Geopsychiatry 4 (2026) 100115

Contents lists available at ScienceDirect

Geopsychiatry

journal homepage: www.sciencedirect.com/journal/geopsychiatry

Concurrent risk profiling of depressive and post-traumatic stress symptoms among war-displaced adults: An explainable machine learning analysis

Seyed-Ali Sadegh-Zadeh [a,*,1], Shayan Saadat [b,1], Zeynab Sadeghimanesh [c], Ommolbanin Soleymani [d], Alireza Soleimani Mamalo [a], Sanam Anoosheh [e], Mohammadreza Shalbafan [f], Elham Khalilian [g], Seyed-Yaser Mousavi [g]

[a] Department of Computing, University of Staffordshire, Stoke-on-Trent, United Kingdom
[b] Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham, United Kingdom
[c] University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
[d] Islamic Azad University, Kashan, Iran
[e] Department of Psychiatry, School of Medicine, Urmia University of Medical Sciences, Urmia, Iran
[f] Mental Health Research Center, Psychosocial Health Research Institute (PHRI), Department of Psychiatry, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[g] Department of Psychiatry, Tehran University of Medical Sciences, Tehran, Iran

A R T I C L E I N F O

Keywords:
Depressive symptoms
PTSD symptoms
Refugees
Machine learning
SHAP
Risk profiling
Humanitarian mental health

A B S T R A C T

Background: Conflict-affected populations face disproportionate psychological distress, yet humanitarian settings lack efficient tools for identifying those most in need. We examined whether routinely collected displacement survey indicators can distinguish adults at high versus low concurrent burden of depressive and post-traumatic stress symptoms.

Methods: Cross-sectional baseline data (2019) from the Cox's Bazar Panel Survey were analysed among camp-resident displaced Rohingya adults (N = 4598). Outcomes were elevated depressive symptoms (PHQ − 9 ≥ 10) and PTSD symptoms (HTQ Part IV ≥ 2.5). L2-logistic regression, elastic net, and XGBoost (with nested hyperparameter tuning) were evaluated via household-grouped 5-fold cross-validation. Geographic validation used held-out primary sampling units (PSUs). SHAP identified key correlates mapped to the IASC MHPSS intervention pyramid.

Results: Among camp adults (56% female; mean age 33.7), 23.7% screened positive for elevated depressive symptoms and 5.9% for PTSD symptoms (Cronbach's α: PHQ-9 = 0.83, HTQ = 0.87). Tuned XGBoost achieved AUC = 0.819 internally and 0.814 in PSU-holdout geographic validation. Risk tertiles showed depression screening prevalence of 0.8% (low), 7.0% (medium), and 62.1% (high). Food insecurity, trauma exposure, crime or conflict experience, and healthcare barriers were the most influential modifiable correlates.

Conclusions: Displacement survey indicators meaningfully stratify concurrent mental health symptom burden. Cross-sectional design precludes prognostic claims; prospective validation and implementation research are needed before operational deployment.

1. Introduction

Armed conflict and forced displacement are among the strongest determinants of population mental health burden. A seminal meta-analysis by Charlson et al. [1] estimated that 22.1% of individuals in conflict-affected settings experience depression, anxiety, or post-traumatic stress disorder (PTSD) at any given time, substantially exceeding global baselines. Steel et al. [2] found that cumulative trauma exposure and post-migration living difficulties independently predict mental health outcomes among refugees and conflict-affected

* Corresponding author.
E-mail addresses: ali.sadegh-zadeh@staffs.ac.uk (S.-A. Sadegh-Zadeh), shayan.saadat@uhb.nhs.uk (S. Saadat), Zeynab.sadeghimanesh@gmail.com (Z. Sadeghimanesh), soleymani54.81@gmail.com (O. Soleymani), anoosheh.sanam@gmail.com (S. Anoosheh), drmrsh@gmail.com (M. Shalbafan), Khalilianelham7@gmail.com (E. Khalilian), Drsymousavi@gmail.com (S.-Y. Mousavi).
[1] Joint First author.

https://doi.org/10.1016/j.geopsy.2026.100115
Received 17 May 2026; Accepted 27 July 2026
Available online 7 August 2026
3050-7138/© 2026 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

populations. Nearly all conflict-exposed individuals experience some psychological distress, making mental health a system-wide concern in humanitarian emergencies [3,4] (see Figs. 1–9).

Despite this burden, mental health services in displacement settings are chronically underfunded. The WHO estimates a treatment gap exceeding 75% in low- and middle-income countries [5]; during active conflict, this gap widens as infrastructure collapses and qualified personnel are lost [6]. Effective mental health and psychosocial support (MHPSS) programming therefore requires efficient resource allocation: identifying who carries the highest symptom burden and where services should be directed 7–9].

In displacement contexts, psychological distress is shaped not only by pre-migration trauma but equally by post-migration living difficulties, food insecurity, housing strain, restricted mobility, safety threats, and barriers to healthcare [2,10,11]. Miller and Rasmussen [12] argued that daily stressors in displacement may be as important as war-related trauma in maintaining psychological distress, a framework supported by subsequent empirical work [13,14]. These contextual determinants are potentially modifiable through humanitarian programming, making them relevant targets for intervention design. The IASC Guidelines on MHPSS in Emergency Settings [9] provide a layered framework from basic services through specialised care that can operationalise such targeting.

Prior research on the Rohingya displacement crisis, the largest refugee camp complex globally with over 900,000 people in Cox's Bazar, Bangladesh, has documented extremely high rates of distress [15,16]. Riley et al. [17] and analyses using the Cox's Bazar Panel Survey (CBPS) [15,18] have identified correlates through conventional regression approaches. However, to our knowledge, no prior study has applied explainable machine learning (ML) with systematic policy translation to this population.

Machine learning approaches have been increasingly applied to mental health prediction in clinical [19,20] and population-based settings [21,22], as well as other clinical domains including neurodegenerative disease screening [23] and surgical outcome prediction [24], typically using SHAP (SHapley Additive exPlanations) [25] for model interpretability, alongside emerging approaches for handling class imbalance and limited training data [26,27]. However, as Christodoulou et al. [28] and van der Ploeg et al. [29] demonstrated, in many clinical prediction contexts, complex ML models offer negligible performance gains over logistic regression, a finding with direct implications for deployability in resource-constrained humanitarian settings.

Several important caveats frame this work. First, the CBPS baseline is cross-sectional; we characterise concurrent risk profiles, not prognostic predictions. The distinction is fundamental: a cross-sectional model identifies who currently has elevated symptoms, not who will develop them [30]. Second, the PHQ-9 and HTQ are screening instruments, not diagnostic tools; depressive symptoms and PTSD symptoms denote symptom burden above screening thresholds [31,32]. Third, these instruments have not been formally validated in the Rohingya language, although internal reliability in the present sample was good (reported below) and prior use in this population has been documented [15,17].

This study has three objectives: 1) to evaluate whether models based on routinely collected displacement survey indicators can meaningfully stratify concurrent symptom burden; 2) to identify which indicators contribute most, distinguishing fixed from modifiable factors; and 3) to map modifiable correlates onto the IASC MHPSS pyramid [9] as hypotheses for programmatic prioritisation, while acknowledging that cross-sectional associations require prospective and interventional evidence before informing causal claims.

2. Methods

2.1. Data source

We used baseline data (March–August 2019) from the CBPS,

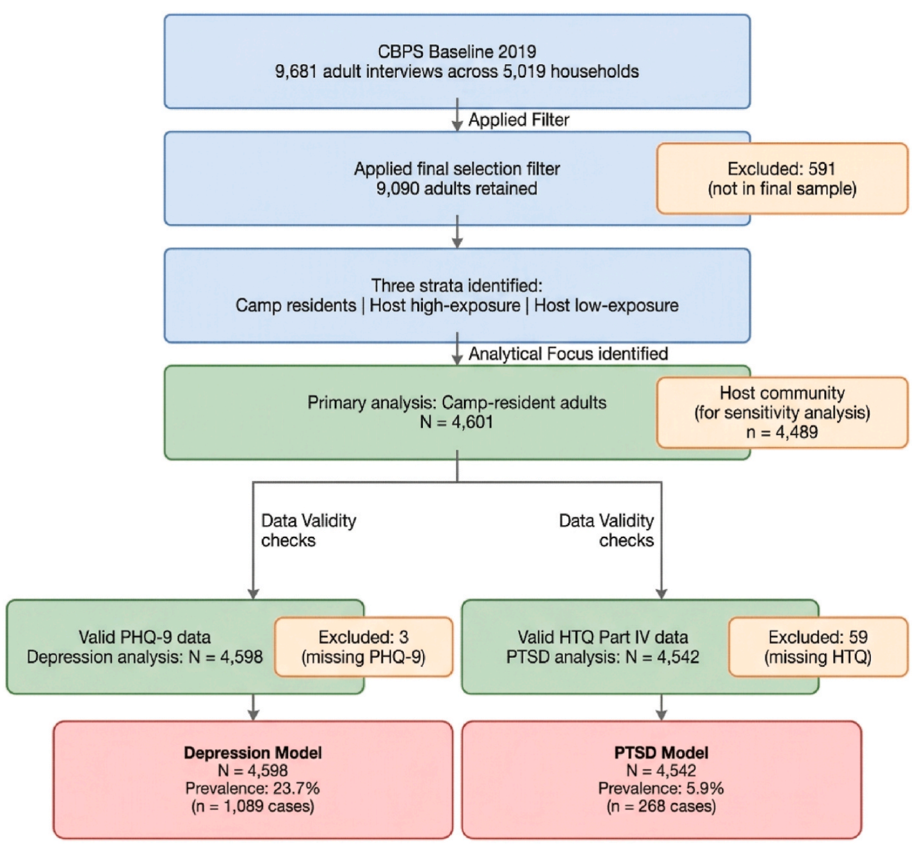

Fig. 1. Sample selection flowchart. CBPS baseline (2019) filtering from 9090 filtered adults to camp-resident analytic sample (N = 4598 with valid PHQ-9).

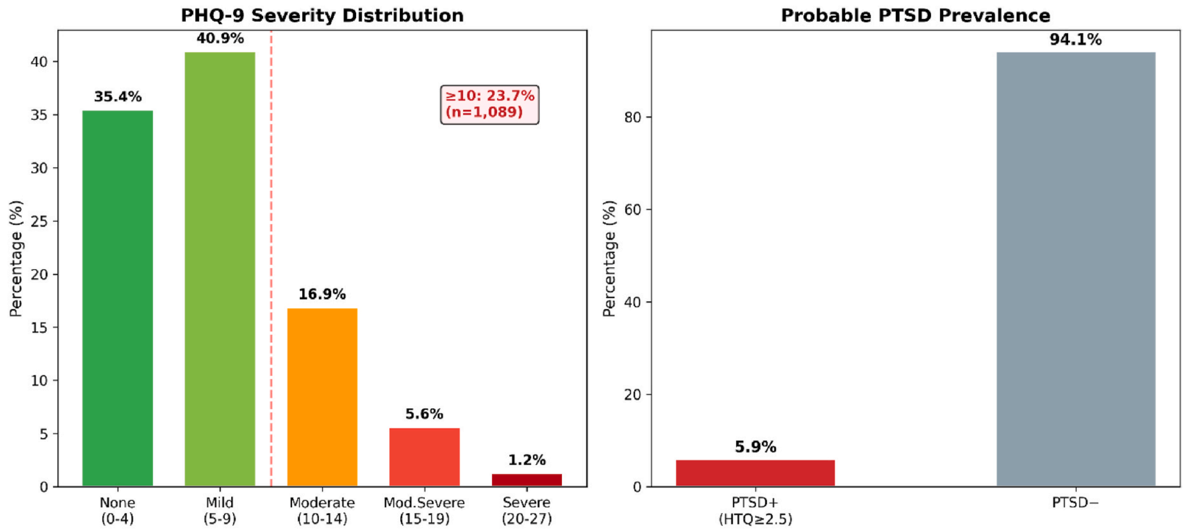

Fig. 2. Distribution of PHQ-9 severity categories and probable PTSD prevalence among camp-resident displaced adults (N = 4598). Left panel: PHQ-9 severity categories; 1089 adults (23.7%) *scored* ≥ 10, indicating moderate-or-greater depressive symptom burden. Right panel: 268 adults (5.9%) met the HTQ Part *IV* ≥ 2.5 threshold for probable PTSD. These estimates reflect screening-level classification, not clinical diagnoses.

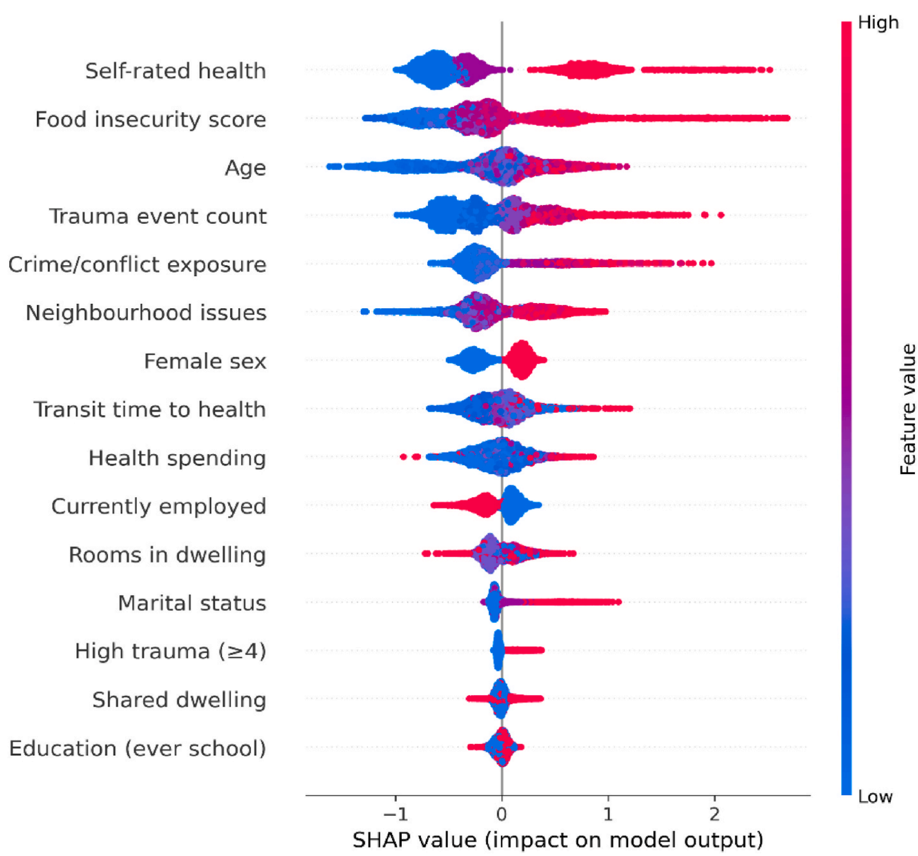

Fig. 3. SHAP feature importance for depression ($PHQ-9 \geq 10$). Self-rated health, food insecurity, and age are the strongest correlates.

implemented by Innovations for Poverty Action in partnership with Yale University, the World Bank, and the GAGE programme [15,18]. Data are available through the Harvard Dataverse (https://doi.org/10.7910/DVN/V614VB). We note that these data precede the COVID-19 pandemic, camp fires, Bhasan Char relocations, and subsequent shifts in conditions [33]; temporal applicability to the current situation is uncertain and represents a key limitation.

2.2. *Study population*

The CBPS surveyed 5019 households across three strata: camp-resident displaced Rohingya, high-exposure host communities (within 15 km), and low-exposure host communities. Adult interviews were conducted with two randomly selected household members aged ≥ 15 years (9681 total). After applying the recommended sample filter (*cbps_final_selection* $\in \{1, 2\}$), 9090 adults were retained: 4601 camp

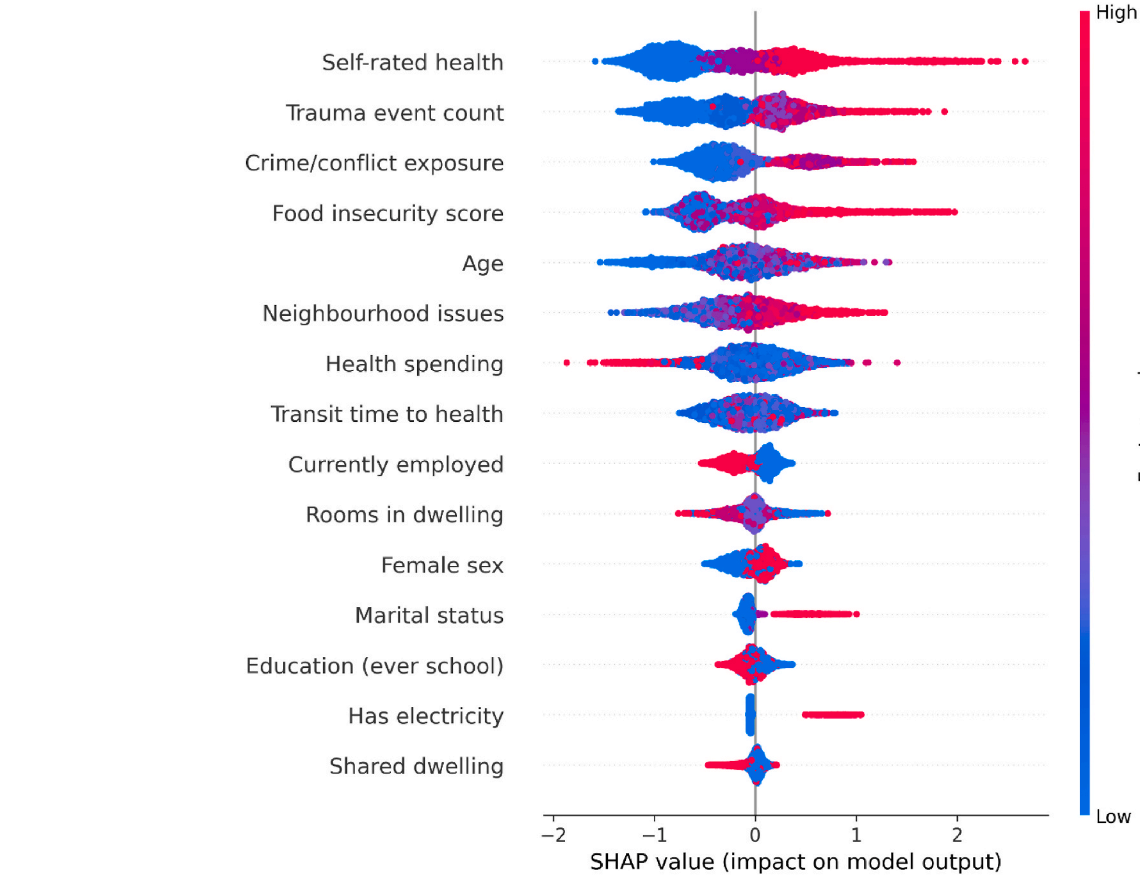

Fig. 4. SHAP feature importance for PTSD ($HTQ \geq 2.5$). Trauma, self-rated health, and crime/conflict exposure dominate.

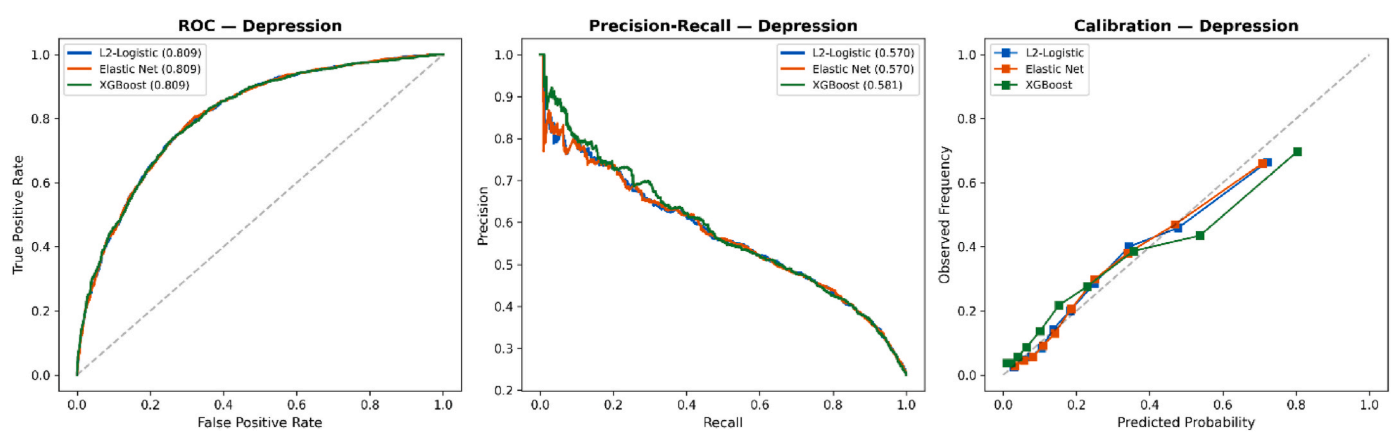

Fig. 5. ROC, precision-recall, and calibration curves for depression models. AUC ranges 0.809–0.819 across architectures.

residents and 4489 host community residents. The primary analysis used the camp sample ($N = 4,598$ with valid PHQ-9). The host community served as a secondary cross-population comparison.

2.3. Outcomes

The primary outcome was elevated depressive symptoms ($PHQ-9\ total \geq 10$). The PHQ-9 comprises nine items scored 0–3 (range 0–27); a threshold of 10 identifies moderate-or-greater symptom burden [31]. We emphasise this is a screening classification, not a clinical diagnosis. Prorated scores were computed when 7–8 of 9 items were answered.

The secondary outcome was elevated PTSD symptoms (HTQ Part IV mean ≥ 2.5). The HTQ Part IV comprises 16 items scored 1-4 [32]; the 2.5 threshold has been applied in prior Rohingya research [15]. We required ≥ 12 items for a valid mean.

Internal consistency was assessed via Cronbach's alpha; one-dimensionality was examined via single-factor analysis.

2.4. Predictors

Because the data are cross-sectional, all predictor-outcome associations are concurrent [30]. We use correlate rather than risk factor throughout. Seventeen candidate variables were drawn from six blocks:

Demographics (4): age, sex, marital status, ever attended school. Trauma history (2): HTQ Part I event count (12 events); binary ≥ 4 events. Post-displacement stressors (2): neighbourhood issue count;

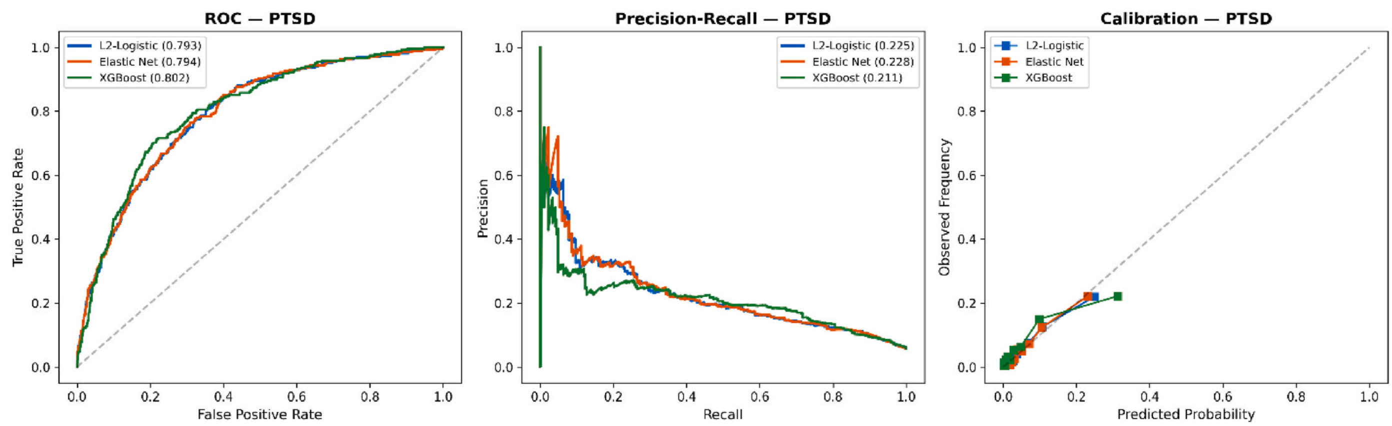

Fig. 6. ROC, precision-recall, and calibration curves for PTSD models. AUC ranges 0.793–0.802.

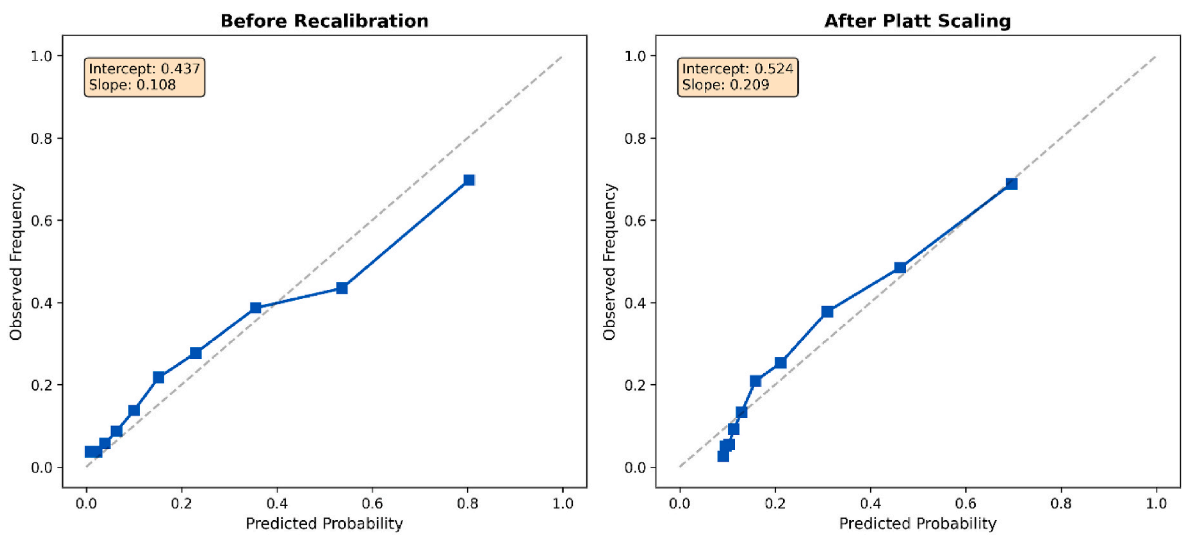

Fig. 7. Calibration before (left) and after Platt scaling (right). Slope improved from 0.11 to 0.21.

experienced crime/conflict count. Access to care (3): transit time to health facility; health spending; self-rated health (SRH). Economic vulnerability (1): current employment. Living conditions and food insecurity (5): rooms; electricity; shared dwelling; food insecurity count; binary food insecurity.

We are pre-specified that SRH may exhibit construct overlap with depression [34] and included a sensitivity analysis excluding it. Multicollinearity was assessed via variance inflation factors (VIF); all were below 5 (maximum 3.01), indicating no problematic collinearity. Variables not available in the CBPS, perceived social support, religious coping, prior psychiatric history, represent unmeasured confounders.

2.5. Addressing the triage tool paradox

This model is not designed to replace direct symptom screening (PHQ-9 = 9 questions). Its value lies in three use cases: 1) population-level stratification using administrative/survey data already collected for other purposes (food security monitoring, protection assessments), guiding where to deploy PHQ-9 screening; 2) identifying which contextual factors carry the strongest concurrent associations with symptom burden, informing programme design; and 3) testing whether ML complexity adds value over simpler approaches. In settings where systematic mental health screening is not yet implemented, which describes most humanitarian emergencies [5,6], risk profiling from existing data could guide resource allocation for screening deployment.

2.6. Missing data

Missingness was minimal: 94.3% complete cases; maximum single-variable missingness was 4.8% (transit time). Primary analysis used within-fold median imputation. Sensitivity analysis used iterative imputation (IterativeImputer, analogous to MICE [35] with 10 iterations per fold.

2.7. Model development

Three models per outcome:

L2-logistic regression (C = 1.0), transparent baseline. Elastic net (C = 0.1, L1 ratio = 0.5), penalised variable selection. XGBoost, hyperparameters tuned via grid search in inner CV folds ($n\_estimators \in \{100,\ 300\}$; $max\_depth \in \{3,\ 5\}$; $learning\_rate \in \{0.05,\ 0.1\}$; $min\_child\_weight \in \{5, 10\}$).

2.8. Validation strategy

Internal: Nested cross-validation with 5 household-grouped outer folds and 3 inner folds for tuning. Within-household PHQ-9 correlation was r = 0.40; grouped splitting prevents leakage [36].

Geographic (PSU-holdout): Twenty percent of primary sampling units (39/195 PSUs) were held out entirely from training. The model was trained on the remaining 80% and tested on the held-out PSUs

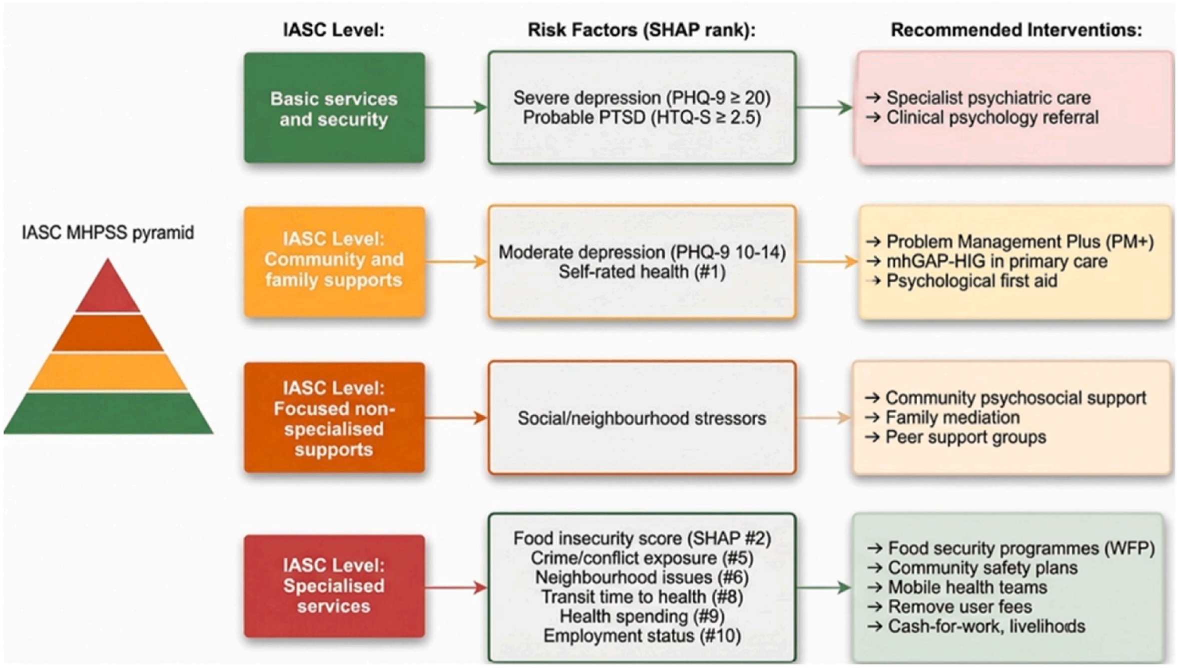

Fig. 8. Modifiable SHAP correlates mapped to the IASC MHPSS intervention pyramid.

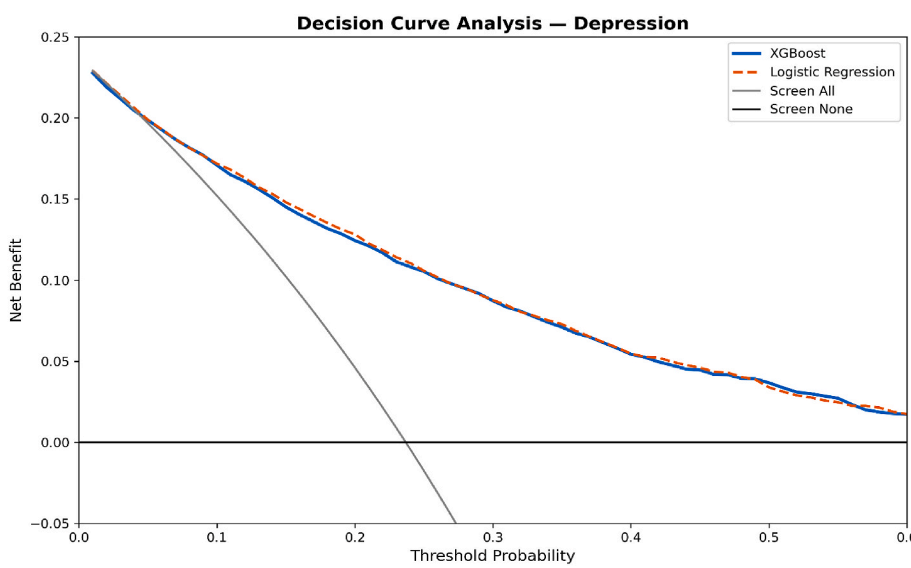

Fig. 9. Decision curve analysis for depression. Positive net benefit across all thresholds 1%–60%.

(N = 941), providing geographic validation within the camp population.

Cross-population: The full camp-trained model was tested on the host community (N = 4486) to assess transportability to a demographically different population. We note this is not independent external validation per TRIPOD + AI criteria [37], it shares survey methodology and time period, but tests population-level transferability.

Recalibration: Platt scaling [38] was applied within the CV framework to address observed miscalibration. Calibration plots are presented before and after recalibration.

2.9. Performance metrics

ROC-AUC with bootstrap 95% CIs (500 resamples); PR-AUC; Brier score; Brier Skill Score (BSS = 1 – Brier or Brier_null) [39]; calibration intercept and slope; sensitivity, specificity, PPV, NPV at 0.5 threshold. For PTSD, Youden-optimal and cost-sensitive thresholds (miss cost = 5× false positive) are reported given the 5.9% base rate [40].

2.10. Explainability

SHAP values [25] from the tuned XGBoost quantified feature importance. Feature contribution analysis has been shown to improve interpretability across diverse clinical domains [41]. Global summary plots, dependence plots for key modifiable correlates, and tertile-based risk stratification are presented.

2.11. Pre-specified sensitivity analyses

1) Excluding SRH (construct overlap); 2) survey-weighted logistic regression; 3) iterative imputation; 4) PSU-grouped CV; 5) excluding

trauma history; 6) modifiable correlates only; 7) continuous outcomes; 8) cross-population test on host community.

2.12. Ethical considerations

The CBPS data are publicly available and de-identified; no additional ethical approval was required.

2.13. Reporting and reproducibility

This study follows the TRIPOD + AI checklist [37] and clinical AI reporting guidelines [42] (supplement). We characterise this as a model development study with internal and geographic validation, not a fully validated tool. Analysis code is available at https://github.com/salisadegh/Symptoms-Among-War-Displaced-Adults.

2.14. Role of generative AI

Claude (Anthropic) assisted with Python code generation and initial manuscript drafting. All study design, analytical decisions, interpretation, and scientific claims were made by the authors. All text was critically reviewed, revised, and verified by the authors, who take full intellectual responsibility. AI was not used to generate, select, or alter results. Specific AI contributions: data processing scripts, model training pipelines, SHAP visualisation code, and paragraph-level drafts subsequently rewritten by authors.

3. Results

3.1. Sample characteristics

Among 4598 camp-resident adults, 56.0% were female, mean age was 33.7 years (SD 14.6), 57.5% had ever attended school, and 43.3% reported current employment. Food insecurity was pervasive (79.9% reporting $\geq$ 1 indicator; mean score 3.7/9). Only 4.3% had electricity. Mean trauma event count was 1.7 (SD 1.9); 15.6% reported $\geq$ 4 events. Mean transit time to health facility was 25 min. Missingness was minimal (maximum 4.8% for transit time; all other variables < 1%). Table 1 presents full characteristics.

Table 1
Sample characteristics of camp-resident displaced adults (N = 4598).

Variable	n (%) or Mean (SD)	Missing n (%)
Demographics		
Age (years)	33.7 (14.6)	0
Female sex	2575 (56.0%)	0
Ever attended school	2644 (57.5%)	0
Currently employed	1992 (43.3%)	0
Trauma and conflict exposure		
Trauma event count	1.7 (1.9)	0
High trauma ($\geq$4 events)	717 (15.6%)	0
Crime/conflict exposure count	0.8 (1.2)	0
Neighbourhood issues count	1.4 (1.7)	0
Food security		
Food insecurity score (0–9)	3.7 (2.8)	0
Any food insecurity	3674 (79.9%)	0
Health access and housing		
Self-rated health (1–5)	3.2 (0.9)	12 (0.3%)
Transit time to health facility (min)	25.0 (18.4)	221 (4.8%)
Health spending (BDT)	482.3 (1247.5)	15 (0.3%)
Rooms in dwelling	1.3 (0.6)	0
Has electricity	198 (4.3%)	0
Shared dwelling	1847 (40.2%)	0
Mental health outcomes		
PHQ-9 total score	6.75 (4.53)	0
Depression (PHQ-9 $\geq$10)	1089 (23.7%)	0
HTQ-IV mean score	1.68 (0.46)	56 (1.2%)
PTSD (HTQ $\geq$2.5)	268 (5.9%)	56 (1.2%)
Instrument reliability		
Cronbach α PHQ-9	0.830	
Cronbach α HTQ-IV	0.872	

BDT = Bangladeshi Taka. PHQ-9 = Patient Health Questionnaire-9. HTQ = Harvard Trauma Questionnaire Part IV. Food insecurity score based on 9 binary indicators from the CBPS food security module. Self-rated health scored 1 (excellent) to 5 (poor).

3.2. Instrument reliability and factor structure

Cronbach's alpha was 0.830 for the PHQ-9 and 0.872 for the HTQ Part IV, indicating good-to-excellent internal consistency [43]. Single-factor analysis confirmed unidimensional structure for both instruments, consistent with their intended use as summary screening scores. These reliability estimates are comparable to those reported in other South Asian refugee populations [44].

3.3. Prevalence of elevated symptoms

Among camp adults, 23.7% (n = 1089) exceeded the PHQ-9 $\geq$ 10 threshold (mean 6.75, SD 4.53). By severity: 36.5% none/minimal, 30.2% mild, 18.3% moderate, 10.3% moderately severe, 4.8% severe. Elevated PTSD symptoms ($HTQ \geq 2.5$) were present in 5.9% (n = 268; mean 1.68, SD 0.46). These are screening-level estimates and may differ from clinical diagnostic prevalence [45].

The depression figure aligns with the meta-analytic estimate of ~22% in conflict-affected populations [1].

3.4. Model performance

Depression, internal validation. Tuned XGBoost: AUC = 0.819 (95% CI 0.805–0.832); BSS = 0.252; Brier = 0.139. L2-logistic regression: AUC = 0.809 (0.795–0.823). Elastic net: AUC = 0.809 (0.797–0.823). The near equivalence confirms that concurrent associations are predominantly linear and additive, consistent with findings from Christodoulou et al. [28] and van der Ploeg et al. [29].

Depression, geographic validation (PSU-holdout). AUC = 0.814 (N = 941 from 39 held-out PSUs). This minimal degradation from internal validation (0.819 to 0.814) indicates geographic stability within the camp population.

Depression, cross-population (host community). AUC = 0.698 (N = 4486; prevalence 30.1%). The substantial decline reflects genuine population-level differences (displacement history, socioeconomic composition, service environment) and underscores limited transportability to non-camp settings. This is expected and does not invalidate within-camp risk profiling.

Calibration. Raw model calibration showed systematic underconfidence (intercept = 0.44, slope = 0.11). After Platt scaling [38], calibration improved (intercept = 0.52, slope = 0.21) though remained imperfect. This underscores that absolute risk estimates require population-specific recalibration before operational use [39].

PTSD, internal validation. XGBoost: AUC = 0.802 (0.777–0.829). At the 0.5 threshold, sensitivity was 0.06, clinically useless. At the Youden-optimal threshold (0.047), sensitivity = 0.716, specificity = 0.779, PPV = 0.169. At the cost-sensitive threshold, sensitivity = 0.959, specificity = 0.332. Threshold selection must weigh over-referral costs against missed-case consequences [40].

3.5. Key correlates (SHAP analysis)

For depression, the top correlates by mean $|SHAP|$ were self-rated health (0.628), food insecurity (0.479), age (0.373), trauma count (0.367), crime/conflict (0.319), neighbourhood issues (0.300), female sex (0.219), transit time to health (0.181), health spending (0.177), employment (0.143). Dependence plots (Supplement) show monotonic relationships for food insecurity and trauma, supporting additive interpretation. Top-feature correlations were low (maximum r = 0.33 between age and SRH), indicating distinct information contributions.

For PTSD, rankings were similar: SRH, trauma, crime/conflict, food insecurity, age, neighbourhood issues, health spending, transit time, employment, rooms.

Among the top 10 for both outcomes, 6 of 10 are potentially modifiable: food insecurity, healthcare access (transit, cost), neighbourhood safety, crime/conflict, and employment. Modifiable correlate denotes concurrent association; causal modifiability requires interventional evidence.

3.6. Construct overlap sensitivity

Excluding SRH reduced depression AUC from 0.819 to 0.782 and PTSD from 0.802 to 0.769. The remaining predictor ranking was stable. We recommend the SRH-excluded model (AUC = 0.782) as the conservative benchmark, free from construct overlap concerns. Even this estimate substantially exceeds chance and retains clinical utility [46].

3.7. Risk stratification and screening efficiency

Tertile-based groups for depression showed: low-risk (33%) prevalence 0.8%; medium (33%) 7.0%; high (34%) 62.1%. Screening only the top tertile captures 72% of cases; NNS = 2.0. In humanitarian contexts where universal PHQ-9 screening is infeasible, this targeted approach could improve detection efficiency by directing limited screening resources to the highest-burden subgroup [7]. We note this is concurrent classification efficiency, not prospective yield, and requires longitudinal validation.

3.8. Further sensitivity analyses

Results were robust. Weighted logistic: AUC = 0.807. MICE-like imputation: AUC = 0.809. PSU-grouped CV: AUC = 0.807. Without trauma: AUC = 0.794 (depression), 0.767 (PTSD). Modifiable only: AUC = 0.714 (depression), 0.685 (PTSD). Continuous outcome: $R^2 = 0.37$ (PHQ-9), 0.26 (HTQ).

3.9. Policy mapping

We map modifiable SHAP correlates to the IASC pyramid [9] as testable hypotheses, not established causal pathways:

Basic services: Food security programmes (correlate #2); mobile health teams and fee removal (correlates #8, #9); community safety (#6); livelihoods support (#10). These align with WHO [5] and UNHCR [40,47,48] guidance.

Community supports: Psychosocial support, peer networks, family strengthening. Informed by neighbourhood stressor findings.

Focused non-specialised: For moderate symptom burden, Problem Management Plus (PM+) [47], a five-session scalable intervention, and mhGAP-HIG [49] integration.

Specialised services: For severe cases (PHQ $\geq$ 20, HTQ $\geq$ 2.5), specialist referral where available (see Figs. 10–12).

4. Discussion

4.1. Summary

Using 17 concurrent indicators from a representative displacement

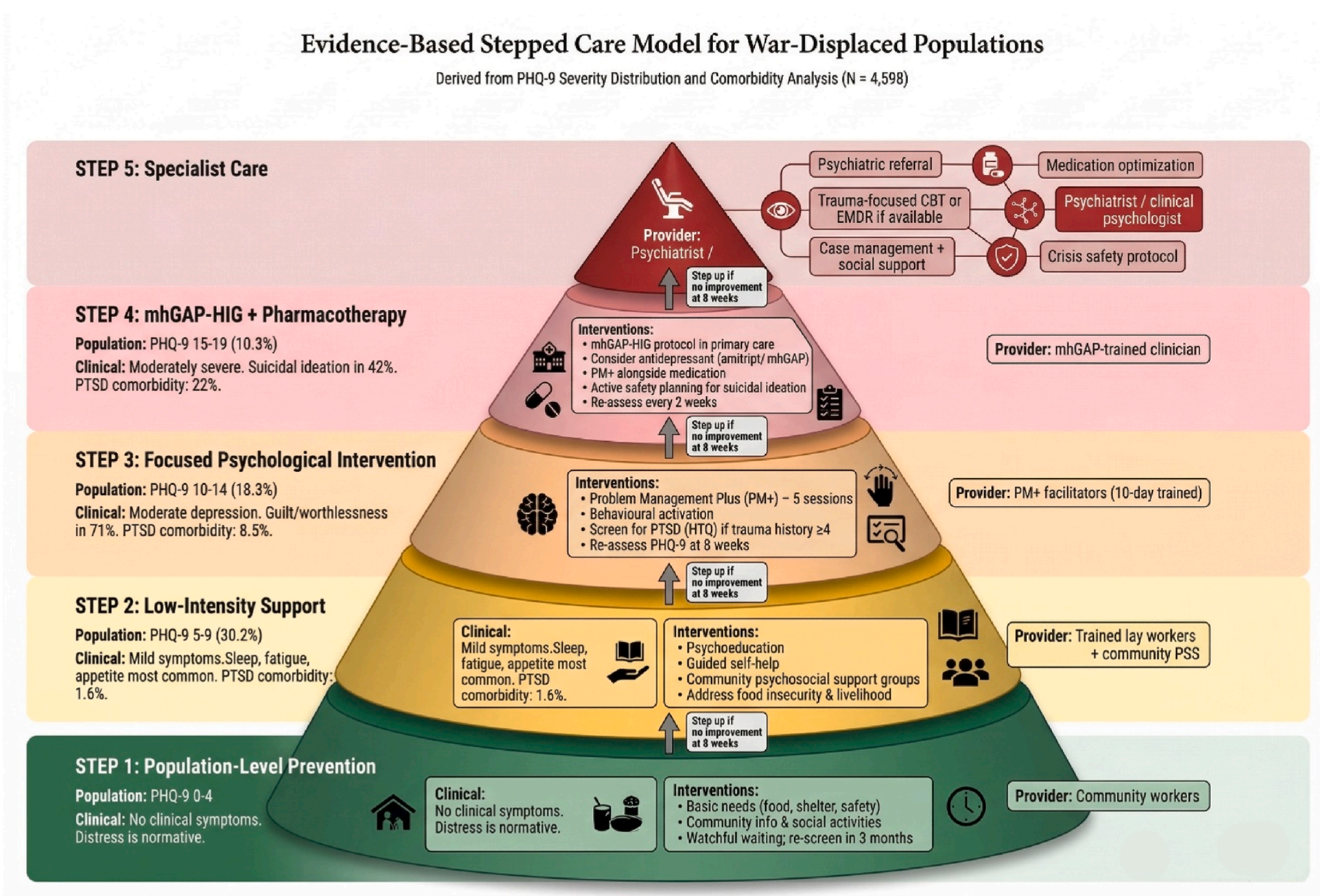

Fig. 10. Stepped care model derived from PHQ-9 severity distribution and comorbidity analysis.

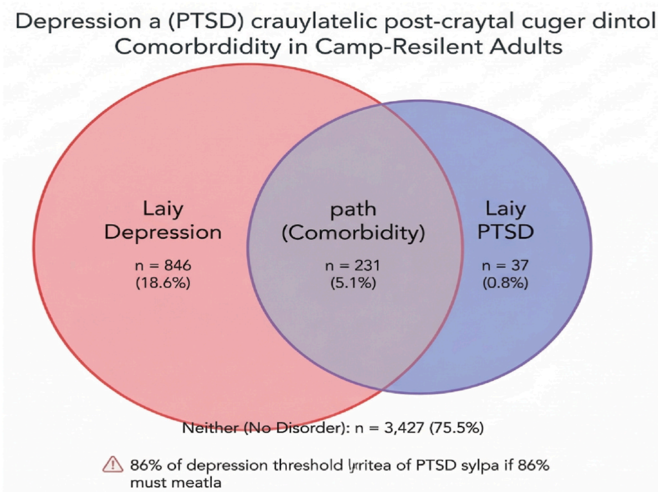

Fig. 11. Depression–PTSD comorbidity. 86% of PTSD cases also meet depression threshold.

survey, we demonstrate that mental health symptom burden can be meaningfully stratified among camp-resident Rohingya adults. All model architectures achieved similar AUC (≈0.81 for depression; 0.80 for PTSD), confirming predominantly additive associations. Geographic validation within camp PSUs was strong (AUC = 0.814), while cross-population transportability to the host community was limited (AUC = 0.698). Risk tertiles showed a 78-fold prevalence gradient (0.8% vs 62.1%), and targeted screening of the top tertile captures 72% of elevated-symptom cases.

4.2. *What this study does and does not claim*

We are explicit about the boundaries of inference. This is a model development study based on cross-sectional data [30,37]. It identifies concurrent correlates of elevated symptom burden and quantifies their relative contributions. It does not: predict future symptom onset; establish causal pathways; provide a validated triage tool ready for deployment; or demonstrate that modifying the identified correlates will reduce symptom burden. These claims require, respectively: longitudinal cohorts, randomised experiments, prospective implementation studies, and intervention trials. Our policy mapping generates hypotheses for these next steps, grounded in the existing IASC framework [9].

4.3. *Model equivalence as a central finding*

The identical performance of logistic regression and XGBoost extends prior work showing ML rarely outperforms regression for structured clinical data [26,28,29]. In humanitarian settings, this has a practical implication: the simplest model suffices, reducing implementation barriers. We acknowledge this finding is not novel per se [28], but its demonstration in a displacement-affected population adds context-specific evidence.

4.4. *Geographic versus cross-population validation*

The PSU-holdout AUC of 0.814 (vs 0.819 internal) demonstrates that within the camp population, the model is geographically stable, not overfit to specific camp blocks. The host community AUC of 0.698 reflects genuine population-level differences rather than methodological failure. This dual finding carries an important message: the model works well within its target population but should not be assumed transportable without local recalibration [39].

We acknowledge this does not satisfy TRIPOD + AI criteria for fully independent external validation [37]. No temporally separate CBPS cohort was available for this analysis. Future work should apply this framework to any future survey waves that include mental health modules, or to independent datasets from other displacement settings.

4.5. *Data vintage and contemporary relevance*

The 2019 data predate COVID-19, campfires, Bhasan Char relocations, and evolving political dynamics [33]. We cannot assume stable predictor–outcome associations. However: 1) the analytical framework, using existing survey data for symptom burden stratification, is transferable regardless of specific parameter estimates; 2) the broad predictor domains (food security, safety, healthcare access) are identified across displacement contexts [2,10–12], not specific to 2019 Cox's Bazar; and 3) replication with more recent data is explicitly recommended as a priority. We frame this as a methodological proof-of-concept, not a contemporary needs assessment.

4.6. *Construct overlap*

SRH is the strongest correlate, but self-rated health and depression share common-method variance (both self-report) and conceptual overlap (somatisation, negative affectivity) [34]. Our conservative model excluding SRH (AUC = 0.782) represents the floor estimate. That food insecurity, trauma, and crime/conflict retain strong contributions

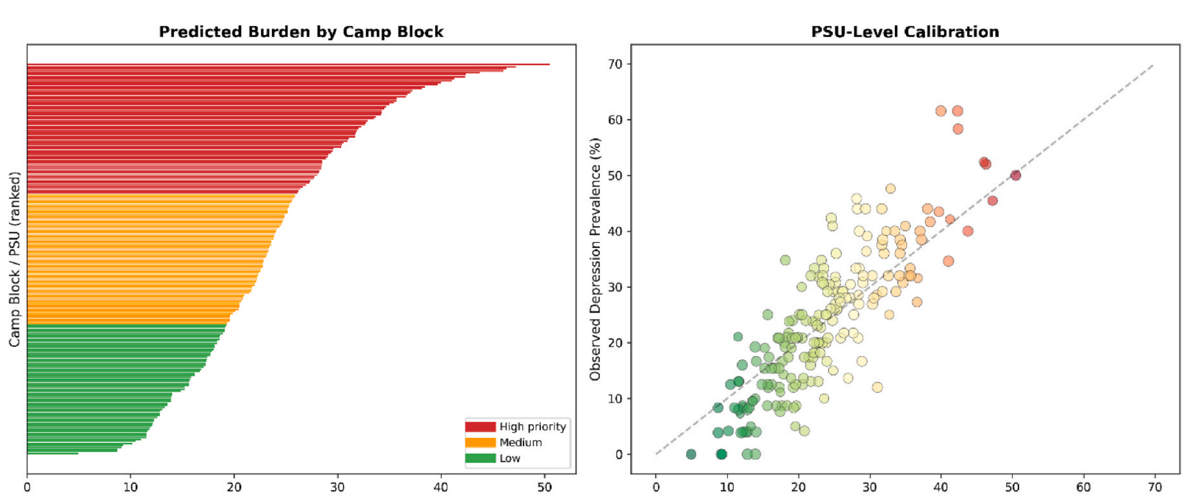

Fig. 12. Geographic priority map by camp block (PSU). Left: ranked predicted burden. Right: predicted vs observed prevalence.

without SRH confirms that the model captures genuine contextual associations, not merely a psychometric artefact.

4.7. Calibration and operational readiness

The observed miscalibration (slope 0.11 raw, 0.21 after Platt scaling) means absolute predicted probabilities should not be used for individual-level decision-making without population-specific recalibration [39]. The model's value lies in risk ranking (discrimination), not risk estimation. For screening triage (high vs low priority), relative ranking suffices; for clinical decision-making, recalibration to the local population is required.

4.8. Comparison with prior work

Our depression screening prevalence (23.7%) aligns with Charlson et al.'s meta-analytic 22.1% [1] and CBPS-based estimates [15,18]. PTSD at 5.9% is lower than some Rohingya studies [16,17], reflecting the conservative HTQ $\geq$ 2.5 threshold. Our methodological contribution, ML with SHAP, IASC mapping, NNS calculation, geographic validation, is incremental rather than transformative. The advance lies in operationalising risk profiling with transparent sensitivity analyses, not in methodological novelty.

4.9. Implications for humanitarian practice

If prospectively validated, this framework could inform two operational questions in humanitarian settings: 1) Where to screen: Deploying PHQ-9/HTQ screening to camp blocks or subpopulations with the highest predicted symptom burden, guided by routinely collected survey data, could improve detection efficiency over universal screening [7]. 2) What to prioritise: The prominence of food insecurity, healthcare access barriers, and safety concerns as correlates supports the IASC emphasis on basic services as the foundation of MHPSS response [9]. This aligns with WHO's building-back-better framework [6] and evidence for PM + [47] and mhGAP-HIG [49].

We note that cost-effectiveness analysis, comparing targeted versus universal screening per case identified, and assessing whether this exceeds GDP-per-capita thresholds, was beyond the scope of this study and represents a critical next step.

4.10. Ethical implications of misclassification

In humanitarian contexts, both Type I errors (false positives causing unnecessary referrals) and Type II errors (false negatives missing genuine distress) carry consequences. False positives burden an already overstretched system; false negatives leave vulnerable individuals unserved. The choice of operating threshold must balance these costs locally, ideally through decision-analytic methods incorporating community-specific values [40].

4.11. Strengths

1) Large representative sample (N = 4598) with validated instruments ($\alpha > 0.80$) [36]. 2) Genuine nested CV with inner-loop tuning. 3) Geographic (PSU-holdout) validation demonstrating within-population stability. 4) Comprehensive metrics including BSS, calibration, and cost-sensitive thresholds. 5) Eight sensitivity analyses including SRH exclusion and MICE imputation. 6) Transparent framing of cross-sectional limitations throughout.

4.12. Limitations

1) Cross-sectional design, the fundamental limitation precluding prognostic or causal claims [30]. 2) No prospective external validation, PSU-holdout is geographic validation within a single cross-section; temporal validation (future CBPS waves) is needed. 3) Screening instruments, not diagnoses, PHQ-9 and HTQ have not been formally validated in Rohingya; reliability was satisfactory but construct validity requires further study [44,45]. 4) Data vintage (2019), conditions have changed substantially [33]; current applicability is uncertain. 5) SRH construct overlap, conservative model excluding SRH provides the floor estimate (AUC = 0.782). 6) Limited predictor set, social support, coping, religious practices, prior psychiatric history were unavailable. 7) Triage paradox, the model requires more items than PHQ-9; value is in population-level stratification, not individual screening replacement. 8) Calibration imperfect, recalibration needed before absolute risk estimation [39]. 9) No cost-effectiveness analysis, screening efficiency (NNS) is reported but formal economic evaluation was not conducted.

4.13. Future directions

Priority next steps: 1) temporal validation using future survey rounds with mental health modules in Cox's Bazar or comparable displacement settings; 2) prospective study assessing whether high-risk classification predicts sustained or worsening symptoms at 6–12 months; 3) implementation research on risk-stratified screening versus universal screening; 4) cross-cultural validation of PHQ-9 and HTQ in the Rohingya language; 5) cost-effectiveness analysis of targeted screening strategies.

5. Conclusion

Routinely collected displacement survey indicators can meaningfully stratify concurrent depressive and PTSD symptom burden among war-displaced adults. Targeted screening of the highest-risk tertile captures 72% of elevated-symptom cases. Model performance is geographically stable within the camp population (PSU-holdout AUC = 0.814) but shows limited cross-population transportability (host AUC = 0.698). This work should be understood as a development study generating testable hypotheses for future prospective validation and implementation research, not a ready-to-deploy triage tool.

CRediT authorship contribution statement

Seyed-Ali Sadegh-Zadeh: Conceptualization, Formal analysis, Investigation, Project administration, Software, Supervision, Writing – review & editing, Writing – original draft. **Shayan Saadat:** Conceptualization, Formal analysis, Investigation, Writing – review & editing, Writing – original draft. **Zeynab Sadeghimanesh:** Formal analysis, Writing – review & editing. **Ommolbanin Soleymani:** Writing – review & editing, Visualization, Writing – original draft. **Alireza Soleimani Mamalo:** Writing – original draft, Validation, Visualization. **Sanam Anoosheh:** Writing – review & editing, Formal analysis, Visualization. **Mohammadreza Shalbafan:** Writing – original draft, Visualization, Software. **Elham Khalilian:** Writing – review & editing, Investigation. **Seyed-Yaser Mousavi:** Writing – review & editing, Formal analysis.

Data availability

Data will be made available on request.

References

[1] Charlson F, Van Ommeren M, Flaxman A, Cornett J, Whiteford H, Saxena S. New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis. Lancet 2019;394(10194):240–8.
[2] Steel Z, Chey T, Silove D, Marnane C, Bryant RA, Van Ommeren M. Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: a systematic review and meta-analysis. JAMA 2009;302(5):537–49.
[3] Tol WA, et al. Mental health and psychosocial support in humanitarian settings: linking practice and research. Lancet 2011;378(9802):1581–91.

[4] Miller KE, Rasmussen A. The mental health of civilians displaced by armed conflict: an ecological model of refugee distress. Epidemiol Psychiatr Sci 2017;26(2): 129–38.
[5] Organization WH. Mental Health Atlas 2020. World Health Organization; 2021.
[6] Organization WH. Building Back Better: Sustainable Mental Health Care After Emergencies. World Health Organization; 2013.
[7] Ventevogel P, van Ommeren M, Schilperoord M, Saxena S. Improving mental health care in humanitarian emergencies. SciELO Pub Health 2015.
[8] Marshall C. The inter-agency standing committee (IASC) guidelines on mental health and psychosocial support (MHPSS) in emergency settings: a critique. Int Rev Psychiatr 2022;34(6):604–12.
[9] Committee I-AS. IASC Guidelines on Mental Health and Psychosocial Support in Emergency Settings. IASC; 2007. https://interagencystandingcommittee.org/iasc-task-force-mental-health-and-psychosocial-support-emergency-settings/iasc-guidelines-mental-health-and-psychosocial-support-emergency-settings-2007.
[10] Li SSY, Liddell BJ, Nickerson A. The relationship between post-migration stress and psychological disorders in refugees and asylum seekers. Curr Psychiatry Rep 2016; 18(9):82.
[11] Porter M, Haslam N. Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: a meta-analysis. JAMA 2005;294(5):602–12.
[12] Miller KE, Rasmussen A. War exposure, daily stressors, and mental health in conflict and post-conflict settings: bridging the divide between trauma-focused and psychosocial frameworks. Soc Sci Med 2010;70(1):7–16.
[13] Silove D, Ventevogel P, Rees S. The contemporary refugee crisis: an overview of mental health challenges. World Psychiatry 2017;16(2):130–9.
[14] Nickerson A, Liddell BJ, Maccallum F, Steel Z, Silove D, Bryant RA. Posttraumatic stress disorder and prolonged grief in refugees exposed to trauma and loss. BMC Psychiatry 2014;14(1):106.
[15] Genoni ME, Khan AU, Krishnaswamy N, Palaniswamy N, Raza W. The Cox's Bazar Panel Survey: Baseline Data on Displaced Rohingya and their Host Population. Washington DC: World Bank, Poverty and Equity Global Practice; 2021 [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/574491613543058730.
[16] Riley A, Varner A, Ventevogel P, Taimur Hasan MM, Welton-Mitchell C. Daily stressors, trauma exposure, and mental health among stateless Rohingya refugees in Bangladesh. Transcult Psychiatry 2017;54(3):304–31.
[17] Riley A, Akther Y, Noor M, Ali R, Welton-Mitchell C. Systematic human rights violations, traumatic events, daily stressors and mental health of Rohingya refugees in Bangladesh. Conflict Health 2020;14(1):60.
[18] Lopez-Pena P, et al. Cox's Bazar Panel Survey. Harvard Dataverse 2021. https://doi.org/10.7910/DVN/V614VB.
[19] Chekroud AM, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016;3(3):243–50.
[20] Kessler RC, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatr 2016;21(10):1366–71.
[21] Nemesure MD, V Heinz M, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 2021;11(1):1980.
[22] Su C, Xu Z, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. Transl Psychiatry 2020;10(1):116.
[23] Sadegh-Zadeh S-A, Nazari M-J, Aljamaeen M, Yazdani FS, Mousavi SY, Vahabi Z. Predictive models for Alzheimer's disease diagnosis and MCI identification: the use of cognitive scores and artificial intelligence algorithms. NPG Neurologie-Psychiatrie-Gériatrie 2024;24(142):194–211.
[24] Jourahmad Z, et al. Machine learning techniques for predicting the short-term outcome of resective surgery in lesional-drug resistance epilepsy. arXiv preprint arXiv:2302.10901 2023.
[25] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
[26] Sadegh-Zadeh S-A, et al. Precision diagnostics in cardiac tumours: integrating echocardiography and pathology with advanced machine learning on limited data. Inform Med Unlocked 2024;49:101544.
[27] Sohrabi MA, Zare-Mirakabad F, Ghidary SS, Saadat M, Sadegh-Zadeh S-A. A novel data augmentation approach for influenza A subtype prediction based on HA proteins. Comput Biol Med 2024;172:108316.
[28] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110: 12–22.
[29] Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14(1):137.
[30] Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? Bmj 2009;338.
[31] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–13.
[32] Mollica RF, Caspi-Yavin Y, Bollini P, Truong T, Tor S, Lavelle J. The Harvard Trauma Questionnaire: validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. J Nerv Ment Dis 1992;180(2):111–6.
[33] Bhatia A, et al. The Rohingya in Cox's Bazar: when the stateless seek refuge. Health Hum Rights 2018;20(2):105.
[34] Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med 2009;69(3):307–16.
[35] Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Software 2011;45:1–67.
[36] Steyerberg EW, Harrell Jr FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2015;69:245.
[37] Collins GS, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. bmj 2024;385.
[38] Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 1999;10(3):61–74.
[39] Steyerberg EW. Clinical Prediction Models, vol 201. Springer; 2019. p. 9.
[40] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak 2006;26(6):565–74.
[41] Sadegh-Zadeh S-A, Sadeghzadeh N, Soleimani O, Ghidary SS, Movahedi S, Mousavi S-Y. Comparative analysis of dimensionality reduction techniques for EEG-based emotional state classification. Am J Neurodegener Dis 2024;13(4):23.
[42] Sadegh-Zadeh SA, Bagheri M. Harnessing the power of clinical data in dentistry: importance and guidelines for dentists in AI modelling for enhanced patient care. Open J Clin Med Images 2024;4(1):1188.
[43] Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011;2: 53.
[44] Tay AK, et al. The culture, mental health and psychosocial wellbeing of Rohingya refugees: a systematic review. Epidemiol Psychiatr Sci 2019;28(5):489–94.
[45] Fazel M, Wheeler J, Danesh J. Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review. The lancet 2005;365 (9467):1309–14.
[46] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5(9):1315–6.
[47] Dawson KS, et al. Problem management plus (PM+): a WHO transdiagnostic psychological intervention for common mental health problems. World Psychiatry 2015;14(3):354.
[48] UNHCR. Global Trends: Forced Displacement in 2022. Geneva: United Nations High Commissioner for Refugees; 2023 [Online]. Available: https://www.unhcr.org/sites/default/files/2023-06/global-trends-report-2022.pdf.
[49] Organization WH. mhGAP Humanitarian Intervention Guide (mhGAP-HIG): Clinical Management of Mental, Neurological And Substance Use Conditions in Humanitarian Emergencies. World Health Organization; 2015.