VIVA MEDIKA
Jurnal Kesehatan, Kebidanan, dan Keperawatan

Homepage: http://ejournal.uhb.ac.id/index.php/vm
P-ISSN: 1979-2026 E-ISSN: 2656-1034
DOI: 10.35960/vm.v18i3.2214

Artificial Intelligence in Biomedical Psychology: A Systematic Review of Clinical and Cognitive Applications

Annastasya Nabila Elsa Wulandari[1], Agung Budi Prasetio[2], Muhammad Ahmad Baballe[3], Taraknath Paul[4]
[1]Peneliti Teknologi Teknik Indonesia, Yogyakarta, Indonesia
[2]Faculty Computer Science, Institut Teknologi Tangerang Selatan, Indonesia
[3]Master of Engineering Academic, Nigerian Defence Academy, Kaduna, Nigeria
[4]Department of Computational Sciences Brainware University, Kolkata-700125, WB, India
[1]anstsya.new@gmail.com*; [2]agung@itts.ac.id, [3]mbaballe@kanopoly.edu.ng,
[4]drtnp.cs@brainwareuniversity.ac.in

ABSTRACT

Biomedical psychology emphasises psychological and neurocognitive assessment through the integration of biological, neurophysiological, and quantitative behavioural data to support clinical decision-making. However, conventional assessment approaches remain limited by issues of objectivity, scalability, and longitudinal monitoring, prompting the utilisation of artificial intelligence (AI) as a computational tool in clinical and cognitive contexts. This systematic review synthesises the application of AI in biomedical psychology with an explicit focus on assessment functions, rather than intervention or therapy, following the PRISMA 2020 guidelines through a systematic search of four major databases. The included studies cover a variety of clinical and cognitive applications with variations in psychological constructs, data modalities, and AI methods. The synthesis results show that AI is most often used for diagnostic classification, risk screening, and continuous estimation of cognitive functions and dimensional constructs. Differences in assessment objectives between clinical and cognitive domains reveal consistent methodological trade-offs related to model selection, validation strategies, and overfitting risks. As a key contribution, this review presents an assessment-oriented cross-domain synthesis and proposes fit-forpurpose design principles as a conceptual framework for developing robust, interpretable, and clinically relevant AI-based assessment systems

Keywords: Artificial Intelligence; AI-Based Assessment; Biomedical Psychology; Clinical Decision Support; Cognitive Assessment; Multimodal Machine Learning

1. INTRODUCTION

Biomedical psychology focuses on understanding and evaluating the relationship between biological processes and psychological functions, particularly in clinical and cognitive contexts. In the clinical realm, mental disorders are evaluated through evidence-based practices that integrate patient assessment, clinical documentation, and scientific findings into the professional decision-making process (Delcea et al., 2020). In cognitive contexts, neuropsychological evaluation assesses functions

***Annastasya Nabila Elsa Wulandari.**
Tel.: -
Email:anstsya.new@gmail.com

such as memory, attention, and executive processes to explain the impact of brain dysfunction on behaviour and functional capacity (Collinson et al., 2010).

Psychological assessment is placed as a key component in this practice, although conventional methods that rely on clinical interviews, self-reports, and standardised tests are known to be prone to subjective bias (Schwartz et al., 2023). Comprehensive psychological evaluations require time, professional expertise, and high costs, limiting the scalability and accessibility of services (Alkan et al., 2025). In addition, periodic assessment procedures often fail to capture dynamic psychological and cognitive changes over time (Zheng et al., 2025). These structural limitations highlight the need for quantitatively grounded assessment approaches that support population-level screening and longitudinal monitoring (Finley, 2024), (Patterson et al., 2025).

Biomedical psychology is operationally defined as an assessment approach that integrates biological, neurophysiological, and quantitative behavioural data to evaluate psychological and neurocognitive conditions that are directly relevant to clinical decision making (Engel et al., 2022). Unlike traditional clinical psychology, which emphasises qualitative interpretation, this approach focuses on the objective quantification of psychological constructs and their relationship with biological indicators (Ren et al., 2025). This approach is also distinguished from computational psychiatry, which is oriented towards mechanistic modelling of disorders, as the emphasis is on developing objective, scalable, and interpretable assessment instruments, without being directed towards intervention or therapy (Baydili et al., 2025).

Advances in AI and machine learning provide computational mechanisms capable of analysing complex behavioural and biological data at scale. Quantitative behavioural modelling has demonstrated potential in distinguishing individuals with mild cognitive impairment from healthy controls (Sikström et al., 2024). AI-based systems further enable decentralised and privacy-preserving assessment architectures, including federated learning approaches for mental health data (Dubey et al., 2025).

AI is positioned as a computational tool used to support psychological assessment without replacing the role of clinical professionals. AIbased systems are applied in initial screening, diagnostic classification, and estimation of the severity of psychological and cognitive disorders (Rakotomanana & Rouhafzay, 2025), and are utilised as decision support systems through the integration of multimodal data that can be interpreted by clinicians (Bae et al., 2025). The application of AI also covers various psychological constructs, such as measuring attachment using the Biometric Attachment Test (Parra et al., 2022), while DeepPerson was developed by integrating psychological theory and deep learning to improve the accuracy of text-based personality dimension detection (Yang et al., 2023).

The available literature shows fragmentation in the application of AI in psychological and neurocognitive assessment, with most reviews focusing on specific disorders, data modalities, or methodological approaches, thus limiting cross-domain synthesis. This situation indicates that reviews explicitly focusing on AI-based assessment functions and the differentiation of objectives, designs, and methodological tradeoffs between clinical and cognitive assessments within the biomedical psychology framework are still relatively limited. This systematic review aims to fill this gap through an assessment-oriented cross-domain synthesis, covering the domains of application, psychological constructs, data modalities, AI methods, and key methodological challenges.

Specifically, this review addresses three research questions:

- RQ1: What types of clinical and cognitive assessment tasks are applied using AI in biomedical psychology?
- RQ2: What data modalities and AI methods are most commonly used in AI-based assessment?
- RQ3: What methodological challenges and validation strategies are most frequently reported in AI-based psychological assessment studies?

2. METHODS

2.1 Review Desain

This review was conducted as a Systematic Literature Review (SLR) with reference to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Haddaway et al., 2022). A systematic literature search was conducted on four major databases, namely PubMed, Scopus, ScienceDirect, and SpringerLink, to obtain broad coverage of studies examining the application of AI in biomedical psychology. The search strategy was designed using a combination of terms representing AI methods, the domains of clinical psychology and neuropsychology, and the assessment context, with the following main search string examples: ("machine learning" OR "deep learning" OR "artificial intelligence") AND ("clinical psychology" OR "neuropsychology" OR "mental health" OR "psychiatry") AND ("assessment" OR "screening" OR "diagnosis").

Terminology adjustments were made on a limited basis according to the characteristics of each database, including the use of indexed terms in PubMed. The search was limited to peer-reviewed articles in English published between 2020 and 2026. The review focused on studies that utilised AI as an assessment tool to evaluate, classify, or estimate clinical conditions and cognitive functions based on human data. Meta-analysis was not applied due to the high level of heterogeneity between studies, so qualitative synthesis was used as the main approach. The extracted data were analysed using a narrative approach supported by categorical mapping based on application type, psychological construct, data modality, and AI method.

2.2 Eligibility Criteria

Inclusion and exclusion criteria were established to ensure that the included studies represented the application of AI as an assessment instrument in biomedical psychology. The selection process focused on studies that utilised AI as the main analytical component to evaluate clinical conditions or cognitive function based on human data through clearly defined quantitative measurements. Studies outside the scope of psychological assessment, purely technical in nature, intervention-oriented, or lacking adequate empirical validation were excluded from the analysis. The scope of this study was narrowed to maintain methodological coherence and clinical relevance. A summary of the inclusion and exclusion criteria is presented in Table 1.

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Machine/deep learning as primary analysis	AI used only for support, visualization, or exploratory analysis
AI applied for psychological or neurocognitive assessment	Intervention-, therapy-, training-, or neurofeedback-focused studies
At least one psychological or neuropsychological construct assessed	No psychological or cognitive construct assessed
Human-subject data	Animal, simulated, or synthetic data without human validation
Quantitative evaluation reported	No quantitative performance metrics
Peer-reviewed journal or conference paper (2020–2026, English)	Reviews, editorials, abstracts only, preprints, theses, or inaccessible full text

2.3 Study Selection

The study selection process was conducted in accordance with the PRISMA 2020 guidelines using predefined inclusion and exclusion criteria. At the identification stage, 1,067 records were retrieved from PubMed (7), Scopus (200), ScienceDirect (572), and SpringerLink (288). Prior to screening, 147 records were removed, including duplicate records (n = 2), book chapters (n = 1), and review articles (n = 144). Following removal, 920 records remained for title and abstract screening. During the screening stage, 863 records were excluded, comprising 602 exclusions based on title screening and 261 based on abstract screening. This process resulted in 57 reports being sought for full-text retrieval. Of these, three reports were not accessible. Consequently, 54 full-text articles were assessed for eligibility.

At the eligibility stage, 35 articles were excluded for the following reasons: AI was not used as the primary assessment tool (n = 12), no psychological or neurocognitive construct was assessed (n = 8), quantitative evaluation was not reported (n = 10), or only non-human or proxy data were used (n = 5). A total of 19 studies met all inclusion criteria and were included in the final qualitative synthesis. The complete study selection flow is presented in Figure 1.

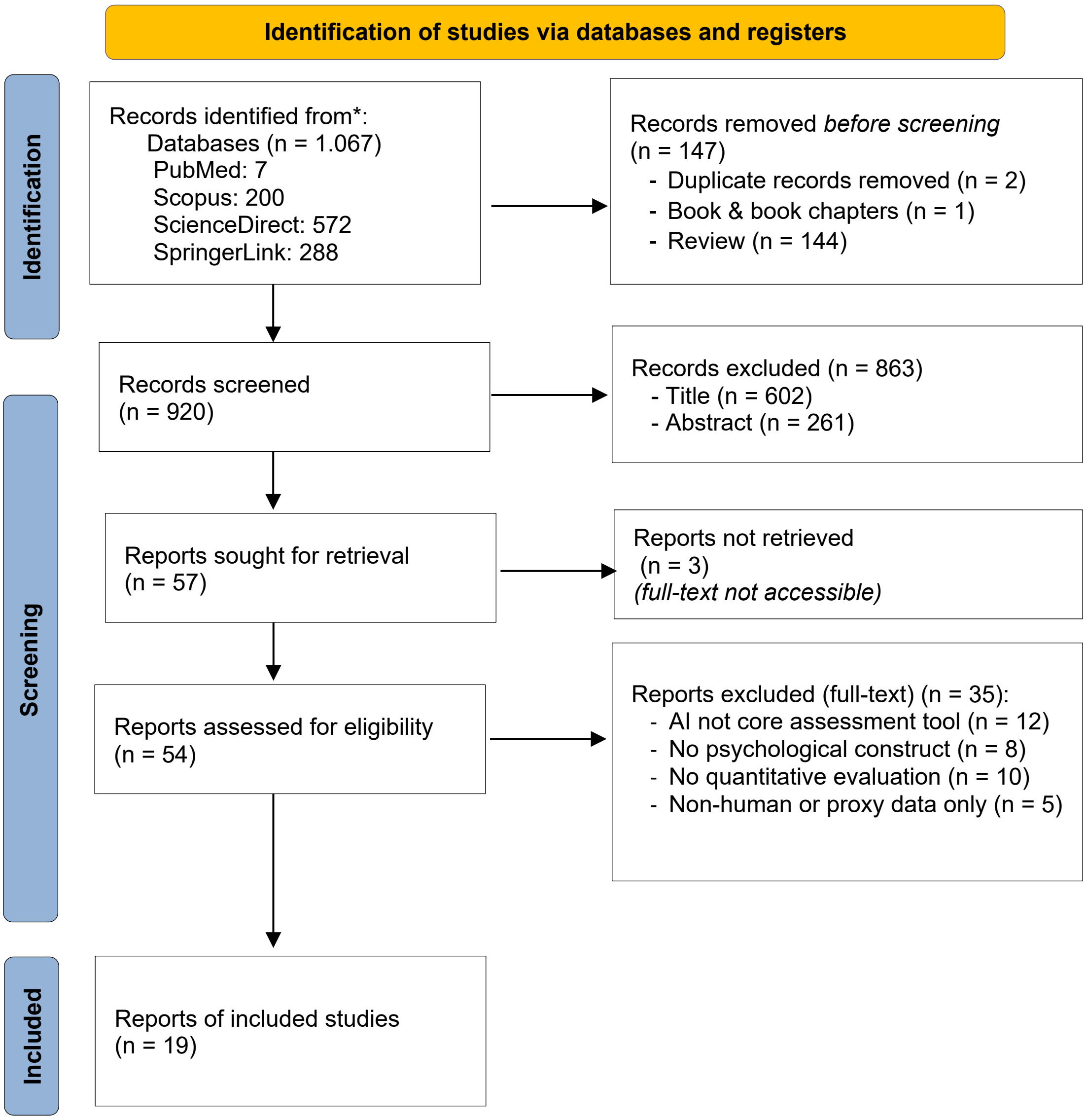

Figure 1. PRISMA 2020 Flow Diagram

2.4 Data Extraction

Data extraction was carried out systematically for all studies that met the inclusion criteria using a predefined extraction framework. The information collected focused on key characteristics relevant to the objectives of this systematic review, including the type of application, psychological constructs assessed, data modality, AI methods used, validation strategies, and sample size. The data extraction framework used in this review is summarised in Table 2.

Table 2. Data Extraction Framework for Included Studies

Category	Description
Application	Clinical or Cognitive application
Psychological Construct	Target psychological or neuropsychological construct
Modality	Data modality used (e.g., EEG, MRI/fMRI, speech, behavioral, multimodal)
AI Method	Machine learning or deep learning approach
Validation	Validation strategy (e.g., cross-validation, hold-out, external validation)
Sample Size	Number of human participants

2.5 Quality Assessment

Study quality was assessed using a structured, indicator-based appraisal approach to accommodate methodological heterogeneity across the included studies. Given the substantial variation in psychological constructs, data modalities (e.g., neuroimaging, EEG, behavioral sensing, multimodal biomarkers), AI architectures (machine learning and deep learning models), and validation strategies, conventional risk-of-bias frameworks developed for intervention-based research were considered methodologically unsuitable for this review.

Unlike intervention-focused studies that evaluate treatment effects within relatively standardised experimental designs, the studies included in this review represent diverse AI-based assessment models applied across heterogeneous clinical and cognitive contexts. The primary objective of this review is methodological mapping and cross-domain synthesis rather than quantitative aggregation of effect sizes. Consequently, applying traditional bias instruments would not adequately capture modelling robustness, validation rigor, and construct operationalisation in AI-based assessment systems.

Instead, quality indicators were applied to support critical interpretation of study findings. Four indicators were evaluated consistently across all studies: sample size adequacy relative to model complexity, validation strategy, use of independent data as a proxy for generalisation, and construct validity. Methodological risk was classified from low to high, with studies employing adequate sample sizes and external validation considered lower risk, and those applying complex models to limited datasets without external validation considered higher risk. Quality assessment outcomes informed the interpretation of methodological strengths and limitations in the discussion rather than serving as criteria for study exclusion.

3. RESULT

3.1 Study Characteristics

A total of 19 studies met the inclusion criteria and were analysed, covering AI-based clinical and cognitive assessment applications in biomedical psychology with variation in application domains, sample sizes, data modalities, and assessment task types. Publications were predominantly from 2023–2026, reflecting increased interest in AI as an assessment instrument. Sample sizes ranged from dozens to several thousand participants, and reported performance metrics varied from moderate to high depending on task complexity, data modality, and validation strategy. Most studies relied on internal validation, primarily cross-validation, whereas independent or external validation was reported in only a small subset, a pattern consistent across clinical and cognitive domains. A summary of the main characteristics of all studies is presented in Table 3.

Table 3. Characteristics of Included Studies

Year (Author)	Domain	Psychological Construct	Sample Size	Modality	Task
Silveira et al., 2020	Clinical	Rumination (brooding)	200	Clinical scales + blood biomarkers	Binary classification
Haining et al., 2021	Clinical	Cognitive deficits & functional outcome	146 (118 FU)	Neuropsychological tests	Regression / classification
Venugopalan et al., 2021	Clinical	Cognitive impairment (CN/MCI/AD)	220 (multimodal subset)	MRI + clinical + genetic	Multiclass classification
Asare et al., 2021	Clinical	Depression severity	629	Smartphone sensing	Binary classification
Wei et al., 2023	Clinical	Anxiety & depression	480	Clinical assessments	Binary classification
Bolla et al., 2023	Cognitive	Mild cognitive impairment	78 + 155 (external)	rs-fMRI	Binary classification
Kumar et al., 2024	Cognitive	Personality traits (Big Five)	98	EEG/ECG/GSR	Binary classification
Naseerullah et al., 2024	Clinical	Obsessive–compulsive disorder	193	Gene expression	Binary classification
J. Zhang et al., 2024	Cognitive	Global cognition, attention, memory	863	CDT images	Binary classification
Gómez-Pascual et al., 2024	Cognitive	Cognitive decline / AD progression	800	Plasma multi-omics	Classification / prognosis
Longo et al., 2024	Cognitive	PD-related MCI	275	Neuropsychological tests	Binary classification
Y. Zhang et al., 2024	Cognitive	Cognitive impairment	823	Community assessments	Binary classification
Díaz-Álvarez et al., 2025	Cognitive	Multidomain cognitive function	360 + 30 (test)	Neuropsychological tests + FDG-PET	Binary prediction
Wu et al., 2025	Cognitive	Cognitive decline (MCI)	56	EEG	Regression / classification
Walker et al., 2025	Cognitive	Irritability severity	6,065	Task-based fMRI	Regression
Gramkow et al., 2025	Clinical	Cognitive impairment & dementia etiology	170	Wearable actigraphy	Multiclass classification
Yeh et al., 2025	Clinical	IGD risk & social anxiety	60	EEG + VR	Binary classification
Wariri et al., 2026	Cognitive	Irritability (high vs low)	1,934	Task-based fMRI	Binary classification

3.2 Clinical Applications

The application of AI in the clinical domain in the included publications shows patterns that can be systematically classified based on the clinical assessment objectives. AI-based clinical applications form several main functional categories, ranging from diagnostic classification to risk and prognosis estimation, as summarised in Table 4 along with related psychological constructs and clinical outcomes. Performance evaluation in most publications was conducted using internal validation, while external or cross-cohort validation was reported in only a small number of publications.

Table 4. Taxonomy of Clinical Applications of AI in Biomedical Psychology

Clinical Application Category	Assessment Purpose	Psychological Construct	Typical Output
Diagnostic Classification	Distinguish clinical status	Depression, OCD, MCI, AD, PD-MCI	Disorder vs. control
Screening and Early Detection	Early identification of at-risk individuals	Cognitive impairment, dementia risk	At-risk vs. non–at-risk
Severity, Risk, and Outcome Estimation	Estimate severity, risk, or clinical outcomes	Depression severity, irritability, functional outcome	Continuous score or risk probability
Clinical Subtyping and Prognosis	Differentiate disease subtypes and progression	MCI conversion, AD progression, dementia etiology	Subtype or progression label

3.3 Cognitive Applications

In the cognitive domain, AI is primarily applied to cognitive function assessment tasks, both in measuring current cognitive status and in estimating future cognitive risk and changes. Unlike clinically oriented applications focused on diagnosis, AI-based cognitive approaches focus on measuring cognitive performance, mapping specific cognitive domains, and quantitatively modelling cognitive decline, covering core functions such as memory, attention, executive function, and social cognition in the context of population screening, cognitive ageing, and longitudinal monitoring. Performance evaluation in this domain is dominated by the use of internal validation, including in studies applying deep learning- s to high-dimensional data such as neuroimaging and EEG. The taxonomy of cognitive applications is summarised in Table 5.

Table 5. Taxonomy of Cognitive Applications of AI in Biomedical Psychology

Cognitive Application Category	Assessment Purpose	Cognitive Construct	Typical Output
Cognitive Status Classification	Classification of cognitive status between groups	Global cognition, MCI, cognitive impairment	Normal vs impaired
Domain-Specific Cognitive Assessment	Assessment of specific cognitive functions	Memory, attention, executive function	Domain score / class label
Cognitive Decline & Progression Modeling	Estimation of cognitive decline and change	Cognitive decline, aging-related impairment	Continuous score / progression label
Dimensional Trait Estimation	Measurement of continuous-dimensional cognitive-affective constructs	Irritability, personality traits	Continuous trait score

3.4 Data Modalities and AI Methods

A range of data modalities was employed across the included studies, reflecting diverse approaches to AI-based psychological assessment. The frequency of modality usage is summarised in Figure 2, while the mapping between application domains, data modalities, and AI method categories is presented in Table 6. As individual studies often incorporated more than one modality, the reported frequencies represent modality occurrences rather than mutually exclusive study counts. Together, these mappings illustrate how different assessment objectives are supported by distinct combinations of data sources and analytical approaches.

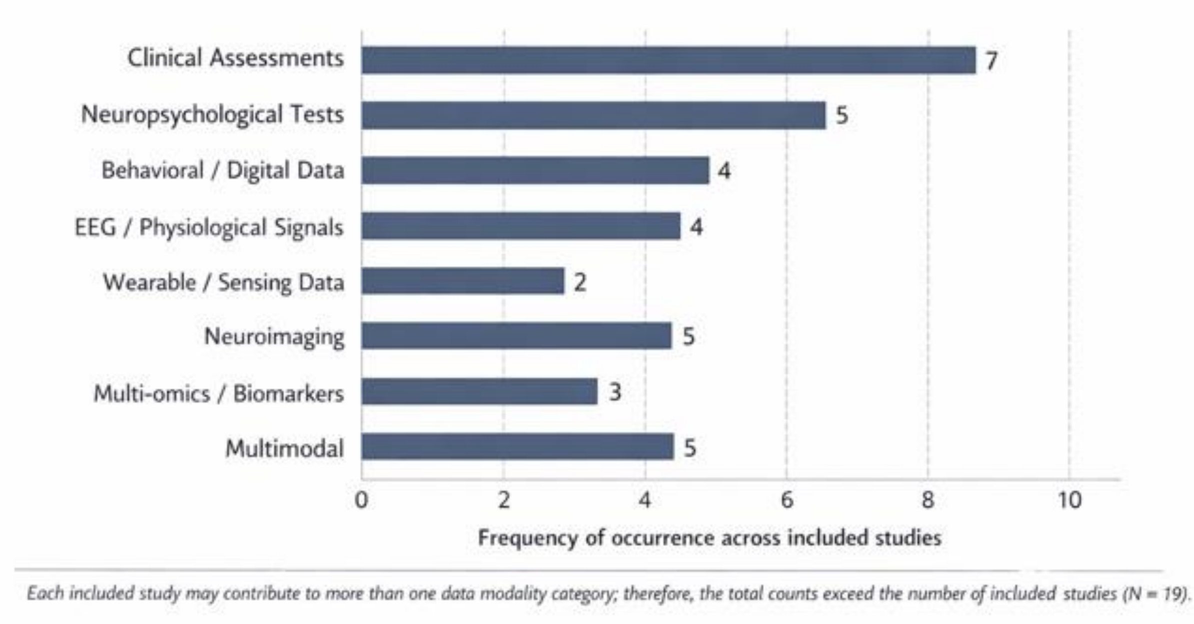

Figure 2. Distribution of Data Modalities Across Included Studies

Table 6. Application Modality Method Mapping

Application Domain	Application Type	Data Modality	AI Method Category	Typical Task
Clinical	Diagnostic classification	Clinical assessments	Machine learning	Binary / multiclass classification
	Screening & early detection	Smartphone sensing / wearable	Machine learning	Binary classification
	Severity / risk estimation	Multimodal (clinical + biomarkers)	Machine learning	Regression / risk estimation
Cognitive	Subtyping & prognosis	Neuroimaging / multi-omics	Deep learning	Multiclass classification
	Cognitive status classification	Neuroimaging / behavioral	Machine learning	Binary classification
	Domain-specific assessment	Neuropsychological tests	Machine learning	Domain-level prediction
	Cognitive decline modeling	EEG / longitudinal data	Deep learning	Regression / classification
	Dimensional trait estimation	Neuroimaging / physiological signals	Deep learning	Continuous score estimation

4. DISCUSSION

4.1 Principal Findings and Design Implications

The findings of this review indicate that artificial intelligence in biomedical psychology is predominantly employed as an assessment tool, with clear methodological differentiation between clinical and cognitive applications. Clinical applications are largely oriented toward categorical decision-making tasks, including diagnostic classification, risk screening, and subtype identification, targeting psychopathological and neurological conditions such as depression, anxiety, obsessive–

compulsive disorder, and neurodegenerative disorders (Silveira et al., 2020), (Venugopalan et al., 2021), (Asare et al., 2021), (Naseerullah et al., 2024). In contrast, cognitive applications more frequently focus on dimensional assessment objectives, including global cognitive status, multidomain cognitive performance, and longitudinal modelling of cognitive decline and disease progression (Bolla et al., 2023), (Díaz-Álvarez et al., 2025), (Gómez-Pascual et al., 2024).

Beyond categorical classification, several studies apply AI for continuous estimation of multidimensional constructs, such as cognitive function, irritability, and personality traits, reflecting a shift toward modelling individual variability rather than discrete clinical labels (Kumar et al., 2024), (Wu et al., 2025), (Walker et al., 2025). This pattern highlights the dual role of AI in biomedical psychology: supporting both threshold-based clinical decisions and fine-grained quantification of psychological and neurocognitive constructs across populations.

Across domains, a recurring methodological issue is the misalignment between assessment objectives, model complexity, and data characteristics. In clinical classification tasks based on limited or standardised datasets, increasing model complexity particularly through deep learning does not consistently improve performance and often amplifies overfitting risk. Conversely, in high-dimensional and continuous cognitive assessment contexts, complex models may be justified only when supported by adequate sample sizes and rigorous validation strategies. Collectively, these findings position the *fit-for-purpose* principle as a core design requirement for AI-based psychological assessment systems that are clinically valid, interpretable, and methodologically robust.

4.2 Clinical and Cognitive Implications

The main implication of these findings is demonstrated by the positioning of AI as a scalable assessment tool, particularly in contexts characterised by resource constraints. The application of AI enables large-scale screening and assessment by utilising relatively easily obtained data, such as smartphone sensing, community-based assessment and wearable-based monitoring (Asare et al., 2021), (Y. Zhang et al., 2024), (Gramkow et al., 2025). This approach expands the potential of psychological assessment to a broader population and enables more continuous monitoring compared to traditional methods.

The use of AI is positioned as a decision support system, not as a replacement for clinical professionals or neuropsychologists. AI models are used to improve the sensitivity and consistency of assessments, while the final interpretation remains within the framework of current clinical practice (Haining et al., 2021), (Bolla et al., 2023), (Longo et al., 2024), (Yeh et al., 2025). The integration of AI into biomedical psychology practice is more focused on augmenting the assessment process rather than fully automating decision-making.

4.3 Methodological Challenges

Despite their substantial potential, the included studies exhibit recurring methodological limitations, primarily related to sample size, model complexity, and validation strategies. These issues are most pronounced in neuroimaging, EEG, and biomarker-based studies, where deep learning models are often applied to limited datasets, increasing the risk of overfitting and performance inflation derived from internal validation (Bolla et al., 2023), (Kumar et al., 2024), (Wu et al., 2025). Across studies, reliance on internal cross-validation remains dominant, while external validation or independent datasets are reported infrequently (Díaz-Álvarez et al., 2025) , (Gómez-Pascual et al., 2024) (Gramkow et al., 2025), with externally validated models typically showing more moderate but stable performance.

Additional challenges concern limited generalisability and construct validity. Many models are developed within specific cohorts, constraining external applicability (Wei et al., 2023), (Naseerullah et al., 2024), (Yeh et al., 2025). The use of clinical labels or instrument scores as proxy representations of complex psychological constructs risks producing assessments that are statistically robust but conceptually weak, increasing vulnerability to bias and fairness issues across populations. Accordingly, construct validity, feature transparency, and bias considerations should be treated as core methodological requirements in AI-based psychological assessment.

4.4 Future Directions

Research findings indicate the need for the development of more standardised benchmark assessments for both clinical and cognitive domains. Standardising the definitions of psychological constructs, assessment protocols, and evaluation metrics is expected to improve comparability between studies and facilitate cross-domain synthesis (Díaz-Álvarez et al., 2025), (Haining et al., 2021), (J. Zhang et al., 2024)Click or tap here to enter text.. These efforts are crucial to reduce the methodological fragmentation currently evident in the literature.

Multimodal approaches are increasingly being applied by integrating clinical, biological, and behavioural data to enrich the representation of psychological conditions (Gómez-Pascual et al., 2024), (Venugopalan et al., 2021), (Wariri et al., 2026). Multimodal integration is balanced with more consistent external validation practices and attention to model interpretability so that AI-based assessment results can be applied responsibly in clinical and cognitive contexts (Walker et al., 2025), (Yeh et al., 2025).

4.5 Limitations

The relatively limited number of included studies reflects the deliberate narrowing of this review to AI-based assessment, rather than the full spectrum of AI applications in psychology. While this focus enhances conceptual and methodological coherence, the included studies span diverse clinical and cognitive domains, psychological constructs, data modalities, and assessment tasks. Such heterogeneity precludes quantitative synthesis or meta-analysis and necessitates a qualitative interpretative approach. Variations in construct definitions and assessment protocols further limit cross-study comparability.

Methodological quality across primary studies varied, particularly regarding sample size, validation strategies, and reporting transparency. Although inclusion criteria ensured the use of human data and quantitative evaluation, limited external validation and potential publication bias may constrain the generalisability of findings. Accordingly, this review should be interpreted as a conceptual and methodological synthesis of AI-based psychological assessment practices, rather than a comparative evaluation of model performance.

CONCLUSION

This systematic review synthesises the application of artificial intelligence in biomedical psychology with a specific focus on clinical and cognitive assessment. Across application domains, AI is consistently employed to support diagnostic classification, risk screening, and the estimation of psychological and neurocognitive outcomes, spanning psychiatric, neurodegenerative, and cognitive assessment contexts. A clear distinction emerges between clinical applications, which are predominantly centred on categorical decision-making, and cognitive applications, which more frequently emphasise dimensional measurement and longitudinal assessment. This distinction underscores the importance of aligning analytical strategies, model outputs, and evaluation designs with the specific objectives of psychological assessment.

From a methodological and conceptual perspective, the reviewed literature consistently positions AI as a decision-support tool rather than a substitute for professional judgement. While advanced machine learning and deep learning models offer increased representational capacity, their application must be justified by the assessment goal and supported by adequate sample sizes and appropriate validation strategies. Inadequate alignment between model complexity, data characteristics, and validation design increases the risk of clinically unstable or inflated performance, particularly in high-dimensional and small-sample settings. These findings reinforce the need for cautious and principled model selection in AI-based psychological assessment.

Accordingly, AI-based assessment systems in biomedical psychology should be developed according to a fit-for-purpose design principle, in which method selection, data modality, and model complexity are explicitly tailored to the target construct and application context. The establishment of standardised assessment benchmarks, consistent cross-population validation practices, and interpretable multimodal integration remains essential to ensure that AI-enabled psychological assessment is robust, responsible, and clinically relevant. Together, these considerations provide a structured foundation for the continued and responsible integration of AI into psychological and neurocognitive assessment practice.

REFERENCES

Alkan, B. B., Kumartas, M., Kuzucuk, S., & Alkan, N. (2025). Fast and effective assessment for individuals with special needs form optimization and prediction models. *BMC Psychology*, *13*(1), 415. https://doi.org/10.1186/s40359-025-02768-z

Asare, K. O., Terhorst, Y., Vega, J., Peltonen, E., Lagerspetz, E., & Ferreira, D. (2021). Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: Exploratory study. *JMIR MHealth and UHealth*, *9*(7). https://doi.org/10.2196/26540

Bae, S., Hong, J., Ha, S., Moon, J., Yu, J., Choi, H., Lee, J., Do, R., Sim, H., Kim, Hanna, Lim, H., Park, M.-H., Ko, E., Yang, C.-M., Lee, D., Yoo, H., Lee, Y., Bong, G., Kim, J. I., … Cheon, K.-A. (2025). Multimodal AI for risk stratification in autism spectrum disorder: integrating voice and screening tools. *Npj Digital Medicine*, *8*(1), 538. https://doi.org/10.1038/s41746-025-01914-6

Baydili, İ., Tasci, B., & Tasci, G. (2025). Artificial Intelligence in Psychiatry: A Review of Biological and Behavioral Data Analyses. *Diagnostics*, *15*(4), 434. https://doi.org/10.3390/diagnostics15040434

Bolla, G., Berente, D. B., Andrássy, A., Zsuffa, J. A., Hidasi, Z., Csibri, E., Csukly, G., Kamondi, A., Kiss, M., & Horvath, A. A. (2023). Comparison of the diagnostic accuracy of resting-state fMRI driven machine learning algorithms in the detection of mild cognitive impairment. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-49461-y

Collinson, S. L., Lam, M., & Hayes, C. J. (2010). The utility and benefits of clinical neuropsychology in Asia. *Asian Journal of Psychiatry*, *3*(2), 50–54. https://doi.org/10.1016/j.ajp.2010.03.007

Delcea, C., Rusu, D., Matei, H., Vica, L., & Siserman, C. (2020). The Evidence-Based Practice Paradigm Applied To Judicial Psychological Assessment In The Context Of Forensic Medicine. *Romanian Journal of Legal Medicine*, *28*(3), 257–262. https://doi.org/10.4323/rjlm.2020.257

Díaz-Álvarez, J., García-Gutiérrez, F., Bueso-Inchausti, P., Cabrera-Martín, M. N., Delgado-Alonso, C., Delgado-Alvarez, A., Diez-Cirarda, M., Valls-Carbo, A., Fernández-Romero, L., Valles-Salgado, M., Dauden-Oñate, P., Matías-Guiu, J., Peña-Casanova, J., Ayala, J. L., & Matias-Guiu, J. A. (2025). Data-driven prediction of regional brain metabolism using neuropsychological assessment in Alzheimer's disease and behavioral variant Frontotemporal dementia. *Cortex*, *183*, 309–325. https://doi.org/10.1016/j.cortex.2024.11.022

Dubey, Parul, Dubey, Pushkar, & Bokoro, P. N. (2025). Federated learning for privacy-enhanced mental health prediction with multimodal data integration. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *13*(1).https://doi.org/10.1080/21681163.2025.2509672

Engel, S., Klusmann, H., Laufer, S., Kapp, C., Schumacher, S., & Knaevelsrud, C. (2022). Biological markers in clinical psychological research - A systematic framework applied to HPA axis regulation in PTSD. *Comprehensive Psychoneuroendocrinology*, *11*, 100148. https://doi.org/10.1016/j.cpnec.2022.100148

Finley, J.-C. A. (2024). Performance validity testing: the need for digital technology and where to go from here. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1452462

Gómez-Pascual, A., Naccache, T., Xu, J., Hooshmand, K., Wretlind, A., Gabrielli, M., Lombardo, M. T., Shi, L., Buckley, N. J., Tijms, B. M., Vos, S. J. B., ten Kate, M., Engelborghs, S., Sleegers, K., Frisoni, G. B., Wallin, A., Lleó, A., Popp, J., Martinez-Lage, P., … Legido-Quigley, C. (2024). Paired plasma lipidomics and proteomics analysis in the conversion from mild cognitive impairment to Alzheimer's disease. *Computers in Biology and Medicine*, *176*. https://doi.org/10.1016/j.compbiomed.2024.108588

Gramkow, M. H., Brink-Kjær, A., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Jennum, P., Hasselbalch, S. G., & Frederiksen, K. S. (2025). Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study. *Alzheimer's Research and Therapy* , *17*(1). https://doi.org/10.1186/s13195-025-01751-5

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producingPRISMA 2020-compliant flow diagrams, with interactivityfor optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, *18*(2). https://doi.org/10.1002/cl2.1230

Haining, K., Brunner, G., Gajwani, R., Gross, J., Gumley, A. I., Lawrie, S. M., Schwannauer, M., Schultze-Lutter, F., & Uhlhaas, P. J. (2021). The relationship between cognitive deficits and impaired short-term functional outcome in clinical high-risk for psychosis participants: A machine learning and modelling approach. *Schizophrenia Research*, *231*, 24–31. https://doi.org/10.1016/j.schres.2021.02.019

Kumar, D., Singh, P., & Raman, B. (2024). All signals point to personality: A dual-pipeline LSTM-attention and symbolic dynamics framework for predicting personality traits from Bio-Electrical signals. *Biomedical Signal Processing and Control*, *96*. https://doi.org/10.1016/j.bspc.2024.106609

Longo, C., Romano, D. L., Pennacchio, M., Malaguti, M. C., Di Giacopo, R., Giometto, B., & Papagno, C. (2024). Are the criteria for PD-MCI diagnosis comprehensive? A Machine Learning study with modified criteria. *Parkinsonism and Related Disorders*, *124*. https://doi.org/10.1016/j.parkreldis.2024.106987

Naseerullah, Hayat, M., Iqbal, N., Tahir, M., AlQahtani, S. A., & Alamri, A. M. (2024). Early diagnosis of obsessives-compulsive disorder through gene expression analysis using machine learning models. *Chemometrics and Intelligent Laboratory Systems*, *248*. https://doi.org/10.1016/j.chemolab.2024.105107

Parra, F., Scherer, S., Benezeth, Y., Tsvetanova, P., & Tereno, S. (2022). Development and Cross-Cultural Evaluation of a Scoring Algorithm for the Biometric Attachment Test: Overcoming the Challenges of Multimodal Fusion with “Small Data.” *IEEE Transactions on Affective Computing*, *13*(1), 211–225. https://doi.org/10.1109/TAFFC.2019.2921311

Patterson, J. D., Pronchick, J., Panchanadikar, R., Fuge, M., van Hell, J. G., Miller, S. R., Johnson, D. R., & Beaty, R. E. (2025). CAP: The creativity assessment platform for online testing and automated scoring. *Behavior Research Methods*, *57*(9), 264. https://doi.org/10.3758/s13428-025-02761-9

Rakotomanana, H., & Rouhafzay, G. (2025). A Scoping Review of AI-Based Approaches for Detecting Autism Traits Using Voice and Behavioral Data. *Bioengineering*, *12*(11), 1136. https://doi.org/10.3390/bioengineering12111136

Ren, W., Xue, X., Liu, L., & Huang, J. (2025). AI Applications in Depression Detection and Diagnosis: Bibliometric and Visual Analysis of Trends and Future Directions. *JMIR Mental Health*, *12*, e79293. https://doi.org/10.2196/79293

Schwartz, B., Uhl, J., & Atzil-Slonim, D. (2023). Editorial: Assessments and measures in psychotherapy research: going beyond self-report data. *Frontiers in Psychiatry*, *14*. https://doi.org/10.3389/fpsyt.2023.1276222

Sikström, S., Valavičiūtė, I., Kuusela, I., & Evors, N. (2024). Question-based computational language approach outperforms rating scales in quantifying emotional states. *Communications Psychology*, *2*(1), 45. https://doi.org/10.1038/s44271-024-00097-2

Silveira, É. de M., Passos, I. C., Scott, J., Bristot, G., Scotton, E., Teixeira Mendes, L. S., Umpierre Knackfuss, A. C., Gerchmann, L., Fijtman, A., Trasel, A. R., Salum, G. A., & Kauer-Sant’Anna, M. (2020). Decoding rumination: A machine learning approach to a transdiagnostic sample of outpatients with anxiety, mood and psychotic disorders. *Journal of Psychiatric Research*, *121*, 207–213. https://doi.org/10.1016/j.jpsychires.2019.12.005

Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer’s disease stage. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-74399-w

Walker, J. C., Swineford, C., Patel, K. R., Dougherty, L. R., & Wiggins, J. L. (2025). Deep learning identification of reward-related neural substrates of preadolescent irritability: A novel 3D CNN application for fMRI. *NeuroImage: Reports*, *5*(2). https://doi.org/10.1016/j.ynirp.2025.100259

Wariri, F. M., Walker, J. C., Wiggins, J. L., & Bi, J. (2026). SHAP-based explainable machine learning analysis of reward-related neural connectivity to predict preadolescent irritability. *NeuroImage: Reports*, *6*(1). https://doi.org/10.1016/j.ynirp.2026.100322

Wei, Z., Wang, X., Ren, L., Liu, Chang, Liu, Chao, Cao, M., Feng, Y., Gan, Y., Li, G., Liu, X., Liu, Y., Yang, L., & Deng, Y. (2023). Using machine learning approach to predict depression and anxiety among patients with epilepsy in China: A cross-sectional study. *Journal of Affective Disorders*, *336*, 1–8. https://doi.org/10.1016/j.jad.2023.05.043

Wu, X., Huang, Q., Liu, Y., Dong, X., Liu, H., & Wen, D. (2025). Multimodal Deep Learning–Based EEG Health Scoring Model for Mild Cognitive Impairment. *IFAC-PapersOnLine*, *59*(35), 697–702. https://doi.org/10.1016/j.ifacol.2025.12.560

Yang, K., Lau, R. Y. K., & Abbasi, A. (2023). Getting Personal: A Deep Learning Artifact for Text-Based Measurement of Personality. *Information Systems Research*, *34*(1), 194–222. https://doi.org/10.1287/isre.2022.1111

Yeh, P. Y., Lin, C. L., Sun, C. K., & Tung, S. Y. (2025). Application of EEG-Based Machine Learning in Time–Frequency and Brain Connectivity Analysis Among Individuals at High Risk of Internet Gaming Disorder with Social Anxiety. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-025-01560-9

Zhang, J., Bandyopadhyay, S., Kimmet, F., Wittmayer, J., Khezeli, K., Libon, D. J., Price, C. C., & Rashidi, P. (2024). Developing a fair and interpretable representation of the clock drawing test for mitigating low education and racial bias. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-68481-w

Zhang, Y., Xu, J., Zhang, C., Zhang, X., Yuan, X., Ni, W., Zhang, H., Zheng, Y., & Zhao, Z. (2024). Community screening for dementia among older adults in China: a machine learning-based strategy. *BMC Public Health*, *24*(1). https://doi.org/10.1186/s12889-024-18692-7

Zheng, Y., Liu, H., Zhang, B., & Ding, Y. (2025). Research and Implementation of a Multi-Modal AI-Driven Psychological Assessment and Analysis System Based on Unity. *Proceedings of the 2025 2nd International Conference on Image Processing, Intelligent Control and Computer Engineering*, 580–587. https://doi.org/10.1145/3768184.3768283