International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

Navigating The Complexities Of Machine Learning In Mental Health Prognosis: Paradigms And Challenges

Daivashala Deshmukh[1], **Dr. Smita Kasar**[2]
[1*]Maharashtra Institute of Technology, Chhatrapati. Sambhajinagar, Maharashtra, India.
[2]Maharashtra Institute of Technology, Chhatrapati. Sambhajinagar, Maharashtra, India.
E-Mail Id's :- daivashaladeshmukh@gmail.com[1]*, smitakasar@gmail.com[2].

Abstract:
Mental disorders are an increasing worldwide problem, and traditional ways of diagnosing mental health are often restricted by subjectivity and use of retrospective information. The problem discussed in this paper is the lack of accuracy and generalizability of existing mental health prognosis systems. The overall goal is to review the opportunities and obstacles of implementing Machine Learning (ML) methods in mental health prediction and treatment optimization. The literature-based approach was used, whereby the research reviewed literature published in 2015-2022 to determine the data, model, implementation, and ethical challenges. Results indicate that challenges related to poor data quality, limited diversity, model bias, issues to do with interpretability and absence of real-world validation impede successful deployment. Findings reveal that although, ML algorithms such as SVM, random forest, and deep learning models have potential, their applicability to clinical practice is limited. The study finds that there is a need to have strong, interpretable, and ethically viable ML strategies. The future looks promising now with the use of deep learning, multimodal data, transfer learning, and federated learning to deploy scalable, accurate and patient-centred mental health solutions.
Keywords: *Electronic Health Records, Federated Learning, Machine Learning, Mental Health, Transfer Learning.*

1. INTRODUCTION

Mental health problems are becoming increasingly pressing around the world and impacting an increasing number of people, which is a very great problem for healthcare. Conventional approaches of evaluating and diagnosing mental disorders and prognosis include question and answer of the patient, interviews, and questionnaires, which are not very accurate and are restricted to past data [1]. Mental disorders are identified as a change in emotion, thought, or behaviour, or a combination of these. Panic and difficulty in acclimatizing to social, occupational, or kin settings may stem from such circumstances [2]. ML approaches can prospect for fresh ways of understanding tendencies of human behavior, as well as risk and sign factors linked with mental health, prognosis of the diseases, optimization, and individualization of treatment [3]. Using algorithmic ML to predict mental health has three advantages: it boosts the accuracy and timeliness by processing numerous Electronic Health Records, social media activity, wearable devices data, in addition to genetic data. This helps in the indication of early onset of mental health disorders, thus facilitating early use of accurate interventions aiming at improving the results of treatment of persons with mental disorders; thus, decreasing the effects of adverse mental health disorders. These strategies should be applied where childhood mental health problems are being identified early enough [4].

In the work of Cho G, Yim [5], the authors explain which ML algorithms are mostly used in the mental health area and provide real examples of their implementation in this sphere. Several ML techniques, such as SVM, GBM, Random Forest, K-Nearest Neighbors (KNN), and Naïve Bayes, are used to predict mental health. A classification method based on training data is called supervised learning. In contrast to supervised learning, unsupervised learning does not need previous advice on how to anticipate things [6,7]. The objective of this work is to investigate the problems and restrictions that hinder the use of ML techniques in this particular sector. Moreover, it will discuss its possible subjects and scopes; also seek to discover what remains to be addressed in this area for new research projects.

2. LITERATURE REVIEW

2.1 Literature Search & Selection

We re-examined research from 2015 to 2022 in order to concentrate this review on freshly released material. This is neither a systematic review nor does it provide a comprehensive list of all published research that satisfies these general requirements. To find studies that carried out novel clinical research in a field related to machine learning and mental health, we used PubMed and Google Scholar. Studies that discussed possible uses of machine learning or the creation of algorithms or systems that had not yet been put to the test in a real-world setting were excluded from our analysis. Additionally, studies on mental illnesses (such as depression, anxiety disorders, etc.) were excluded. Diagrammatic representation, Figure 1 shows the process of literature Search & Selection, how we have reviewed the existing studies.

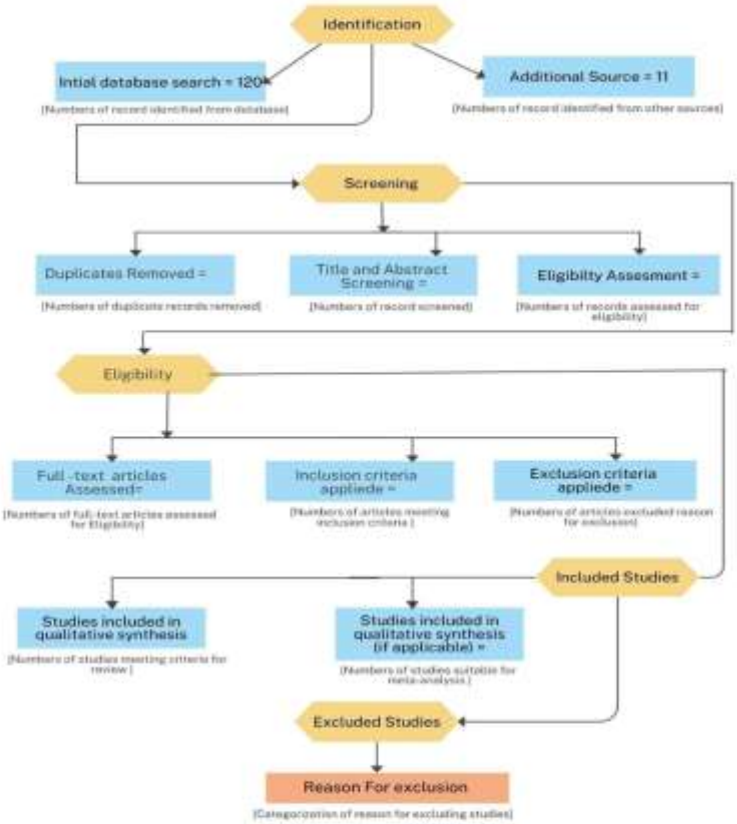

Figure 1: Process of Literature Search & Selection

2.2 Review Objectives

The primary goal of this research is to comprehensively identify and address the main issues associated with the use of machine learning in mental health. Achieving this requires an analysis of the current state of the research, involving important challenges such as data quality and accessibility, and ethical considerations. In order to enhance the efficiency and acceptability of ML technologies in mental health treatment, the review aims to fully understand these barriers and present insights and recommendations that can guide future research and implementation work in Figure 2.

Figure 2: Review Objective: Key Challenges

This research paper adheres to the established framework of a systematic literature review, as depicted in Figure 3. Identifying the main obstacles to applying machine learning to mental health entails a number of crucial areas, as the graphic illustrates. Issues with data availability, variety, and quality are examples of data-related obstacles that may have a big influence on ML models' performance and generalizability. Model-related challenges encompass the difficulties in selecting, training, and validating appropriate models to ensure they are both accurate and robust. Implementation challenges highlight the barriers in translating ML models into real-world clinical practice, including integration with existing systems and gaining clinician acceptance. Furthermore, ethical and privacy obstacles include worries about data privacy, ethical issues, and regulatory compliance. These issues must be resolved in order to preserve confidence and guarantee the appropriate use of machine learning technology in mental health treatment.

Figure 3: Existing Reviews in Various Mental Health Problems& ML approaches

2.3 Data-Related Challenges

There are multiple issues that may occur when using the methodology of ML to predict mental health, in particular in regard to data characteristics, e.g. quality, diversity and availability in Figure 4.

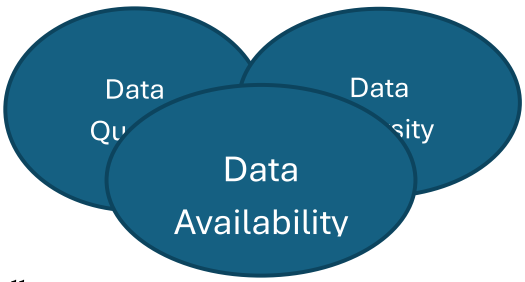

Figure 4: Data-related challenges

Data Quality:

The problems that are associated with the quality of data and preprocessing have been discussed by several research article authors. They had concerns of missing data, noise and any form of inconsistency [8][9]. A lot of it has to do with the quality of raw data which we are inputting into models. The latter, a mental health predictor, can be collected with the aid of a great variety of resources, including questionnaires, electronic health records, and smartphone logs.Appropriate methods to address such kind of errors are pre-processing techniques, noise removal, and gap filling by imputation. Such data must be free of any biased or misleading results, therefore, there should be accuracy and consistency [8]. Jakobsen et al. (2017) provide a detailed account of the MICE method of dealing with missing data and its application to mental health. The lack of data in trials lowers statistical power, so it is necessary to carefully analyze and describe such limitations.. Preventing missing data is better than a cure and should be practically targeted to implement effective prevention strategies in order to have more high-quality trials. It is necessary to treat the results obtained from datasets with incomplete values [9]. The authors Smalheiser and Torvik (2008) discuss several signal processing methods that can be used to reduce noise within data, like wavelet transforms, filtering techniques, among others. Data standardization and normalization are important for consistency across a range of mental health studies, including ML, as argued by Bishop (2006). Chu et al.'s (2016) study also looks at how automated data cleaning tools are used to detect and resolve inconsistencies in large datasets. In Suresh's paper "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle", Suresh and Guttag (2021) talk about methodological issues such as sample size calculations, appropriate statistical tests for various types of medical research papers, addressing confounding factors when comparing different groups; these issues directly affect the validity of conclusions drawn from such studies [10].

Data Diversity:

The availability of different data sets is critical to develop models that perform well across groups. Generally, this means that mental health issues should cover a range of ages; generations of generations; gender and race; and economic status. Without accounting for the fact that there is diversity, one can have models that perform poorly for some groups but are poor for others, and therefore of limited usefulness and meaning (BioMed Central, Nature). This also shows as clinical issues, behavioral issues, environmental factors, etc. The algorithm could be better able to forecast her mental health condition if it included a range of data types. A study by Mosteiro P., Kuiper J, Masthoff J., Scheepers F., and Spruit M (2020) highlights how machine learning models used in clinical psychiatry can be biased against a population. Authors require how they can be trained on data sets that do not represent them, so such training sets, when used, lead to inappropriate predictions, especially with small populations. Multiple anti-bias strategies, such as using the same data collection methods or standardizing results across research projects, are suggested. This paper employs data from university students of 17 institutions in Southeast Asia for forecasting psychological well-being and demonstrates the significance of the data set in model functioning and its implications for highlighting the general methodology. Research indicates that the addition of various information types adds a significant amount of robustness to predictive models of psychological well-being. By using health behavior data from large populations, researchers have sought to create models that are more valid and generalizable across populations [11]. In that paper, it explores the implications of ML algorithms on populations and information obtained through mobile sensing technologies. Although longitudinal studies have identified symptoms of mental illness in several demographic groups, the research aims to compare data from several longitudinal studies to test the utility and validity of models. The findings demonstrate the need to use data sets to construct more accurate and cross-sectional predictive models of mental health [12]. ML Strategies for Missing Data (Invisible Record): Some assumptions, goals, documentation, and analysis practices, and rationalizations for making the missing record visible can be altered with (ML) system New methods complex and holistic and so common in other ways that the probability of relatively simple delivery methods, with such populations, are shown to be limited to periodic rolls. These unclear pricing algorithms can be applied to the analysis of straightforward datasets, such as data-compressible devices; electronic health records (EHRs), records, and ergonomic records [13]. While statistical improvement methods can ameliorate the effects of missing or incomplete records, unmediated dependency on them can once again lead to models that are less amenable to a new, novel record or even a type of record that can be confusing to a reader but can also appeal to their confidence.

Data Availability:

Another challenge is to obtain the specific and long-term data. Psychological well-being is determined by complex interactions of trends over time. Thus, the denseness of data is a prerequisite to the possibility of explaining such developments and the creation of useful predictive models. The data access may be limited but due to privacy issues, standard data derivation methods, and unavailability of quality data in some places [17]. [18]. Combined efforts of research institutions, health care providers and technology companies, such as the use of good data collection instruments, good sharing of shared data to ensure that the data delivery is confidential, and better proactive measures to address quality issues will contribute to improving data quality and availability, ultimately; it is highly effective and equitable Complex predictive models.

2.4 Model-Related Challenges

The key elements in the prediction of mental health are probably the proper choice of ML models that should be created and their proper training. Many algorithms have different performances on different datasets and prediction tasks. This alternative can cause gigantic transformations in accuracy, generalizability and stability of predictions in mental health. The best approach to mental health prediction can be determined by a literature review of a variety of ML algorithms and their performance. Comparison of ML Algorithms: Shatte et al. compared various ML algorithms for the prediction of different mental health conditions, including SVM, DT, and neural networks. In general, the results indicated that ensemble methods, mainly RF and GB Machines, performed better in terms of accuracy and were more robust than single algorithms [19].

Deep Neural Network Models:

Study by Choi et al. (2016): In this study, deep learning models were pitted against traditional ML approaches in the task of mental health prediction using EHR records. The finding in this study is that RNN and CNN are better performers on patterns that identify mental health conditions [20].

Hybrid Models and Feature Engineering:

Lin et al. (2020) have contributed a study on the application of hybrid models in which SVMs are combined with deep learning for mental health prediction. So, we're honing in on feature engineering in order to increase model performance [21]. It showed that the integration of engineered features into the deep learning model improves predictive accuracy. Study by Torous et al. (2018): The current study assessed how transfer learning and domain adaptation techniques can aid in improving model performance across different mental health datasets [22]. Thus, the findings demonstrated that these approaches are able to generalize models to new populations and conditions since they borrow knowledge from related tasks.

Feature Selection and Engineering:

In our research class, we found that there are a ton of studies out there looking into different methods to pick and engineer features into mental health prediction models to increase their accuracy.

Domain Knowledge-Based Feature Selection:

Study by Saeb et al. (2016): According to this paper, domain knowledge is needed in feature selection. So, based on what we've learned in the clinical expertise labs and in the old studies, we found that sleep patterns, social activity, and physical activity are important features that are associated with mental health conditions. Saeb S, Zhang M, Kwasny MM, Mohr DC. "Mobile sensor zero correlates of depressive symptom severity in daily-life behavior: An exploratory study. Journal of Medical Internet Research. 2015".

Automated Feature Selection Techniques:

Study by Iniesta et al., 2016: In the study, the researchers explored automated feature selection methods, such as mutual information and recursive feature reduction, for selecting the most relevant characteristics from big datasets. These methods are useful for reducing the dimensionality of models and optimizing performance.

2.5 Implementation-Related Challenges

Feature Selection and Engineering:

The determinants of mental health conditions are exceptionally broad, genetic, environmental, psychological, which makes the task of selecting the appropriate features and engineering them a rather complicated one. The nature of subjectivity of mental health diagnosis further complicates this, as this can bring noise to the data models that will be trained on. The decision to apply which features is important, and the decision of what representation to apply can take ages of thought but basically, all you need to know is: how did the mental health disorder you are studying come to be [24].

Model Interpretability:

Predictive model interpretability is a critical dimension of mental health prognosis and in particular in the implementation of complex machine-learning systems with latent representations, such as a deep neural network, which is often a black box. Although these models may be highly predictive, they tend to be opaque, thus, rendering them inappropriate to be adopted in the clinical practice because doctors require assurance that the inferences which have been generated using such systems can be understood, hence, predictable. As suggested in reference [25], models ought to comply with the interpretability requirements, founded on clinical knowledge, which ought to be inseparable in the implementation of the same in the real life. Clinicians must be in a position to describe, as well as reason out treatment to their patients.

Generalization and Bias:

A big challenge is the generalization of machine-learning models across different populations. However, these models are often trained on specific datasets and can become biased to the idiosyncrasies of the training data, generating biased predictions when applied to cohorts beyond the intended demographics. To make matters worse, cultural and socio-economic preconceptions are inherent in mental-health datasets, which the algorithms may learn and reproduce. It is important to acknowledge these biases during the development stage to ensure the models have the correct generalizability and therefore serve as a basis for effective and equitable mental-health interventions [27].

Incorporation into Clinical Practice: It can be a formidable endeavour to incorporate the machine-learning models in clinical practice. Indeed, the clinical community often resists and ignores such novel technologies, in large part because they do not understand or have confidence in the internal mechanics of the algorithms. Moreover, infrastructural conditions necessary for the application of these models in regular use are often daunting and not universally available. The actual difficulty, though, lies in the necessity to keep learning and updating the models as data accrues over time brings new insight: Mental health is, by its very nature, a moveable and constantly changing field. The important question therefore is, how will it be ensured that these models are smoothly integrated with the existing clinical processes?

Metrics and Benchmarks:

This renders it very difficult to measure the efficacy of the ML applications within the mental health sector since these measurements lack standardization. Despite the existence of metrics, including accuracy, precision and recall, these metrics are not indicative of the complexity of prognosticating mental disorders. The measures also lack the multidimensional nature of psychological outcomes and therefore, their validity in predicting mental health cannot be considered valid. Most of the models that exist though are validated with short term measures and this begs the question of whether the models are able to show long term performance. This time limit confuses our notion of the predictive gains not only being evanescent, but also very persistence in the context of the patient continuum. This next stage in the development of this immature region, then, is to establish robust gauges of appraisal and standards pegged on the novel circumstances of mental-health prognostication. Much later, and only then, will we be in a position to quantitatively compare actual effects of new computational interventions against strict domain specific benchmarks.

2.6 Ethical & Privacy Challenges

Data Quality and Availability: Data quality and availability are two significant issues that will be solved in applying machine-learning to mental-health prognosis. Empirical mental health data has been termed as small, disproportional (in terms of class), and not representative enough of population variation and thus the resulting models tend to be biased and not highly representative. Secondly, it is crippled by the growing ethical issues of privacy and confidentiality, which may impede the use of high-quality data in sensitive fields, such as mental health. Clinical records, self-report measures, and online social media data are most heterogeneous, and therefore, it is inconsistent and unachievable to integrate them into one modelling framework.

Ethical and Legal Issues: ML-based methods of prognostication of mental health diseases are never discussed without raising profound ethical and legal concerns. There is a particular concern in securing informed consent to the use of patient data in the extraction and use of these data in model-training protocols, particularly when vulnerable populations are involved. Secondly, there is a possibility that these models may reinforce institutionalized biases, and in that regard, they will become a source of discriminatory mental healthcare practices. [33]. These requirements compound the anxieties; therefore, the agents operating in this field are advised to develop and deploy machine learning tools with high care and careful consideration. The requirement is, therefore, to enhance the rights of patients and ensure that there is fairness in the execution.

Patient Engagement and Feedback: Last but not least, patient engagement and feedback are essential elements of mental-health interventions, and such introduce special issues when introducing them with machine-learning methods. The data available to train models are not always full because patients do not always fully report their symptoms or behaviors. The other issue that has existed is the safeguarding of patient autonomy through the use of ML in making decisions. Simultaneously, there is an immediate need to continue patient interaction through active requesting of sufficient feedback, which can be incorporated into the model predictions, and in this way, the depiction of the final tools will be functioning as anticipated without violating patient rights [34, 35].

2.7 Gaps in the Research

The researchers would have to specify the challenges and limitations that confront them in learning the gaps in the literature of ML approaches in this field.

Small Sample Size: The majority of the evaluated papers, the authors note, either do not mention a sample size or only employ a small sample size in their research. Accordingly, depending on the parameters for the model used in the experiments, certain methods may function with a small sample without sacrificing accuracy, even if machine learning can be quite reliable when evaluating huge sample sizes.

Vabalas et al. observed that small sample sizes are often used in mental health since human subjects are expensive to obtain data on and experimental protocols for different diseases are still being developed [37].

Inadequate Validation: Many forms of research are still in the proof-of-concept stage because of tiny sample sizes and a lack of independently vetted sources. For instance, structural neuroimaging studies have often been conducted on individuals with a history of mental illness, which makes it very challenging to identify whether structural changes in the brain are the cause, outcome, or risk factor for the condition. A clinical expert who can provide crucial information on validation, truth, biases, etc. that might influence data analysis should work with the researchers to improve accuracy and manage deployment risks [39].

Limited Research in Deep Learning: Deep learning (DL) algorithms are thought to have worked well for a small number of applications, particularly in fields connected to health. The use of DL algorithms in mental health is still not well explored, however. In addition, DL algorithms are seen as a mystery, making it difficult for academics to define their purpose and mechanism of operation. According to recent research, efforts are being made to unlock the mystery of DL algorithms [40]. Researchers must persuade medical professionals to use the predictive mental health system, and this kind of investigation is crucial to their success.

Still, Lack of Real-Life Testing: Despite ML's ability to predict mental health outcomes for researchers, real-world testing is still lacking for a number of reasons. Numerous medical experts continue to question the precision of automated techniques, such as ML, and bring up concerns about consistency and complexity when integrating machine learning prediction systems into actual medical procedures. According to Dang et al., there is no set method for gathering high-quality data, difficulties in assigning labels, and a failure to recognize the best practices for managing ML models, all of which contribute to the inconsistent nature of supervised learning techniques [41]. The practical use of ML models in the area of mental health may be diminished by these obstacles and causes.

3.6 Avenues for Future Research.

Next, the review will indicate the specific methodologies that should help in progressing the research toward the effectiveness of the ML application in this field.

Research in DL: The success of using machine learning techniques to predict mental health may be expanded to include DL techniques. In addition to diagnosing additional long-term diseases like diabetes, cancer, and so on, these apps may also be able to forecast mental health issues. DL architectures in image processing may be used to recognize and forecast mental health issues based on face expressions. To improve clinical designs with greater accuracy, deep learning architectures should be able to integrate with memory [43] and attention processes [44]. In addition to brain MRI, photographs or recordings of speech, sociodemographic variables, clinical histories, and facial microexpressions are often used. This multimodal approach creates a broader data set and maximizes the applicability and relevance of deep-learning methods. The very development of such extensive and abundant data highlights an old concern in the mental-health circles and, by extension, the necessity of interinstitutional cooperation [45].

Nonetheless, the other two fundamental directions in research are the development of neural network architectures, particularly the architectures that tackle the complexity of data collected on mental health. Advanced neural networks like transformers and graph neural networks and the temporal and contextual relationship that are highly prized in the prediction of mental health have been used to model complex relations in high-dimensional data. These architectures support the high level longitudinal data analysis that can be utilized to arrive at a subtle knowledge regarding the dynamics of mental health conditions over the years. Secondly, multimodal learning is a methodology that integrates different kinds of information, such as text, images and physiology, which has emerged swiftly and is a useful tool.

The combination of these various, non homogenous modalities allows researchers to create the most holistic models and greater representation of the multidimensional mind of mental health.

Precise Predictive Models: Future research should focus on the use of new machine-learning models to make prognostic clinical predictions. In order to apply the analytically robust prediction models to clinical decision support (differentiating psychotic disorders, optimising pharmacotherapy, and prophylaxis) a web-based prognostication portal and combined medical analytics system is required. An example of this is the creation of Psycho Web, an app that is based on machine learning to aid in the collection of information about mental health patients, and possibly predict their data [46]. However, this program is still on its initial phases, and more improvements are being done.

XAI-Explainable Model: ML models must be executable and, at the very least, able to explain mental health issues. Before using the underlying system of categorization and prediction in the real world with patients, humans, including medical professionals, need to have a thorough grasp of it. Ensuring that the outcomes of these models are comprehensible is the primary goal in developing trustworthy systems. More specifically, as described in Holzinger et al. [47], the paper introduces a novel, interactive conditional explanation model for counterfactual graphs that can help to facilitate constructive human-machine interactions with artificial intelligence-based systems. XAI is well-positioned within this framework to bridge the gap between complex algorithmic design and the need for interpretability among mental health practitioners whose role is to make good clinical decisions. Finally, XAI can be used to establish confidence and adoption of ML tools within the clinical environment because it offers clinicians with a clear explanation of why the model made the predictions it did and because they were not just statistically sound but also interpretable and actionable [48].

Transference learning and adaptive algorithms

Transfer learning, a paradigm that supports the generalization capabilities of machine-learning constructs, is an important paradigm, designed with versatile applicability. It will be possible to significantly enhance clinical milieus by its implementation since the technique has undergone a broad exploitation in other fields where the image processing paradigms are heavily relied upon [49]. Perhaps the most crippling liability to the mental health projects in the transition phase will be the inherently volatile nature of ingestible data, and unproblematic algorithmic plasticity of critical concern. To reduce the risk of disastrous forgetting, the ML systems are to be instantiated within a continuous-learning framework, therefore, to infuse the tenets of a lifelong learning paradigm [50]. To exploit such imminent opportunities, there will be a need to engage in a leading partnership among clinicians, computer scientists, and data scholars.

Moreover, two of the most interesting methods that might influence the generalizability and application of machine learning models to a broad field the most are transfer learning and federated learning. In simple terms, transfer learning enables one to use models that are trained on source tasks to a related target task with data scarcity as the problem; which is highly useful in the psychiatric and mental-health domain [52], [53]. Conversely, federated learning allows the creation of predictive modelling models within data repositories that are geographically distributed and the patient major of concern secret in mental health studies. Such methods will give a good foundation of more robust and scalable ML solutions, since the challenges associated with data availability and bias will be alleviated.

Real-Time Data Analysis and Tailored Interventions: The final, but, by no means, the least is the emerging scholarly enthusiasm in real-time data analytics with customized interventions in the Mental Health Care system. It is enabled by the creation of wearable objects and mobile health apps that continuously monitor the physiological and psychological status of patients and provide them with personalized feedback and interventions. It is possible to provide a responsive system that actively responds to the changes in the mental condition of a patient through this live data stream and machine learning to gain a better understanding of it, in order to provide more timely and person-centered assistance. They are important courses of action which can be capitalized on to improve the accuracy, interpretability and clinical usefulness of the machine-learning models in the field of mental-health prognosis through these methodological advances and new technologies [54].

CONCLUSION

As it has been stressed, machine learning implementation in the sphere of mental health outcomes prediction is fraught with several important barriers that hinder its application. The availability and quality of data is feared, and mental health data is restricted and disproportional in general. This makes the selection and engineering of features a tricky task since the diagnosis in its current form is subjective and the factors that can affect mental health are heterogeneous. Continued attempts at explaining AI practices, with a perspective to comprehending predictive mechanisms, is a powerful poke at the significance of interpretability to machine learning models in healthcare. The recent research is an attempt to make sure that predictive models are not only fair, but they also generalize well across different groups of people - an attempt to remove long-standing worries on the issue of generalizability and algorithmic bias. The development of robust evaluation scales and the straightforward connection of these models to the clinical workflow can be used to advance this field. Nevertheless, even with the advancements, a number of concerns have been raised that underscore the need to continue research on how to enhance the efficiency and generalizability of machine-learning approaches to the mental health domain, in general.

REFERENCES:

1. A. Thieme, D. Belgrave, and G. Doherty, “Machine Learning in Mental Health,” ACM Transactions on Computer-Human Interaction, vol. 27, no. 5, pp. 1–53, Oct. 2020, doi: https://doi.org/10.1145/3398069.
2. American Psychiatric Association (APA) (n.d.). What Is Mental Illness? Retrieved from https://www.psychiatry.org/patients-families/what-is-mental-illness.
3. Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ‘18). ACM, Paper 582, 18 pages. DOI: https://doi.org/10.1145/3173574.3174156
4. A. J. Xu, M. A. Flannery, Y. Gao, and Y. Wu, “Machine learning for mental health detection,” 2019, https://digitalcommons.wpi.edu/mqp-all/6732/.
5. Cho G, Yim J, Choi Y, Ko J, Lee S-H (2019) Review of Machine Learning Algorithms for Diagnosing Mental Illness. Psychiatry Investigation 16:262–269. doi: https://doi.org/10.30773/pi.2018.12.21.2
6. Adrian B.R. Shatte, Delyse M. Hutchinson, and Samantha J. Teague. 2019. Machine learning in mental health: a scoping review of methods and applications. Psychological Medicine, 1-23. DOI: https://doi.org/10.1017/S0033291719000151
6. Adrian B.R. Shatte, Delyse M. Hutchinson, and Samantha J. Teague. 2019. Machine learning in mental health: a scoping review of methods and applications. Psychological Medicine, 1-23. DOI: https://doi.org/10.1017/S0033291719000151
7. A. Sau and I. Bhakta, “Screening of anxiety and depression among seafarers using machine learning technology,” Informatics in Medicine Unlocked, vol. 16, Article ID 100228, 2019.
8. Su C, Xu Z, Pathak J, Wang F (2020) Deep learning in mental health outcome research: a scoping review. Translational Psychiatry 10. doi: https://doi.org/10.1038/s41398-020-0780-3
9. Jakobsen JC, Gluud C, Wetterslev J, Winkel P (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology 17. doi: https://doi.org/10.1186/s12874-017-0442-1
10. Suresh H, Guttag J (2021) A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. Equity and Access in Algorithms, Mechanisms, and Optimization. doi: https://doi.org/10.1145/3465416.3483305
11. Hanif Abdul Rahman, Kwicklis M, Mohammad Ashraf Ottom, Areekul Amornsriwatanakul, Abdul-Mumin KH, Rosenberg M, Dinov ID (2023) Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students. Bioengineering 10:575–575. doi: https://doi.org/10.3390/bioengineering10050575
12. Adler DA, Wang F, Mohr DC, Choudhury T (2022) Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies. PLOS ONE 17:e0266516. doi: https://doi.org/10.1371/journal.pone.0266516
13. Phung S, Kumar A, Kim J. A deep learning technique for imputing missing healthcare data. Conf Proc IEEE Eng Med Biol Soc. 2019;2019:6513–6.https://doi.org/10.1109/EMBC.2019.8856760 Epub 2020/01/18PubMed PMID:31947333.
14. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, Freitag DF, Benoit J, Hughes MC, Khan F, Slater P, Shameer K, Roe M, Hutchison E, Kollins SH, Broedl U, Meng Z, Wong JL, Curtis L, Huang E (2021) The role of machine learning in clinical research: transforming the future of evidence generation. Trials 22. doi: https://doi.org/10.1186/s13063-021-05489-x
15. Liu Y, Gopalakrishnan V. An overview and evaluation of recent machine learning imputation methods using cardiac imaging data. Data (Basel).2017;2(1):8. https://pubmed.ncbi.nlm.nih.gov/28243594/.
16. Luo Y, Szolovits P, Dighe AS, Baron JM. 3D-MICE: integration of crosssectional and longitudinal imputation for multi-analyte longitudinal clinical data. J Am Med Inform Assoc. 2018;25(6):645–53. https://doi.org/10.1093/jamia/ocx133.
17. Brnabic A, Hess LM (2021) Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. BMC Medical Informatics and Decision Making 21. doi: https://doi.org/10.1186/s12911-021-01403-2
18. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, Freitag DF, Benoit J, Hughes MC, Khan F, Slater P, Shameer K, Roe M, Hutchison E, Kollins SH, Broedl U, Meng Z, Wong JL, Curtis L, Huang E (2021) The role of machine learning in clinical research: transforming the future of evidence generation. Trials 22. doi: https://doi.org/10.1186/s13063-021-05489-x
19. Shatte ABR, Hutchinson DM, Teague SJ (2019) Machine learning in mental health: a scoping review of methods and applications. Psychological Medicine 49:1426–1448. doi: https://doi.org/10.1017/s0033291719000151
20. Choi E, Schuetz A, Stewart WF, Sun J (2016) Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association 24:ocw112. doi: https://doi.org/10.1093/jamia/ocw112
21. J. Lin, Y.H. Kuo, and W. Chou, "Predicting mental health problems among college students using a hybrid machine learning/deep learning approach," IEEE Access, vol. 8, 2020.
22. Torous J, Onnela J-P, Keshavan M (2017) New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Translational Psychiatry 7:e1053–e1053. doi: https://doi.org/10.1038/tp.2017.25
23. Ayanna Howard, Cha Zhang, and Eric Horvitz. 2017. Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. In IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, 1-7. DOI: https://doi.org/10.1109/ARSO.2017.8025197
24. Alhuwaydi AM. Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions - A Narrative Review for a Comprehensive Insight. Risk Manag Healthc Policy. 2024 May 21;17:1339-1348. doi: 10.2147/RMHP.S461562. PMID: 38799612; PMCID: PMC11127648.
25. Doshi-Velez F, Kim B (2017) Towards A Rigorous Science of Interpretable Machine Learning. arXiv (Cornell University). doi: https://doi.org/10.48550/arxiv.1702.08608

26. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N (2015) Intelligible Models for HealthCare. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15. doi: https://doi.org/10.1145/2783258.2788613
27. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021) A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys 54:1–35. doi: https://doi.org/10.1145/3457607
28. Topol EJ (2019) High-performance medicine: the Convergence of Human and Artificial Intelligence. Nature Medicine 25:44–56
29. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science. 2014 Mar 14;343(6176):1203-5. doi: 10.1126/science.1248506. PMID: 24626916.
30. Chicco, D., Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6 (2020). https://doi.org/10.1186/s12864-019-6413-7
31. Cohen ZD, DeRubeis RJ. Treatment Selection in Depression. Annu Rev Clin Psychol. 2018 May 7;14:209-236. doi: 10.1146/annurev-clinpsy-050817-084746. Epub 2018 Mar 1. PMID: 29494258.
32. Squires M, Tao X, Elangovan S, Gururajan R, Zhou X, Acharya UR, Li Y. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Inform.
33. Char, Danton & Shah, Nigam & Magnus, David. (2018). Implementing Machine Learning in Health Care - Addressing Ethical Challenges. The New England journal of medicine. 378. 981-983. 10.1056/NE.
34. Lavallee DC, Lee JR, Austin E, Bloch R, Lawrence SO, McCall D, Munson SA, Nery-Hurwit MB, Amtmann D. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. Mhealth. 2020 Jan 5;6:8. doi: 10.21037/mhealth.2019.09.
35. Coghlan, S.; Leins, K.; Sheldrick, S.; Cheong, M.; Gooding, P.; D'Alfonso, S. To chat or bot to chat: Ethical issues with using chatbots in mental health. Digit Health 2023 9. doi: 10.20552076231183542.
36. Wykes, Til & Lipshitz, Jessica & Schueller, Stephen. (2019). Toward the Creation of Ethical Guidelines Endorsement for the Realm of Digital Mental Health and Its Applications. Current Treatment Options in Psychiatry. 6. 10.1007/s40501-019-00180.
37. A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, Article ID e0224365, 2019.
38. S. K. Ernala, M. L. Birnbaum, K. A. Candan et al., "Methodological gaps in predicting mental health states from social media," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, UK, May 2019
39. S. K. Ernala, M. L. Birnbaum, K. A. Candan et al., "Methodological gaps in predicting mental health states from social media," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, May 2019
40. W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K. R. Muller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2660–2673,2017.
41. Y. Dang, Q. Lin, and P. Huang, "AIOps: real-world challenges and research innovations," in *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering:Companion Proceedings (ICSE-Companion)*, Montreal, Canada,May 2019
42. Vaswani, Ashish, et al. "Attention Is All You Need." ArXiv.org, 5 Dec. 2017, arxiv.org/abs/1706.03762.
Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4-24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.
43. S. Hochreiter and J. Schmidhuber, "Long short-term memory,"*Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
44. K. M. Hermann, T. Kocisk´y, E. Grefenstette et al., "Teaching machines to read and comprehend," 2015, http://arxiv.org/abs/1506.03340.
45. S. Graham, C. Depp, E. E. Lee et al., "Artificial intelligence for mental health and mental illnesses: an overview," *Current Psychiatry Reports*, vol. 21, no. 11, 2019.
46. P. Morillo, H. Ortega, D. Chauca, J. Proaño, D. Vallejo-Huanga, and M. Cazares, "Psycho web: a machine learning platform for the diagnosis and classification of mental disorders,"in *Advances in Neuroergonomics and Cognitive Engineering*,pp. 399–410, Springer International Publishing,Berlin, Germany, 2019.
47. A. Holzinger, B. Malle, A. Saranti, and B. Pfeifer, "Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI," *Information Fusion*, vol. 71, pp. 28–37, 2021.
48. Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." Nature Machine Intelligence, vol. 15, May 2019, pp. 206–215, https://doi.org/10.1038/s42256-019-0048-x.
49. B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *RadioGraphics*, vol. 37,no. 2, pp. 505–515, 2017.
50. R. Kemker, A. Abitino, M. McClure, and C. Kanan, "Measuring catastrophic forgetting in neural networks," 2017,http://arxiv.org/abs/1708.02072.
51. W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk,"*BMC Medicine*, vol. 9, no. 1, 2011.
52. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
53. T. Baltrušaitis, C. Ahuja and L. -P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 1 Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.
[54]. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, Murphy SA. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med. 2018 May 18;52(6):446-462. doi: 10.1007/s12160-016-9830-8. PMID: 27663578; PMCID: PMC5364076