Computational Models of Mental Health Disorders

Milica Kiš

Fakulta matematiky, fyziky a informatiky
Univerzity Komenského
Mlynská dolina F1
842 48 Bratislava
Email: kis24@uniba.sk, milicakis@gmail.com

Abstract

This contribution has three main aims: (1) to provide a brief overview of types of computational models used in computational psychiatry (data-driven, theory-driven and combined approaches); (2) to offer a brief overview of all data types used in computational psychiatry (i.e. neuroimaging, genetic, digital, and behavioral data) and highlight possible problems in their use; (3) to synthesize findings from the application of these approaches to depression and to identify possible research gaps.

Keywords: computational psychiatry, mental health disorders, depression

1 Introduction

This paper is based on the Master's thesis entitled „Computational Models of Mental Health Disorders: A Review" (Kiš, 2025). The aim of this paper is to introduce the topic of computational approach to studying and analyzing mental health disorders as an alternative to current approaches in psychiatric theory and practice. Current use of diagnostic manuals (DSM-5-TR, ICD-11) in clinical practice is based on a theoretical and descriptive approach, whereby each disorder is characterized by a list of possible symptoms. Alternative approaches, such as the Research Domain Criteria (RDoC) initiative, or computational models, represent the effort to move away from the more traditional, symptom-based categories.

Computational psychiatry (CP) is an emerging interdisciplinary field 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). 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, the 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.

2 Types of models used in computational psychiatry

Computational psychiatry encompasses three broad approaches: data-driven, theory-driven, and combined models. ***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 (Huys et al., 2016).

2.1 Data-driven models

Data-driven modeling refers to the application of ML techniques to diverse data, which are generally agnostic in regards to the underlying mechanisms of the studied disorders and are used for finding patterns in data. In the context of computational psychiatry, the focus is mainly on supervised and unsupervised ML 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.

Therefore, supervised methods are particularly useful for the following problems in psychiatric practice: diagnostic classification (i.e., automating the diagnostic process), prediction of treatment outcomes (prognosis), and treatment selection. On the other hand, unsupervised methods are well-suited to solving the problems of stratifying (subtyping) mental health disorders, both within and across different diagnoses. Clustering approaches, which belong to unsupervised methods, are also significant for establishing new descriptions and classifications, beyond traditional, symptom-based categories, when applied to various data.

2.2 Theory-driven models

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.

2.2.1 Synthetic models

Synthetic, biophysically realistic neural-network models are commonly used to elucidate how biological abnormalities found in mental health disorders affect neurobehavioral dynamics. Synthetic models have been successfully used for explaining the disturbances in OCD, schizophrenia, and addiction.

2.2.2 Reinforcement learning models

Reinforcement learning (algorithmic) models address how an agent (in either natural or artificial system) optimizes behavior in a complicated environment, which 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.

2.2.3 Bayesian (optimal) models

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, can also be related to the imbalance between incoming sensory information and prior beliefs and expectations. Similarly to RL models, these models are typically validated through quantitative statistical means.

3 Types of data used in computational models

A special part of the thesis was dedicated to the overview of various types of data used in computational models, as well as the problems and challenges encountered during the processes of data collection, implementation into computational models and data validation. The rationale for this was that more systematic presentation of data types in literature was lacking, but the aim was also to highlight the use of novel types of data, especially those collected via digital devices, termed ***digital data***. For the purposes of this review, data were grouped into three categories: (1) ***clinical***, (2) ***laboratory-based***, and (3) ***digital*** data (Figure 1). Clinical data refers to data elicited from patients by means of clinical assessments or behavioral experiments. Laboratory-based data refers to neuroimaging (e.g. M(EEG), (f)MRI)) and genetic data. Both these data types are highly reliable (being obtained in highly controlled conditions) and denoised. However, their collection may be costly and time-consuming. The third type of data, i.e. digital data, differs from the more „traditional“ data types in that it is much faster and easier to collect, but it might also need additional validation.

Digital data may be roughly divided into ***passive*** and ***active*** data and includes any data collected from participants via digital devices (Hauser et al., 2022). 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 (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. Bringing together passive and active data sources, e.g., by collecting eye-tracking data during game play, could yield further insights in future studies.

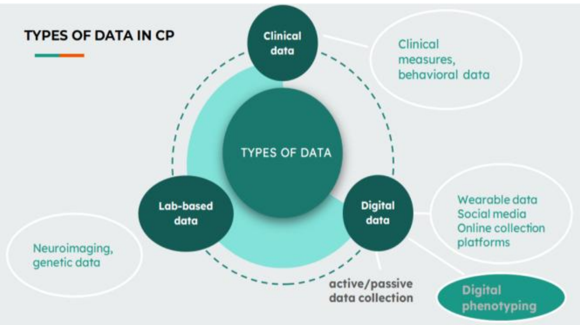

Fig. 1: Types of data used in CP

4 Application of the models to the study of depression

In order to illustrate how computational modeling can be applied both to testing specific theoretical hypotheses about mental health disorders and to real problems in clinical practice, we used examples of studies related to one of the most common disorders globally, i.e., depression.

Depression can have a very heterogeneous presentation, and certain aspects of the disorder are well-suited for the application of theory-driven computational models. Most common aspects include rumination, anhedonia, learned helplessness, and various cognitive deficits (working memory, executive function).

Deficits in executive function

Dillon et al. (2015) used the drift diffusion models (DDMs)[1] 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.
In clinical practice, researchers and clinicians are often challenged to improve patients' responsiveness to certain therapies, which shortens the treatment duration and alleviates patient suffering. Various data-driven approaches have been proven successful in predicting the treatment outcomes, which is also one of the goals of computational psychiatry.
Chekroud et al. (2016) developed an ML model to predict whether patients would achieve clinical remission from major depressive disorder (MDD) after a 12-week course of citalopram. The model was trained on data from STAR*D[2] 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 critical steps in the model development. The 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. The 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[3] 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-minute questionnaire.

Conclusion

Computational models are showing great potential for understanding the underlying mechanisms of mental health disorders, as well as for solving various problems in psychiatric practice. However, translating scientific findings into clinically usable interventions is a slow process. 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.

References

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.

[1] In drift diffusion models (DDMs), a process of making a decision between two choices is 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).

[2] 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) (Chekroud et al., 2016).

[3] COMED (Combining Medications to Enhance Depression Outcomes) was a single-blind, randomized, placebo-controlled trial comparing the efficacy of medication combinations in the treatment of MDD (Chekroud et al., 2016).

(2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250.

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.

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.

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.

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.

Kiš, M. (2025). Computational Models of Mental Health Disorders: A Review.

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