Dr. R Ashok Kumar, 2026, 14:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

MindCare: An Integrated On-Device Affective Computing, Hybrid Adaptive Screening, and Clinician-in-the-Loop Tele-Consultation Framework for Mental Health Triage

Dr. R Ashok Kumar[1], Satavisa Deka[2], Shuchi Sharma[3], Varun R[4]
Department of Computer Science and Business Systems, B.M.S. College of Engineering, Bengaluru, India

Abstract-. Digital mental health screening, affective text analysis and access to professional support are usu-ally delivered through separate systems. Fixed questionnaires adapt poorly to what an individual writes, while cloud-based conversational tools raise privacy and governance concerns. This paper presents MindCare, a soft-ware framework that combines on-device lexical text analysis, adaptive screening, clinician-facing summaries, specialist recommendation and tele-consultation in one application. The text analysis extracts sentiment, emo-tion and risk indicators from free-text responses and combines them with a 15-question bank covering anxiety, depression, stress and interpersonal difficulties. Question selection follows a hybrid scheme with general, spe-cific and confirmation phases, a safety override and an evidence-based early-stopping rule. The prototype is built with Flutter and Dart, using local storage, Cloud Firestore and Agora video calls. It was checked with 64 unit and widget tests across 13 modules, all of which passed. These results show consistent software be-haviour only; they do not show clinical validity, accuracy or effectiveness. Clinical testing, privacy and security hardening, and validation against recognised instruments are required before clinical use.

Keywords- Digital Mental Health, Adaptive Screening, Sentiment Analysis, Emotion Detection, Risk Assessment, Natural Language Processing (NLP), On-Device Text Analysis, Mental Health Assessment, Clinician-Facing Summaries, Specialist Recommendation, Tele-consultation, Flutter, Dart, Cloud Firestore, Privacy-Preserving Healthcare.

I. INTRODUCTION

Mental health problems, including anxiety, depression, chronic stress and social difficulties, can greatly affect a person's life, including education and employment. Early intervention is key to providing these individuals with the correct support. However, recognising that there is a prob-lem and linking the person to a professional can be diffi-cult. Questionnaires, self-help apps and appointment sys-tems each address part of this.

Short screening tools such as the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder-7 (GAD-7) are commonly used to help organise symptoms [1, 2]. They are standardised to ensure consis-tent administration, but their fixed format may not adapt well to what the individual writes in a free-text response. Chatbots can provide a more flexible interaction, but they raise concerns about how sensitive information is handled, how responses are interpreted, and whether the interaction is intended for screening or diagnosis [3, 4].

Another practical challenge is the gap between screen-ing and access to professional care. A screening tool may generate a score without guiding the user towards a suitable mental health professional. Conversely, a tele-consultation platform may allow users to book appoint-ments without giving the clinician a structured summary of the concerns shared beforehand. Bringing these processes together could create a smoother and more connected ex-perience, linking the initial

© 2026 Dr. R Ashok Kumar. his is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

conversation and preliminary screening with a professional assessment. The MindCare software platform addresses these is-sues by combining rule-based text processing on the de-vice, adaptive question selection, clinical summaries, spe-cialist recommendations, appointment booking and video calling (Fig. 1). It aims to provide initial triage support and to improve communication between users and health professionals. It is not intended to diagnose any mental health condition or to act as a substitute for professional judgement.

The principal contributions of this work are:

- A built-in text-processing system that understands infor-mal language and identifies the emotions, overall senti-ment and possible signs of concern in the user's mes-sages.
- An adaptive screening approach that brings together in-formation from everyday conversation and guided ques-tions to better understand the user's emotional well-being across four key areas.
- A clinician-facing reporting and specialist-matching system that helps mental health professionals review user reports and connect individuals with suitable spe-cialists, while ensuring that clinical assessments and diagnostic information are only accessible to qualified professionals.
- A consultation system that manages appointments, pre-vents scheduling conflicts, prioritises users waiting for calls, and allows them to connect with professionals through video consultations.
- A software evaluation carried out through automated tests of individual functions and user interface compo-nents, together with a clear statement of the system's clinical and security limitations.

The rest of the paper reviews related work, explains how the system is designed and developed, presents the implementation and testing results, and highlights the im-provements and further steps needed before the application can be used in real clinical settings.

II. RELATED WORK

Computerized adaptive mental health screening.
Tools such as the PHQ-9, GAD-7 and Perceived Stress Scale allow a structured assessment of certain symptoms and perceived stress [1, 2, 5]. While standardised tools are useful, it may not always be necessary to ask every possi-ble question of every participant in a digital environment. This can be addressed through computerised adaptive test-ing, where the questions asked depend on the previous an-swers given by the individual. Gibbons et al. demonstrated how this approach can be used for depression assessment [6]. Other digital interventions have also explored conver-sational and branching-based methods [7]. Within Mind-Care, the next set of questions is adapted based on previ-ous user answers. However, unlike validated tools based on item response theory, MindCare currently uses heuris-tic rules to select the next questions. It should therefore be considered a prototype approach rather than a scientifi-cally validated adaptive test.

1. Lexical text analysis.
Lexical approaches use pre-defined sentiment or emotion values for certain words and phrases. Examples include AFINN, a lexical resource in which sentiment scores are assigned to individual words [8], and VADER, which uses both lexical information and rules to account for factors such as negation and intensity [9]. Past work in mental health informatics has also investigated the use of lan-guage patterns for linking to depression and other emo-tional distress [10, 11]. Recent work has also consid-ered machine-learning and transformer-based models for analysing mental-health-related text [3, 12].

MindCare uses a simpler, rule-based methodology. The Dart implementation uses a subset of the AFINN re-source, combined with extra lexicons covering domain-specific terms, emotions, slang and risk terms. This makes scoring more straightforward to understand and allows the analysis to run within the application. There are, however, drawbacks. Rule-based matching of words and phrases handles complex sentences, sarcasm and variation in lan-

guage less well than context-aware language models.

2. Clinician involvement in digital mental health. Reviews of digital mental-health tools emphasise that ap-plications should be evaluated carefully and should not be treated as substitutes for professional care [13, 14]. Au-tomated outputs can be misinterpreted when they are pre-sented as diagnoses rather than as preliminary indicators. Decision-support systems therefore benefit from clear ex-planations of how an output was produced and from work-flows that leave clinical interpretation to qualified profes-sionals [15]. MindCare follows this principle by separat-ing the patient interface from the clinician-facing report summary and by connecting screening to a consultation workflow. The present implementation demonstrates this integration at the software level; it does not establish that the generated summaries can improve clinical decisions or intake efficiency.

III. SYSTEM ARCHITECTURE

As Fig. 2 shows, MindCare follows a layered MVVM and service-oriented architecture implemented in Flutter and Dart. It consists of a role-aware presentation and state management layer, an on-device clinical NLP layer, a hy-brid adaptive screening engine, a clinician-in-the-loop re-porting and tele-consultation layer, and a dual-mode per-sistence layer.

1. Client Presentation and State Management Layer
The frontend is built with Flutter and Dart and uses Provider-managed reactive services (AuthService, ConsultationService, MoodService and CallS ervice) with dedicated ViewModels (ChatViewModel and QuizViewModel). SessionGate resolves the ac-tive session and routes users to role-specific interfaces (UserRole.patient or UserRole.psychologist). The patient environment provides the conversational screening chatbot (ChatScreeningScreen), daily mood check-in and trend analytics (MoodWidgets), guided box-breathing exercises (BreathingScreen), specialist recommendations (RecommendationsScreen), con-sultation messaging (ConsultationChatScreen) and video call rooms (VideoCallScreen). The psychol-ogist workspace (PsychologistDashboardScreen) provides consultation queue management, explainable pa-tient triage reports (PatientReportScreen), interactive calendar scheduling (DoctorCalendar) and analytics dashboards.

2. On-Device Clinical Natural Language Processing Layer
The TextAnalysisPipeline manages the entire linguis-tic analysis process within the client application using a deterministic approach. User responses, whether typed manually, selected through quick-reply chips, or converted from speech to text using VoiceInputService (speech_ to_text), are first checked by AnswerValidator to fil-ter out random keystrokes or meaningless input. The re-sponses are then normalized using TextUtils. Three spe-cialized analyzers then run one after another:

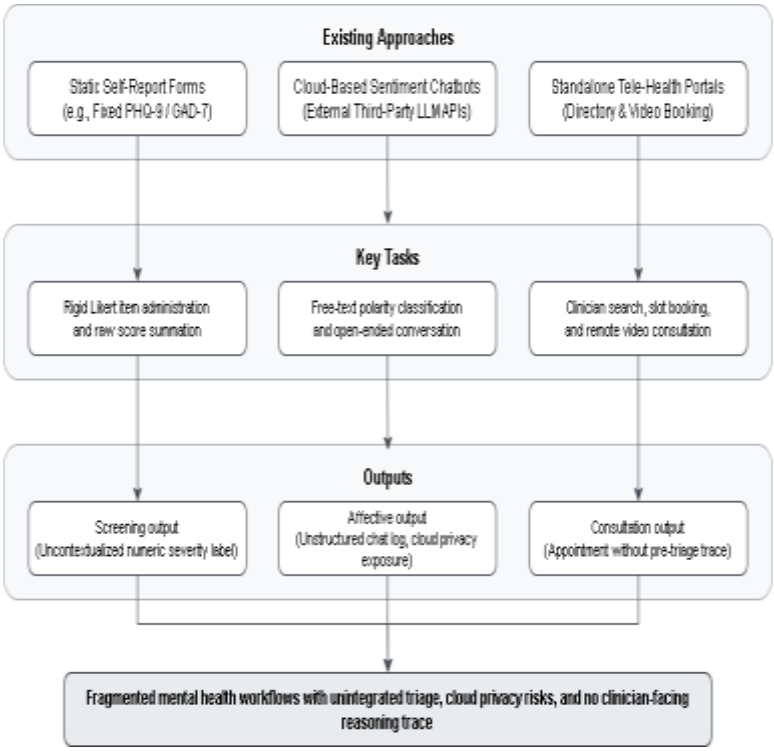


Figure 1. Fragmented approaches to digital mental health screening, affective text analysis, and tele-consultation versus the integrated MindCare workflow.

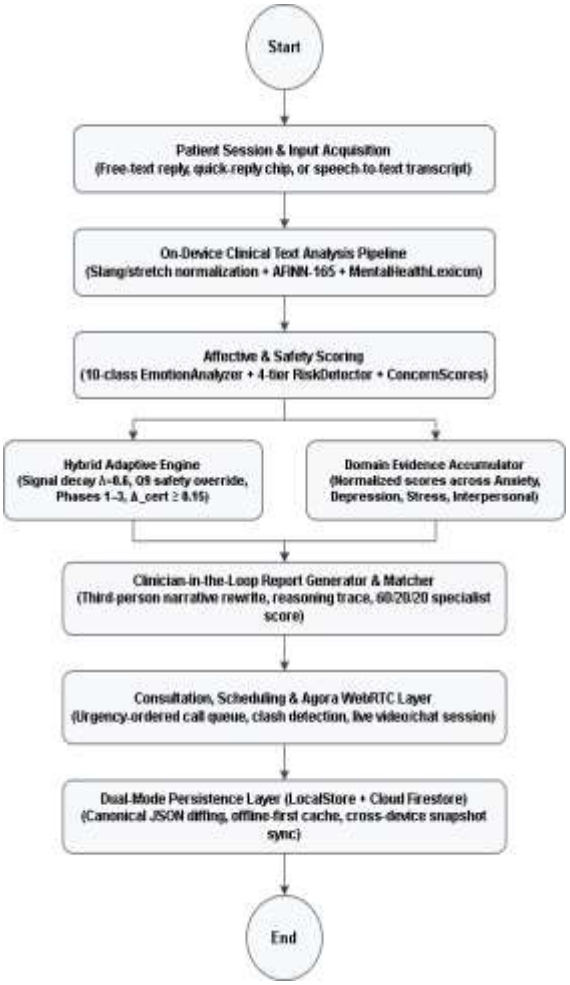


Figure 2. Layered architecture and end-to-end workflow of the MindCare application, illustrating the interaction between on-device clinical NLP, hybrid adaptive screening, clinician-in-the-loop reporting, and dual-mode persistence services.

- SentimentAnalyzer evaluates valence polarity using AfinnLexicon (394 words), MentalH ealthLexicon (132 clinical domain keywords, 29 negators, 31 intensifiers, 26 affirmation phrases and 23 denial phrases), and question-context severity boosting.
- EmotionAnalyzer scores ten affective states (Sadness, Fear/Anxiety, Anger/Frustration, Guilt/Shame, Loneliness, Hopelessness, Ex-haustion, Overwhelm, Joy and Calm) using EmotionLexicon (23 multi-word phrases and 130 single-word hints).
- RiskDetector checks the user's responses for signs of self-harm, suicidal thoughts and emotional distress using 71 predefined multi-word phrases. Based on the findings, the system classifies the re-sponses into four risk levels (Critical, High, Mod-erate and Low). It also accounts for overlapping phrases and negation before calculating an overall risk score.

3. Hybrid Adaptive Screening Engine

The AdaptiveEngine maintains the sum of DomainE vidence scores from four areas of clinical screening (Anxiety, Depression, Stress and Interpersonal/Trauma), exponentially decaying free-text concerns, the risk at the peak of the conversation (peakRisk), summed emotion scores, and a history of why questions were selected (selectionReasons). After a required general opening question (Q1), questions are selected using a hybrid ap-proach. First, safety is prioritised by asking the hopeless-ness question (Q9) if there is a moderate or higher level of risk. Next, the primary concern from the user's written answers is prioritised, but no more than two consecutive questions can be asked from the same area. If neither of these applies, questions are selected using a series of rules based on the evidence gathered across three phases.

4. Clinician-in-the-Loop Reporting, Matching, and Tele-Consultation Layer

After screening is completed, ReportGenerator pro-duces a ScreeningResult that contains normalised scores, severity levels, clinical observations, explanations of the methodology, an emotion summary, and the rea-soning for each question. An enforced perspective filter

$$c \in C(w),\ c \in V$$

$$r=1$$

(narrativeNeedsRewrite / rewriteNarrative) en-sures that all text is written in the third person ("The pa-tient. . . ") and is not visible to anyone other than the psy-chologist. RecommendationService ranks the available psychologists by specialisation, rating and experience rel-ative to the primary and secondary screening domains. ConsultationService and CallService handle asyn-chronous messaging, appointment creation, calendar clash checking, urgency-based prioritisation of video calls, and real-time video calls using agora_rtc_engine (with an HTML5 WebRTC fallback on the web).

5. Dual-Mode Data Persistence and Cloud Synchronization

LocalStore implements an offline-first JSON document store backed by SharedPreferences (browser localS torage on the web and local disk on desktop and mobile) and mirrored to Cloud Firestore (mindcare_ consultations, mindcare_time_blocks, mindcare_ call_requests and mindcare_accounts). Records are serialized into canonical key-sorted JSON strings so that only modified or deleted documents trigger Firestore batch writes. Real-time Firestore snapshot listeners and a 3- second cross-tab polling timer keep patient and psychol- ogist views synchronized across browser tabs and devices degrading gracefully to local-only operation when net-work connectivity is unavailable.

IV. METHODOLOGY

MindCare follows a step-by-step process to understand users' responses. It begins by cleaning and processing the language, then identifies emotions and possible risk indicators. Based on these findings, it selects relevant screening questions, combines the collected information, and uses it to support recommendations with the involve-ment of mental health professionals. Each step passes its output to the next stage (Fig. 3).

1. Conversational Text Normalization and Tokenization

Raw user input Traw frequently contains expressive char-acter repetition (e.g., "yesssss", "sooooo tireddd"), youth abbreviations ("ngl im so stressed fr"), emoji, and coded crisis euphemisms ("unalive", "kms"). TextUtils. normalize converts Traw to lower case, maps 13 affective emoji to canonical words, and repairs stretched tokens not already present in the reference vocabulary V. For a token w decomposed into R character runs {(cr, ℓr)} , candidate reductions retain either one or two characters per stretched run (ℓr ≥ 2) when the number of stretched runs is at most five, selecting the longest candidate recognized in the vo-cabulary:

$$\hat{w} = \arg\max_{c \in C(w),\ c \in V} |c|, \quad C(w) = \prod_{r=1}^{R} \{c_r,\ c_r c_r\}. \qquad (1)$$

If w contains only double-letter runs (ℓr = 2), reduction is restricted to a 17-word stretch-safe whitelist (yes, no, so, hi, ok, etc.), so that legitimate English double-letter words such as "off", "stuff" and "will" remain unaltered. Subsequently, 40 slang and euphemism mappings expand abbreviations (including mapping "unalive" to "kill" and "kms" to "kill myself"), the pattern "X af/asf" is rewrit-
ten as "very X", and 19 English contractions (can't → cannot, don't → do not, etc.) are expanded before punc-tuation removal and whitespace tokenization into

2. Lexicon-Driven Sentiment and Clinical Domain Hint Extraction

For a tokenized message W = [w1, . . . , wM], SentimentA nalyzer computes a compound valence score by scan-ning each token wi against AfinnLexicon. A 3-token lookback window checks whether any preceding token wi− j (1 ≤ j ≤ min(3, i)) belongs to the 29-element nega-
tor set Nneg, while the immediately preceding token wi−1
is checked against the 31-entry intensifier map γ(wi−1) ∈
[0.6, 2.0] (defaulting to 1.0). The token-level valence con-
tribution is

$$V_i = V(W_{i-1}) \cdot \begin{cases} -0.75\,\mathrm{AFINN}(w_i), & \exists j \in \{1,2,3\} : w_{i-j} \in N_{neg}, \\ \mathrm{AFINN}(w_i), & \text{otherwise.} \end{cases} \qquad (2)$$

Because a brief reply such as "yes, constantly" to a distress question has neutral lexical valence in isolation but confirms the presence of a symptom, the analyzer also computes phrase-level affirmation and denial scores
A(T), D(T) ∈ [0, 3]. The normalized compound sentiment score S norm ∈ [−8, 8] is obtained by clamping the adjusted sum:

$$S = \sum_{i=1}^{M} v_i + \begin{cases} +1.5\,A(T), & A(T) > D(T) \wedge v_i \geq -0.5, \\ -1.0\,D(T), & D(T) > A(T) \wedge v_i \leq 0.5, \\ 0, & \text{otherwise,} \end{cases} \qquad (3)$$

$$S_{norm} = \mathrm{clamp}(S_{comp}, -8.0, 8.0). \qquad (4)$$

The base sentiment severity sbase ∈ {0, 1, 2, 3} maps S norm

via piecewise thresholds: sbase = 0 if S norm ≥ 1.5 (positive), 1 if −0.5 ≤ S norm < 1.5 (neutral), 2 if −3.0 ≤ S norm < −0.5 (negative), and 3 if S norm < −3.0 (very negative). Simultaneously, any token match-ing MentalHealthLexicon.domainKeywords accumu-lates a domain hint weight Hd(T) for domain d ∈ D = {Anxiety, Depression, Stress, Interpersonal}.



Figure 3. Five-stage on-device clinical NLP pipeline and hybrid adaptive question selection workflow.

Table 1. Composition of the on-device clinical NLP lexicons and question bank.

Knowledge resource	Entries	Scale / categories	Primary role in pipeline
AFINN-165 subset (AfinnLexicon)	394 words	Valence ∈ [−5, +5]	General affective polarity scoring
Clinical domain lexicon (MentalHealthLexicon)	132 words	Weight ∈ [0.3, 3.0] across 4 domains	Domain-specific symptom keyword hints
Context modifiers (negators / intensifiers)	29 / 31 terms	Flip (−0.75×) / scale ∈ [0.6, 2.0]	Local 3-token negation and preceding intensifier
Affirmation and denial phrases	26 / 23 phrases	Clamped score ∈ [0, 3]	Contextual agreement with distress prompt
Emotion lexicon (words + phrases)	130 + 23 = 153	Weight ∈ [0.8, 2.5] across 10 emotions	Multi-class emotional state extraction
Tiered crisis lexicon (RiskDetector)	17 / 23 / 21 / 10	71 phrases; weights ∈ [1.5, 10.0]	Critical, High, Moderate and Low risk gating
Slang and emoji lexicon (SlangLexicon)	40 + 13 = 53	Canonical word/phrase expansion	Informal youth vernacular and euphemism handling
Adaptive question bank (QuestionBank)	15 questions	Phases 1 (3), 2 (11), 3 (1)	Multi-domain adaptive clinical probing

3. Multi-Class Emotion Recognition and Four-Tier Risk Detection

EmotionAnalyzer evaluates the 10-element emotion set E = {Sadness, Fear, Anger, Shame, Loneliness, Hopeless-ness, Exhaustion, Overwhelm, Joy, Calm}. Multi-word phrases in EmotionLexicon.phrases are matched first;

$$R_{msg} = \text{clamp}\left(w_k + 0.2\sum_{j=1}^{L} w_k + 0.8\,\mathbb{I}(S_{\text{label}} = \text{very negative}) + \text{clamp}\cdot E(\text{hopelessness})/2.5,\ 0,\ 1\right),\ 0.0,\ 10.0\ . \quad (5)$$

if a phrase starting at index i is not negated, its constituent token indices are marked as consumed so that single-word tokens are not double-counted, and its weight wp·γ(wi−1) is added to the target emotion. Unconsumed single tokens wi are then matched against EmotionLexicon.words with asymmetric negation handling: negating a negative emo-tion (e.g., "not sad", "not anxious") contributes zero, whereas negating a positive emotion (e.g., "not happy") contributes 0.6 weγ(wi−1) to Emotion.sadness. The pri-mary emotion e∗ = arg maxe∈E E(e) has normalized inten-sity Iemo = clamp(E(e∗)/4.0, 0.0, 1.0).

RiskDetector scans W against 71 risk phrases acrossfour tiers (Critical: weights 8.5–10.0; High: 6.5–8.5; Moderate: 4.5–5.5; Low: 1.5–2.5). Any phrase match preceded within three tokens by a negator (e.g., "I would never hurt myself", "not suicidal") is discarded as protec-tive. Remaining hits are sorted by descending weight wk, and greedy interval scheduling retains a non-overlapping set of spans K = {k1, k2, . . . , kL} (wk1 ≥ wk2 ≥ . . .), so that nested substrings (such as "cannot take it" inside "cannot take it anymore") are never double-counted. The message-level risk score Rmsg ∈ [0, 10] is

Although the phrase lexicon contains four non-zero tiers, the composite message score is mapped to five output lev-els, with None representing the absence of a qualifying signal: Critical (Rmsg ≥ 9.0), High (7.0 ≤ Rmsg < 9.0), Moderate (4.0 ≤ Rmsg < 7.0), Low (1.5 ≤ Rmsg < 4.0), and None (Rmsg < 1.5). Levels at or above High set needsSafetyResponse=true, surfacing immediate cri-sis helpline resources in the patient interface.

4. Hybrid Signal-Driven Adaptive Question Selection and Domain Evidence Aggregation

For each user message answering question q, TextA nalysisPipeline fuses lexical hints, emotion-to-domain affinities Ae,d ∈ [0.3, 1.0], and question-context weights into a message-level concern vector Cd(T) for each domain d ∈ D:

$$C_d(T) = H_d(T) + 0.6 \quad E(e)A \quad +I(s \quad \geq 2)\beta(s \quad) W \quad , \quad (6)$$

where β(3) = 1.2, β(2) = 0.8, and Wq,d is the pri-mary or secondary weight of question q on domain d (Table 2). The dominant message concern is d∗ = arg maxd∈D Cd(T), provided Cd∗ (T) ≥ 0.8, with confi-dence Cd∗ (T)/ .d∈D Cd(T). The final message severity sfinal ∈ {0, 1, 2, 3} elevates sbase to at least 2 when a neg- ative primary emotion has Iemo ≥ 0.6 or the risk is at least Moderate, and to 3 whenever needs Safety Response is true.

Within AdaptiveEngine, running concern signalsS (t) are updated after turn t using an exponential decay factor λ = 0.6:

$$S_d^{(t)} = \lambda S_d^{(t-1)} + C_d^{(t)}, \qquad \lambda = 0.6. \qquad (7)$$

Meanwhile, the Likert-equivalent response value rt ∈ {0, 1, 2, 3, 4} (derived from sfinal in conversational mode or from direct Likert selection in structured mode) updates the accumulated domain evidence Eraw(d) and the maxi-mum possible evidence

$$E_{\text{raw}}^{(t)}(d) = E_{\text{raw}}^{(t-1)}(d) + r_t W_{q,d},$$

$$E_{\max}^{(t)}(d) = E_{\max}^{(t-1)}(d) + 4.0\, W_{q,d}.$$

Emax(d): The normalized domain score is Enorm(d) = Eraw(d)/Emax(d) ∈ [0, 1], which the prototype maps to four internal evidence bands: High (≥ 0.65), Moderate (0.40–0.64), Mild (0.20–0.39) and Low (< 0.20). Let E(1) ≥ E(2) ≥ E(3) ≥ E(4) denote the sorted normalized domain scores. The inter-domain certainty gap is

$$\Delta_{\text{cert}} = E_{(1)} - E_{(2)}. \qquad (9)$$

nates when $N_{\text{asked}} \geq 12$ (maxQuestions). Screening terminates when Nasked ≥ 12 (maxQuestions) or when Nasked ≥ 7 (minQuestions) and Δcert ≥ 0.15 (certaintyThreshold). While screening is active, selectNextQuestion() applies a three-level priority hi-erarchy:

- Safety override: if the conversation-peak risk is at least Moderate and the hopelessness probe Q9 has not yet been asked, select Q9 immediately.
- Linguistic concern steering: if maxd S (t)≥ 1.0(_signalThresholdd) and the last two consecutivequestions did not already probe that same domain anti-tunnelling constraint), select the next unasked question for that domain across Phases 1, 2 and 3.
- Rule-based phase progression: otherwise, exe-cute Phase 1 broad openers (Q1–Q3), Phase 2 focused exploration (Q4–Q14, prioritizing uncer-tain domains where |E(1) − Enorm(d)| < 0.15 and
- Enorm(d) > 0.10), and Phase 3 cross-cutting confir-mation (Q15).

5. Third-Person Narrative Synthesis, Specialist Matching, and Clash-Free Scheduling

To ensure that reports are ready for clinical review and never address the clinician as if they were the patient, ReportGenerator.narrativeNeedsRewrite scans all observation, methodology and recommenda-tion strings against the case-insensitive regular expression \b(you|your|yours)\b and rewrites any legacy second-person records into third-person clinical prose (“The pa- tient. . . ”).

Given a primary domain d1 and an optional sec-ondary domain d2 (included when Enorm(d2) ≥ 0.20), RecommendationService evaluates each available psy-chologist p ∈ P with specialization set S p, rating Rp ∈ [1, 5] and years of experience Yp. The specialization match factor S spec(p, d1, d2) is

$$S_{\text{spec}} = \begin{cases} 1.0, & d_1 \in S_p \wedge d_2 \in S_p, \\ 0.7, & d_1 \in S_p \wedge d_2 \notin S_p, \\ 0.4, & d_1 \notin S_p \wedge d_2 \in S_p, \\ 0.1, & \text{otherwise.} \end{cases} \qquad (10)$$

With normalized rating Rnorm(p) = (Rp − 1.0)/4.0 and nor-malized experience Xnorm(p) = clamp(Yp/20.0, 0.0, 1.0), the composite specialist match score M(p) ∈ [0, 1] is

$$M(p) = 0.6\, S_{\text{spec}}(p, d_1, d_2) + 0.2\, R_{\text{norm}}(p) + 0.2\, X_{\text{norm}}(p).$$

M(p) = 0.6 S spec(p, d1, d2) + 0.2 Rnorm(p) + 0.2 Xnorm(p). (11)

When a patient requests a consultation or video call, CallService orders the psychologist's pending call queue by descending urgency tier u ∈ {low, medium, high} and then by ascending request timestamp treq (first-come, first-served within each urgency tier). Before scheduling an appointment or video call over the interval [tstart, tend) with tend = tstart + Δt, clashesFor checks all existing scheduled consultations, video calls and personal calendar blocks [cstart, cend) for the clinician using strict half-open interval intersection:

Overlap ⇐⇒ tstart < cend ∧ cstart < tend. (12)

V. IMPLEMENTATION

The MindCare codebase comprises 19,875 physical lines of Dart across 72 source files in lib/ (excluding tests) and 1,381 physical lines of Dart across 14 test files in test/. The automated test suite contains 64 test functions (test and testWidgets) across 13 test mod- ules. Core dependencies declared in pubspec.yaml(targeting Dart SDK ^3.7.0 and verified on Flut-ter 3.47.2 / Dart 3.13.2) include provider 6.1.2 forstate management; firebase_core 3.12.1, firebase_ auth 5.5.1 and cloud_firestore 5.6.5 for authentica-tion and cloud synchronization; shared_preferences 2.5.5 and crypto 3.0.7 for local persistence and SHA-256 credential hashing; speech_to_text 7.5.0 for voice input transcription; agora_rtc_engine 6.6.4 and permission_handler 13.0.2 for real-time audio/video tele-consultation; and fl_chart 0.70.2, animations 2.0.11 and google_fonts 6.2.1 for data visualization and interface motion.

1. On-Device Lexicon and Pipeline Integration

All linguistic resources (AfinnLexicon, MentalH ealthLexicon, EmotionLexicon, SlangLexicon and Vocabulary) are compiled directly into con-stant Dart hash maps and sets within lib/data/ and lib/services/vocabulary.dart. Consequently, TextAnalysisPipeline.analyze() makes no net-work request within the lexical lookup routine; execution latency has not been benchmarked and is therefore not quantified here. ChatViewModel persists the entire se-rialized AdaptiveEngine state (sessionToJson()) to LocalStore after each conversational turn, so that inter-rupted screening sessions can resume across app restarts with the accumulated domain evidence, concern signals and selection rationales retained.

2. Cross-Platform Tele-Consultation andDual-Mode Storage Integration

Video calls use conditional compilation (web_call. dart, web_call_web.dart and web_call_stub.dart)

Table 2. Question bank distribution and cross-domain evidence weights across Phases 1–3.

ID	Phase	Primary domain ($W_{q,p}$)	Secondary domain ($W_{q,s}$)	Clinical assessment purpose
Q1	1 (Broad)	Stress (0.8)	Anxiety (0.5)	Broad opener: gauges general emotional strain
Q2	1 (Broad)	Anxiety (1.0)	Stress (0.3)	Early worry probe: differentiates anxiety from stress
Q3	1 (Broad)	Depression (1.0)	—	Early anhedonia probe: loss of interest or pleasure
Q4	2 (Focused)	Anxiety (1.0)	Stress (0.4)	Somatic restlessness and psychomotor agitation
Q5	2 (Focused)	Anxiety (1.0)	—	Acute panic episodes and sudden fear onset
Q6	2 (Focused)	Depression (1.0)	Stress (0.5)	Persistent fatigue and unrefreshing rest
Q7	2 (Focused)	Stress (1.0)	Anxiety (0.4)	Task rumination, deadlines and cognitive overload
Q8	2 (Focused)	Anxiety (1.0)	Stress (0.3)	Autonomic arousal (racing heart, sweating, trembling)
Q9	2 (Focused)	Depression (1.0)	—	Hopelessness probe: triggered early on risk signals
Q10	2 (Focused)	Stress (1.0)	—	Chronic pressure and inability to recharge
Q11	2 (Focused)	Interpersonal (1.0)	Depression (0.4)	Social withdrawal and isolation from close contacts
Q12	2 (Focused)	Depression (0.8)	Stress (0.5)	Executive dysfunction, indecisiveness and memory
Q13	2 (Focused)	Interpersonal (1.0)	—	Interpersonal trust and emotional safety
Q14	2 (Focused)	Interpersonal (1.0)	Stress (0.3)	Relational conflict impact on emotional wellbeing
Q15	3 (Confirm)	Stress (0.6)	Depression (0.6)	Cross-cutting sleep architecture disruption

alongside agora_rtc_engine, so the application is struc-tured to support native and browser targets (web 1.1.2). When a psychologist schedules a call request c, the de-terministic channel identifier mindcare_<callId> is as-signed, enabling both patient and clinician to join the room starting five minutes before scheduledAt. LocalStore synchronizes four shared collections (consultations, time_blocks, call_requests and accounts) with Cloud Firestore using 8-second connection timeouts. If Firestore is unreachable or unconfigured, the prototype is intended to fall back to local persistence through SharedP references; this behaviour should be verified under each target platform and failure condition.

VI. EVALUATION

1. Experimental Setup

The evaluation covers four areas: (1) conversational text normalization, stretched-word repair and adversarial gib-berish filtering; (2) on-device sentiment, multi-class emo-tion, four-tier risk detection and hybrid adaptive question routing; (3) clinician-in-the-loop perspective enforcement and patient UI safety auditing; and (4) dual-mode storage persistence, calendar clash detection and UI layout respon-siveness across screen geometries. Testing was conducted using the repository's automated flutter_test suite of 64 test functions across 13 test files (Table 3).

2. Results

Text normalization, slang and adversarial input validation.

Across the 15 test cases in slang_test.dart and answer_validator_test.dart, TextUtils. normalize accurately restored stretched affirmative, negative and affective tokens ("YESSS" → "yes", "yesssssssssss" → "yes", "sooooo tireddd" → "so tired", "goooood" → "good") while leaving all eight tested standard English double-letter words ("off", "stuff", "will", "all", "too", "see", "pass", "cross") untouched (Table 4). Youth slang phrases such as "ngl im so stressed fr" and "anxious af" normalized to "i am so stressed" and "very anxious", yielding Emotion.fear and ScreeningDomain.anxiety. Meanwhile, AnswerValidator accepted all stretched natural replies ("yeaaaah", "okkkkk") while rejecting all keyboard-mashed inputs ("aaaaaaa", "asdfghjkl", "lolololol1234").

Sentiment, emotion, risk and hybrid routing verification.As summarized in Table 4, RiskDetector correctly dis-tinguished genuine crisis expressions and coded youth eu-phemisms ("i want to unalive myself", "honestly kms") from negated protective statements ("I would never un-alive myself") and benign academic hyperbole ("ngl this exam is killing me lol"). In end-to-end simulations of 12-turn conversations (text_analysis_pipeline_test. dart), the AdaptiveEngine never repeated a questionID, terminated within [7, 12] questions once Δcert ≥ 0.15, routed interpersonal disclosures after Q1 directly to in-terpersonal items, and preempted standard phase rules with Q9 whenever moderate-or-higher hopelessness sig-nals were detected.

Clinician-in-the-loop perspective enforcement and persistence audit.Static source-code scanning in perspective_test. dart inspected all 10 patient-facing view and widget files (quiz_intro_screen.dart, chat_screening_screen.dart,screening_complete_screen.dart,share_more_screen.dart,recommendations_screen.dart, psychologist_profile_screen. dart, chat_viewmodel.dart, mood_widgets.dart, and all files under lib/views/patient/), confirm-ing 0 instances of strings offering diagnostic reports or self-assessment summaries to the patient (Fig. 4). Concurrently, 100% of generated and legacy-migrated ScreeningResult narratives passed third-person validation ("The patient. . . " with zero matches for \b(you|your|yours)\b), as summarized in Table 5 and illustrated in Fig. 5. In call_service_test.dart and overflow_sweep_test.dart, urgency queue sorting, interval overlap detection (clashesFor) and multi-screen layout rendering passed with zero layout overflow errors (Fig. 6).

Table 3. Repository-level software metrics and automated verification audit

Test module (test/)	Test functions	Physical lines	Primary subsystem verified	Status
slang_test.dart	13	102	Stretched-letter repair, slang/emoji expansion, crisis euphemisms	Passed (13/13)
text_analysis_pipeline_test.dart	13	130	Emotion classes, negation, 4-tier risk gating, hybrid Q9/concern routing	Passed (13/13)
motion_test.dart	10	163	UI transition animations, reduced-motion accessibility compliance	Passed (10/10)
call_service_test.dart	6	110	Urgency call queue ordering, interval clash detection, lifecycle states	Passed (6/6)
local_persistence_test.dart	4	114	LocalStore CRUD, canonical JSON serialization, session recovery	Passed (4/4)
perspective_test.dart	4	134	Third-person clinician narrative rewrite and patient UI AST/string audit	Passed (4/4)
screens_render_test.dart	4	120	Patient and psychologist screen widget tree rendering	Passed (4/4)
calendar_chat_test.dart	3	106	Doctor calendar time blocks, consultation chat message synchronization	Passed (3/3)
answer_validator_test.dart	2	21	Minimum-content and keyboard-mashing rejection rules	Passed (2/2)
local_auth_test.dart	2	59	Role-based authentication, SHA-256 credential hashing, session gate	Passed (2/2)
chat_busy_test.dart	1	32	Concurrent message submission lock during bot typing	Passed (1/1)
overflow_sweep_test.dart	1	264	Multi-viewport layout sweep across all screens (zero pixel overflow)	Passed (1/1)
widget_test.dart	1	18	Root MindCareApp initialization and welcome screen smoke test	Passed (1/1)
Total (13 modules listed)	64	1,373[a]	Full-stack application and on-device NLP verification	100% (64/64)

a Lines in the 13 listed modules. The repository has 1,381 physical test lines across 14 test files; the remaining 8 lines are in a file with no test functions.

Table 4. Evaluation of on-device text normalization, sentiment, emotion, risk and hybrid adaptive routing across benchmark cases.

Category level	Raw input utterance	Normalized text / key tokens	Primary emotion	Risk (R_{level})	Dominant concern / next action
Stretched slang	"ngl im so anxious af rn"	"i am so very anxious right now"	Fear / Anxiety	None (0.0)	Anxiety (s_{final} =
2) Negated affect		"I'm not sad at all"		"i am not sad at all" (NOT_sad)	
None (E(sad) = ∅)	None (0.0)	Neutral (s_{final} = 1)			
Interpersonal	"I feel so lonely and isolated lately"	"i feel so lonely and isolated lately"	Loneliness	Low (3.5)	Interpersonal → probes Q11/Q13
Hopelessness	"I feel hopeless, there is no hope"	"i feel hopeless there is no hope" selects	Hopelessness	Moderate (6.0)	Safety override → Q9
Passive ideation (needsS	"everyone would be better off without me"	"everyone would be better off without me"		—	High (8.0)Safety banner afetyResponse)
Explicit crisis	"I want to kill myself"	"i want to kill myself"	—	Critical (10.0)	Safety banner (s_{final} = 3)
Coded euphemism	"i wanna unalive myself fr"	"i want to kill myself"	—	Critical (10.0)	Safety banner (s_{final} = 3)
Protective negation	"I would never hurt myself, I'm not suicidal"	Negated lookback on "hurt myself", "suicidal"	—	None (0.0)	No false-positive safety trigger
Benign hyperbole	"ngl this exam is killing me lol"	"this exam is killing me"	—	None (0.0)	No false-positive safety trigger

Table 5. Clinician-in-the-loop perspective enforcement, scheduling and UI audit results.

Verification dimension	Audited scope / test mechanism	Metric / criterion	Observed outcome
Patient non-diagnostic isolation	Regex scan across 10 patient-facing Dart source files	Forbidden report/summary string literals	0 violations across all patient views
Clinician narrative perspective	ReportGenerator + legacy ConsultationService migration	2nd-person pronoun match \b(you|your|yours)\b	0 matches; 100% 3rd-person rewrite
Adaptive non-repetition and bounds	12-turn conversational simulation (AdaptiveEngine)	Unique question IDs; $7 \le N_{total} \le 12$	100% unique IDs; valid early stopping
Call queue priority and clash check	CallService.queueFor and clashesFor unit suite	Urgency-first ordering; half-open interval overlap	100% pass across 6 scheduling cases
Multi-screen visual layout sweep	overflow_sweep_test.dart across app screens	Flutter RenderFlex pixel overflow exceptions	0 layout overflows across all screens

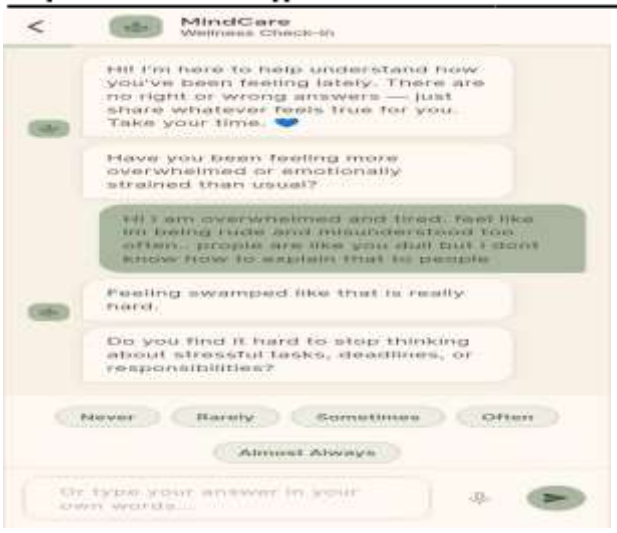


Figure 4. Patient-facing conversational screening interface (ChatScreeningScreen) illustrating empathetic bot acknowl-edgement, quick-reply chips, voice input toggle, and real-time on-device crisis safety resource intervention.



Figure 5. Psychologist-facing patient triage report (PatientReportScreen) displaying normalized four-domain evidence bars, severity badges, third-

person clinical observa-tions, extracted emotion/risk indicators, and the chronological question-by-question reasoning trace.

VII. DISCUSSION AND LIMITATIONS

The prototype brings several functions into one work-flow: local lexical analysis, adaptive question selec-tion, clinician-facing reporting, specialist recommenda-tion, scheduling and video consultation. The automated tests provide evidence that selected software behaviours operate as intended under the tested conditions. In particular, the tests cover text normalization, risk-rule handling, question routing, narrative perspective checks, appointment-conflict detection, persistence operations and interface rendering.

These findings should be interpreted as software-verification results. They do not demonstrate that the lex-ical scores correspond to a person's clinical state, that the adaptive question sequence is psychometrically reliable, or that the system can identify risk with a known sensi-tivity or specificity. The current domain weights, decay factor, stopping threshold and specialist-matching weights are manually selected engineering parameters. They re-quire empirical calibration and validation before they can be used to support clinical decisions.

The current approach also has limitations in language understanding. A finite lexicon and a three-token negation window cannot reliably resolve all forms of sarcasm, am-biguity, indirect expression or long-distance negation. The slang map covers a limited set of informal expressions and does not provide comprehensive support for code-mixed or regional Indian languages. A phrase-based risk detector may therefore miss expressions that are not represented in its lexicon and may misinterpret expressions whose mean-ing depends on context. A positive or negative lexical match must not be treated as a clinical risk assessment on its own.

Privacy claims also require careful qualification. The lexical analysis is implemented in the client application, but MindCare includes account management, Firestore synchronization, speech-to-text functionality and third-party video communication. The handling of data by each component depends on platform behaviour, configuration, access controls and service-provider terms. The proto-type's use of demonstration Firestore rules and a static Agora development App ID is not suitable for real patient data. Before deployment, the system would require strict role-based access control, secure credential and token management, appropriate encryption, data-minimisation policies, retention and deletion procedures, and a docu-mented threat model.

Finally, the evaluation is limited to automated tests. It does not include participant recruitment, clinician review, comparison with validated instruments, usability testing with intended users, or prospective monitoring of safety outcomes. The manuscript therefore makes no claim of diagnostic accuracy, treatment benefit or clinical effective-ness. Any future study involving people or sensitive health information must obtain the applicable ethics approval and informed consent, and must include a predefined clinical safety and escalation protocol.

VIII. CONCLUSION AND FUTURE WORK

This paper has presented MindCare, a software frame-work that integrates on-device lexical text analysis, hy-brid adaptive screening, clinician-facing reporting, spe-cialist recommendation, appointment scheduling and tele-consultation. The implementation combines rule-based processing of informal text with structured question re-sponses across four screening domains. Its current evalua-tion consists of 64 automated unit and widget tests across 13 test modules, all of which are reported as passing.

The contribution is an integrated software prototype rather than a clinically validated screening instrument. The next stage of work should prioritise independent veri-fication of the implementation, security hardening of cloud and communication services, and a carefully governed evaluation with intended users and qualified clinicians. Such an evaluation should compare the system with estab-lished screening instruments, assess agreement with clin-ician assessments, measure usability and

respondent bur-den, and examine false-positive and false-negative safety cases. Multilingual support and longitudinal analysis may

be considered after the core safety, privacy and validity re-quirements have been addressed.

The authors thank the Department of Computer Science and Business Systems, B.M.S. College of Engineering, Bengaluru, for their academic support, infrastructure and guidance through-out this research.

REFERENCES

1. " K. Kroenke, R.L. Spitzer, J.B.W. Williams, The PHQ-9: Validity of a brief depression severity mea-sure. J. Gen. Intern. Med. 16(9), 606–613 (2001). https://doi.org/10.1046/j.1525-1497. 2001.016009606.x
2. R.L. Spitzer, K. Kroenke, J.B.W. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch. Intern. Med. 166(10), 1092–1097 (2006). https://doi.org/10.1001/ archinte.166.10.1092
3. A.N. Vaidyam, H. Wisniewski, J.D. Halamka, M.S. Kashavan, J.B. Torous, Chatbots and conversational agents in mental health: A review of the psychiatric landscape. Can. J. Psychiatry 64(7), 456–464 (2019). https://doi.org/10.1177/0706743719828977
4. K. Kretzschmar, H. Tyroll, G. Pavarini, A. Manzini,I. Singh, Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chat-bots) in mental health support. Biomed. Inform. Insights 11, 1–9 (2019). https://doi.org/10. 1177/1178222619829083
5. S. Cohen, T. Kamarck, R. Mermelstein, A global measure of perceived stress. J. Health Soc. Be-hav. 24(4), 385–396 (1983). https://doi.org/ 10.2307/2136404
6. R.D. Gibbons et al., Development of a computerized adaptive test for depression. Arch. Gen. Psychiatry 69(11), 1104–1112 (2012). https://doi.org/10. 1001/archgenpsychiatry.2012.14
7. K.K. Fitzpatrick, A. Darcy, M. Vierhile, Deliver-ing cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A ran-domized controlled trial. JMIR Ment. Health 4(2), e19 (2017). https://doi.org/10.2196/mental. 7785
8. F.Å. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, in Proc. ESWC Workshop Making Sense of Microposts (#MSM2011), CEUR-WS 718 (2011), pp. 93–98
9. C.J. Hutto, E. Gilbert, VADER: A parsimonious rule-based model for sentiment analysis of social media text, in Proc. 8th Int. AAAI Conf. Weblogs and Social Media (ICWSM) (2014), pp. 216–225
10. M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, in Proc. 7th Int. AAAI Conf. Weblogs and Social Media (ICWSM) (2013), pp. 128–137
11. R.A. Calvo, D.N. Milne, M.S. Hussain, H. Chris-tensen, Natural language processing in mental health applications using non-clinical texts. Nat. Lang. Eng. 23(5), 649–685 (2017). https://doi.org/ 10.1017/S1351324916000383
12. T. Zhang, A.M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental ill-ness detection: A narrative review. npj Digit. Med. 5, 46 (2022). https://doi.org/10.1038/ s41746-022-00589-7
13. J. Torous et al., The growing field of digital psychi-atry: Current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychia-try 20(3), 318–335 (2021). https://doi.org/10. 1002/wps.20883
14. E.G. Lattie, E.C. Adkins, N. Winquist, C. Stiles-Shields, Q.E. Wafford, A.K. Graham, Digital mental health interventions for depression, anxiety, and en-hancement of psychological well-being among col-lege students: Systematic review. J. Med. Internet Res. 21(7), e12869 (2019). https://doi.org/10. 2196/12869
15. A. Holzinger, C. Biemann, C.S. Pattichis, D.B. Kell, What do we need to build explainable AI systems for the medical domain? arXiv:1712.09923 (2017). https://arxiv.org/abs/1712.09923