

Supplementary document

Supplementary Table S1. Detailed definitions and coding of variables used in the machine-learning models

Variable Name	Description	Type
Predictor variables		
Feature Description	Type	
Age (years)	Raw value	Continuous
Gender	(0=Male, 1=Female)	Binary
Marital status	(0=Single, 1=In a relationship/Engaged, 2=Married)	Categorical
Displaced during conflict	(0=No, 1=Yes)	Binary
Left country during conflict	(0=No, 1=Yes)	Binary
Number of individuals living with you	Raw value	Continuous
Fear of your own and family safety	(0=No, 1=Yes)	Binary
Private counseling/therapy	(0=No, 1=Yes)	Binary
Take medications for mental health	(0=No, 1=Yes)	Binary
Family members injured due to the conflict	(0=No, 1=Yes)	Binary
Rely on religious/spiritual practices	(0=Not at all to 4=To a very large extent)	Ordinal
Cigarette smoking	(0=Do not practice, 1=Reduced, 2=No change, 3=Increased)	Categorical
Shisha smoking	(0=Do not practice, 1=Reduced, 2=No change, 3=Increased)	Categorical
Alcohol consumption	(0=Do not practice, 1=Reduced, 2=No change, 3=Increased)	Categorical
Physical activity	(0=Do not practice, 1=Reduced, 2=No change, 3=Increased)	Categorical
Healthy eating	(0=Do not practice, 1=Reduced, 2=No change, 3=Increased)	Categorical
Number of hours sleep per day	(0=<7 hrs/day, 1=7-9 hrs/day, 2=>9 hrs/day)	Categorical
Time spent on social media during the conflict	(0=Decreased, 1=No change, 2=Increased)	Categorical
Coping responses (RFSS avoidance and approach)	Raw value	Continuous
Outcome variables		
Depression (PHQ_Binary_ML_10)	(0 = PHQ-9 score < 10 (Low–mild depressive symptoms), 1 = PHQ-9 score ≥ 10 (Moderate to severe depressive symptoms))	Binary
Depression (PHQ_Binary_ML_15)	(0 = PHQ-9 < 15 (Low-moderate depressive symptoms), 1 = PHQ-9 ≥ 15 (Moderately severe to severe depressive symptoms))	Binary
Anxiety (GAD_7_Binary_ML_10)	(0 = GAD-7 < 10 (Low-mild anxiety symptoms), 1 = GAD-7 ≥ 10 (Moderate to severe anxiety symptoms))	Binary

Anxiety (GAD_7_Binary_ML_15)	(0 = GAD-7 < 15 (Low-moderate anxiety symptoms), 1 = GAD-7 ≥ 15 (Severe anxiety symptoms))	Binary
Stress (PSS_Binary_ML_20)	(0 = PSS-10 < 20 (Below-median perceived stress), 1 = PSS-10 ≥ 20 (Above-median elevated perceived stress))	Binary
Stress (PSS_Binary_ML_27)	(0 = PSS-10 < 27 (Low–moderate perceived stress), 1 = PSS-10 ≥ 27 (High perceived stress))	Binary
Stress (PSS_Categorical_3)	(0 = PSS-10 ≤ 13 (low stress), 1 = PSS-10: 14–26 (Mild-moderate stress), 2 = PSS-10 ≥ 27 (High stress))	Categorical

Supplementary Table S2. Hyperparameter search grids used for model tuning

Algorithm	Hyperparameter	Values evaluated
Random Forest	Number of trees	100, 200
	Maximum tree depth	None, 10, 20
	Minimum samples required to split an internal node	2, 5
	Minimum samples required at leaf nodes	1, 2
	Number of features considered at each split	sqrt, log2, 0.5
XGBoost	Number of estimators	100, 200
	Maximum depth of individual trees	3, 6
	Learning rate	0.01, 0.1
Support Vector Machine	Regularization strength (C)	0.1, 1, 10
	Kernel function	linear, rbf
	Gamma parameter	scale, auto
Logistic Regression	Penalty type	none, l1, l2
	Regularization strength (C)	0.01, 0.1, 1, 10
Decision Tree	Maximum depth	None, 10, 20
K-Nearest Neighbors	Number of neighbors	3, 5, 7
Naïve Bayes	Variance smoothing	1e-9, 1e-8
AdaBoost	Number of estimators	50, 100, 200
	Learning rate	0.01, 0.1, 1
Multi-Layer Perceptron	Hidden-layer configuration	(50,), (100,), (100, 100)
	Activation function	relu, tanh

Abbreviations: l1, least absolute shrinkage and selection operator penalty; l2, ridge penalty; rbf, radial basis function.

Supplementary Table S3. Best-performing models across balanced and SMOTE sampling for all binary outcomes

Outcome	Approach	Best model (by PR-AUC)	PR-AUC	F1	Recall	Accuracy
PHQ_Binary_ML_10	Balanced	Random Forest	0.932	0.839	0.929	0.778
PHQ_Binary_ML_10	SMOTE	Random Forest	0.920	0.833	0.893	0.778
PHQ_Binary_ML_15	Balanced	Random Forest	0.486	0.435	0.357	0.711
PHQ_Binary_ML_15	SMOTE	Random Forest	0.502	0.414	0.429	0.622
GAD_7_Binary_ML_10	Balanced	Random Forest	0.879	0.746	0.815	0.667
GAD_7_Binary_ML_10	SMOTE	XGBoost	0.844	0.787	0.889	0.711
GAD_7_Binary_ML_15	Balanced	Naïve Bayes	0.645	0.491	0.813	0.400
GAD_7_Binary_ML_15	SMOTE	Support Vector Machine	0.686	0.625	0.625	0.733
PSS_Binary_ML_20	Balanced	AdaBoost	0.954	0.882	0.909	0.822
PSS_Binary_ML_20	SMOTE	Random Forest	0.939	0.836	0.848	0.756
PSS_Binary_ML_27	Balanced	Random Forest	0.716	0.571	0.500	0.800
PSS_Binary_ML_27	SMOTE	Logistic Regression	0.759	0.645	0.833	0.756

Supplementary Table S4. Full binary model comparison for all cut-offs, algorithms, and sampling strategies

PHQ_Binary_ML_10_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.778	0.765	0.929	0.839	0.884	0.932
XGBoost	Balanced	0.711	0.692	0.964	0.806	0.785	0.881
Support Vector Machine	Balanced	0.733	0.735	0.893	0.806	0.800	0.888
Logistic Regression	Balanced	0.689	0.694	0.893	0.781	0.819	0.900
Decision Tree	Balanced	0.644	0.731	0.679	0.704	0.633	0.696
K-Nearest Neighbors	Balanced	0.733	0.767	0.821	0.793	0.794	0.845
Naïve Bayes	Balanced	0.533	1.000	0.250	0.400	0.777	0.876
AdaBoost	Balanced	0.756	0.758	0.893	0.820	0.803	0.834
MLP Neural Net	Balanced	0.711	0.759	0.786	0.772	0.765	0.846

PHQ_Binary_ML_10_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.778	0.781	0.893	0.833	0.853	0.920
XGBoost	SMOTE	0.689	0.733	0.786	0.759	0.790	0.886
Support Vector Machine	SMOTE	0.689	0.769	0.714	0.741	0.786	0.876
Logistic Regression	SMOTE	0.800	0.852	0.821	0.836	0.824	0.899
Decision Tree	SMOTE	0.644	0.714	0.714	0.714	0.622	0.688
K-Nearest Neighbors	SMOTE	0.733	0.900	0.643	0.750	0.868	0.903
Naïve Bayes	SMOTE	0.533	0.818	0.321	0.462	0.744	0.814

AdaBoost	SMOTE	0.733	0.786	0.786	0.786	0.850	0.913
MLP Neural Net	SMOTE	0.711	0.742	0.821	0.780	0.758	0.858

PHQ_Binary_ML_15_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.711	0.556	0.357	0.435	0.647	0.486
XGBoost	Balanced	0.622	0.385	0.357	0.370	0.565	0.427
Support Vector Machine	Balanced	0.622	0.364	0.286	0.320	0.539	0.340
Logistic Regression	Balanced	0.600	0.375	0.429	0.400	0.548	0.353
Decision Tree	Balanced	0.556	0.333	0.429	0.375	0.521	0.321
K-Nearest Neighbors	Balanced	0.533	0.267	0.286	0.276	0.571	0.360
Naïve Bayes	Balanced	0.644	0.250	0.071	0.111	0.521	0.371
AdaBoost	Balanced	0.689	0.500	0.500	0.500	0.584	0.413
MLP Neural Net	Balanced	0.556	0.333	0.429	0.375	0.502	0.374

PHQ_Binary_ML_15_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.622	0.400	0.429	0.414	0.615	0.502
XGBoost	SMOTE	0.600	0.400	0.571	0.471	0.598	0.395
Support Vector Machine	SMOTE	0.689	0.500	0.714	0.588	0.647	0.438
Logistic Regression	SMOTE	0.667	0.474	0.643	0.545	0.629	0.441
Decision Tree	SMOTE	0.556	0.350	0.500	0.412	0.502	0.311
K-Nearest Neighbors	SMOTE	0.511	0.367	0.786	0.500	0.510	0.318
Naïve Bayes	SMOTE	0.644	0.375	0.214	0.273	0.537	0.422
AdaBoost	SMOTE	0.622	0.435	0.714	0.541	0.570	0.355
MLP Neural Net	SMOTE	0.533	0.316	0.429	0.364	0.493	0.355

GAD_7_Binary_ML_10_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.667	0.688	0.815	0.746	0.763	0.879
XGBoost	Balanced	0.533	0.575	0.852	0.687	0.559	0.746
Support Vector Machine	Balanced	0.644	0.641	0.926	0.758	0.757	0.852
Logistic Regression	Balanced	0.622	0.614	1.000	0.761	0.794	0.863
Decision Tree	Balanced	0.600	0.680	0.630	0.654	0.593	0.650
K-Nearest Neighbors	Balanced	0.733	0.714	0.926	0.806	0.730	0.746
Naïve Bayes	Balanced	0.511	0.727	0.296	0.421	0.685	0.794

AdaBoost	Balanced	0.600	0.622	0.852	0.719	0.678	0.783
MLP Neural Net	Balanced	0.644	0.704	0.704	0.704	0.691	0.823

GAD_7_Binary_ML_10_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.578	0.643	0.667	0.655	0.700	0.839
XGBoost	SMOTE	0.711	0.706	0.889	0.787	0.737	0.844
Support Vector Machine	SMOTE	0.600	0.680	0.630	0.654	0.689	0.788
Logistic Regression	SMOTE	0.622	0.692	0.667	0.679	0.658	0.726
Decision Tree	SMOTE	0.511	0.619	0.481	0.542	0.519	0.609
K-Nearest Neighbors	SMOTE	0.622	0.708	0.630	0.667	0.679	0.706
Naïve Bayes	SMOTE	0.533	0.750	0.333	0.462	0.679	0.764
AdaBoost	SMOTE	0.644	0.657	0.852	0.742	0.745	0.838
MLP Neural Net	SMOTE	0.556	0.652	0.556	0.600	0.650	0.786

GAD_7_Binary_ML_15_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.689	0.667	0.250	0.364	0.714	0.642
XGBoost	Balanced	0.689	0.625	0.313	0.417	0.664	0.598
Support Vector Machine	Balanced	0.689	0.600	0.375	0.462	0.670	0.587
Logistic Regression	Balanced	0.711	0.636	0.438	0.519	0.707	0.616
Decision Tree	Balanced	0.622	0.467	0.438	0.452	0.581	0.404
K-Nearest Neighbors	Balanced	0.622	0.467	0.438	0.452	0.568	0.443
Naïve Bayes	Balanced	0.400	0.351	0.813	0.491	0.664	0.645
AdaBoost	Balanced	0.667	0.556	0.313	0.400	0.755	0.621
MLP Neural Net	Balanced	0.667	0.545	0.375	0.444	0.588	0.567

GAD_7_Binary_ML_15_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.733	0.750	0.375	0.500	0.716	0.654
XGBoost	SMOTE	0.644	0.500	0.313	0.385	0.688	0.582
Support Vector Machine	SMOTE	0.733	0.625	0.625	0.625	0.720	0.686
Logistic Regression	SMOTE	0.733	0.625	0.625	0.625	0.731	0.591
Decision Tree	SMOTE	0.667	0.529	0.563	0.545	0.643	0.453
K-Nearest Neighbors	SMOTE	0.533	0.400	0.625	0.488	0.582	0.490

Naïve Bayes	SMOTE	0.422	0.361	0.813	0.500	0.647	0.665
AdaBoost	SMOTE	0.711	0.579	0.688	0.629	0.721	0.588
MLP Neural Net	SMOTE	0.689	0.563	0.563	0.563	0.588	0.559

PSS_Binary_ML_20_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.778	0.829	0.879	0.853	0.773	0.907
XGBoost	Balanced	0.733	0.800	0.848	0.824	0.813	0.932
Support Vector Machine	Balanced	0.711	0.738	0.939	0.827	0.765	0.920
Logistic Regression	Balanced	0.756	0.806	0.879	0.841	0.838	0.941
Decision Tree	Balanced	0.711	0.833	0.758	0.794	0.670	0.809
K-Nearest Neighbors	Balanced	0.622	0.767	0.697	0.730	0.649	0.824
Naïve Bayes	Balanced	0.333	1.000	0.091	0.167	0.732	0.855
AdaBoost	Balanced	0.822	0.857	0.909	0.882	0.866	0.954
MLP Neural Net	Balanced	0.689	0.828	0.727	0.774	0.750	0.893

PSS_Binary_ML_20_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.756	0.824	0.848	0.836	0.828	0.939
XGBoost	SMOTE	0.711	0.833	0.758	0.794	0.760	0.905
Support Vector Machine	SMOTE	0.711	0.813	0.788	0.800	0.707	0.868
Logistic Regression	SMOTE	0.267	0.000	0.000	0.000	0.500	0.733
Decision Tree	SMOTE	0.622	0.786	0.667	0.721	0.583	0.768
K-Nearest Neighbors	SMOTE	0.600	0.895	0.515	0.654	0.691	0.834
Naïve Bayes	SMOTE	0.311	0.667	0.121	0.205	0.722	0.835
AdaBoost	SMOTE	0.778	0.871	0.818	0.844	0.833	0.937
MLP Neural Net	SMOTE	0.711	0.857	0.727	0.787	0.750	0.886

PSS_Binary_ML_27_Balanced

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	Balanced	0.800	0.667	0.500	0.571	0.833	0.716
XGBoost	Balanced	0.800	0.636	0.583	0.609	0.770	0.568
Support Vector Machine	Balanced	0.667	0.385	0.417	0.400	0.672	0.443
Logistic Regression	Balanced	0.822	0.750	0.500	0.600	0.803	0.646
Decision Tree	Balanced	0.778	0.571	0.667	0.615	0.742	0.470
K-Nearest Neighbors	Balanced	0.800	0.667	0.500	0.571	0.721	0.588
Naïve Bayes	Balanced	0.422	0.316	1.000	0.480	0.807	0.540

AdaBoost	Balanced	0.711	0.455	0.417	0.435	0.783	0.575
MLP Neural Net	Balanced	0.778	0.625	0.417	0.500	0.773	0.519

PSS_Binary_ML_27_SMOTE

Model	Approach	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Random Forest	SMOTE	0.822	0.700	0.583	0.636	0.826	0.593
XGBoost	SMOTE	0.822	0.643	0.750	0.692	0.870	0.681
Support Vector Machine	SMOTE	0.756	0.524	0.917	0.667	0.225	0.214
Logistic Regression	SMOTE	0.756	0.526	0.833	0.645	0.881	0.759
Decision Tree	SMOTE	0.689	0.450	0.750	0.563	0.708	0.404
K-Nearest Neighbors	SMOTE	0.600	0.375	0.750	0.500	0.711	0.492
Naïve Bayes	SMOTE	0.511	0.353	1.000	0.522	0.818	0.536
AdaBoost	SMOTE	0.711	0.462	0.500	0.480	0.770	0.480
MLP Neural Net	SMOTE	0.800	0.615	0.667	0.640	0.806	0.559

Supplementary Table S5. Multiclass classification performance for three-level perceived stress (PSS-10 low, moderate, and high), by algorithm and sampling strategy

PSS_Multiclass_ML_Balanced

Model	Approach	F1_macro	Recall_macro	Precision_macro
Random Forest	Balanced	0.530	0.524	0.538
XGBoost	Balanced	0.498	0.528	0.490
Support Vector Machine	Balanced	0.537	0.548	0.531
Logistic Regression	Balanced	0.513	0.532	0.507
Gradient Boosting	Balanced	0.520	0.520	0.520
Decision Tree	Balanced	0.409	0.429	0.400
K-Nearest Neighbors	Balanced	0.560	0.575	0.557
Naive Bayes	Balanced	0.437	0.560	0.540
MLP Neural Net	Balanced	0.489	0.492	0.491

PSS_Multiclass_ML_SMOTE

Model	Approach	F1_macro	Recall_macro	Precision_macro
Random Forest	SMOTE	0.608	0.631	0.600
XGBoost	SMOTE	0.517	0.556	0.520
Support Vector Machine	SMOTE	0.536	0.536	0.536
Logistic Regression	SMOTE	0.448	0.484	0.449
Gradient Boosting	SMOTE	0.538	0.548	0.535
Decision Tree	SMOTE	0.512	0.532	0.510

K-Nearest Neighbors	SMOTE	0.406	0.452	0.397
Naive Bayes	SMOTE	0.451	0.548	0.550
MLP Neural Net	SMOTE	0.557	0.575	0.565