Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Xaant Resting Autonomic Signature (X-RAS): prospective clinical validation of an autonomic biomarker for anxiety and depressive disorders

Dharmendra Singh
Batra Hospital and Research Center
Cijo Alex
Believers Church Medical College Hospital
Philip Matthew
Believers Church Medical College Hospital
Saritha Susan Vargese
Believers Church Medical College Hospital
Priyadarshini Arambam
Clicebo
Natwar Sharma
TerraBlue XT
Priya Ranjan Pattanaik
TerraBlue XT
Rajlakshmi Borthakur
raji@teblux.com

TerraBlue XT

Research Article

Keywords: artificial intelligence, autonomic nervous system, heart rate variability, electrodermal activity, anxiety disorders, depression, diagnostic biomarkers, resting-state monitoring

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11744905/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. NS, PRP, and RB are employees of TerraBlue XT. The project was funded by TerraBlue XT.

The Xaant Resting Autonomic Signature (X-RAS): prospective clinical validation of an autonomic biomarker for anxiety and depressive disorders

Authors: Dharmendra Singh[1], Cijo Alex[2], Philip Matthew[3], Saritha Susan Varges[4], Priyadarshini Arambam[5], Natwar Sharma[6], Priya Ranjan Pattanaik[7], Rajlakshmi Borthakur[8]*

[1] Department of Psychiatry, Batra Hospital and Research Center, New Delhi, India

[2] Believers Church Medical College Hospital, Thiruvalla, Kerala, India

[3] Department of Critical Care, Believers Church Medical College Hospital, Thiruvalla, Kerala, India

[4] Department of Community Medicine, Believers Church Medical College Hospital, Thiruvalla, Kerala, India

[5] Clicebo, New Delhi, India

[6] TerraBlue XT, Bengaluru, India

[7] TerraBlue XT, Bengaluru, India

[8] TerraBlue XT, Hyderabad, India

Corresponding author: raji@teblux.com

Abstract

Background: Anxiety and depressive disorders affect over 500 million people globally, with a treatment gap exceeding 75% in low- and middle-income countries. Current diagnostic frameworks rely on clinical interview and self-report, with test–retest reliability of 0.20 to 0.67

for these disorders in DSM-5 field trials. Autonomic nervous system dysregulation is a well-established biological correlate of anxiety and depression. We report the first, to our knowledge, prospective clinical validation of an autonomic biomarker-based artificial intelligence system against a blinded psychiatrist reference standard in routine clinical practice.

Methods: We conducted a prospective, two-site diagnostic accuracy study at tertiary hospitals in Delhi and Kerala, India (CTRI/2023/07/054984). The Xaant Resting Autonomic Signature (X-RAS) platform, a three-layer LSTM with self-attention previously developed on an independent cohort of 2,349 participants and frozen prior to the trial, analysed electrodermal activity and blood volume pulse from 15-minute resting-state recordings and produced a four-class diagnostic classification. Of 180 participants enrolled, 170 completed the protocol. Out of these, 114 clinical participants with DSM-5 depression, anxiety, or comorbid disorders, and 56 reference group participants recruited from a long-term meditation practice tradition were considered. The reference standard was the DSM-5 diagnosis established by the site psychiatrist, who was blinded to the platform output. The primary outcome was Cohen's kappa for AI-psychiatrist diagnostic agreement within the clinical cohort.

Results: Within the clinical cohort, the platform achieved 93.0% agreement with clinical diagnosis ($\kappa = 0.884$, 95% CI 0.789 to 0.956), with 100% sensitivity for generalised anxiety disorder (GAD) detection in the GAD subgroup (n = 15, 95% CI 78.2 to 100). All eight classification disagreements fell along the depression–comorbid boundary, and the AI classification agreed with the patient's own self-categorisation in five of the eight. Performance was consistent across the two sites (Delhi 94.6%, Kerala 91.4%, $p = 0.527$) and across age, sex, and education subgroups. Classical autonomic markers separated the clinical and reference groups with moderate to large effect sizes ($d = 0.53$ to 0.84). Two of the physiological constructs

of the development-cohort, namely Hypoactivity and Experience, were reproduced at the group level in independently diagnosed patients. A per-minute slope analysis of the tonic EDA signal showed a transdiagnostic reversal, where the baseline settled downward during rest in the reference group and drifted upward across all three clinical subgroups (Kruskal–Wallis $p = 3.0 \times 10^{-22}$).

Conclusions: Resting-state autonomic AI analysis provides objective diagnostic support for anxiety and depressive disorders, with Cohen's $\kappa = 0.884$ (95% CI 0.789 to 0.956) against psychiatrist-assigned DSM-5 diagnoses in routine outpatient practice. Direct comparison with published inter-rater or test-retest reliability estimates is not appropriate given differences in methodology and reference standards. The platform's low cost, portability, and suitability for non-specialist operators support scalability in resource-limited settings. External validation in primary care, community, telehealth, and non-Indian populations is the next step for the programme.

Trial registration: Clinical Trials Registry of India, CTRI/2023/07/054984, registered July 2023.

Keywords: artificial intelligence, autonomic nervous system, heart rate variability, electrodermal activity, anxiety disorders, depression, diagnostic biomarkers, resting-state monitoring

Background

Neuropsychiatric disorders account for 14% of the global disease burden and affect nearly one billion people worldwide [1]. Depression alone affects 322 million (4.4% of the global population) and anxiety disorders affect 264 million, which makes them the two most prevalent psychiatric conditions [2]. Together they account for the majority of disability-adjusted life years

lost to mental illness, exceeding the combined burden of many communicable diseases [3]. The health burden extends beyond the psychiatric presentation. Depression and anxiety disorders are associated with elevated risk of cardiovascular disease, metabolic syndrome, premature all-cause mortality, and impaired recovery from medical illness, all mediated in part through chronic autonomic dysregulation. The treatment gap is stark: 76 to 85% of affected individuals in low- and middle-income countries do not receive any treatment [4].

The crisis is particularly acute in India. The Global Burden of Disease Study estimated that 197.3 million Indians live with mental disorders, which is approximately one in seven citizens [19]. The treatment gap in India remains substantial, at 83% [5]. With only 0.3 psychiatrists per 100,000 population, compared with the WHO-recommended minimum of 1.0 [6], the shortage of mental health professionals poses a major barrier to expanding access to conventional psychiatric care.

Psychiatric diagnosis relies primarily on clinical interviews and self-reported symptoms, both of which have recognised limitations in reliability. DSM-5 field trials have reported test-retest reliability coefficients ranging from 0.20 to 0.67 for common mood and anxiety disorders [7], while self-report instruments show considerable overlap between depressive and anxiety symptoms [8]. The Patient Health Questionnaire-9 (PHQ-9) and Generalised Anxiety Disorder-7 (GAD-7) are widely used screening instruments, with scores of ≥10 commonly indicating clinically significant symptoms. However, their interpretation may be influenced by cultural

differences in symptom expression and reporting. In South Asian populations, psychological distress may be expressed predominantly through somatic symptoms, potentially affecting questionnaire responses and threshold-based classification. These limitations highlight the need for complementary assessment approaches that are less dependent on subjective symptom reporting.

The need for technological augmentation of psychiatric diagnosis has been discussed by major international bodies for over a decade. The World Health Organization's Comprehensive Mental Health Action Plan (2013 to 2030) identifies scalable screening tools as central to closing the global treatment gap, particularly in low- and middle-income countries where workforce shortages preclude conventional approaches[23]. The Lancet Commission on Global Mental Health and Sustainable Development placed objective screening and digital health platforms among its core recommendations[3]. The United States National Institute of Mental Health has prioritised biomarker development and precision psychiatry through its Research Domain Criteria initiative, calling for measurement frameworks grounded in observable biology rather than symptom clusters alone[24]. A 2026 Lancet Psychiatry Commission on multidisciplinary research priorities for artificial intelligence in mental health articulated the need for rigorous prospective validation of AI tools before clinical deployment[25]. Professional bodies including the Royal College of Psychiatrists have issued position statements on the integration of digital monitoring technology into clinical practice[26]. Despite this convergent mandate, routine diagnostic workflow for anxiety and depressive disorders has changed little since DSM-5 was

introduced in 2013, with assessment continuing to rely on clinical interview and self-report questionnaires that demonstrate known reliability limits[7,8].

Among candidate objective measurement frameworks, the autonomic nervous system presents a particularly compelling target for psychiatric assessment. Anxiety and depression produce measurable, persistent alterations in autonomic function that outlast the acute episode. Anxiety disorders are characterised by sympathetic hyperactivation and parasympathetic withdrawal, which manifest as elevated skin conductance, reduced heart rate variability (HRV), and altered cardiac pre-ejection periods [11,12]. Depression presents a more complex autonomic profile, with predominant vagal withdrawal reflected in reduced RMSSD and high-frequency HRV power, often accompanied by blunted sympathetic reactivity [13].

HRV, which measures beat-to-beat variations in cardiac rhythm, provides a non-invasive indicator of autonomic regulation. Meta-analyses involving over 20,000 individuals have reported reduced HRV in psychiatric populations compared with healthy controls, with effect sizes ranging from moderate in depression ($d = 0.45$) to large in generalised anxiety disorder ($d = 0.75$) [14]. These alterations may persist beyond acute symptomatic episodes and partially normalise following successful treatment, suggesting both trait-related and state-dependent components of autonomic dysregulation [15,16]. Unlike stress-induction protocols that assess autonomic reactivity, resting-state recordings characterise baseline autonomic function and may provide information about the physiological correlates of chronic psychological distress [15].

The evidence base for autonomic assessment in psychiatric disorders is extensive. Several decades of research have produced substantial literature on HRV in mood and anxiety disorders, alongside parallel investigations of electrodermal activity (EDA), pupillometry, and respiratory

sinus arrhythmia. Recent umbrella reviews synthesising multiple meta-analyses involving large participant populations have identified consistent associations between autonomic dysregulation and anxiety and depressive disorders [9,12–14]. The scale of this research is reflected in Google Scholar, which indexes more than 120,000 records relating to HRV and mental health and more than 47,000 relating to EDA and mental health, at the time of writing. Yet this substantial body of scientific knowledge has not translated into routine clinical screening, highlighting a persistent gap between physiological research and clinical implementation.

This gap also reflects both the technical complexity of autonomic signal analysis and the practical constraints of clinical implementation. Autonomic signals are inherently non-stationary, exhibiting continuous fluctuations that require high-quality acquisition, appropriate sampling frequencies, and analysis over sufficiently long observation windows. High-resolution recordings, including sampling at 250 Hz or higher for selected EDA and blood volume pulse (BVP) analyses, can support the extraction of detailed temporal and physiological features. Clinically relevant information is often embedded in variability indices, frequency-domain decompositions, and relationships across multiple physiological channels rather than in patterns readily identifiable through visual inspection. Extracting and integrating these features requires specialised signal-processing methods and substantial analytical expertise. Consequently, comprehensive autonomic assessment has largely remained within specialist autonomic laboratories and research settings, where dedicated equipment, trained personnel, and time-intensive interpretation are available. These requirements may have contributed to the limited development of scalable, standardised protocols suitable for routine psychiatric screening. To our knowledge, no published platform has so far combined multi-channel autonomic acquisition, machine-learning distillation of the signal into interpretable classifications, and prospective

clinical validation in routine care against psychiatrist-assigned reference diagnoses at a scale sufficient to establish real-world diagnostic utility.

Access to psychiatric diagnosis in India is markedly unequal across socioeconomic, educational, and geographic groups, and an objective, low-cost screening tool could help reduce these disparities. The Xaant Resting Autonomic Signature (X-RAS) was built to close this translation gap. The platform integrates the electrodermal and the BVP-derived cardiac photoplethysmographic signals through a single frozen model, pre-developed on 2,349 Indian participants, and operates on a 15-minute recording that can be administered by non-specialist staff. An LSTM architecture with integrated self-attention was designed specifically for the signal properties described above. It learns to weight temporal segments by diagnostic relevance, aggregates cross-channel features at scales that the unaided eye cannot resolve, and produces a single interpretable four-class output from what would otherwise be a comprehensive volume of raw physiological data. Each sensing modality is grounded in decades of established autonomic physiology research. The engineering contribution of the present work lies in integrating these signals, applying machine learning to extract clinically relevant patterns, and translating the resulting framework into a standardised assessment protocol suitable for use outside controlled laboratory settings.

This paper reports the prospective clinical validation of the integrated X-RAS platform, with the aim of evaluating whether a portable, resting-state autonomic AI tool can provide clinically meaningful diagnostic information to support psychiatric assessment. The platform is intended as an adjunct to, rather than a replacement for, clinician-led diagnosis and decision-making. Our approach differs from previous work in three key respects: (1) brief resting-state recordings requiring no active patient participation beyond remaining at rest, (2) the use of accessible

physiological sensors rather than specialised neuroimaging equipment, and (3) prospective validation in naturalistic clinical settings involving medicated patients and the diagnostic complexity encountered in routine psychiatric practice.

Results

Participant characteristics and demographics

Across the two study centres, we enrolled 180 participants, of whom 170 (94.4%) completed the protocol and provided suitable data. The final cohort comprised 114 participants with a DSM-5 psychiatric diagnosis (67.1%), who together constitute the clinical cohort, and 56 participants in the reference group (32.9%) recruited from the Heartfulness Institute (Fig. 1).

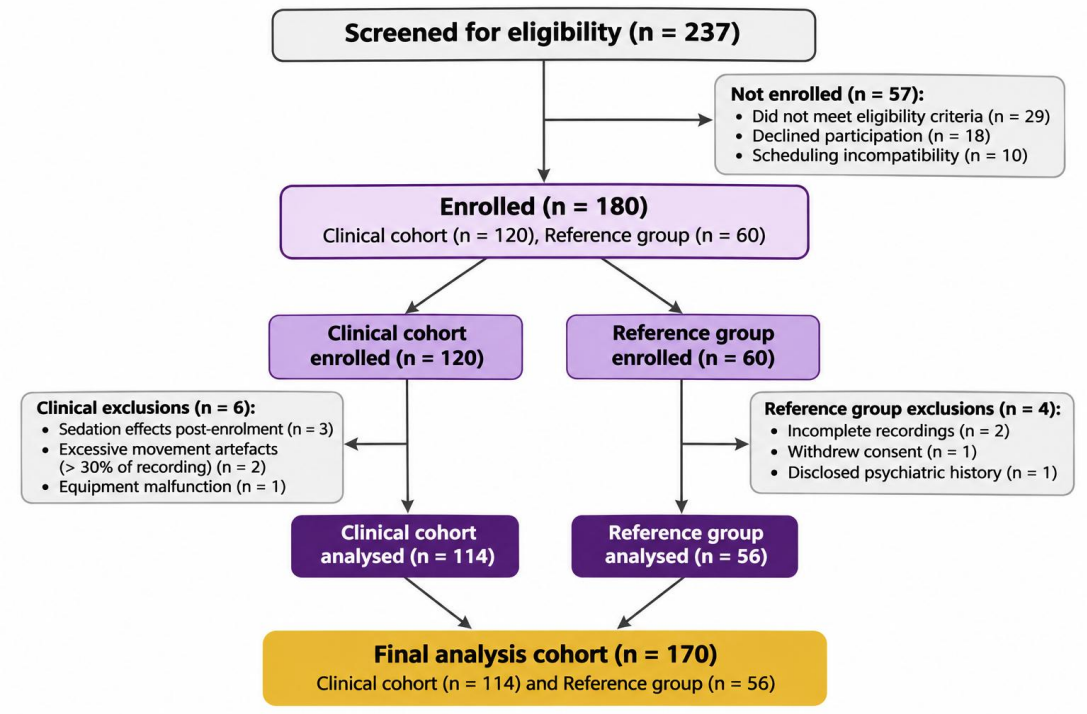


Fig. 1 | STARD flow diagram of participant recruitment and selection. A total of 237 individuals were screened, 180 enrolled, and 170 included in the final analysis (114 autonomically dysregulated clinical participants, 56 autonomically regulated reference participants).

Age averaged 37.8 ± 10.9 years in clinical participants and 36.4 ± 11.2 years in the reference group ($t(168) = 0.77$, $p = 0.443$, $d = 0.13$). Sex distribution differed between groups, with a higher proportion of female participants in the reference group (67.9%) than in the clinical cohort (45.6%, $p = 0.006$). The 35 to 49 age band was overrepresented in the clinical group (36.8% versus 14.3%, $p = 0.002$). Current smoking was more prevalent in the clinical group than in the reference group (13 versus 0 participants; $p = 0.009$), whereas alcohol use did not differ significantly between groups ($p = 0.161$). Given the established influence of smoking and sex on autonomic function, both factors are considered as potential confounders in the Limitations section.

Medical comorbidities occurred at similar rates in the two groups (27.2% clinical versus 21.4% reference, $p = 0.416$). The most common were hypertension (10 clinical, 2 reference), diabetes (7 clinical, 5 reference), and thyroid disorders (9 clinical, 3 reference), all managed with stable medications. None of the individual comorbidity categories reached statistical difference between groups.

Within the clinical group, 52 participants (45.6%) met criteria for depression, 15 (13.2%) for anxiety, and 47 (41.2%) for comorbid anxiety and depression. The diagnostic distribution differed between sites ($\chi^2(2) = 40.52$, $p < 0.001$). The Delhi site saw predominantly depression presentations (73.2% versus 19.0% in Kerala), while the Kerala site saw predominantly comorbid presentations (69.0% versus 12.5% in Delhi). This distribution reflects the presenting pattern at each referral centre rather than any diagnostic drift between sites, as discussed in the Inter-site reliability section below.

The pharmacological profile of the clinical group was heterogeneous, with the full distribution of medication classes detailed in Supplementary Table S1.

The physiological recording and clinical assessment were completed without adverse events.

Primary outcome: diagnostic agreement between AI and clinical assessment

The AI platform reached Cohen's kappa of 0.884 (95% CI 0.789 to 0.956, $p < 0.001$) with clinical psychiatric diagnosis. Overall diagnostic accuracy reached 93.0% (106 of 114 correct classifications) with a weighted F1-score of 0.930 (Table 1). Every recording that passed quality control produced a determinate four-category classification. A prespecified sensitivity analysis assessed the robustness of the primary agreement estimate to recruitment-site differences. Bootstrap resampling was performed separately within each site to preserve the original site composition. The site-stratified bootstrap yielded a 95% confidence interval for Cohen's κ of 0.798 to 0.956, similar to the primary analysis (95% CI 0.789 to 0.956).

Table 1. Diagnostic performance metrics of the AI platform.

Category	n	Sensitivity	Specificity	PPV	NPV	F1
Depression	52	90.4%	95.2%	94.0%	92.2%	0.922
Anxiety	15	100%	100%	100%	100%	1.000
Comorbid	47	93.6%	92.5%	89.8%	95.4%	0.917
Overall (macro)	114	94.7%	95.9%	94.6%	95.9%	0.930

Overall: accuracy = 93.0% (106/114); Cohen's κ = 0.884 (95% CI 0.789 to 0.956); weighted F1 = 0.930; macro-averaged PPV = 94.6%, NPV = 95.9%.

Clinical diagnosis served as the reference standard. PPV, positive predictive value; NPV, negative predictive value. Per-category metrics computed from the confusion matrix (Table 2). Overall row: accuracy, weighted F1, and macro-averaged PPV and NPV.

Performance varied across diagnostic categories but remained clinically useful in all three (Fig. 2). Detection of generalised anxiety disorder (GAD), the sole anxiety diagnosis assessed in this trial, achieved perfect sensitivity (15 of 15 cases correctly identified) with 100% specificity and a positive predictive value of 100%, meaning that every patient with a primary GAD diagnosis in the study was identified by the platform. Depression detection reached 90.4% sensitivity and

95.2% specificity. The comorbid category (GAD co-occurring with depression), which is the most diagnostically difficult in clinical practice, achieved 93.6% sensitivity and 92.5% specificity. This performance was obtained in a fully medicated cohort, which reflects the patients clinicians actually encounter in routine outpatient care rather than treatment-naive research populations.

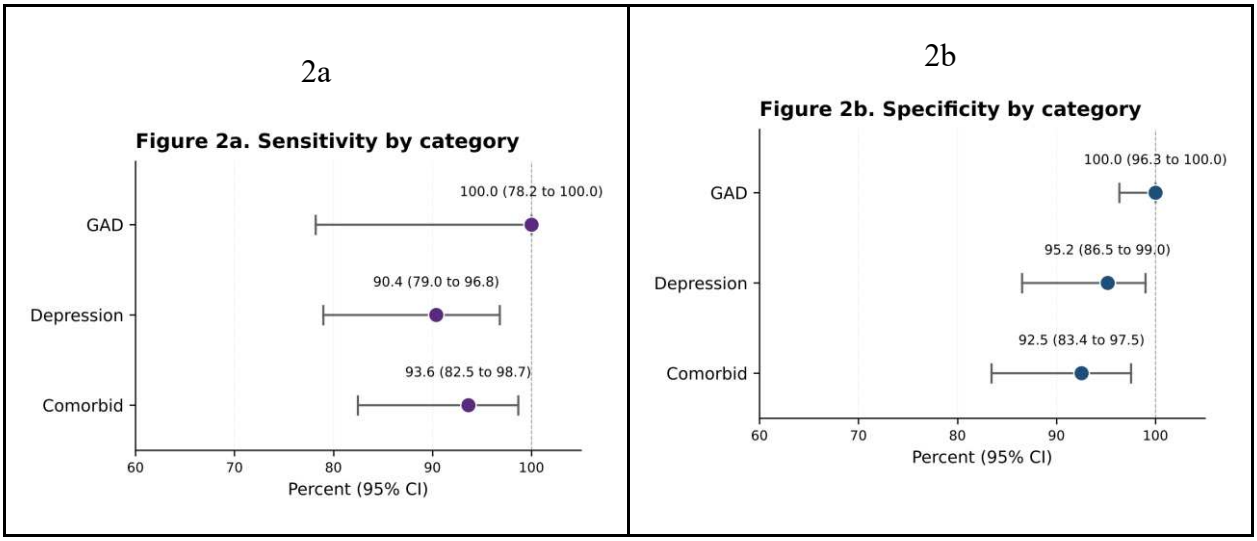


Fig. 2 | Forest plot of sensitivity and specificity with 95% confidence intervals by diagnostic category. Point estimates with exact binomial 95% confidence intervals. Detection of generalised anxiety disorder (GAD, the sole anxiety diagnosis assessed in this trial) achieves perfect sensitivity, with wider confidence intervals reflecting the smaller subgroup size (n = 15).

Classification patterns and misclassification analysis

The confusion matrix (Table 2, Fig. 3) shows strong concordance between AI and clinical diagnoses across all three categories. The McNemar–Bowker test for symmetry of disagreements returned $\chi^2 = 0.50$, $p = 0.480$, consistent with no systematic directional bias between AI and clinical assessment.

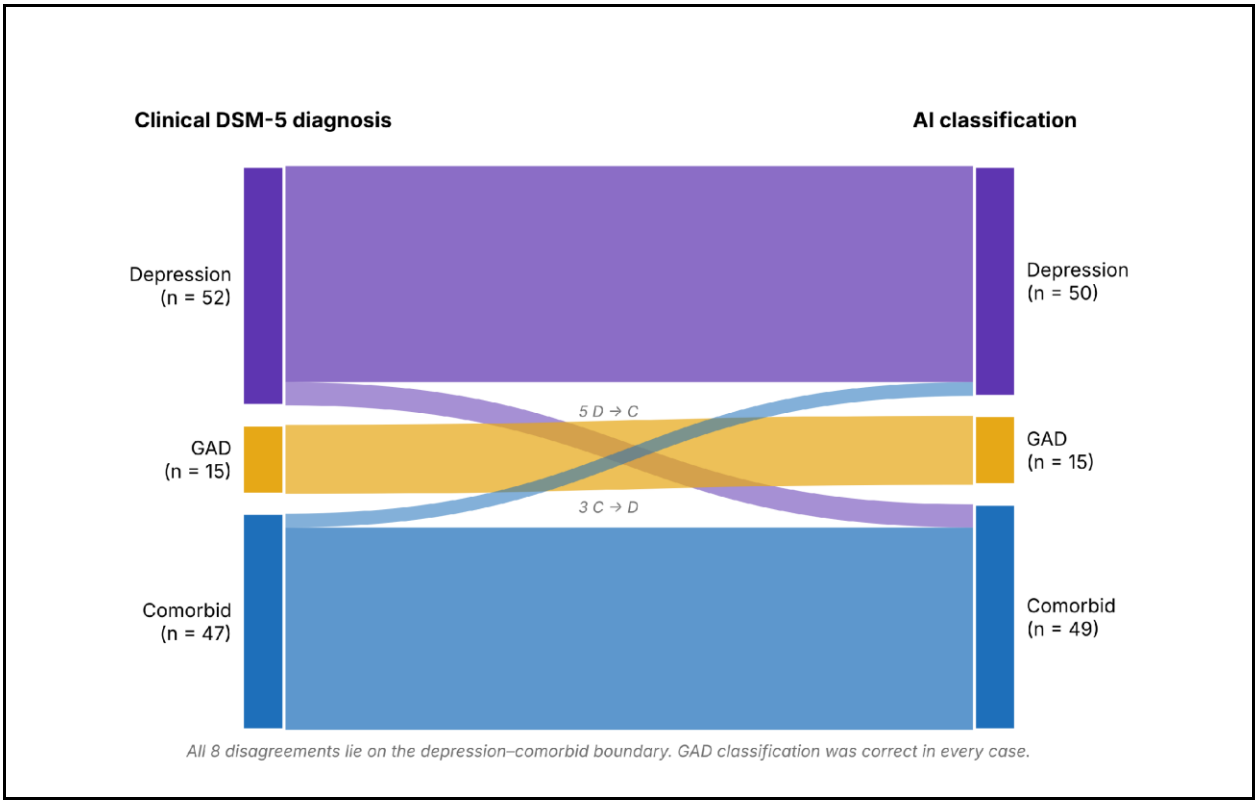


Fig. 3 | Sankey flow diagram of classification agreement between the clinical DSM-5 diagnosis (left) and the AI classification (right). Ribbon width is proportional to the number of participants. Correct classifications flow horizontally along the three diagonal ribbons. The two crossing ribbons show the eight misclassifications: five participants with a clinical depression diagnosis classified by the AI as comorbid, and three participants with a clinical comorbid diagnosis classified by the AI as depression. All eight disagreements lie on the depression-comorbid boundary; the GAD category was correctly classified in every case.

Table 2. Cross-classification of AI and clinical diagnoses.

Clinical diagnosis	AI depression	AI anxiety	AI comorbid	Total
Depression	47	0	5	52
Anxiety	0	15	0	15
Comorbid	3	0	44	47
Total	50	15	49	114

Rows: clinical (reference) diagnosis. Columns: AI classification. Diagonal cells indicate agreement; off-diagonal cells show misclassifications.

Case-level analysis of the eight classification disagreements revealed a consistent pattern, with all misclassifications occurring between the depression and comorbid categories and none involving the GAD category. Five patients diagnosed with depression were classified by X-RAS as having comorbid depression and anxiety (three in Delhi and two in Kerala), while three

patients diagnosed with comorbid conditions were classified as having depression alone (all in Kerala). The disagreements were distributed across the two sites (three in Delhi and five in Kerala), without an apparent concentration at either site after considering differences in case mix.

A further observation emerged from comparing the misclassified cases with patient self-reports. In five of the eight disagreements, the AI classification aligned with the patient's own self-categorisation of their primary condition, and in three the clinical diagnosis aligned with the patient's self-categorisation.

Secondary analysis: four-category classification across the full cohort

As a secondary analysis, the evaluation was extended to all 170 participants, including the 56 reference group participants, using the platform's four-category classification framework (depression, GAD, comorbid depression and anxiety, and regulated reference). This analysis assessed the platform's ability to distinguish the reference group from the clinical diagnostic categories, in addition to differentiating among the three clinical categories.

Across the full cohort, X-RAS achieved 82.9% agreement with the assigned reference categories (Cohen's $\kappa = 0.765$, 95% CI 0.686–0.838), indicating substantial agreement. Among the 56 reference participants, 35 were classified into the regulated reference category, while 21 were assigned to one of the clinical categories.

The classification of these 21 reference participants warrants further examination. All had PHQ-9 and GAD-7 scores below the conventional screening thresholds (≥10) and were considered free of

clinically identified psychiatric disorders at enrolment. Nevertheless, most reported some degree of symptom burden, including endorsement of individual symptom items despite subthreshold total scores. Exploratory examination suggested that the platform's classifications tended to correspond to the symptom domain showing relatively greater elevation. However, these observations do not establish the presence of undiagnosed psychiatric disorders or confirm that the platform detected subclinical psychopathology. Alternative explanations, including physiological heterogeneity within the selected reference population and false-positive classification, cannot be excluded.

Inter-site reliability and generalisability

The two sites presented distinct case mixes, with depression predominating in Delhi and comorbid presentations predominating in Kerala, as described in the Participant characteristics section. Despite these population differences, the platform maintained robust performance at both sites. Agreement with clinical diagnosis reached 94.6% at the Delhi site ($\kappa = 0.884$, 95% CI 0.732 to 1.000) and 91.4% in Kerala ($\kappa = 0.821$, 95% CI 0.639 to 0.955), with the between-site difference not statistically significant ($p = 0.527$).

The sites also represented distinct linguistic and cultural contexts within India, with Hindi-speaking participants in Delhi and Malayalam-speaking participants in Kerala. The comparable performance observed across these settings provides preliminary support for the applicability of

the autonomic signature across different clinical and sociocultural contexts within India, although broader generalisability requires further validation.

Exploratory subgroup analyses revealed no statistically significant differences in diagnostic agreement across age, sex, or education strata. Agreement was similar between participants aged 18–35 years ($\kappa = 0.86$) and 36–60 years ($\kappa = 0.88$; $p = 0.743$), between female ($\kappa = 0.87$) and male participants ($\kappa = 0.88$; $p = 0.880$), and between participants with a high-school education or below ($\kappa = 0.85$) and those with a college education ($\kappa = 0.89$; $p = 0.625$). Following Benjamini–Hochberg false discovery rate correction across four prespecified comparisons (site, sex, age group, and education), none of the between-subgroup differences was statistically significant. No statistically significant interaction between participant sex and diagnostic accuracy was detected. These findings suggest broadly comparable performance across the examined subgroups, although the available sample sizes limit the ability to detect smaller differences and do not establish statistical equivalence. Fig. 4 presents the subgroup-specific agreement estimates, at 95% confidence intervals, and unadjusted and FDR-adjusted p-values.

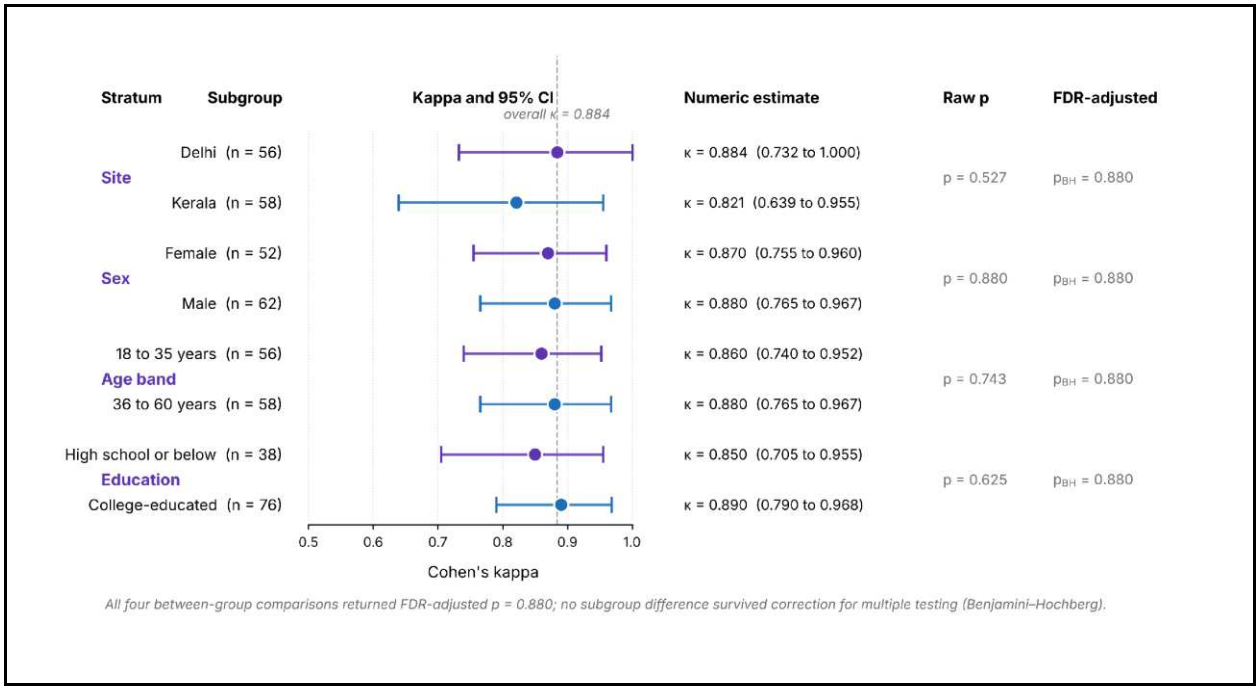


Fig. 4 | Subgroup forest plot of diagnostic agreement (Cohen's kappa) across the four pre-specified demographic strata: site (Delhi versus Kerala), sex (female versus male), age band (18 to 35 versus 36 to 60 years), and education (high school or below versus college-educated). Point estimates and 95% confidence intervals are shown; the dashed vertical line marks the overall primary kappa (0.884). Raw between-group p-values and Benjamini-Hochberg FDR-adjusted p-values are reported in the right-hand columns. No subgroup difference survived correction for multiple testing.

Physiological evidence for the autonomic signature

Measurable autonomic differences between the clinical and reference populations provide the physiological basis for the AI classification. We examined this premise through three complementary analyses. First, we compared established sympathetic and parasympathetic markers between the two cohorts. Second, we assessed whether the physiological constructs that guided feature engineering during model development were also evident in the independently diagnosed validation cohort. Third, we characterised temporal changes in tonic electrodermal activity (EDA), focusing on a signal pattern identified during model development as a potential indicator of altered autonomic regulation.

Classical autonomic markers across the regulated and dysregulated poles

Autonomic parameters showed substantial differences between the clinical cohort and the reference group (Table 3). The differences emerged consistently across sympathetic and parasympathetic measures and remained significant after stringent correction for multiple comparisons.

Table 3. Autonomic nervous system parameters in the autonomically regulated reference group and the autonomically dysregulated clinical cohort.

Parameter	Clinical cohort (mean ± SD)	Reference group (mean ± SD)	Cohen's d	p
Tonic EDA (μS)	0.054 ± 0.077	0.005 ± 0.012	0.77	<0.001
RMSSD (ms)	34 ± 35	67 ± 48	0.83	<0.001
SDNN (ms)	38 ± 26	59 ± 30	0.77	<0.001
pNN50 (%)	8 ± 14	22 ± 21	0.84	<0.001
LF:HF ratio	1.724 ± 1.301	1.098 ± 0.909	0.53	<0.001

Data presented as mean ± standard deviation. Effect sizes computed as Cohen's d. All five parameters remained significant after Bonferroni correction ($\alpha = 0.01$). The two-group comparison characterises the biological dynamic range between the regulated and dysregulated poles. Four-group analyses (reference, depression, GAD, comorbid) for Hypoactivity (SMNA), Experience (LF/HF), and the tonic baseline trajectory are reported in the Results narrative above, with group-level distributions shown in Figs. 5 and 6. SMNA, sudomotor nerve activity; HR, heart rate; RMSSD, root mean square of successive differences; SDNN, standard deviation of NN intervals; pNN50, percentage of successive NN intervals differing by >50 ms; LF:HF, low-frequency to high-frequency power ratio.

Sympathetic indicators were markedly elevated in the clinical cohort. Tonic EDA, a measure of baseline sympathetic arousal, was more than ten-fold higher in clinical participants than in reference participants ($p < 0.001$, $d = 0.77$). Phasic EDA responses and sudomotor nerve activity followed the same direction with moderate effect sizes ($d = 0.52$ and 0.49 respectively).

Parasympathetic markers showed even larger differences. RMSSD, the primary time-domain measure of vagal tone, was reduced by 49% in clinical participants ($p < 0.001$, $d = 0.83$). The large effect size reflects substantial parasympathetic withdrawal. SDNN showed a comparable reduction ($d = 0.77$), and pNN50 showed the largest proportional difference ($d = 0.84$).

Frequency-domain analysis revealed altered autonomic balance. The LF/HF ratio was elevated by 57% in clinical participants ($p < 0.001$, $d = 0.53$). After Bonferroni correction across the five autonomic parameters in Table 3 (adjusted $\alpha = 0.01$), all five remained statistically significant, with moderate to large effect sizes throughout.

Reproducibility of H2SEC constructs in the validation cohort

The H2SEC framework, which is the mathematical foundation used during AI model development, organised 47 features derived from the EDA and BVP signals, into five physiologically interpretable constructs: Habituation, Hypoactivity, Synchronization, Experience, and Calmness (elaborated in Methods, Development of the AI platform). Two of these constructs, Hypoactivity and Experience, was reassessed directly from the 15-minute protocol used in the trial. Reproducing them in independently diagnosed patients, who were scanned under a shortened protocol and against a different reference standard than the development cohort, tested whether the constructs capture stable physiological properties of the dysregulated state rather than artefacts of the development pipeline.

Hypoactivity, indexed by the amplitude of the sudomotor nerve activity (SMNA), differed across the four groups (Kruskal–Wallis $p = 2.7 \times 10^{-5}$). The comorbid subgroup was markedly elevated above the reference group (reference 1.0×10^{4} versus comorbid 3.6×10^{4}; Mann–Whitney U $p = 0.005$), consistent with the development-cohort finding of raised resting sympathetic drive in clinical autonomic dysregulation.

Experience, indexed by the LF/HF ratio of HRV, was lowest in the reference group (0.78 ± 0.37) and elevated across all three clinical subgroups (depression 2.51, anxiety 2.31, comorbid 1.98). The four-group Kruskal–Wallis reached $p = 0.011$, and the reference-versus-comorbid pairwise contrast reached $p = 1.6 \times 10^{-4}$. Group-level distributions for both constructs are shown in Fig. 5.

The reproducibility of the two constructs in independently diagnosed patients is revisited in the Discussion as evidence of physiological interpretability.

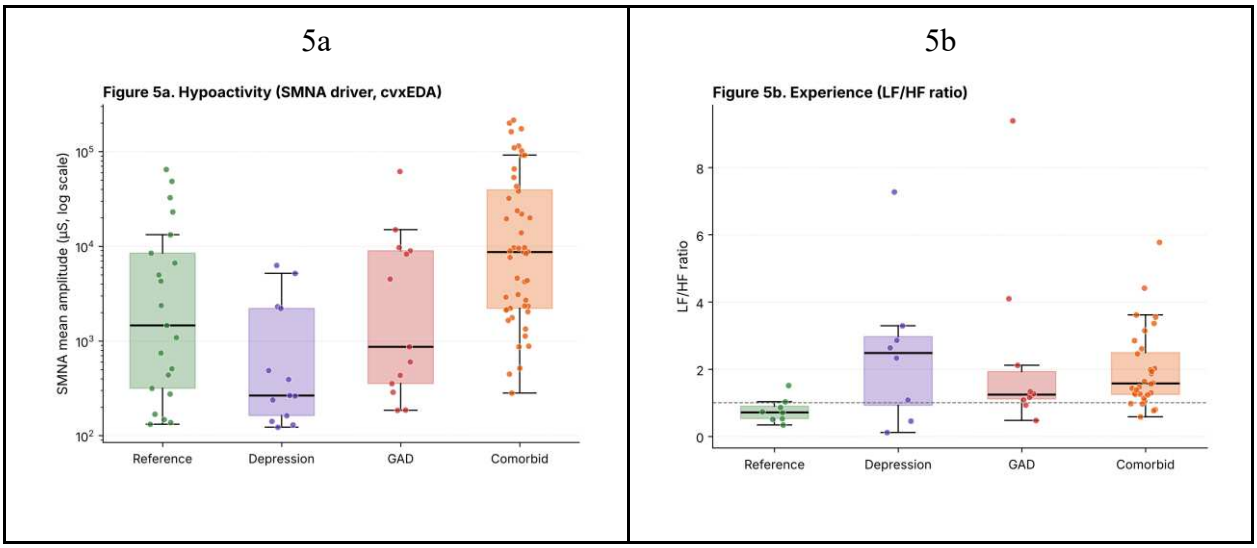


Fig. 5 | Group-level distributions of two H2SEC physiological constructs that could be computed from the 15-minute validation protocol. Panel a: Hypoactivity, indexed by the mean SMNA driver amplitude (log scale). Panel b: Experience, indexed by the LF/HF ratio of heart rate variability. Individual data points overlaid on the box plots.

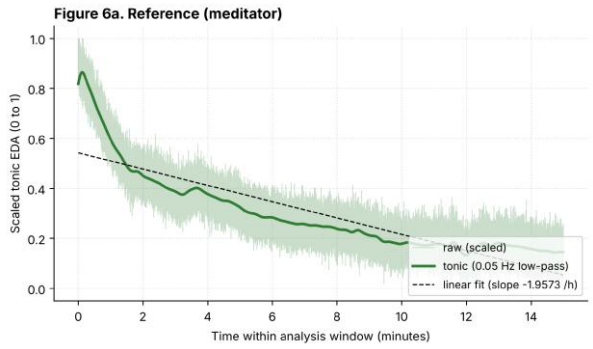
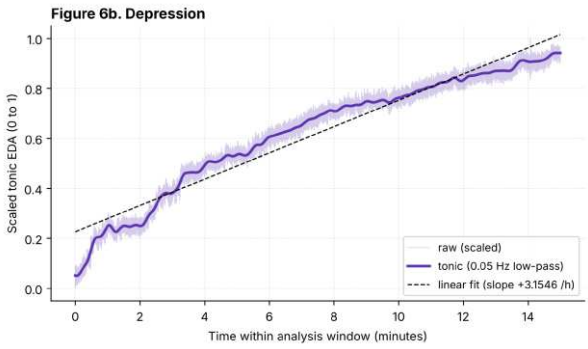
A transdiagnostic marker: direction of the tonic EDA baseline during rest

During development, we had observed that recordings from the clinical group and the recordings from the reference group differed not only in absolute level but in direction over time. The reference group baselines tended to settle downward during rest, while clinical baselines tended to drift upward. We examined this pattern systematically across the full validation cohort using the per-minute slope analysis developed in the foundational Xaant work.

For each recording, we computed the linear regression slope of the scaled tonic EDA within each non-overlapping 1-minute segment and averaged the slopes across all segments in the recording. In the autonomically regulated reference group, the mean per-window slope was -1.66×10^{-6} per sample (SD 1.42×10^{-5}), which corresponds to a progressive downward drift of the tonic baseline during rest, which indicates that physiologically the subject is settling into a calmer state

over time. Every one of the three clinical subgroups showed the opposite pattern. Mean per-window slopes were $+5.23 \times 10^{-6}$ (SD 0.98×10^{-5}) in depression, $+2.11 \times 10^{-6}$ (SD 2.02×10^{-5}) in anxiety, and $+2.63 \times 10^{-6}$ (SD 2.11×10^{-5}) in comorbid, all rising rather than settling.

The four-group Kruskal–Wallis test confirmed a highly significant effect ($H = 103.3$, $p = 3.0 \times 10^{-22}$). Pairwise Mann–Whitney U tests showed significant reference-versus-clinical separations for depression ($p < 0.001$), anxiety ($p = 0.014$), and comorbid ($p = 1.6 \times 10^{-4}$). Within the clinical cohort, depression showed the most consistent monotonic rise. Comorbid presentations produced the largest subject-level excursion over the 15-minute window and a within-window standard deviation approximately twice that of the other subgroups, consistent with the visually jagged morphology observed during data collection. Representative tonic trajectories from each group are shown below:

<table>
<tr><td>6a</td><td>6b</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>6c</td><td>6d</td></tr>
</table>

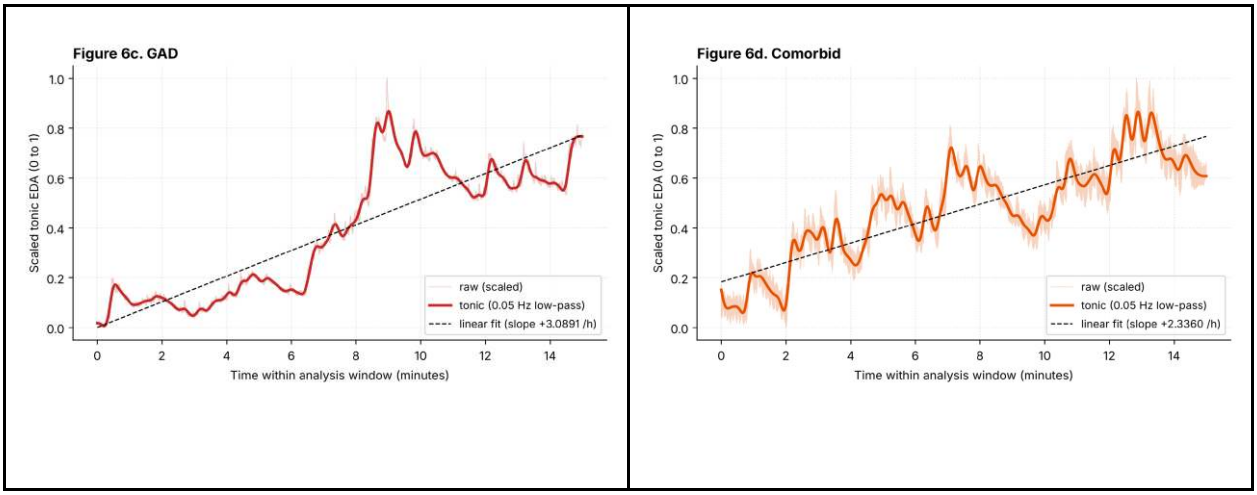


Fig. 6 | Representative tonic electrodermal activity trajectories for one participant from each group (panels a to d: reference, depression, GAD, comorbid). The raw EDA trace is shown in light shading with the low-pass tonic component overlaid, and a linear fit across the analysis window indicates the per-window trajectory.

This result identifies the direction of the tonic baseline as a transdiagnostic physiological marker of autonomic dysregulation: a rising baseline is observed across anxiety, depression, and comorbid presentations, and is discriminated from the settling pattern seen in the regulated reference group. The clinical interpretation is revisited in the Discussion.

Comparison with self-reported symptoms

We examined the validity of the two main self-report routes used in routine clinical practice (Fig. 7). The first was the standardised PHQ−9 and GAD−7 cutoff classifier (PHQ−9 ≥ 10, GAD−7 ≥ 10). Against the psychiatrist DSM−5 reference standard, this questionnaire classifier produced Cohen's κ = 0.073 (95% CI −0.029 to 0.175, p = 0.161), which corresponds to essentially no agreement beyond chance. Against the same reference standard the AI platform produced κ = 0.105 (95% CI 0.001 to 0.209, p = 0.048) when evaluated against the questionnaire labels rather than the clinical

labels (Table 4). Of the 114 clinical patients, 81 (71.1%) exceeded both questionnaire thresholds, compared with 47 (41.2%) who received a comorbid clinical diagnosis ($\chi^2 = 18.42$, $p < 0.001$).

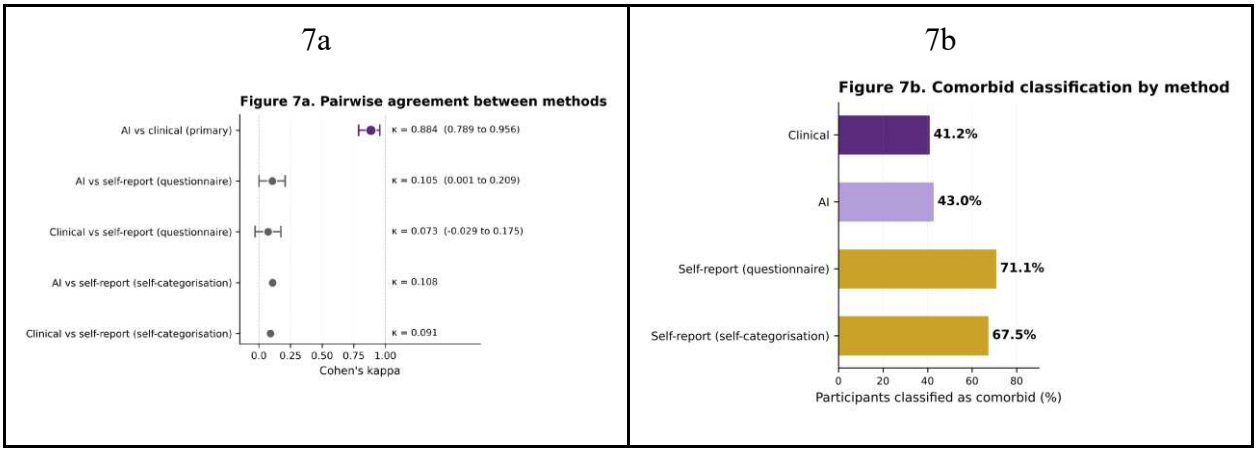


Fig. 7 | Three-way agreement visualisation: kappa comparison (panel a) and diagnostic distribution (panel b). Panel a: Cohen's kappa coefficients for pairwise agreement between AI, clinical, and self-report (both operationalisations). Panel b: distribution of diagnostic categories across methods. Self-report (questionnaire) shows substantial inflation of comorbid classifications relative to clinical and AI assessments.

Table 4. Agreement between diagnostic methods.

Comparison	κ	95% CI	p
AI vs clinical diagnosis †	0.884	0.789 to 0.956	<0.001
AI vs self-report (questionnaire)	0.105	0.001 to 0.209	0.048
Clinical vs self-report (questionnaire)	0.073	-0.029 to 0.175	0.161

†Primary outcome. Self-report (questionnaire) classification based on PHQ-9 ≥10 for depression and GAD-7 ≥10 for anxiety. The two self-categorisation comparisons (clinical κ = 0.091; AI κ = 0.108) are reported in the Results narrative rather than tabulated here.

The second route used each patient's own identification of their primary condition, as anxiety, depression, or both. This also agreed poorly with the other two assessments (clinical $\kappa = 0.091$; AI $\kappa = 0.108$), and 67.5% of patients self-identified as comorbid. Both self-report routes therefore over-classified patients as comorbid relative to the structured psychiatric assessment, which inflates the apparent prevalence of dual diagnosis in this cohort.

Summed questionnaire cutoffs can also obscure the symptoms a patient actually reports. A patient who scores the maximum on three GAD-7 items, such as feeling on edge, uncontrollable worry, and trouble relaxing, still gets a total of only 9, which falls below the cutoff of 10 and is therefore recorded as no anxiety. The underlying issue is that the summed-cutoff classifier treats symptoms as interchangeable and collapses clinically distinct profiles into a single total.

Discussion

Principal findings

In this multicenter study, the Xaant Resting Autonomic Signature (X-RAS) platform classified patients with depression, GAD, and comorbid presentations in naturalistic clinical settings. Within the clinical cohort, the platform reached Cohen's $\kappa = 0.884$ (95% CI 0.789 to 0.956) with psychiatrist-assigned DSM-5 diagnoses. This result should be read against the recognised variability of clinical psychiatric diagnosis, and a direct comparison with published inter-rater or test–retest reliability estimates is not appropriate. The reference diagnoses were made in routine practice at two geographically distinct tertiary-care centres, independently of the platform, which reflects the settings in which such a tool would operate potentially in real life.

The study is a prospective validation of an AI model that was developed in advance and frozen before the trial, rather than the development and validation of a new architecture. The question it addresses is whether the physiological patterns learned during development retain clinical validity when the model is applied prospectively to independently diagnosed patients. The model reached 97% cross-validation accuracy during development against a composite questionnaire-derived label, and 93% diagnostic agreement in the present validation against psychiatrist DSM-5 diagnosis. These findings were obtained using different reference standards and are therefore

not directly comparable. Nevertheless, the performance of the same frozen model across two independently assessed cohorts provides supportive evidence of its generalisability, although differences in recruitment procedures, cohort characteristics, and assessment protocols warrant cautious interpretation.

Classification of generalised anxiety disorder (GAD) was a notable finding. The platform correctly identified all 15 participants with a primary GAD diagnosis, yielding a sensitivity of 100% (95% CI 78.2 to 100). However, the small GAD subgroup and the exclusion of other anxiety disorders limit the generalisability of this result.

Physiological interpretation

Autonomic dysregulation is a well-established feature of anxiety and depressive disorders, supported by a substantial body of evidence. The neurovisceral integration model of emotion regulation [39] posits bidirectional coupling between central regulatory circuits and peripheral autonomic output, and frames resting-state autonomic assessment as a measure of trait-level vulnerability. Meta-analyses report reduced resting HRV in major depression [35] and anxiety disorders [36], and umbrella reviews across mental disorders confirm the robustness of these findings [9]. The approach is validated in other clinical domains, where resting HRV is an established predictor of cardiovascular morbidity [37] and a standard assessment of autonomic function in diabetes [38].

The elevated LF/HF ratio observed in clinical participants reflects differences in the frequency-domain characteristics of heart rate variability. Together with reductions in RMSSD and pNN50, these findings are consistent with altered autonomic regulation, particularly reduced cardiac vagal modulation. Resting-state assessment provides a measure of baseline autonomic activity,

complementing stress-induction protocols that assess physiological reactivity. This distinction may be clinically relevant, as resting autonomic patterns could provide information about regulatory function without requiring an externally induced stress response.

The reproduction of two H2SEC constructs identified during model development in the independently diagnosed validation cohort provides physiological support beyond the observed diagnostic agreement. The differences in conventional autonomic markers presented in Table 3 characterise physiological variation between the reference group and the pooled clinical cohort. However, no individual marker independently distinguishes depression, GAD, and comorbid presentations. The ability of X-RAS to classify participants across these three clinical categories suggests that the integrated model captures information beyond that provided by individual autonomic markers, although the specific contribution of the LSTM architecture requires confirmation through comparative modelling and further ablation analyses.

The reproduction of group-level physiological patterns associated with these constructs in independently diagnosed patients provides additional support for their physiological relevance. However, these findings do not establish the relative contribution of individual constructs to the model's predictions and should be regarded as complementary to, rather than a substitute for, feature-attribution methods such as SHAP or permutation importance. The remaining two H2SEC constructs, Habituation and Synchronization, could not be reassessed using the 15-minute validation protocol but remain incorporated in the model. Their independent evaluation will require future studies incorporating longer recording periods and a dedicated respiratory measurement channel.

Analysis of tonic electrodermal activity (EDA) revealed a distinctive temporal pattern that may have clinical relevance. During the resting-state recording, the tonic EDA baseline generally declined in the reference group but increased across all three clinical subgroups. Previous research on electrodermal activity in depression has primarily described reduced phasic reactivity, often characterised as electrodermal flatness. The present findings extend this perspective by examining the direction of tonic baseline changes over time, rather than focusing exclusively on phasic responses. At the one-minute analysis scale, the upward trend was most consistent among participants with depression, whereas those with comorbid depression and anxiety exhibited the greatest baseline excursion and within-window variability. These observations suggest that the tonic EDA trajectory may represent a candidate transdiagnostic physiological feature associated with the clinical presentations examined. However, its reproducibility, diagnostic specificity, and potential contribution to classification performance require independent validation.

Diagnostic boundaries and questionnaires

The prevalence of comorbid depression and anxiety in the clinical cohort was 41% according to psychiatrist-assigned DSM-5 diagnoses and 68% according to patient self-categorisation, highlighting differences between clinical diagnosis and patients' perceptions of their symptoms. All eight classification disagreements occurred between the depression and comorbid categories, while all participants with a primary diagnosis of GAD were correctly classified. In five of these eight cases, the X-RAS classification aligned with the patient's self-categorisation. This pattern may reflect overlapping symptom profiles and shared autonomic characteristics of depression and anxiety. Although these findings raise the possibility that autonomic measurements capture

anxiety-related physiological features below the threshold for a formal comorbid diagnosis, this interpretation requires further clinical investigation. Similarly, the assignment of certain reference group participants to clinical categories may reflect physiological heterogeneity within the reference group. However, false-positive classifications and other sources of autonomic variation cannot be excluded.

Self-report measures showed low agreement with psychiatrist-assigned diagnoses under both classification approaches, although the reasons for this discordance cannot be determined from the present study. Potential contributing factors include mental health stigma in India [17], cultural differences in the expression of psychological distress, including somatic presentations documented in South Asian populations [18], and possible differences in how patients interpret and respond to items on the PHQ-9 and GAD-7 [40]. These explanations remain hypotheses, as their contributions were not directly evaluated in this cohort. As autonomic assessment does not depend on patients explicitly reporting their psychological symptoms, it may provide complementary information in settings where linguistic, cultural, or individual factors influence symptom disclosure. However, its potential advantage over questionnaire-based screening requires further evaluation in appropriately designed comparative studies.

Positioning against prior AI work

The present findings contribute to the growing literature on artificial intelligence applications in psychiatry while addressing several methodological limitations identified in previous research. Supplementary Table S2 situates X-RAS alongside representative approaches based on autonomic signals, neuroimaging, electrophysiology, and multimodal data. Previous systematic and umbrella reviews have highlighted challenges in translating laboratory-based AI

performance into clinical practice, including risks of bias and a scarcity of prospective validation studies [9,10,32–34]. The present study contributes to this evidence base through prospective evaluation of a model frozen before enrolment, comparison against independently assigned psychiatrist diagnoses, and assessment in naturalistic clinical settings involving patients receiving psychotropic medication. Nevertheless, differences in study populations, diagnostic reference standards, and classification tasks limit direct comparisons with previously published systems. These findings should therefore be interpreted as evidence supporting the clinical feasibility and diagnostic agreement of X-RAS within the populations studied, rather than as a direct benchmarking.

Strengths and limitations

The study has three principal methodological strengths: prospective validation of a model developed and frozen before the trial, recruitment across two geographically and linguistically distinct clinical centres, with psychiatrist-assigned diagnoses obtained independently of the platform and without centralised adjudication, and evaluation of the primary diagnostic agreement endpoint exclusively within the clinical cohort, thereby avoiding dependence on the composition of the reference group.

Several limitations warrant consideration. Further validation is required across primary-care and community settings, rural and resource-limited populations, telehealth environments, and healthcare systems outside India. Both the clinical and reference cohorts were recruited within India, and differences in predictor distributions between the development and validation cohorts were not formally assessed, limiting conclusions regarding model transportability. Furthermore, the anxiety classification was evaluated specifically in patients with generalised anxiety disorder

(GAD), with a relatively small GAD-only subgroup (n = 15), and consequently, the findings should not be generalised to anxiety disorders as a broader diagnostic category.

The reference group comprised long-term contemplative practitioners recruited on the basis of their documented meditation experience. Although sustained contemplative practice may be associated with differences in autonomic regulation, it does not constitute a diagnostic reference standard. In contrast, the clinical cohort was defined by independently assigned psychiatrist diagnoses. Accordingly, the observed physiological differences represent comparisons between two distinct, externally defined populations. The selected reference group also limits the ability to estimate the platform's diagnostic specificity in unselected community populations.

All 114 clinical participants were receiving psychotropic medication, which reflects routine outpatient practice. The contribution of individual medication classes is confounded within the dataset, and the distribution is given in Supplementary Table S1. Common classes, including SSRIs, SNRIs, and benzodiazepines, modulate autonomic function, which bears on the interpretation of the physiological comparisons. Differences in sex distribution and current smoking between the groups remain potential confounders, and the cross-sectional design limits causal inference about the relationship between autonomic dysfunction and psychiatric symptoms. The subgroup analyses reported here are exploratory and do not constitute a comprehensive fairness evaluation. This evaluation, together with feature-level interpretability through SHAP or permutation importance, is a priority for future work.

Clinical implications and future validation

The intended role for X-RAS is as an adjunctive assessment that precedes rather than replaces psychiatric evaluation. A patient referred for mental health evaluation would first undergo the

15-minute resting-state recording, administered by trained non-specialist staff. A psychiatrist or primary care clinician would then review the four-class output and its physiological interpretation alongside the clinical interview. Final diagnostic and treatment decisions remain with the clinician.

Several features support use in routine and resource-limited settings. The protocol requires portable sensors, with a hardware cost below USD 200 per sensor set and per-assessment consumable costs below USD 10. Each assessment comprises a 5-minute acclimatisation period and a 15-minute recording, about 20 minutes of participant time with no active task. Research staff holding bachelor's degrees administered the protocol after 8 hours of training, which indicates that existing clinic staff could adopt the assessment without specialist expertise.

These observations describe operational feasibility rather than a formal usability evaluation. A structured usability study and a cost-effectiveness analysis, covering operator time, infrastructure, licensing, and downstream follow-up, remain for the subsequent deployment phase, alongside the external validation identified above.

Conclusions

This prospective clinical validation shows that artificial intelligence analysis of brief resting-state autonomic signals can detect anxiety and depressive disorders in real-world clinical settings, reaching $\kappa = 0.884$ (95% CI 0.789 to 0.956) with independent psychiatrist diagnosis in a fully medicated cohort across two geographically and culturally distinct Indian tertiary care centres. The convergence between the 93 percent agreement reported here and the 97 percent cross-validation accuracy reported during development, against methodologically distinct reference standards and in independent cohorts, is a form of convergent validity for the underlying

autonomic signature. Substantial physiological differences between the clinical and the reference groups provide biological grounding for the approach. External validation in primary care, community, telehealth, and non-Indian populations is the clear next step for the programme.

Methods

Study design and ethical approval

We conducted a prospective, multi-center diagnostic accuracy study following Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines between July 2023 and December 2023. Clicebo, an independent contract research organisation, coordinated with both hospitals and managed regulatory approvals and site liaison. No patients or members of the public were involved in the design, conduct, or reporting of this study. The completed STARD checklist and the completed TRIPOD+AI checklist are available as Supplementary Information.

The design is built around five methodological safeguards. First, recruitment at two geographically distant sites, Batra Hospital in Delhi (North India) and Believers Church Medical College Hospital in Thiruvalla, Kerala (South India), sampled two culturally and linguistically distinct clinical populations within one country. Second, clinical assessments were performed by hospital-employed psychiatrists with no affiliation to TerraBlue XT, the company that developed the AI platform. Third, data collection, processing, and statistical analysis were carried out by Clicebo, an independent clinical research organisation, which separates platform development from study execution. Fourth, the AI platform processed physiological recordings alone, without ingesting clinical diagnoses, demographic information, or self-report scores. Fifth, the model and its decision thresholds were frozen before the trial cohort was unblinded, so that the version evaluated here is the version that was committed in advance.

The target enrolment of 120 clinical participants was sized to provide at least 80% power to detect a kappa of 0.70 or greater against a null hypothesis of 0.40 (moderate agreement) at a two-sided alpha of 0.05, using simulation methods for multi-category kappa statistics.

Participant recruitment and selection

We initially screened 237 individuals from outpatient psychiatry departments, enrolling 180 participants who met eligibility criteria. The clinical group, comprised of 120 individuals with DSM-5 diagnosed anxiety and/or depressive disorders. The autonomically regulated reference group comprised 60 long-term meditation practitioners recruited from the Heartfulness Institute, a meditation-focused wellness organisation.

Potentially eligible clinical participants were identified as a consecutive series of patients attending the outpatient psychiatry departments at each site during the enrolment period. Reference group participants were recruited as a convenience sample from eligible members of the Heartfulness Institute meeting the inclusion criteria specified below.

Inclusion criteria for the clinical group included: (1) age 18 to 60 years, (2) a DSM-5 diagnosis of a unipolar depressive disorder (major depressive disorder or persistent depressive disorder) and/or generalised anxiety disorder, confirmed through clinical interview by a qualified psychiatrist, and (3) willingness to undergo physiological recording.

The reference group

The composition of the reference group is a central methodological decision, and we set out the reasoning in detail. A conventional healthy-control cohort is defined by the absence of self-reported symptoms, that is, by a subjective statement of mental health. For a study whose central claim is that autonomic physiology provides an objective measurement framework, admitting a

self-report criterion into the reference comparator would reintroduce the measurement property that the primary signal is designed to overcome. We therefore chose a reference group whose autonomic regulation status is evidenced externally, through documented long-term contemplative practice and institutional attestation. This provides an objective comparator for the autonomically dysregulated clinical cohort, which itself is anchored by independent psychiatrist diagnosis rather than by self-report.

Inclusion criteria required: (1) age 18 to 60 years, (2) at least seven years of continuous meditation practice within a single tradition, practised twice daily, (3) authorisation from the organisation to train others in that practice, and (4) absence of any current or past major psychiatric disorder. These criteria identify practitioners with sustained engagement and demonstrated competence, consistent with published literature linking long-term meditation to distinct autonomic patterns, including enhanced parasympathetic tone, reduced sympathetic reactivity, and elevated HRV relative to non-practising adults [27-31]. Reference-group participants followed the identical recording protocol used for the clinical cohort and did not meditate during the recording. Exclusion criteria, detailed below, were applied identically across both groups. The reference group was chosen to characterise the regulated end of the autonomic regulation spectrum, against which the autonomically dysregulated clinical cohort is contrasted.

Exclusion criteria

Exclusion criteria, applied to all participants, eliminated confounding factors that could affect autonomic function: (1) psychotic disorders, bipolar disorder, or severe personality disorders (2) neurological conditions affecting autonomic function (epilepsy, Parkinson's disease, multiple sclerosis), (3) substance use disorders within the past six months, (4) medication changes within four weeks of assessment, (5) acute medical illness, and (6) pregnancy or lactation.

From 180 enrolled participants, 10 were excluded from analysis: six from the clinical group (three under sedation effects discovered post-enrolment, two with excessive movement artifacts compromising >30% of recording, one with equipment malfunction) and four from the reference group (two with incomplete recordings, one withdrew consent, one disclosed psychiatric history). The final sample comprised 170 participants (114 clinical, 56 reference group).

Clinical assessment protocol

Clinical assessments were performed at each participating site by licensed psychiatrists, who established a DSM-5 diagnosis through clinical interview and Mental State Examination, supplemented by review of relevant medical records. The clinical interview followed the routine outpatient diagnostic process used at each institution, rather than a structured research interview such as the Structured Clinical Interview for DSM-5 (SCID-5) or the Mini-International Neuropsychiatric Interview (MINI). This choice reflects the study's naturalistic design, which was intended to validate the platform under the conditions of routine clinical practice. The implication for inter-rater reliability of the reference standard is discussed in the Limitations. Psychiatrists assessed only participants recruited at their respective institutions. No cross-site diagnostic adjudication or harmonisation process was performed. Psychiatrists were blinded to the physiological recordings and AI outputs. The site psychiatrist's DSM-5 diagnosis served as the reference standard against which the AI classification was compared.

Participants completed validated self-report measures in their preferred language. The Patient Health Questionnaire-9 (PHQ-9) assessed depression severity, with scores ≥10 indicating clinically significant symptoms. The Generalised Anxiety Disorder-7 (GAD-7)

assessed anxiety severity, with scores ≥10 as the clinical threshold. Additionally, participants self-categorised their primary condition as depression, anxiety, or both.

Final diagnostic categorisation classified participants as having: (1) Depression, meeting DSM-5 criteria for a unipolar depressive disorder, major depressive disorder or persistent depressive disorder, without comorbid anxiety. Presentations included melancholic and anhedonic features and chronic, pervasive low mood, and excluded bipolar and psychotic depression (2) Anxiety, meeting DSM-5 criteria for generalised anxiety disorder without comorbid depression (3) Comorbid, meeting DSM-5 criteria for both a unipolar depressive disorder and generalised anxiety disorder, as determined by the assessing psychiatrists using DSM standard.

Medication documentation

To account for potential pharmacological modulation of autonomic function, we recorded all active psychotropic prescriptions for each clinical participant. Medication data are provided in Supplementary Table S1 to characterise the pharmacological heterogeneity of the cohort.

Development of the AI platform

The Xaant AI platform (pronounced ‘Shaant’, meaning ‘calmness’ in Sanskrit/Hindi) emerged from a four-year development programme involving physiological recordings from 5,000 participants, comprising approximately 8,400 hours of clinical recordings and 1,600 hours of recordings from individuals without clinical symptoms. The broader dataset encompassed multiple subsets representing diverse psychological conditions and characteristics. A defined subset of 2,349 participants, comprising individuals with generalised anxiety disorder (GAD),

depressive symptoms, comorbid anxiety and depression, and a calm reference group, was selected for development of the classification model evaluated in the present study. Data for this subset were collected between 2017 and 2021 from clinical and community settings across four Indian cities (Hyderabad, Bangalore, New Delhi, and Guwahati). The development cohort included participants across a range of ages, sexes, socioeconomic backgrounds, and symptom severity levels. All development data were collected with institutional ethics approval and written informed consent.

The training dataset underwent rigorous quality control to ensure diagnostic accuracy. Clinical participants (n = 1,937) included individuals with anxiety disorders (n = 712), depressive disorders (n = 662), and comorbid presentations (n = 563), diagnosed through structured clinical interviews by a clinical team. The control group (n = 412) comprised meditation practitioners (n = 248) from multiple traditions and non-meditators (n = 164). A total of 287 participants were excluded for data quality issues.

Quality assurance involved multi-step validation. All training participants completed comprehensive assessments including demographic questionnaires, sleep quality measures (Pittsburgh Sleep Quality Index, among others), and validated psychiatric scales (including GAD-7, PHQ-9, WHO-5 Well-Being Index, Five Facet Mindfulness Questionnaire, and so on). Questionnaire–physiology concordance was verified and participants with poor concordance were excluded from training to ensure label accuracy.

Reference labels for model development were derived through a composite framework that combined the questionnaire battery into two latent symptom dimensions. Each instrument was normalised to a common 0 to 100 scale. An anxiety severity index was formed as a weighted combination of the GAD-7, the DASS anxiety subscale, the STAI, and BASIS-24, and a

depression severity index was formed as a weighted combination of the PHQ-9, the DASS depression subscale, the Subjective Happiness Scale, and BASIS-24. A bounded contextual modifier derived from fatigue, recovery, resilience, and negative-affect measures adjusted both indices.

Each participant was then assigned to one of four classes, healthy, anxiety, depression, or comorbid, by comparing the two indices against decision thresholds. These thresholds were derived empirically on the development cohort by receiver operating characteristic analysis against clinical ground truth, with the operating point selected by the Youden index. The thresholds were frozen after development and applied unchanged thereafter. The composite framework and its thresholds were used only during model development and played no part in the validation study reported here, where the reference standard was the independent diagnosis established by the site psychiatrist with no access to the autonomic recordings or the platform output.

During development, the network was trained on continuous 45-minute recordings sampled at 250 Hz, extracting 47 features spanning EDA parameters (tonic level, phasic responses, amplitude, peak count, rise time, recovery time) and HRV indices across time-domain (RMSSD, SDNN, pNN50), frequency-domain (LF power, HF power, LF:HF ratio), and geometric measures, supplemented by statistical features derived from proprietary algorithms. The prospective validation reported here used 15-minute recordings, supported by the window-duration analysis described below.

The H2SEC framework

Autonomic physiology can be characterised through several established regulatory processes. Habituation describes the attenuation of autonomic responses following repeated or sustained stimulation, while sympathetic activity reflects aspects of physiological arousal and its regulation. Cardiorespiratory coupling characterises the temporal relationship between cardiac and respiratory rhythms and is influenced by parasympathetic modulation. Frequency-domain measures of heart rate variability (HRV), including the low-frequency (LF) and high-frequency (HF) components, provide additional information about autonomic regulation, although their ratio should not be interpreted as a direct measure of sympathovagal balance.

Although the 47 extracted physiological features capture different aspects of these processes, treating them as an undifferentiated feature vector limits their physiological interpretability. To provide a structured framework for feature engineering and physiological analysis, we organised the development feature set into five constructs, collectively termed the H2SEC framework: Habituation (H_1), Hypoactivity (H_2), Synchronization (S), Experience (E), and Calmness (C). The first four constructs represent distinct aspects of autonomic function, while Calmness provides an integrated scalar summary derived from the other four. These constructs are operational definitions developed for the present framework and should not be interpreted as independently validated clinical measures.

The H2SEC framework is implemented in four stages: (i) range calibration, in which admissible values for each physiological feature are established using the development cohort, (ii) temporal segmentation of each recording into Sympathetic Transition Point (STP) and non-STP intervals, (iii) computation of construct-specific scores separately within STP and non-STP intervals, and

(iv) integration of these construct scores into a single participant-level Calmness Score. The four stages are described below.

Stage 1: Range calibration

A single feature value carries little clinical meaning on its own. A tonic EDA level of 2 μS, for example, can reflect healthy baseline arousal in one person and abnormal elevation in another, depending on the distribution of values observed in the reference population. To scale each feature to its physiologically plausible range, we computed an admissible interval θ for each feature on the development cohort using the interquartile range (IQR):

$$\theta = [\, Q_1 - 1.5 \cdot \mathrm{IQR},\ Q_3 + 1.5 \cdot \mathrm{IQR} \,]$$

where Q_1 and Q_3 are the first and third quartiles of the feature distribution across the development participants. Values outside θ were treated as physiologically implausible and excluded from downstream scoring. The IQR-based interval is a standard non-parametric choice and is robust to the moderate skew observed in several of the autonomic markers.

Stage 2: Sympathetic Transition Point (STP) segmentation

Resting-state autonomic activity is not stationary. Under quiet recording conditions sympathetic tone still rises and falls across the 15-minute window as the subject settles, as attention wanders, and as internal physiological rhythms unfold. A construct score averaged across the whole window therefore conflates calm moments with aroused moments and loses clinically relevant structure. We introduced Sympathetic Transition Point (STP) segmentation to separate calm intervals from aroused intervals so that each construct could be scored independently in the two states. An STP interval is a window in which the subject's tonic sympathetic activity, measured through EDA, sits below its own elapsed trajectory.

STP boundaries are identified from the EDA signal itself through a two-signal decomposition. Let $EMA_m(x)$ denote the exponential moving average of signal x with span m minutes, defined by the standard recursion:

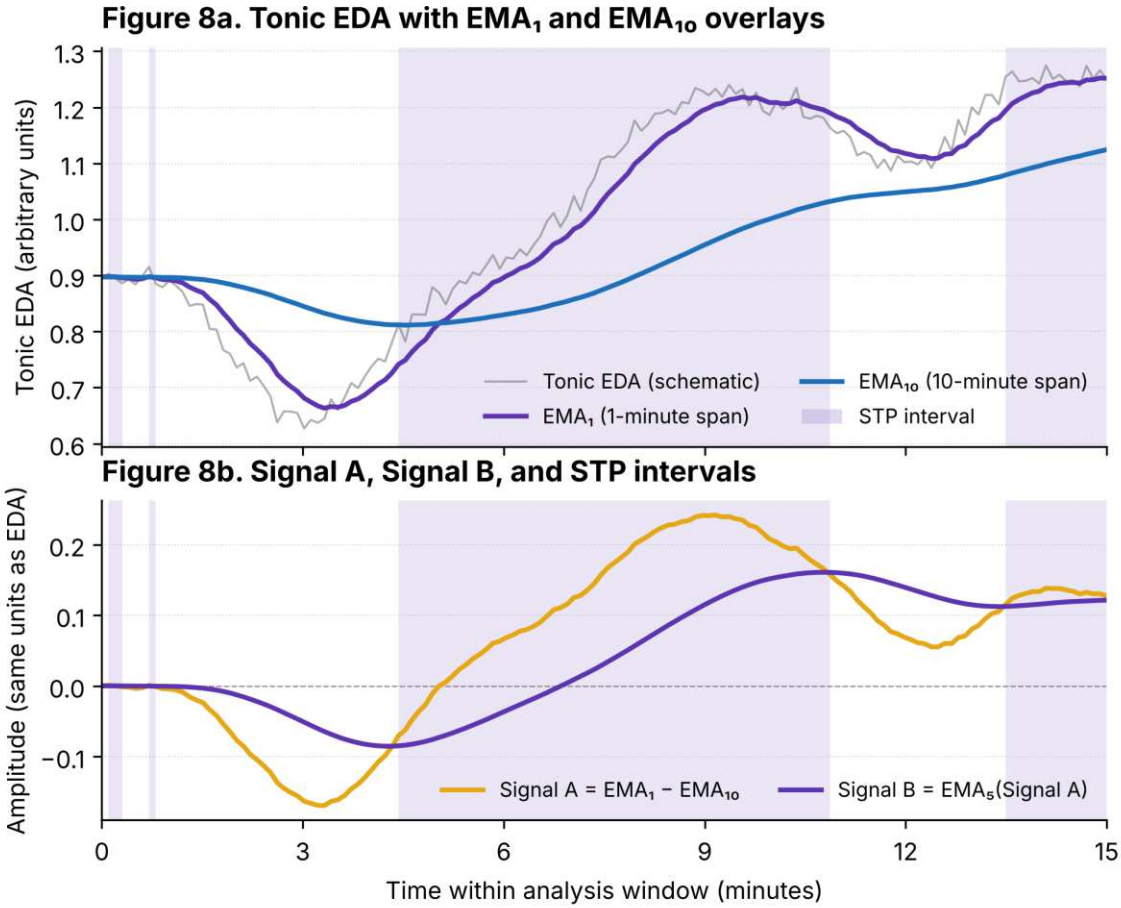
$$\mathrm{EMA}[n] = \alpha \cdot x[n] + (1 - \alpha) \cdot \mathrm{EMA}[n - 1], \quad \alpha = 2 \,/\, (m + 1)$$

Two derived signals are then computed from the tonic EDA. The first captures the short-term fluctuation around the slow trajectory of EDA; the second is a smoothed version of the first on the same time axis:

$$\text{Signal A} = \mathrm{EMA}_1(\mathrm{EDA}) - \mathrm{EMA}_{10}(\mathrm{EDA})$$

$$\text{Signal B} = \mathrm{EMA}_5(\text{Signal A})$$

At each time index, the subject is assigned to the STP state when Signal B sits below Signal A, and to the non-STP state otherwise. Full operational detail, including the hysteresis around zero crossings, the ordering of convergence and divergence tests, and the treatment of edge intervals, is retained in internal documentation. A conceptual schematic of the segmentation is shown in Fig. 8.



822

823 **Fig. 8 |** Conceptual schematic of Sympathetic Transition Point (STP) segmentation. Panel a: a synthetic tonic EDA
824 trace with its EMA1 and EMA10 overlays, with STP intervals shaded. Panel b: Signal A = EMA1(EDA) -
825 EMA10(EDA) and Signal B = EMA5(Signal A) on the same time axis, with STP intervals shaded. Drawn on
826 representative synthetic data.

827 **Stage 3: Construct scoring**

828 Within each recording, we identified T_{in} minutes of STP time and T_{out} minutes of non-STP

829 time. The three H_2E constructs, Habituation (X_1), Hypoactivity (X_2), and Experience (X_3),

830 were each computed independently within STP and outside STP. For each construct $X \in$

831 $\{X_1, X_2, X_3\}$ and each time index t, we scored the construct as the fraction of its constituent

832 features whose current value fell within its admissible interval:

$$X(t) = (1/6) \cdot \Sigma_{j=1}^{6} \mathbb{1}[x_j(t) \in \theta_j]$$

where $\mathbb{1}[\cdot]$ is the indicator function (1 when the feature value falls within its admissible interval θ_j, 0 otherwise). Each feature contributes equally and the construct score at any instant ranges from 0 (no features within range) to 1 (all six features within range). The assignment of the features to each construct is given in Table 5.

Table 5. Feature assignments per H2SEC construct.

Construct	Physiological interpretation	Features
X_1 Habituation	Return to baseline after sympathetic activation	EDA recovery time; EDA rise time; EDA standard deviation; RMSSD; combined HF rate of change and EDA rate of change
X_2 Hypoactivity	Magnitude of resting sympathetic drive	SMNA (sudomotor nerve activity); phasic EDA; EDA amplitude; RMSSD; combined HF rate of change and EDA rate of change
X_3 Experience	Vagal–sympathetic balance during rest	HF power; LF/HF ratio; tonic EDA standard deviation; RMSSD; total HRV power; combined HF rate of change and EDA rate of change

Each construct X_1 (Habituation), X_2 (Hypoactivity), and X_3 (Experience) is scored from six features chosen for their physiological interpretation. The Synchronization (S) and Calmness (C) constructs are computed separately as described in Methods, Stage 3 and Stage 4. HF, high-frequency HRV power; LF/HF, low-frequency to high-frequency HRV ratio; RMSSD, root mean square of successive differences; SMNA, sudomotor nerve activity.

The Synchronization construct (S) is computed separately as a scalar per time window rather than per time index. It captures the fraction of minutes during which the subject exhibits positive cardiorespiratory synchronization. A minute is labelled positively synchronous when $\lambda \geq 0.7$ and the mean heart rate is below 72 BPM, where λ is the mean resultant vector of the heart-rate–respiration phase difference. Φ_{in} and Φ_{out} denote the within-STP and outside-STP synchronization fractions respectively.

Stage 4: Calmness Score

The final integration produces a single scalar per subject, the Calmness Score C. For each subject we compute two aggregates, one within STP and one outside STP, by averaging the three H_2E construct scores over their respective time windows and modulating by the corresponding Synchronization fraction:

$$F(X) = (1 / T_{in}) \cdot \Sigma_{t \in STP} [X_1(t) + X_2(t) + X_3(t)] / 3 \cdot \Phi_{in}$$

$$G(X) = (1 / T_{out}) \cdot \Sigma_{t \notin STP} [X_1(t) + X_2(t) + X_3(t)] / 3 \cdot \Phi_{out}$$

$$C = F(X) - G(X)$$

A positive C indicates that the subject's H_2E construct averages are higher during lowered-sympathetic (STP) segments than during elevated-sympathetic (non-STP) segments, which is the physiological signature of a subject who achieved and sustained calmness during the recording. A negative C indicates the reverse pattern, in which construct scores fall during calm intervals and rise during arousal, a profile consistent with chronic autonomic dysregulation.

All construct thresholds and scoring parameters were derived on the development cohort and frozen before the prospective validation reported here.

Signal processing and feature extraction were implemented in Python 3.11 using neurokit2 [41] for EDA decomposition [42] and for standard time-domain and frequency-domain HRV analysis on the BVP-derived inter-beat series, with complementary routines from BioSPPy [43].

The network architecture comprised three LSTM layers with 128, 64, and 32 hidden units respectively, incorporating dropout regularisation (rate = 0.2) between layers to prevent overfitting. A custom self-attention mechanism weighted temporal segments by diagnostic relevance, enabling the model to focus on physiologically informative windows within the

recording. The softmax output layer provided probability distributions across four diagnostic categories.

Training employed the Adam optimiser with initial learning rate 0.001, utilising exponential decay (factor = 0.95 per epoch). Class weights addressed imbalanced representation, with minority classes weighted inversely to frequency. Early stopping monitored validation loss with patience of 10 epochs. Model validation followed rigorous protocols ensuring generalisation. Five-fold stratified cross-validation maintained diagnostic proportions across folds, achieving mean accuracy of 0.97 (SD = 0.02). The cross-validation split was performed at the participant level: all segments from a given participant were assigned to a single fold, and no participant contributed data to both the training and the held-out set of any fold. This guards against the form of data leakage that would otherwise arise when multiple segments from one participant appear in both training and validation partitions.

Choice of the 15-minute recording window

The development programme used 45-minute recordings, and the prospective validation reported here used 15-minute recordings. We shortened the protocol deliberately, after a sensitivity analysis that compared the HRV and EDA features computed over 15, 30, and 45-minute epochs on the development data. The feature distributions did not differ meaningfully between the three durations, which indicates that a 15-minute window captures the relevant physiological indices as reliably as the longer windows while cutting the acquisition time to a third.

This sensitivity result is consistent with the measurement properties of the markers used for classification. RMSSD, pNN50, tonic EDA, and the LF/HF ratio are summary statistics over the recording window and reflect trait-level regulatory patterns that stabilise within minutes of physiological acclimatisation rather than requiring prolonged observation [20-22]. SDNN rises

with recording length because it accumulates slower oscillations [20], we therefore normalised SDNN prior to model input, and the sensitivity analysis confirmed that the normalised metric was stable across the 15 to 45-minute range. Once the initial 5-minute acclimatisation period has elapsed and physiological stabilisation has occurred, the subsequent 15-minute recording captures a representative sample of the individual's baseline autonomic regulatory state. The self-attention layer of the LSTM reinforces this design by weighting the most diagnostically informative temporal segments within the recording. The 15-minute protocol was adopted for prospective validation to maximise clinical feasibility while preserving diagnostic accuracy.

Role of questionnaires in development versus validation

The questionnaire battery described above served a specific role in model development. It provided a structured framework for aligning physiological recordings with psychological constructs in a large non-clinical and clinical population. The intent was to identify stable physiological signatures associated with the latent anxiety and depressive dimensions, rather than to reproduce questionnaire scores. The prospective validation reported here was designed to evaluate whether those physiology-derived signatures generalise to independent DSM-5 diagnoses established by psychiatrists. Validation is therefore an assessment of convergence between the autonomic biomarkers and expert clinical diagnosis, rather than a replication of the labels used during model development.

Physiological data collection protocol

Participants underwent standardised physiological assessment in quiet clinical rooms maintained at 23 °C (±2 °C) with ambient lighting. The protocol began with a 5-minute acclimatisation period allowing physiological stabilisation, followed by 15-minute resting-state recording with

participants seated comfortably, eyes open, without external stimuli (phones, reading materials, and electronic devices removed).

Physiological signals were acquired using research-grade sensors (BiosignalPlux, PLUX Wireless Biosignals, Portugal). The sensor set is lightweight, battery-powered, and non-invasive, comprising skin-surface electrodes and a finger clip connected by a slim cable to a small acquisition hub placed beside the participant. EDA sensors with Ag/AgCl electrodes were placed on the palmar surface of the index and middle fingers, measuring skin conductance at 250 Hz. BVP sensors utilised photoplethysmography on the ring finger at the same sampling rate.

Signal processing employed validated algorithms implemented in Python 3.11. EDA and HRV features were estimated using cvxEDA and neurokit2. EDA preprocessing applied Butterworth bandpass filtering (0.045–0.25 Hz) to isolate sympathetic responses, followed by stationary wavelet transform removing motion artifacts with 750-sample moving average smoothing. BVP preprocessing utilised two-stage filtering (40 Hz lowpass, 0.5 - 40 Hz bandpass), adaptive filtering for motion artifacts, and spline interpolation for discontinuities. From these signals, the platform extracted time-domain HRV metrics (RMSSD, SDNN, pNN50), frequency-domain measures (LF, HF, LF:HF ratio), Poincaré plot parameters (SD1, SD2), and EDA features (tonic, phasic, amplitude, rise/recovery times).

Quality control metrics ensured data integrity. Recordings with >10% artifacts, signal dropout >30 seconds, or excessive motion identified through accelerometry were excluded. Real-time visualisation allowed technicians to identify and address issues during recording.

For each participant, the physiological recording and the clinical DSM-5 assessment were performed on the same day.

Statistical analysis

Statistical analyses were performed in Python 3.11 using SciPy, StatsModels, and scikit-learn.

The primary outcome, diagnostic agreement between the AI platform and clinical diagnosis within the clinical cohort, was assessed using Cohen's kappa with bootstrap 95% confidence intervals (2,000 resamples). As a pre-specified sensitivity analysis addressing potential site-level clustering, a site-stratified bootstrap of the primary kappa was performed by resampling with replacement within each site to preserve the site case mix across replicates. The study used two assessment sites with one psychiatrist per site, which precludes a well-identified random-effects model with site as the random intercept. The site-stratified bootstrap preserves the observed site composition and examines sensitivity to site stratification. It cannot establish robustness to all site-level sources of uncertainty. A secondary analysis evaluated four-class agreement across all 170 participants with the reference group included. Agreement was interpreted using the Landis and Koch criteria: slight ($\kappa < 0.20$), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80), and almost perfect (0.81 to 1.00).

Diagnostic performance metrics were calculated for each category against the clinical diagnosis. Sensitivity, specificity, positive predictive value, and negative predictive value used exact binomial confidence intervals. F1-scores quantified the harmonic balance between precision and recall. Positive and negative likelihood ratios were reported as indicators of clinical utility.

Group comparisons of physiological parameters used independent-samples t-tests for normally distributed variables and Mann–Whitney U tests for non-normal distributions, with normality assessed by the Shapiro–Wilk test. Effect sizes were reported as Cohen's d for parametric comparisons and rank-biserial correlation for non-parametric tests. A Bonferroni correction was applied across the five autonomic parameters reported in Table 3 (adjusted $\alpha = 0.01$).

Agreement between diagnostic methods (AI, clinical, self-report) was assessed pairwise using Cohen's kappa. The McNemar–Bowker test evaluated the symmetry of disagreements to screen for systematic bias. Chi-square tests compared diagnostic distributions between sites. Across the four pre-specified subgroup analyses (site, sex, age band, education), raw between-group p-values were adjusted for multiple testing using the Benjamini–Hochberg procedure to control the false discovery rate at 0.05, following current reporting practice for subgroup heterogeneity analyses in prospective AI validation studies. Missing data, comprising under 5% of questionnaire responses, were handled through multiple imputation by chained equations (10 imputations). All tests were two-tailed with significance set at $p < 0.05$, adjusted where appropriate for multiple comparisons.

Relationship to the Preliminary Preprint

A preliminary report of this study was posted on Research Square in October 2024 (DOI: 10.21203/rs.3.rs-5293897/v1), to encourage early scientific scrutiny and feedback from the international research community. The feedback and subsequent extensive scientific, statistical, and clinical reviews contributed to the identification and correction of reporting and analytical errors, resulting in a substantially revised and more rigorously evaluated manuscript. The reported agreement consequently changed from 91.23% ($\kappa = 0.838$) in the preliminary preprint to 93.0% ($\kappa = 0.884$) in the present manuscript.

The earlier report also described a broader development dataset of approximately 5,000 individuals encompassing diverse psychological characteristics and multiple condition-specific subsets. An editorial omission in that report resulted in the specific development subset used for the psychiatric classification model not being identified. The present manuscript clarifies that

this model was developed using a defined subset of 2,349 individuals representing generalised anxiety disorder, depression, comorbid anxiety and depression, and a calm reference group.

Declarations

Ethics approval and consent to participate

The study protocol received approval from institutional review boards at Batra Hospital and Research Center, New Delhi, and at Believers Church Medical College Hospital, Thiruvalla, Kerala, and was registered with the Clinical Trials Registry of India (CTRI/2023/07/054984). All participants provided written informed consent prior to enrolment. Consent forms were provided in the participant's preferred language (English, Hindi, or Malayalam).

Consent for publication

Consent forms signed by all participants included explicit permission for publication of anonymised, aggregate data. No individual participant data or images that would permit identification are presented in this manuscript.

Availability of data and materials

Individual-level physiological recordings and clinical assessments contain sensitive mental health information. In accordance with institutional review board requirements and participant consent agreements, access to raw data is restricted. De-identified summary statistics are provided in the article tables and figures. De-identified diagnostic classification data for all 114 participants (AI diagnosis, clinical diagnosis, self-report category) are available from the corresponding author on reasonable request, subject to a data use agreement.

The Xaant model is proprietary to TerraBlue XT. Access to the trained model and model weights is restricted under commercial licensing terms. The Methods describe the architecture (three-layer LSTM with self-attention), the signal processing pipeline, the 47-feature set, and the training procedure in sufficient detail to allow independent replication of the analytical approach.

The complete STARD 2015 checklist and the completed TRIPOD+AI checklist are provided as Supplementary Information.

Competing interests

NS, PRP, and RB are employees of TerraBlue XT, the company that developed the Xaant AI platform evaluated in this study. DS, CA, PM, SSV, and PA declare no competing interests.

Funding

This work was funded by TerraBlue XT, developer of the Xaant platform.

Authors' contributions

DS: Clinical investigation, participant assessment, DSM-5 diagnostic evaluation, and data curation (Delhi site).

CA: Clinical investigation, participant assessment, DSM-5 diagnostic evaluation, and data curation (Kerala site).

PM: Clinical oversight, study review, and supervision (Kerala site).

SSV: Clinical guidance, study review, and methodological input.

PA: Study conceptualisation, study coordination, ethics and regulatory management, data acquisition, data processing, analysis support, and liaison between ethics committees, hospitals, analysis teams, and platform teams (on behalf of Clicebo).

NS: Clinical guidance for platform and study, and study review.

PRP: Study design support, signal processing methodology, model design and development, model validation, sensor calibration and troubleshooting, quality control, and technical review.

RB: Platform conceptualisation, platform research and design, computational model design and architecture, X-RAS framework definition, funding acquisition and project administration. All authors reviewed, revised, and approved the final manuscript.

Acknowledgements

The authors thank the participating patients and reference-group participants, the Heartfulness Institute for facilitating reference-group recruitment, and the clinical and data teams at Batra Hospital and Research Center and Believers Church Medical College Hospital for their contribution to study execution. Clicebo acted as the independent contract research organisation coordinating data collection, management, and statistical analysis.

References

1. GBD 2019 Mental Disorders Collaborators. 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. Lancet Psychiatry. 2022;9:137-150.

2. World Health Organization. Depression and other common mental disorders: global health estimates. Geneva: World Health Organization; 2017.

3. Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. Lancet. 2018;392:1553-1598.

4. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull. World Health Organ. 2004;82:858-866.

5. Gururaj G, Varghese M, Benegal V, et al. National Mental Health Survey of India, 2015-16: Summary. Bengaluru: National Institute of Mental Health and Neurosciences; 2016.

6. Garg K, Kumar CN, Chandra PS. Number of psychiatrists in India: baby steps forward, but a long way to go. Indian J. Psychiatry. 2019;61:104-105.

7. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, Part II: test–retest reliability of selected categorical diagnoses. Am. J. Psychiatry. 2013;170:59-70.

8. Meyer GJ, Finn SE, Eyde LD, et al. Psychological testing and psychological assessment: a review of evidence and issues. Am. Psychol. 2001;56:128-165.

9. Abd-Alrazaq A, Alhuwail D, Schneider J, et al. The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. npj Digit. Med. 2022;5:87.

10. Chen Z, Liu X, Yang Q, et al. Evaluation of risk of bias in neuroimaging-based artificial intelligence models for psychiatric diagnosis: a systematic review. JAMA Netw. Open. 2023;6:e231671.

11. Thayer JF, Åhs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies. Neurosci. Biobehav. Rev. 2012;36:747-756.

12. Chalmers JA, Quintana DS, Abbott MJ, Kemp AH. Anxiety disorders are associated with reduced heart rate variability: a meta-analysis. Front. Psychiatry. 2014;5:80.

13. Kemp AH, Quintana DS, Gray MA, et al. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. Biol. Psychiatry. 2010;67:1067-1074.

14. Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. Int. J. Psychophysiol. 2015;98:338-350.

15. McEwen BS, Stellar E. Stress and the individual: mechanisms leading to disease. Arch. Intern. Med. 1993;153:2093-2101.

16. Hamilton JL, Alloy LB. Atypical reactivity of heart rate variability to stress and depression. Biol. Psychol. 2016;118:108-118.

17. Shidhaye R, Kermode M. Stigma and discrimination as a barrier to mental health service utilization in India. Int. Health. 2013;5:6-12.

18. Raguram R, Weiss MG, Channabasavanna SM, Devins GM. Stigma, depression, and somatization in South India. Am. J. Psychiatry. 1996;153:1043-1049.

19. India State-Level Disease Burden Initiative Mental Disorders Collaborators. The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017. Lancet Psychiatry. 2020;7:148-161.

20. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93:1043-1065.

21. Boucsein W, Fowles DC, Grimnes S, et al. Publication recommendations for electrodermal measurements. Psychophysiology. 2012;49:1017-1034.

22. Munoz ML, van Roon A, Riese H, et al. Validity of (ultra)-short recordings for heart rate variability measurements. PLoS ONE. 2015;10:e0138921.

23. World Health Organization. Comprehensive Mental Health Action Plan 2013-2030. Geneva: World Health Organization; 2021.

24. Insel T, Cuthbert B, Garvey M, et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry. 2010;167:748-751.

25. Linardon J, et al. Multidisciplinary research priorities for artificial intelligence in mental health: a call to action. Lancet Psychiatry. 2026;13:792-803.

26. Royal College of Psychiatrists. Use of digital monitoring technology in mental health settings [position statement]. London: Royal College of Psychiatrists; 2025 Jan 22. Available from: https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2025/01/22/use-of-digital-monitoring-technology-in-mental-health-settings.

27. Peng CK, et al. Heart rate dynamics during three forms of meditation. International Journal of Cardiology. 2004;95:19-27.

28. Krygier JR, Heathers JAJ, Shahrestani S, Abbott M, Gross JJ, Kemp AH. Mindfulness meditation, well-being, and heart rate variability: a preliminary investigation into the impact of intensive Vipassana meditation. International Journal of Psychophysiology. 2013;89:305-313.

29. Tang YY, Hölzel BK, Posner MI. The neuroscience of mindfulness meditation. Nature Reviews Neuroscience. 2015;16:213-225.

30. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? Frontiers in Psychology. 2014;5:756.

31. Thimmapuram J, Pargament R, Sibliss K, Grim R, Risques R, Toorens E. Effect of heartfulness meditation on burnout, emotional wellness, and telomere length in health care professionals. Journal of Community Hospital Internal Medicine Perspectives. 2017;7:21-27.

32. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA. 2017;318:1215-1216.

33. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors. Annual Review of Clinical Psychology. 2017;13:23-47.

34. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25:44-56.

35. Koch C, Wilhelm M, Salzmann S, Rief W, Euteneuer F. A meta-analysis of heart rate variability in major depression. Psychological Medicine. 2019;49:1948-1957.

36. Cheng Y-C., Huang Y-C., Huang W-L. Heart rate variability in patients with anxiety disorders: a systematic review and meta-analysis. Psychiatry and Clinical Neurosciences. 2022;76:292-302.

37. Hillebrand S, Gast KB, de Mutsert R, Swenne CA, Jukema JW, Middeldorp S, Rosendaal FR, Dekkers OM. Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose-response meta-regression. Europace. 2013;15:742-749.

38. Vinik AI, Ziegler D. Diabetic cardiovascular autonomic neuropathy. Circulation. 2007;115:387-397.

39. Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. Journal of Affective Disorders. 2000;61:201-216.

40. Sumathipala A, Siribaddana SH, Bhugra D. Culture and somatisation: clinical presentation of somatoform disorders in Sri Lanka. British Journal of Psychiatry. 2004;184:200-209.

41. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behavior Research Methods. 2021;53:1689-1696.

42. Greco A, Valenza G, Lanata A, Scilingo EP, Citi L. cvxEDA: a convex optimisation approach to electrodermal activity processing. IEEE Transactions on Biomedical Engineering. 2016;63:797-804.

43. Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, Fred A. BioSPPy: biosignal processing in Python. 2015 [software]. Available from: https://github.com/scientisst/BioSPPy.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary1STARDChecklist.docx
- Supplementary2TRIPODAIChecklist.docx
- Supplementary3MedicationTables.docx
- Supplementary4ComparisonXaantAIvsExistingWorkTable.docx