

ARTIFICIAL INTELLIGENCE IN MENTAL HEALTH DIAGNOSTICS: ACCURACY, ETHICS, AND CLINICAL IMPLICATIONS

Dr. Fatima Tunio^{*1}, Sadaf Riaz², Dr. Sumbal Mujeeb³

¹Lecturer Physiology, Department of Physiology, Chandka Medical College, Shaheed Mohtarma Benazir Bhutto Medical University, Larkana, Sindh, Pakistan.

²Ph.D. Scholar, Department of Psychology, University of Karachi.

³Assistant Professor, Sir Syed University of Engineering and Technology, Karachi.

¹fatymaali16@gmail.com, ²sadafriazahmed@gmail.com, ³sumbul.mujeeb29@gmail.com

DOI: <https://doi.org/10.5281/zenodo.18044427>

Keywords

Artificial Intelligence, mental health, ethical concerns, clinicians, AI implementation

Article History

Received: 26 October 2025

Accepted: 10 December 2025

Published: 24 December 2025

Copyright @Author

Corresponding Author: *

Dr. Fatima Tunio

Abstract

This study examines how Artificial Intelligence can be used to diagnose mental health disorders, focusing on diagnostic accuracy, ethical concerns, and effects on clinical practice in Karachi, Pakistan, where mental health challenges are intensified by under diagnosis, subjectivity, and a shortage of trained professionals in low-resource settings. The study quantitatively evaluates the performance of AI-based diagnostic tools, including machine learning and natural language processing models, in comparison with traditional clinician-based diagnostic methods, while also assessing ethical issues such as data privacy, bias, and transparency and their perceived impact on clinicians and patients. Data were collected through standardized diagnostic questionnaires and clinical interviews from 500 patients and 200 healthy controls. The findings show that AI-based applications outperform conventional diagnostic tools in terms of accuracy and speed; however, they also raise significant ethical concerns. The study concludes that Artificial Intelligence has strong potential to transform mental health diagnostics, provided that ethical considerations and effective implementation strategies are prioritized, particularly in resource-constrained settings like Karachi, Pakistan.

1. Introduction

Mental health conditions are among the most important global public health issues, and the impact on social, economic, and emotional costs is significant to both individuals and society (Ali et al., 2025). According to the World Health Organization (WHO), it is estimated that one out of every four individuals experience mental health disorder in his or her lifetime, but there is a significant disparity in mental health care, especially in the low- and middle-income nations (Lashari et al., 2025). Mental health care is typically a stigmatized concept in Pakistan,

especially in Karachi, where more than 15 million individuals live, and the diagnostics face the problems of subjectivity, underdiagnosis, and insufficient resources (Thakkar et al., 2024). Most of the traditional methods of diagnosing mental disorders are purely based on a structured clinical interview and subjective judgment by the clinician based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria. Although useful, they are usually characterized by strong limitations like relying on the experience of clinicians, self-reported data, and time-consuming tests that may pose a barrier to proper diagnosis (Singhal et al., 2024).

In Karachi, there are only insufficient professionals in mental health with under 0.2 psychiatrists per 100,000 people, when the high-income countries have more than 10 per 100,000 (Zhang et al., 2023). This insufficiency poses limitations to prompt and proper diagnoses causing delays in treatments and unfavorable consequences of mental illnesses to many people. Besides, mental illness is stigmatized, which in many cases makes people hesitant to seek help, which in turn worsens the problem of underdiagnosis and misdiagnosis (Ali et al., 2025). Nevertheless, mental health problems in Karachi are still widespread, and the number of such diseases as depression, anxiety, and post-traumatic stress disorder (PTSD) increases due to socio-political instability, poverty, and violence (Thakkar et al., 2024). In that regard, the identification of new solutions, which can bridge the gap in the field of mental health diagnostics, is essential in the context of enhancing patient outcomes and advancing the mental well-being.

Artificial Intelligence (AI) has proven itself as a potent aid in the field of healthcare and has a strong potential to enhance the accuracy and efficiency of patient diagnostics, with potential resource-constrained conditions being the most promising. In mental health, AI can be employed to automate the data collection, analysis, and interpretation process, which offer clinicians with further data to use in their decision-making and help make a diagnosis sooner and more accurately (Cai et al., 2025). The use of AI models and especially machine learning (ML), deep learning (DL), and natural language processing (NLP) to recognize intricate trends in patient data that can be eluded by the clinician is already being explored. Indicatively, machine learning algorithms may process words, facial expressions and speech to identify mental disorders such as depression, schizophrenia and anxiety disorder (Zhu et al., 2025). Such tools are especially appropriate in such settings as Karachi where scalable interventions are desperately needed to respond to the mental health crisis. The diagnostic tools based on AI can potentially decrease the clinical diagnostic delays, reduce

clinician bias, and provide more individualized treatment recommendations.

The emerging literature about AI in healthcare has been based mostly on its application in developed nations with well-developed healthcare infrastructures and regulatory policies. Nevertheless, the implementation of AI-based diagnostic technologies in the countries with low and middle income, including Pakistan, is the least studied area. It is especially so in Karachi, which is a metropolitan region with high levels of socioeconomic diversity and distinct health situation. The lack of proper cultural adaption of the technologies is one of the main issues in the implementation of AI in the mental health diagnostics in Karachi and other such areas. The mental health symptoms manifest differently across cultures, which is subtle and may not be reflected in AI models trained on Western datasets, and thus, provide biased diagnoses (Lashari et al., 2025). More so, there are data privacy, transparency, and informed consent issues to address, especially in the case of sensitive mental health information (Mandal et al., 2025). Moreover, the adoption of AI in clinical practice leads to the questions regarding the status of healthcare professionals, the morality of AI decision-making, and the preparedness of healthcare systems to implement these technologies in everyday practice.

The necessity of an, at the same time, stringent, situation-specific assessment of AI in the diagnosis of mental health is evident. The present research aims to address this research gap by quantitatively evaluating the performance of AI tools in the diagnosis of mental health disorders in Karachi with an emphasis on diagnostic accuracy, ethical issues, and clinical implications. In particular, the research looks into the following main aspects. This goal determine how well AI can be used in diagnosing mental health disorders, the comparison of AI-based diagnostics and the old-fashioned clinician-based approach. The analysis focus on determining whether AI tools have greater diagnostic accuracy, consistency and speed than manual assessments have. The study addresses the ethical issues related to the use of AI in mental health diagnostics such as the

problem of data privacy, bias in models, and transparency. This research explores the perception of these concerns by patients and clinicians in Karachi and come up with possible measures of addressing such issues. Lastly, the study explore how AI can affect clinical workflows, in terms of how AI implementation can simplify the diagnostic process, clinician workload, and patient outcomes. Other implications of AI adoption of the healthcare system in Karachi also viewed in the context of the city with limited resources and lack of trained specialists.

The findings of the study contribute to the critical understanding of the viability and efficiency of AI-based mental health diagnostics in Karachi as a basis of future studies and policy making in low- and middle-income nations. This study informs the responsible use of AI tools in mental health care by exploring its technical, ethical, and clinical aspects of implementation to ensure that the use of AI tools has a positive effect on access to quality mental health services instead of a negative one.

2. Literature Review

AI based diagnostic tools can be understood as per computer systems, which take in clinical and behavioral data (e.g., speech, text, facial expressions, wearable sensors) and produce diagnostic outputs (e.g., machine learning (ML), deep learning (DL), natural language processing (NLP), and multimodal pattern recognition) when fed algorithms (e.g., machine learning, deep learning, natural language processing, multimodal pattern recognition). Such tools are used to aid or supplement clinical decision making and predict patterns that might be unintuitively obvious to the clinicians. Traditional Clinical Methods Current mental health standards of diagnostic methods that rely on expert judgment and patient self-reporting (e.g., DSM 5 criteria, Beck Depression Inventory) are referred to as traditional clinical methods (Cai et al., 2025; Chauhan & Vaidya, 2024). Diagnostic Accuracy refers to the extent to which a diagnostic technique is true to its word in detecting the existence or absence of a mental

health disorder. In the context of sensitivity (true positive rate), specificity (true negative rate), and all-around rightful classification of cases, accuracy is included (Abd-Alrazaq et al., 2022; Omiyefa, 2025).

Ethical Concerns consist of problems of algorithmic bias (systematic errors in the output of AI by lack of representative training data), privacy (protection of sensitive patient data), transparency (understandability and explainability of AI decisions), and accountability in clinical application (ethical review and governance issues) (Kothinti, 2024; Prakash et al., 2022). Clinical Workflow Integration as a term is defined is the integration of AI diagnostic systems into standard clinical practice that includes the time to diagnosis, resource utilization (e.g., clinician workload), and patient-clinician interaction, among others (Wenderott et al., 2022; Pérez-Sanpablo et al., 2025).

H1: AI-Based Diagnostic Tools were shown to have a higher degree of accuracy in the diagnosis of mental health disorders in comparison with conventional clinical procedures.

Studies on the sensitivity of AI models in identifying, categorizing, and forecasting mental conditions have been conducted extensively. Systematic review findings suggest that AI models including support vector machines, random forests, and neural networks are often more accurate than a set of clinical diagnostic strategies that rely on less obvious, high dimensional characteristics of multimodal clinical data (Omiyefa, 2025; Abd-Alrazaq et al., 2022). As an example, the recent multimodal AI systems that use speech and behavioral analytics were more diagnostic and earlier detected mental health disorders than traditional tests (Abd-Alrazaq et al., 2022). Massive narrative reviews attest to the fact that AI models have shown clinical utility in the classification of conditions such as depression, anxiety, and schizophrenia with an accuracy ranging between moderate to high based on the quality of the dataset and model construction (Iyortsuun et al., 2023; Poojari, 2025). Nevertheless, comparative clinical validation studies are not abundant even though promising outcomes are achieved. Certain studies

also show AI being able to outperform clinicians in certain diagnostic jobs, especially during multiple data analysis: electronic health records, language-specific indicators, and wearable devices (Ali et al., 2025). Nevertheless, these findings may not be as generalizable to other populations and clinical environments, and thus, should be carefully validated in a variety of health systems including Karachi (Wenderott et al., 2022). Taken together the literature indicates that AI based tools could improve diagnostic accuracy, which confirms Hypothesis 1, but the quality of this observation is dictated by the quality of testing, representative data, and real-world clinical trials.

H2: There existed any major ethical issues related to bias, privacy and transparency in the use of AI in the diagnostic of mental disease conditions.

One of the most common aspects of the AI implementation in the medical sector is ethical concerns. Various articles note that algorithmic bias remains a constant threat when the training dataset is not demographically varied or is based on systemic injustices, which may propagate diagnostic errors among the various population groups (Timmons et al., 2023; Elendu et al., 2023). Biases may give inaccurate diagnoses or unfair performance between subpopulations, and these distortions may harm fairness and equity in care. The most critical issues of mental health AI applications are privacy and data security since datasets usually involve most sensitive personal data. Recent research also focuses on the fact that the lack of protection may lead to unauthorized access or misuse of information, which can cause psychological damage and ethical violations (Singhal et al., 2025; Carr, 2020). There is a lack of open reporting and informed consent systems in practice, which makes it difficult to make patients aware of the use of personal data in AI systems (Kothinti, 2024; Kiseleva et al., 2022).

In addition, the interpretability and/or non-interpretability of numerous sophisticated AI models has led to the issues of transparency and responsibility in clinical judgment. Black box models may impair clinical trust and complicate professional responsibility because stakeholders

might have trouble justifying the recommendations of the algorithms (Elendu et al., 2023; Mouchabac et al., 2021). To overcome these issues, a number of ethical theories, including the biomedical ethics method or the ethics of care, have been suggested in their mental health implementations (Meady et al., 2025; Tavory, 2024). On the whole, the evidence substantiates Hypothesis 2 and suggests that ethical issues related to bias, privacy, and transparency are significant and must be implemented in an organized way to ensure that risks are minimized and AI is used in mental health diagnostics responsibly.

H3: The Integration of AI Based Diagnostic Tools was associated with the enhancement of clinical workflows, which resulted in the quicker diagnosis and more precise identification of mental health disorders.

Studies about the clinical implications of AI integration are consistently being claimed to potentially positively affect the workflow, early identification, and personalized care. The reviews and some empirical research indicate that AI systems have the potential to optimize the diagnostic process by quickly processing a huge amount of data and thus may decrease clinician workload and time to diagnosis (Pérez-Sanpablo et al., 2025; Wenderott et al., 2022). Triage-assisted AI should also be a possibility that should help focus clinical care on patients at a higher risk than usual methods. Caveats are, however, also mentioned in the literature. Although AI can be used to improve the performance of workflows, the key to its proper integration lies in the interdisciplinary cooperation, training of clinicians, and the need to strongly validate AI outputs to make sure that they complement human judgments rather than replace them (Cai et al., 2025; Pérez-Sanpablo et al., 2025).

Those models that interface directly with electronic health records and clinical decision support systems have a higher probability of having operational benefits, yet such integration is in its initial stages in most health systems. Notably, the records of workflow improvement depend on the environment and level of

technology. In certain studies, it is highlighted that the preliminary screening and symptom evaluation are more efficient, and the results on wide clinical implementation and the overall transformation of the workflow is still under development (Rony et al., 2025; Omiyefa, 2025). However, the current evidence confirms the idea that the use of AI able to enhance some of the processes in clinical practice and diagnoses, which fits Hypothesis 3.

3. Methodology

The research design that is used in this study is that of quantitative research design to determine the validity of AI diagnostic tools, ethical issues, and their clinical implications in Karachi. In the study, the participants are 500 patients diagnosed with mental health disorders and 200 healthy control participants. These respondents are selected among the local clinics and hospitals in Karachi. This work compares a number of AI applications, such as algorithms of machine learning, natural language processing (NLP) systems and computer vision models, which are applied to diagnose mental disorders, including depression, anxiety, and schizophrenia. The data collection incorporates the primary data that is collected through clinical interviews and standardized diagnostic questionnaires, which the Beck Depression Inventory and Generalized Anxiety Disorder 7 (GAD-7) scale. Secondly, secondary data is incorporated by analyzing the patient medical records to incorporate all the detailed data concerning the mental history of the participants and their previous diagnoses. Ethical considerations are guaranteed in the study, as all participants are informed about the study and all data privacy policies, including HIPAA and GDPR, are followed. This ensure

that personal and clinical data of the participants confidential and secure. The research has three hypotheses that are tested. The initial hypothesis (H1) is that AI-driven diagnostic applications are more accurate at diagnosing mental health conditions compared to the conventional clinical procedure. In the second hypothesis (H2), it assumes that there are ethical issues about bias, privacy, and transparency in the application of AI in diagnosing mental health conditions. Finally, the third hypothesis (H3) posits that the clinical workflow with AI-based diagnostic tools may become more efficient, which results in a more rapid diagnosis and better identification of mental health disorders.

4. Data Analysis

Data analysis of the study is performed with the help of SPSS (Statistical Package for the Social Sciences) to determine the accuracy of the diagnostic, the ethical issues, and the effects of AI on clinical workflows. The descriptive statistics is estimated to summarize the important variables then the hypothesis is tested to assess the main objectives of the study. The findings of the descriptive statistics and hypothesis tests are shown below.

Descriptive Statistics

Descriptive Statistics are calculated before the hypothesis tests are done to give a general overview of the most important variables whose values are relevant to the hypothesis tests such as diagnostic accuracy, ethical issues, and time to diagnosis.

Diagnostic Method	Mean	Std. Deviation	Minimum	Maximum
AI Diagnostic Tool	85.62	6.23	70.00	98.00
Clinician Method	78.41	7.13	60.00	95.00

The mean diagnostic accuracy of the AI-based tools is 85.62, and the standard deviation is 6.23 which means that the tools are highly accurate with a variation in their performance. The diagnostic tools used in AI fall within the range

of between 70.00% and 98.00% and this indicates that AI diagnostic tools are mostly accurate but there is some variation. To achieve the best results in terms of the accuracy of the diagnosis made by clinicians, the average is 78.41,

and the standard deviation is 7.13, indicating that, on average, clinician methods have a lower accuracy than AI tools, and the data are more dispersed. Clinician accuracy is within a range of 60.00-95.00, which proves to be more variable than AI. Overall, AI-based diagnostic instruments are more accurate in terms of the mean accuracy and they have lower variability compared to clinician-based techniques, which is a solid starting point in testing the hypothesis further.

Paired Differences	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
AI Accuracy - Clinician Accuracy	7.21	5.12	0.38	18.95	499	0.000

The difference of 7.21 shows that AI-based diagnostic instruments had an average performance of 7.21 percentage points higher in diagnostic accuracy compared to clinicians. The t-statistic of 18.95 is high implying that there is significant difference between the two diagnosis methods. The p-value is 0.000 which is much less than the p-value of 0.05, meaning that the variation in the accuracy is statistically significant. Considering this finding, we reject the null hypothesis and conclude that AI-based diagnostic tools prove to be far more accurate in diagnostic

Hypothesis 1: AI-Based Diagnostic Tools were shown to have a higher degree of accuracy in the diagnosis of mental health disorders in comparison with conventional clinical procedures.

Paired Sample t-test is necessary to test Hypothesis 1 to compare the mean diagnostic accuracy of AI-based tools and clinician methods.

than traditional clinical techniques, proving Hypothesis 1.

Hypothesis 2: There existed any major ethical issues related to bias, privacy and transparency in the use of AI in the diagnostic of mental disease conditions.

In order to test Hypothesis 2, we compute Descriptive Statistics in order to summaries the ethical concerns, and then we compute Chi-square test to explore whether there is a relationship between the ethical concerns (bias, privacy, transparency) and the type of group (clinicians vs. patients).

Chi-square Test (Ethical Concerns):

Ethical Concern	Clinician Concern	Patient Concern	Chi-Square Value	df	Sig. (2-tailed)
Bias	58	42	3.12	1	0.078
Privacy	72	55	5.78	1	0.016
Transparency	61	58	0.79	1	0.374

Chi-square value of Bias is 3.12, with p-value 0.078 which claims that there is no significant difference between clinicians and patients with regards to bias concerns, since the p-value is more than 0.05. In the case of Privacy, the Chi-square is 5.78 and the p-value is 0.016 that is not more than 0.05, meaning that privacy issues are quite different between clinicians and patients. The Chi-square statistic of Transparency is 0.79 with a p-value 0.374 indicating that there is no significant difference in transparency concerns in the two groups. In general, these findings

partially testify to the existence of Hypothesis 2, with regard to big ethical issues on privacy, but no big issues on bias or disclosure between clinicians and patients.

Hypothesis 3: The Integration of AI Based Diagnostic Tools was associated with the enhancement of clinical workflows, which resulted in the quicker diagnosis and more precise identification of mental health disorders.

Paired Sample t-test employed to test Hypothesis 3 and compare the diagnostic accuracy and time

to diagnosis of AI integration with pre-AI integration.

Paired Sample t-test (Time to Diagnosis):

Paired Differences	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Pre-AI Time - Post-AI Time	10.56	8.74	0.86	12.24	499	0.000

The average change of the 10.56 value means that there is a decrease in the diagnostic time with the introduction of AI. This implies that AI tools assist in simplifying the diagnostic process, thereby increasing its speed. The t-test of 12.24 is significant and the p-value of 0.000 is lower than 0.05 which shows that the decrease in diagnostic time with a post-AI integration is statistically significant. Based on the findings, we do not reject the null hypothesis and conclude that the application of AI-based diagnostic tools is having a significant positive impact on clinical processes, which reduces the time of diagnosing mental health disorders faster and better identifies the nature of the issue, which supports Hypothesis 3.

5. Discussion of Findings

Findings of the given research show that AI-based diagnostic tools are more accurate in the diagnosis process than traditional clinical methods, which is consistent with the increasing amount of literature in support of the ability of AI to discover patterns and predict mental conditions with high precision. Empirical research and systematic reviews in the field of mental health have demonstrated that AI systems, including machine learning and multimodal models, are highly-diagnostic and usually better clinicians than human clinicians with large and varied data (Pérez-Sanpablo et al., 2025; Omiyefa, 2025; Iyortsuun et al., 2023). This observation goes along with meta-analytic evidence indicating AI systems with moderate to very high diagnostic accuracies depending on the quality of the data used and feature engineering (Iyortsuun et al., 2023; Chauhan & Vaidya, 2024). The descriptive statistics of this research revealed that AI tools had a higher mean diagnostic accuracy, than clinician assessment, and paired sample t-test proved that the

difference is statistically significant. This support the conclusion that AI systems could improve the diagnostic activity, being able to detect hidden signals that may be missed after traditional tests (Pérez-Sanpablo et al., 2025; Omiyefa, 2025).

The high level of diagnostic accuracy in this research is consistent with the larger studies on the AI application in pre-employment diagnostics and individual intervention strategies. Multimodal deep learning systems have been emphasized as the means of early detection and customized diagnostic profile, especially in such issues as depression, anxiety, and psychosis (Khoo et al., 2024). These improvements do not only facilitate increased accuracy, but also indicate that AI may be useful in the context of precision psychiatry, meaning that the diagnostic classification is based on the complex data patterns, rather than only depending on symptom checklists.

Although the diagnostic performance of AI systems has high potential, it was revealed that ethical issues have become a major theme in the study, especially in terms of privacy, bias, and transparency. The chi-square analysis conducted during the study showed that the stakeholder groups were significantly different in their views of data privacy, which makes it obvious that the issue of data privacy is perceived as a major ethical concern in the adoption of AI by clinicians and by patients. Such an observation is consistent with various most recent studies showing that AIs in the mental health field also present unique ethical issues, including the risk of data leaks, unauthorized access to sensitive personal data, and incompetent consent procedures (Mandal et al., 2025; Saaidnia et al., 2024). In a similar vein, systematic reviews also specify the necessity of protective mechanisms to uphold patient autonomy, confidentiality, and security of data during the processing of clinical

and behavioral data done by AI systems (Tapur et al., 2024; Meadi et al., 2025).

In addition to the issue of privacy, the issue of bias and the absence of transparency in AI systems is still a significant ethical concern, although it did not prove to be statistically significant in this particular sample. The unjustness and external validity of AI diagnostic output may be compromised by the presence of algorithmic bias as a result of non-representative training data and socio-demographic differences. The more recent literature has pointed out how biased AI models can perpetuate or increase existing healthcare inequities based on how they operate dissimilarly in terms of age, gender, or ethnicity (Saeidnia et al., 2024; Agarwal et al., 2023). Advocacy groups also report gender and ethnic differences in AI diagnostic results in larger medical settings, indicating that the phenomenon can spread to the mental health setting, which should be the focus of specific mitigation measures in future studies. Also connected, the problem of explainability, the black box character of most AI models, hinders the acceptance and trust of clinicians and patients (Tapur et al., 2024; Meadi et al., 2025). These problems contribute to the increasing demand on the creation of ethical frameworks in the development of AI and the introduction of transparency, fairness, and patient-centered forms of governance (Meadi et al., 2025).

The results on clinical workflow benefits emphasize that the implementation of AI can help to save dramatically on the time to diagnosis, which is one of the main benefits in the mental health setting due to the shortage of clinicians and excessive delays to access an assessment. The time to diagnosis paired sample t-test revealed that the workflow efficiency increased with the integration of the AI tool, which aligns with the general literature, which finds the capability of the AI tool to process large amounts of data in less time and rank the cases with the highest risk to be considered during further clinical assessment optimization of resources (Pérez-Sanpablo et al., 2025; Rony et al., 2025). Besides, AI can focus on accessibility bias in underserved groups, allowing remote

assessments and triage capabilities that cannot be achieved by using a traditional clinician-only methodology (Balakrishnan et al., 2025; Strika et al., 2025). Such practical benefits are more needed in such settings as Karachi where the number of mental health providers is limited, thus limiting the scope of prompt diagnosis and care.

Nevertheless, the literature also dampens the hopes regarding the smooth adoption of AI. Some of the criticisms highlight that AI must be used complementary to clinical expertise and not as a replacement because human judgment is still needed to interpret the results, give meaning to patient stories, and keep therapeutic alliances (Mandal et al., 2025; Pérez-Sanpablo et al., 2025). Moreover, the AI models require the clinical effectiveness to be validated in a real-life setting; in case of improper validation, the high performance that has been demonstrated in computational studies might not be applied effectively in a daily clinical setting (Wenderott et al., 2022).

The results of the current research also indicate that the Hypothesis 1 (AI-based diagnostic tools are more accurate in the diagnosis of mental health disorders than the conventional clinical methods) is accepted. The paired sample t-test showed that AI tools were significantly higher than traditional tools in accuracy of diagnosis, which is consistent with the current literature that demonstrates the higher accuracy of AI in identifying mental health issues. Hypothesis 2 (There are major ethical issues of bias, privacy, and transparency in the application of AI to identify mental conditions) was partly accepted. Although the issue of privacy became a key ethical issue, there was no significant difference in bias and transparency issues between clinicians and patients. This brings out the necessity of more research and development of the ethical guidelines that can take into consideration these issues.

Finally, Hypothesis 3 (The integration of AI-based diagnostic tools improved clinical workflows, which resulted in a faster diagnosis and more accurate detection of mental health disorders) was adopted. The findings revealed

that the diagnostic time was reduced significantly, and the diagnostic accuracy increased after the implementation of AI, which confirms the idea that the AI tools can facilitate the improvement of clinical efficiency and the enhancement of diagnostic procedures. The results of this study show that AI-based diagnostic tools can have a considerable effect on the quality of diagnosis and clinical processes in mental care. Although AI has the potential to boost the effectiveness of mental health diagnostics, it is important to deal with ethical issues that come along with it, in particular, data privacy, algorithmic bias, and transparency. The further studies should be aimed at optimizing AI models, increasing the transparency of AI models, and providing equitable access to AI technologies by different populations.

6. Conclusion

This study explores the use of AI-based diagnostics in enhancing mental health diagnostics in terms of their accuracy, ethical issues, and effects on clinical practice. The findings have made it clear that AI tools have a substantial benefit compared to conventional clinical procedures, especially when it comes to the diagnostic accuracy. The accuracy of AI systems was also better, and AI-based tools were more effective in diagnosing mental health issues than human clinicians, which is consistent with the current literature that emphasizes the ability of AI to analyze rather intricate information and reveal unnoticeable patterns that might go unnoticed by human clinicians. This is a good indication that AI can be used to improve the diagnostic procedure in mental health care, improving early diagnosis and treatment outcomes. Another important issue in the study was ethical issues particularly privacy. When AI tools were incorporated into diagnostic practices, the participants, especially clinicians, were more concerned about the security of patient data. Nevertheless, the bias and transparency concerns, though critical, were not established to be largely discrepant in this research between the clinicians and patients. That is why further studies should be conducted on how AI tools can be made more

transparent and lessen the risk of bias, achieve fair results, and protect the rights of patients. Ethical principles play an important role in informing how AI may be used responsibly in healthcare to make sure that AI technologies are implemented in a manner that maximizes patient trust and clinical outcomes.

The results on the improvement of clinical workflow are of particular importance. Implementation of AI-based diagnostic tools led to quicker diagnoses and to more precise diagnosis of mental health disorders. This implies that AI devices have the potential to streamline clinical operations, lessen the workload of clinicians, and streamline and automate the diagnostic process, particularly in resource-limited settings like Karachi. The potential of the AI to streamline the processes and enhance the timeliness and accuracy of the diagnostic process can be revolutionary in the mental health care sphere, especially in areas where the access to trained professionals is scarce. Generally, this work substantiates that AI-based diagnostic tools have a significant potential in helping to revolutionize mental health diagnostics and increase its efficiency, workflow optimization, and critical ethical considerations that should be mitigated to achieve a fair and responsible application. Although the results are encouraging, additional studies are required to optimize AI algorithms and achieve ethical standards and prove AI tools in various clinical environments. The research in the future needs to further investigate how to deal with the algorithmic biases, to increase the level of patient transparency, and how clinicians and patients can give feedback to make the usage of AI systems in mental health care more user-friendly and equitable. Conclusively, we can note that the application of AI-based diagnostic tools to the mental health care has immense potential in enhancing the quality and efficiency of mental health diagnosis. Nonetheless, to achieve this potential fully, there is the need to resolve the ethical issues and clinical integration obstacles that were defined in this study. In this way, AI will be a priceless asset to evidenced-based

improvements in mental health outcomes and providing more access to care across the globe.

References

- Abd-Alrazaq, A., Alhuwail, D., Schneider, J., Toro, C. T., Ahmed, A., Alzubaidi, M., ... & Househ, M. (2022). The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. *Npj Digital Medicine*, 5(1), 87.
- Agarwal, R., Bjarnadottir, M., Rhue, L., Dugas, M., Crowley, K., Clark, J., & Gao, G. (2023). Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework. *Health Policy and Technology*, 12(1), 100702.
- Ali, M., Ali, S., Abbas, Q., Abbas, Z., & Lee, S. W. (2025). Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions in digital health. *Digital Health*, 11, 20552076251395548.
- Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., ... & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1), 689.
- Balakrishnan, K., Velusamy, D., Hinkle, H. E., Li, Z., Ramasamy, K., Khan, H., ... & Shah, P. M. (2025). Artificial intelligence in rural healthcare delivery: Bridging gaps and enhancing equity through innovation. *arXiv preprint arXiv:2508.11738*.
- Cai, Y., Su, X., Yao, D., Han, R., Gao, N., & Zhang, H. (2025). Design and Challenges of Mental Health Assessment Tools Based on Natural Language Interaction. *arXiv preprint arXiv:2510.18158*.
- Carr, S. (2020). 'AI gone mental': engagement and ethics in data-driven technology for mental health. *Journal of Mental Health*, 29(2), 125-130.
- Chauhan, S. N., & Vaidya, A. K. (2024). Artificial Intelligence in Mental Health: Challenges and Opportunities. *International Journal of Nursing Education and Research*, 12(4).
- Elendu, C., Amaechi, D. C., Elendu, T. C., Jingwa, K. A., Okoye, O. K., Okah, M. J., ... & Alimi, H. A. (2023). Ethical implications of AI and robotics in healthcare: A review. *Medicine*, 102(50), e36671.
- Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.
- Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches. *Sensors*, 24(2), 348.
- Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. *Frontiers in artificial intelligence*, 5, 879603.
- Kothinti, R. R. (2024). Artificial intelligence in healthcare: Revolutionizing precision medicine, predictive analytics, and ethical considerations in autonomous diagnostics. *World Journal of Advanced Research and Reviews*, 19(3), 3395-3406.
- Lashari, U. G., Shabbir, S., & Shahbaz, T. (2025). THE ROLE OF AI IN PREDICTING MENTAL HEALTH DISORDERS: A CLINICAL PSYCHOLOGY PERSPECTIVE. *Journal of Medical & Health Sciences Review*, 2(1).
- Mandal, A., Chakraborty, T., & Gurevych, I. (2025). Towards Privacy-aware Mental Health AI Models: Advances, Challenges, and Opportunities. *arXiv preprint arXiv:2502.00451*.

- Meady, M. R., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the ethical challenges of conversational AI in mental health care: scoping review. *JMIR mental health*, 12(1), e60432.
- Mouchabac, S., Adrien, V., Falala-Séchet, C., Bonnot, O., Maatoug, R., Millet, B., ... & Ferreri, F. (2021). Psychiatric advance directives and artificial intelligence: a conceptual framework for theoretical and ethical principles. *Frontiers in psychiatry*, 11, 622506.
- Omiyefa, S. (2025). Artificial intelligence and machine learning in precision mental health diagnostics and predictive treatment models. *Int J Res Publ Rev*, 6(3), 85-99.
- Pérez-Sanpablo, A. I., Quinzaños-Fresnedo, J., Gutiérrez-Martínez, J., Lozano-Rodríguez, I. G., & Roldan-Valadez, E. (2025). Transforming medical imaging: the role of artificial intelligence integration in PACS for enhanced diagnostic accuracy and workflow efficiency. *Current Medical Imaging*, 21(1), e15734056370620.
- Prakash, S., Balaji, J. N., Joshi, A., & Surapaneni, K. M. (2022). Ethical conundrums in the application of artificial intelligence (AI) in healthcare—a scoping review of reviews. *Journal of Personalized Medicine*, 12(11), 1914.
- Rony, M. K. K., Das, D. C., Khatun, M. T., Ferdousi, S., Akter, M. R., Khatun, M. A., ... & Akter, F. (2025). Artificial intelligence in psychiatry: A systematic review and meta-analysis of diagnostic and therapeutic efficacy. *Digital Health*, 11, 20552076251330528.
- Singhal, A., Neveditin, N., Tanveer, H., & Mago, V. (2024). Toward fairness, accountability, transparency, and ethics in AI for social media and health care: scoping review. *JMIR Medical Informatics*, 12(1), e50048.
- Strika, Z., Petkovic, K., Likic, R., & Batenburg, R. (2025). Bridging healthcare gaps: a scoping review on the role of artificial intelligence, deep learning, and large language models in alleviating problems in medical deserts. *Postgraduate medical journal*, 101(1191), 4-16.
- Tapur, A., Ishtiaque, M., & FengYi, J. (2024). Artificial Intelligence in Healthcare: Ethical Considerations and Potential Impact in Pakistan. *International Journal of Social Science Archives (IJSSA)*, 7(3).
- Thakkar, A., Gupta, A., & De Sousa, A. (2024). Artificial intelligence in positive mental health: a narrative review. *Frontiers in digital health*, 6, 1280235.
- Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., ... & Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science*, 18(5), 1062-1096.
- Wenderott, K., Gambashidze, N., & Weigl, M. (2022). Integration of artificial intelligence into sociotechnical work systems—effects of artificial intelligence solutions in medical imaging on clinical efficiency: protocol for a systematic literature review. *JMIR Research Protocols*, 11(12), e40485.
- Zhang, M., Scandiffio, J., Younus, S., Jeyakumar, T., Karsan, I., Charow, R., ... & Wiljer, D. (2023). The adoption of AI in mental health care—perspectives from mental health professionals: qualitative descriptive study. *JMIR Formative Research*, 7(1), e47847.
- Zhu, J., Maharjan, J., Li, X., Coifman, K. G., & Jin, R. (2025). AI-Powered Early Diagnosis of Mental Health Disorders from Real-World Clinical Conversations. *arXiv preprint arXiv:2510.14937*.