

Due to complexity of biological mechanisms of mental health disorders, their heterogeneous representation, and different ways of probing behavioral and cognitive components of the disorders, computational psychiatry deals with large and very diverse datasets (Figure 4.1). Psychiatric practice may have started with collecting self-report and behavioral data (which contributed to the establishment of current classification and diagnostic systems), but advances in neuroimaging and genetics enabled collection and interpretation of biological data that underlie the disorders. Finally, mobile devices, social media and online data collection platforms offer a more ecologically valid, cheaper and possibly more efficient way of collecting data from participants compared to clinics or laboratories. Additional advantage is tracking or collecting data in real time. We discuss these types of data in more detail below, as well as the problems and challenges encountered during the processes of data collection, implementation into computational models and data validation.

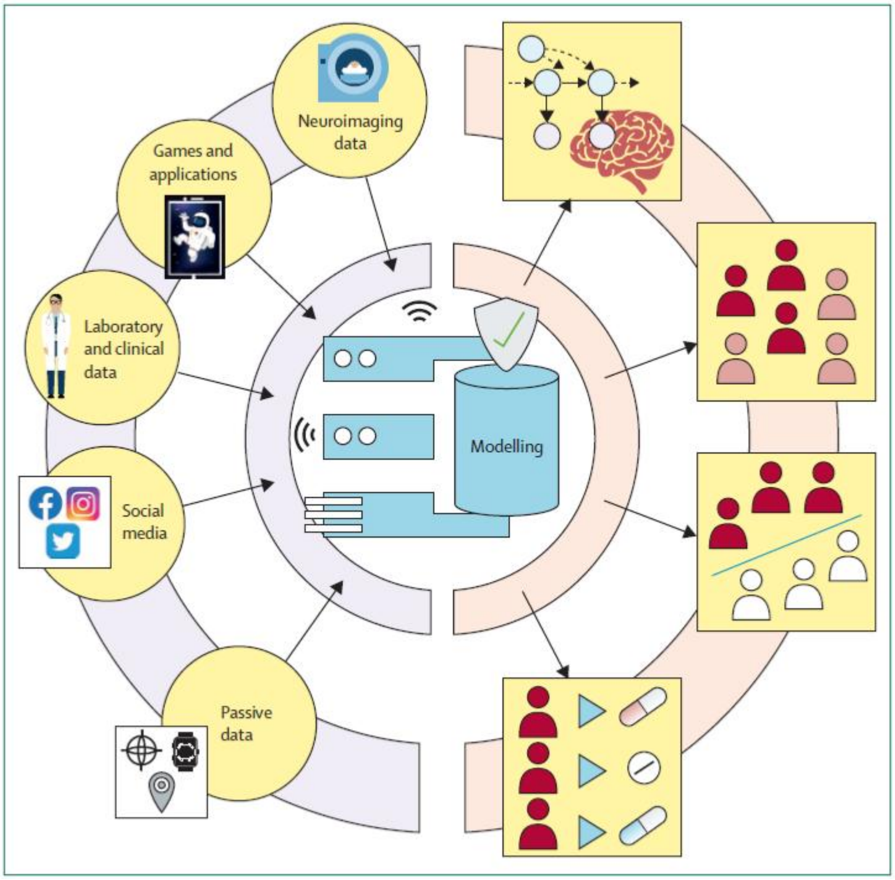


Figure 4.1 Types of data used in CP. Adapted from Hauser et al. (2022)

Based on the place or source of collection, we can distinguish between these types of data: (1) clinical, (2) laboratory-based and (3) digital data. Clinical data are collected by clinical staff and provide information about the process of diagnosis and treatment in healthcare facilities (e.g., detailed notes on patients, records of self-report and interviews, assessment scales, questionnaires, etc.). However, privacy concerns and missing data infrastructures make it challenging to harvest such data for modeling purposes (Hauser et al., 2022). Laboratory-based data is data collected in controlled environments for scientific studies. These often entail behavioral and biology-derived data. Due to controlled conditions and selective participant recruitment, these data are reliable and denoised. However, due to expensive methods, sample sizes are usually quite small and biased, which has implications for model generalizations, statistical power and translation of findings. Digital data is collected via digital devices, social media and online data collection platforms. Compared to other two methods of collection, digital data collection is faster, has fewer limitations regarding the diversity of the sample (various demographics) or collection in multiple points in time (suitable for longitudinal research).

4.2 Clinical data

4.2.1 Clinical measures

Self-reported or clinician-rated symptom data provide qualitative and quantitative measures of mental health states. The primary psychiatric assessment tool is a direct face-to-face interview, with the emphasis on both form and content, and it is tailored to the needs of the individual patient. Other clinical assessment tools that complement the interview include (semi)-structured interviews[21], standardized data forms, questionnaires, and rating scales[22]. Assessment tools are used for diagnostic reasons, for assessing the severity of the disorder and for recording the change during treatment (Koen Demyttenaere & Heirman, 2023). Although these tools are widely used in clinical research, they are less often incorporated into daily practice, but due to rigorous development and validation process, quantitative and qualitative data obtained this way is highly reliable.

4.2.2 Behavioral data

Behavioral paradigms in psychiatry

In psychiatric research, experimenters can use various behavioral paradigms in order to elicit and test various aspects of observable behavior. Paradigms that are often used include: ***decision-making***, ***social behavior*** and ***emotional processing***, since these are the areas of functioning that are often compromised in mental health disorders.

Tasks that can probe the function of decision-making in patients include Iowa Gambling Task (IGT)[23], delay discounting and two-step tasks. IGT evaluates risk and reward

[21] Fully structured interviews have detailed standardized questions while semi-structured interviews resemble a guided diagnostic conversation.

[22] Important distinction has to be made between the observer-rating scales and self-rating scales, whereby the latter provides more “subjective” and possibly biased information regarding the patient’s condition.

[23] The Iowa Gambling Task (IGT, designed by Bechara et al., 1994) involves probabilistic learning via monetary rewards and punishments, where advantageous task performance requires subjects to

processing, which is useful for studying addiction and impulsivity. Similarly, delay discounting measures preference for immediate versus delayed rewards, which is often altered in disorders such as ADHD and addiction. Two-step task is a reinforcement learning paradigm used to dissociate model-based and model-free learning, and can be applied to disorders such as OCD or schizophrenia (Castro-Rodrigues et al., 2022).

Tasks related to assessing social behavior are usually game-based, e.g., ultimatum game or trust game[24]. The elements of trust, reciprocity, cooperation, but also risk-taking (which has been recognized as a confounding variable involved in such games) made them suitable for studying disorders such as schizophrenia or autism spectrum disorders (ASD) (Robson et al., 2019). Importance of correctly recognizing emotions and intentions of others in social contexts is particularly highlighted in ASD. Facial emotion recognition tasks measure the ability to identify emotions, which might be impaired in disorders such as depression (Krause et al., 2021). In some cases, patients show attentional biases towards "sad faces". Affective Go/No-Go tasks (especially inhibition of the prepotent response in the No-Go element) enable measuring impulsivity in a variety of disorders, such as ADHD or eating disorders. Simpler psychometric measures, such as reaction times, are useful for studying certain aspects of the disorders, e.g., slower response time in decision-making tasks in depression, reflecting affected cognitive effort and/or possibly psychomotor retardation.

4.3 Laboratory-based data

4.3.1 Neuroimaging data

Neuroimaging data used in computational psychiatry may be divided into three distinct categories: (1) ***structural*** imaging (e.g., magnetic resonance imaging (MRI), diffusion tensor imaging (DTI)), (2) ***functional*** imaging (functional MRI (fMRI), electro- and magnetoencephalograpy ((M)EEG)), and (3) ***molecular*** imaging (positron emission

forego potential large immediate rewards for small longer-term rewards to avoid larger losses (Bull et al., 2015).

[24] An Ultimatum Game is defined as a behavioral economics exchange game where two players, a proposer and a responder, decide how to split a sum of money. If the responder rejects the proposer's offer, neither player receives any money. Trust game (designed by Berg et al., 1995) is another example of a neuroeconomic game whereby the amount given by the investor to the trustee may be multiplied.

tomography (PET)). Neuroimaging in CP is mostly used for studying brain connectivity or for biomarker discovery[25] (diagnostic or prognostic) in mental health disorders. Development of biomarkers in psychiatry seems especially tantalizing because of the lack of objective measures or "gold standard" diagnostic markers. However, neuroimaging-based markers for CP have shown varying degrees of validity and clinical utility.

MRI is primarily used in research for establishing quantitative differences between patients with mental health disorders and healthy controls in specific ROIs, thereby enabling the formulation of hypotheses regarding pathophysiology[26]. More recently, a variation of structural MRI called DTI has been utilized in psychiatry for assessment of white matter tracts, and indirectly, as a measure of connectivity in psychiatric conditions (Teixeira et al., 2023). Recent application of ML techniques for identification of patterns in MRI data enables the discrimination of patients versus controls, however with variable degrees of sensitivity and specificity.

In the group of functional imaging methods, fMRI is the most used and studied technique and it enables mapping of functional connectivity between specific brain regions based on the BOLD signal. Task-based fMRI studies engage particular networks (e.g., working memory, emotional processing) and measure the BOLD signal changes between the task and control states. On the other hand, resting state fMRI measures low-frequency changes and it is useful for the characterization of the functional architecture of the brain (Teixeira et al., 2023).

Finally, PET imaging is a valuable tool for assessing the accumulation or distribution of certain neurotransmitters and neuromodulators suspected to be implicated in a variety of mental health disorders (e.g., dopamine, serotonin, glutamate). PET radioligands can be used for characterization of functional anatomy and pathophysiology, diagnosis, early detection and prognosis, disease monitoring and pharmacological advancement in (neuro)psychiatry (Teixeira et al., 2023).

[25] A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group, 2001).

[26] MRI has also been used for elimination of the dichotomy between functional and organic mental health disorders.

4.3.2. Genetic data

Biobanks are collections of human biological materials (biospecimens) alongside personal health information that are stored for scientific research. Biological specimens usually include blood, but also saliva, hair, feces, cerebrospinal fluid (CSF), and tissue. These samples are then linked to individuals' personal medical records such as history, lifestyle and genetic markers. Samples need to be properly collected, stored and maintained. For mental health disorders, the process usually involves comparing genetic markers of individuals with the same diagnosis or identifying inflammatory pathways in specific conditions (Govind et al., 2024). The best known biobanks are UK Biobank (voluntary collected data from 500,000 participants) and All of Us Research Program (US) (longitudinal data from over a million participants). Ethical considerations and possible limitations regarding the use of potentially sensitive data will be discussed in Chapter 7.

Heritability has been assumed to play a role in the development of mental health disorders for a long time. Over the past decade, advances in psychiatric genetics have provided significant insights into the genetic etiology of psychiatric disorders (e.g., pathobiology)[27] and the effects of gene-environment interplay, which is considered to be the most likely mechanism for the emergence of mental health disorders. Genome-wide association studies (GWAS) have become the most successful approach for linking genetic variants to human phenotypes. GWAS test hundreds of thousands of genetic variants across many genomes to find those statistically associated with a specific trait or disease.

In psychiatry, genetic data is useful for the assessment of heritability for the potential development of mental health disorders, i.e. determining the risk factor. The estimated heritability is generally higher in psychotic and neurodevelopmental disorders such as schizophrenia and autism (74-85%) than in mood and anxiety disorders (37-58%). However, current polygenic risk score tools, which predict individual genetic susceptibility to illness, do not yet provide clinically actionable information, e.g., are not ready to use for early detection, prevention or prediction (Andreassen et al., 2023).

[27] The most convincing biological interpretation of the genetic findings implicates altered synaptic function in ASD and schizophrenia.

4.4 Digital data

Digital data may be roughly divided into ***passive*** and ***active data*** and includes any data collected from participants via digital devices. Data are most commonly collected via mobile phone applications, social media and online collection platforms. Active data requires the participant to interact with a request from the experimenters, while passive data is obtained from social media activity and sensor data from smartphones and wearable devices[28] (recording physiological responses and other responses, e.g., capturing information about circadian rhythms). Passive data collection is unobtrusive (requires minimal participation) and as such it is especially suitable for obtaining longitudinal data. In active data collection, participants most commonly engage in self-report as a means of assessing their mood and experiences or in game-like activities used for cognitive assessment. Data collected by both approaches meet the three criteria of "big data": velocity, volume and variety (Torous et al., 2015).

A move towards online-based task assessments in recent years is a first step towards clinically usable data assessment tools. Online services, i.e. crowdsourcing sites like Amazon Mechanical Turk (AMT) or Crowdflower allow large samples to be rapidly tested on cognitive tasks, allowing robust assessment of novel task characteristics and their relationship to self-reported clinical symptoms (Rutledge et al., 2019). The use of gamified smartphone applications[29], such as Brain Explorer (UCL, London) and Neureka (Trinity College, Dublin) has also proven to be promising.

Combining the data from mobile assessment platforms with self-reports constitutes ecological momentary assessments (EMA)[30]. However, despite obvious advantages of having

[28] Mental health has been linked to various types of information obtainable from mobile and wearable devices, such as geolocation (movement), sleep pattern data, smartphone and keyboard usage patterns (e.g., typing speed). For example, accelerometer data on 91,105 UK Biobank participants were used to derive circadian rhytmicity parameters related to sleep patterns. Circadian disruptions has been shown to be associated with increased lifetime risk of both major depression and bipolar disorder (Lyall et al., 2018). Also, tracking of sleep-wake cycles in real-time may indicate the onset of manic episode in bipolar disorder and enable timely interventions (Huang et al., 2021).

[29] Gamification refers to the approach of making cognitive (and other) tasks more game-like using the design principles implemented in electronic games, thus making them more entertaining, which increases user engagement (Hauser et al., 2022).

[30] Ecological momentary assessments include methods of repeated sampling of an individual's behavior and experiences in real-time and in natural environments (Hauser et al., 2022).

a more diverse participant pool[31], being cheaper and faster to implement, data collected this way should nevertheless be validated in some way. Promising results obtained in anonymous online studies should be replicated in more controlled laboratory studies (Rutledge et al., 2019)[32]. Bringing together passive and active data sources, e.g., by collecting eye-tracking data during game play could yield insights in future studies.

All these developments played a role in the emergence and need for digital phenotyping. ***Digital phenotyping*** (or personal sensing) is the moment-by-moment, in situ quantification (and prediction) of the individual-level human phenotype using data from personal digital devices (Huckvale et al., 2019).

Apart from data collection, attempts to leverage online data acquisition and testing for therapeutic purposes are also of great significance for mental health treatment. For instance, a smartphone study of digital cognitive behavioral therapy (CBT) showed alleviated symptoms of insomnia in participants compared with usual practice (Freeman et al., 2017). These approaches have gained additional popularity due to restrictions imposed by COVID-19 pandemic (i.e. the lack of direct contact between the patients and mental health care providers, which can be overcome by telepsychiatry, chatbots etc.).

4.5 Possible problems with data

In the process of computational modeling, various problems associated with the use of data may arise during different stages of data collection, implementation and validation.

Data collection process

Some of the problems that arise during the data collection process include working with noisy, missing and sparse data and small sample sizes. Noise in the data affects model quality and reliability and can add bias. Measurement noise can arise from poorly controlled data collection environments, imprecise data collection (e.g., MRI artifacts) or insensitive task

[31] Participants can be sourced worldwide and from diverse demographic backgrounds, they can participate anonymously (or not), belong to both control or patient groups, etc.

[32] For example, depressive symptoms collected via AMT had a high test-retest reliability (r = 0.87) after one week (Shapiro et al., 2013). Furthermore, data collected this way has the advantage of capturing momentary reactions compared to traditional clinical testing, whereby patients need to rely on their memory about past events when answering a questionnaire.

measures. Missing data is one of the main concerns in model building and it often requires statistical preprocessing and corrections, especially for longitudinal data. Sparse data (e.g., imbalanced samples) also lead to substantial biases. As mentioned previously, laboratory studies usually have smaller and biased sample sizes, which can lead to non-reproducible effects and low statistical power. On the other hand, online data collection can ensure large sample sizes, but such data tends to be noisy (Hauser et al., 2022).

Data implementation process

Collected data sometimes needs preprocessing or other preparation procedures in order to be implemented in computational models. Dimensionality of the data is determined by the number of input features. ***Highly dimensional data*** are rich in information (i.e. they contain many data points per participant) and potentially more robust against noise, but often require special treatment prior to model implementation, such as unsupervised dimensionality reduction or regularization techniques[33]. Sheer volume or dimensionality of the data may be reflected in the requirement of computational power to run such a model or in model scalability.

In addition to high dimensionality, data included in computational models are often ***multimodal***. Some research has shown that combining and integrating various types of data (e.g., MRI in combination with other data sources, such as clinician ratings, genetic data and neuropsychological tests) improved predictive ability in ML approaches, compared to using MRI data alone (Koutsouleris et al., 2021).

Data validation process

When using computational models, it is crucial that the model's performance is validated against an independent test dataset, usually by applying procedures such as ***cross-validation***. If such an approach is not used (i.e. within-sample prediction), then the accuracy might be inflated and the results prone to ***overfitting*** (Hauser et al., 2022).

[33] Regularization can be described as a set of constraints imposed on model parameters (e.g., weights or coefficients) to prevent them from taking too large values, thus eventually reducing model complexity and preventing overfitting. These models can account for the redundancy and high covariance between features.

Latent or hidden variables are those variables that are not immediately observable in the behavioral data (e.g., values of different choices in a task), but which the theory assumes are important for the computations occurring in the brain (Wilson & Collins, 2019).

Possible implications of the use of various types of data in different stages of the modeling process will be discussed in the next chapter, which outlines the process of computational model building and highlights possible problems along the way.

5 Methodologies of computational modeling and model building in psychiatry

Computational psychiatry encompasses three broad approaches: data-driven, theory-driven, and combined models (Figure 4.1). ***Data-driven*** approach consists of theoretically agnostic data analysis and methods from machine learning (ML) including, but also extending, standard statistical methods. ***Theory-driven*** models mathematically specify mechanistically interpretable relations between variables, often including both observable variables and postulated, theoretically meaningful hidden variables. However, these approaches are not mutually exclusive, and may be combined if necessary. For example, in high-dimensional datasets, theory-driven approach may be employed in the preprocessing step (for dimensionality reduction), in order to choose theoretically meaningful parameters for prediction and classification (Huys et al., 2016).

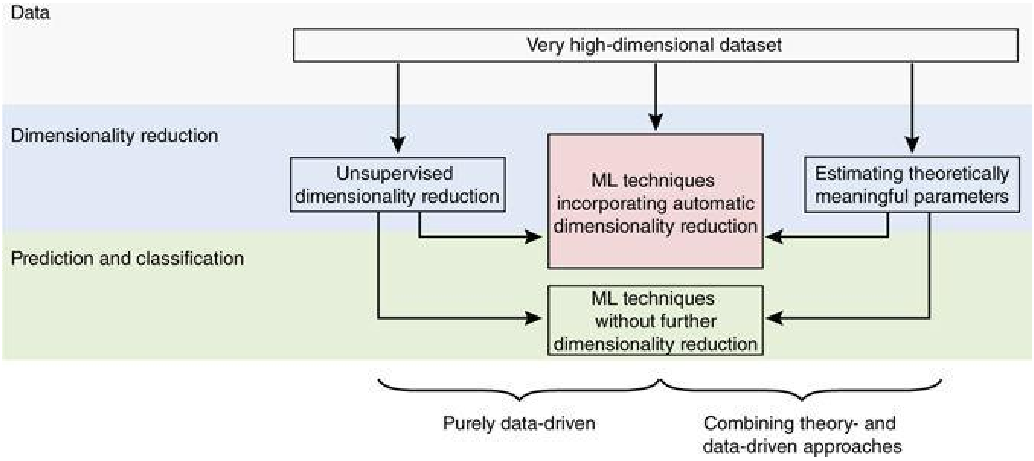

Data
Very high-dimensional dataset
Dimensionality reduction
Unsupervised dimensionality reduction
ML techniques incorporating automatic dimensionality reduction
Estimating theoretically meaningful parameters
Prediction and classification
ML techniques without further dimensionality reduction
Purely data-driven
Combining theory- and data-driven approaches

Figure 5.1 Combining theory- and data-driven approaches in high-dimensional datasets. Adapted from Huys et al. (2016).

5.1 Data-driven modeling

In computational psychiatry, it is not unusual to work with high-dimensional, multimodal datasets, including clinical, genetic, cognitive, neuroimaging, behavioral and other data types. ML techniques applied to such diverse data are generally agnostic in regards to the underlying mechanisms of the studied disorders and are used for finding patterns in data.

Machine learning (ML) is a field of artificial intelligence (AI), related to the study, design and development of algorithms and statistical models that can learn from data without explicit instructions (Silva & Zhao, 2016). The main aim of ML methods is to devise models capable of enhancing their precision over time by training on extensive datasets, followed by making predictions or decisions on unseen data. ML encompasses several learning types: supervised, unsupervised, semi-supervised and reinforcement learning (RL). In this section, we focus only on supervised and unsupervised methods. In supervised learning, input data and correct outputs are labeled with the aim of generalizing the association between the two. This enables the algorithm to predict unseen data via classification or regression. Unsupervised learning, conversely, uncovers hidden patterns within unlabeled data, commonly used for clustering, dimensionality reduction, and feature extraction. Other AI methods, such as deep learning (DL) models, an extension of learning to multilayer artificial neural networks (ANNs), are successfully used for the analyses of more extensive, complex and structured data, such as images or textural features (Wu et al., 2023).

In data-driven approach, the neuroimaging data are one kind of data that can be used to identify some neuropsychiatric disorders. Two common types of neuroimaging data analyzed in mental health studies are functional magnetic resonance imaging (fMRI) and structural MRI (sMRI). DL models have been successful in finding patterns in neuroimaging data for detection and prediction of disorders such as ADHD, schizophrenia and to some extent, depression (Su et al., 2020). In recent years, approaches such as natural language processing (NLP) have been applied to textual, audio or video data collected from patients/participants and show great potential for proactive mental healthcare, i.e. for early diagnosis, prevention and other mental health interventions[34] (Zhang et al., 2022).

[34] The examples of such data include social media posts, interviews, clinical and non-clinical notes. NLP approaches facilitate various tasks such as information extraction, sentiment analysis, emotion

In general, the ability of pattern recognition by ML is particularly useful for the following problems in computational psychiatry: (1) predicting clinical variables, (2) stratifying psychiatric disorders, and (3) learning mappings between behavior and brain systems.

The first group of problems generally consists of diagnostic classification (i.e. automating the diagnostic process), prediction of treatment outcomes (prognosis) and treatment selection. These problems can mostly be addressed by ***supervised learning***. For example, these methods can be used to automatically classify patients versus controls (e.g., schizophrenia) or to differentiate between the disorders (e.g., overlap between anxiety and depression) (Richter et al., 2020). Prediction of treatment response and outcome bears great clinical significance for the whole course of the treatment. For example, in depression, although up to three quarters of patients eventually respond to a particular antidepressant, two thirds require multiple treatment trials before responding (Huys et al., 2016). In order to make this process more efficient, response to various medications can be optimized by first performing referenced-EEG (rEEG) procedure[35] and then submitting it to automated analysis for medication ranking (DeBattista et al., 2011). Similarly, multiple regression analysis can be used for treatment selection; for instance, choosing between cognitive behavioral therapy (CBT) and antidepressants based on other relevant variables (marital or employment status, presence of comorbid disorders etc.) (Huys et al., 2016).

Clustering methods (an example of ***unsupervised learning***) have been used extensively for stratifying (subtyping) mental health disorders, both within and across different diagnoses. These methods perform very well if the disorder can be clearly separated into subgroups. On the other hand, the problem with this approach so far is that it always yields a result and separates data into a specified number of clusters regardless of underlying data distribution. Therefore, the number and validity of clusters must be specified *a priori* or assessed *post hoc* (Marquand et al., 2016). As an example of hierarchical clustering, Checkroud et al. (2017) stratified the symptoms of common depression into three statistically robust and replicable

detection, and mental health surveillance (i.e. they are useful for different types of screening, e.g., suicide, risk of self-harm).

[35] Referenced-Electroencephalogram (rEEG), or medication sensitivity testing, is a test that provides physicians a treatment guide for more effective medication treatment. Medication sensitivity testing can be completed during a quantitative (qEEG) brain mapping.

clusters: a mood/emotional cluster, a sleep/insomnia cluster, and an atypical symptom cluster. The utility of this finding was reflected in different responsiveness to antidepressants.

Clustering approaches are also significant for establishing new descriptions and classifications, beyond traditional, symptom-based categories. Starting from these descriptions, but taking into consideration a variety of collected data for these cohorts (genetic, brain activity, physiological etc., pointing out to possible underlying mechanisms), new clusters may emerge that are perhaps more homogeneous than the original classification (Figure 4.2). Also, clustering methods like **DBSCAN (Density-Based Spatial Clustering of Applications with Noise**) have not been evaluated yet. This density-based clustering algorithm does not require specifying the number of clusters. Instead, it relies on the density of data points to form clusters, making it robust to noise and capable of finding arbitrarily shaped clusters (Ester et al., 1996).

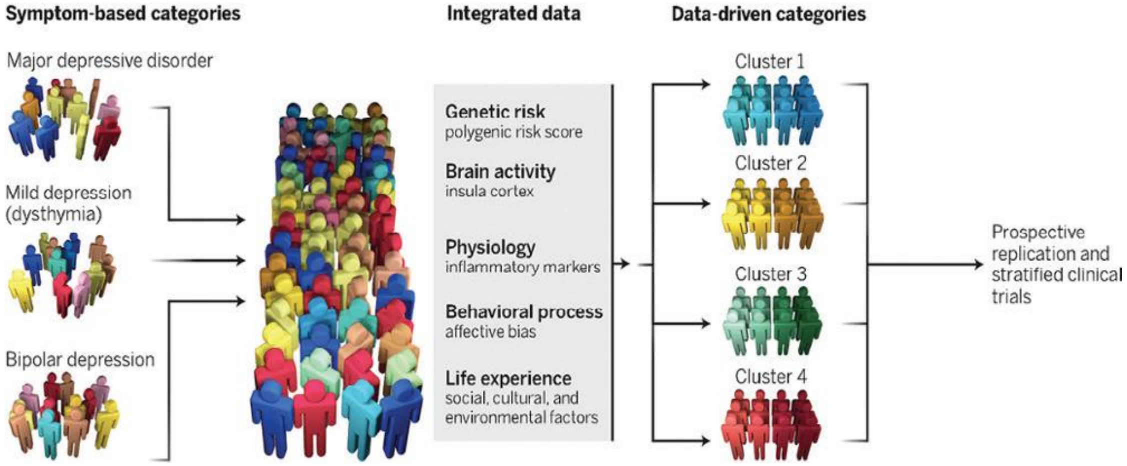


Figure 5.2 A hypothetical example illustrating how precision medicine might deconstruct traditional symptom-based categories. Adapted from Insel & Cuthbert (2015).

Finally, technical advances and large-scale neuroimaging data sets have allowed for the development of models capable of predicting individual differences in traits and behavior using ***brain connectivity*** measures derived from neuroimaging data. These mainly data-driven models, based on regression methods, use approaches such as principle component regression (PCR), connectome predictive modeling (CPM) and canonical correlation analysis (CCA) for ***linking brain and behavior***, i.e. creating models able to generate predictions of behavioral measures in novel subjects based on brain connectivity data (Rutherford et al., 2021).

5.2 Theory-driven modeling

In contrast to theoretically agnostic models discussed above, theory-driven models rely on existing theoretical knowledge (from brain anatomy and/or physiology to higher level functions such as mechanisms of perception, learning or decision-making) to test particular hypotheses about psychiatric phenomena against experimental data. In case of discrepancies between the two, assumptions can be made that there are some hidden/unobserved variables that may account for observations, thus pointing to gaps in the knowledge. Huys et al., 2016 propose three broad groups of theoretically-driven models: ***synthetic*** (biophysically realistic neural-network models), ***algorithmic*** (RL models) and ***optimal*** (Bayesian) models. They will be presented in further detail below.

5.2.1 Synthetic models

Synthetic, biophysically realistic neural-network models are commonly used to elucidate how biological abnormalities found in mental health disorders affect neurobehavioral dynamics. Therefore, they are the most intuitive and straightforward in terms of model building. These models are validated by qualitatively examining their predictions, which may include multiple levels of analysis (e.g., neural activity and behavior). If the biological mechanism is too complex or realistic and if it requires a multitude of parameters to answer the scientific question at hand, this increases computational power required for the task, so a more reductionist approach might be needed. However, despite these limitations, synthetic models have been successfully used for explaining the disturbances in OCD, schizophrenia, addiction and the like. For example, attractor models[36] have been used to explore the effects of glutamatergic and serotonergic disturbances in OCD. Decreased levels of serotonin and increased levels of glutamate, two suspected abnormalities in OCD, led to the strong and persistent activity patterns towards which the network tended to settle (and could not get out of). Models of control subjects were reported to be able to flexibly switch to a new stimulus,

[36] Attractor network is a network of nodes (i.e., neurons in a biological network), often recurrently connected, whose time dynamics settle to a stable pattern (Eliasmith, 2007).

while in models of OCD patients, obsessive thoughts proved to be resistant to switch (Huys et al., 2016).

Synthetic or neural network models originate from the connectionist framework in cognitive science. Donald Hebb introduced the term "connectionism" to describe the set of approaches that models mental or behavioral phenomena as emergent processes in interconnected networks of simple units, i.e. in neural networks comprised of simplified models of neurons. This eventually led to the idea that possible impairments of the cognitive function, such as those observed in mental health disorders, could be explained by impairments in either the structure or the elements of the underlying neural networks (e.g., the destruction of certain connections or an increase of noise in some nodes). For example, patients with schizophrenia or mania experience hallucinations and delusions, as well as rapidly changing, loose associations in thought and speech. Working under the assumptions made in Hopfield networks, the increase in noise may lead to less specific (broadening of associations) and less stable (constantly altering) memories. Similarly, destruction of connections, which resembles excessive pruning, or overload of network with memories, produces localized, spurious attractors, which correspond to hallucinations or delusions[37] (Hoffman & Mcglashan, 2001).

Another intriguing approach to tackle molecular processes involved in psychiatric and neurological disorders is computational neurogenetic modeling (Benuskova & Kasabov, 2007). The authors outlined how expression of genes that code for proteins which neurotransmitter receptors and ion channels are made of, can be linked to parameters of model neurons. In addition, they took into account that genes do not work in isolation but instead they themselves are nodes in the internal gene-protein regulatory networks that have their own temporal dynamics. Thus, the expression of genes and consequently concentrations and properties of neuronal proteins are not constant, but instead a complex function of intracellular and extracellular influences. In line with this approach, Mäki-Martunen et al. (2024) used biochemically detailed computational modeling of synaptic plasticity to investigate how schizophrenia-associated genes can affect synaptic plasticity in the cortex. They showed that the gene expression alterations lead to impaired protein kinase A - pathway and consequently

[37] Excessive pruning caused the network to produce percepts spontaneously, that is, in the absence of inputs, thereby simulating hallucinations. *Note*: Delusions are distorted beliefs, while hallucinations are imaginary sensations (visual or auditory).

to specific changes in characteristics of synaptic plasticity. Such models provide insights into possible genetic mechanisms for plasticity impairments in mental health disorders, which in turn can lead to improved understanding of their mechanisms, and ultimately their pharmacological treatment.

5.2.2 Reinforcement learning models

Reinforcement learning (RL) is a field that spans mathematical psychology, artificial intelligence, operations research, statistics and control theory. Reinforcement learning (algorithmic) models address how an agent (in either natural or artificial system) optimizes behavior in a complicated environment, that presupposes transitions between states, i.e. how it can learn to gain rewards and avoid punishments. When applied to psychiatry, dysfunctional behavior can be understood in terms of flaws, inefficiencies, or miscalibration of RL mechanisms. RL approaches have been applied to the issues of affect, motivation and emotional decision-making in psychiatry[38].

These models are usually simpler than synthetic models, with comparatively smaller number of parameters and are typically validated through quantitative statistical means. They are useful for measuring hidden variables and processes that are difficult or impossible to measure directly (Huys et al., 2016). There are three main control systems in RL framework: (1) model-based, (2) model-free and (3) Pavlovian. Pavlovian control involves involuntary actions on the basis of prediction of outcome, whether or not the actions are appropriate for gaining or avoiding consequences. On the other hand, model-based and model-free systems link the choice of actions directly to affective consequences[39]. Model-based systems are computationally costly, they make predictions based on the previously built internal model of the environment (a form of cognitive map), are thought to capture goal-directed actions and

[38] Learning and decision-making are highly intertwined processes. If learning mechanisms are impaired, maladaptive decisions will be taken, which in turn will influence what will be learned. Also, decision-making involves the accumulation of evidence associated with the utilities of possible options and choosing one based on the evidence. This is similar to the basic concept of drift diffusion models (DDMs), a process of making a decision between two choices based on accumulation of evidence toward one of the possible outcomes. A decision is made when the accumulation process reaches a certain threshold (Seriès, 2020).

[39] These two approaches reflect the difference between the classical (Pavlovian) conditioning and operant (instrumental) conditioning. Classical conditioning involves approaching or withdrawing behaviors, which is why it appears to be involuntary, while instrumental conditioning requires taking action towards optimizing the behavior.

rely on cognitive and limbic cortico-striato-thalamo-cortical (CSTC) loops. Conversely, model-free systems learn values by iteratively updating them with prediction errors through experience, are thought to capture habits and rely on sensorimotor CSTC loops (Huys et al., 2016). However, both are used for choice valuation. Table 4.1 summarizes the differences between these two approaches.

Table 5.1 Differences between model-based RL and model-free RL approach

Model-based RL	**Model-free RL**
Building a statistical model of the environment	Without building a model, relies on trial-and-error
Huge memory and computational resources	Lower demands on memory and computation
Improves predictions by optimizing the model, flexible	Less flexible to changes in environment
Goal-oriented behavior	Habit-forming behavior
Forward planning	Decision is made based on present state

Distinctions between model-based and model-free learning appear to be especially relevant for psychiatry. It has been proposed that addictive and compulsive disorders might involve a shift from model-based to model-free decision-making, which could explain inflexible behavior in patients. The process of RL presumes that learning is guided by a signal called reward prediction error (RPE), which represents the difference between the actual and expected reward. In the brain, this signal is decoded by the firing of dopamine neurons (Montague et al., 1996; Montague et al., 2004). The link between dopamine and prediction error has important consequences for understanding maladaptive behaviors such as addiction. As most addictive substances release dopamine, they may boost learning based on RPE and speed up the establishment of drug-related habits. In other words, dopamine release may interfere with the RPE signal by giving increasingly higher values to actions leading to the obtaining of the drug (Seriès, 2020).

The notion of reward is one of the central concepts in RL and sensitivity to reward seems to be altered in many psychiatric conditions. For example, anhedonia is a common feature of depression and it is generally defined as the inability to feel pleasure in normally pleasurable activities. Therefore, anhedonia may be related to impairments in the motivation, and consequently, exerting effort to obtain a reward. Another, slightly more complicated example, is the interpretation of impulsivity in ADHD as the reduced delay aversion to over-discounting of delayed rewards (Sonuga-Barke, 2003).

5.2.3 Bayesian (optimal) models

Bayesian (optimal) models attempt to link observed behavior to the Bayes-optimal solution of the problem. Bayesian decision theory (BDT) allows for formulating optimal behavior during a task and then analyzing how suboptimal behavior can arise.

These models, similar to RL models, can be used for quantitative assessment of differences between controls and patients. The central idea of Bayesian models applied to psychiatry is that internal models of patients, in particular their prior beliefs, differ from those in healthy subjects. For example, positive symptoms of schizophrenia, i.e. hallucinations and delusions (discussed above) can also be related to imbalance between incoming sensory information and prior beliefs and expectations[40]. Similarly, in autism, it has been proposed that prior expectations might be attenuated compared to actual sensory inputs, which might explain why patients experience the environment as overwhelming and "too real" (Seriès, 2020).

Following the similar vein, another question that might be answered by Bayesian models is whether a given symptom is related to suboptimal inference. For instance, a study conducted by Browning et al. (2015) showed that subjects with high trait anxiety cannot update optimally on how volatile an aversive situation is, while low anxiety controls presented close to Bayes-optimum behavior. Therefore, this element of ***uncertainty*** (of the environment) is another important aspect of Bayesian models. It has been shown that the statistics of aversive experience play an important role in several processes, from learned helplessness and depression to familiarity in fear conditioning.

[40] Within Bayesian framework, hallucinations and delusions are explained as the disruption in the mechanism for the minimization of prediction error. It is assumed that this mechanism applies to both perception (hallucinations) and beliefs (delusions). In hallucinations, it is common for patients with schizophrenia to attribute agency to external forces, thereby perceiving absent sensory input (e.g., voices, inner speech) as externally generated (Fletcher & Frith, 2009).

Additional note – Integrating approaches

As we have seen from the examples presented above, the mechanisms of psychiatric conditions can be very complex, which is why sometimes knowledge of more than one framework is needed to explain the entire process. For example, ***learned helplessness***[41] is one of the prominent features of anxiety and depression. However, if we want to explain it in computational terms, we can conceptualize this initially as impaired learning due to inability to consistently predict outcomes of one's actions, because of uncontrollable rewards and punishments. The onset of learned helplessness (as a kind of conditioned response) is marked by unwillingness to explore or making no attempts to escape when placed in a new environment, even if the options are available to the subjects. However, the whole process can be observed from Bayesian perspective as well: the prior belief that the environment is uncontrollable will discourage exploration. Additional interpretation is that due to the impairment, prior beliefs are not updated correctly based on new information, i.e. new information does not have any discriminatory value for the subject. Also, in the context of RL, the exploration-exploitation dilemma seems to be attenuated (or non-existent), as learned helplessness can be seen as avoiding exploration in favor of the current unsatisfactory situation (Teodorescu & Erev, 2014). Furthermore, it would also be interesting to explore how a one-time traumatic event and one of the possible responses (e.g., freeze[42]) is different from the repeated exposure to aversive stimuli that leads to learned helplessness.

[41] ***Learned helplessness*** is a behavior exhibited by a subject after enduring repeated aversive stimuli beyond their control. It was initially studied in animals undergoing ***experimental neurosis*** (induced by presenting them with an insoluble learning problem or subjecting them to inescapable electric shocks). In the 1970s, Martin E. P. Seligman extended the concept from nonhuman animal research to clinical depression in humans and proposed a learned helplessness theory to explain the development of or vulnerability to depression. According to this theory, people repeatedly exposed to stressful situations beyond their control develop an inability to make decisions or engage effectively in purposeful behavior (APA Dictionary of Psychology, 2014).

[42] The link seems to be the ability of vmPFC to turn off the dorsal raphe nuclei response (which stimulate amygdala and sensorimotor cortex, i.e., „freeze“ response) if and when we experience that taking purposeful action leads to a desired result (Maier & Seligman, 2016) .

5.3 Combined approach

Based on the approaches presented above, it seems that there is a tendency to use theoretically agnostic ML approaches when developing clinically useful applications, and theory-driven approaches when trying to better understand the mechanism of the studied disorders (Huys et al., 2016). As it was previously mentioned (see Figure 4.1), these approaches are not mutually exclusive, but rather complementary. Existing theoretical knowledge can significantly reduce the dimensionality of the data set (and capture non-random variation in the data), thus preparing the data for the application of ML techniques. According to some findings, combining theory-driven and data-driven approaches can outperform data-driven approaches alone (Rutledge et al., 2019, Huys et al., 2016).

5.4 Computational model building and development

In this subsection, we will present steps that are necessary for model development in CP as well as to point out to some problems or challenges in this process. Irrespective of the choice of approaches outlined above (theory-driven, data-driven or combined), the process of model building needs to start with the formulation of the research question or hypotheses and identifying variables and their relationships (based on existing psychiatric theories and theoretical frameworks). Any computational model may be described as trying to capture associations (formulated as mathematical equations) between a set of input variables and one or more output variables. In psychiatry, the example of input variable could be neural activity or self-report data, while diagnosis or treatment response would correspond to output variables[43]. Furthermore, computational models can quantify how well these associations are reflected by output variables (i.e. model fit).

However, before the incorporation of data into the model, they need to be collected and preprocessed. Data collection process and possible problems with data have been outlined in the previous chapter. In the preprocessing step, some issues that need to be addressed include noise and artifact removal (applying techniques such as PCA/ICA), standardization and

[43] Both input and output variables can be ***numeric*** (e.g., continuous, such as duration of treatment), ***categorical*** (e.g., whether an individual will develop a disorder or not) or ***complex*** (e.g., text strings).

normalization of data[44], handling missing values and transforming complex data (e.g., dimensionality reduction for neuroimaging).

Depending on the research question, the key step and challenge is to determine the ***right level of abstraction*** (from modeling impairments on the level of ion channels to interactions between different brain regions or even the whole brain connectivity). It is also possible to move between different levels of abstraction, allowing models to map processes spanning different layers of disease pathology (Hauser et al., 2022). This obviously has implications for incorporating data from different brain imaging modalities.

Modeling process should contain three general steps: building a model, simulating the model with artificial data (simulation) and applying the model to real data (validation) (Seriès, 2020). Furthermore, a model can then be compared to other models in the process of model comparison. ***Model comparison*** involves evaluation of which set of possible models best describes the data, as a way to understand which mechanisms are more likely to underlie the behavior (Wilson & Collins, 2019). Essentially, we are trying to evaluate differences and relative performance of the models, e.g., by using statistical methods. However, it should be noted that accuracy of a model is not a necessary indicator of its clinical utility. It is also important to consider the ***interpretability*** of the models which are not easily quantifiable.

After data collection and preprocessing (ensuring the adequate quality of data), ***model features*** need to be established. These are data or aggregated substrates thereof which are used to train a computational model to predict a ***label*** (outcome variable in a supervised model). A feature is any characteristic that can be extracted from the data and that is believed to be informative about the class labels (Wolfers et al., 2015). Sometimes, it is also necessary to select certain features (***feature selection and extraction***)[45]. ***Model fitting*** is a process of

[44] Normalization refers to rescaling data to a specific range, while standardization transforms data to mean 0 and variance 1 (when data has varying distributions and requires Gaussian-like scaling in order to be similar to normally distributed data).

[45] According to Gao et al. (2018), *feature selection* and *feature extraction* can be grouped into feature reduction methods. *Feature selection* is performed in supervised models when the most discriminant features are selected with the help of labels in the training data to reduce noise. One strategy is to use prior knowledge to decrease dimensionality. *Feature extraction* occurs when the original high-dimensional data is projected onto a lower dimension while maintaining the feature's discriminative abilities. One typical example is PCA.

finding model parameters[46] so that the model's predictions maximally match the data (parameter optimization). ***Parameter optimization*** is a set of procedures for finding a set of parameter values in the model that will maximize the model's objective function (e.g., likelihood in probabilistic models) or that will minimize the error between the model predictions and actual data (Hauser et al., 2022).

It is useful to simulate the data on several models before running the model on real data. ***Simulation*** involves running the model with particular parameter settings to generate "fake" data. Simulation is a way to make theoretical predictions more precise and testable (Wilson & Collins, 2019). Essentially, we are trying to establish whether the model can answer the question in theory. If the answer to this question is satisfactory, one can proceed to validating the chosen model against real data. A model is chosen in a process of model selection. It considers the model fit and the model complexity, to avoid ***underfitting*** or ***overfitting***[47]. Another, more general principle that is implicated in model building is Occam's razor, i.e. the fact that a simple theory or model is favored over a complex one, if the former can explain the phenomena and capture the data adequately (Hauser et al., 2022).

Datasets are split into two subsets: a training data set, which is used to estimate the model parameters, and a validation data set, which is used to test how well those parameters predict "new" data. It is crucial that the model's performance is validated against an independent test dataset, usually by applying procedures such as ***cross-validation***[48]**.** It is essential that the training and testing set are kept independent from one another to avoid overfitting. Validation enables us to establish whether the model can account for the new data.

Generalisability refers to the ability to use models beyond the data that were used to develop the original model (i.e., predicting the labels correctly in new data). This is crucial for

[46] Parameters represent aspects of the model that control how it behaves or predicts outcomes (e.g., a learning rate in reinforcement learning model or the connectivity weights in a neural network).

[47] Bias-variance trade-off is a conflict between two types of errors in computational model development. High bias arises due to underfitting, when the model is not capturing relevant associations between features and output labels. Variance error arises when a model is overfitting the training set and interprets random noise as meaningful variation, thus generalizing poorly on new data (Hauser et al., 2022).

[48] If the validation procedure is repeated using multiple different training and test partitions, the procedure is called *k-fold cross-validation*, where *k* denotes the number of data partitions (folds). The special case, where *k* is equal to the number of samples is referred to as leave one out-cross validation (LOO-CV) (Wolfers et al., 2015). By averaging the results from all the tests we get a more reliable estimate of how well the model performs.

the clinical success of modeling efforts, i.e. for translating the findings into clinical practice. There are several instances in which such procedures have proved to be useful: (1) predictive use (a model is used to predict clinical outcomes such as risk of relapse or treatment response), (2) personalized interventions (tailoring interventions based on individual model parameters) and (3) decision support (e.g., for treatment selection).

It should also be pointed out that modeling is an ***iterative process***, i.e. model may require revision or refining based on validation results, new data, or feedback. For instance, this can be done through latent variable analysis or inference (e.g., adding a latent variable for attention in a model of decision-making). Also, modifications can consist of updating theoretical assumptions if the model reveals some unexpected insights.

5.5 Areas of application of CP models

Hauser et al. (2022) propose four application areas of computational modeling in psychiatry: uncovering the mechanism (theory-driven approach), subtyping, status prediction and treatment stratification/selection (data-driven approaches).

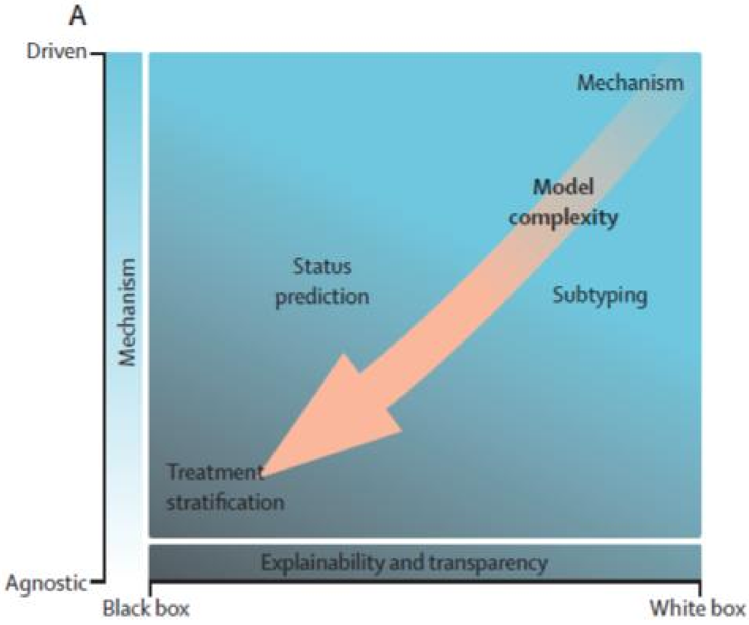


Figure 5.3 Correlations between model characteristics and explainability. Models differ in the transparency of mechanisms, which determines their best use. Although most complex models often achieve higher predictive performance, "white box" models allow for an understanding of the underlying mechanisms. Adapted from Hauser et al. (2022).

Mechanism-agnostic (i.e. data-driven) models are termed "black box" in ML because they provide no information on how input variables meaningfully relate to or explain output variables. On the other hand, mechanism-driven models, also known as "white box" or "glass box" models, enable understanding of the underlying mechanisms (Figure 6.1). For mechanism-agnostic models, the key challenge is understanding *how* these models operate and *what* they predict. Their complexity renders them opaque, but there are ways to move from "black box" to "grey box" models. For example, this can be done with the use of causal ML models that allow advancement beyond simple correlational effects, and thus improving interpretability (e.g., XGBoost[49]). For mechanism-driven models, the biggest challenge is their predictive performance. One solution is using mechanism-driven algorithms as a dimensionality reduction step before the subsequent generation of optimally predictive mechanism-agnostic models.

6 Overview of models applied to depression

This section outlines some applications of CP models to the study of depression. We start by providing the description of the clinical picture of major depressive disorder, followed by the description of the neural bases of the disorder. Various aspects of depression can be explored via RDoC matrix. Theory-driven models will be used to explain or interpret the aspects of depression such as anhedonia, rumination, cognitive deficits and learned helplessness. Data-driven approaches will be used to illustrate the most interesting findings arising from some of the most pressing issues in clinical practice.

[49] XGBoost (Extreme Gradient Boosting) allows better interpretability due to its tree-based structure and measures of feature importance, among other characteristics. The structure of sequentially built decision trees can be visualized to understand how specific features influence predictions. Metrics such as gain, frequency and weight allow users to evaluate the relative importance of features in the model (Sagi & Rokach, 2021). This approach is particularly useful for healthcare providers who have to make informed decisions about patients.

6.1 Depression – a clinical picture

Major depressive disorder (MDD) represents the classic condition in the group of depressive disorders outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR, 2022). The manual lists two core symptoms of MDD: depressed mood and loss of interest or pleasure (anhedonia), of which at least one needs to be present for diagnosis. Other symptoms (somatic, emotional and cognitive) include significant fluctuations of body weight, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or guilt, problems in thinking and concentrating, and recurrent thoughts of death or suicide. Overall, five or more symptoms have to be present for at least two weeks (most of the day, nearly every day), cause significant disturbances in daily functioning, and should not be better explained by other disorders. A diagnosis based on a single episode is possible, although the disorder is a recurrent one in the majority of cases. Careful consideration should be given to the discrimination of normal sadness and grief (due to adverse life events) from MDD. A chronic form of depression, persistent depressive disorder, can be diagnosed when the mood disturbance continues for at least two years in adults and one year in children.

6.2 Neuroscience of depression

As we have seen from the symptoms listed above, depression can have a very heterogeneous representation. Despite being considered primarily a mood disorder, depression is also characterized by cognitive and decision-making deficits. Therefore, processes involving emotional regulation and processing, memory and executive function are usually impaired in people suffering from depression. These dimensions of functioning are associated with certain brain structures. To get the overview of the neural correlates of depression, the following structures need to be considered: certain brain regions, brain networks, neurotransmitter systems and structural abnormalities.

Brain regions

Most common brain regions implicated in depression are summarized in Table 6.1.

Table 6.1 Most common brain regions implicated in depression.

Brain region	**Role**	**Findings in depression**
Prefrontal cortex (PFC)	Executive function, emotion regulation, decision-making	Reduced DLPFC activity, increased VMPFC activity
Amygdala	Emotional reactivity	Hyperactive to negative valence stimuli
Hippocampus	Learning, memory, cognition	Reduced volume, impaired function (both working and episodic memory)
Anterior Cingulate Cortex (ACC)	Emotion cognition and integration	Hyperactive sgACC
Striatum	Reward processing	Hypometabolism, anhedonia
Hypothalamus	Control of HPA axis	Altered response to stress (stress → anxiety → depression)
Orbitofrontal cortex (OFC)	Assessment of stimulus value and reward, representation of internal values	Negative sense of self

The table contains main brain regions implicated in the depression, their corresponding roles and findings in depression. Abbreviations: **DLPFC** (dorsolateral prefrontal cortex), **VMPFC** (ventromedial prefrontal cortex), **sgACC** (subgenual anterior cingulate cortex)

Brain networks

Li et al., (2018) point out to the involvement of at least four networks in patients with depression (see Figure 6.1). Elevated connectivity of a ventral limbic affective network (AN) appears to be associated with excessive negative mood (***dysphoria***) in patients. Decreased connectivity of a frontal-striatal reward network (RN) has been suggested to account for the loss of interest, motivation and pleasure (***anhedonia***). Enhanced default mode network (DMN) connectivity seems to be associated with depressive ***rumination***. Finally, diminished connectivity of a dorsal cognitive control network (CCN) is related to ***cognitive deficits***, especially ineffective top-down control of negative thoughts and emotions.

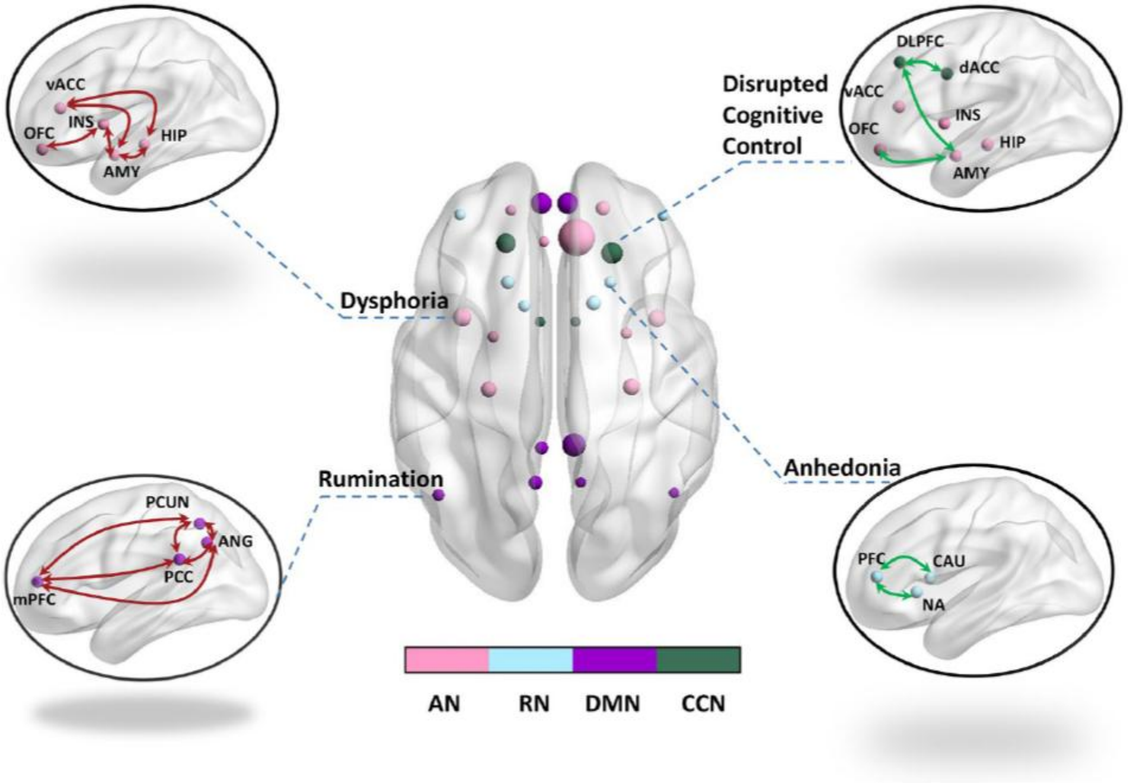


Figure 6.1 Major brain networks implicated in depression.
Affective network (AN) (hyperconnectivity, (red)); Reward network (RN) (attenuated connectivity (green); Default mode network (DMN) (hyperconnectivity, (red)); Cognitive control network (CCN) (attenuated connectivity (green)).
Main structures: **OFC**: orbitofrontal cortex; **INS**: insula; **AMY**: amygdala; **HIP**: hippocampus; **vACC**: ventral anterior cingulate cortex; **mPFC**: medial prefrontal cortex; **PCC**: posterior cingulate cortex; **PCUN**: precuneus; **ANG**: Angular; **DLPFC**: dorsolateral prefrontal cortex; **dACC**: dorsal anterior cingulate cortex; **PFC**: prefrontal cortex; **CAU**: caudate; **NA**: nucleus accumbens. Adapted from Li et al. (2018).

Neurotransmitter systems

Monoamine neurotransmitters, such as serotonin (5-HT), norepinephrine (NE) and dopamine (DA), are implicated in different processes in depression. Noradrenergic neurons in the brain form a neurotransmitter system, which when activated, exerts effects on large areas of the brain. The effects are manifested in alertness, arousal, and readiness for action. Dopamine is associated with reinforcement learning processes, while serotonin is linked to processing of aversive stimuli. Learned helplessness and depression may be related to a failure to stop such aversive processes. Antidepressant medications (SSRI/SNRI) commonly target serotonin or norepinephrine receptors by inhibiting the reuptake of 5-HT and NE (Cui et al., 2024).

Structural abnormalities

Large meta-analyses have concluded that MDD is associated with alterations in cortical thickness, especially in OFC. Cortical thickness measurements showed greater differences than surface area measures in adult MDD, but consistent surface area deficits were found in adolescent MDD. Cortical thickness and surface area represent distinct morphometric features of the cortex and may be differentially affected by depression at various stages of life (Schmaal et al., 2016). Decrease in the total gray matter volume (Grieve et al., 2013) and alterations in cortical thickness are the two most consistent findings in structural neuroimaging studies.

6.3 Exploring depression with RDoC matrix

Depression is a highly heterogeneous mental health disorder, in terms of etiology, symptoms (varying degrees of severity and levels of dysfunction) and underlying neural bases and mechanisms. RDoC framework has two main advantages for studying depressive disorders: (1) it encourages interdisciplinary research, so that the systemized knowledge from fields such as neuroscience, psychology and genetics contributes to a more complete picture of the disorder; (2) it provides transdiagnostic insight, i.e. it enables comparison with other similar disorders and identifies overlapping areas[50].

There have been various attempts to apply RDoC framework to the study of aspects of depression (Woody & Gibb, 2015; Gibb et al., 2015; Park & Kim, 2021). Following the general RDoC principle, we can try to conceptualize depression as a deviation from otherwise normal dimensions of psychological functioning. In the negative valence system, researchers have been focused on the hypertrophy of the feelings of loss, while maladaptive reward patterns have been studied within the negative valence system. Depression is characterized by the impairment of cognitive systems, particularly in the area of working memory and executive control (ineffective top-down control of negative thoughts and emotions). In the sensorimotor domain, there are also two dimensions: psychomotor retardation or agitation, as

[50] We may conceptualize this similarly to the within- and between-subject (disorder) research design, both of which can potentially provide clinical utility (e.g., improved classification).

well as irregularity in sleep-wake patterns (either insomnia or hypersomnia (and other impairments along the continuum)).

Regarding different units of analysis, depressive disorders have been linked with a variety of indicators. For example, at the genetic level, research focuses on the genes that are known to regulate neurotransmission of monoamines including serotonin, dopamine and norepinephrine (e.g., 5-HTTLPR, 5-HT receptor genes, MAOA and COMP) (Woody & Gibb, 2015)[51]. On the cellular level, depression is characterized by deficits in synaptic plasticity (Duman et al., 2016). Studies that investigate the influence of antidepressant (AD) therapy on synaptic plasticity reveal that the increase in number of extrasynaptic receptors leads to enhancement of synaptic efficiency by promoting new synapse growth. These are adaptive responses of neurons to the decrease in neurotransmitter levels. However, these newly formed synapses are either stabilized if they receive appropriate synaptic input, or pruned if they are inactive. The desired effect is the enhancement of synaptic transmission efficiency, which is crucial for alleviating depressive symptoms and better understanding of the AD mechanism (Beňušková, 1991; Castrén & Hen, 2013). At the level of neuronal circuits, depressive symptoms have been linked to altered functional and effective connectivity[52] (either enhanced or diminished). Enhanced connectivity of a ventral limbic affective network seems to be associated with excessive negative mood (dysphoria), while diminished connectivity of a frontal-striatal reward network has been suggested to account for the loss of interest, motivation and pleasure (anhedonia) (Li et al., 2018).

At the physiological level, peripheral measures of autonomic nervous system (ANS), hypothalamic-pituitary-adrenal (HPA) axis, and neuroimmune dysregulation seem to be connected to depression (Woody & Gibb, 2015). In terms of observable behavior, many features are congruent with symptoms listed by the DSM (anhedonia, social withdrawal, fatigue and low energy, deficits in cognitive functioning etc.). Finally, the self-report level of analysis highlights attributional styles and hopelessness.

[51] Some research suggests that joint consideration of multiple genetic and environmental factors has much greater explanatory power than separate studies of genetic or environmental causation. Multi-factorial gene-environment interactions (GxE) are likely to be a generic mechanism involved in the majority of cases of mental illness (Uher & Zwicker, 2017).

[52] Functional connectivity refers to (undirected) correlations between the activity of two brain regions, while effective connectivity refers to (directed and usually reciprocal) causal inferences among brain regions within a network (Li et al., 2018).

6.4 Theoretically-driven models applied to specific aspects of depression (rumination, anhedonia, learned helplessness, cognitive deficits)

This section will provide some examples of different types of theory-driven models that have been applied to specific aspects of depression. These are either symptoms (such as anhedonia) or other constructs associated with the disorder (learned helplessness) that are helpful for understanding the mechanisms and impairments in depression, and possibly determining approaches to treatment.

Rumination

Rumination may be defined as a tendency to repetitively think about the causes, situational factors, and consequences of one's negative emotional experience. This aspect of depression has been shown to predict the onset of depression, prolong the duration, exacerbate negative thinking, and impair problem-solving. Hyperconnectivity of the DMN[53] may represent excessive self-referential processes and maladaptive rumination in patients.

Siegle and Hasselmo (2002) provided an example of how a neural network, i.e. connectionist models, can be used to better understand deficits in depression during (negatively biased) emotional information processing. A neural network model was used to simulate classification of emotional word stimuli (labeled positive, negative or neutral). It was able to reproduce typical behavior in depression, i.e. it was quicker to identify negative information and showed larger sustained activity when confronted with negative words[54]. The authors proposed ***overlearning of negative information*** as one of the mechanisms related to rumination. A network that had "overlearned" on negative information could be retrained using positive information (corresponding to CBT), which resulted in normalization of network activity in response to negative content.

There is another interesting link between rumination and certain cognitive deficits, i.e. working memory. Namely, depressive rumination increases cognitive load, reducing available resources for working memory tasks, resulting in impaired attention and concentration,

[53] Elevated DNM functional connectivity appears to be a robust marker of MDD that is evident even in remitted and recovered state.

[54] Depressed individuals have been shown to pay excessive attention to negative information, ruminate about it, preferentially remember it, and interpret information as negative (Siegle & Hasselmo, 2002).

reduced processing speed, difficulty in task switching (cognitive flexibility) and poor recall and retention of information (Onraedt & Koster, 2014).

Anhedonia

Anhedonia is one of the core symptoms of depression. Evidence from neuroimaging studies suggests that anhedonia may be attributed to diminished connectivity in frontal-striatal reward network[55] (Li et al., 2018). On the neuronal level, prediction-error signals appear to be reduced in the striatum and other dopamine-rich regions of the brain, thereby suggesting that depressive symptoms are associated with impairment in encoding of reward-learning signals.

In CP, anhedonia has primarily been studied within the RL framework. Temporal-difference (TD)[56] prediction-error learning signals have been linked to the firing of dopamine neurons in the brain (Montague et al., 2016). Kumar et al. (2008) found blunted reward-prediction error signals in patients versus controls, as well as the correlation between such blunting and illness severity.

In a meta-analysis by Huys et al. (2013), different variants of RL models were used to explore different mechanisms in anhedonia. The goal was to find out whether anhedonia was associated with the initial rewarding experience of stimuli, or the subsequent learning from these rewards. Two mechanisms are important to disentangle, as they probably correspond to different etiologies and possible strategies for therapies. The result suggested that reward sensitivity rather than learning rate is primarily impaired in this phenomenon. The model also allowed them to make a distinction between the absence of reward and punishment, i.e. that participants could interpret the absence of reward on a given trial as punishment. They did that by including a punishment sensitivity parameter in the model.

Learned helplessness

For the discussion on learned helplessness, please see 4.2.3 (*Additional note – Integrating approaches*).

[55] Reduction in connectivity has been found to be in proportion to depression severity and an important predictor of depressive relapse.

[56] TD learning is an unsupervised technique in which the learning agent learns to predict the expected value of a variable occurring at the end of a sequence of states.

Cognitive deficits

Negative perceptual and cognitive biases

Commonly observed pessimistic cognitive biases in depression have been explained using prior beliefs within the framework of Bayesian decision theory. Huys, Vogelstein, and Dayan (2009) fitted a Bayesian learning model to the behavior of depressed and healthy participants and included two parameters in the model: sensitivity to reward and a prior belief about control (i.e. helplessness). Individuals with strong priors (i.e. belief that they have the control over their environment) would predict that previously rewarded actions will likely be rewarded again, while depressed patients would expect weaker associations between actions and rewards. The formulation of the model enabled them to use a simple linear classifier to distinguish between healthy and MDD population based purely on behavioral measures, and avoiding any verbal reports. This obviously has important implications for the psychiatric practice, since reliable classification is one of the important goals in CP.

Deficits in executive function

Dillon et al. (2015) used the drift diffusion models (DDMs) to explore a seemingly counterintuitive notion that enhanced executive functioning in depression is sometimes observed during tasks that require careful thought and precision. Depression can lead to increased analytical information processing (similar to rumination), yielding worse performance in tasks requiring fast decisions, but higher accuracy in more detail-oriented tasks. Drift rate for the executive control mechanism was lower, but there was an additional decreased drift rate in the reflexive mechanism (signaling to inhibition). In other words, they found that patients were more accurate but slower on trials with incongruent stimuli. This approach enabled the study of the regulation of speed-accuracy trade-offs in depression.

Memory deficits

Various memory impairments are common in depression. For example, episodic memory is disrupted in unipolar depression. Depressed individuals typically show impaired recollection, biased memory performance for positive and negative events, and “overgeneral” autobiographical retrieval (Dillon & Pizzagalli, 2018). However, these deficits seem still largely unexplored computationally (Seriès, 2020). Additionally, there is some research that suggests that cognitive deficits persist even in remitted patients. In their meta-analysis, Rock et al. (2014) demonstrate that cognitive impairment represents a core feature of depression, and not an epiphenomenon secondary to symptoms of low mood.

Other aspects of depression

Other aspects of clinical symptoms in depression, such as ***bodily or somatic symptoms*** and ***social deficits***[57], are also central to the disorder yet remain underinvestigated with computational modeling (Saez & Gu, 2022).

In addition to symptom overlap between the disorders and their heterogeneous presentation, difficulties in studying mental health disorders also arise from the need to decompose and analyze some of the more complex constructs associated with the disorder. Including them in computational models enables testing some very specific hypotheses about various facets of those constructs or their possible interaction (e.g., reward sensitivity and control over priors).

Additional note

In the majority of RL studies, rewards were presented as monetary gains or points. Considering the fact that motivation and exerting effort are also compromised in depression, it would be interesting to explore whether a more (emotionally) salient reward would correspond to a more ecologically valid representation of a rewarding experience, and whether it might consequently affect learning.

6.5 Data-driven approach: which models are suitable for predicting the treatment outcomes?

Antidepressant treatment efficacy is low and it usually involves a process of trial-and-error to achieve adequate responsiveness in patients. All this delays clinical improvement and increases risks and costs of treatment. Chekroud et al. (2016) developed an ML model to predict whether patients would achieve clinical remission from MDD after a 12-week course

[57] Social skill, defined as the emission of behaviors which are positively reinforced by others, is seen as an area of deficit especially important in the development of depressive behaviors (Lewinsohn & Atwood, 1969)

of citalopram. The model was trained on data from STAR*D and identified 25 variables that were most predictive of treatment outcomes from a total of 164 patient-reportable variables. The choice of variables was one of the most important steps in the model development. Top 25 predictive items were chosen by using elastic net regularization (supervised dimensionality reduction), which is a method that avoids issues of correlated predictors and overfitting. Validation method was a repeated 10-fold cross-validation. The model demonstrated statistically significant predictive accuracy, achieving an internal validation accuracy of 64.6% in the STAR*D cohort ($p<0.0001$). It was also externally validated in the COMED[58] trial, where it showed an accuracy of 59.6% ($p=0.043$) in the escitalopram treatment group. Ultimately, researchers came up with an ML model optimized to detect future responders for a specific, first-line antidepressant (citalopram), with a simple 10-minutes questionnaire. The model uses easy to obtain (patient-reportable) information, and could be hosted online or in a clinical setting. One of the advantages of this model is that it was developed by mining existing clinical trial data, thus reducing the time, effort and costs of data collection. Alternative method of predicting the antidepressant treatment outcome is quantitative EEG (QEEG) biomarker, the Antidepressant Treatment Response index (ATR). QEEG power in the theta and alpha frequency bands may identify patients who are most likely to respond to tricyclic antidepressants (TCAs) or SSRIs (Leuchter et al., 2009). However, it still requires EEG data collection and processing, even with a limited electrode array in the prefrontal region.

[58] STAR*D (Sequenced Treatment Alternatives to Relieve Depression) is the largest prospective, randomized controlled study of outpatients with MDD (data was collected from June, 2001, to April, 2004). COMED (Combining Medications to Enhance Depression Outcomes) was a single-blind, randomized, placebo-controlled trial comparing efficacy of medication combinations in the treatment of MDD (Chekroud et al., 2016).

6.6 Application of CP models to brain stimulation techniques (rTMS) and discovery of discriminative biomarkers

Repetitive transcranial magnetic stimulation (rTMS) is a noninvasive neurostimulation treatment for treatment-resistant depression (TRD) that modulates functional connectivity in cortical networks[59]. Although the left dorsolateral prefrontal cortex (DLPFC) is the most common target for stimulation, recent studies have demonstrated efficacy for a dorsomedial prefrontal (DMPFC) target, too. This raises the intriguing possibility that differences in dysfunctional connectivity at the DMPFC target site may give rise to different treatment outcomes. Drysdale et al. (2016) conducted an extensive study to differentiate the so-called biotypes obtained from rs-fMRI and tested their utility for predicting the outcome of rTMS treatment. Namely, by using rs-fMRI in a large multi-site sample, the authors were able to differentiate four neurophysiological subtypes ('biotypes') defined by distinct patterns of dysfunctional connectivity in limbic and frontostriatal networks. To select connectivity features for clustering, they used canonical correlation analysis to define a low-dimensional representation of these features and associated them with weighted combinations of clinical symptoms, as quantified by the 17-item Hamilton Depression Rating Scale (HAMD)[60]. This data-driven approach to feature selection and dimensionality reduction identified two sets of functional connectivity features that were correlated with distinct clinical-symptom combinations. The next step was using hierarchical clustering to discover clusters of patients, by assigning them to nested subgroups with similar connectivity patterns. Clustering patients on this basis enabled the development of diagnostic classifiers (biomarkers) with high

[59] rTMS involves the repeated application of electromagnetic pulses delivered by a magnetic coil placed on the scalp to depolarize cortical neurons and modulate neuronal activity. In large, real-world studies, response rates in TRD ranging from 50% to 80% have been reported. The most common stimulation protocol is 10 Hz stimulation applied to the left DLPFC, lasting for 20-40 minutes. (Chen et al., 2023).

[60] Hamilton Depression Rating Scale (HDRS or HAMD) is the most widely used clinician-administered assessment scale. The original version (Hamilton, 1960) contains 17 items pertaining to symptoms of depression experienced over the past week.

sensitivity and specificity[61] (82-93%) for depression subtypes in multisite validation and out-of-sample replication datasets[62].

The usefulness of the newly discovered biotypes was tested by assessing their responsiveness to rTMS treatment. The protocol (repetitive high-frequency stimulation of DMPFC for 5 weeks) was most effective for patients with biotype 1 (82.5% of whom improved significantly with more than 25% HAMD score reduction). The authors compared predictions based solely on clinical symptoms and on functional connectivity biotypes. Classification according to connectivity features plus biotype diagnosis yielded the highest predictive accuracy (89.6%).

Furthermore, the authors wanted to ascertain whether the newly discovered biotypes correspond to any other disorders similar to depression. Namely, they studied whether patients diagnosed with generalized anxiety disorder (GAD) shared similar patterns of abnormal connectivity with one or more depression biotypes. They applied optimized classifiers to GAD cohort (without overlapping clinical depression). Out of this sample, 69.2% of participants were still classified as belonging to one of the depression biotypes, with the majority of these (59.3%) assigned to anxiety-associated biotype 4. Interestingly, when applied to a group of patients with schizophrenia, which is not considered close to depression (not primarily a mood disorder), only 9.8% of patients tested positive for a depression biotype.

This pioneering study showed that computational approach can indeed bring significant advancements both on theoretical and on the more practical, translational level. On theoretical level, defining novel subtypes of an existing disorder that transcend current diagnostic boundaries enhances the understanding of how various brain dysfunctions contribute to the diverse clinical presentations of depression and discrimination from other similar disorders. On the more practical, translational level, the utility of these findings can be translated to optimization of treatment protocols.

[61] Sensitivity measures how well a test can identify true positives and specificity measures how well a test can identify true negatives.

[62] The authors also tackled the problem of the capacity of classifiers trained on one data set at a single site to generalize to data collected at multiple sites. They tested the most successful classifiers in an independent replication data set collected from 13 sites.

6.7 Looking ahead: predicting the transition to psychosis from recent onset depression (ROD)

Another concern in psychiatric practice is determining the prognosis and implementing prevention strategies for certain disorders, especially in young population. Koutsouleris et al., (2021) tried to predict transition to psychosis in patients with clinical high-risk states (CHR)[63] and recent onset depression (ROD) using multimodal ML. To that end, they conducted a multisite, longitudinal prognostic study that followed up patients with CHR, ROD and healthy volunteers. Their models integrated clinical-neurocognitive data, structural MRI data, and polygenic risk scores for schizophrenia (PRS)[64].

Clinical implementation. To facilitate clinical implementation, a *sequential prediction model* was developed that optimized the ordering and number of data modalities along with prognostic uncertainty thresholds to decide whether a patient needed further testing, thus lowering diagnostic burden. The initial condensed clinical-neurocognitive model was streamlined from 141 variables to only 7 key variables.

Cybernetic model. The authors tested the performance of the clinical-neurocognitive, sMRI-based, and PRS-based models separately, but also created a *cybernetic model* that combined all algorithmic and human components. Human input consisted of clinician-rated estimates on transition to psychosis, which showed a pronounced optimism bias toward risk estimation. However, it provided valuable contextual insight. Therefore, because algorithms showed exactly the inverted bias (high sensitivity and low specificity), the cybernetic model presented a superior predictive system. The final model, optimizing the diagnostic workflow, achieved accuracy of 85.9% (sensitivity, 84.6%; specificity, 87.3%).

Validation. Models were validated through internal and external data sets. As a part of the PRONIA initiative (Personalized Prognostic Tools for Early Psychosis Management, EU), data was available from 7 academic early recognition services in 5 European countries. The

[63] The construct of a clinical high-risk (HR) state for psychosis has evolved to capture the prepsychotic phase, describing people presenting with potentially prodromal symptoms (Fusar-Poli et al., 2013). In DSM-5, Attenuated Psychosis Syndrome (APS) was recognized as a condition for further research.

[64] A polygenic risk score (abbreviated PRS) uses genomic information alone to assess a person's chances of having or developing a particular medical condition. A person's PRS is a statistical calculation based on the presence or absence of multiple genomic variants, without taking environmental or other factors into account (Polygenic Risk Score (PRS), 2024).

study employed techniques such as nested cross-validation and label permutation testing to assess the model's performance and prevent overfitting.

Other findings – identification of risk factors and neuroanatomical biomarkers. The study identified significant predictors of psychosis transition, i.e. clinical variables such as APS, motor disturbances and non-supportive family environment during childhood out of total number of variables. Based on available sMRI data, the researchers also identified specific neuroanatomical biomarkers which differentiated participants at risk for psychosis from those that were not. Namely, a psychosis predictive brain signature was discovered that generalized well on independent cohorts. Interestingly, participants labeled as non-transitional showed reversed temporo-occipital volume reductions compared to healthy controls, which might point out to a compensatory mechanism of resilience to psychosis. Therefore, sMRI-based models may have prognostic and observational utility in clinical settings. These findings support the neurobiological proximity between the early-onset affective and psychotic disorders.

The study results suggest that an individualized prognostic workflow integrating artificial and human intelligence may facilitate personalized prevention of psychosis in young patients with CHR or ROD. Considering the fact that transition to a more serious illness, i.e. psychosis is debilitating, costly, harder to treat and potentially long-lasting, such prevention measures are certainly worth exploring.

Additional note

Sample sizes

The majority of studies based on neuroimaging data have the problem of small sample sizes. Obtaining reliable neuroimaging data is both costly and time-consuming. Small sample sizes have implications for the statistical power of the findings, generalizability of the model and introduction of bias (due to highly selective participant criteria). It is possible to pool neuroimaging data from multiple sites by means of data sharing initiatives (e.g., 1000 Functional Connectomes Project International Data Sharing Initiative). However, uniformity in respect to data acquisition and (pre)processing has to be taken into account. This can be addressed by applying adequate pre-registration procedures.

7 Challenges and limitations and future directions

7.1 Challenges and limitations in the field of computational psychiatry

Computational psychiatry, as an emerging, interdisciplinary field, certainly shows considerable potential and promise for tackling current problems in psychiatric theory and practice. However, it has been met with some challenges and limitations along the way. This section will outline some major roadblocks and suggest which steps have been taken to overcome them.

Translation of findings to clinical practice

Despite extensive research efforts, it seems that the field of CP is very slowly progressing from validation of theoretical constructs to implementation of tools in psychiatric practice. In general, translational research is a bidirectional concept in which the knowledge generated from the "benches" of laboratory science can be translated to "bedside" (or the population) and vice-versa. When applied to psychiatry, it involves the translation of (neuro)scientific discoveries into clinically meaningful interventions (Weissman et al., 2011). Translational efforts in psychiatry are mostly aimed at precision medicine, i.e. tailoring treatments and interventions to individual patients based on their unique characteristics, including genetic makeup, biomarkers, clinical symptoms, and personal preferences. The aim is to move from "one-treatment-fits-all" to a more personalized treatment, rendering it more effective. However, this process has been implemented very slowly so far.

Problems with data

This thesis has already outlined some challenges regarding the use of very diverse data in computational models (see Chapter 4). For example, in neuroimaging studies, ***sample sizes*** are usually ***small and biased***. One possible solution is to pool data from multiple sites. However, in order to ensure reproducibility and comparability of the results, it is necessary to adhere to adequate ***preregistration protocols***[65], for example. These practices are also in accordance with

[65] The key features of preregistration are: (1) *a priori* specification of the research design and analysis plan; (2) posting the plan in discoverable repositories prior to observing the outcomes of the study,

principles of open science and transparency in research. Another option is using data from biobanks. The collected data and biosamples in biobanks are readily available for scientific research. However, there have also been some concerns about the participation bias, i.e. that the data submitted voluntarily by (mainly) healthy subjects does not adequately represent more general population (including the underrepresentation of different ethnic groups)[66]. It is also costly and time consuming to collect ***longitudinal data*** in clinical settings. However, they are essential for monitoring or studying the progression of the disorder. One way to obtain longitudinal data with minimal imposition and cost can be via (wearable) digital devices, social media and data collection platforms.

Interpretability of computational models

As it has already been mentioned, ML models in CP are usually termed "black box" because they offer no explanation as to how input variables relate to output of such models. ***Explainable AI (XAI)*** provides a rationale that allows users to understand why a model has produced a given output, which can then be interpreted in a given context based on user's expertise. Some XAI techniques include decision trees, SHapley additive explanation, i.e. SHAP (enables observing how all model features collectively influence its output), various data analytics and visualization tools. One area that is in great need of XAI is that of Clinical Decision Support Systems (CDSSs), in medical field in general, but also in psychiatry. AI-based clinical support tools (CST) can be divided into two groups: (1) those that assist in establishing diagnosis (classification), prognosis or treatment selection (e.g., STAR*D for choice of antidepressants); (2) AI-based services used as auxiliary treatment tools (e.g., therapy or telepsychiatry platforms (BetterHelp), mood or behavior/physiological data trackers (MindStrong, MoodGym), etc.). Explainability or interpretability is particularly relevant for the first group of tools. The absence of explainability may lead to issues of underreliance (or

e.g., Open Science Framework (OSF; https://osf.io/); and (3) reporting all of the planned analyses. Preregistration practices should prevent ordinary confirmation, hindsight, and outcome biases that affect human reasoning. Commonly used in clinical trials, the practice is now gaining popularity in other fields, particularly the social and behavioral sciences like psychology (Bakker et al., 2020).

[66] However, there are examples of databases specialised for mental health, such as the Munich Mental Health Biobank (MMHB). It was established in 2019 and as of 2021, it contains a continuously growing set of data from 578 patients and 104 healthy controls (46.37% women; median age, 38.31 years) (Kalman et al., 2022).

more specifically untrustworthiness) or overreliance on computational models (Antoniadi et al., 2021). Incorporating CSTs routinely in clinical workflows is still not too common. For example, a recent study (Maslej et al., 2023) showed that summaries of clinical notes about a patient with MDD were rated less favorably when psychiatrists believed the notes were generated with AI as compared to another psychiatrist, regardless of whether the notes provided correct or incorrect information.

One of the unintended adverse consequences of predictive algorithms (for prognosis) is potential stigmatization or denial of services based on probabilistic assessments. Models need to be validated and perform well on unseen data, which can be problematic if they were trained on biased datasets. For instance, this may lead to underdiagnosis or misdiagnosis in minority populations. Figure 8.1 illustrates how the differences in model interpretability affect the users.

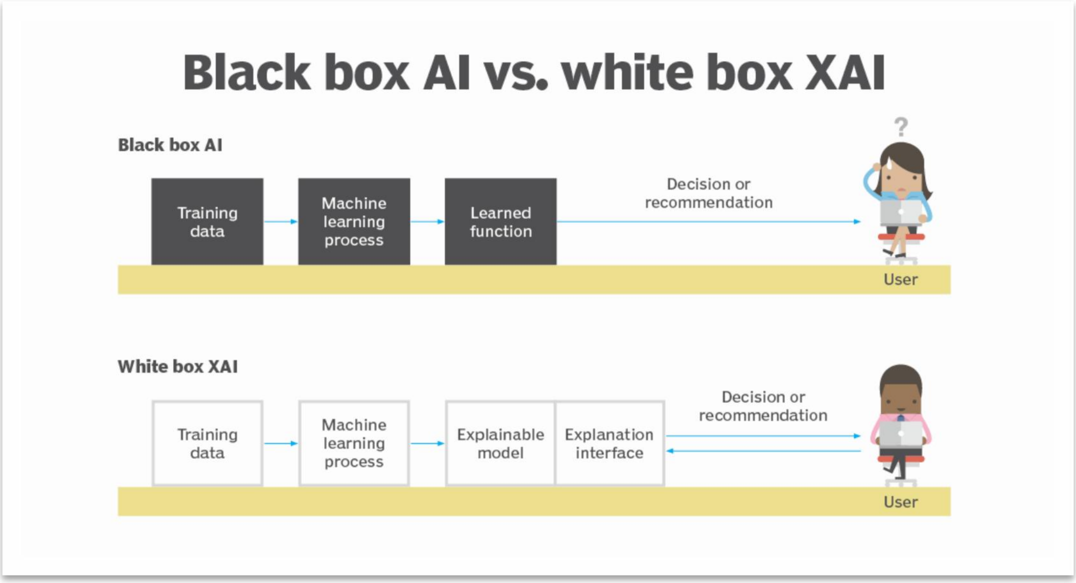


Figure 8.1 More interpretable models for better decision-making with XAI. Explainability helps mental health care providers make better informed decisions regarding treatment selection, diagnosis or prognosis. (Source: https://www.techtarget.com/whatis/definition/explainable-AI-XAI)

***Ethical considerations** – data and patient confidentiality*

Using healthcare data from biobanks, electronic health records (EHR)[67] and other sources in CP carries significant implications for data and patient confidentiality. Anonymizing, controlling access and safely transferring healthcare data is usually a complex task. Although these types of data do not ordinarily enable making direct inferences about the patient identity, they are still deemed sensitive. For example, leaking of private information can affect personal lives, including bullying, high insurance premiums, loss of jobs due to medical or psychiatric history, and other forms of discrimination and stigmatization (Fusar-Poli et al., 2022).

Privacy-enhancing technologies (PET) aim to prevent data leaks while balancing privacy and usability. ***Federated learning*** (FL) is one of the PETs with the primary concept of protecting the privacy of clients' data. FL is a machine learning approach that allows multiple devices to train a shared model without exchanging raw data. Instead of centralizing the data in one location, each device holds a portion of data and collaborates with other devices to update the shared model. This enables training on more extensive and diverse datasets, while preserving the privacy of individual data (Samar Samir Khalil et al., 2024). In this process, similar challenges to collection of data from multiple sites may be encountered.

Ecological validity

Another significant limitation of current tasks in CP, as well as in more traditional neuropsychological testing, is the fact that they are not (sufficiently) ecologically valid. For instance, decisions aimed at maximizing small monetary rewards or minimizing small shocks in the lab do not necessarily reflect prior expectations or decisions in more complex, real-world contexts. Experimental tasks are designed to isolate certain behaviors or cognitive skills so that they can be effectively measured, modeled, and interpreted. However, this necessity sometimes ignores the fact that in real-world scenarios, people might employ multiple skills for solving the task (either sequentially or in parallel), might be affected by the volatility of the environment and similar factors.

[67] An electronic health record includes information about a patient's health history, such as diagnoses, medicines, tests, allergies, immunizations, and treatment plans. Electronic health records can be seen by all authorized health care providers who are taking care of a patient and can be used by them to help make recommendations about the patient's care.

Gamification could possibly be seen as a solution to this problem. By creating a game world, participants are able to make different choices, approach problems in various ways and generally act with greater agency and flexibility compared to isolated computational tasks (Benrimoh et al., 2023). Furthermore, digital tools, such as electronic games, are explicitly designed to be easy to disseminate and require minimal training for the participants, thus facilitating user engagement. Hosting these tools online can potentially reach out to much greater and diverse pool of participants than lab-based studies. Additionally, costs of implementing such tools are significantly lower.

7.2 Future research directions

Systematic review of data in CP

One possible direction in by which this thesis topic could be extended is by delving into a more systematic overview of the data types used in CP, as well as problems in their collection, implementation into CP models and validation. Majority of articles on CP provide descriptions of the data that are needed for a particular experiment or research question at hand, while more systematic reviews are lacking. This is one of the possible research gaps which have been encountered while writing this thesis. In the light of the fact that various types of digital data will likely continue to be collected and used more often in modeling (and might possibly supplant or supplement more traditional data types)[68], it would be interesting to follow-up on this trend.

Adopting and interpreting computational models – building better understanding between computational modelers and clinicians

Another, more practical future research direction would be establishing the degree openness and susceptibility of mental health providers towards employing some practical solutions from CP in their work. For example, a (pilot) questionnaire that would be aimed at therapists and

[68] The use of some parameters obtained from wearable devices (e.g., movement patterns) could be used as a ***proxy*** for light or DLMO parameter in modeling circadian rhythms (Huang et al., 2021). DLMO indicates dim light melatonin onset. The authors assume that activity reaches a plateau at 500 lux and that there is no activity at less than 50 lux, so that would mean inferring a physiological parameter from the recorded movement patterns.

psychiatrists (in our particular ethnic and socioeconomic surroundings) could be designed and implemented to assess their attitude towards computational approaches. In addition, or more specifically, to assess their degree of trustworthiness towards AI-based clinical support tools (CSTs), for diagnosis, treatment selection or prediction. Mental health care providers are effectively the end-users of these tools, so their previous knowledge, training or general familiarity with these approaches may provide useful insights into the feasibility of incorporating such tools into the clinical workflow. The questions could also be used to probe into the reasons for or against trustworthiness or utility of these tools in their work. The results could potentially point out to gaps in knowledge or misconceptions that could be addressed with additional education or training.

Other findings and directions

This extensive literature review enabled identification of some research gaps or understudied phenomena, especially in the theory-driven computational approach (which was one of the objectives of this thesis). Namely, various memory deficits, bodily or somatic symptoms and social deficits are still largely underinvestigated with computational modeling.

However, some authors have begun to recognize the importance of interoceptive processes in mental health and in maintaining the overall homeostasis (Petzschner et al., 2017; Khalsa et al., 2018; Seth & Friston, 2016). Sometimes overlooked in diagnostics, bodily or somatic symptoms expression across the spectrum of psychiatric disorders has helped to motivate the extension of tools from computational psychiatry to address interoception and body regulation via an approach termed ***computational psychosomatics***. This implies a joint computational approach to characterizing disease mechanisms in exteroceptive (psychiatry) and interoceptive (psychosomatics) domains (Petzschner et al., 2017). For example, this approach is important for establishing causality of symptoms such as fatigue (if it does not have organic or neurological origin) in some mental health disorders, e.g., depression. Can depression be caused by fatigue or is fatigue a symptom of depression? They are sometimes difficult to disentangle because they form a closed loop. Findings about the interaction between interoceptive states (brain-body interactions) and environment, i.e. brain-world interactions, could be valuable for understanding the mechanisms of achieving allostasis or homeostasis[69], which might be impaired in mental health disorders.

[69] If it is conceptualized within the predictive coding framework, homeostasis can be described as waiting for errors and correcting them, while allostasis uses prior knowledge, both innate and learned, to prevent errors and minimize them. (Sterling, 2014).

References

Andreassen, O. A., Hindley, G. F. L., Frei, O., & Smeland, O. B. (2023). New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World Psychiatry*, *22*(1), 4–24. https://doi.org/10.1002/wps.21034

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences*, *11*(11), 5088. https://doi.org/10.3390/app11115088

APA Dictionary of Psychology. (2014). *APA Dictionary of Psychology*. Apa.org. https://dictionary.apa.org/learned-helplessness

Bakker, M., Veldkamp, C. L. S., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., Soderberg, C. K., Mellor, D., & Wicherts, J. M. (2020). Ensuring the quality and specificity of preregistrations. *PLOS Biology*, *18*(12), e3000937. https://doi.org/10.1371/journal.pbio.3000937

Benrimoh, D., Fisher, V., Mourgues, C., Sheldon, A. D., Smith, R., & Powers, A. R. (2023). Barriers and solutions to the adoption of translational tools for computational psychiatry. *Molecular Psychiatry*, *28*(6), 2189–2196. https://doi.org/10.1038/s41380-023-02114-y

Benuskova L. (1991). Antidepressants and synaptic plasticity: A hypothesis. *Medical Hypotheses*, *35*(1), 17–22. https://doi.org/10.1016/0306-9877(91)90077-c

Benuskova L & Kasabov N (2007) Computational Neurogenetic Modeling. Springer, New York. ISBN 978-0-387-48353-5.

Biomarkers Definitions Working Group. (2001). Biomarkers and Surrogate endpoints: Preferred Definitions and Conceptual Framework. *Clinical Pharmacology & Therapeutics*, *69*(3), 89–95. https://doi.org/10.1067/mcp.2001.113989

Bull, P. N., Tippett, L. J., & Addis, D. R. (2015). Decision making in healthy participants on the Iowa Gambling Task: new insights from an operant approach. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00391

Castrén, E., & Hen, R. (2013). Neuronal plasticity and antidepressant actions. *Trends in Neurosciences*, *36*(5), 259–267. https://doi.org/10.1016/j.tins.2012.12.010

Castro-Rodrigues, P., Akam, T., Snorasson, I., Camacho, M., Paixão, V., Maia, A., Barahona-Corrêa, J. B., Dayan, P., Simpson, H. B., Costa, R. M., & Oliveira-Maia, A. J. (2022). Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature Human Behaviour*, *6*(8), 1126–1141. https://doi.org/10.1038/s41562-022-01346-2

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/s2215-0366(15)00471-x

Chen, L., Deborah, Tik, M., Elizabeth, Downar, J., Fitzgerald, P. B., Williams, N. R., & Baeken, C. (2023). Accelerated Repetitive Transcranial Magnetic Stimulation to Treat Major Depression: The Past, Present, and Future. *Harvard Review of Psychiatry*, *31*(3), 142–161. https://doi.org/10.1097/hrp.0000000000000364

Computational Psychiatry Course. (2024, September 30). GitHub. https://github.com/computational-psychiatry-course

Cui, L., Li, S., Wang, S., Wu, X., Liu, Y., Yu, W., Wang, Y., Tang, Y., Xia, M., & Li, B. (2024). Major Depressive disorder: hypothesis, mechanism, Prevention and Treatment. *Signal Transduction and Targeted Therapy*, *9*(1). https://doi.org/10.1038/s41392-024-01738-y

Dillon, D. G., & Pizzagalli, D. A. (2018). Mechanisms of Memory Disruption in Depression. *Trends in Neurosciences*, *41*(3), 137–149. https://doi.org/10.1016/j.tins.2017.12.006

Dillon, D. G., Wiecki, T., Pechtel, P., Webb, C., Goer, F., Murray, L., Trivedi, M., Fava, M., McGrath, P. J., Weissman, M., Parsey, R., Kurian, B., Adams, P., Carmody, T., Weyandt, S., Shores-Wilson, K., Toups, M., McInnis, M., Oquendo, M. A., & Cusin, C. (2015). A computational analysis of flanker interference in depression. *Psychological Medicine*, *45*(11), 2333–2344. https://doi.org/10.1017/s0033291715000276

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., Schatzberg, A. F., Sudheimer, K., Keller, J., Mayberg, H. S., Gunning, F. M., Alexopoulos, G. S., Fox, M. D., Pascual-Leone, A., Voss, H. U., & Casey, B. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Duman, R. S., Aghajanian, G. K., Sanacora, G., & Krystal, J. H. (2016). Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. Nature Medicine, 22(3), 238–249. https://doi.org/10.1038/nm.4050

Ester, M, Kriegel, H P, Sander, J, & Xiaowei, Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise.

First, M. B. (2005). Mutually Exclusive versus Co-Occurring Diagnostic Categories: The Challenge of Diagnostic Comorbidity. Psychopathology, 38(4), 206–210. https://doi.org/10.1159/000086093

First, M. B., Rebello, T. J., Keeley, J. W., Bhargava, R., Dai, Y., Kulygina, M., Matsumoto, C., Robles, R., Stona, A.-C., & Reed, G. M. (2018). Do mental health professionals use diagnostic classifications the way we think they do? A global survey. World Psychiatry, 17(2), 187–195. https://doi.org/10.1002/wps.20525

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. https://doi.org/10.1038/nrn2536

Frässle, S., Yao, Y., Schöbi, D., Aponte, E. A., Heinzle, J., & Stephan, K. E. (2018). Generative models for clinical applications in computational psychiatry. *WIREs Cognitive Science*, *9*(3). https://doi.org/10.1002/wcs.1460

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L.-M., Nickless, A., Harrison, P. J., Emsley, R., Luik, A. I., Foster, R. G., Wadekar, V., Hinds, C., Gumley, A., Jones, R., Lightman, S., Jones, S., Bentall, R., Kinderman, P., Rowse, G., Brugha, T., & Blagrove, M. (2017). The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. *The Lancet Psychiatry*, *4*(10), 749–758. https://doi.org/10.1016/s2215-0366(17)30328-0

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, *1*(2), 148–158. https://doi.org/10.1016/s2215-0366(14)70275-5

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., Keshavan, M., Wood, S., Ruhrmann, S., Seidman, L. J., Valmaggia, L., Cannon, T., Velthorst, E., De Haan, L., Cornblatt, B., Bonoldi, I., Birchwood, M., McGlashan, T., Carpenter, W., & McGorry, P. (2013). The Psychosis High-Risk State. *JAMA Psychiatry*, *70*(1), 107–120. https://doi.org/10.1001/jamapsychiatry.2013.269

Fusar-Poli, P., Manchia, M., Koutsouleris, N., Leslie, D., Woopen, C., Calkins, M. E., Dunn, M., Tourneau, C. L., Mannikko, M., Mollema, T., Oliver, D., Rietschel, M., Reininghaus, E. Z., Squassina, A., Valmaggia, L., Kessing, L. V., Vieta, E., Correll, C. U., Arango, C., & Andreassen, O. A. (2022). Ethical considerations for precision psychiatry: A roadmap for research and clinical practice. *European Neuropsychopharmacology*, *63*, 17–34. https://doi.org/10.1016/j.euroneuro.2022.08.001

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neuroscience & Therapeutics*, *24*(11), 1037–1052. https://doi.org/10.1111/cns.13048

Gibb, B. E., McGeary, J. E., & Beevers, C. G. (2015). Attentional biases to emotional stimuli: Key components of the RDoC constructs of sustained threat and loss. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(1), 65–80. https://doi.org/10.1002/ajmg.b.32383

Govind, N. S., Gillespie, K. M., & Branjerdporn, G. (2024). Mental Health Biobanks—A Systematic Review on the Prevalence, Creation, and Implementation of Mental Health Biobanks Globally. *Psychiatry International*, *5*(1), 1–14. https://doi.org/10.3390/psychiatryint5010001

Grieve, S. M., Korgaonkar, M. S., Koslow, S. H., Gordon, E., & Williams, L. M. (2013). Widespread reductions in gray matter volume in depression. *NeuroImage: Clinical*, *3*, 332–339. https://doi.org/10.1016/j.nicl.2013.08.016

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56

Hauser, T. U., Skvortsova, V., De Choudhury, M., & Koutsouleris, N. (2022). The promise of a model-based psychiatry: building computational models of mental ill health. *The Lancet Digital Health*, *4*(11), e816–e828. https://doi.org/10.1016/S2589-7500(22)00152-2

Huang, Y., Mayer, C., Cheng, P., Siddula, A., Burgess, H. J., Drake, C., Goldstein, C., Walch, O., & Forger, D. B. (2021). Predicting circadian phase across populations: a comparison of mathematical models and wearable devices. *Sleep*, *44*(10). https://doi.org/10.1093/sleep/zsab126

Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0166-1

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. Nature Neuroscience, 19(3), 404–413. https://doi.org/10.1038/nn.4238

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1). https://doi.org/10.1186/2045-5380-3-12

Huys, Q. J., Vogelstein, J., & Dayan, P. (2008). Psychiatry: Insights into depression through normative decision-making models. 21, 729–736.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. American Journal of Psychiatry, 167(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Kalman, J. L., Burkhardt, G., Adorjan, K., Barton, B. B., Jonge, S. D., Eser-Valeri, D., Falter-Wagner, C. M., Heilbronner, U., Jobst, A., Keeser, D., Koenig, C., Koller, G., Nikolaos Koutsouleris, Kurz, C., Landgraf, D., Merz, K., Musil, R., Nelson, A. M., Padberg, F., & Sergi Papiol. (2022). Biobanking in everyday clinical practice in psychiatry—The Munich Mental Health Biobank. *Frontiers in Psychiatry*, *13*. https://doi.org/10.3389/fpsyt.2022.934640

Kämpfen, F., Kohler, I. V., Ciancio, A., Bruine de Bruin, W., Maurer, J., & Kohler, H.-P. (2020). Predictors of mental health during the Covid-19 pandemic in the US: Role of economic concerns, health worries and social distancing. *PLOS ONE*, *15*(11), e0241895. https://doi.org/10.1371/journal.pone.0241895

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., & Stephan, K. E. (2018). Interoception and Mental Health: A Roadmap. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 3(6), 501–513. https://doi.org/10.1016/j.bpsc.2017.12.004

Kirill Fedorovich Vasilchenko, & Egor Maksimovich Chumakov. (2023). Current status, challenges and future prospects in computational psychiatry: a narrative review. Consortium Psychiatricum/Consortium Psychiatricum, 4(3), 33–42. https://doi.org/10.17816/cp11244

Koen Demyttenaere, & Heirman, E. (2023). Assessment Tools in Psychiatry. *Springer EBooks*, 1–32. https://doi.org/10.1007/978-3-030-42825-9_101-1

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., & Rosen, M. (2021). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*, *78*(2), 195. https://doi.org/10.1001/jamapsychiatry.2020.3604

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics. Psychophysiology, 53(3), 286–297. https://doi.org/10.1111/psyp.12518

Krause, F. C., Linardatos, E., Fresco, D. M., & Moore, M. T. (2021). Facial emotion recognition in major depressive disorder: A meta-analytic review. *Journal of Affective Disorders*, *293*, 320–328. https://doi.org/10.1016/j.jad.2021.06.053

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain*, *131*(8), 2084–2093. https://doi.org/10.1093/brain/awn136

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133–145. https://doi.org/10.1002/wps.20941

Leuchter, A. F., Cook, I. A., Marangell, L. B., Gilmer, W. S., Burgoyne, K. S., Howland, R. H., Trivedi, M. H., Zisook, S., Jain, R., McCracken, J. T., Fava, M., Iosifescu, D., & Greenwald, S. (2009). Comparative effectiveness of biomarkers and clinical indicators for predicting outcomes of SSRI treatment in Major Depressive Disorder: Results of the BRITE-MD study. *Psychiatry Research*, *169*(2), 124–131. https://doi.org/10.1016/j.psychres.2009.06.004

Lewinsohn, P. M., & Atwood, G. E. (1969). Depression: A clinical-research approach. *Psychotherapy: Theory, Research & Practice*, *6*(3), 166–171. https://doi.org/10.1037/h0088744

Li, B.-J., Friston, K., Mody, M., Wang, H.-N., Lu, H.-B., & Hu, D.-W. (2018). A brain network model for depression: From symptom understanding to disease intervention. CNS Neuroscience & Therapeutics, 24(11), 1004–1019. https://doi.org/10.1111/cns.12998

Lyall, L. M., Wyse, C. A., Graham, N., Ferguson, A., Lyall, D. M., Cullen, B., Celis Morales, C. A., Biello, S. M., Mackay, D., Ward, J., Strawbridge, R. J., Gill, J. M. R., Bailey, M. E. S., Pell, J. P., & Smith, D. J. (2018). Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *The Lancet Psychiatry*, *5*(6), 507–514. https://doi.org/10.1016/s2215-0366(18)30139-1

Maj, M. (2005). “Psychiatric comorbidity”: an artefact of current diagnostic systems?. British Journal of Psychiatry, 186(3), 182–184. https://doi.org/10.1192/bjp.186.3.182

Mäki-Marttunen, T, Blackwell, K. T., Ibrahim Akkouh, Alexey Shadrin, Mathias Valstad, Torbjørn Elvsåshagen, Linne, M.-L., Srdjan Djurovic, Einevoll, G. T., & Andreassen, O. A. (2024). Genetic mechanisms for impaired synaptic plasticity in schizophrenia revealed by computational modeling. *Proceedings of the National Academy of Sciences*, *121*(34). https://doi.org/10.1073/pnas.2312511121

Maslej, M. M., Kloiber, S., Ghassemi, M., Yu, J., & Hill, S. L. (2023). Out with AI, in with the psychiatrist: a preference for human-derived clinical decision support in depression care. *Translational Psychiatry*, *13*(1), 1–9. https://doi.org/10.1038/s41398-023-02509-z

Montague, P. R. 2007. Neuroeconomics: A View from Neuroscience. Functional Neurology 22 (4): 219.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Montague, P., Dayan, P., & Sejnowski, T. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/jneurosci.16-05-01936.1996

Nordgaard, J., Nielsen, K. M., Rasmussen, A. R., & Henriksen, M. G. (2023). Psychiatric comorbidity: a concept in need of a theory. Psychological Medicine, 53(13), 5902–5908. https://doi.org/10.1017/S0033291723001605

Onraedt, T., & Koster, E. H. W. (2014). Training Working Memory to Reduce Rumination. PLoS ONE, 9(3), e90632. https://doi.org/10.1371/journal.pone.0090632

Park, S.-C., & Kim, Y.-K. (2021). Challenges and Strategies for Current Classifications of Depressive Disorders: Proposal for Future Diagnostic Standards. *Major Depressive Disorder*, 103–116. https://doi.org/10.1007/978-981-33-6044-0_7

Petzschner, F. H., Weber, L. A. E., Gard, T., & Stephan, K. E. (2017). Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. *Biological Psychiatry*, *82*(6), 421–430. https://doi.org/10.1016/j.biopsych.2017.05.012

Polygenic Risk Score (PRS). (2024). Genome.gov. https://www.genome.gov/genetics-glossary/Polygenic-Risk-Score-PRS

Predictive coding. (2010). Frontiers. https://www.frontiersin.org/research-topics/599/predictive-coding

Rajkumar, R. P. (2022). The Correlates of Government Expenditure on Mental Health Services: An Analysis of Data From 78 Countries and Regions. *Cureus*, *14*(8). https://doi.org/10.7759/cureus.28284

Ritchie, H., Roser, M., Dattani, S., & Rodes-Guirao, L. (2018, April). *Mental health*. Our World in Data. https://ourworldindata.org/mental-health

Robson, S. E., Repetto, L., Gountouna, V.-E., & Nicodemus, K. K. (2019). A review of neuroeconomic gameplay in psychiatric disorders. *Molecular Psychiatry*, *25*(1), 67–81. https://doi.org/10.1038/s41380-019-0405-5

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. https://doi.org/10.1017/s0033291713002535

Roser, M., Ritchie, H., & Spooner, F. (2024). Burden of Disease. *Our World in Data*. https://ourworldindata.org//burden-of-disease

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology*, *55*, 152–159. https://doi.org/10.1016/j.conb.2019.02.006

Saez, I., & Gu, X. (2022). Invasive Computational Psychiatry. *Biological Psychiatry*, *0*(0). https://doi.org/10.1016/j.biopsych.2022.09.032

Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, *572*, 522–542. https://doi.org/10.1016/j.ins.2021.05.055

Samar Samir Khalil, Tawfik, N. S., & Spruit, M. (2024). Exploring the potential of federated learning in mental health research: a systematic literature review. *Applied Intelligence*. https://doi.org/10.1007/s10489-023-05095-1

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., Cheung, J. W., van Erp, T. G. M., Bos, D., Ikram, M. A., Vernooij, M. W., Niessen, W. J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H. J., Janowitz, D., Bülow, R., Selonke, M., & Völzke, H. (2016). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular Psychiatry*, *22*(6), 900–909. https://doi.org/10.1038/mp.2016.60

Schurr, R., Reznik, D., Hillman, H., Rahul Bhui, & Gershman, S. J. (2024). Dynamic computational phenotyping of human cognition. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-024-01814-x

Seriès, P. (2020). *Computational psychiatry : a primer*. The MIT Press.

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*, *1*(2), 213–220. https://doi.org/10.1177/2167702612469015

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for Sensory Prediction Deficits in Schizophrenia. *American Journal of Psychiatry*, *162*(12), 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Siegle, G. J., & Hasselmo, M. E. (2002). Using connectionist models to guide assessment of psycological disorder. *Psychological Assessment*, *14*(3), 263–278. https://doi.org/10.1037/1040-3590.14.3.263

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current Opinion in Neurobiology*, *25*, 85–92. https://doi.org/10.1016/j.conb.2013.12.007

Sterling, P. (2014). Homeostasis vs Allostasis. *JAMA Psychiatry*, *71*(10), 1192. https://doi.org/10.1001/jamapsychiatry.2014.1043

Teixeira, A. L., Rocha, N. P., & Berk, M. (2023). *Biomarkers in Neuropsychiatry*. Springer Nature.

The Lancet Global Health. (2020). Mental Health Matters. *The Lancet Global Health*, *8*(11). https://doi.org/10.1016/s2214-109x(20)30432-0

Torous, J., Staples, P., & Onnela, J.-P. (2015). Realizing the Potential of Mobile Mental Health: New Methods for New Data in Psychiatry. *Current Psychiatry Reports*, *17*(8). https://doi.org/10.1007/s11920-015-0602-0

Uher, R., & Zwicker, A. (2017). Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. World Psychiatry, 16(2), 121–129. https://doi.org/10.1002/wps.20436

Weissman, M. M., Brown, A., & Talati, A. (2011). Translational Epidemiology in Psychiatry. Archives of General Psychiatry, 68(6), 600–600. https://doi.org/10.1001/archgenpsychiatry.2011.47

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*. https://doi.org/10.7554/elife.49547

Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., & Marquand, A. F. (2015). From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, *57*, 328–349. https://doi.org/10.1016/j.neubiorev.2015.08.001

Woody, M. L., & Gibb, B. E. (2015). Integrating NIMH Research Domain Criteria (RDoC) into depression research. Current Opinion in Psychology, 4, 6–12. https://doi.org/10.1016/j.copsyc.2015.01.004