

# Explainable Machine Learning for Predicting Depression Risk Using Sleep Patterns, Self-Esteem, and Psychosocial Factors

Muslihat Adejoke Gaffari  
East Tennessee State  
University  
USA

**Abstract:** Depression presents a substantial public health challenge, shaped by interacting behavioural, psychological and social conditions. Earlier identification of vulnerability requires approaches that capture these relationships while remaining interpretable and appropriate for the populations assessed. This proposed study will examine whether sleep patterns, self-esteem and psychosocial factors improve depression prediction beyond demographic and health-related characteristics. Validated measures will characterise sleep quality, duration and regularity, self-esteem, perceived stress, loneliness and social support. An interpretable regression baseline will be compared with selected machine learning algorithms using domain-specific and combined predictor sets. Model development will incorporate missing-data procedures, leakage prevention, tuning and independent evaluation. Discrimination, calibration, prediction error and subgroup performance will assess reliability, while complementary explanation methods will examine influential features, nonlinear patterns and explanation stability. Analyses will address overlap between sleep predictors and depression questionnaire items. Longitudinal data will support future-risk prediction; concurrent measurements will support identification of elevated symptoms only. Findings will clarify the incremental value and limitations of combining predictor domains, without treating explanations as causal evidence. External validation and evaluation of clinical usefulness will remain necessary before routine implementation.

**Keywords:** Depression risk; Sleep patterns; Self-esteem; Psychosocial factors; Explainable machine learning; Predictive modelling.

## 1. INTRODUCTION

### 1.1 Depression Risk as a Multidimensional Prediction Problem

Depression can impair daily functioning, disrupt relationships and increase demands on healthcare, with consequences extending beyond individual symptoms to families, education and employment [1]. Identifying people who may need further assessment is therefore important, although useful identification requires an appropriate response pathway rather than a numerical score alone. Symptom severity, duration and presentation vary, as do the social circumstances in which difficulties arise.

Screening identifies possible symptoms warranting assessment; diagnosis involves clinical evaluation; prediction estimates an outcome from information available at a defined time. Validated questionnaires can support symptom assessment but should not be conflated with comprehensive diagnostic evaluation [3]. Likewise, a prediction model cannot independently establish a clinical diagnosis or determine treatment. Transparent reporting must specify the target population, outcome and intended use [2]. This distinction matters when evaluating models that classify existing symptoms while being described as forecasting future depression, particularly when predictors and outcomes are measured simultaneously.

### 1.2 Connecting Sleep Patterns, Self-Esteem and Psychosocial Conditions

Sleep duration, regularity, perceived quality and disturbances provide complementary descriptions of sleep behaviour. Prospective research has linked insomnia symptoms with subsequent depression treatment, supporting investigation of sleep as a predictive domain without establishing that every sleep characteristic has equivalent relevance [4]. Self-esteem adds information about self-evaluation; longitudinal findings suggest that lower self-esteem can precede depressive symptoms in some populations [5].

Psychosocial circumstances include social support, loneliness, perceived stress and adverse experiences. Longitudinal evidence connects loneliness with subsequent depressive symptoms [6], while observational research associates childhood adversity with adult depression [7]. These findings justify measurement, but their applicability depends on population, instruments and study design.

Potential interactions warrant examination: stressful circumstances may coincide with disrupted sleep, while supportive relationships may accompany greater resilience. However, depressive symptoms may themselves influence sleep, self-perception and relationships, making causal direction uncertain in concurrent data. Analysing isolated predictors can miss combinations of vulnerability and protection. Combined modelling will therefore test complementary information while accounting for overlapping constructs, correlated measurements and the possibility that variables contribute little value.

### 1.3 Study Questions and Intended Contribution

The proposed study will focus on adults represented in the selected dataset and evaluate prediction within its documented assessment setting. It will ask whether combining sleep, self-esteem and psychosocial domains improves performance beyond demographic and health-related baselines, and whether explanations remain stable across individuals, subgroups and validation samples. Feature-attribution methods can describe contributions to model predictions, but they do not establish causal mechanisms [8].

The intended contribution will be an interpretable, rigorously evaluated approach with explicit uncertainty and limitations. Longitudinal observations will permit assessment of future onset or symptom escalation; concurrent measurements will support identification of elevated symptoms only. Consistent with transparent reporting principles [2], the prediction target, measurement timing and intended use will be defined before algorithm selection. These decisions will establish how outputs could responsibly inform further assessment.

## 2. DEFINING THE PREDICTION TARGET AND INTENDED USE

### 2.1 Population, Prediction Moment and Outcome Horizon

The proposed population will comprise adults aged eighteen years or older in the selected dataset, subject to documented consent, measurement availability and eligibility requirements. Recruitment setting, geographical coverage and demographic composition will be reported once the source is confirmed. Exclusions will be justified because restrictions based on language, complete questionnaires or follow-up availability may narrow applicability. Prediction research requires explicit attention to participant selection and intended use [13].

The prediction moment will occur after eligible baseline measurements become available and before any future outcome assessment. With longitudinal data, the protocol will specify the follow-up interval before analysis. Incident depression will concern participants without the defined condition initially; symptom worsening will concern change from baseline severity. Prognostic research distinguishes a defined starting point from subsequent endpoints [9].

If measurements are concurrent, the task will remain classification of elevated symptoms. Intended users will be professionals considering further evaluation or referral, not automated diagnosis. Clinical evaluation will need to examine how predictions influence actual decisions within the proposed service pathway before implementation [12].

### 2.2 Depression Measurement and Outcome Construction

Outcome construction will depend on the validated measure actually available. If the dataset contains the PHQ-8, its eight items will be scored from zero to three, producing totals from zero to twenty-four; a prespecified threshold of ten will identify elevated depressive symptoms [10]. This

questionnaire-based classification will not be described as a clinically confirmed depressive disorder.

The primary outcome will be binary elevated symptoms, with continuous severity examined secondarily where justified. Longitudinal analyses will define incident threshold crossing separately from prespecified symptom worsening. Incomplete responses will follow instrument guidance and a documented missing-data plan; unanswered items will not automatically represent absent symptoms. Outcome definition and analysis decisions will be assessed for potential bias [13].

Because the questionnaire includes sleep disturbance, using sleep predictors may partly reproduce outcome content. Sensitivity analyses will compare results after excluding overlapping predictors and examining a sleep-item-excluded symptom score. Any modified score will be labelled exploratory, without assuming that the original validated threshold remains appropriate. These checks will distinguish predictive information from measurement circularity and prevent misleading claims about sleep as an independent explanatory mechanism [8].

### 2.3 Boundaries of Responsible Use

Predictive classification identifies statistical patterns rather than underlying causes; feature contributions explain model outputs, not the effects of changing personal circumstances [8]. Consent arrangements, lawful access, confidentiality and restricted handling of sensitive responses will follow applicable governance requirements. WHO guidance emphasises autonomy, privacy, transparency and accountability in health artificial intelligence [11].

Model-generated estimates will support professional assessment rather than determine diagnosis or treatment. Where researchers contact participants directly, an approved response protocol will specify trained personnel, timely assessment and escalation of concerning disclosures. Participants will receive clear information about monitoring arrangements and confidentiality limits.

Deployment will require external validation and evaluation within the intended response pathway, including human oversight and consequences of errors [12]. Prediction performance alone will not establish treatment benefit. With these boundaries defined, subsequent analysis will organise sleep, self-esteem and psychosocial measurements into interpretable domains and establish their quality, timing and suitability for the specified prediction task.

## 3. BUILDING A DOMAIN-BASED ANALYTICAL DATASET

### 3.1 Study Design, Recruitment and Data Provenance

The study will use an observational dataset containing depression outcomes and relevant predictor domains. Because no source has been confirmed, recruitment setting, sampling method, observation dates and follow-up availability will be specified before analysis. Questionnaires, interviews, clinical

records and wearable measurements will be distinguished rather than treated as interchangeable sources. Participant selection and outcome ascertainment will be examined using prediction-study bias principles [13].

Eligibility will follow Section 2, with exclusions and repeated observations documented in a participant flow record. Age distribution, socioeconomic composition, recruitment access and attrition will inform assessment of selection bias. The applicable ethics approval, waiver or secondary-analysis determination will be reported without presuming exemption; sensitive information will require appropriate governance [11]. A versioned data dictionary, processing scripts and exclusion log will preserve provenance. Reproducible workflows will support scrutiny of transformations and potential leakage across analytical stages [17].

### 3.2 Sleep Pattern Features

Sleep duration will represent estimated time asleep; timing will describe sleep onset, awakening or midpoint; regularity will capture variation across observed nights. Quality and disturbances will reflect validated questionnaire responses where available. The Pittsburgh Sleep Quality Index offers a recognised assessment of subjective sleep characteristics, but will be used only if present in the dataset [14].

Self-reported sleep will remain distinguishable from device-derived estimates. Actigraphy infers sleep from movement and is not equivalent to polysomnographic measurement [18]. Device model, algorithms, recording coverage and valid-night rules will therefore be documented.

Aggregation windows will be prespecified and limited to information available before prediction. Weekday–weekend differences will be retained where coverage permits. Shift work, medication, caregiving and other contextual influences will be considered when recorded. Missing recordings will not automatically indicate normal sleep. Recall error, nonwear and algorithmic misclassification will be acknowledged. Duration will be examined flexibly rather than assuming that longer sleep or one universal threshold is protective across participants and circumstances.

### 3.3 Self-Esteem and Psychosocial Features

Self-esteem will be operationalised using the instrument available in the selected source. If the Rosenberg Self-Esteem Scale is present, negatively worded items will be reverse-scored and responses combined according to its documented scoring convention [15]. The coding scheme and score direction will be stated, avoiding comparisons between differently coded totals.

Psychosocial domains will distinguish perceived support, subjective loneliness, perceived stress, relationship circumstances and relevant adverse experiences. Where available, a validated Perceived Stress Scale will quantify perceived unpredictability, uncontrollability and overload rather than objective exposure alone [16]. Instrument

versions, recall periods and population suitability will be reviewed before harmonisation.

Construct validity will be considered alongside language, cultural interpretation and response patterns. Related constructs will not be collapsed merely because they correlate. Sensitive responses may reflect recall, social desirability or reluctance to disclose; confidentiality safeguards will support appropriate handling [11]. Protective resources, including supportive relationships and coping opportunities, will accompany adversity measures where recorded. Predictor inclusion will follow the research questions and measurement quality, rather than availability or association with depression alone.

### 3.4 Covariates, Missingness and Leakage Controls

Covariates will include age, recorded sex or gender, education, employment, income and relevant health characteristics where justified and available. Categories, reference groups and score directions will be prespecified. Continuous variables will retain information, with nonlinear transformations considered within model development rather than arbitrary categorisation. Participant and predictor definitions will be reviewed for applicability concerns [13].

Missingness will be described by domain, outcome group and relevant participant characteristics. Imputation, scaling, encoding and feature selection will be fitted within training partitions and applied to validation data without refitting. Such separation addresses leakage that can inflate apparent performance [17].

Depression questionnaire totals and constituent outcome items will be excluded from predictor matrices. Post-prediction diagnoses, treatment changes and later measurements will also be excluded. Repeated observations from the same participant, and linked observations where identifiable, will remain together during partitioning. Source identifiers will support grouping without becoming predictors. Harmonisation rules, exclusions, implausible values and corrections will be logged. These safeguards will establish the dataset’s analytical boundaries before testing whether combined domains provide information beyond simpler alternatives, while preserving a reproducible record of processing decisions.

**Table 1. Predictor domains, operational definitions and measurement safeguards.**

| Domain         | Variable or instrument                  | Operational definition                              | Measurement timing             | Expected analytical role             | Main limitation or leakage concern                 |
|----------------|-----------------------------------------|-----------------------------------------------------|--------------------------------|--------------------------------------|----------------------------------------------------|
| Sleep duration | Questionnaire, diary or device estimate | Estimated hours asleep per night; aggregated over a | Available by prediction moment | Capture sleep quantity and nonlinear | Recall error; device misclassification; incompatib |

| Domain                         | Variable or instrument                     | Operational definition                                                | Measurement timing                                                      | Expected analytical role                         | Main limitation or leakage concern                                             |
|--------------------------------|--------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------------------------------|
|                                |                                            | prespecified window                                                   |                                                                         | patterns                                         | le measurement windows                                                         |
| Sleep timing and regularity    | Sleep onset, awakening and midpoint        | Typical timing, nightly variability and weekday – weekend differences | Pre-prediction recording; concurrent for cross-sectional classification | Describe timing and consistency                  | Sparse recordings, shift work and nonwear may distort estimates                |
| Sleep quality and disturbances | PSQI or available validated measure        | Instrument-defined scores or prespecified components                  | Recall period ending by prediction moment                               | Capture perceived quality and sleep difficulties | Overlap with the depression outcome's sleep item                               |
| Self-esteem                    | Rosenberg scale, if available              | Total following documented reverse-scoring and coding                 | Baseline or concurrent assessment                                       | Represent global self-evaluation                 | Cultural interpretation; response style; overlap with negative self-perception |
| Perceived stress               | Available validated Perceived Stress Scale | Version-specific score using the documented recall period             | Available by prediction moment                                          | Represent perceived demands and coping imbalance | Subjective appraisal may be influenced by current symptoms                     |
| Social support and loneliness  | Available validated instruments            | Separate measures of perceived support and subjectiv                  | Baseline or concurrent assessment                                       | Assess social resources and perceived disconn    | Related constructs are not interchangeable; shared reporting                   |

| Domain                      | Variable or instrument                               | Operational definition                                                      | Measurement timing                                     | Expected analytical role                               | Main limitation or leakage concern                                     |
|-----------------------------|------------------------------------------------------|-----------------------------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------------------|
|                             |                                                      | isolation                                                                   |                                                        | ction                                                  | bias                                                                   |
| Relationships and adversity | Structured questionnaire, interview or records       | Prespecified relationship circumstances and adverse experiences             | Events occurring before prediction                     | Characterise psychosocial context                      | Sensitive disclosure, recall error and incomplete event timing         |
| Background and covariates   | Demographic, socioeconomic and health variables      | Prespecified coding, units and reference groups                             | Available by prediction moment                         | Establish comparison models and contextual information | Inappropriate proxies, missingness and post-outcome information        |
| Depression outcome          | Validated symptom instrument or confirmed assessment | Prespecified elevated-symptom classification and secondary severity measure | Follow-up for prognosis; concurrent for classification | Target variable only                                   | Outcome totals and constituent items must not enter predictor matrices |

## 4. TESTING THE ADDED PREDICTIVE VALUE OF EACH DOMAIN

### 4.1 Nested Predictor Sets and Prespecified Comparisons

A common baseline will contain justified demographic, socioeconomic and health-related covariates available at prediction. Four additions will be evaluated: baseline plus sleep, baseline plus self-esteem, baseline plus psychosocial measures, and the fully combined set. Psychosocial measures will retain their construct definitions; perceived stress, for example, will not be treated as interchangeable with objective adversity [16].

Domain-removal analyses will compare the full model with versions omitting sleep, self-esteem or psychosocial information. These comparisons will estimate conditional predictive contributions rather than independent biological

effects. Correlated domains may contain overlapping information, making apparent importance dependent on the remaining predictors and the explanation method [24].

All feature sets will use identical participant partitions, outcome definitions and evaluation procedures. Comparisons will use a common evaluation sample so differences in missingness do not masquerade as predictive gains. Training resources and tuning opportunities will remain comparable. Preprocessing will respect partition boundaries to prevent leakage [17]. Differences will be reported with uncertainty and measurement burden. Improved prediction will not establish that changing sleep, self-esteem or social circumstances would change depression outcomes; intervention effects require a different research design.

#### 4.2 Candidate Algorithms and Model Complexity

Penalised logistic regression will provide an interpretable baseline for the primary binary outcome. Continuous predictor effects will be represented using prespecified flexible terms where justified. A limited comparison will include random

Algorithm selection will reflect sample size, outcome frequency, predictor dimensions and intended interpretation. Regression penalties, tree depth, minimum leaf size, learning rate and boosting iterations will constrain complexity as appropriate. Hyperparameter ranges will be prespecified and tuned within development data.

Regression interactions will be limited to justified hypotheses; tree ensembles may learn interactions implicitly, subject to stability checks. Nonlinear models will not be presumed superior merely because they are flexible. Selection and performance assessment must remain separated to avoid optimistic estimates [21]. Highly complex architectures will be excluded unless sufficient data and a clear methodological rationale support their inclusion. Computational burden and reproducibility will inform the comparison.

#### 4.3 Training, Tuning and Independent Evaluation

Model development will be separated from final evaluation before preprocessing. Where observations repeat, all records from one participant will remain within the same partition.

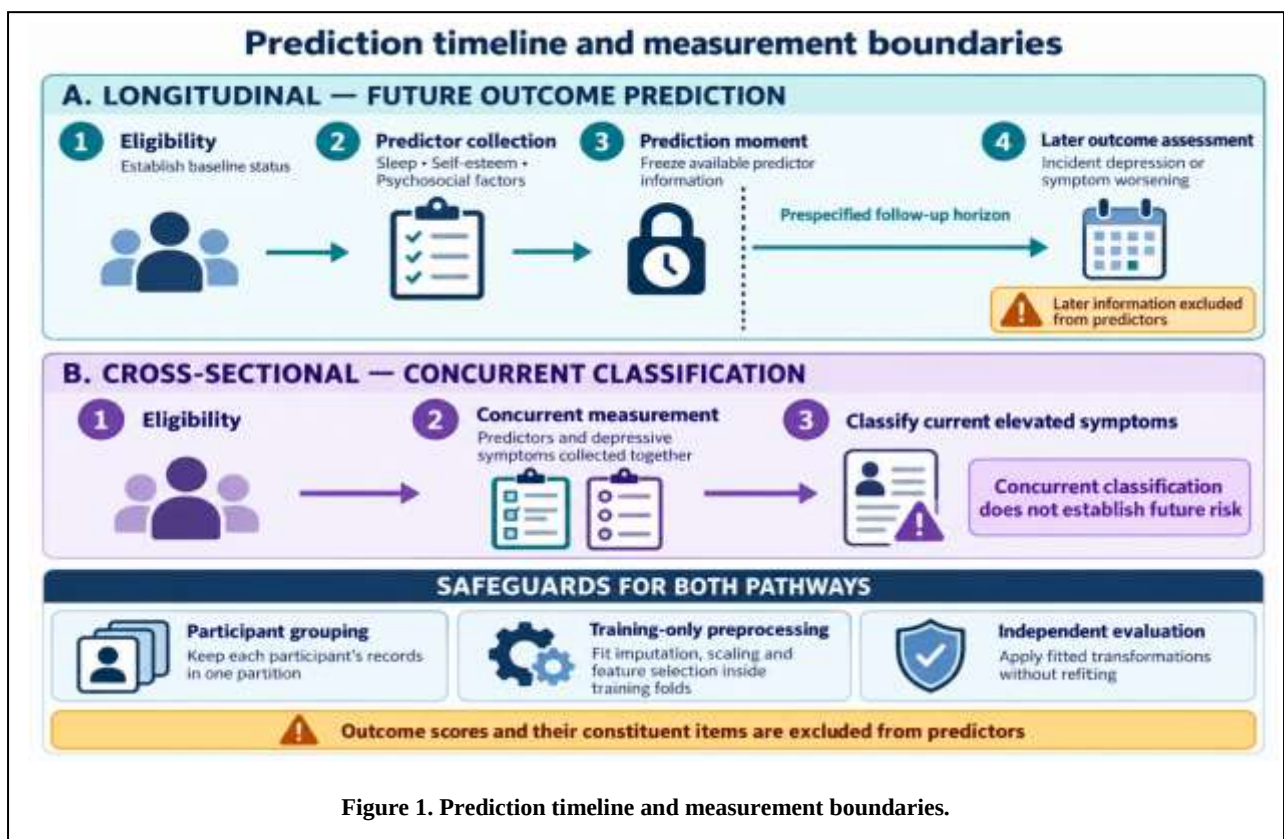


Figure 1. Prediction timeline and measurement boundaries.

forests and gradient boosting, avoiding an unnecessarily broad algorithm search. Random forests combine multiple randomised trees [19], whereas gradient boosting builds an additive model through sequential improvement of a loss function [20].

For longitudinal prediction, chronological separation will place later observations in evaluation data, with grouping rules preventing participant overlap. Leakage can otherwise inflate apparent generalisation [17].

Within development data, nested cross-validation will separate inner hyperparameter tuning from outer performance estimation. All imputation, encoding, scaling and feature selection will be refitted inside the relevant training fold. Nested evaluation addresses optimism arising when model selection and assessment share data [21]. The partitioning

Sensitivity, specificity and predictive values will be reported at prespecified or development-derived thresholds. Thresholds will reflect the intended assessment pathway and consequences of missed cases and unnecessary referrals, rather than maximising a convenient statistic on evaluation data.

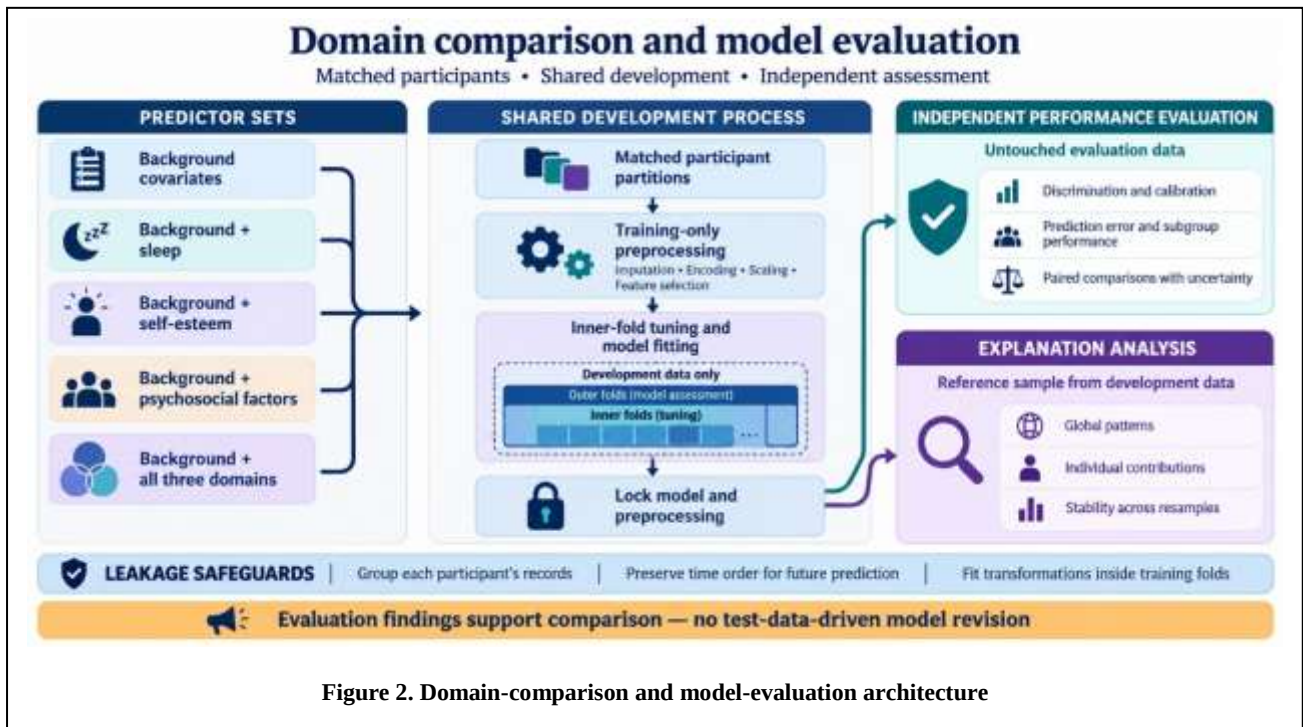


Figure 2. Domain-comparison and model-evaluation architecture

scheme will be adapted to sample size and event counts rather than fixed percentages alone.

Class weighting or resampling, if justified, will operate only during training. Evaluation samples will retain their observed outcome distribution. Any probability recalibration will use development data and be evaluated separately [22].

An untouched evaluation set will be preserved where feasible. Uncertainty will use paired resampling at the participant level, preserving the prediction horizon and relevant grouping. External validation will be undertaken only if an independent dataset supports compatible predictors and outcomes. Otherwise, the absence of external evidence will remain an explicit limitation.

#### 4.4 Performance, Calibration and Decision Relevance

Discrimination will be assessed using receiver operating characteristic area under the curve, supplemented by precision–recall performance when outcome prevalence warrants it. Accuracy alone will not determine model selection. Calibration plots, calibration intercept and slope, and the Brier score will assess probability reliability and overall error. Calibration requires attention beyond discrimination because misleading probabilities can undermine decisions [22].

Decision-curve analysis will be considered only when an actionable decision and plausible threshold range are defined. Its net-benefit framework compares model-guided decisions with relevant default strategies [23]. Confidence intervals and subgroup results will accompany overall estimates. Small statistical gains will be weighed against measurement burden, operational feasibility and uncertain benefit. Predictive improvement alone will not demonstrate better care or justify deployment without further evaluation in practice.

#### 4.5 Explainability and Explanation Robustness

Global explanations will combine grouped importance summaries with plots examining nonlinear patterns and selected interactions. Individual explanations will describe contributions to particular predictions using a stated reference sample. Dependence between predictors can materially affect Shapley-based explanations and requires explicit consideration [24].

Stability will be assessed across resamples, feature sets and reasonable model specifications. Explanation generation will remain distinct from performance evaluation, with reference data drawn from development samples. Feature contributions will describe model behaviour, not personalised causes or treatment recommendations. Results will therefore begin with data quality and comparative predictive value before

examining explanations, their consistency and remaining interpretive limitations.

## 5. RESULTS: WHAT IMPROVES PREDICTION, AND HOW RELIABLE ARE THE EXPLANATIONS?

### 5.1 Sample Characteristics and Outcome Distribution

Participant records were assessed against prespecified eligibility criteria, and exclusions were documented by reason. The analytical sample was partitioned into development and independent evaluation sets, with repeated observations grouped by participant to prevent overlap.

Sample characterisation covered demographic composition, outcome frequency, symptom severity and the distributions of sleep, self-esteem and psychosocial measures. Missingness was examined across predictor domains and outcome groups. Development and evaluation samples were compared to identify differences that could affect interpretation. Evaluation precision was considered in relation to sample composition and outcome counts [26].

For longitudinal data, follow-up completeness and differences between retained participants and those lost to follow-up formed part of the assessment. Concurrent observations were interpreted as cross-sectional measurements rather than evidence of subsequent outcomes.

### 5.2 Added Value of Sleep, Self-Esteem and Psychosocial Domains

The analytical comparisons contrasted the background-covariate baseline with models incorporating sleep, self-esteem and psychosocial information separately and jointly. Evaluation considered discrimination, calibration and overall probability error rather than inferring superiority from a single performance statistic [27].

Differences between predictor sets were assessed using paired comparisons on the same evaluation participants, with uncertainty intervals. Repeated validation examined the consistency of these differences. Precision–recall performance was interpreted alongside outcome prevalence, recognising the importance of appropriate evaluation under class imbalance [25].

Domain-removal analyses examined the contribution of each domain conditional on the remaining predictors. These comparisons assessed potential redundancy and complementary predictive information without establishing independent causal contributions [35].

Interpretation considered additional measurement burden alongside performance differences. Small, uncertain or inconsistent differences were distinguished from reproducible improvements. Table 2 was structured to present algorithm-specific comparisons, including domain additions that offered no clear improvement. Comparisons involving different complete-case samples were considered separately because

differences in sample composition could affect apparent performance.

**Table 2. Illustrative model-performance example—synthetic values, not study findings.**

| Predictor set                      | Algorithm           | Evaluation N | ROC-AUC | Precision-AUC | Brier score | Calibration intercept/slope | $\Delta$ ROC-AUC vs baseline |
|------------------------------------|---------------------|--------------|---------|---------------|-------------|-----------------------------|------------------------------|
| Background covariates: baseline    | Logistic regression | 800          | 0.68    | 0.35          | 0.158       | −0.05<br>0.91               | /Reference                   |
| Baseline + sleep                   | Logistic regression | 800          | 0.75    | 0.44          | 0.145       | −0.03<br>0.94               | /+0.07                       |
| Baseline + self-esteem             | Logistic regression | 800          | 0.76    | 0.46          | 0.141       | −0.04<br>0.93               | /+0.08                       |
| Baseline + psychosocial measures   | Logistic regression | 800          | 0.74    | 0.42          | 0.148       | −0.02<br>0.95               | /+0.06                       |
| Baseline + all domains             | Logistic regression | 800          | 0.81    | 0.54          | 0.128       | −0.02<br>0.97               | /+0.13                       |
| Combined model without sleep       | Logistic regression | 800          | 0.79    | 0.50          | 0.134       | −0.03<br>0.95               | /+0.11                       |
| Combined model without self-esteem | Logistic regression | 800          | 0.78    | 0.48          | 0.137       | −0.04<br>0.94               | /+0.10                       |

### 5.3 Global Patterns, Interactions and Individual Explanations

Global interpretation will identify influential predictors and assess the stability of their rankings and contributions across resamples. Nonlinear plots will examine relationships between predictor values and model outputs. Interpretation will account for feature dependence, which can alter attribution and complicate comparisons between correlated predictors [24].

Prespecified analyses will examine interactions involving sleep, self-esteem and psychosocial conditions. Patterns that vary substantially across resamples or model specifications will be treated as exploratory. Comparisons across algorithms and explanation methods will identify agreement and disagreement without assuming that one explanation represents an underlying causal mechanism.

Anonymised examples, selected using prespecified criteria, will illustrate individual predictions, including correctly and incorrectly classified observations where appropriate [26].

were identified without ranking unstable estimates, recognising that validation precision depended on adequate information for each performance measure [26].

Sensitivity analyses examined alternative depression thresholds, exclusion of overlapping sleep predictors and an exploratory symptom score excluding the sleep item. Modified outcome definitions were considered separately, without automatically retaining thresholds validated for the original instrument. Comparisons considered discrimination,

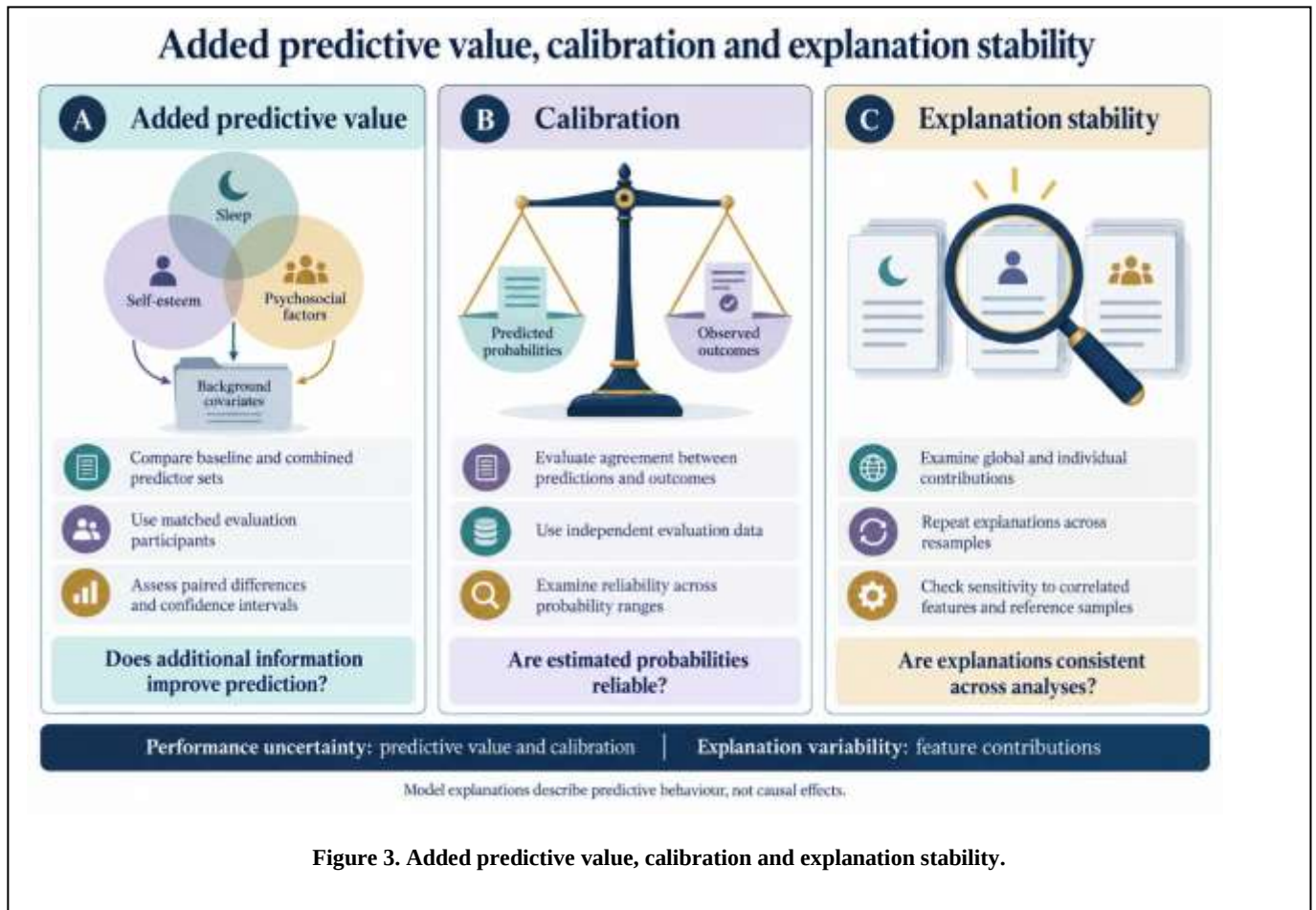


**Figure 3. Added predictive value, calibration and explanation stability.**

Each example will report the predicted probability, observed outcome and reference population used for explanation. Participant confidentiality will be maintained throughout. Individual contributions will be examined across alternative reference samples and defensible model specifications. Material changes will be reported transparently rather than selecting explanations for visual clarity [20]. Stable rankings will indicate consistency of model interpretation, but will not establish causal validity or justify recommendations to modify individual predictors.

#### 5.4 Subgroup Performance and Sensitivity Findings

Calibration and threshold-specific errors were assessed across prespecified demographic and contextual groups. Estimates included subgroup denominators, outcome counts and uncertainty intervals. Groups with insufficient observations

calibration and overall prediction error [27].

Alternative missing-data procedures were compared, with their assumptions and implementation documented. Where chained-equation imputation was used, the imputation model was specified and fitted within training partitions [28]. Additional analyses assessed performance when predictors were restricted to measures available in the intended assessment setting.

Findings were evaluated for consistency across analytical specifications. Similar overall accuracy was not treated as sufficient evidence of fairness, given the potential for healthcare algorithms to reproduce inequities [31]. Interpretation considered subgroup errors, measurement overlap and population differences, recognising that

performance in one setting did not establish transportability to another [33].

## **6. DISCUSSION: MEANING, LIMITATIONS AND RESPONSIBLE TRANSLATION**

### **6.1 Interpreting Complementarity Across Predictor Domains**

Whether combined domains improve prediction will depend on the independent comparisons reported in Section 5. Consistent gains in discrimination, calibration and probability error would support complementary information; small or uncertain gains would favour a more restrained interpretation. No superiority can be claimed because analytical results remain unavailable.

Sleep describes behavioural patterns, self-esteem concerns self-evaluation, and psychosocial measures represent perceived demands and social circumstances. These distinctions provide a rationale for combination, but shared reporting processes and overlapping constructs may inflate apparent complementarity [32]. Depressive symptoms may also influence sleep, self-perception and relationships, particularly when measurements are concurrent.

Regression coefficients describe model-specific conditional associations, whereas predictive explanations allocate contributions according to the fitted algorithm and explanatory assumptions. Neither automatically answers what would happen if a predictor changed; prediction and causal inference address different questions [35].

Interactions will warrant substantive interpretation only when supported across resamples and defensible specifications. Exploratory patterns will remain hypotheses. If a simpler model performs comparably, its lower measurement burden may outweigh marginal complexity. Decisions about usefulness must ultimately consider whether predictions improve an assessment pathway, rather than whether more variables produce detectable gains in a research dataset [34].

### **6.2 Measurement, Design and Generalisability Limitations**

Self-reported sleep, self-esteem and psychosocial circumstances may be affected by recall error, current mood and socially desirable responding. Collecting predictors and symptoms through similar questionnaires can introduce common-method variation, potentially exaggerating relationships [32]. A questionnaire threshold may also misclassify depressive disorder because symptom screening and clinical diagnosis are not interchangeable.

Sleep-related outcome items create a circularity concern. Excluding overlapping predictors or exploring modified symptom scores can assess sensitivity, but altered scores require cautious interpretation. Concurrent measurements cannot resolve whether disturbed sleep precedes symptoms or follows them; causal direction requires a different inferential design [35].

Selection bias may arise through recruitment access, language restrictions, nonparticipation and incomplete questionnaires. Longitudinal attrition may further distort the evaluated population. Imputation depends on assumptions and cannot restore unmeasured experiences or guarantee representativeness.

Overfitting, limited outcome counts and repeated model comparisons may produce uncertain performance differences. Correlated features can redistribute attribution across otherwise similar models, so stable predictions need not imply stable explanations. Conversely, stable explanations do not establish causal validity.

Transportability across ages, cultures, service settings and observation periods remains untested until independent evaluation is completed. External validation should examine differences in case mix, measurement and setting rather than assume that successful internal validation establishes generalisability [33]. These limitations will constrain interpretation even if the selected model demonstrates strong aggregate performance within the available sample.

### **6.3 Equity, Acceptability and Practical Application**

False positives may generate anxiety, stigma or unnecessary referrals, while false negatives may delay appropriate assessment. Their consequences will depend on available services and how professionals communicate predictions. Research on healthcare algorithms demonstrates that design choices can reproduce inequities despite apparently useful overall performance [31].

Collecting sensitive information about self-worth, relationships or adversity will require privacy protections and participant understanding. Acceptability cannot be inferred from questionnaire completion alone. People may differ in willingness to disclose, literacy, language access and confidence that information will remain confidential.

Models requiring wearables or extensive questionnaires could disadvantage individuals with limited digital access, missing recordings or constrained time. These concerns will require subgroup assessment and accessible alternatives rather than exclusion of participants with incomplete information.

Accessible assessment and timely follow-up must accompany any proposed application. Research feasibility will not establish deployment readiness. Impact evaluation should determine whether model-supported decisions improve care compared with existing practice [34]. Until such evidence exists, the model will remain a research approach whose usefulness depends on a defined support pathway, professional oversight and explicit consideration of errors, costs and equity.

### **6.4 Future Research and Validation Priorities**

Future research should evaluate incident depression and symptom worsening longitudinally, with predictor collection preceding outcome assessment. Repeated sleep and

psychosocial measurements could clarify whether trajectories improve prediction beyond baseline summaries without establishing causal effects.

External and temporal validation should examine calibration, subgroup errors and measurement compatibility across settings [33]. Studies should also test whether shorter questionnaires or fewer domains retain performance, reducing burden and dependence on digital resources.

Explanation stability should be examined under population change, alternative reference samples and correlated-feature handling. Participants should help evaluate privacy, comprehensibility, stigma and acceptable follow-up arrangements. Their involvement should influence design decisions rather than simply endorse completed systems.

Comparative impact studies should assess whether model-supported assessment improves decisions, referral completion and outcomes beyond established screening [34]. Intervention studies or other credible causal designs will be necessary before claiming that modifying influential predictors reduces depression [35]. These priorities define the evidence required before the proposed analytical contribution can inform routine practice or service planning.

## 7. CONCLUSION

### 7.1 Integrated Contribution and Principal Findings

Combining sleep patterns, self-esteem and psychosocial information offers a structured way to examine whether behavioural, psychological and social domains contribute complementary predictive information. However, no completed analyses are available, so added value over simpler models remains unestablished. The proposed contribution is a transparent comparison of predictor sets, supported by calibration assessment, independent evaluation and examination of explanation stability. Any eventual conclusions must reflect uncertainty, measurement burden and the possibility that regression performs comparably to more complex algorithms. Concurrent measurements can identify elevated symptoms but cannot establish future risk. Forecasting onset or worsening requires predictors measured before subsequent outcomes, with a clearly defined prediction horizon and appropriate longitudinal evaluation in practice.

### 7.2 Conditions for Responsible Use

Responsible use will require external validation, reliable measurement and assessment of calibration and errors across relevant subgroups. Model explanations describe algorithmic behaviour rather than causes or the benefits of changing individual predictors. Privacy, participant acceptability and accessible professional follow-up must inform any proposed application. Deployment should depend on evidence that predictions improve a defined assessment pathway beyond existing practice. Carefully evaluated models may support earlier recognition of assessment needs, but improved predictive performance alone does not demonstrate better

mental health, effective treatment or equitable outcomes for the people being assessed.

## 8. REFERENCE

1. Chung JO, Lam KK, Ho KY, Cheung AT, Ho LL, Gibson F, Li WH. Relationships among resilience, self-esteem, and depressive symptoms in Chinese adolescents. *Journal of health psychology*. 2020 Nov;25(13-14):2396-405.
2. Portugal LC, Gama CM, Gonçalves RM, Mendlowicz MV, Erthal FS, Mocaiber I, Tsirlis K, Volchan E, David IA, Pereira MG, Oliveira LD. Vulnerability and protective factors for PTSD and depression symptoms among healthcare workers during COVID-19: A machine learning approach. *Frontiers in psychiatry*. 2022 Jan 12;12:752870.
3. C Manikis G, Simos NJ, Kourou K, Kondylakis H, Poikonen-Saksela P, Mazzocco K, Pat-Horenczyk R, Sousa B, Oliveira-Maia AJ, Mattson J, Roziner I. Personalized risk analysis to improve the psychological resilience of women undergoing treatment for breast cancer: Development of a machine learning-driven clinical decision support tool. *Journal of Medical Internet Research*. 2023 Jun 12;25:e43838.
4. Usen J. Food safety practices, policy interventions and the reduction of public health risks in the US. *Magna Sci Adv Res Rev*. 2021;3(2):152-166. doi:10.30574/msarr.2021.3.2.0095.
5. Zaman A, Acharyya R, Kautz H, Silenzio V. Detecting low self-esteem in youths from web search data. In *The World Wide Web Conference 2019 May 13* (pp. 2270-2280).
6. Du C, Liu C, Balamurugan P, Selvaraj P. Deep learning-based mental health monitoring scheme for college students using convolutional neural network. *International Journal on Artificial Intelligence Tools*. 2021 Dec 29;30(06n08):2140014.
7. de Siqueira Rotenberg L, Borges-Junior RG, Lafer B, Salvini R, da Silva Dias R. Exploring machine learning to predict depressive relapses of bipolar disorder patients. *Journal of Affective Disorders*. 2021 Dec 1;295:681-7.
8. Tyas SL, Pederson LL. Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tobacco control*. 1998 Dec 1;7(4):409-20.
9. Lim H, Lee H, Kim J. A prediction model for childhood obesity risk using the machine learning method: a panel study on Korean children. *Scientific Reports*. 2023 Jun 21;13(1):10122.
10. Kırcaburun K, Kokkinos CM, Demetrovics Z, Király O, Griffiths MD, Çolak TS. Problematic online behaviors among adolescents and emerging adults: Associations between cyberbullying perpetration, problematic social media use, and psychosocial factors. *International Journal of Mental Health and Addiction*. 2019 Aug 15;17(4):891-908.
11. Ogan RK. Machine-learning-based API failure prediction and root-cause analytics for improving reliability of enterprise banking integration architectures at scale. *Int J*

- Comput Artif Intell. 2021;2(2):142-153. DOI:10.33545/27076571.2021.v2.i2a.378.
12. Kiss O, Baker FC, Palovics R, Dooley EE, Pettee Gabriel K, Nagata JM. Using explainable machine learning and fitbit data to investigate predictors of adolescent obesity. *Scientific reports*. 2024 May 31;14(1):12563.
13. Salah H, Srinivas S. Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents. *Scientific Reports*. 2022 Dec 19;12(1):21905.
14. Alebiosu O. Engineering stakeholder alignment frameworks for resolving land rights, environmental compliance, and project delivery conflicts. *Int J Appl Res*. 2024;10(12):401-412.
15. Haghighi EF, Obaidi M, Strømme T, Bjørge T, Grønnerød C. Mental health, well-being, and adolescent extremism: a machine learning study on risk and protective factors. *Research on child and adolescent psychopathology*. 2023 Nov;51(11):1699-714.
16. Bhandari, Ambika Baniya. 2026. "Bridging the Public Health–Cloud Security Divide: Workforce Implications of the HIPAA Security Rule Modernization for Public Health Informatics". *Current Journal of Applied Science and Technology* 45 (8):33-44. <https://doi.org/10.9734/cjast/2026/v45i84730>.
17. Tayarani-N MH, Shahid SI. Detecting anxiety via machine learning algorithms: a literature review. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2025 Mar 12;9(4):2634-57.
18. Kim K, Yoon Y, Shin S. Explainable prediction of problematic smartphone use among South Korea's children and adolescents using a machine learning approach. *International journal of medical informatics*. 2024 Jun 1;186:105441.
19. Sandra Darkwa. (2023). FEDERATED DEEP REINFORCEMENT LEARNING SYSTEMS FOR ADAPTIVE COUNTER-DISINFORMATION MESSAGING DURING CROSS-BORDER ELECTORAL INFLUENCE OPERATIONS GLOBALLY. *International Journal Of Engineering Technology Research & Management (IJETRM)*, 07(12), 692–706. <https://doi.org/10.5281/zenodo.20237639>
20. Pinto SJ, Parente M. Comprehensive review of depression detection techniques based on machine learning approach. *Soft Computing*. 2024 Sep;28(17):10701-25.
21. Mittal S, Mahendra S, Sanap V, Churi P. How can machine learning be used in stress management: A systematic literature review of applications in workplaces and education. *International Journal of Information Management Data Insights*. 2022 Nov 1;2(2):100110.
22. Bhandari, Ambika Baniya. 2026. "Quality Control Frameworks for Electronic Health Record Data Workflows in Resource-Constrained Settings: A Structured Review With Implications for Global Health Information Management". *Current Journal of Applied Science and Technology* 45 (8):1-16. <https://doi.org/10.9734/cjast/2026/v45i84728>.
23. Kim SS, Gil M, Min EJ. Machine learning models for predicting depression in Korean young employees. *Frontiers in Public Health*. 2023 Jul 12;11:1201054.
24. Tariq R, Orozco-del-Castillo MG, Zamir MT, Ramírez-Montoya MS, Wilberforce T. RETRACTED ARTICLE: Explainable artificial intelligence for predictive modeling of student stress in higher education. *Scientific Reports*. 2025 Nov 3;15(1):38375.
25. ShamsEldin T, Gaber S, Ansari S, Elgohary R, Shawky MA, Elbahnasawy M, Abdrabou M. Artificial intelligence for predicting depression anxiety and stress using psychometric data. *Scientific Reports*. 2025 Oct 24;15(1):37282.
26. Iyanuoluwa Oluwafemi. (2024). ENVIRONMENTAL MONITORING STRATEGIES INTEGRATING MOLECULAR DETECTION TECHNOLOGIES TO ENHANCE AGRICULTURAL BIOSECURITY AND SUSTAINABLE CROP PRODUCTION SYSTEMS. *International Journal Of Engineering Technology Research & Management (IJETRM)*, 10(12), 757–772. <https://doi.org/10.5281/zenodo.21152414>
27. Hosseinzadeh Kasani P, Lee JE, Park C, Yun CH, Jang JW, Lee SA. Evaluation of nutritional status and clinical depression classification using an explainable machine learning method. *Frontiers in Nutrition*. 2023 May 9;10:1165854.
28. Khoo LS, Lim MK, Chong CY, McNaney R. Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches. *Sensors*. 2024 Jan 6;24(2):348.
29. Luo L, Yuan J, Wu C, Wang Y, Zhu R, Xu H, Zhang L, Zhang Z. Predictors of depression among Chinese college students: a machine learning approach. *BMC Public Health*. 2025 Feb 5;25(1):470.
30. Emmanuel Uzoh. (2023). PREDICTIVE OPERATIONS INTELLIGENCE: AN AUTOMATED MACHINELEARNING SYSTEM FOR BOTTLENECK DETECTION AND DYNAMIC WORKSTREAM PRIORITIZATION IN SMALL AND MEDIUM-SIZED ENTERPRISES. *INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)*, 07(12), 793–806. <https://doi.org/10.5281/zenodo.22089397>
31. Price GD, Heinz MV, Song SH, Nemesure MD, Jacobson NC. Using digital phenotyping to capture depression symptom variability: detecting naturalistic variability in depression symptoms across one year using passively collected wearable movement and sleep data. *Translational Psychiatry*. 2023 Dec 9;13(1):381.
32. Saylam B, İncel ÖD. Multitask learning for mental health: depression, anxiety, stress (DAS) using wearables. *Diagnostics*. 2024 Feb 26;14(5):501.
33. Chu X, Yang X, Ye Z, Hu X, Chen L, Deng Y, Zheng Y, Chen L. An interpretable machine learning model for predicting sleep disturbances in adolescents with mood disorders: the key role of psychological factors. *BMC psychiatry*. 2025 Dec 3;26(1):19.
34. Sandra Darkwa. Transformer-Driven Strategic Communication Analytics for Identifying Synthetic Media Manipulation and Coordinated Narrative Amplification Networks Digitally. *Int J Literacy Educ* 2024;4(2):333-346. DOI: [10.22271/27891607.2024.v4.i2d.434](https://doi.org/10.22271/27891607.2024.v4.i2d.434)

35. Niu B, Wan M, Zhou Y. Development of an explainable machine learning model for predicting depression in adolescent girls with non-suicidal self-injury: A cross-sectional multicenter study. *Journal of Affective Disorders*. 2025 Jun 15;379:690-702.