Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol.14 , Issue 32s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Mental Health Assessment and Computational Frameworks

Shagufta Farzana[1], Amrita Verma[1]*, S. R. Tandan[2], Rohit Kumar Miri[3]

[1]Department of Computer Science and Engineering, Dr. C.V. Raman University, Kota, Bilaspur, Chhattisgarh 495113, India.

[2]Department of Computer Science, Government Rajmata Vijaya Raje Sindhiya Kanya Mahavidyalaya, Kawardha, Chhattisgarh 491995, India.

[3]Department of Biomedical Engineering and Bioinformatics, Chhattisgarh Swami Vivekanand Technical University, Bhilai, Chhattisgarh, 491107, India

***Corresponding Author,**

Dr. Amrita Verma

Department of Computer Science and Engineering, Dr. C.V. Raman University, Kota, Bilaspur, Chhattisgarh 495113, India.

Email ID: amrita.85024@gmail.com

Cite this paper as: Shagufta Farzana, Amrita Verma, S. R. Tandan, Rohit Kumar Miri (2025) Mental Health Assessment and Computational Frameworks. Journal of Neonatal Surgery, 14, (32s) 10374-10384

ABSTRACT

Common mental health issues such as depression and anxiety are leading casuse of disability globally. An early and accurate assessment enables people to receive the right care sooner and reduces long-term harm. This paper presents an overview of traditional questionnaires and interviews and explains how new computer- based methods allow for faster, more accurate, and more scalable assessments. Clinician interviews (such as the SCID) are still the reference standard but require time and trained personnel. Self-report instruments (PHQ-9, GAD-7, HADS, K10) are quick and well tested and therefore useful for screening and monitoring. Transferring these tools to phones and the web enables remote use, autoscoring and tracking over time but has demanding privacy and security appended to them. Computer adaptive testing (CAT) can reduce the length of the questionnaire by selecting the most informative questions for each respondent, thus achieving high accuracy at low respondent burden. New approaches use language and data science. Natural language processing (NLP) can be used to analyse short free-text answers and capture numerous details that fixed response options might miss. Computational models which are machine learning based (like support vector machines, random forests, neural networks) can take into account questionnaire score, text and electronic health record data to identify risk, recommend triage and monitor treatment. These tools should be transparent and tested in new settings for fairness because bad data can damage under-represented groups. We must also tackle hindrances to accessing the digital world, like the internet and skills. Clinical judgment shouldn't be replaced by computational methods but supported. A helpful strategy is to choose relevant tools, digitize with a privacy- by-design focus, combine computer-aided translation (CAT) with free-text modules, develop models with clear reporting and strong validation and fairness checks, and embed outputs into routine clinical workflows. Future work should adapt and personalize screening, join text and speech and biosignals, push cross-cultural testing further, and adopt shared datasets and reporting standards..

Keywords: *Mental health assessment; Digital questionnaires; Machine learning; Natural language processing*

INTRODUCTION

A significant and increasing global public health challenge pertains to mental health disorders. About 850 million people worldwide suffer from depression, anxiety or any other common mental disorder (CMD). In fact, CMDs account for more than 5% of all disability-adjusted life years (DALYs). In turn, this causes trillions in monetary costs. According to the World Health Organization, depression and anxiety are the leading causes of disability across the world, with 46.9 million DALYs due to depression in 2019 [1]. The harms caused by mental illnesses can be minimised through early diagnosis and correct assessment. Recognizing persons who run the risk, or are in the early stages of mental disorders, enables timely intervention that can improve prognostic outcome and prevent the chronicity and severity of symptoms [2]. Earlier evaluations enable the distribution of limited healthcare resources to guide treatment and help in avoiding crises leading to more severe states of mental illness.

When there is a delay or inadequate diagnosis, it can lead to prolonged pain and suffering, worsening the clinical outcome and increase in healthcare utilization. For instance, failing to treat depression can worsen the other chronic illnesses, harm physical health, and raise the mortality risk from suicide. On the other hand, appropriate early screening at the level of primary care and the community can reduce these risks through referral to services. Also, a population- wide screening is good for public health surveillance and increases epidemiologic understanding and policy. Various epidemiological studies done on populations like military personnel highlight these issues. Studies of military cohorts demonstrate that service members have high trajectories of PTSD and distress. There is a resilient rate of 84% and a rate of 9.6% worsening distress across four years. Prior trauma and sleep problems predicts their worsened course [3], [4]. In addition, analyses of primary care datasets suggest that common mental disorders can be identified from routine clinical data with fair accuracy, given the right algorithms [5]. Algorithms using PHQ-9 scores along with diagnostic codes from a multi-million primary-care patient database have over 96% specificity and 0.29–0.32 sensitivity with and without symptom codes, respectively [5]. Collectively, these findings highlight the need for timely and accurate assessment of mental health in order to improve societal and individual health outcomes.

A. Overview of Assessment Modalities

For a long time now, mental health assessment, in general, has happened through structured clinical interviews and standardized, mostly self-report, questionnaires. The Structured Clinical Interview for DSM (SCID), which is clinician-administered, is the gold standard for diagnosis. They use the psychiatric interview tools specified in the DSM [6]. According to Shabani et al. (2021) [7], the clinician is given the chance to investigate the symptomatology very thoroughly, clarify ambiguities and examine contextual factors through interviews in such diagnostic scheme. Nonetheless, the conduct of such interviews requires special training, significant administration time, and is impractical in scale or resource constraints.In contrast, standardized tools like the PHQ-9 facilitates rapid screening and quantifies severity of depression and anxiety. Most self-report tools make use of symptom criteria adopted in the DSM-5 and International Classification of Diseases (ICD- 11). Also, the items are meant for capturing the symptoms for depression, anxiety, etc. Response Forms provide measurable information that can be delivered and scored quickly. They can be used in monitoring the course of therapy, in epidemiology, and in primary care. It is frequently checked against structured clinical interviews to make certain about clinical relevance and psychometric reliability.The way the assessment is given differs according to clinician- versus self-report. These two alternatives have different compromises with respect to diagnosis" depth and efficiency. For example, structured interview often generate richer clinical data. This data is useful for diagnostic certainty and comorbidity detection. In contrast, self-report scales focus on scalability and accessibility. Validation studies con- firm the efficacy and limitations of both methods. Validation of the PHQ-9 in primary care indicates the measure is highly sensitive and specific for major depressive disorder compared to structured interview standard [8]. Meanwhile, the PHQ- 9, GAD-7, and Work and Social Adjustment Scale (WSAS) have been translated and adapted for specific populations, such as Deaf British Sign Language users [9]. This unique strategy involves employing both after mentioned intervention strategies which are interview or structured interview and questionnaire.

B. Need for Computational Frameworks

Although rating scalers and interviews are useful, they fail to capture some of the less obvious, nuanced features of mental health gathered from patients with a variety of presentations. Fixed-item responses, in particular, hide the complexities of emotional and cognitive states, are subject to central tendency bias, and do not provide a mechanism to evolve with the individual patient. As a result of this development, new computational frameworks could help in screening for mental health using machine learning (ML) and natural language processing (NLP).Digital tools using artificial intelligence (AI) can analyse both structured data from questionnaires as well as unstructured information, such clinical narratives or other open text. For example, subtle forms of scoring such as machine learning symptom patterns or risk profiles can be detected. Recent advances in natural language processing (NLP) offer the possibility of assessing natural language responses, therefore overcoming the limitations posed by forced-choice scales by enabling the capture of richer, person-centred data. Through this approach, mental health evaluation can be more personalized and nuanced for users' diagnostics and engagement.Moreover, utilizing computational methods allows for scalable and real-time analyses integrated in electronic health records (EHR) will facilitate clinical decision support. Through the use of smart algorithms that analyze clinical notes and patient- reported outcomes, rapid screening, triaging and monitoring can occur within routine care workflows [10]. Machine-learning models have shown efficiency in distinguishing diagnosis and predicting treatment response although interpreter ability and translation into clinical practice are problematic for clinical researchers.These frameworks can potentially enhance the assessment of mental health by delivering tailored and understandable insights that function alongside diagnostic tools. Nonetheless, strong validation, ethical use, and clinician oversight are vital to achieving these benefits entirely [11].

1. QUESTIONNAIRE-BASED MENTAL-HEALTH EVALUATION

Mental-health questionnaires are the most surveyed tools for screening, case finding and tracking symptoms over time. They use lived experiences and convert them into a structured form of information that employed by doctors and expert researchers, allowing for comparing information between settings and within settings as well. Not one instrument fits all needs because use-cases can vary e.g. rapid screening in general practice to precise diagnosis in specialist clinics. A tool's choice should balance psychometric quality (reliability, validity, measurement invariance), burden on respondents, cultural/linguistic appropriateness, mode of delivery (paper, web, mobile) and feasible constraints (licensing, training).This section summarizes major types of questionnaires and what they are commonly used for. It shows how instrument design and administration impact the diagnostic accuracy, equity, and feasibility of questionnaires.

A. Classification of Questionnaires

To structure the landscape, we categorize the instruments along three axes: (i) who gives the assessment (clinician vs self-report);

(ii) what they assess (disorder-specific vs transdiagnostic constructs); and (iii) how they are delivered (format and mode of administration). Every axis involves compromises related to depth, speed, cost, and broad applicability.

1) Clinician-Administered Interviews vs. Self-Report Scales: Mental health questionnaire is mainly of two types, clinician- administered diagnostic interviews and self-report rating scales. Clinician-administered interviews such as the SCID are elaborate semi-structured interviews created to systematically assess clinical disorders based on DSM diagnostic criteria. As clinical judgment is allowed and one can probe with open-ended questions, it captures the comorbidities and contextual features that symptom checklists miss. Nevertheless, these interviews take time and resources and need trained individuals, limiting their large-scale or routine use.On the other hand, self-report scales provide a fast and standardized way of data collection. Instruments such as the PHQ-9 or GAD-7 provide fixed-response options that relate to the frequency or severity of symptoms. These can be scored easily and are useful for screening or monitoring in populations. Extensive validation studies have demonstrated that these instruments have high sensitivity and specificity to detect common mental disorders in primary care and community settings. As Michopoulos pointed out, the two-factor structure of the HADS (Hospital Anxiety and Depression Scale), as well as its reliability (Cronbach's $\alpha \geq 0.80$), was found in Greek inpatients; Arabic version shows reliability ($\alpha \geq 0.85$) and responsiveness and validity in surgical patients [12].These contrasting modalities suit different contexts. Structured interviews work best where diagnostic precision is needed and in particular settings for treatment planning or research requiring a gold-standard diagnosis. Self-report questionnaires are more practical for use in screening, epidemiology or treatment monitoring for symptom changes. These formats, according to validation studies, are complementary. For instance, while the SCID remains the dominant diagnostic instrument in clinical trials, the PHQ-9 has reliable endorsements for screening for depression in primary care due to its brevity and diagnostic utility, the composite International Diagnostic Interview (CIDI) in 2,642 primary-care patients: PHQ-9 replaced with ≥ 10 yields 74% sensitivity and 91% specificity; PHQ-2 ≥ 2 shows 86% sensitivity and 78% specificity [8]. Moreover, in community samples, 89% sensitivity and 82% specificity for generalized anxiety disorder is achieved when Cronbach's $\alpha > 0.90$ and threshold $\geq$ 10 [8]. Algorithm development using EHR questionnaires and clinical profile data can further extend and deepen identification of mental disorders [5].

2) Disorder-Specific vs. Transdiagnostic Instruments: Some mental health questionnaires are disorder-specific while others are transdiagnostic in nature. Measures that focus only on the order assess only the symptoms that are characteristics of the order. The PHQ-9 has nine items which ask about the severity of a person's depressive symptoms. The GAD-7 scale examines symptoms that could be generalized anxiety disorder.In comparison, transdiagnostic measures are designed to capture higher-order constructs like general distress, psychological well- being, or multiple symptom domains. The PHQ-ADS is designed as a screening tool for comorbid depression and anxiety and consists of two sets of symptoms (depression and anxiety) of comparable size [13]. The Four-Dimensional Symptom Questionnaire (4DSQ) is an instrument that distinguishes between distress, depression, anxiety and somatization. It is a multidimensional instrument to measure psychopathology [14].The existence of more than 280 questionnaires measuring depression highlights the undoubted complexity involved in choosing the right measure for use in either clinical or research settings. Both disorder specificity and transdiagnostic measures are capable of providing precise and holistic information, respectively. Nonetheless, they each hold potential downsides mainly arising from lack of sensitivity or lack of specificity. To optimize assessment, it is essential to know the psychometric properties and purpose of each instrument [15]. Initial validation in three trials ($n \approx 4500$) is showing evidence for internal consistency ($\alpha = 0.86$–0.92) and correlation with clinician assessments ($r = 0.68$–0.75) according to Gibbson [16]. According to Terluin's study conducted among 1147 working Dutch populations, the confirmatory four-factor CFI = 0.98, RMSEA = 0.04 [14].

3) Format and Administration: Mental health instruments can differ in coverage and the ways they are put together and administered affecting their feasibility and data quality. The traditional paper-and-pencil formats have always been used that are commonly computerized these days. They are administered through web portals or mobile applications as they make it easier to scale up and ensure automated scoring. The format of the questions also differs. Many questionnaires make use of graded-response or Likert-type scales (for example, frequency 'not at all' to 'nearly every day') in which the participant chooses one from a number of options. Thus, in such cases, it becomes possible to quantify the severity of a person's symptoms. Other people include open-ended items or free-text responses, which collect more rich qualitative information but add complexity to scoring and interpretation.

The type of respondent also governs administration, given that patient self-report is the most common. However, parent- or clinician- rated scales exist to offer alternatives, and are useful in paediatric populations or where patient insight is limited. A number of British Sign Language versions of the PHQ-9 and GAD-7 are currently being piloted among Deaf people in the UK [9]. Patient-Reported Outcome (PRO) requires assessing the respondent burden, as excessive burden causes missing data of poor quality [17]. Terkawi Digital PHQ-9/GAD-7 maintain psychometrics in web/mobile for- mats (ICC¿0.85; equivalence across modes) [18].

B. Common Disorder-Specific Scales

Questionnaires that focus on symptoms of a specific disorder are called disorder-specific questionnaires. These questionnaires may focus on a singular syndrome only like depression or anxiety. It is utilized for screening and case identification. Further, the questionnaires also assist in tracking symptom severity over time. Here is a detailed overview of the major psychological assessment tools that are currently in use. Each of these tools has been either tested or recommended as useful and helpful for the assessment of a particular psychological health condition.

1) *Depression Questionnaires:* The PHQ9 is a scale that someone answers themself about their depression symptoms. The MDD checklist consists of nine items that correspond to the major depressive disorder criteria. Participants give a rate on how much they experienced the symptoms in the last 2 weeks from 0 (not at all) to 3 (nearly every day). This one is measured on a total score. The cut-points of the scale are 5, 10, 15, and 20 that are indexed to mild, moderate; moderately severe and severe depression respectively. Studies have shown this tool to have a score of 10 or above with the approximate sensitivity and specificity of 88% for major diagnosis of depression [8]. The same tool is accurate for screening for use in primary care. The BDI is a self-report scale of 21 items that assesses cognitive, affective, and somatic symptoms of depression. DASS-21 is a standardized measurement scale used to measure illness and physiological burden. It depicts the presence and severity each item has. HADS-D , the depression subscale of the Hospital Anxiety and Depression Scale, a scale designed for use in hospital populations. These measures have been validated in many populations, which is why they are used [19]. Representatives from Deaf Access reportedly claim that British Sign Language translations for PHQ-9 and GAD-7 show measurement invariance and criterion validity against SCID (sensitivity = 0.88, specificity = 0.91) [20]. Various instruments have different lengths, coverage, and psychometric properties. This range allows researchers and clinicians to choose a tool in keeping with the aim and setting.

2) *Anxiety Questionnaires:* GAD-7 is a brief 7 item self- administered questionnaire designed to screen for generalised anxiety disorder and the severity of anxiety symptoms. The tool is made up of seven items corresponding to the anxiety criteria from the DSM- IV. It has excellent internal consistency as evidenced by a Cronbachs $\alpha > 0.9$. Additionally, the tool demonstrates measurement invariance across ages and sexes. Evidence from population-based studies, such as among Peruvian adults [9]. According to three studies of diagnostic accuracy, GAD-7 identifies cases of generalized anxiety disorder. It is suitable for use in primary care and epidemiological studies. Other anxiety assessment tools include the Beck Anxiety Inventory (BAI) which focuses on the somatic symptoms of anxiety; the State-Trait Anxiety Inventory (STAI) which studies both state and trait models of anxiety; and the HADS Anxiety subscale (HADS-A) which is mainly designed for use in hospital settings. Each measure has different emphases, administration times, and psychometrics. The instrument is usually chosen depending on the target population, clinical pertinence and administration limitations [15].

3) *Dual-Domain Scales:* The HADS measures a hospital patient's anxiety and depression. The Hospital Anxiety and Depression Scale is designed to measure anxiety and depression symptomatology at the same time in hospital and primary care populations. Created to reduce bias from bodily symptoms that accompany physical illness, it has 14 items with equal numbers in anxiety and depression subscales. Several confirmatory factor analyses across cultures have shown its two-factor structure, internal consistency, and construct validity. For instance, studies conducted on Greek hospital patients showed Cronbach's alpha coefficients of over 0.8 for both subscales, while concurrent validity was shown with the Beck Depression Inventory [12]. Arabic versions validated in a surgical inpatient sample have also shown good reliability and responsiveness [18]. In clinical situations, HADS is to facilitate the detection of comorbid anxiety and depression for referral and intervention in the context of the lack of expert mental health. Due to its power and shortness, it is widely used in hospital mental health screening.

4) *Psychological Distress Scales:* The Kessler Psychological Distress Scale (K10, and the shorter 6-item version K6). Measures nonspecific psychological distress. Kessler scale includes anxiety and depression symptoms. Created for the evaluation of populations, the test only takes a few minutes and has excellent psychometric characteristics. According to studies among Canadian military personnel, the instrument has a Cronbach's alpha around 0.88 with a good unidimensional factor structure (RMSEA $\approx$ 0.05; CFI $\approx$ 0.99) [21]. In addition, a K10 score greater than or equal to 10 provides an optimal balance of sensitivity (86%) and specificity (83%) for any past-month mental disorder. These scales play a significant role in mental health monitoring, helping monitor levels for distress in the general and clinical populations. Due to their non-specific nature, they allow wide screening, but may require follow-up with disorder-specific measures for a conclusive diagnosis [22].

5) *Challenges in Questionnaire Selection and Validity:* Choosing suitable mental health questionnaires is fraught with problems because of the overlap of symptoms among different disorders and the heterogeneous character of screening instruments. The existence of similar symptoms like sleep disturbance which are common among depression and anxiety creates a dilemma in diagnosis. This kind of variation can cause misclassification to happen and reduce diagnostic specificity. Also, comparing findings across studies can also be challenging because of this particular reason. Cultural and linguistic adaptation is important to ensure valid measurement across populations. Testing of measurement invariance like that of GAD-7 confirms that the instrument functions similarly across sex and age, thus not biased [9]. Rigorous translations like forward and backward translation and pilot testing improves cross-cultural applicability. Practical aspects such as the respondent's burden, if excessive, will lead to incomplete data and reduce reliability. The choice of clinical or research tools is influenced by interpretation and licensing requirements. A challenge remains to balance a thorough assessment with shortness and ease of use. These parameters mean that instrument assessment and validation are required for the context of use [13].

2. DIGITIZATION OF QUESTIONNAIRE-BASED ASSESSMENTS

Digital delivery is transforming the mental health assessment sector by moving verified tools from paper to web and mobile platforms. Using the above-mentioned standards requires strong privacy, security, usability, and proof of measurement equivalence across modes, along with automated scoring and longitudinal monitoring in clinical systems. This section provides a brief overview of electronic versions of established scales; Q&A platforms; computer-adaptive testing; NLP add-ons; and practical standards for safe, equitable use.

A. Electronic Versions of Conventional Scales

Many mental health questionnaires which were originally developed for paper are now available in the digital format on a website or mobile app. Digital management facilitates the management of large scale population, automated assessment and reporting, and the potential for real-time monitoring of symptoms enabled through electronic health records. Patients can complete assessments remotely for improved accessibility. Though things are going digital, the digital literacy of the users is not uniform, there are issues with privacy and data; and need safety of data storing and transmission. Difficulty engaging with technology-based assessment, especially for the elderly or those not accessing digital devices; said users. Moreover, to demonstrate that electronic and paper measures are equivalent, rigorous psychometric verification is essential with a measurement invariance test to show that scoring and interpretation are the same [9]. Researches that evaluate the digital versions of PHQ-9 and GAD-7 show that they are still valid and reliable. Thus, their implementation into clinical workload and population health monitoring is justifiable [18]

B. Question-and-Answer–Based Digital Tools

A number of systematic reviews of digital mental health assessments from about 2005–2021 already show a burgeoning world of question-and-answer–based tools aimed at adults. These digital assessments deploy self-report symptom questionnaires, along with proprietary and computational scoring algorithms, validated against gold-standard diagnostic interviews, such as the SCID. Measures of screening and diagnostic accuracy, including sensitivity, specificity, Youden index, and area under the curve (AUC), show promising performance for depression, anxiety, PTSD and ADHD. However, the reviews stress under- representation of marginalized populations and diversity within clinical settings (e.g. LGBQQ and transgender people) [23]. Some online platforms combine interactive decision- support, real-time behavioral monitoring, and long-term symptom tracking (e.g. depression and anxiety scales) with standalone digital questionnaires. The provision of these tools can broaden mental health screening out of conventional healthcare settings, yet requiring further validation and data governance scrutiny [10].

C. Computer Adaptive Testing (CAT)

Computer adaptive testing refers to a method of testing mental health in which items are selected based on previous answers. This tailor item selection helps to maximize accuracy and minimize burden on the respondent. An example of the CAT-MH suite, which uses multidimensional item response theory, selects items from large banks for measuring such constructs as depression (CAT-DI), anxiety (CAT- ANX), PTSD (CAT-PTSD), and substance use disorders (CAT-SUD). CAT instruments normally need 10 to 12 items for each new administration. The efficiency gains are almost double those of a fixed-length questionnaire of the same scale. Also, they do not seem to decrease measurement precision in any way. The algorithm, which is adaptive, picks initial items at mid-severity level. It then refines the estimates again and again to meet a pre-set target threshold for measurement uncertainty. This helps improve sensitivity to an individual's specific symptom profile [24]. According to studies done in comparison, CAT is relatively more efficient and acceptable in comparison to the traditional scales. Moreover, CAT can use an OS integration web or mobile clinical and research usability. These systems can help in monitoring mental health and allocating resources based on needs [15].

D. NLP-Based Assessment of Open-Ended Responses

NLP techniques can now be used to analyse open-ended narrative responses, as an alternative to fixed-item questionnaires. Studies have shown that free-text descriptions of one's emotional state provide higher classifications, through machine-learning classification, of mental health conditions than rating-scale symptom scores on their own. Recent research shows that using NLP methodologies on descriptive words, a classification accuracy of 64% is achieved for depression and anxiety states. This outperformed a 44% classification accuracy on traditional rating scales [10] This method tackles the issues present in rating scales like central tendency error and generalization of complex feeling. Person-centered language data capture captures nuances, variability and individual symptom expression. Advances in computational linguistics and ML have made it feasible to conduct scalable automated qualitative analysis of such free- text responses. Based on the benefits they promise to deliver, it is still too early to validate these tools across different populations and clinical settings. As such, the developing integration strategies are designed to integrate into existing diagnostic strategies, without comprising confidence and performance [10] The digitization of the question-based assessment is briefly illustrated in the Figure 1.

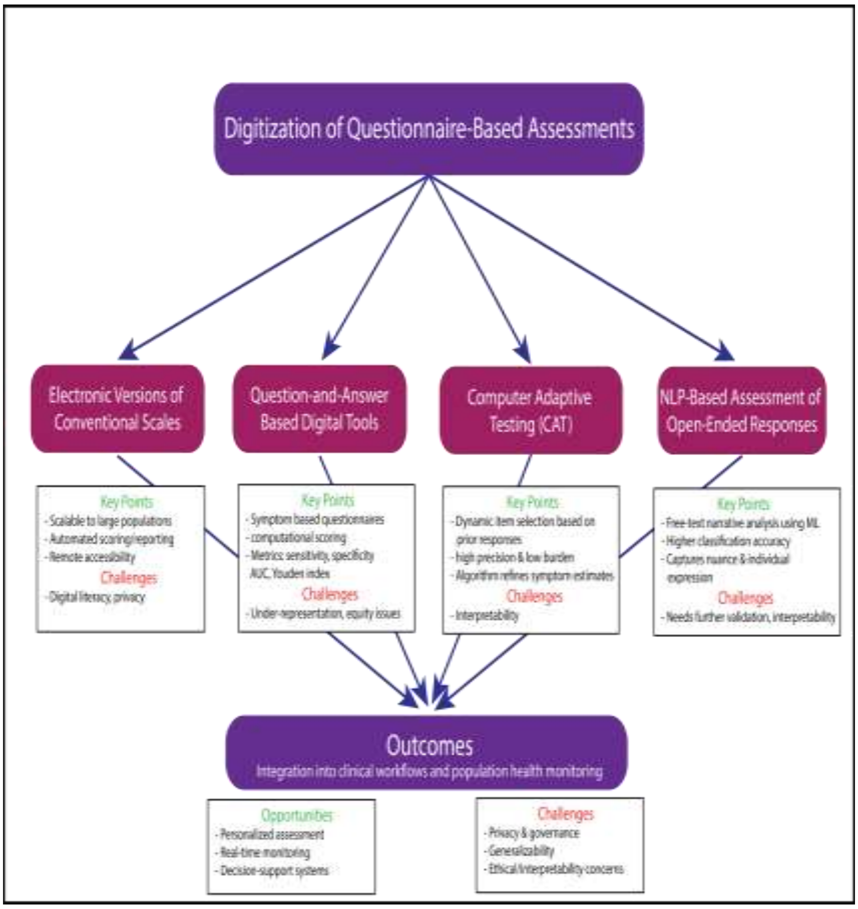

Fig. 1. Digitization of Questionnaire-Based Assessments

3. COMPUTATIONAL FRAMEWORKS FOR ANALYZING QUESTIONNAIRE DATA

Computational frameworks turn questionnaire data into decision-ready insights through a reproducible pipeline: (i) curate multi-source data and preprocess to handle noise and missingness; (ii) engineer features from items and text (including embeddings); (iii) train and validate machine-learning models with appropriate metrics and cross- validation; and (iv) integrate outputs into clinical workflows via EHRs with human oversight. This section outlines each step, emphasizing data quality, transparency, fairness, and security.

A. Data Sources and Preprocessing

Experts who study computational mental health use different types of data. This can be data from a clinical trial study, an electronic health record (EHR), a digital mental health app, social media and a wearable sensor output data. The different kinds of data require careful ways of getting them to ensure that they are quality data, representative data, and complete data. The ability to deal with missing and noisy data is important for obtaining robust models. Preprocessing varies by data type. Responses to Likert- scale questionnaires are coded to numerical values and normalized to the same scale. Open-ended textual data need natural language preprocessing techniques that include tokenization (breaking the text into words or phrases), lemmatization, stop-word removal and embedding to numerical vectors suitable for feeding into ML models. The complex preprocessing pipeline is very important for the data integrity. Moreover, it is very important for the feature extraction and model training [10].

B. Feature Extraction and Representation

Feature extraction turns raw questionnaire or textual responses into informative input variables for machine learning models. The raw item scores or composite factor scores obtained from exploratory or confirmatory factor analyses could be used as features for traditional questionnaires. Dimensionality reduction methods can extract key features and avoid multicollinearity. The embedding methods alternate with textual data representation whenever there are words or phrases. They range from a simple bag of words model to more sophisticated distributed representations including Word2Vec and BERT. Machine-learning algorithms can use embeddings to detect subtle language related to mental health. In applied studies, features are extracted from descriptive free-text using language models. Thereafter, the features are classified using logistic regression or neural networks to obtain high-accuracy symptom classification [10].

C. Machine-Learning Models

Most progresses in the prediction and classification of mental health are driven by algorithms that fall under the category of supervised machine learning. Some common methods used include convolutional neural networks (CNNs), random forests, support vector machines (SVMs), deep neural networks. Every one of them offers trade-offs in Interpretability, computational complexity, and performance.According to an extensive literature review article published in 2024, the accuracy of CNNs is more than any other method for the diagnosis of bipolar disorder. Also, CNNs can detect complex features in the data. In general for prediction random forests achieve superior performance compared to other classifiers as they are robust to overfitting and easy to interpret [10]. Unsupervised learning, including clustering and anomaly detection, is still less frequently used clinically but has exciting potential for exploratory data analysis, subtyping, and discovery of new patterns.Assessment of machine learning models requires metrics such as accuracy, sensitivity, specificity, area under the AUC curve, Youden index and rigorous cross-validation frameworks to evaluate generalizability. The insufficiency of evidence to support the choice of algorithms, model opacity hindering clinical implementation, and lack of congruence between ML approach and data characteristics are some of the challenges. Tackling these challenges through transparency and elaborate designs of models is very important for the implementation [10].

D. Integration with Healthcare Workflows

The usefulness of computational frameworks clinically de- pends on their integration in a healthcare workflow. Clinical practice can benefit by prioritizing the three domains of ML models namely risk stratification, triaging and monitoring response over time.Interoperability with electronic health systems can ensure that predictive insights are available at the point of care. Involve clinicians in the collaborative design of these essential systems to trust and use them. Model outputs require human oversight to ensure appropriate use and management of ethics. According to Afshar et al., 2023, user-centered frameworks, institution approvals, and cybersecurity safe- guards are essential for deployment studies.This section is summerized in the figure 2.

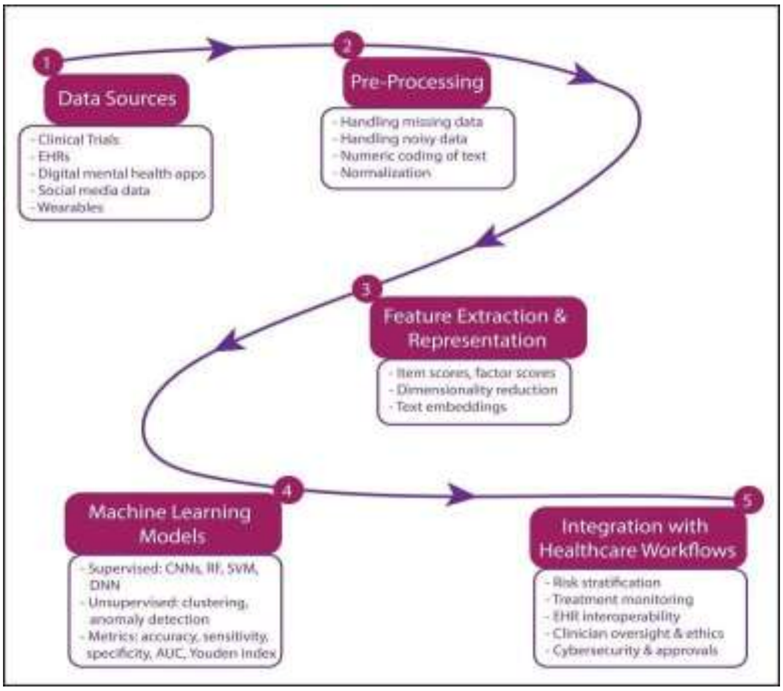

Fig. 2. Computational Frameworks for Analyzing Questionnaire Data

4. ETHICAL, LEGAL AND SOCIAL CONSIDERATIONS

A. Data Privacy and Security

We must take the highest precautions while handling sensitive mental health information. Digital mental health assessments need strong encryption, secure data storage and limit access protocols. Moreover, the participants must know how their data will be used. When data is compromised, it leads to leaking of confidential information or an increase in stigma. It also reduces engagement with services. The analyses of privacy risks in digital mental health research highlight the need to balance the utility of data with patient rights [25].

B. Bias and Fairness

AI and ML models trained on data that is biased or non-representative risk perpetuating and amplifying social inequities. To identify and alleviate harm to marginalized demographics, the performance of algorithms must be disaggregated by demographic. Researchers say we should constantly monitor for bias and use fairness-aware machine-learning techniques. Unless such measures are taken, disadvantaged groups may remain poorly serviced or miscategorized affecting the public health aims [21].

C. Transparency and Accountability

For clinical utilization and regulatory oversight, disclosure of algorithm development, validation procedures, and limitations is essential. Trusted relationships among clinicians, patients and regulators can arise through open-source model architectures, external validations and documentation.

Regulatory frameworks have to change in order to overcome the unique issues that will arise due to the influence of AI on healthcare and more importantly who will be held accountable and responsible for such decisions. The ethical guidelines suggest involving developers, clinicians and ethicists to ensure responsible deployment of AI in mental health [26].

D. Digital Divide and Health Disparities

The difference in access to technology can widen the gap in mental health care, known as the digital divide. Populations with limited access to the internet or digital resources may be unlikely to benefit from digital assessment tools, thus widening the care gap. Efforts such as providing access devices, digital literacy training and alternatively maintaining traditional assessment modes are crucial for inclusivity. Studies of disparities in telehealth show a necessity of implementation strategies that focus on equity [27].

5. FUTURE DIRECTIONS AND RESEARCH GAPS

A. Multi-Modal Assessment

The future research directions include the merging of various forms of data, such as text, speech rhythms, bio- physiological data (and heart rate variability, galvanic skin response), and devices that record behaviors. Multi-modal frameworks can help improve diagnostic accuracy, enhance the timing of detection, and improve the phenotyping of mental health states. This is because they allow for the detailed capturing of complex multi-faceted presentations of disease in an integrated manner. Pilot study results that aggregated these data streams show feasibility although synchronization of the data, privacy of the subject and analytic strategies for heterogeneous data amalgamation remain challenges [10].

B. Adaptive and Personalized Screening

Assessments that change as per the client's problem and risk level (for instance, CAT-MH) will continue to improve. Using real-time personalization in Screening instruments is expected to reduce respondent burden, enhance measurement accuracy, and increase patient engagement [24].

C. Cross-Cultural Validation and Inclusivity

Low- and middle-income countries and other linguistic and cultural populations are validated studies where lacking. Assessing measurement invariance, such as evidence for GAD-7 invariance across sex and age, which acts as a model, ensures that instrument performs well. These efforts must consider cultural representations of mental health symptoms, issues with translation, socio-economic situations affecting symptoms' expression and reporting [21], [9].

D. Interdisciplinary Collaboration

Harnessing the full potential of computational mental health assessments requires close collaboration with clinicians, data scientists, behavioral scientists, ethicists, and policy makers. Involving end-users in development through co-design boosts clinical relevance, ethical reliability and usability [10].

E. Standardization of Datasets and Reporting

There is an urgent need for common benchmarks, open-access datasets, and standardized reporting guidelines to enhance reproducibility and comparability in mental health computational research. The inconsistency in data sets and evaluation takes a toll on capacity development. Efforts calling for transparency in data sharing, harmonised protocols and agreed performance metrics can hasten advancement and clinical translation [28].

CONCLUSION

The mental health evaluation on the basis of questionnaires remains a high point of clinic assessment and public health surveillance. Disorder-specific scales such as PHQ-9, GAD- 7, BDI, and HADS have strong psychometric properties and are widely used in practice [29], [14]. Nonetheless, the emergence of these tools diminished homogeneity and comparability across studies [30]. The digitization and computational techniques, which include adaptive testing and NLP assessment of open-ended responses, as well as machine learning models, offer advantages in terms of scalability, personalization, and predictive power [31], [32]. It is necessary to be guided by ethical considerations about privacy, bias, transparency and access. Future research should try for a multi-modal approach along with a culturally inclusive and interpretable approach to be integrated into the healthcare system and, above all, to benefit the patient [33]. Future work should pursue multimodal, culturally inclusive and interpretable approaches that integrate seamlessly into healthcare systems and prioritize patient well-being.

REFERENCES

1. Jinnan Liu, Wei Ning, Ning Zhang, Bin Zhu, and Ying Mao. Estimation of the Global Disease Burden of Depression and Anxiety between 1990 and 2044: An Analysis of the Global Burden of Disease Study 2019. Healthcare, 12(17):1721, August 2024.
2. Glory Okwori, Steven Stewart, Megan Quinn, and Delaney Lawson. Health Care Burden and Expenditure Associated with Adverse Childhood Experiences in Tennessee and Virginia. Journal of Child & Adolescent Trauma, 15(3):727– 739, September 2022.
3. Lisa Dell, Carolina Casetta, Helen Benassi, Sean Cowlishaw, James Agathos, Meaghan O'Donnell, Monique Crane, Virginia Lewis, Belinda Pacella, Sonia Terhaag, David Morton, Alexander McFarlane, Richard Bryant, and David Forbes. Mental health across the early years in the military. Psychological Medicine, 53(8):3683–3691, June 2023.
4. Katie J. S. Lewis, Catrin Lewis, Alice Roberts, Natalie A. Richards, Claudia Evison, Holly A. Pearce, Keith Lloyd, Alan Meudell, Bethan M. Edwards, Catherine A. Robinson, Rob Poole, Ann John, Jonathan I. Bisson, and Ian Jones. The effect of the COVID-19 pandemic on mental health in individuals with pre-existing mental illness. BJPsych Open, 8(2):e59, March 2022.
5. Ann John, Joanne McGregor, David Fone, Frank Dunstan, Rosie Cornish, Ronan A. Lyons, and Keith R. Lloyd. Case- finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. BMC Medical Informatics and Decision Making, 16(1):35, December 2016.
6. Koen Demyttenaere and Elke Heirman. Assessment Tools in Psychiatry. In Allan Tasman, Michelle B. Riba, Renato D. Alarco´n, Ce´sar A. Alfonso, Shigenobu Kanba, Dusica Lecic-Tosevski, David M. Ndetei, Chee H. Ng, and Thomas G. Schulze, editors, Tasman's Psychiatry, pages 1333–1364. Springer International Publishing, Cham, 2024.
7. Amir Shabani, Samira Masoumian, Somayeh Zamirinejad, Maryam Hejri, Tahereh Pirmorad, and Hooman Yaghmaeezadeh. Psychometric properties of Structured Clinical Interview for DSM-5 Disorders- Clinician Version (SCID-5-CV). Brain and Behavior, 11(5):e01894, May 2021.
8. B. Arroll, F. Goodyear-Smith, S. Crengle, J. Gunn, N. Kerse, T. Fishman, K. Falloon, and S. Hatcher. Validation of PHQ-2 and PHQ-9 to Screen for Major Depression in the Primary Care Population. The Annals of Family Medicine, 8(4):348–353, July 2010.
9. K. D. Rogers, A. Young, K. Lovell, M. Campbell, P. R. Scott, and S. Kendal. The British Sign Language Versions of the Patient Health Questionnaire, the Generalized Anxiety Disorder 7-Item Scale, and the Work and Social Adjustment Scale. Journal of Deaf Studies and Deaf Education, 18(1):110–122, January 2013.
10. Majid Afshar, Sabrina Adelaine, Felice Resnik, Marlon P Mundt, John Long, Margaret Leaf, Theodore Ampian, Graham J Wills, Benjamin Schnapp, Michael Chao, Randy Brown, Cara Joyce, Brihat Sharma, Dmitriy Dligach, Elizabeth S Burnside, Jane Mahoney, Matthew M Churpek, Brian W Patterson, and Frank Liao. Deployment of Real- time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults. JMIR Medical Informatics, 11:e44977, April 2023.
11. Md Manjurul Ahsan, Shahana Akter Luna, and Zahed Siddique. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. Healthcare, 10(3):541, March 2022.
12. Ioannis Michopoulos, Athanasios Douzenis, Christina Kalkavoura, Christos Christodoulou, Panayiota Michalopoulou, Georgia Kalemi, Katerina Fineti, Paulos Patapis, Konstantinos Protopapas, and Lefteris Lykouras. Hospital Anxiety and Depression Scale (HADS): validation in a Greek general hospital sample. Annals of General Psychiatry, 7(1):4, December 2008.

13. Kurt Kroenke, Jingwei Wu, Zhangsheng Yu, Matthew J. Bair, Jacob Kean, Timothy Stump, and Patrick O. Monahan. Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials. Psychosomatic Medicine, 78(6):716–727, July 2016.
14. Berend Terluin, Harm Wj Van Marwijk, Herman J Ade`r, Henrica Cw De Vet, Brenda Wjh Penninx, Marleen Lm Hermens, Christine A Van Boeijen, Anton Jlm Van Balkom, Jac Jl Van Der Klink, and Wim Ab Stalman. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. BMC Psychiatry, 6(1):34, December 2006.
15. Laura J. Julian. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care & Research, 63(S11), November 2011
16. Robert D. Gibbons, David J. Weiss, David J. Kupfer, Ellen Frank, Andrea Fagiolini, Victoria J. Grochocinski, Dulal K. Bhaumik, Angela Stover, R. Darrell Bock, and Jason C. Immekus. Using Computerized Adaptive Testing to Reduce the Burden of Mental Health Assessment. *Psychiatric Services*, 59(4):361–368, April 2008.
17. Olalekan Lee Aiyegbusi, Jessica Roydhouse, Samantha Cruz Rivera, Paul Kamudoni, Peter Schache, Roger Wilson, Richard Stephens, and Melanie Calvert. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nature Communications*, 13(1):6026, October 2022.
18. AbdullahSulieman Terkawi, Siny Tsang, GhadahJumaan AlKahtani, SumayaHussain Al-Mousa, Salma Al Musaed, UsamaSaleh AlZoraigi, EsraaM Alasfar, KhalidS Doais, Anas Abdulrahman, and KhaildAli Altirkawi. Development and validation of Arabic version of the Hospital Anxiety and Depression Scale. *Saudi Journal of Anaesthesia*, 11(5):11, 2017.
19. Lenore Sawyer Radloff. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3):385–401, June 1977.
20. Rachel A. Belk, Mark Pilling, Katherine D. Rogers, Karina Lovell, and Alys Young. The theoretical and practical determination of clinical cut-offs for the British Sign Language versions of PHQ-9 and GAD-7. *BMC Psychiatry*, 16(1):372, December 2016.
21. Scott D. Easton, Najwa S. Safadi, Yihan Wang, and Robert G. Hasson. The Kessler psychological distress scale: translation and validation of an Arabic version. *Health and Quality of Life Outcomes*, 15(1):215, December 2017.
22. Jeremy Aldworth, Lisa J. Colpe, Joseph C. Gfroerer, Scott P. Novak, James R. Chromy, Peggy R. Barker, Kortnee Barnett-Walker, Rhonda S. Karg, Katherine B. Morton, and Katherine Spagnola. The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration analysis. *International Journal of Methods in Psychiatric Research*, 19(S1):61–87, June 2010.
23. A Aidala, J Havens, Ca Mellins, S Dodds, K Whetten, D Martin, L Gillis, and P Ko. Development and validation of the Client Diagnostic Questionnaire (CDQ): a mental health screening tool for use in HIV/AIDS service settings. *Psychology, Health & Medicine*, 9(3):362–380, August 2004.
24. Chad Ebesutani, Adam Bernstein, Bruce F. Chorpita, and John R. Weisz. A transportable assessment protocol for prescribing youth psychosocial treatments in real-world settings: Reducing assessment burden via self-report scales. *Psychological Assessment*, 24(1):141– 155, 2012.
25. Abayomi Arowosegbe and Tope Oyelade. Application of Natural Language Processing (NLP) in Detecting and Preventing Suicide Ideation: A Systematic Review. *International Journal of Environmental Research and Public Health*, 20(2):1514, January 2023.
26. Kate Churruca, Chiara Pomare, Louise A. Ellis, Janet C. Long, Suzanna B. Henderson, Lisa E. D. Murphy, Christopher J. Leahy, and Jeffrey Braithwaite. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expectations*, 24(4):1015–1024, August 2021.
27. Beth A. Smith, Anna M. Georgiopoulos, Amy Mueller, Janice Abbott, Paula Lomas, Enid Aliaj, and Alexandra L. Quittner. Impact of COVID-19 on mental health: Effects on screening, care delivery, and people with cystic fibrosis. *Journal of Cystic Fibrosis*, 20:31–38, December 2021.
28. Sebastian Raschka. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning, 2018. Version Number: 3.
29. K. Kroenke, R. L. Spitzer, and J. B. Williams. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606–613, September 2001.
30. Jennifer J. Newson, Daniel Hunter, and Tara C. Thiagarajan. The Heterogeneity of Mental Health Assessment. *Frontiers in Psychiatry*, 11:76, February 2020.
31. Lauren M. O'Reilly, Azhar I. Dalal, Serena Maag, Matthew T. Perry, Alex Card, Max B. Bohrer, Jackson Hamersly, Setarah Mohammad Nader, Kelli Peterson, David G. Beiser, Robert D. Gibbons, Brian M. D'Onofrio, and Paul I. Musey. Computer adaptive testing to assess impairing behavioral health problems in emergency

department patients with somatic complaints. *Journal of the American College of Emergency Physicians Open*, 3(5):e12804, October 2022.

32. Sverker Sikstro¨m, Ieva Valavicˇiu¯te˙, Inari Kuusela, and Nicole Evors. Question-based computational language approach outperforms rating scales in quantifying emotional states. *Communications Psychology*, 2(1):45, May 2024.
33. Nicole Martinez-Martin. A broader approach to ethical challenges in digital mental health. World psychiatry: official journal of the World Psychiatric Association (WPA), 23(3):394–395, October 2024.