Machine Learning: How ML Can Transform Psychotherapy

CS4991 Capstone Report, 2025

Franz Charles Silva
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
kxb3ex@virginia.edu

ABSTRACT

The field of psychotherapy faces a significant challenge of incorporating new technology while maintaining private, personalized therapeutic practices. To address this, I conducted a meta-study exploring how machine learning (ML) is being integrated into mental health treatment, exploring its potential to enhance or devalue therapy. I analyzed a range of academic studies and journals, identifying the benefits and limitations of ML within mental health treatment. The research reveals that ML shows promise in improving accessibility and personalization in mental health care, while also raising concerns regarding data privacy and bias. Future work includes addressing ethical concerns towards the use of ML and collaborative efforts between ML developers and mental health experts to ensure that the technology created assists, rather than replaces, current mental health practices.

1. INTRODUCTION

In the United States, millions of Americans struggle to access timely and effective mental health treatment due to a shortage of trained professionals, financial barriers and discrimination. While traditional in-person therapy models are effective, access for rural and under-served areas is limited. Advancements of telehealth, although they expand access, do not fully address the challenge presented by the lack of trained professionals.

ML presents a promising avenue to enhance psychotherapy by improving diagnosis, personalizing treatment, and providing alternate forms of therapy like chatbots. Advanced algorithms can detect patterns from social media posts and patient data to identify risks for certain mental disorders, while AI-powered chatbots and virtual therapists provide immediate support to patients outside of therapists' hours, offering an alternative for those hesitant to seek traditional care due to financial or accessibility issues. ML can not only increase accessibility to underserved areas, but increase the accuracy of predicting risks for mental health conditions and tailoring unique treatment plans for patients.

2. RELATED WORKS

A number of experts in the field of machine learning and mental health have proposed the use of ML as a potential solution to the growing demand of accessible psychotherapy. The major advantages of adopting this approach include improved detection and diagnosis of mental health conditions, as well as enhanced treatment personalization through predictive analytics (Shatte, et al., 2019). However, potential drawbacks including the black box nature of machine learning models used within psychotherapy and biases in training data, could negatively affect the output.

Specifically, deep learning algorithms function as black boxes, making it difficult for

therapists and patients to understand how predictions or recommendations are generated. Without clear explanations for how a model reaches a conclusion, distrust in machine learning assisted therapy can drive patients away from seeking treatment. Additionally, bias present in training data could be perpetuated through the incorporation of machine learning. If technologies are trained on non-representative data, they may fail to recognize symptoms accurately within under-represented populations, decreasing the personalization and accessibility machine learning is supposed to provide to under-served areas (Gooding & Kariotis, 2021). This could disproportionately affect marginalized groups, deepening the gap in mental health treatment.

While machine learning holds great promise in enhancing psychotherapy, its limitations must be carefully addressed. Moving forward, efforts should focus on increasing transparency for patients and ensuring diverse, representative datasets to mitigate risks of bias.

3. PROJECT DESIGN

This section discusses the research approach used to explore applications of machine learning in mental healthcare, giving an overview of the meta-study's goals and outcomes.

3.1 Introduction

This meta-study aims to systematically review and analyze the applications of ML in mental healthcare, focusing on their clinical effectiveness. The study will explore how specific ML models are integrated into diagnostic tools, predictive analytics, and chatbots.

The study is guided by the following research questions: How are machine learning models currently being used in mental health treatment? What are the benefits and limitations of these machine learning applications?

This meta-study will focus on identifying and categorizing ML models used in mental health care, primarily focusing on their integration into predictive analytics and chatbots. Research will be conducted through academic journals and studies obtained from Google Scholar. Along with studies that discuss practical applications rather than purely theoretical research, studies will be selected based on their empirical use of ML models for mental health diagnostics and patient support.

3.2 Deep Neural Networks

Durstewitz, et al. (2019) discusses the current and potential use of deep neural networks (DNNs) in mental health treatment. Previously, diagnosis of a mental illness based on neuroimaging and health data was difficult, as the statistical relationships between features inside of the image were too complex to handle and interpret. However, the ability of DNNs to extract patterns across multiple data sources, such as functional magnetic resonance imaging (fMRI), structural MRI (sMRI), and position emission tomography (PET), allows the integration of raw neuroimaging data directly. This capability of DNNs have shown promise in classifying psychiatric disorders, such as dementia, Alzheimer's disease and attention deficit hyperactivity disorder (ADHD), which could be attributed to the large availability of neuroimaging datasets capable of being fed into the neural network. Current findings suggest that DNNs can achieve, and also surpass, the diagnostic accruacy of both shallow models like support vector machines (SVM) and experienced clinicians.

3.3 Chatbots

Casu, et al. (2024) explains how diverse the technologies used in chatbot development are,

including natural language processing (NLP), reinforcement learning (RL), deep learning (DL), and a rule-based system, along with emphasizing the promising results of recent studies. For example, the XiaoE chatbot significantly reduced depressive symptoms in college studies, utilizing a combination of NLP, RL, DL, and a rule-based system. Similarly, Peuters, et al. (2024) found positive effects on physical activity, sleep, and mood through a chatbot designed to promote mental health in adolescents, mainly using NLP. Other studies, like Ulrich, et al. (2024), showed that a chatbot-based app improved mental well-being in individuals with migraines, while Vereschagin, et al. (2024) found that the Minder app reduced anxiety, depressive symptoms and substance use among university students.

4. RESULTS

The findings demonstrate that ML, particularly DNNs and NLP, are becoming increasingly researched and integrated into diagnostic tools and patient support systems (chatbots). DNNs, for example, have shown great promise in analyzing neuroimaging data to classify psychiatric disorders. Chatbot-based interventions, through standalone methods along with multi-component mental health intervention methods, demonstrated significant reductions in depressive symptoms and improvements in well-being. The effectiveness of chatbots and diagnostic tools was generally high, with the majority showing improvements in mental health outcomes.

Despite these promising results, challenges remain, particularly in terms of explainability and user engagement. Many ML models, especially those using deep learning, are often considered "black boxes," making it difficult for clinicians to trust their outputs without a clear understanding of how the model arrived at its conclusions. Engagement with ML-based chatbots, though beneficial, has also been inconsistent, with some users reporting difficulties in maintaining interaction. These challenges indicate that while ML-based tools hold substantial promise for mental health applications, further work is required to improve their usability, ensure model transparency, and increase user engagement. Future work should focus on enhancing the personalization of these tools, refining the models to be more user-friendly, and expanding their application to diverse patient populations to ensure broader effectiveness.

5. CONCLUSION

This meta-study shows the potential of machine learning within mental healthcare, analyzing different applications and how their validity is tested. Machine learning-assisted tools, like chatbots and predictive analytics, show promise for increasing accessibility and accuracy within psychotherapy. The aim of this study was to contribute to the growing discussion around the usage of machine learning based tools in sensitive fields such as psychotherapy.

6. FUTURE WORK

Future research should involve complete transparency within clinical settings, building trust among clinicians and those seeking mental health treatment. Ensuring machine learning-assisted tools result in safe, unbiased outcomes for patients should be prioritized. This would involve collaboration between experts in fields like psychotherapy, machine learning and ethics. Ultimately, future implementation of machine learning in mental health should revolve around the human experience and focus on enhancing, rather than completely altering, the field of psychotherapy.

REFERENCES

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI chatbots for mental health: A scoping review of

effectiveness, feasibility, and applications. *Applied Sciences,* 14(13), 5889. https://doi.org/10.3390/app14135889

Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Molecular Psychiatry,* 24(11), 1583–1598. https://doi.org/10.1038/s41380-019-0365-9

Gooding, P., & Kariotis, T. (2021). Ethics and law in research on algorithmic and data-driven technology in mental health care: Scoping review. *JMIR Mental Health,* 8(6), e24668. https://doi.org/10.2196/24668

He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., & Tian, T. (2022). Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: Single-blind, three-arm randomized controlled trial. *Journal of Medical Internet Research,* 24(11), e40719. https://doi.org/10.2196/40719

Peuters, C., Maenhout, L., Cardon, G., De Paepe, A., DeSmet, A., Lauwerier, E., & Crombez, G. (2024). A mobile healthy lifestyle intervention to promote mental health in adolescence: A mixed-methods evaluation. BMC public health, 24(1), 44.

Shatte, A., Hutchinson, D., & Teague, S. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine,* 49(09), 1426–1448. https://doi.org/10.1017/S0033291719000151

Ulrich, S., Gantenbein, A., Zuber, V., Von Wyl, A., Kowatsch, T., & Künzli, H. (2024). Development and evaluation of a Smartphone-based chatbot coach to facilitate a balanced lifestyle in individuals with headaches (BalanceUP App): Randomized controlled trial. *Journal of Medical Internet Research*, 26, e50132. https://doi.org/10.2196/50132

Vereschagin, M., Wang, A. Y., Richardson, C. G., Xie, H., Munthali, R. J., Hudec, K. L., Leung, C., Wojcik, K. D., Munro, L., Halli, P., Kessler, R. C., & Vigo, D. V. (2024). Effectiveness of the Minder Mobile Mental Health and Substance Use Intervention for university students: Randomized controlled trial. *Journal of Medical Internet Research,* 26, e54287. https://doi.org/10.2196/54287