Implementation Issues AI - Mental Healthcare

Anne-Kathrin Kleine[1] &

[1] LMU

Author Note

LMU

The authors made the following contributions. Anne-Kathrin Kleine: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; : , .

Correspondence concerning this article should be addressed to Anne-Kathrin Kleine, . E-mail: Anne-Kathrin.Kleine@psy.lmu.de

Implementation Issues AI - Mental Healthcare

Relevant terms:

- The International Medical Device Regulators Forum (IMDRF) has defined *Software* as a Medical Device as "software intended to be used for medical purposes that performs its objectives without being part of a hardware medical device."
- *Automation vs. decision support tasks*:
  - Automation tasks are cases in which "a machine operates independently to complete a task," whereas clinical decision support tasks are cases in which "a machine is concerned with providing information or assistance to the primary agent responsible for task completion."

Issues in the implementation of AI in mental healthcare practice

- large amount of academic knowledge and developed algorithms not integrated into clinical care (Sendak et al., 2020)
  - "This narrative review was unable to provide standard metrics of adoption, because many of the figures marketed by product developers have no peer-reviewed evidence" (Sendak et al., 2020)
- Big data confidentiality (Aafjes-van Doorn et al., 2021)
- Black box problems Kelly et al. (2019)
- In addition, black-box predictive models combined with (similarly complex) explanatory methods may yield complicated decision pathways that increase the likelihood of human error (Chekroud et al., 2021)
- ethical challenges:
  - responsibility (Chekroud et al., 2021)
  - dehuminization (Chekroud et al., 2021)
  - in clinical settings: transparency highly values - opposing black box problem (Chekroud et al., 2021)
  - erronous outcomes for underrepresented groups (Chekroud et al., 2021)

    - misuse of personal and sensitive data (Chekroud et al., 2021)
- diagnostic challenges (Lee et al., 2021)
    - Performance of supervised algorithms depends on the quality of the diagnostic labels used to train a model; Given the heterogeneity characteristic of mental illnesses, labels of disease states may not be specific enough to yield AI algorithms with high sensitivity and specificity
        - One possibility is to use ML algorithms to predict specific symptoms or functional consequences rather than diagnoses
        - Another opportunity lies in leveraging the strength of deep neural networks that can operate without human oversight to identify novel biomarkers for detecting specific diseases (29)
        - When the results of ML algorithms are published, they must include information regarding the quality of the data used to train the model as well as any potential biases in it, which is rarely done at present.
- Specificity vs sensitivity tradeoffs:
    - e.g., However, the positive predictive value (PPV) (number of correctly predicted positive cases divided by the number of predicted positive cases) of prediction models for suicide attempts and deaths remains extremely low. In a systematic review of 17 studies, Belsher et al. found a PPV of less than 1% for suicide mortality despite good accuracy (greater than or equal to 80%) [193]. In other words, ML algorithms still deliver a high rate of false alarms despite a high level of accuracy (Roth et al., 2021)
        - shown for multiple areas (assessment of suicide risk, depression, psychosis)
    - More complex ML models often have greater accuracy but lower interpretability
    - Generally, there is a trade-off between explainablity and performance. For instance, a constrained linear or bilinear model will fit many of these criteria, but the linear model does not warrant a good performance. Additionally, a model

that is potentially explainable does not guarantee explainability. For example, co-dependence of input variables may make explanations ambiguous (Chen et al., 2022)

- Furthermore, each mental disorder has various types of overlapping symptoms with varying degrees, bringing an additional challenge to uniquely define the disorder in psychiatry (unlike a clear cut in cardiology or oncology) (Chen et al., 2022)
- many mental disorders have overlapping symptoms with other physical or mental disorders (Chen et al., 2022)

Issues in application research

- few studies test algorithms in independent samples Kelly et al. (2019)
- when randomizing patients to algorithm-informed care or usual care, clinicians may override algorithm recommendations and choose alternative treatments (Chekroud et al., 2021)
- Patients may refuse the algorithm-recommended treatment, or have restrictions to its use that were not contemplated by the decision support tool (e.g., prohibitive cost of therapy) (Chekroud et al., 2021)
- In light of this, effect sizes for these interventions will often vary when applied in different settings (Chekroud et al., 2021)
- the development of data-driven decision tools should be informed by extensive consultation and coproduction with the intended users, in order to implement models that maximize acceptability and compatibility with other clinical guidelines (i.e., risk management procedures, norms about safe dosage or titration of medications) (Chekroud et al., 2021)
- fear of being substituted by AI systems?
- research environments must encourage large-scale, collaborative, interdisciplinary consortia (Browning et al., 2020)
- performance metrics:

- The selected factors may include both specific computational properties such as parameter identifiability as well as practical features of an assay (e.g. duration to complete, complexity) and clinical validity (e.g. correlation with symptoms or treatment response) (Browning et al., 2020)
- Longitudinal observational studies may be used to assess whether an assay covaries with mental state changes or traits of interest and whether it has predictive validity, for example by predicting response to treatment (Browning et al., 2020)
- Regardless of whether the goal of using a computational assay is to predict a clinical outcome or to guide the development of a novel treatment, the efficacy of computationally informed approaches must ultimately be assessed in clinical trials. Such trials may, for example, randomly assign patients to be treated according to a predictive algorithm or standard treatment, or to receive a computationally informed intervention vs. a control (Browning et al., 2020)
- difficulty of comparing different algorithms and AI systems (Kelly et al., 2019)
- "products listed in Table 2 that predict the same outcome cannot be easily compared. Reporting of machine learning models often fails to follow establish best practices and model performance measures are not standardised across publications" (Sendak et al., 2020)
- there is no current standard definition of accuracy and patient health outcomes against which to measure the products.
- metrics may not reflect clinical applicability: e.g., AUC not the most useful metric and difficult to understand by clinicians (Kelly et al., 2019)
- However, none of these measures ultimately reflect what is most important to patients, namely whether the use of the model results in a beneficial change in patient care Shah et al. (2019)
- possible solution: decision curve analysis

Ways out and forward

- When conducted with care for ethical considerations, ML research can become an essential complement to traditional psychotherapy research (Chekroud et al., 2021)
- highlight AI as a chance and addition to common practice (supporting, not substituting):
  - It is important to highlight that none of the identified ML applications were developed to replace the therapist, but instead were designed to advance the therapists' skills and treatment outcome (Chekroud et al., 2021)
- educating about limitations AND chances (Roth et al., 2021)

Multimodality of sources

- ML methods provide an opportunity for multimodal analyses of patient and therapist moment-by-moment changes in word use, speech, body movements, and physiological states, that are not (yet) usually considered in clinical decision making (Chekroud et al., 2021)
- Illustrations include Instagram photographs to predict risk of developing depression (51), speech data to predict psychosis onset in high-risk youth (52), and identifying individuals with PTSD (53) (Lee et al., 2021)
- Mental illnesses may be observable in online contexts, and social media data have been leveraged to predict diagnoses and relapses (51,72,76,77), with accuracies comparable to clinician assessments and screening surveys (78) (Lee et al., 2021)
- leveraging "big data" from a longitudinal perspective offers a promising way to track the trajectories of neural phenotypes that have been rarely examined in previous cross-sectional studies of psychiatric disorders (Chen et al., 2022)
- AI methodology can also incorporate both genetic and environmental risks (54), accounting for complex environment-gene interactions and psych-bio-social factors, particularly relevant in PTSD (55) (Lee et al., 2021)
- Furthermore, AI methodologies are well-suited for deciphering patterns from

longitudinal data (56), critical for honing the accuracy of diagnoses based on evolving psychiatric symptoms (Lee et al., 2021)

- Lastly, AI methods may have a growing role in gathering sensitive and accurate data from patients. One study found that individuals were more forthcoming disclosing sensitive information with a computer system than with a person (57) (Lee et al., 2021)

Precision psychiatry

- finer grained diagnoses possible: First, AI approaches can bolster the ability to differentiate between diagnoses with similar initial clinical presentations but divergent treatment approaches (43) – e.g., identifying bipolar versus unipolar depression based on brain imaging features (44), or differentiating between types of dementia using structural MRI scans (45) (Lee et al., 2021)
- Secondly, data-driven AI methods can help identify novel disease subtypes based on heterogeneity of presentations, demographic features, and environmental factors (43). Examples include neurocognitive profiles in bipolar disorder (46), genetic profiles in schizophrenia (47), biomarker profiles in psychoses (48), and neuroimaging subtypes in depression (49) (Lee et al., 2021)
- Thirdly, AI approaches can build models from unusual/novel data sources and reconcile data from multiple heterogeneous datastreams, e.g., EHR, behavioral data from digital phenotyping and wearable sensors, speech, social media feeds, neurophysiology, imaging, and genetics (50), to coalesce explanatory and mechanistic models of mental illness across self-report to molecular assessments (Lee et al., 2021)
- that existing clinical diagnostic categories could misrepresent the causes underlying mental disturbance and the case-control study design has limited strength in delineating the significant clinical and neurobiological heterogeneity of psychiatric disorders (Chen et al., 2022)
- The ultimate goal of RDoC is to find “new ways of classifying psychiatric diseases based on multiple dimensions of biology and behavior” (Chen et al., 2022)

- Thanks to the advancement in cuttingedge techniques in ML/AI, psychiatrists and investigators now have an unprecedented opportunity to benefit from complex patterns in brain, behavior, and genes using machine learning tools (Chen et al., 2022)
- Increasing evidence suggests that datadriven subtyping could drive novel neurobiological phenotypes associated with distinctive behavior and cognitive functioning
- Chances for *precision psychiatyry*: categorization of psychiatric patients into new data-driven subgroups (Roth et al., 2021)
  - less stigmatization
  - homogenous disease classification, early diagnosis, prediction of disease trajectory, and tailored, more effective, safer, and predictable treatment, potentially at the individual level
- Clinical decision support (CDS) provides clinicians with knowledge (e.g., treatment guidelines) and patient-specific information (e.g., clinical and laboratory data), specifically selected and presented in a timely fashion, to enhance the quality of medical care (Roth et al., 2021)

Importance of practitioner training:

- To improve understanding, medical students and practising clinicians should be provided with an easily accessible AI curriculum to enable them to critically appraise, adopt and use AI tools safely in their practice (Kelly et al., 2019)
- Thus, it will be important for psychotherapy researchers to become better-versed in the ML methods and how to interpret this research literature (Chekroud et al., 2021)
- Accessible ML education and tool development is required to facilitate understanding and usage in the wider clinical research community. Besides formal education on ML in psychology graduate programs, it might also be helpful for psychotherapy researchers to attend (online and freely available) courses on ML (Chekroud et al., 2021)
- Sendak et al. (105) have proposed four phases of translation necessary to bridge this

gap: design and development of ML products that can support clinical decision-making and are actionable; evaluation and validation; diffusion and scaling across settings such that the tools are more widely applicable; and continued monitoring and maintenance to remain current with clinical practice needs (Lee et al., 2021)

- For instance, a classification function learned by the machine to predict a disease outcome would not only need to report a probability outcome but also need to address additional questions for the end-user: why is this outcome instead of the alternative? How reliable is the outcome? When does it fail if something is missing or misrepresented? When and why the prediction is wrong? Accordingly, a model with improved interpretability comes with parameter/structure/connectivity constraints and some prior domain knowledge (Chen et al., 2022)
- The translation milestones (Sendak et al., 2020)
  - To map between individual products and the translational path, milestones for each product are marked within four phases:
  - 1) design and develop
    - The setting and funding of the team shapes many aspects of how the machine learning product is designed and developed.
    - For example, in an academic setting it may be easier to cultivate collaborations across domains of expertise early on in the process. However, academic settings may have difficulty recruiting and retaining the technical talent required to productise complex technologies.
  - 2) evaluate and validate
    - Clinical utility: can the product improve clinical care and patient outcomes?
    - Statistical validity: Can the machine learning product perform well on metrics of accuracy, reliability, and calibration?
    - Economic utility: Can there be a net benefit from the investment in the machine learning product?

  - 3) diffuse and scale
    - diffuse and scale across settings, which requires special attention to deployment modalities, funding, and drivers of adoption.
    - To scale, machine learning products must be able to ingest data from different EHR and must also support on-premise and cloud deployments. For this reason, many models are also distributed as stand-alone web applications that require manual entry to calculate risk.
  - 4) continuing monitoring and maintenance
    - Data quality, population characteristics, and clinical practice change over time and impact the validity and utility of models.
    - Model reliability and model updating are active fields of research and will be integral to ensure the robustness of machine learning products in clinical care.
- machine learning technologies are referred to as products rather than models, recognising the significant effort required to productise and operationalise models that are often built primarily for academic purposes
- The ‘inconvenient truth’ of machine learning in healthcare was pointedly described as “at present the algorithms that feature prominently in research literature are in fact not, for the most part, executable at the front lines of clinical practice.”
- machine learning is initially expected to impact healthcare through augmenting rather than replacing clinical workflows
- Machine learning technologies were included as case studies if they met two criteria: 1) they tackle a clinical problem using solely EHR data; and 2) they are evaluated and validated through direct integration with an EHR to demonstrate clinical, statistical, or economic utility
- Case studies were selected amongst 1,672 presentations at 9 informatics and machine learning conferences between January 2018 and October 2019

Research Ideas

Focus Group Psychiatrists

- Educate about chances and limitations
- Discuss implementation possibilities and difficulties

Meta-analysis AI performance in fields for which none exists (see Roth et al. (2021))

- PTSD
- Delirium
- Substance use

AI for EHR (electronic health records) handling: Possibilities (review)

Interviews about implementation possibilities and issues with practitioners (health care specialists, data scientists, (patients?))

RCTs in mental health applications

- Physical health: Randomized controlled trials (RCTs) and prospective studies can bridge this gap between theory and practice, more rigorously demonstrating that AI models can have a quantifiable, positive impact when deployed in real healthcare settings (Rajpurkar et al., 2022)
- currently largely missing for mental health applications

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Browning, M., Carter, C. S., Chatham, C., Ouden, H. D., Gillan, C. M., Baker, J. T., Chekroud, A. M., Cools, R., Dayan, P., Gold, J., Goldstein, R. Z., Hartley, C. A., Kepecs, A., Lawson, R. P., Mourao-Miranda, J., Phillips, M. L., Pizzagalli, D. A., Powers, A., Rindskopf, D., . . . Paulus, M. (2020). Realizing the Clinical Potential of Computational Psychiatry: Report From the Banbury Center Meeting, February 2019. *Biological Psychiatry*, *88*(2), e5–e10. https://doi.org/10.1016/j.biopsych.2019.12.026

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170. https://doi.org/10.1002/wps.20882

Chen, Z. S., Prathamesh, Kulkarni, Galatzer-Levy, I. R., Bigio, B., Nasca, C., & Zhang, Y. (2022). *Modern Views of Machine Learning for Precision Psychiatry*. arXiv. http://arxiv.org/abs/2204.01607

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H.-C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(9), 856–864. https://doi.org/10.1016/j.bpsc.2021.02.001

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Roth, C. B., Papassotiropoulos, A., Brühl, A. B., Lang, U. E., & Huber, C. G. (2021). Psychiatry in the Digital Age: A Blessing or a Curse? *International Journal of Environmental Research and Public Health*, *18*(16), 8302. https://doi.org/10.3390/ijerph18168302

Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., Ratliff, W., & Balu, S. (2020). A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innovations*. https://doi.org/10.33590/emjinnov/19-00172

Shah, N. H., Milstein, A., & Bagley, P., Steven C. (2019). Making Machine Learning Models Clinically Useful. *JAMA*, *322*(14), 1351–1352. https://doi.org/10.1001/jama.2019.10306