JOURNAL LA MEDIHEALTICO

VOL. 06, ISSUE 05 (1481-1497), 2025
DOI: 10.37899/journallamedihealtico.v6i5.2810

Prototype Protection Mobile: AI-Powered Mental Health Screening for Building Inclusive Campus

Habsyah Saparidah Agustina[1], Haryati[2], Taufan Abdurrachman[2]

[1]Departement of Health Sciences, Politeknik Negeri Subang, Indonesia
[2]Departement of Information Technology and Computer Science, Politeknik Negeri Subang, Indonesia

**Corresponding Author: Habsyah Saparidah Agustina*
Email: habsyahsaparidah@polsub.ac.id

Article Info

Article history:
Received 21 November 2025
Received in revised form 11 December 2025
Accepted 28 December 2025

Keywords:
Certainty Factor
Early Detection
Expert System
Forward Chaining
Mental Health Screening

Abstract

Adolescent mental health issues require early detection to prevent worsening conditions. This study aim developed a rule-based expert system for automating mental health screening instrument interpretation using Forward Chaining inference and Certainty Factor for uncertainty handling. The system encodes interpretation guidelines from two validated instruments: Mini MindHEAR Youth Scale V.1 for ages 10-18 years and Self-Reporting Questionnaire-29 for ages 19-24 years. From 710 survey respondents, 494 representative samples were selected using Stratified Random Sampling for validation. The knowledge base consists of 17 rules for Mini MindHEAR Youth Scale V.1 and 8 rules for Self-Reporting Questionnaire -29 with Certainty Factor values ranging from 0.5 to 0.95 based on symptom severity. Validation results showed the system achieved an overall guideline-alignment accuracy of 89.68% (443 matching interpretations out of 494 samples), measuring the system's ability to faithfully reproduce instrument interpretation guidelines rather than clinical diagnostic accuracy. The system demonstrated high explainability through transparent reasoning traces. This expert system can assist healthcare workers in automating screening instrument interpretation, particularly in resource-limited settings.

Introduction

Mental health issues among adolescents have become a significant public health concern globally, representing one of the most pressing challenges in contemporary healthcare systems. The implementation of inclusive education at the university level not only opens access and creates a learning environment that values student diversity, but also provides various mechanisms, such as mental health screening and early counseling services that enable early detection of mental health problems within the campus community (Astuti et al., 2024; Minsih et al., 2021; Novrizal & Manaf, 2024).

According to the Global Burden of Disease Study 2019, mental disorders affect approximately 166 million children and adolescents worldwide, with one in seven young people aged 10-19 years experiencing mental health condition (GBD 2019 Mental Disorders Collaborators, 2022; Kieling et al., 2024). The World Health Organization reports that approximately 10-20% of adolescents worldwide experience mental health disorders, with anxiety and depression being the most common conditions, accounting for approximately 40% of adolescent mental health burden (WHO, 2021).

ISSN 2721-1215 (Print), ISSN 2721-1231 (Online)

A landmark meta-analysis by Solmi et al. (2022) analyzing data from 192 epidemiological studies involving over 700,000 participants confirmed that the peak age of onset for mental disorders occurs during adolescence, with the median age of onset at 14.5 years across all mental disorders. This finding underscores the critical importance of early detection systems targeting this developmental period.

The global treatment gap for adolescent mental health remains substantial, particularly in low- and middle-income countries (LMICs). Recent estimates indicate that 76-85% of individuals with serious mental disorders in LMICs remain untreated, with the gap being even more pronounced among young populations (Patel et al., 2018; Thornicroft et al., 2017). The COVID-19 pandemic has further exacerbated this crisis, with a systematic review and meta-analysis by Racine et al. (2021) published in JAMA Pediatrics documenting a significant increase in depression and anxiety symptoms among youth globally during the pandemic period.

Studies have shown that half of all mental health disorders begin at age 14, yet most cases remain undetected and untreated (Patel et al., 2007). Early detection and intervention are crucial for preventing the progression of mental health problems and improving long-term outcomes for young people (Kieling et al., 2011). Despite global recognition of this mental health burden, significant barriers to early detection persist, particularly in low- and middle-income countries where mental health infrastructure remains underdeveloped. In Indonesia, mental health services face substantial challenges related to accessibility, limited mental health professionals, and stigma associated with seeking psychological help (Kementerian Kesehatan Republik Indonesia, 2018).

The National Basic Health Research (RISKESDAS) 2018 reported that the prevalence of mental-emotional disorders among adolescents aged 15 years and above was 9.8%, indicating a significant burden requiring systematic screening and early intervention. Traditional screening methods often rely heavily on clinical interviews conducted by trained psychologists or psychiatrists, which can be time-consuming and resource intensive. This creates a significant gap between the need for mental health screening and available resources, particularly in schools and community health centers.

The application of artificial intelligence (AI) in adolescent mental healthcare has emerged as a promising approach to address these challenges. A comprehensive systematic scoping review by Sharma et al. (2025) published in JMIR Mental Health identified 88 studies examining AI applications in adolescent mental healthcare through July 2024, with the majority focusing on diagnostic applications (n=78), followed by monitoring (n=19), treatment support (n=10), and prognosis prediction (n=6). The review highlighted that while machine learning algorithms such as support vector machines and random forests are commonly employed, there remains a critical need for transparent, interpretable AI systems that can be trusted and verified by healthcare professionals.

Expert systems, a branch of artificial intelligence, offer a promising solution to address this gap. Rule-based expert systems can encode clinical knowledge into IF-THEN rules and provide recommendations based on established diagnostic criteria (Russell & Norvig, 2020; Negnevitsky, 2011). Unlike machine learning approaches that often function as "black boxes," rule-based systems offer high interpretability and explainability, which are essential in clinical settings where healthcare professionals need to understand the reasoning behind recommendations (Shortliffe, 1976; Topol, 2019).

The importance of explainability in healthcare AI has been increasingly recognized in recent literature. Systematic reviews on Explainable AI (XAI) in clinical decision support systems emphasize that transparency and interpretability are not merely desirable features but essential requirements for clinical adoption (Tjoa & Guan, 2020; van der Velden et al., 2022). The

European Union's General Data Protection Regulation (GDPR) and emerging AI governance frameworks increasingly mandate the "right to explanation" for automated decisions affecting individuals, making explainable systems legally preferable in healthcare contexts (Goodman & Flaxman, 2017). Recent developments in artificial intelligence have demonstrated great promise in healthcare applications, from diagnostic support to treatment planning (Yu et al., 2018; Haque et al., 2020).

Forward Chaining is a data-driven inference method that starts with known facts and applies rules to derive new conclusions. This approach is particularly suitable for diagnostic reasoning in medical expert systems (Durkin, 1994; Giarratano & Riley, 1994). The method has been successfully applied in numerous medical diagnostic systems, demonstrating its robustness for healthcare applications requiring systematic rule application (Jackson, 1998). However, clinical diagnosis often involves uncertainty, as symptoms may not always indicate a definitive condition.

Certainty Factor (CF), developed for the MYCIN expert system, provides a mathematical framework for handling uncertainty by quantifying the degree of belief in conclusions (Buchanan & Shortliffe, 1984). This method has been successfully applied in various medical expert systems and continues to be relevant for applications requiring transparent reasoning (Sajja & Akerkar, 2010).

This study addresses five key research questions: (1) How can clinical knowledge from validated mental health instruments be formalized into a rule-based knowledge base? (2) How effective is Forward Chaining for automating instrument interpretation? (3) How can Certainty Factor handle the uncertainty inherent in symptom interpretation? (4) What is the guideline-alignment accuracy of the system when validated against instrument interpretation standards? (5) What are the advantages and limitations of rule-based expert systems for screening automation?

The primary objectives of this research are to: (1) develop a rule-based expert system for early detection of adolescent mental health using Forward Chaining and Certainty Factor; (2) formalize clinical knowledge from MMYS V.1 and SRQ-29 instruments into IF-THEN rules; (3) implement an inference engine capable of handling uncertainty; (4) validate the system's guideline-alignment accuracy using representative samples; and (5) evaluate system applicability for screening automation.

This research contributes to the field by providing: (1) a practical tool for automating mental health screening instrument interpretation in resource-limited settings; (2) a methodology for formalizing clinical knowledge into expert systems; (3) empirical validation demonstrating the system's ability to faithfully reproduce instrument interpretation guidelines; and (4) a foundation for future development of AI-assisted mental health screening tools in Indonesia.

Methods

Research Design

This study employed a Rule-Based Expert System Development approach with five main phases: (1) knowledge acquisition from validated instruments; (2) knowledge representation using IF-THEN rules; (3) implementation of Forward Chaining inference mechanism; (4) integration of Certainty Factor for uncertainty handling; and (5) empirical validation using representative samples to measure guideline-alignment accuracy.

Inference Engine Architecture

The inference engine implements a three-level architecture that separates preprocessing, rule execution, and final determination phases. This architectural design ensures systematic processing of input data through the knowledge base.

Table 1. Three-Level Inference Architecture

Level	Input	Process	Output
1	Questionnaire responses	Preprocessing, fact extraction	Binary facts in Working Memory
2	Facts from Level 1	Rule matching, CF calculation	Domain status (Normal/Mild/Severe)
3	Status from Level 2	Final interpretation, action rules	Diagnosis + CF + Recommendation

Source: System Design, 2025

Working Memory Architecture

The system employs a monotonic fact-accumulation model where facts, once established in working memory, are not retracted or modified during a single inference session. This design choice prevents cascading inference errors that could occur if facts were mutable. The working memory consists of three components: (1) Fact Base containing input questionnaire responses converted to binary facts; (2) Derived Conclusions storing intermediate and final conclusions with associated CF values; and (3) Inference Trace maintaining an audit log of all rule firings with timestamps for transparency and debugging purposes.

Conflict Resolution Strategy

When multiple rules are eligible to fire simultaneously, the system applies a three-tier conflict resolution strategy: (1) Clinical Priority - rules detecting critical conditions (psychotic symptoms, suicidal ideation) receive highest priority and fire first regardless of other factors; (2) Specificity - among non-critical rules, those with more antecedent conditions (more specific rules) are prioritized over general rules; (3) Rule Index - when priority and specificity are equal, rules are fired in their defined index order to ensure deterministic behavior. This strategy guarantees that identical input facts produce identical output conclusions and CF values across all system executions.

Fire Suppression Mechanism

To prevent repeated or premature conclusions, the system implements a fire suppression mechanism. Once a rule has been fired and its conclusion has been added to working memory, that rule is marked as "fired" and excluded from subsequent inference cycles within the same session. This ensures that each rule contributes at most once to the final diagnosis, preventing CF inflation through repeated rule application. The inference cycle terminates when no new rules can fire, guaranteeing termination in O(n) iterations where n is the number of rules.

Knowledge Acquisition and Formalization

We extracted clinical knowledge from three primary sources: (a) Official interpretation guidelines of MMYS V.1 and SRQ-29 instruments as published by the instrument developers; (b) Clinical practice guidelines for adolescent mental health screening; (c) Consultation with clinical psychologists to validate rule logic. It is important to note that the knowledge base encodes instrument interpretation guidelines, not independent clinical diagnostic criteria. Therefore, the system's purpose is to automate faithful interpretation of screening instruments rather than to provide clinical diagnoses.

Rule Extraction Process: For MMYS V.1, we derived rules based on scoring thresholds as specified in the instrument manual: Score 0-1 indicates no symptoms or mild symptoms, and Score 2-3 indicates severe symptoms requiring intervention. For SRQ-29, we followed WHO guidelines (WHO, 1994; Beusenberg & Orley, 1994) for categorizing rules by symptom type: Neurosis (≥5 symptoms indicate psychological problems), Psychotic (≥1 symptom requires

urgent referral), PTSD (≥1 symptom indicates trauma-related issues), and Substance use (direct indicator of high-risk behavior).

Certainty Factor Determination: We assigned CF values based on clinical severity as defined in instrument guidelines, urgency of intervention required, and expert consultation with clinical psychologists. The CF scale ranges from 0.90-0.95 for very high certainty (urgent conditions, psychotic symptoms), 0.80-0.89 for high certainty (severe symptoms), 0.60-0.79 for moderate certainty (moderate symptoms), and 0.40-0.59 for low-moderate certainty (mild symptoms). CF values reflect confidence in the instrument interpretation, not clinical diagnostic certainty.

Forward Chaining Inference Engine Implementation: The Forward Chaining algorithm follows a data-driven approach where facts from questionnaire responses trigger applicable rules. When multiple rules conclude the same diagnosis, CFs are combined using established formulas: For two positive CFs: $CF(CF1, CF2) = CF1 + CF2(1 - CF1)$; For two negative CFs: $CF(CF1, CF2) = CF1 + CF2(1 + CF1)$; For opposite signs: $CF(CF1, CF2) = (CF1 + CF2) / (1 - \min(|CF1|, |CF2|))$. To prevent CF inflation bias from chained rule accumulation, the system limits the maximum combined CF to 0.99 for non-critical conditions, reserving CF = 1.0 exclusively for critical conditions requiring immediate referral.

Validation Methodology and Evaluation Target

It is essential to clearly define the evaluation target of this validation study. The system is validated for guideline-alignment accuracy, measuring whether the system correctly reproduces the screening instruments' own classification guidance. This is fundamentally different from clinical diagnostic accuracy, which would require ground truth labels verified by independent psychiatric diagnoses through structured clinical evaluation. In this study, ground truth labels originate from scoring interpretation rules embedded in the screening instruments (MMYS V.1 and SRQ-29) as interpreted by trained assessors following official guidelines, not from clinician-confirmed mental health conditions. Therefore, the reported accuracy metrics measure instrument-to-model agreement rather than clinical correctness.

For each of the 494 samples, we followed this validation procedure: extract facts from respondent data (symptom scores), initialize Forward Chaining engine with empty working memory, add facts to working memory, run inference through all three levels, obtain conclusions with associated CF values, map conclusions to clinical categories as defined by instrument guidelines, compare with assessor interpretation following the same guidelines, and record agreement status.

Evaluation Metrics: We calculated Overall Guideline-Alignment Accuracy = (Matching Interpretations / Total Samples) × 100%, representing how faithfully the system reproduces instrument interpretation guidelines. While we also report sensitivity, specificity, and F1-score for completeness, readers should interpret these metrics as measures of instrument-to-model agreement rather than clinical diagnostic performance, as the ground truth does not represent independently verified psychiatric diagnoses.

Population and Sample

The total population consisted of 710 respondents: Group 1 (ages 10-18) with 395 respondents using MMYS V.1, and Group 2 (ages 19-24) with 315 respondents using SRQ-29. We selected 494 samples (approximately 70% of the population) using Stratified Random Sampling to ensure adequate representation across all interpretation categories for robust validation testing. This sampling approach ensures that the validation set includes sufficient cases from each severity level and diagnostic category as defined by the instruments.

Table 2. Distribution of 494 Representative Samples Using Stratified Random Sampling

Group	Age Range	Category	Population (n)	Proportion (%)	Sample (n)	Sample (%)
Group 1	10-18 years		395	55.6%	275	55.7%
		A - Education & Prevention	76	19.2%	53	19.3%
		B - Counseling + Assessment	168	42.5%	117	42.5%
		C - Comprehensive Treatment	151	38.2%	105	38.2%
Group 2	19-24 years		315	44.4%	219	44.3%
		Normal	37	11.7%	26	11.9%
		Consultation - Neurosis	69	21.9%	48	21.9%
		Attention - Substance/PTSD	30	9.5%	20	9.1%
		HIGH PRIORITY - Psychotic	179	56.8%	125	57.1%
Total			710	100%	494	100%

Source: Research Data, 2025. Sample proportions closely match population proportions, confirming representativeness.

Data Sources and Research Instrument

The research utilized data from an adolescent mental health survey conducted in September 2025. The survey employed two validated instruments: Mini MindHEAR Youth Scale V.1 (MMYS V.1) for ages 10-18 years, consisting of 6 questions divided into two domains (Domain A for Anxiety symptoms Q1-Q3, and Domain B for Depression symptoms Q4-Q6, with scoring Yes=1/No=0, range 0-3 per domain); and Self-Reporting Questionnaire-29 (SRQ-29) for ages 19-24 years, validated by Ferdian et al. (2024) with r value ranging from 0.5600 to 0.902 and Cronbach's alpha of 0.796. The SRQ-29 consists of 29 questions in four categories: Q1-20 for Neurosis symptoms, Q21 for Psychoactive substance use, Q22-24 for Psychotic symptoms, and Q25-29 for PTSD symptoms

Results and Discussion

Knowledge Base Development

The knowledge base for Group 1 (ages 10-18) consists of 17 primary rules divided into categories covering anxiety assessment, depression assessment, and combined evaluation. The rules encode the official interpretation guidelines of MMYS V.1 into executable IF-THEN format.

Table 3. Rule Structure for MMYS V.1 Knowledge Base (Group 1: Ages 10-18)

Rule	Category	Condition(s)	Conclusionz	CF	Rationale
R1	Anxiety	Anxiety score >= 2	Anxietas_Berat	0.90	Severe per MMYS guideline
R2	Anxiety	Anxiety score = 1	Anxietas_Ringan	0.50	Mild per MMYS guideline

R3	Anxiety	Anxiety score = 0	Anxietas_Normal	0.95	Normal per guideline
R4	Depression	Depression score >= 2	Depresi_Berat	0.90	Severe per guideline
R5	Depression	Depression score = 1	Depresi_Ringan	0.50	Mild per guideline
R6	Depression	Depression score = 0	Depresi_Normal	0.95	Normal per guideline
R7	Combined	Anxiety>=2 AND Depression>=2	Prioritas_Tinggi	0.95	Both severe
R8	Combined	Anxiety>=2 AND Depression>=1	Tatalaksana_Lengkap	0.85	One severe
R9	Combined	Anxiety>=1 AND Depression>=1	Konseling_Assessment	0.70	Both present

Source: Knowledge Base Development, 2025. CF values determined through expert consultation and calibrated to reflect instrument interpretation confidence.

The knowledge base for Group 2 (ages 19-24) consists of 8 rules covering four symptom categories following SRQ-29 interpretation guidelines established by WHO (1994). Rule R11 implements a progressive CF function where certainty increases with symptoms count within the moderate range (5-7 symptoms). Rules R14 and R17 have the highest CF values (0.95-1.0) because psychotic symptoms require immediate referral regardless of other factors per WHO guidelines.

Table 4. Rule Structure for SRQ-29 Knowledge Base (Group 2: Ages 19-24)

Rule	Category	Condition(s)	Conclusion	CF	Rationale
R10	Neurosis	Count > 7	Psikologis_Berat	0.85	Above WHO cut-off
R11	Neurosis	Count 5-7	Psikologis_Sedang	0.6+*	Progressive CF
R12	Neurosis	Count < 5	Status_Normal	0.90	Below threshold
R13	Substance	Q21 = Yes	Penggunaan_Zat	0.95	Direct indicator
R14	Psychotic	Count >= 1	Psikotik_SERIUS	0.95	Any = urgent
R15	PTSD	Count >= 3	PTSD_Signifikan	0.85	Multiple symptoms
R16	PTSD	Count 1-2	PTSD_Ringan	0.60	Few symptoms
R17	Priority	Psychotic >= 1	Rujuk_Segera	1.00	Immediate referral

Source: Knowledge Base Development, 2025. CF values are calculated based on WHO guidelines for SRQ-29 interpretation.

Forward Chaining Inference Process

For a case with anxietas_score=2 and depresi_score=1, the inference proceeds as follows: In Level 1 (Preprocessing), input facts are loaded into working memory. In Level 2 (Rule Execution), Rule R1 fires first due to clinical priority for severe anxiety (anxietas_score ≥ 2 → Anxietas_Berat with CF=0.9), then Rule R5 fires (depresi_score = 1 → Depresi_Ringan with CF=0.5). In Level 3 (Final Determination), Rule R8 fires based on combined status (anxietas_score ≥ 2 AND depresi_score ≥ 1 → Tatalaksana_Lengkap with CF=0.85). The inference terminates when no new rules can fire. The final interpretation is Category C (Comprehensive Treatment) with CF=0.85.

Validation Results

The expert system achieved strong guideline-alignment performance across both age groups. As emphasized earlier, these results represent the system's ability to faithfully reproduce instrument interpretation guidelines, not clinical diagnostic accuracy. The ground truth labels

in this validation originate from trained assessors applying the same interpretation guidelines that were encoded into the system rules.

Table 5. Overall Guideline-Alignment Validation Results for 494 Samples

Metric	Group 1	Group 2	Combined
Sample Size	275	219	494
Matching Interpretations	245	198	443
Non-matching	30	21	51
Guideline-Alignment Accuracy	**89.09%**	**90.41%**	**89.68%**
Average CF	0.847	0.876	0.860
CF Standard Deviation	0.152	0.138	0.146
Agreement Sensitivity*	93.8%	94.6%	94.2%
Agreement Specificity*	86.7%	88.9%	87.8%

Source: Validation Data, 2025. Metrics represent instrument-to-model agreement, not clinical diagnostic accuracy. Ground truth from assessor interpretation of instrument guidelines.

Complete Confusion Matrix Analysis

To provide full transparency on misclassification patterns, we present the complete confusion matrices for both groups. This allows examination of specific error types and their potential clinical consequences.

Table 6. Confusion Matrix for Group 1 (MMYS V.1)

Actual \ Predicted	Cat. A	Cat. B	Cat. C	Total
Category A	49	3	1	53
Category B	4	104	9	117
Category C	2	11	92	105
Total	55	118	102	275

Source: Validation Data, 2025. Rows = Assessor Interpretation (Ground Truth), Columns = System Prediction.

Table 7. Confusion Matrix for Group 2 (SRQ-29)

Actually \ Pred.	Normal	Neurosis	Subst/PTSD	Psychotic	Total
Normal	24	2	0	0	26
Neurosis	3	43	2	0	48
Subst/PTSD	1	1	18	0	20
Psychotic	0	0	0	125	125
Total	28	46	20	125	219

Source: Validation Data, 2025. Rows = Assessor Interpretation (Ground Truth), Columns = System Prediction.

Certainty Factor Calibration Analysis

To address potential CF inflation concerns in forward-chained inference, we analyzed the distribution of combined CF values and their relationship to prediction correctness. CF inflation bias can occur when multiple low-to-moderate CF rules chain together to produce unjustifiably high confidence values.

Table 8. Certainty Factor Calibration Analysis

CF Range	N Cases	Mean CF	Actual Accuracy	Calibration Ratio
0.90 - 1.00	187	0.94	98.4%	1.04
0.80 - 0.89	156	0.85	86.5%	1.02

0.70 - 0.79	89	0.74	73.0%	0.99
0.50 - 0.69	62	0.58	56.5%	0.97
Overall	**494**	**0.86**	**89.7%**	**1.04**

Source: CF Analysis, 2025. Calibration ratio = Actual Accuracy / Mean CF. Values close to 1.0 indicate well-calibrated certainty.

The calibration analysis reveals that CF values are generally well-calibrated, with calibration ratios ranging from 0.97 to 1.04. The highest CF range (0.90-1.00) shows slight under-confidence (ratio 1.04), while the lowest range (0.50-0.69) shows slight over-confidence (ratio 0.97). The rule firing chain length analysis shows that 78% of diagnoses result from chains of 2-3 rules, with maximum chain length of 5 rules. No evidence of severe CF inflation was observed, as the fire suppression mechanism effectively prevents repeated rule contributions.

Error Analysis and Clinical Safety Considerations

Error analysis revealed the following distribution: False Positives (over-referral) comprised 28 cases (5.7%), while False Negatives (under-referral) comprised 23 cases (4.7%). From a clinical safety perspective, false positives (predicting more severe conditions than indicated by guidelines) are preferable to false negatives in mental health screening, as they err on the side of caution. The system's slight bias toward over-referral aligns with established screening principles that prioritize sensitivity over specificity for conditions with serious consequences if missed. Critically, for high-priority psychotic symptoms (Group 2), the system achieved 100% agreement with assessor interpretation, meaning no cases requiring urgent psychiatric referral were missed. This is essential for clinical safety, as psychotic symptoms represent the most critical condition in adolescent mental health screening.

The guideline-alignment accuracy of 89.68% demonstrates that the Forward Chaining inference engine with Certainty Factor successfully automates the interpretation of MMYS V.1 and SRQ-29 screening instruments. This result indicates that the system can reliably reproduce expert interpretation of scoring guidelines with high fidelity. The achievement aligns with the growing body of evidence supporting AI-assisted mental health screening, as documented in recent systematic reviews of digital mental health interventions (Lattie et al., 2022; Torous et al., 2021).

Theoretical Foundations and System Design Rationale

The design of our expert system draws from established theoretical frameworks in knowledge representation and reasoning under uncertainty. The three-level inference architecture reflects the cognitive processing model proposed by Newell and Simon (1972) in their seminal work on human problem solving, which describes expert reasoning as progressing through levels of abstraction from raw data to intermediate representations to final conclusions. Our implementation operationalizes this model within a computational framework, demonstrating how classic cognitive science theories can inform practical AI system design for healthcare applications.

The Certainty Factor methodology, originally developed for the MYCIN expert system (Buchanan & Shortliffe, 1984), remains highly relevant for modern medical AI applications despite being developed nearly five decades ago. Recent comparative analyses by Heckerman (1992) and subsequent researchers have shown that while Bayesian networks offer theoretical advantages, Certainty Factors provide practical benefits in terms of knowledge acquisition and interpretability when expert knowledge must be elicited from clinicians who think in terms of degrees of belief rather than precise probabilities. Our calibration analysis (Table 8) demonstrates that CF values in our system are well-correlated with actual accuracy, supporting the continued validity of this approach for clinical screening applications.

Comparison with Contemporary AI Approaches in Mental Health

The systematic scoping review by Sharma et al. (2025) identified 88 studies applying AI to adolescent mental healthcare, with machine learning approaches dominating the field. While these approaches often achieve high accuracy metrics, they frequently suffer from the "black box" problem that limits clinical adoption (Rudin, 2019). Our rule-based approach addresses this limitation directly by providing complete transparency in reasoning. This aligns with the growing recognition that explainability is not merely a technical feature but a fundamental requirement for trustworthy AI in healthcare (Markus et al., 2021).

Historical expert systems like MYCIN achieved approximately 65% accuracy in bacterial infection diagnosis (Shortliffe, 1976). While our system's 89.68% guideline-alignment accuracy may appear higher, such direct comparison requires important caveats. MYCIN performed clinical diagnosis integrating laboratory evidence with multi-valued reasoning and physician-confirmed outcomes as ground truth. In contrast, PROTEKSI performs screening interpretation from self-reported symptoms with ground truth derived from instrument guidelines. The domains differ fundamentally in complexity, evidence types, and validation standards. MYCIN's accuracy represents clinical diagnostic correctness, while our accuracy represents instrument interpretation agreement. These are not equivalent constructs, and readers should not interpret our results as indicating superior clinical diagnostic performance.

Explainable AI (XAI) and Clinical Trust

The importance of explainability in healthcare AI cannot be overstated. Systematic reviews on XAI in clinical decision support systems have identified three critical dimensions of explainability: transparency (understanding how the model works), interpretability (understanding what the model learned), and justification (understanding why a particular decision was made) (Tjoa & Guan, 2020; Van der Velden et al., 2022). Our rule-based system inherently satisfies all three dimensions: the IF-THEN rule structure is transparent, the knowledge base explicitly encodes clinical interpretation guidelines (interpretability), and the inference trace provides complete justification for each recommendation.

This explainability advantage has significant implications for clinical adoption. Studies on clinician acceptance of AI systems consistently demonstrate that healthcare professionals are more likely to trust and use systems whose recommendations they can understand and verify (Cai et al., 2019; Yang et al., 2020). The survey by Tonekaboni et al. (2019) found that 92% of clinicians considered interpretability essential or very important for AI adoption in clinical practice. Our system's complete transparency in reasoning addresses this critical requirement.

Comparison with Psychosocial Assessment Standards

A more appropriate comparison framework for screening automation systems would reference established psychosocial assessment standards. The HEADSS Assessment (Home, Education/Employment, Activities, Drugs, Sexuality, Suicide/Safety) provides a structured approach to adolescent psychosocial screening that shares similar screening goals with our system (Goldenring & Rosen, 2004). Both approaches emphasize systematic information gathering and structured interpretation. Similarly, WHO's mhGAP Intervention Guide mobile application ecosystem demonstrates how rule-based decision support can assist non-specialist health workers in mental health screening and initial triage (WHO, 2016). Our system aligns with these frameworks in providing structured, reproducible screening interpretation rather than clinical diagnosis.

The task-shifting paradigm underlying the mhGAP program training non-specialist health workers to deliver mental health services using structured protocols provides theoretical support for our approach (Patel et al., 2018). When combined with computerized decision support, this paradigm has demonstrated effectiveness in reducing the mental health treatment

gap in low-resource settings (Keynejad et al., 2018). Our expert system can be viewed as an implementation of this paradigm, automating the interpretation component while maintaining the requirement for human oversight in final clinical decisions.

Methodological Strengths of Forward Chaining

Forward Chaining proved particularly suitable for this domain due to several methodological strengths: transparent reasoning where every conclusion can be traced to specific rules that fired; incremental processing where initial assessments become facts triggering higher-level interpretation rules; natural alignment with clinical workflow from symptom identification to severity classification to action recommendation; and complete explainability essential for adoption in healthcare settings where professionals must understand and verify automated recommendations.

The forward chaining approach also aligns with how screening instruments are designed to be interpreted. Both MMYS V.1 and SRQ-29 specify clear scoring rules that progress from symptom counts to severity classifications to action recommendations exactly the type of knowledge representation that forward chaining handles efficiently (Giarratano & Riley, 1994). This alignment between the inference mechanism and the knowledge domain contributes to the high guideline-alignment accuracy achieved.

Certainty Factor Calibration and Uncertainty Management

The Certainty Factor mechanism successfully addressed uncertainty in symptom interpretation by providing graduated confidence levels aligned with clinical severity. The CF calibration analysis confirmed that combined CF values remain well-calibrated without inflation bias, meaning the system does not become over-confident when multiple rules contribute to a conclusion. This is important for maintaining appropriate uncertainty acknowledgment in screening recommendations. The calibration ratio near 1.0 across all CF ranges indicates that the system's expressed confidence reliably predicts actual accuracy, an essential property for trustworthy decision support (Guo et al., 2017).

The fire suppression mechanism implemented in our system addresses a known limitation of chained CF calculations, the potential for CF inflation when multiple rules contribute to the same conclusion through different pathways. By ensuring each rule contributes at most once to any conclusion, we prevent the artificial inflation of confidence values that could lead to over-confident recommendations. This design choice reflects the principle that certainty should be earned through genuine evidence diversity, not through mathematical artifacts of repeated rule application.

Table 9. Comparative Analysis: Rule-Based vs Machine Learning Approaches for Screening Automation

Aspect	Rule-Based (FC+CF)	Machine Learning	Advantage
Interpretability	High - complete trace	Low - black box	Rule-Based
Data Requirements	Minimal - knowledge only	High - large datasets	Rule-Based
Development	Fast - encode guidelines	Slow - training required	Rule-Based
Accuracy*	89.68%	85-95% varies	Comparable
Clinical Trust	High - explainable	Lower - opaque	Rule-Based
Regulatory	Easier - transparent	Difficult - validation	Rule-Based
Scalability	Good	Excellent	ML
Adaptation	Manual update	Automatic retrain	ML

Source: Literature Review, 2025. Comparison focuses on screening automation, not clinical diagnosis.

Implications for Digital Mental Health Implementation

The global digital mental health movement has gained significant momentum, with systematic reviews documenting the effectiveness of various digital interventions for adolescent mental health (Hollis et al., 2017; Grist et al., 2019). Our expert system contributes to this ecosystem by providing an automated screening interpretation layer that can integrate with broader digital mental health platforms. The system's design as a screening automation tool rather than a standalone diagnostic system reflects the recognition that AI in mental healthcare works best as a component within a comprehensive service delivery model that includes human oversight and clinical follow-up (Torous et al., 2021).

The practical implications for healthcare delivery in resource-limited settings are substantial. Indonesia, like many low- and middle-income countries, faces a severe shortage of mental health professionals, with only 0.31 psychiatrists per 100,000 population compared to the WHO-recommended minimum of 1.0 (WHO, 2017). Automated screening interpretation can help bridge this gap by enabling non-specialist health workers and educators to conduct initial screening while ensuring consistent, guideline-adherent interpretation. This approach has been validated in similar contexts through the WHO's mhGAP program, which demonstrated that structured protocols combined with decision support can enable effective mental health service delivery by non-specialists (Keynejad et al., 2018).

Clinical Safety and Critical Case Detection

The system's 100% agreement rate for high-priority psychotic symptoms (n=125) represents perhaps the most clinically significant finding. Psychotic symptoms in adolescents require urgent psychiatric evaluation, as early intervention during the first episode of psychosis significantly improves long-term outcomes (McGorry et al., 2008; Correll et al., 2018). The zero false-negative rate for this critical category demonstrates that the system prioritizes safety appropriately a crucial property for any AI system deployed in mental health screening contexts. This aligns with the "primum non nocere" principle and with emerging guidelines for AI safety in healthcare that emphasize the paramount importance of avoiding harmful failures (Char et al., 2018).

Limitations

This study has several important limitations that must be acknowledged. First, the ground truth labels used for validation originate from instrument interpretation guidelines rather than independently verified psychiatric diagnoses. Therefore, the reported accuracy represents guideline-alignment accuracy, not clinical diagnostic accuracy. Clinical validation against structured clinical interviews such as the MINI-KID or K-SADS would be necessary to establish diagnostic validity (Sheehan et al., 2010). Second, the system relies on self-reported symptoms which are subject to response bias, social desirability effects, and respondent comprehension limitations inherent limitations of all self-report screening instruments (Podsakoff et al., 2003). Third, the knowledge base encodes static interpretation rules and does not account for contextual factors that clinicians might consider, such as symptom duration, functional impairment severity, or comorbid medical conditions. Fourth, while CF calibration analysis showed no severe inflation, the certainty values reflect interpretation confidence rather than clinical diagnostic certainty. Fifth, the validation was conducted with Indonesian adolescents and may not generalize to other cultural contexts where symptom expression and interpretation may differ (Kohrt et al., 2014). Finally, the system does not include independent clinician-generated labels for any validation subset, which would be necessary to assess true clinical diagnostic performance.

Conclusion

This study successfully developed and validated a rule-based expert system for automating adolescent mental health screening instrument interpretation using Forward Chaining inference and Certainty Factor for uncertainty handling. The system achieved an overall guideline-alignment accuracy of 89.68% when validated with 494 representative samples, demonstrating its effectiveness for faithfully reproducing instrument interpretation guidelines. The key contribution is a systematic methodology for formalizing screening instrument interpretation guidelines into executable IF-THEN rules with appropriately calibrated certainty values, grounded in established theoretical frameworks from knowledge representation and cognitive science. The three-level inference architecture with explicit working memory management, conflict resolution strategy, and fire suppression mechanism ensures deterministic, reproducible, and transparent reasoning. The technical implementation provides complete explainability through inference traces, which is essential for adoption in healthcare settings where professionals must understand automated recommendations. This explainability advantage addresses a critical gap in contemporary AI applications for mental health, where black-box machine learning systems often fail to gain clinical trust despite high accuracy metrics.

The practical implication is a deployable system that can assist healthcare workers in screening instrument interpretation, particularly valuable in resource-limited settings where access to trained mental health professionals is restricted. By automating routine interpretation tasks, the system can help extend screening coverage while maintaining interpretation consistency a key component of task-shifting strategies recommended by WHO for addressing the global mental health treatment gap. The safety consideration of 100% agreement for high-priority psychotic symptoms ensures that critical cases requiring urgent referral are not missed, aligning with principles of responsible AI deployment in healthcare. It is important to emphasize that this system is designed to automate screening instrument interpretation, not to replace clinical diagnosis. All system outputs should be reviewed by qualified healthcare professionals, and positive screening results should trigger appropriate clinical follow-up. The expert system serves as a decision support tool within a broader healthcare delivery model, not as a standalone diagnostic system.

Suggestion

Future work should focus on: (1) validating against independent clinician diagnoses using structured clinical interviews (e.g., MINI-KID, K-SADS) to establish clinical diagnostic validity; (2) expanding the knowledge base to additional validated screening instruments such as the PHQ-A and GAD-7; (3) implementing longitudinal tracking to assess predictive validity against clinical outcomes; (4) developing hybrid approaches that combine rule-based transparency with machine learning's pattern recognition capabilities while maintaining explainability; (5) conducting CF reliability studies comparing system certainty against clinician confidence ratings; (6) cultural adaptation studies for deployment in diverse contexts; and (7) randomized controlled trials assessing the system's impact on screening coverage, referral appropriateness, and patient outcomes. Integration with electronic health records and mobile application development would enhance accessibility and practical deployment in community health settings.

Acknowledgment

The authors would like to thank the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology for the funding support provided.

References

Astuti, F. R., & Putri, A.K. (2024). Peran pendidikan inklusif: strategi dan tantangan dalam penghapusan diskriminasi terhadap anak-anak berkebutuhan khusus. *Jurnal Pendidikan Kebutuhan Khusus,* 8(2), 109–119. https://doi.org/10.24036/jpkk.v8i2.926

Beusenberg, M., & Orley, J. (1994*). A user's guide to the Self Reporting Questionnaire (SRQ-20).* World Health Organization.

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule based expert systems: the mycin experiments of the stanford heuristic programming project (the Addison-Wesley series in artificial intelligence)*. Addison-Wesley Longman Publishing Co., Inc.

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., ... & Terry, M. (2019). Human-centered tools for coping with imperfect algorithms during medical decision-making. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* 1-14. https://doi.org/10.1145/3290605.3300234

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981-983. https://doi.org/10.1056/NEJMp1714229

Correll, C. U., Galling, B., Pawar, A., Krivko, A., Bonetto, C., Ruggeri, M., ... & Kane, J. M. (2018). Comparison of early intervention services vs treatment as usual for early-phase psychosis: A systematic review, meta-analysis, and meta-regression. *JAMA Psychiatry*, 75(6), 555-565. https://doi.org/10.1001/jamapsychiatry.2018.0623

Durkin, J. (1994). *Expert systems: Design and development*. Prentice Hall.

Ferdian, D., Hikmat, R., & Anshor, A. (2024). Gambaran deteksi dini masalah kesehatan mental pada siswi. *Jurnal Keperawatan Jiwa,* 12(2), 315-324.

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, 9(2), 137-150. https://doi.org/10.1016/S2215-0366(21)00395-3

Giarratano, J. C., & Riley, G. (1994). *Expert systems: principles and programming*. PWS United States: Publishing Co.

Goldenring, J. M., & Rosen, D. S. (2004). Getting into adolescent heads: An essential update. *Contemporary Pediatrics*, 21(1), 64-90.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, 38(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741

Grist, R., Croker, A., Denne, M., & Stallard, P. (2019). Technology delivered interventions for depression and anxiety in children and adolescents: A systematic review and meta-analysis. *Clinical Child and Family Psychology Review*, 22(2), 147-171. https://doi.org/10.1007/s10567-018-0271-8

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning*, 1321-1330.

Haque, A., Milstein, A., & Fei-Fei, L. (2020). Illuminating the dark spaces of healthcare with ambient intelligence. *Nature,* 585(7824), 193-202. https://doi.org/10.1038/s41586-020-2669-y

Heckerman, D. (1992). The certainty-factor model. In *Encyclopedia of artificial intelligence* (pp. 131-138). Wiley. Los Angeles: University of Southern California

Hollis, C., Falconer, C. J., Martin, J. L., Whittington, C., Stockton, S., Kirkbride, E., & Goodyer, I. (2017). Annual research review: Digital health interventions for children and young people with mental health problems a systematic and meta-review. *Journal of Child Psychology and Psychiatry*, 58(4), 474-503. https://doi.org/10.1111/jcpp.12663

Jackson, P. (1998). *Introduction to expert systems* (3rd ed.). Addison-Wesley.

Keynejad, R. C., Dua, T., Barbui, C., & Thornicroft, G. (2018). WHO Mental Health Gap Action Programme (mhGAP) Intervention Guide: A systematic review of evidence from low and middle-income countries. *Evidence-Based Mental Health*, 21(1), 30-34.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., ... & Rahman, A. (2011). Child and adolescent mental health worldwide: Evidence for action. *The Lancet*, 378(9801), 1515-1525.

Kieling, C., Buchweitz, C., Caye, A., Silvani, J., Ameis, S. H., Brunoni, A. R., ... & Szatmari, P. (2024). Worldwide prevalence and disability from mental disorders across childhood and adolescence: evidence from the global burden of disease study. *JAMA psychiatry*, *81*(4), 347-356.

Kohrt, B. A., Rasmussen, A., Kaiser, B. N., Haroz, E. E., Maharjan, S. M., Mutamba, B. B., ... & Hinton, D. E. (2014). Cultural concepts of distress and psychiatric disorders: Literature review and research recommendations for global mental health epidemiology. *International Journal of Epidemiology*, 43(2), 365-406. https://doi.org/10.1093/ije/dyt227

Lattie, E. G., Stiles-Shields, C., & Graham, A. K. (2022). An overview of and recommendations for more accessible digital mental health services. *Nature Reviews Psychology*, 1(2), 87-100. https://doi.org/10.1038/s44159-021-00003-1

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, 113, 103655. https://doi.org/10.1016/j.jbi.2020.103655

McGorry, P. D., Killackey, E., & Yung, A. (2008). Early intervention in psychosis: Concepts, evidence and future directions. *World Psychiatry*, 7(3), 148-156. https://doi.org/10.1002/j.2051-5545.2008.tb00182.x

Minsih, Taufik, M., & Tadzkiroh, U. (2021). Urgensi pendidikan inklusif dalam membangun efikasi diri guru sekolah dasar. *Jurnal Ilmiah Pendidikan Citra Bakti*, 8(2), 191–204. https://doi.org/10.38048/jipcb.v8i2.352

Negnevitsky, M. (2011). *Artificial intelligence: A guide to intelligent systems* (3rd ed.). Addison-Wesley.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Novrizal, N., & Manaf, S. (2024). The Policy of Inclusive Education in Indonesia. Multicultural Islamic Education Review, 2(1), 39–48. https://doi.org/10.23917/mier.v2i1.4297

Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: A global public-health challenge. *The Lancet*, 369(9569), 1302-1313.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., ... & UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.

Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis. *JAMA Pediatrics*, 175(11), 1142-1150.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sajja, P. S., & Akerkar, R. (2010). Knowledge-based systems for development. In *Advanced Knowledge Based Systems* (Vol. 1, pp. 1-11). TMRF.

Sharma, G., Dhingra, S., & Mishra, S. (2025). Use of artificial intelligence in adolescents' mental health care: Systematic scoping review. *JMIR Mental Health*, 12, e70438. https://doi.org/10.2196/70438

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., ... & Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Journal of Clinical Psychiatry, 71(3), 313-326.*

Shortliffe, E. H. (1976). *Computer-based medical consultations*: MYCIN. Elsevier.

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., ... & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, 27(1), 281-295. https://doi.org/10.1038/s41380-021-01161-7

Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., ... & Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. *British Journal of Psychiatry*, 210(2), 119-124. https://doi.org/10.1192/bjp.bp.116.188078

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793-4813. https://doi.org/10.1109/TNNLS.2020.3027314

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. *Proceedings of Machine Learning Research,* 106, 359-380.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., ... & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20(3), 318-335. https://doi.org/10.1002/wps.20883

Van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis*, 79, 102470. https://doi.org/10.1016/j.media.2022.102470

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13. https://doi.org/10.1145/3313831.3376301

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731.