

Predicting Mental Health in People Living with HIV Using Machine Learning Uncovers the Central Role of Resource Insecurity

Godfred O. Boateng

gboateng@yorku.ca

York University

Caleb Osae-Morgan

York University

Sylvia E. Gyan

University of Ghana

Patrick Owuor

Wayne State University

Ellis Adjei Adams

University of Notre Dame

Article

Keywords: Machine Learning, Resource Insecurity, Mental Health, PLHIV, Kenya

Posted Date: June 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9809249/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study examines the predictive effect of critical resource deprivations at the household level: food, water, energy and housing insecurity, using machine learning models to build resource insecurity scores for three mental health outcomes. Data for this study were collected from the Resource Insecurity and Wellbeing study, which focused on PLHIV in Kenya. We analyzed 1,132 adults infected with HIV. Data were collected on participants' experiences with food, water, energy and housing insecurity, mental health outcomes including anxiety, stress and depressive symptomatology, and socio-demographic characteristics. Descriptive and regression analyses were first used to assess the relationships between the variables. Using Random Forest and XGBoost, we computed resource insecurity scores and predicted their effect on mental health outcomes. XGBoost was selected to extract local feature contribution using Shapley values and derive feature importance and interactions. Machine learning models revealed that resource insecurity and sleep disorder were significant predictors of mental health risk among PLHIV in Kenya. SHAP value analysis enhanced model interpretability by identifying water insecurity as a primary driver of depression and stress, with food insecurity, energy insecurity, and housing satisfaction also contributing substantially to mental health risk. Insomnia emerged as the strongest predictor of generalized anxiety disorder, followed by overall resource insecurity. Socio-demographic variables demonstrated relatively low predictive importance across all models. These findings underscore the importance of integrated interventions targeting household resource insecurity and sleep disorders, as well as the need for mental health strategies in resource-limited settings to address broader structural determinants alongside individual-level factors.

Introduction

Mental health is a critical global public health concern, particularly given its strong association with long-term disability and the consequences for individuals, families and the health systems [1]. Globally, more than 1 billion people are estimated to be living with a mental health condition [2, 3]. Mental disorders are among the leading contributors to the global burden of disease [GBD] [4], accounting for 7% of the GBD [2]. In low- and middle-income countries, the prevalence is high, but the budget for mental healthcare is less than 2%, creating a serious public health concern [4–6].

Aside from being among the leading causes of disability globally, it also imposes high costs on households, employers and economies [1, 7]. The burden of mental health conditions in any population, irrespective of age, socio-economic status or place of residence, is not borne by the affected individual alone [8]. Instead, it creates a ripple effect that extends to families, communities and society at large. Families caring for relatives with mental health challenges often shoulder the social and economic costs of health conditions. Globally, disability adjusted life years (DALYs) due to mental disorders are estimated at 125.3 million, representing 4.9% of the total global DALYs [9]. When these conditions remain untreated or are difficult to manage, they can progress into long-term disability and contribute to loss of healthy life.

Among adolescents and young adults, mental disorders may disrupt educational attainment and hinder participation in the workforce [10, 11]. For PLHIV, poor mental health can compromise engagement with HIV care, including adherence to the treatment regimens, leading to negative health outcomes [12]. Given these consequences and social and economic costs of mental health conditions, it is important to broaden research on mental health by examining the associations between structural inequities and mental health. Such evidence is critical for shaping targeted interventions, strengthening policy responses, and advancing progress toward Sustainable Development Goal (SDG) 3 on ensuring good health and promoting well-being, particularly Target 3.4, which calls for the prevention and treatment of mental health conditions and the promotion of mental well-being.

Resource insecurity is a multifaceted condition that shapes individuals' well-being through scarcity or limited access to basic needs, such as food, water, and housing, and affects the well-being of individuals [13]. In recent times, a limited number of studies have identified that resource insecurity is associated with poor mental health outcomes among PLHIV [14–17]. Most of these studies have examined the association between specific resource insecurities and mental health. In a systematic review on the impact of food insecurity on mental health in Africa, it was observed that food insecurity was associated with poor mental health, especially among pregnant women and PLHIV [16]. Similarly, evidence from rural Brazil and urban Ethiopia indicates that food insecurity affects mental health directly through basic needs deprivation, such as the persistent worry about where the next meal will come from [18].

The emerging evidence suggests that these resource insecurities often co-occur. For instance, food and water insecurity frequently overlap and pose public health challenges [19]. Housing and water insecurity may also co-exist and jointly influence the mental health of people, particularly PLHIV. In Canada Logie et al. [20] observed that women living with HIV who experienced both food and housing insecurity also reported HIV-related stigma. Similarly, in LMICs, researchers found an association between food and housing insecurity and poor mental health among women living with HIV [21]. Nawfal et al. [21] in their study identified a strong association between food and housing insecurity and poor mental health, particularly symptoms of depression and anxiety among women living with HIV. Yet, no study has examined and quantified simultaneously the different forms of resource insecurity's influence on mental health outcomes. Consequently, it is difficult to tell which form of insecurity matters most in the face of scarce resources.

Extending the study of the social determinants of mental health to include structural resource insecurities, especially among PLHIV, this study examines the predictive effect of four important resources: food, water, energy and housing insecurity using machine learning. These resources represent the basic needs for human well-being, and the lack of access or availability of such resources can have cascading effects on the mental health of PLHIV. Thus, unlike traditional regression models, the use of machine learning offers a superior framework for evaluating mental health consequences of resource insecurity because it can detect nonlinearities [22, 23], capture complex interactions [24, 25], manage multicollinearity [26, 27], and deliver higher predictive accuracy [28, 29]. Additionally, the use of modern interpretability tools (such as Shapley Additive Explanations values) will allow us to identify which

specific dimension of resource insecurity contributes most to mental health risk [30, 31], providing deeper insights than conventional parametric models.

Methods

Study Participants

This study uses data from the Resource Insecurity and Wellbeing study conducted among PLHIV in peri-urban communities in Kenya. Data were collected from two sub-counties, Seme (n = 766) and Kisumu West (n = 366). Overall, 1,132 PLHIV were examined. Samples were selected using a stratified two-stage cluster sampling design, in which each sub-county was treated as a cluster, and households were randomly selected within each cluster. Data collection started with a manual mapping exercise in areas where the Pamoja Community-Based Organization (CBO) has been providing services to PLHIV since 2007. This process was used to identify households with PLHIV. Subsequently, participants were randomly selected from eligible households within each sub-county for inclusion in the study. The inclusion criteria for participants were age (16 years or older) and knowledge of household-level resource insecurities. Data were collected on participants' experiences with food, water, energy, and housing insecurity, as well as mental health outcomes, including anxiety, stress and depressive symptoms, along with socio-demographic characteristics.

Data

Health Outcomes

Health outcomes of PLHIV were measured using a variety of validated standardized scales administered through a questionnaire to participants during fieldwork from 10th July 2023 to 3rd August 2023.

Depressive Symptomatology

Outcome I: Depression

The five-item World Health Organization Well-Being Index (WHO-5) was used to assess depressive symptoms. The WHO-5 is a self-report instrument consisting of five statement items that assess mental well-being, such as feeling cheerful, feeling calm and relaxed, feeling active and vigorous, feeling fresh and well-rested, and having an interesting daily life. Each statement is rated on a scale from 0 ("At no time") to 5 ("All of the time"), with higher scores indicating good mental well-being. In line with established guidelines, participants who scored less than or above 13 out of a total score of 25 were classified as having or not exhibiting symptoms of depression, respectively. For analytical purposes, participants were classified into two categories based on their cumulative scores on the WHO-5, yielding

a binary variable: participants who scored less than 13 were considered to exhibit symptoms of depression, and those who scored 13 or higher were considered not to exhibit symptoms of depression.

Outcome II: Anxiety

The Generalized Anxiety Disorder 7-item scale (GAD-7), a valid and efficient tool for screening and assessing the severity of generalized anxiety disorder in both clinical practice and research, was used to assess participants' anxiety symptoms [32]. The GAD-7 demonstrates excellent internal consistency (Cronbach's $\alpha = 0.92$). In this study, GAD-7 scores were treated as a continuous variable to capture variation in the severity of anxiety symptoms across participants. The scale consists of 7 questions that assess how often participants have been bothered by the problems listed on the GAD-7 over the past 2 weeks. The responses ranged from "Not at all = 0" to "Nearly every day = 3", yielding a total score of 21. The severity of anxiety symptoms was marked by higher scores on the scale.

Outcome III: Stress

The Perceived Stress Scale (PSS-4) is a widely used self-report instrument that assesses individuals' perceptions of recent stress. It measures the extent to which respondents self-evaluate situations in their lives as unpredictable, uncontrollable, and/or overwhelming. Questions on the PSS-4 ask the participants to rate their feelings and thoughts over the past month, capturing stress within a recent time frame. Due to the limited number of items, the PSS-4 demonstrates a lower internal consistency (Cronbach's $\alpha = 0.60$).

Resource Insecurity Data

Four types of interrelated resource insecurities experienced by participants were measured: energy, food, housing, and water. These insecurities were assessed relying on existing scales that have been adapted and validated in low-income countries and informal settlement contexts. The various scales were coded such that higher scores indicated greater insecurity. Resource insecurity was assessed by focusing on participants' experiences during the 30 days preceding their participation in the study.

Energy insecurity. An adapted version of the Household Energy Poverty Experience Scale (HEPES) was used to measure participants' energy insecurity [33]. The scale consists of a 10-item questionnaire about individual experiences with energy within the last 4 weeks for the research. Responses were coded based on the frequency of the experience within the 4 weeks as follows: 1 = (Never "0 times in the last 4 weeks"), 2 = (Rarely "1–2 times in the last 4 weeks"), 3 = (Sometimes "3–10 times in the last 4 weeks"), 4 = (Often "11–20 times in the last 4 weeks") and 5 = (Always "More than 20 times in the last 4 weeks"). The HEPE Scale produced a Cronbach's alpha of 0.94, with an average interitem covariance of 0.42. The frequency of each experience and item responses was summed to generate a continuous energy insecurity score ranging from 0 to 30, with lower scores indicating greater energy insecurity.

Food insecurity. Food insecurity was assessed using [34] Household Food Insecurity Access Scale (HFIAS). This scale measures multiple dimensions of food insecurity, such as the inability to afford food, anxiety about food availability, a lack of nutritious foods and dietary diversity, reductions in meal size or frequency, and experiences of hunger, such as going to bed hungry or not eating for an entire day. The HFIAS measured the frequency of participants' experiences with access to food in the last four weeks and the responses were coded as follows: 1 = Never (0 times in the last 4 weeks), 2 = Rarely (1–2 times in the last 4 weeks), 3 = Sometimes (3–10 times in the last 4 weeks), 4 = Often (11–20 times in the last 4 weeks) and 5 = Always (More than 20 times in the last 4 weeks). The HFIAS produced a score range of 0 to 27, with a Cronbach's alpha of 0.90. Higher scores suggested greater food insecurity, while lower scores suggested lower food insecurity.

Housing Insecurity. The Multilevel Multidimensional Housing Insecurity Scale (MMHIS), developed by Boateng and Adams [35] was used to measure housing insecurity. The MMHIS consists of 17 questions assessing three interconnected dimensions of housing insecurity in low-income countries. Participants responded to 17-item questions that measured their experience of housing insecurity over the past year, and a composite score was calculated, with higher scores indicating greater housing insecurity. The first dimension captures essential utility deficiencies, including access to tap water, toilet facilities and electricity. The second dimension assesses built-environment challenges such as overcrowding, flooding, and leaking roofs. The third dimension examines tenancy-related challenges, such as fear of eviction by a landlord and the inability to pay rent. Among the three dimensions, this study focused on built-environment challenges because they are well established to be associated with mental well-being outcomes.

Water insecurity. The Household Water Insecurity Experiences Scale (HWISE) was adapted to measure water insecurity [36]. The HWISE Scale is a 12-item questionnaire that examines the dimensions of water insecurity, particularly in low-income countries, by focusing on people's lived experiences with the accessibility, availability, and use of safe and reliable water. The HWISE scale was used to measure the intensity of water insecurity, with responses measuring the frequency with which the condition was experienced: "Never" = 0, "Rarely" = 1, "Sometimes" = 2, or "Often" = 3. This yielded a composite score ranging between 0 and 36. The Cronbach's alpha of the 12 items was 0.97 with an average inter-item covariance of 0.59, indicative of the strength of the scale [37]. Higher scores indicate higher experiences of water insecurity in the past month.

Mediators: Sleep Health

Three measures were used to assess sleep health. The first measure was Acute Sleep Deprivation. Participants were asked, "In the past few days, have you had no sleep or a reduction of sleep in your total sleep time?" with response options such as (0 = No, 1 = Yes). The second measure focused on Insomnia, derived from the first five questions of the Sleep Disorder Questionnaire (SDQ), which together form the diagnostic domain for Insomnia [38]. Participants were asked questions such as: "Do you have trouble falling asleep? Do you have trouble staying asleep? Do you take anything to help you sleep? Do you use

alcohol to help you sleep? Do you have any medical conditions that disrupt your sleep?” The response options for each question were: never, rarely, occasionally, most nights/days, and always (coded 1, 2, 3, 4, 5). This produced Cronbach’s alpha of 0.75. Higher scores on this measure suggest severe insomnia, and lower scores suggest minor experiences of insomnia. The third measure was chronic sleep insufficiency. Participants were asked, “In the past month, do you routinely sleep less than the total amount of time you are required to sleep for optimal functioning? Response to this question was coded as “0” for No and “1” for Yes.

Covariates

We computed for both biosocial and socio-cultural factors. Biosocial variables included age, measured as a continuous count variable, and gender, coded as male (0) and female (1). Socio-cultural factors comprised the number of children under five years of age in the household; marital status, coded as single/unmarried (1), separated/divorced/widowed (2), and married/cohabiting (3); and neighborhood of residence, coded as Seme sub-county (1) and Kisumu sub-county (2). Educational attainment was coded as no education (0), primary (1), middle/junior high (2), secondary/senior high (3), and tertiary (4). Participants were also asked to self-rank their socio-economic status on a scale from 0 to 10, with higher scores indicating higher socio-economic standing. Additional measures included average monthly income, self-rated health assessed on a scale from very good (1) to very bad (5), and the primary source of water, which was classified as improved or unimproved in accordance with the Joint Monitoring Programme for Water Supply and Sanitation.

Statistical Analysis

All data preprocessing and modelling were performed in Python 3.12.3, using open-source libraries including scikit-learn (for Random Forests), XGBoost, and SHAP.

Data preparation. The data selection process and comprehensive study workflow are illustrated in Fig. 1. To ensure a robust analysis, we prioritized an efficient modelling approach that balances high predictive accuracy with overfitting mitigation. The preliminary stages involved systematic data refinement, including initial feature selection, handling missing observations through deletion and imputation, and categorical encoding of variables to facilitate statistical analysis.

From an initial dataset of 37 variables, a subset of 23 features was selected based on conceptual and established theoretical frameworks in social epidemiology. Administrative and technical metadata were excluded because they lacked explanatory power for mental health outcomes. Furthermore, variables such as Neighbourhood and Ethnicity were excluded due to insufficient variance within the sample. The data were collected from two geographically adjacent neighbourhoods with highly similar socio-environmental characteristics. The study population was predominantly composed of indigenes who identified with the same ethnic group. Consequently, these variables exhibited near-zero variance and could not provide meaningful statistical disparity or explanatory power regarding the observed mental health outcomes. By removing homogenous features, we reduced the risk of multicollinearity and sharpened the model’s focus on the heterogeneous drivers of resource insecurity and household

dynamics. The final feature set was categorized into four primary domains: **Mental Health Outcomes** (Anxiety, Stress, Depression), **Sleep Mediators** (Insomnia, Reduced Sleep, Sleeplessness), **Resource Insecurity Indicators** (e.g., Food Insecurity, Energy Insecurity, Water Insecurity, Housing Insecurity), and **Socio-demographic Covariates** (e.g., Age, Gender, Education, and Household composition). This targeted selection ensures that the predictive models remain parsimonious while capturing the multi-dimensional nature of household vulnerability.

To ensure data quality and analytical consistency, raw variables were transformed into formats suitable for rigorous statistical modelling. Continuous variables were retained in their raw count form, while categorical variables with an inherent rank order were treated as ordinal features to preserve their monotonic structure. Nominal variables were maintained as categorical entities, with effects interpreted relative to established reference categories. To safeguard model stability and interpretability, a stringent exclusion criterion was applied. Participants with more than 15% missing data across all features were removed. This process resulted in the exclusion of less than 1% of the total sample, suggesting that missingness had a negligible impact on the study's overall power. However, the HEPEScale (an essential metric for resource insecurity) exhibited a higher missingness rate of approximately 11%. Given its theoretical significance, we retained the feature and addressed the gaps using MissForest, a nonparametric random-forest-based imputation method. MissForest was selected for its ability to capture complex, non-linear relationships and inter-variable interactions without requiring strong parametric assumptions about data distribution. Pairwise correlation diagnostics were conducted on both the pre- and post-imputation datasets to assess potential multicollinearity and verify that the MissForest procedure preserved the data's underlying dependence structure. No feature pairs exceeded commonly used correlation thresholds ($r = 0.9$) in either dataset, and therefore no variables were excluded based on multicollinearity.

Modeling. We employed supervised machine learning to quantify the predictive relationships between resource insecurity, socio-demographic variables, sleep mediators, and mental health outcomes. To maintain data integrity, we inspected the distributions of each outcome to justify our modelling frameworks. Anxiety and Stress were treated as continuous variables and modelled as regression tasks, while Depression was treated as a binary classification task. Model training and validation were conducted using a 10-fold cross-validation (CV) procedure. This specific CV structure was selected to optimize the balance between computational efficiency and the robustness of performance estimates, providing a more reliable generalization than 5-fold or leave-one-out methods. For each outcome, hyperparameter optimization was executed using the Tree-structured Parzen Estimator^[ref] (TPE) algorithm via the Optuna library. We conducted 1,000 trials per model to exhaustively explore the hyperparameter space. For regression tasks, the models were optimized to maximize the Coefficient of Determination (R^2). This ensures that the tuning process prioritizes capturing systematic variance in the response rather than merely minimizing pointwise error. For Depression classification, the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) was used as the primary optimization metric. By prioritizing ROC-AUC over simple accuracy, the tuning process focused on the model's ranking

performance and discriminative power, which is critical given the observed class imbalance in the depression outcome. (insert table/supplementary data for hyperparameters).

We inspected the outcome distribution to justify modelling choices and retain the natural measurement scale. We predicted the mental health outcomes (Anxiety, Stress, Depression) with supervised machine learning methods using multiple sleep mediators, resource insecurity and socio-demographic variables as predictors. Training was first performed using a 10-fold cross-validation (CV) procedure, in which the hyperparameters of the methods are optimized, and performance is evaluated. A 10-fold CV was chosen over, for instance, a leave-one-out or a 5-fold CV to balance the computational efficiency and robustness of our performance estimates. Anxiety and Stress were modelled as a regression task, while Depression was modelled as a binary classification. Both regression and classification methods use matrices specific to the method, which efficiently and accurately report the model's performance. Hyperparameter optimization for each model (the three) was performed using the Tree-structured Parzen Estimator from the Optuna library, which optimizes the path through the hyperparameter space. R2 score and ROC_AUC were used as optimization metrics, both in the maximizing direction, with 1,000 trials to explore all possible paths. With regression, this ensures that hyperparameter tuning prioritizes configurations that best capture systematic variations in the response rather than minimizing pointwise error alone. Regarding classification, optimizing ROC-AUC focuses on ranking performance rather than a single cutoff, which is particularly important given the class imbalance in the depression outcome.

For predictions, we fit two non-parametric and nonlinear tree-based model families: eXtreme Gradient Boosting (XGB) and Random Forest (RF). These models were selected for their ability to flexibly capture complex, nonlinear relationships and higher-order interactions among socio-demographic characteristics, resource insecurity indicators, sleep mediators, and mental health outcomes without requiring strong parametric assumptions. Tree-based ensembles are also robust to outliers, mixed data types, and moderate departures from normality, making them well-suited to the heterogeneous structure of household-level survey data.

On the one hand, Random Forest provides a strong baseline through variance reduction via bootstrap aggregation, while XGB offers a complementary boosting framework that sequentially corrects residual errors and often achieves superior predictive performance under careful regularization. Using both approaches allowed us to assess the stability of predictive patterns across ensemble strategies, reduce reliance on a single modelling paradigm, and compare predictive performance without relying on a single modelling paradigm. Additionally, both models support internally consistent measures of feature importance, enabling post-hoc interpretation of key drivers of resource insecurity and mental health outcomes.

On the other hand, Tree-based models capture complex nonlinear relationships and interactions that cannot be summarized by a single coefficient per feature. To interpret these models, we employed Shapley Additive exPlanations (SHAP), a local explanation framework grounded in cooperative game theory that assigns Shapley values to quantify each feature's contribution to an individual prediction.

This approach preserves the nonlinear structure learned by the models while enabling consistent and additive explanations. For a given observation, a Shapley value represents the extent to which an associated feature shifts the model prediction from the expected (mean) prediction, either positively or negatively. These contributions are additive and sum to the model output, allowing for transparent decomposition of predictions at the individual level. Local explanations were further aggregated to derive global measures of feature importance by averaging absolute Shapley values across individuals. Because of the additive property of Shapley values, contributions from related features could be combined to assess the importance of feature groups or conceptual domains.

Global feature importance estimates were computed within the 10-fold cross-validation framework to account for sampling variability and to construct confidence intervals. For each fold, SHAP attributions were generated on the validation set using TreeExplainer, with a 1,000-observation background sample drawn from the corresponding training data to stabilize the expected value estimation. Feature importance was assessed using the best-performing model, XGBoost. Final local explanations and stratified analyses were conducted using an XGBoost model trained on the full dataset, with hyperparameters selected via cross-validation.

Ethics Approval

The study received ethical approval from the Africa Medical Research Institute (AMREF) Ethics and Scientific Research Committee (ESRC-P1396-2023), the Institutional Review Boards at the University of Texas at Arlington (2021 – 0630) and the University of Notre Dame (21-05-6613). All participants provided informed consent before participating in the study. The study adhered to the ethical principles outlined in the Belmont Report and the Declaration of Helsinki.

Results

Population Characteristics

This study examines the predictive effect of four resource insecurities: food, water, energy, and housing. In this study, resource insecurity refers to uncertain or limited access to necessities. To assess their impact, machine learning models were used to build insecurity scores for three mental health outcomes among PLHIV in Kenya. The study included 1,132 participants (11.9% male, 88.1% female), with an average age of 44 years (range: 16–85). Most participants (73.6%) had only primary education, and 67.7% lived in Seme Sub-Counties.

Figure 2 presents the empirical distributions of the three mental health outcomes. Anxiety scores show a moderately dispersed, multimodal distribution, indicating diverse symptoms of severity. Stress scores are more tightly concentrated, with a pronounced peak that suggests clustering at a specific scale and few extreme values. Depression is binary by design, with a class imbalance where positive cases outnumber negative cases by almost two to one. These patterns guided the modelling choices. Anxiety and Stress were treated as continuous outcomes and analyzed with regression, while Depression was

modelled as a binary classification task. The imbalance in Depression prompted the use of threshold-independent performance metrics, such as ROC-AUC, during model training and evaluation.

Predictive performance and model comparison

Figure 3 compares the cross-validated performance of eXtreme Gradient Boosting (XGB) and Random Forest (RF) models for the three mental health outcomes. Optimized parameters were used in both models. Results are shown as distributions across 10 cross-validation folds, with mean values labelled. For Anxiety and Stress, performance is measured by R2. For Depression, ROC-AUC is used. Model performance across folds remains stable, without extreme degradation.

For Anxiety, XGB outperformed RF with a mean R2 of 0.28 versus 0.23. For Stress, both performed better, but XGB led slightly (mean R2 = 46% vs. 45%). Stress outcomes were more predictable than Anxiety for both models. For Depression, both models had mean ROC-AUCs close to 0.84 and minor differences. This suggests that the results were not dependent on the chosen algorithm. XGB had a slightly higher mean and lower variability, supporting its use for interpretability analysis. Overall, XGB matched or exceeded RF and was selected as the final model for SHAP-based analysis.

To identify key mental health predictors, we aggregated mean absolute SHAP values across individuals and folds for all three XGBoost models. Figure 4 shows these rankings. Sleep mediators were highest, followed by resource insecurity indicators. Both outperformed socio-demographic variables, such as age, gender, and education.

Across all outcomes, sleep variables were most influential. For General Anxiety, Insomnia ranked highest. Reduced sleep and sleeplessness followed, accounting for most of the model's variation. This pattern appeared for both Stress and Depression, highlighting the mediating role of sleep between resource insecurity and mental health. The dominance of sleep mediators shows that sleep health is the key determinant of psychological distress in this group.

Beyond sleep variables, resource insecurity indicators ranked high in all models. Composite measures of household hardship, such as HEPEScale and HFIAScale, and specific domains, such as energy utility insecurity and housing satisfaction, consistently appeared among the top predictors. For Anxiety, HEPEScale (0.66 ± 0.06) and HFIAScale (0.44 ± 0.05) followed the sleep variables. This suggests subjective environmental stressors increase anxiety risk. For Stress and Depression, economic and environmental factors, including housing satisfaction and food security, interacted with sleep quality to influence the probability of outcomes. Their consistent presence among high-importance features shows that material deprivation and unstable living conditions have both direct and indirect effects on mental health, including through sleep disruption.

Socio-demographic covariates showed relatively modest global SHAP values in all three models. Age, education, gender, and household composition contributed to predictions but were less important than sleep mediators and resource insecurity variables. The low importance of demographic factors suggests

that sleep deprivation and resource scarcity impact mental health across all demographic groups in the studied neighbourhoods.

Ranking patterns were consistent across Anxiety, Stress, and Depression, suggesting a shared structure. Minor differences in order reflect outcome sensitivities. Stress was more closely linked to economic factors (Hwiscscale, HFIA Scale), while Anxiety was mainly influenced by sleep. The main predictors were sleep and resource insecurity for all outcomes. This suggests interventions on these factors may benefit several aspects of mental health.

Figure 5 shows SHAP beeswarm plots for depression, anxiety, and stress, with dependence plots for main sleep mediators. Global importance shows an average effect size. Local explanations reveal individual-level feature effects, showing heterogeneity, direction, and nonlinearity.

The beeswarm plots show large vertical dispersion in sleep mediators. This indicates substantial variability between individuals in their contributions to predicted outcomes. Higher values of Insomnia, reduced sleep, and sleeplessness yielded positive SHAP values, leading to upward shifts in predicted symptoms. Lower values produced negative SHAP values, offering protection. This spread confirms that sleep health is the main differentiator of mental health risk across individuals. The results consistently support the hypothesis that sleep disruption increases psychological distress.

Resource insecurity features also showed broad SHAP distributions. High insecurity levels consistently pushed predictions toward worse mental health outcomes. However, their SHAP magnitudes showed greater heterogeneity than those of sleep mediators. This suggests that resource insecurity's impact may be conditional on other factors, such as current sleep status and household composition. This variability is consistent with a pathway where resource insecurity directly affects mental health and also has indirect effects through sleep disruption.

Dependence plots for the sleep mediators further clarified these relationships. For all three outcomes, predicted risk increased steadily as sleep problems worsened. Nonlinearities appeared as steeper slopes with higher disturbance levels. These patterns show that small sleep problems cause minor outcome changes, but severe sleep disruption leads to large increases in mental health burden.

Local patterns were consistent for depression, anxiety, and stress, supporting a shared mechanism. Stress was more clustered and had steeper gradients, aligning with higher predictability. Depression had a broader spread, reflecting its binary nature and the influence of unmeasured factors.

Discussion

This study examined the predictive effect of four forms of resource insecurity, food, water, energy and housing, on three mental health outcomes among PLHIV in Kenya using machine learning models. The application of machine learning provided deeper insights into the mental health consequences of resource insecurity and enabled the identification of the specific insecurities contributing most

significantly to mental health risk, beyond what conventional parametric models could reveal. Consistent with previous studies demonstrating links between sleep disorders, resource insecurities and mental health disorders, our findings indicate that both resource insecurity and sleep disorders were positively associated with multiple dimensions of mental health disorders.

Resource insecurity ranked among the most influential predictors of mental health disorders across all models via the SHAP value analysis. Composite measures of household resource deprivation (e.g., energy insecurity (HEPEScale), food insecurity (HFIAScale) and water insecurity (HWIScale) as well as specific domains of housing insecurity, such as essential utility and housing satisfaction), consistently appeared among the top predictors. For anxiety, the household energy insecurity and food insecurity ranked immediately after sleep variables, suggesting that subjective experiences of environmental stressors substantially increase anxiety risk. Comparable findings have been reported in the US, where individuals experiencing energy insecurity had higher odds of reporting symptoms of depression and anxiety [39]. Although the health impacts of energy insecurity in LMICs are often framed in terms of respiratory illnesses and their associated illnesses [40, 41], our findings highlight its potential role as a predictor of mental health outcomes. This finding further corroborates a study by [33], which reported that in communities where women and children bear the primary responsibility of meeting household energy needs, the stress and uncertainty associated with securing energy for cooking and domestic use can contribute to psychological problems.

Most significantly, the global feature importance assessment showed that water insecurity was the primary driver for both depression and stress symptomatology among participants, making it a robust predictor of stress and depression. This is consistent with a study in rural Uganda, where water insecurity was significantly associated with severe depression symptoms [42]. In our models, water insecurity exhibited the highest global importance for stress and depression, followed by food insecurity, energy insecurity, and housing satisfaction. The significant associations of both water and food insecurity with depressive symptomatology among people living with HIV (PLHIV) in Kenya further corroborate findings from earlier research documenting a syndemic effect of water and food insecurity on depressive symptoms among women living with HIV in Kenya, in which the interaction of water insecurity and food insecurity had a multiplicative effect on depressive symptomatology [43]. This makes structural determinants such as food and water insecurity key targets for any form of depression interventions among PLHIV in Kenya.

Another significant finding of this study is the strong influence of sleep health on anxiety, which further demonstrates the importance of SHAP value analysis in enhancing the interpretability of machine learning models. SHAP analysis revealed that insomnia exhibited the highest global importance, closely followed by overall resource insecurity, with these factors collectively explaining the majority of model-predicted variation. By providing transparent insights into the relative contribution of each predictor, the machine learning approach was able to identify sleep health as the primary driver of generalized anxiety disorder among PLHIV in Kenya, thereby improving real-world applicability and interpretability of the findings. The findings show that sleep was the primary driver of generalized anxiety disorder among

PLHIV in Kenya. This was followed by specific resource insecurity indicators, namely energy insecurity, food insecurity, housing satisfaction, and water insecurity. This finding aligns with previous research reporting a strong association between insomnia and anxiety [44]. Given the study population, this result is not unexpected, as prior studies have documented a high prevalence of insomnia and sleep disturbances among PLHIV [45, 46]. Consistently, our study findings also support the hypothesized mediating role of sleep in the relationship between household resource insecurity and mental health outcomes. A systematic review and meta-analysis by Arenas et al. [47] reported strong associations between food insecurity and depression, anxiety and sleep disorders, highlighting the interconnected nature of these conditions. These results have important implications for healthcare providers working with PLHIV. Sleep disorders have been identified as underlying contributors to various mental health disorders, and evidence suggests that treating insomnia can lead to improvements in other mental health symptoms [48]. Accordingly, we recommend that PLHIV presenting with sleep disorders should be routinely screened for symptoms of anxiety, depression and stress to facilitate early identification and integrated management of comorbid conditions.

Socio-demographic covariates showed relatively modest global SHAP values across all three models. Although variables such as age, education, gender, and household composition contributed to predictive performance, their relative importance was considerably lower than that of sleep-related mediators and resource insecurity variables. The limited influence of demographic factors indicates that the effects of resource insecurity and sleep deprivation on mental health are widespread across demographic groups within the studied neighbourhoods, rather than being restricted to specific groups.

Limitation

Although the use of machine learning enhanced predictive performance and enabled the identification of key contributors to mental health outcomes, the findings should be interpreted with caution due to potential limitations related to data dependency, model complexity, and limited causal inference. In addition, the cross-sectional design restricts the ability to establish temporal and causal relationships between resource insecurity, sleep health, and mental health outcomes. Nonetheless, the findings provide important insights into the clustering of structural vulnerabilities and psychological distress among PLHIV in low-resource communities. Future longitudinal studies are needed to further examine the directionality and dynamic interplay of these relationships over time.

Conclusion

This study identifies resource insecurity as a significant determinant of mental health outcomes among PLHIV in Kenya, partly mediated through sleep-related pathways. This suggests that material deprivation and unstable living conditions have both direct and indirect effects on mental health. Using stable and generalizable machine learning models, the study further highlights the urgency of addressing basic resource needs and sleep quality as possible interventions that can improve mental health outcomes independent of demographic characteristics. The low importance of demographic factors suggests that

resource insecurity and sleep deprivation impact mental health across all demographic groups in the studied neighbourhoods. These findings underscore the importance of designing mental health interventions for PLHIV in resource-limited settings that address broader structural determinants alongside individual-level factors.

Declarations

Acknowledgements

This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program, grant number (102849). The funding source had no involvement in the study design, analysis, or interpretation of the data. We wish to appreciate Hasbergs Analytics and Researchers from Pamoja Community-Based Organization for their assistance with data collection and management.

Author contributions

Godfred O. Boateng: Conceptualization, Writing – Original draft, Methodology, Formal analysis, Investigation, Project administration, writing – review and editing, supervision; **Caleb Osae-Morgan**: Conceptualization, Writing –Original draft, Methodology, Formal analysis; **Sylvia E. Gyan**: Writing – original draft, methodology, writing – review and editing; **Patrick Owuor**: Methodology, Investigation, Project administration, Writing -review and editing; **Ellis Adjei Adams**: Methodology, Investigation, Project administration, Writing -review and editing

Competing interests

The authors declare no competing interests.

Funding

This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program, grant number (#102849). The funding source had no involvement in the study design, review or interpretation of data.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics declarations

The researcher confirms that all research was performed in accordance with relevant guidelines/regulations applicable when human participants are involved (e.g., Declaration of Helsinki). The participants provided written consent before participating in this study.

References

1. Arias, D., Saxena, S. & Verguet, S. Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine* **54**, 101675 (2022).
2. Rehm, J. & Shield, K. D. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr. Psychiatry Rep.* **21**, 10 (2019).
3. World Health Organization (WHO). Over a billion people living with mental health conditions – services require urgent scale-up. (2025). <https://www.who.int/news/item/02-09-2025-over-a-billion-people-living-with-mental-health-conditions-services-require-urgent-scale-up>
4. Whiteford, H. A. et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
5. Becker, A. E., Raviola, G. & Kleinman, A. Introduction: How Mental Health Matters. *Daedalus* 152, 8–23 (2023).
6. World Health Organization (WHO). World Mental Health Day 2022 | WHO | Regional Office for Africa. (2022). <https://www.afro.who.int/regional-director/speeches-messages/world-mental-health-day-2022>
7. World Health Organization. WHO releases new reports and estimates highlighting urgent gaps in mental health. (2025). <https://www.who.int/news/item/02-09-2025-who-releases-new-reports-and-estimates-highlighting-urgent-gaps-in-mental-health>
8. Kupferberg, A. & Hasler, G. The social cost of depression: Investigating the impact of impaired social emotion regulation, social cognition, and interpersonal behavior on social functioning. *J. Affect. Disorders Rep.* **14**, 100631 (2023).
9. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry.* **9**, 137–150 (2022).
10. Gariépy, G., Danna, S. M., Hawke, L., Henderson, J. & Iyer, S. N. The mental health of young people who are not in education, employment, or training: a systematic review and meta-analysis. *Soc. Psychiatry Psychiatr Epidemiol.* **57**, 1107–1121 (2022).
11. Porru, F., Schuring, M., Hoogendijk, W. J. G., Burdorf, A. & Robroek, S. J. W. Impact of mental disorders during education on work participation: a register-based longitudinal study on young adults with 10 years follow-up. *J. Epidemiol. Community Health.* **77**, 549–557 (2023).
12. Remien, R. H. et al. Mental health and HIV/AIDS: the need for an integrated response. *AIDS* **33**, 1411 (2019).
13. Boateng, G. O. & Geography Health and Sustainability. in Resource insecurity and gendered inequalities in health: A challenge to sustainable livelihood 127–144 (Routledge, (2021).
14. Benzekri, N. A. et al. Food insecurity and mental health outcomes among people with HIV in West Africa. *AIDS* **39**, 1262 (2025).

15. Onyeaka, H. et al. The Intersection of Food Security and Mental Health in the Pursuit of Sustainable Development Goals. *Nutrients* **16**, 2036 (2024).
16. Trudell, J. P., Burnet, M. L., Ziegler, B. R. & Luginaah, I. The impact of food insecurity on mental health in Africa: A systematic review. *Social Science Medicine*. **278**, 113953 (2021).
17. Weaver, T. L., Kelton, K. & Riebel, J. The Relationship between Women's Resources and Health-Related Quality of Life in a Sample of Female Victims of Intimate Partner Violence. *J. Soc. Serv. Res.* **47**, 565–578 (2021).
18. Weaver, L. J., Owens, C., Tessema, F., Kebede, A. & Hadley, C. Unpacking the black box of global food insecurity and mental health. *Social Science Medicine*. **282**, 114042 (2021).
19. Miller, J. D. et al. Household Water and Food Insecurity Are Positively Associated with Poor Mental and Physical Health among Adults Living with HIV in Western Kenya. *J. Nutr.* **151**, 1656–1664 (2021).
20. Logie, C. H. et al. Does resource insecurity drive HIV-related stigma? Associations between food and housing insecurity with HIV-related stigma in cohort of women living with HIV in Canada. *J. Int. AIDS Soc.* **25**, e25913 (2022).
21. Nawfal, E. S. et al. Associations of Unmet Food and Housing Needs with Mental Health and Overall Perceived Health Among Women with HIV: Is There a Moderating Effect of Social Support? *Women's Health Reports* **6**, (2025). whr.2024.0120.
22. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
23. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer New York, 2009). 10.1007/978-0-387-84858-7
24. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals Stat.* **29**, 1189–1232 (2001).
25. Hooker, G. Discovering additive structure in black box functions. in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* 575–580 (ACM, Seattle WA USA, (2004). 10.1145/1014052.1014122
26. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. Royal Stat. Soc. Ser. B (Methodological)*. **58**, 267–288 (1996).
27. Zou, H. & Hastie, T. Regularization and Variable Selection Via the Elastic Net. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **67**, 301–320 (2005).
28. Kuhn, M. & Johnson, K. *Applied Predictive Modeling* (Springer New York, 2013). 10.1007/978-1-4614-6849-3
29. Yarkoni, T. & Westfall, J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
30. Lundberg, S. M. & Lee, S. I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., (2017).
31. Molnar, C. *Interpretable Machine Learning* (Lulu.com, 2020).

32. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).
33. Boateng, G. O., Balogun, M. R., Dada, F. O. & Armah, F. A. Household energy insecurity: dimensions and consequences for women, infants and children in low- and middle-income countries. *Soc. Sci. Med.* **258**, 113068 (2020).
34. Coates, J., Swindale, A. & Bilinsky, P. Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide: Version 3: (576842013-001). (2007). <https://doi.org/10.1037/e576842013-001>
35. Boateng, G. O. & Adams, E. A. A multilevel, multidimensional scale for measuring housing insecurity in slums and informal settlements. *Cities* **132**, 104059 (2023).
36. Young, S. L. et al. The Household Water InSecurity Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low-income and middle-income countries. *BMJ Glob Health.* **4**, e001750 (2019).
37. Boateng, G. O. & Boateng, M. O. Optimal Test Construction. *Encyclopedia of Measurement in Social Sciences, Second Edition* (2026).
38. Douglass, A. B. et al. The Sleep Disorders Questionnaire I: Creation and Multivariate Structure of SDQ. *Sleep* **17**, 160–167 (1994).
39. Graff, M. & Aung, T. W. Energy Insecurity and Mental Health Symptoms in US Adults. *JAMA Netw. Open.* **8**, e2539479 (2025).
40. Ang’u, C., Muthama, N. J., Mutuku, M. A. & M’IKiugu, M. H. Analysis of energy poverty in Kenya and its implications for human health. *Energy Policy.* **176**, 113506 (2023).
41. Phoumin, H. & Kimura, F. Cambodia’s energy poverty and its effects on social wellbeing: Empirical evidence and policy implications. *Energy Policy.* **132**, 283–289 (2019).
42. Mushavi, R. C. *When you have no water, it means you have no peace*: A mixed-methods, whole-population study of water insecurity and depression in rural Uganda. *Social Science Medicine.* **245**, 112561 (2020).
43. Boateng, G. O. et al. The syndemic effects of food insecurity, water insecurity, and HIV on depressive symptomatology among Kenyan women. *Social Science Medicine.* **295**, 113043 (2022).
44. Merrill, R. M. Mental Health Conditions According to Stress and Sleep Disorders. *IJERPH* **19**, 7957 (2022).
45. Faraut, B. et al. Sleep Disturbance and Total Sleep Time in Persons Living with HIV: A Cross-Sectional Study. *AIDS Behav.* **22**, 2877–2887 (2018).
46. Osiyemi, A. O. et al. Sleep disturbance and associated factors among Nigerian adults living with HIV in the dolutegravir era. *Front Sleep* **1**, (2022).
47. Arenas, D. J., Thomas, A., Wang, J. & DeLisser, H. M. A Systematic Review and Meta-analysis of Depression, Anxiety, and Sleep Disorders in US Adults with Food Insecurity. *J. Gen. Intern. Med.* **34**, 2874–2882 (2019).

48. Freeman, D., Sheaves, B., Waite, F., Harvey, A. G. & Harrison, P. J. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry*. 7, 628–637 (2020).

Figures

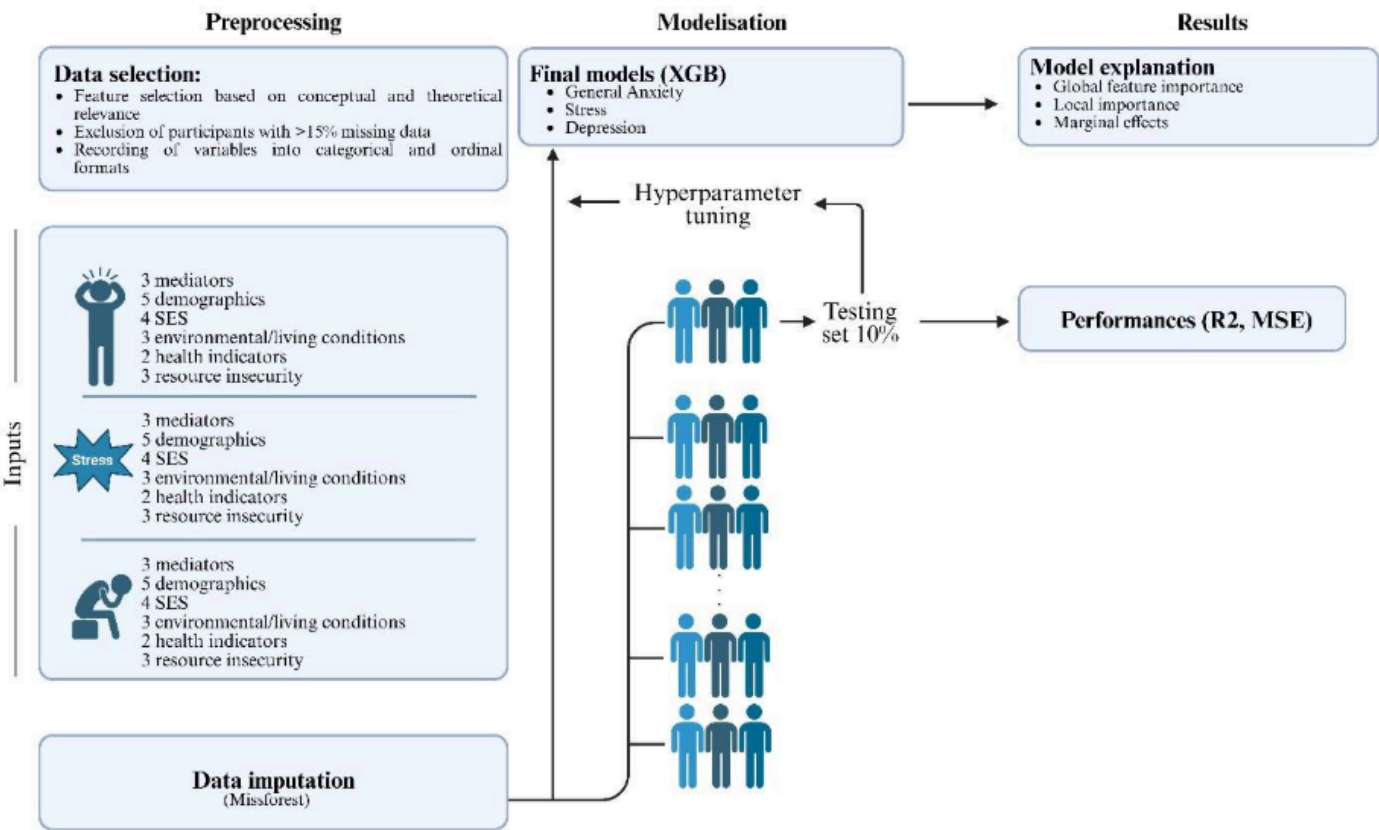


Figure 1
Analytical workflow chart of the predictors of mental health

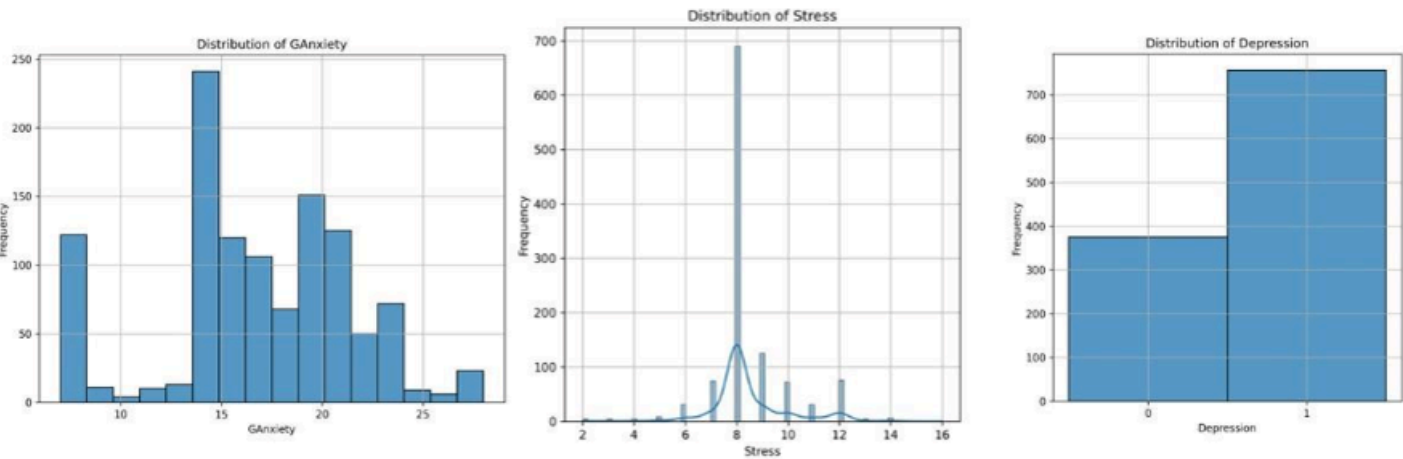


Figure 2
Page 19/21

Normal distribution of mental health outcomes

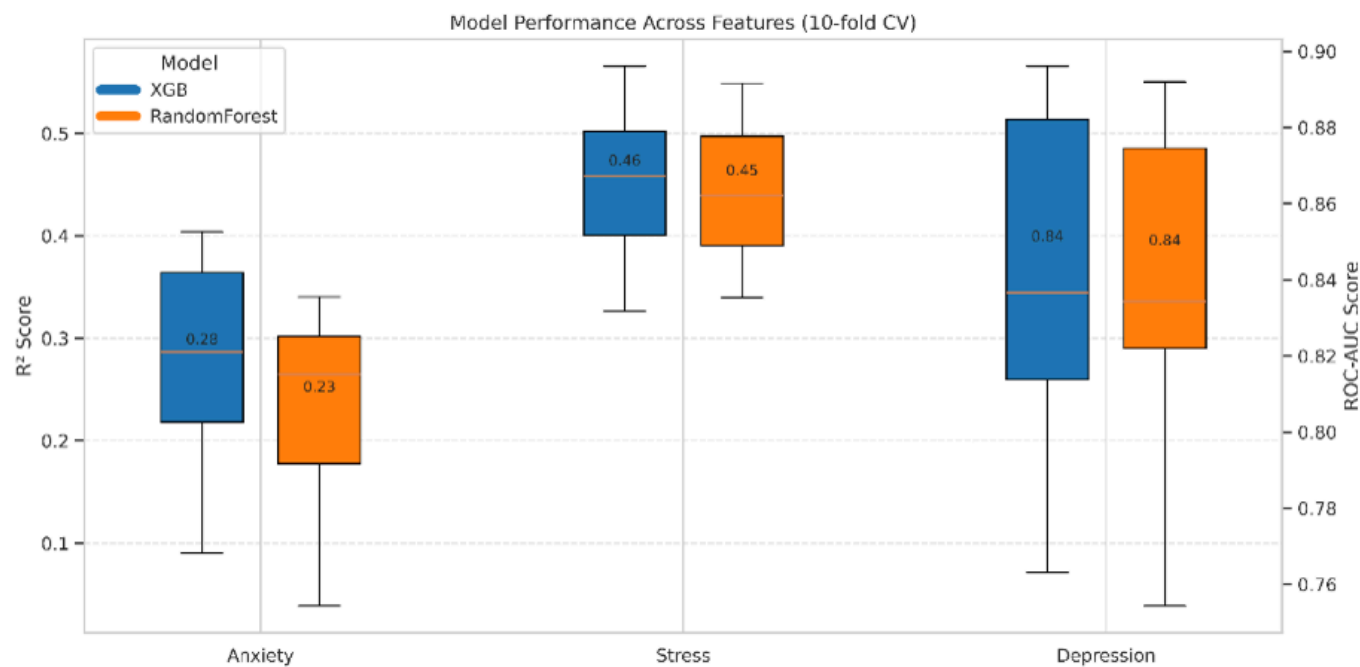


Figure 3

Models' performance comparison obtained after CV and hyperparameter optimization

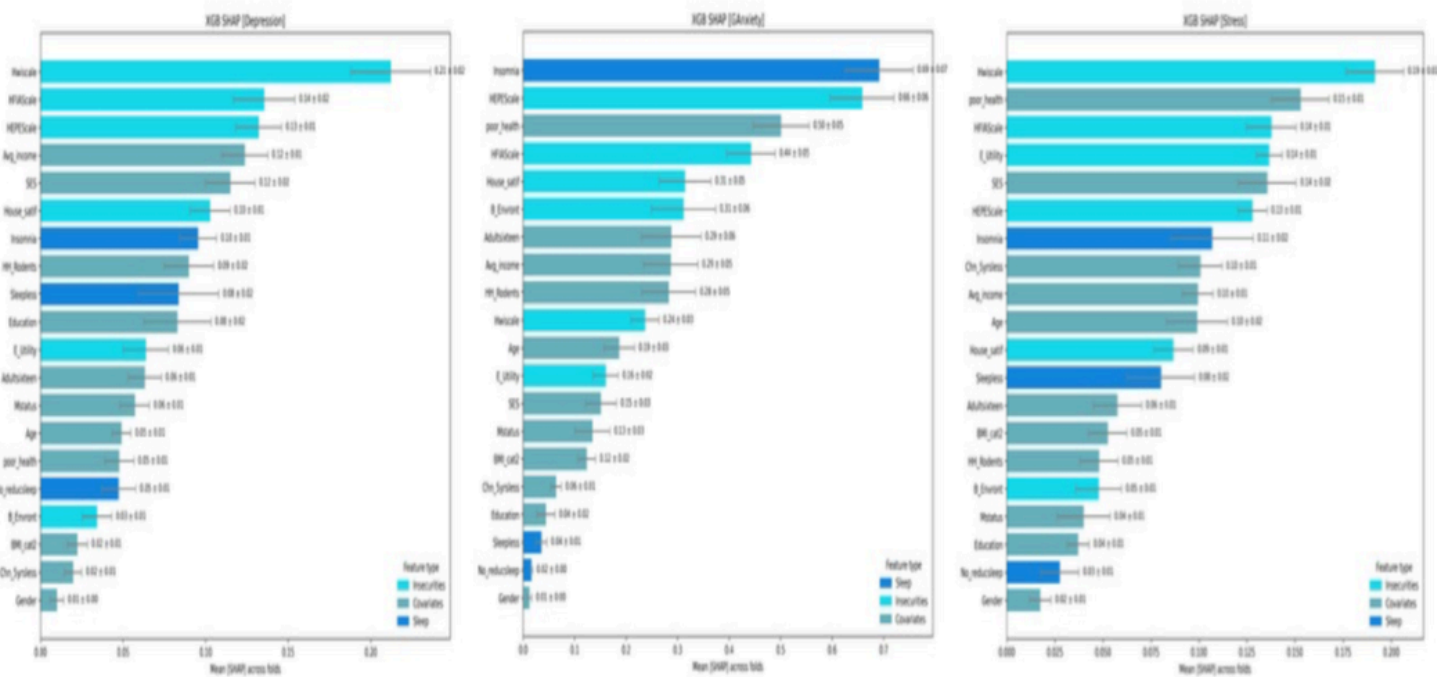


Figure 4

Global feature importance of depression, anxiety, and stress

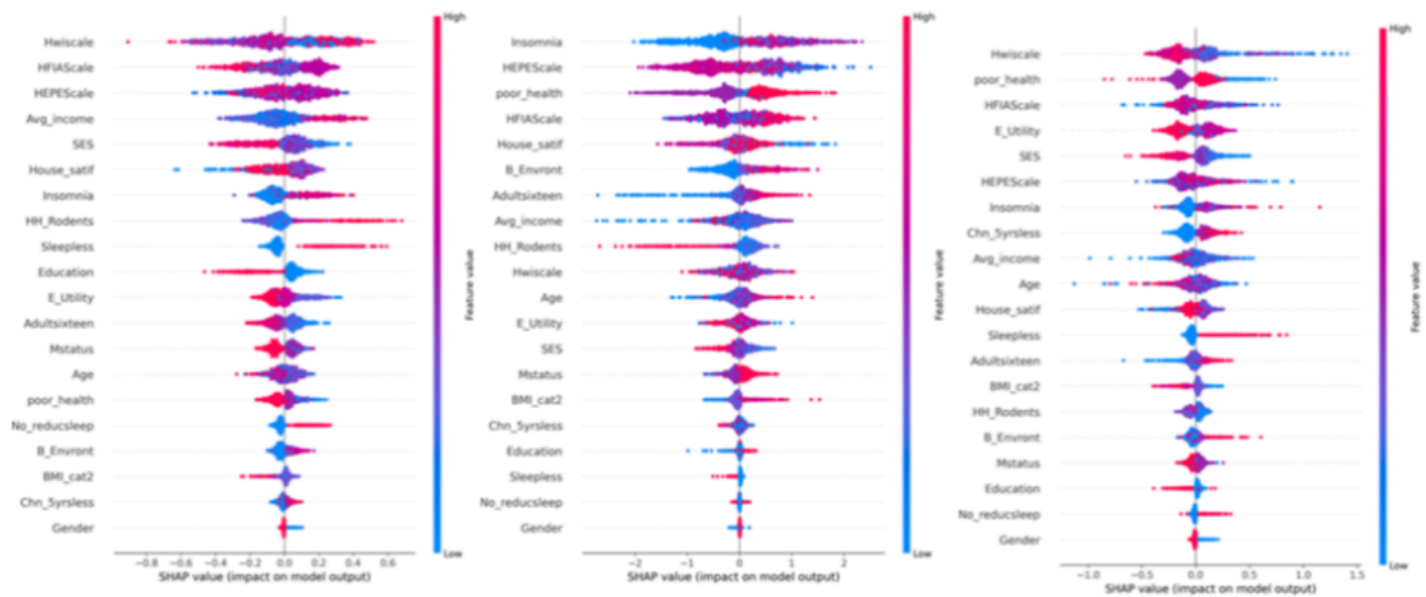


Figure 5

Local SHAP values (beeswarm plot) of depression symptomatology, anxiety, and stress with sleep dependence plots