International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Using a Polynomial Regression Machine Learning Model to Predict Depression Severity Among People Living with HIV

Geofrey Nyabuto, Peters Anselemo Ikoha, Samuel Mungai Mbuguah
Department of Information Technology, Kibabii University, Bungoma, Kenya

Abstract- Background: Binary depression screening does not distinguish patients with mild symptoms from those with clinically urgent symptom burden. Among people living with HIV (PLHIV), depressive-symptom severity may reflect nonlinear interactions among anxiety, immune status, treatment adherence, stigma, behavioural exposures, and demographic characteristics. Routine electronic medical records (EMRs) provide an opportunity to model these relationships using computationally reproducible methods. Objective: To develop and internally validate a polynomial regression machine learning model for predicting continuous PHQ-9 depressive-symptom severity among PLHIV using routine HIV-care EMR data. Methods: A cross-sectional patient-level analytical dataset was constructed from 54,301 de-identified EMR records from Bungoma and Busia counties, Kenya. Fifteen clinical, treatment, psychosocial, behavioural, and demographic predictors were processed using training-derived imputation, encoding, log transformation, and standardisation. Stratified training (n=38,010), validation (n=8,145), and held-out test (n=8,146) partitions were used. Ordinary least-squares regression with degree-1, degree-2, and degree-3 polynomial feature expansions was compared using R^2, adjusted R^2, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Results: The degree-3 model comprised 816 fitted parameters and achieved the strongest held-out performance: R^2 0.753, adjusted R^2 0.730, MAE 1.808, MSE 6.359, and RMSE 2.522 PHQ-9 points. The linear model achieved R^2 0.729, MAE 1.939, and RMSE 2.641. Mean and median cubic-model residuals were 0.040 and 0.005, respectively, although the maximum positive residual was 15.186, indicating important underprediction in a small number of high-severity cases. The largest reported terms were anxiety × CD4 (β=−0.444) and anxiety² (β=0.413). Conclusions: Degree-3 polynomial regression modestly improved prediction of PHQ-9 depressive-symptom severity over linear and quadratic alternatives. Its average error may support broad risk stratification, but threshold crossing, model complexity, coefficient instability, and severe case underprediction preclude autonomous clinical use.

Keywords- Depression severity; PHQ-9; HIV; polynomial regression; machine learning; electronic medical records; nonlinear prediction; Kenya; clinical decision support.

I. INTRODUCTION

The relationship between mental health and HIV is well established, with depression occurring more frequently among people living with HIV (PLHIV) than in the general population (Remien et al., 2019; Molapo et al., 2025; Tadesse et al., 2025). Depressive effects extend beyond emotional suffering as symptoms can reduce concentration, motivation, treatment adherence, appointment attendance, self-care, social functioning, and quality of life. Reviews from Africa continue to document a substantial burden of depressive symptoms among PLHIV (Boakye et al., 2024). Facility-based evidence from Ethiopia and Nigeria likewise demonstrates that symptom burden is associated with interrelated clinical and sociodemographic conditions (Adedeji et al., 2023; Gebru et al., 2024). Depression detection is therefore central to comprehensive HIV care.

The clinical consequences of missed or underestimated symptoms are particularly important. Depression has been associated with ART non-adherence and poorer viral-suppression outcomes, creating a feedback loop in which psychological distress affects treatment behaviour while treatment difficulty and concern about disease progression may

worsen distress (Huang et al., 2025; Zeleke et al., 2025). HIV-related stigma, limited social support, violence exposure, comorbidity, nutritional vulnerability, and advanced disease can add further burden (Abebe et al., 2025; Dessie & Zewotir, 2024; Duko et al., 2024). Integrated mental-health and HIV services are consequently recommended, but limited specialist capacity, time pressure, confidentiality concerns, and weak referral systems continue to impede routine implementation (WHO, 2022).

Clinical need is not adequately represented by a binary depressed/not-depressed label. A patient with mild symptoms may require monitoring and brief intervention; a patient with moderate symptoms may require structured assessment and follow-up; and a patient with moderately severe or severe symptoms may require urgent clinical review, including direct assessment of self-harm risk. The nine-item Patient Health Questionnaire (PHQ-9) provides a summed severity score from 0 to 27 and conventional bands of none/minimal, mild, moderate, moderately severe, and severe symptoms (Kroenke et al., 2001; Kroenke & Spitzer, 2002). Predicting the continuous score can therefore preserve gradations that are lost when the outcome is collapsed into a binary screening class.

The PHQ-9 nevertheless remains a screening and symptom-severity instrument rather than a substitute for psychiatric diagnosis. A recent evaluation among PLHIV showed that its operating characteristics against psychiatric assessment depend on context and threshold (Yotebieng et al., 2024). This distinction defines the intended output of this study where the model estimates the PHQ-9 depressive-symptom severity score that might be observed at a clinical encounter. Any information-system implementation must preserve these boundaries in its terminology, user interface, alerts, and referral logic.

Depressive-symptom severity in HIV care is plausibly nonlinear. The association of anxiety with PHQ-9 score may become steeper at high anxiety levels; the relationship between immune compromise and distress may vary by WHO stage; adherence difficulty may be more consequential when combined with virological instability; and the influence of stigma or violence may depend on social and treatment support. Biological, psychological, and social determinants are therefore better understood as a connected system than as isolated additive inputs (Engel, 1977; Zaongo et al., 2025).

Machine learning offers a computational approach for estimating complex relationships from structured clinical data. In mental-health research, machine-learning methods have been applied to clinical, behavioral, activity, textual, biomarker, and electronic health-record data (Ahmed et al., 2025; Ali et al., 2025; Dwyer et al., 2018; Shatte et al., 2019). The value of these methods is not that they replace clinical judgement, but they can learn multivariable functions intended to generalize to previously unseen patients. In HIV informatics, artificial intelligence has also been proposed for risk stratification, service targeting, and decision support, although transportability, fairness, privacy, and workflow fit remain decisive constraints (Xiang et al., 2022).

Routine EMRs are especially relevant in resource-constrained health systems because they contain information already collected during care. Kenya's national HIV data infrastructure demonstrates the analytical value of standardised electronic records for monitoring and programme management (Ndisha et al., 2023). Extending these records from descriptive reporting to predictive analytics could help prioritise direct screening when personnel and time are limited. However, EMR prediction introduces familiar risks: missingness may reflect care processes rather than random absence; clinical coding can vary between facilities; labels may be measured concurrently with predictors; and apparent performance can be inflated by information leakage or insufficient validation (Garriga et al., 2022; Swinckels et al., 2024).

Polynomial regression is a useful intermediate method between a strictly additive linear model and a highly flexible black-box learner. It retains an explicit regression equation while expanding the design matrix with squared, cubic, and cross-product terms. The approach can represent smooth curvature and prespecified interactions, but its dimensionality grows rapidly with the number of predictors and degree. It can therefore become difficult to interpret, sensitive to scaling and extrapolation, and vulnerable to instability among correlated derived terms (Cohen et al., 2003; Harrell, 2015).

II. LITERATURE REVIEW

Depression severity in the context of HIV care

Review of literature consistently characterizes depression among PLHIV as a biopsychosocial outcome. Clinical factors such as immune status, advanced HIV disease, comorbidity, and treatment experience coexist with psychosocial factors

such as stigma, violence, social support, economic stress, and disclosure concerns. Previous studies show that no single domain adequately explains symptom burden (Boakye et al., 2024; Duko et al., 2024; Tadesse et al., 2025). This multidomain structure supports a computational model in which clinical, treatment, behavioral, psychosocial, and demographic variables enter the same prediction function. It also cautions against interpreting a high-ranked coefficient or interaction as a complete explanation of a patient's mental health state. Associations between depression, ART adherence, and viral suppression provide a further rationale for severity prediction. Meta-analytic and prospective evidence suggests that greater depressive burden can undermine adherence, while poorer treatment control may contribute to anxiety and distress (Huang et al., 2025; Zeleke et al., 2024). Because the present analysis is cross-sectional, direction cannot be determined.

Many depression-prediction studies convert PHQ-9 into a binary or multiclass label. Classification is useful for screening decisions, but it discards within category variation. A continuous model estimates the full score and allows the information system to display both a numeric value and an associated severity band. This is particularly useful near clinical thresholds, where a small prediction error can change the implied category. The PHQ-9's bounded 0–27 range and five-point bands provide a familiar interpretation framework, while its validation history supports its use as a severity measure (Kroenke & Spitzer, 2002; Kroenke et al., 2001).

2.2 Machine learning and EMR-based Depression Prediction
Machine learning in psychiatry has expanded from small biomarker studies to large, heterogeneous clinical datasets. Reviews describe classification, prognosis, symptom monitoring, and treatment-response applications, while also identifying common weaknesses: small samples, weak external validation, inconsistent outcome definitions, and limited integration into care (Dwyer et al., 2018; Shatte et al., 2019). More recent work demonstrates explainable depression detection and severity classification from activity data and general population prediction using clinical and behavioral variables (Ahmed et al., 2025; Vu et al., 2025). These studies support algorithmic feasibility but do not eliminate the need for disease-specific data and locally relevant validation.

Within HIV research, Marathe et al. (2022) showed that supervised machine learning can predict depressive symptoms in an HIV–hepatitis C cohort, while Ssentumbwe et al. (2026) reported machine-learning prediction among adolescents living with HIV in Uganda. These studies indicate that structured HIV and psychosocial variables contain predictive signal. The present study extends that literature by estimating continuous PHQ-9 severity in a large routine-care cohort and by explicitly comparing linear, quadratic, and cubic feature spaces. Direct comparison of performance values remains inappropriate because studies differ in population, instruments, prevalence, predictors, and validation design.

Polynomial regression is Machine learning algorithm that is linear in its fitted coefficients but nonlinear in its original inputs. A degree-2 expansion adds squared terms and pairwise products; a degree-3 expansion adds cubes and third-order products. This allows the model to represent curvature, synergy, attenuation, and conditional relationships while using ordinary least squares for estimation. Polynomial response-surface methods have been used to study nonlinear and congruence effects in psychological research (Ye et al., 2025). In a machine-learning workflow, the polynomial transformer is treated as feature engineering and the fitted regression is evaluated on data not used to estimate coefficients.

Evaluation, transparency, and clinical translation
Good internal prediction requires separation between model development and evaluation. Preprocessing should be learned from training data; model complexity should be selected using validation data or resampling; and final performance should be reported once on a held-out test set. R^2 quantifies variance explained, whereas MAE and RMSE express error in PHQ-9 points. RMSE gives greater weight to large errors, making it useful when severe underprediction is clinically important. Residual plots and summaries assess bias, spread, outliers, and heteroscedasticity (Cook & Weisberg, 1982; Yang et al., 2019).

Contemporary guidance emphasizes that a model is not clinically ready merely because internal metrics are favorable. TRIPOD+AI specifies transparent reporting, PROBAST+AI evaluates bias and applicability, and FUTURE-AI addresses robustness, usability, fairness, traceability, and deployability (Collins et al., 2024; Lekadir et al., 2025; Moons et al., 2025). External validation should examine calibration and subgroup error in new locations and time periods. Responsible health machine learning further requires data protection, clinician oversight, harm monitoring, and a clear pathway for confirmatory assessment (Rajkomar et al., 2018; Rajpurkar et

al., 2022; Wiens et al., 2019; World Health Organization, 2021).

III. METHODOLOGY

Study Design, Setting, and Analytical Cohort

A cross-sectional prediction-model development and internal-validation design was used. The analytical dataset comprised 54,301 de-identified patient-level EMR records from HIV-care facilities in Bungoma and Busia counties, Kenya, covering encounters from 2022 through April 2026. One index encounter was retained per patient; where several eligible visits existed, the most recent record with sufficient PHQ-9, GAD-7, clinical, treatment, behavioural, and psychosocial information was selected. The model consequently estimates severity at the encounter and does not predict onset, recurrence, or within-patient change.

Outcome, Predictors, and Preprocessing

The outcome was the PHQ-9 total score from 0 to 27. For interpretation, scores were grouped as none/minimal (0–4), mild (5–9), moderate (10–14), moderately severe (15–19), and severe (20–27). Fifteen predictors were fixed before expansion: GAD-7 anxiety rating, CD4 count, current WHO stage, BMI, ART adherence, HIV-related stigma, gender-based violence history, alcohol-use frequency, treatment-supporter status, log viral load, age, sex, duration on ART, comorbidity status, and marital status. Data cleaning included duplicate and plausibility checks, categorical encoding, log(VL+1), and standardisation. CD4 was missing for 16,290 patients (30.0%) and imputed using WHO-stage-stratified training medians; 174 illogical ART-duration values (0.32%) were validated and median-imputed.

Table 1. Predictor Vector for PHQ-9 severity modelling

Symbol	Predictor	Coding/preprocessing
X_1	GAD-7 anxiety rating	Ordinal category; standardised
X_2	CD4 count	Continuous; WHO-stage-stratified imputation; standardised
X_3	Current WHO stage	Ordinal; standardised
X_4	Body mass index	Continuous; standardised
X_5	ART adherence	Binary/ordered coding; standardised
X_6	HIV-related stigma	Binary coding
X_7	Gender-based violence history	Binary coding
X_8	Alcohol-use frequency	Ordinal; standardised
X_9	Treatment supporter	Binary coding
X_{10}	Viral load	log(VL+1); standardised
X_{11}	Age at visit	Continuous; standardised
X_{12}	Sex	Binary coding (female=1)
X_{13}	Duration on ART	Continuous; standardised
X_{14}	Comorbidity status	Binary coding
X_{15}	Marital status	Categorical/ordinal coding

Data Partitioning and Model Specification

The cohort was partitioned using a stratified 70:15:15 split into training (n=38,010), validation (n=8,145), and held-out test (n=8,146) sets. Training data were used to estimate imputation constants, encoders, scaling parameters, and regression coefficients; validation data supported model-degree selection; and the test set was reserved for final internal evaluation. The computational workflow was implemented using Python and scikit-learn (Pedregosa et al., 2011). Degree-1, degree-2, and degree-3 ordinary least-squares models were fitted. For 15 predictors, the degrees contained 16, 136, and 816 fitted parameters including the intercept. The degree-3 equation included 15 linear, 120 second-degree, and 680 third-degree terms.

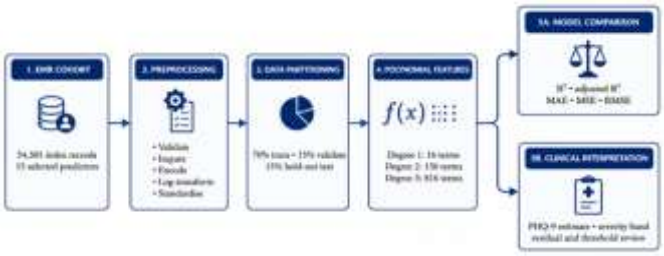

Figure 1. Electronic-medical-record pipeline for polynomial regression severity prediction.

Table 2. Data partitions and polynomial feature-space dimensions

Component	Degree 1	Degree 2	Degree 3
Training records	38,010	38,010	38,010
Validation records	8,145	8,145	8,145
Held-out test records	8,146	8,146	8,146
Intercept	1	1	1
Linear terms	15	15	15
Second-degree terms	0	120	120
Third-degree terms	0	0	680
Total fitted parameters	16	136	816

Evaluation and Ethics

The models were evaluated using R^2, adjusted R^2, MAE, MSE, and RMSE. Residuals were defined as observed minus predicted PHQ-9; positive values therefore indicated underprediction. Summary statistics and extreme residuals were examined because average performance can conceal safety-relevant errors. The study followed prediction-model principles that separate development, internal validation, and future external validation (Collins et al., 2024; Efthimiou et al., 2024; Steyerberg, 2019). Proper research approvals were obtained from Kibabii University, NACOSTI, Bungoma and Busia Counties.

IV. RESULTS, DATA ANALYSIS AND DISCUSSION

PHQ-9 Severity Distribution

Across 54,301 patients, the mean PHQ-9 score was 5.12 (SD 4.98), the median was 3, and the observed range was 0–27. Most records were in the none/minimal band (63.7%); 14.2% were mild, 17.7% moderate, 3.6% moderately severe, and 0.7% severe. The distribution was therefore right skewed, with 403 severe cases. This imbalance matters for regression because ordinary least squares minimizes average squared error across the whole cohort. The dense low-severity region exerts more influence on the fitted surface than the sparse severe tail, even though underprediction in that tail may have the greatest clinical consequence.

Table 3. Observed PHQ-9 severity distribution

Severity band	**Score**	**n**	**%**
None/minimal	0–4	34,590	63.7
Mild	5–9	7,735	14.2
Moderate	10–14	9,613	17.7
Moderately severe	15–19	1,960	3.6
Severe	20–27	403	0.7
Total	0–27	54,301	100.0

Note. Percentages are rounded as reported in the source thesis. The observed distribution supports continuous modelling but also highlights a limitation of aggregate metrics. A model can achieve a strong R^2 by fitting the dominant none/minimal, mild, and moderate ranges while remaining less reliable for severe observations.

Comparison of Polynomial Model Degrees

The degree-3 model achieved the highest R^2 and lowest error in training, validation, and held-out test data. On the test set, its R^2 was 0.753, compared with 0.733 for degree 2 and 0.729 for the linear model. Relative to degree 1, the cubic expansion increased R^2 by 0.024, reduced MAE by 0.131 PHQ-9 points, and reduced RMSE by 0.119 points. Its held-out MSE was 6.359. The improvement is consistent across metrics but modest in absolute magnitude: most of the explainable variance was already present in the 15 linear main effects, and the additional 800 non-intercept terms produced incremental rather than transformative gain.

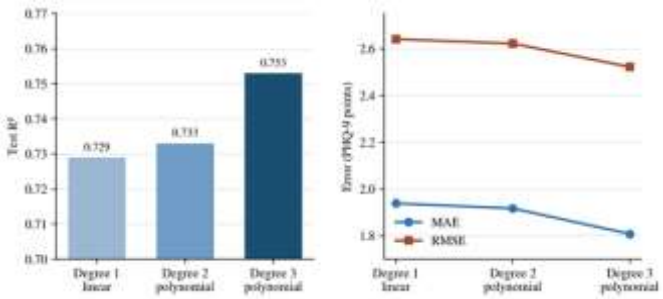

Figure 2. Held-out performance of linear, quadratic, and cubic polynomial regression models.

Table 4. Regression performance across training, validation, and held-out test sets

Model	Partition	R^2	Adjusted R^2	MAE	RMSE
Degree 1	Training	0.721	0.721	1.926	2.625
Degree 1	Validation	0.725	0.725	1.917	2.597
Degree 1	Test	0.729	0.728	1.939	2.641
Degree 2	Training	0.727	0.726	1.901	2.597
Degree 2	Validation	0.728	0.724	1.904	2.581
Degree 2	Test	0.733	0.729	1.917	2.622
Degree 3	Training	0.759	0.754	1.764	2.440
Degree 3	Validation	0.754	0.732	1.788	2.454
Degree 3	Test	0.753	0.730	1.808	2.522

Training, validation, and test R^2 values for degree 3 were 0.759, 0.754, and 0.753. Their proximity provides no obvious evidence of large internal overfitting under the selected split. The validation and test records came from the same broader data environment and period, so shared documentation, facility, and population characteristics remain. External evaluation in another county or later time-period is necessary before the degree-3 performance can be treated as generalizable (Collins et al., 2024; Steyerberg, 2019).

Adjusted R^2 fell from 0.753 unadjusted to 0.730 on the test set because the model contains 816 parameters. This complexity penalty is informative: the cubic terms improve fit, but each improvement must be weighed against implementation cost, coefficient instability, and maintainability. A comparison with restricted cubic splines, ridge regression, elastic net, gradient boosting, ordinal regression, and quantile regression could determine whether similar or better error is achievable with greater stability.

Dominant Fitted Terms and Nonlinear Response Structure

The fitted intercept was 5.120, representing the expected PHQ-9 score at the standardized reference point. The largest reported interaction was anxiety × CD4 (β=−0.444), and the largest nonlinear single-predictor term was anxiety² (β=0.413). Anxiety also appeared in interactions with adherence, age, viral load, WHO stage, alcohol use, stigma, treatment support, sex, ART duration, and gender-based violence. The pattern indicates that the model used anxiety not as a constant additive predictor but as a variable whose contribution changed across clinical, treatment, behavioural, and social contexts.

Table 5. Selected largest-magnitude coefficients from the degree-3 model

Term	β	Careful interpretation
Intercept	5.120	Fitted PHQ-9 at the standardised reference point
Anxiety × CD4	−0.444	Association of anxiety with severity changes across immune status
Anxiety²	0.413	Upward curvature at higher standardised anxiety
Anxiety × adherence²	0.175	Conditional nonlinear anxiety–adherence pattern
CD4²	0.148	Curvature in the CD4–severity relation

Term	β	Careful interpretation
Anxiety × age	0.147	Age modifies the fitted anxiety association
CD4 × adherence²	−0.138	Immune status modifies the nonlinear adherence pattern
Anxiety × viral load	0.135	Combined psychological and virological signal
Anxiety × WHO stage	0.113	Disease stage modifies the anxiety–severity relation
Anxiety × alcohol use	0.108	Combined anxiety and alcohol-use pattern
Anxiety × stigma²	0.107	Conditional anxiety–stigma pattern
ART adherence³	−0.079	Threshold-shaped fitted adherence relation

Note. Coefficients describe the fitted multivariable response surface. They are not causal estimates, and the source analysis did not report bootstrap stability or penalised shrinkage.

The negative anxiety × CD4 coefficient does not mean that anxiety or CD4 is protective. In a polynomial model, an interaction coefficient describes how the fitted slope for one variable changes with the value of another while all powers and products remain in the equation. The net prediction for a patient is the sum of 816 terms. Signs should therefore be interpreted through predicted response surfaces or marginal contrasts across realistic covariate ranges. Causal statements would require temporal ordering, confounding control, and a causal design that are absent from this cross-sectional analysis.

The anxiety² coefficient suggests upward curvature where increases in standardised anxiety have a progressively larger fitted association with severity at higher values, conditional on other terms. This is clinically plausible because high anxiety may reflect broad distress and functional impairment. It may also partly reflect construct proximity between GAD-7 and PHQ-9.

CD4² and interactions involving WHO stage, viral load, and adherence support a nonlinear biological-treatment component. Recent reviews describe HIV-associated depression as arising from interacting neuroimmune, treatment, psychosocial, and environmental mechanisms rather than a single pathway (Tadesse et al., 2025; Zaongo et al., 2025). The fitted terms are consistent with this broad framework but do not validate individual mechanisms. Their primary value is predictive and hypothesis-generating.

Residual Behavior and Clinically Important Error

The cubic model showed little average bias: mean residual was 0.040 and median residual was 0.005. The interquartile range extended from −1.496 to 1.239 points, indicating relatively tight central fit. The residual standard deviation was 2.522 points, matching the RMSE scale because mean bias was close to zero. These findings suggest that most predictions were reasonably close to observed scores and that systematic over- or underprediction was small in aggregate.

Table 6. Held-out residual summary for the degree-3 model

Statistic	PHQ-9 points	Interpretation
Mean	0.040	Minimal average bias
Median	0.005	Central residual near zero
Standard deviation	2.522	Equal to the reported RMSE scale
Minimum	−11.909	Largest overprediction
1st percentile	−5.954	Lower residual tail
25th percentile	−1.496	Lower quartile
75th percentile	1.239	Upper quartile
95th percentile	4.316	Upper-tail underprediction
Maximum	15.186	Largest underprediction; greatest safety concern

The tails tell a more cautious story. Residuals ranged from −11.909 to 15.186. Because residual equalled observed minus predicted score, the maximum represents severe underprediction. The 95th percentile was 4.316, showing that a nontrivial upper tail was underestimated by more than four PHQ-9 points. A four-point error can move a patient across a severity threshold; the maximum error could move a genuinely

severe observation into a much lower predicted band. Mean residual near zero therefore cannot be used as evidence of safety.

Prediction thresholds are particularly sensitive because PHQ-9 bands are five points wide. A model prediction of 9.2 may be labelled mild even if the observed score is 10.5; a prediction of 14.4 may obscure moderately severe symptoms; and a prediction near 20 may affect urgency. An EMR interface should display the continuous estimate, flag proximity to a threshold, and prompt direct PHQ-9 administration whenever the category would change management. Any report of self-harm thoughts must bypass the algorithm and trigger direct safety assessment.

Comparison with Related Computational Studies

The results accord with broader mental-health machine-learning research showing that nonlinear algorithms can extract clinically useful information from multivariable data (Ahmed et al., 2025; Salahudeen et al., 2025; Vu et al., 2025). Salahudeen et al. (2025) predicted depression severity using machine-learning models and clinical plus mitochondrial-peptide factors; the present study instead relies on routine HIV-care EMR data. Its comparative advantage is potential availability in ordinary service settings, while its limitation is dependence on the quality and completeness of routine documentation.

HIV-specific evidence is smaller. Marathe et al. (2022) demonstrated supervised prediction of depressive symptoms in an HIV–HCV population, and Ssentumbwe et al. (2026) applied machine learning among adolescents living with HIV in Uganda. The current study adds a much larger adult routine-care dataset and a continuous severity outcome. Direct metric comparisons are inappropriate because the studies use different populations, labels, predictor sets, and validation methods.

Polynomial regression also provides a different interpretability profile from tree ensembles and neural networks. It exposes an explicit equation, yet an 816-term equation is not automatically understandable. Global mathematical transparency can coexist with poor human interpretability. Response-surface visualization, partial dependence, accumulated local effects, and clinically defined contrasts would communicate the model more effectively than a ranked coefficient list. Explainability should be evaluated for comprehension and appropriate reliance, not merely produced as a technical output (Ahmed et al., 2025; Rajpurkar et al., 2022).

Strengths, Limitations, and Validation Priorities

Strengths include the large sample relative to the cubic parameter count, complete PHQ-9 outcomes, a fixed set of 15 predictors, one index encounter per patient, training-derived preprocessing, explicit training–validation–test separation, comparison of three model degrees, reporting of complementary error metrics, and examination of residual extremes. The use of routine EMR variables strengthens practical relevance, and the continuous outcome preserves severity information that binary classification would discard.

The limitations are substantial. The design is cross-sectional, and predictors were measured concurrently with PHQ-9. There was no external validation, temporal split, facility-cluster analysis, or repeated-measures modelling. Thirty percent of CD4 values were single-imputed, and missingness may have been informative. The degree-3 model was unpenalised; confidence intervals, bootstrap stability, expanded-matrix condition indices, and prediction intervals were not reported. The original-predictor VIF values at or below 1.002 do not demonstrate absence of collinearity among the 816 derived terms. Important variables such as psychiatric history, suicidality, food insecurity, disclosure stress, and household vulnerability were not consistently available.

The dataset came from two counties and one broader health-system context. Differences in screening practice, EMR configuration, laboratory availability, patient mix, and referral access can change performance. Validation is context-specific rather than a permanent property of a model, and contemporary guidance requires evaluation beyond the development environment (Collins et al., 2024; Moons et al., 2025; Van Calster et al., 2023).

Table 7. Evidence boundary and validation roadmap

Domain	Current evidence	Required next step
Target outcome	Concurrent continuous PHQ-9 score	Psychiatric-reference and longitudinal outcomes
Internal performance	Test R^2 0.753; MAE 1.808; RMSE 2.522	Bootstrap intervals and repeated internal validation

Domain	Current evidence	Required next step
Calibration	Mean residual near zero	Calibration plot, intercept/slope, and severity-stratified error
Model complexity	816 unpenalised coefficients	Ridge/elastic-net, spline, ordinal, and boosting comparisons
Missing data	Stage-stratified median CD4 imputation	Multiple imputation and MNAR sensitivity analysis
Transportability	Bungoma and Busia records	Temporal and geographic external validation
Fairness	Aggregate metrics only	Subgroup error and calibration analysis
Clinical utility	Not evaluated	Decision analysis, silent pilot, and prospective impact study
Implementation	Model pipeline specified	Versioned preprocessing bundle, range checks, audit, and rollback

V. CONCLUSIONS AND RECOMMENDATIONS

This study developed a polynomial regression machine learning model for predicting PHQ-9 depressive-symptom severity among 54,301 PLHIV using routine EMR data from western Kenya. The degree-3 model provided the best internal performance, achieving held-out R^2 0.753, adjusted R^2 0.730, MAE 1.808, MSE 6.359, and RMSE 2.522. It outperformed the linear and quadratic alternatives, demonstrating that nonlinear powers and interactions contain additional predictive information. The largest reported terms involved anxiety, CD4 count, adherence, WHO stage, viral load, alcohol use, stigma, age, and violence exposure.

The practical gain over linear regression was real but modest: R^2 increased by 0.024 and RMSE decreased by 0.119 PHQ-9 points. Most predictable variation was already represented by linear effects, while the cubic expansion increased the model to 816 coefficients. The near-zero mean residual indicates little average bias, but the maximum underprediction of 15.186 points reveals a clinically important tail. The model is therefore best regarded as a severity-estimation and screening-prioritisation component, not an autonomous diagnostic or treatment system.

Five recommendations follow. First, externally validate the locked pipeline across time, counties, facilities, and EMR configurations. Second, compare unpenalised degree-3 regression with ridge, elastic net, splines, ordinal models, and boosting under a prespecified validation design. Third, report calibration, prediction intervals, threshold-crossing rates, and severity-stratified errors. Fourth, conduct sensitivity analyses for GAD-7 overlap and CD4 missingness. Fifth, evaluate a clinician-supervised implementation prospectively, with patient co-design, privacy safeguards, fairness monitoring, auditability, direct safety assessment, and reliable referral pathways.

The computing contribution is a reproducible nonlinear EMR analytics pipeline that links data preparation, polynomial feature engineering, internal validation, residual analysis, and implementation boundaries. Its clinical value will depend on evidence beyond internal accuracy. Until external and prospective studies are completed, predicted scores should prompt confirmatory PHQ-9 and clinical assessment rather than replace them.

REFERENCES

1. Abebe, G. F., Alie, M. S., Adugna, A., Shifera, N., Agegnehu, W., Yosef, T., & Girma, D. (2025). Intimate partner violence among women living with HIV in East Africa: A systematic review and meta-analysis. BMC Public Health, 25(1), 2421. https://doi.org/10.1186/s12889-025-23609-z
2. Adedeji, W. A., Ma, Q., Raji, A. M., Cha, R., Rasaki, O. M., Hutson, A., Taiwo, B. O., Charurat, M. E., Yusuf, O. B., Fehintola, F. A., Gureje, O., & Morse, G. D. (2023). Prevalence of depression among people living with HIV in rural hospitals in South-Western Nigeria: Association with clinico-demographic factors. AIDS Research and Therapy, 20(1), 89. https://doi.org/10.1186/s12981-023-00586-0
3. Ahmed, I., Brahmacharimayum, A., Ali, R. H., Khan, T. A., & Ahmad, M. O. (2025). Explainable AI for depression detection and severity classification from activity data: Development and evaluation study of an interpretable

framework. JMIR Mental Health, 12, e72038. https://doi.org/10.2196/72038

4. Ali, M., Ali, S., Abbas, Q., Abbas, Z., & Lee, S. W. (2025). Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions in digital health. Digital Health, 11. https://doi.org/10.1177/20552076251395548
5. Boakye, D. S., Setordzi, M., Dzansi, G., & Adjorlolo, S. (2024). Mental health burden among females living with HIV and AIDS in sub-Saharan Africa: A systematic review. PLOS Global Public Health, 4(2), e0002767. https://doi.org/10.1371/journal.pgph.0002767
6. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Lawrence Erlbaum Associates.
7. Collins, G. S., Dhiman, P., Ma, J., Schlussel, M. M., Archer, L., Van Calster, B., Harrell, F. E., Jr., Martin, G. P., Moons, K. G. M., van Smeden, M., Sperrin, M., Bullock, G. S., & Riley, R. D. (2024). Evaluation of clinical prediction models (part 1): From development to external validation. BMJ, 384, e074819. https://doi.org/10.1136/bmj-2023-074819
8. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378
9. Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. Chapman and Hall.
10. Dessie, Z. G., & Zewotir, T. (2024). HIV-related stigma and associated factors: A systematic review and meta-analysis. Frontiers in Public Health, 12, 1356430. https://doi.org/10.3389/fpubh.2024.1356430
11. Duko, B., Belayhun, Y., & Bedaso, A. (2024). Prevalence of common mental disorder and its association with perceived stigma and social support among people living with HIV/AIDS in Ethiopia: A systematic review and meta-analysis. International Journal of Mental Health Systems, 18, Article 18. https://doi.org/10.1186/s13033-024-00641-x
12. Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. Annual Review of Clinical Psychology, 14, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037
13. Efthimiou, O., Seo, M., Chalkou, K., Debray, T. P. A., Egger, M., & Salanti, G. (2024). Developing clinical prediction models: A step-by-step guide. BMJ, 386, e078276. https://doi.org/10.1136/bmj-2023-078276
14. Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. Science, 196(4286), 129–136. https://doi.org/10.1126/science.847460
15. Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., & Matic, A. (2022). Machine learning model to predict mental health crises from electronic health records. Nature Medicine, 28(6), 1240–1248. https://doi.org/10.1038/s41591-022-01811-5
16. Gebru, T., Ejara, D., Yalew, A., & Deyessa, N. (2024). Prevalence of depression and associated factors among HIV/AIDS patients attending antiretroviral therapy clinic at Adama Hospital Medical College, Adama, Central Ethiopia. Scientific Reports, 14, 2119. https://doi.org/10.1038/s41598-024-52142-z
17. Harrell, F. E., Jr. (2015). Regression modeling strategies (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7
18. Huang, B., Younger, A., Gallant, M. P., & O'Grady, T. J. (2025). Depressive symptoms and HIV viral suppression: A systematic review and meta-analysis. AIDS and Behavior, 29(3), 870–883. https://doi.org/10.1007/s10461-024-04571-0
19. Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. Psychiatric Annals, 32(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06
20. Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
21. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., Weicken, E., Asselbergs, F. W., Prior, F., Collins, G. S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J. A., Kushibar, K., Riklund, K., Marias, K., … Starmans, M. P. A. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ, 388, e081554. https://doi.org/10.1136/bmj-2024-081554

22. Marathe, G., Moodie, E. E. M., Brouillette, M.-J., Cox, J., Cooper, C., Delaunay, C. L., Conway, B., Hull, M., Martel-Laferrière, V., Vachon, M.-L., Walmsley, S., Wong, A., Klein, M. B., & Canadian Co-Infection Cohort Study Investigators. (2022). Predicting the presence of depressive symptoms in the HIV-HCV co-infected population in Canada using supervised machine learning. BMC Medical Research Methodology, 22, 223. https://doi.org/10.1186/s12874-022-01700-y
23. Molapo, D. M., Mokgalaboni, K., & Phoswa, W. N. (2025). Prevalence of depression among people living with HIV on antiretroviral therapy in Africa: A systematic review and meta-analysis. Healthcare, 13(1), 85. https://doi.org/10.3390/healthcare13010085
24. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., … van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505. https://doi.org/10.1136/bmj-2024-082505
25. Ndisha, M., Hassan, A. S., Ngari, F., Munene, E., Gikura, M., Kimutai, K., Muthoka, K., Murie, L. A., Tolentino, H., Odhiambo, J., Mwele, P., Odero, L., Mbaire, K., Omoro, G., & Kimanga, D. O. (2023). Leveraging electronic medical records for HIV testing, care, and treatment programming in Kenya: The national data warehouse project. BMC Medical Informatics and Decision Making, 23, 162. https://doi.org/10.1186/s12911-023-02265-6
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html
27. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872. https://doi.org/10.7326/M18-1990
28. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28, 31–38. https://doi.org/10.1038/s41591-021-01614-0
29. Remien, R. H., Stirratt, M. J., Nguyen, N., Robbins, R. N., Pala, A. N., & Mellins, C. A. (2019). Mental health and HIV/AIDS: The need for an integrated response. AIDS, 33(9), 1411–1420. https://doi.org/10.1097/QAD.0000000000002227
30. Salahudeen, T., Maalouf, M., Elfadel, I. M., & Jelinek, H. F. (2025). Predicting depression severity using machine learning models: Insights from mitochondrial peptides and clinical factors. PLOS ONE, 20(5), e0320955. https://doi.org/10.1371/journal.pone.0320955
31. Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. Psychological Medicine, 49(9), 1426–1448. https://doi.org/10.1017/S0033291719000151
32. Ssentumbwe, V., Matovu, F., Namatovu, P., Najjuuko, C., Kizito, S., Nabayinda, J., Nartey, P., Namuwonge, F., Nabunya, P., Lu, C., Mutumba, M., & Ssewamala, F. M. (2026). Using machine learning to predict depression among adolescents living with HIV in Uganda. Global Social Welfare. https://doi.org/10.1007/s40609-026-00456-3
33. Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ, 338, b2393. https://doi.org/10.1136/bmj.b2393
34. Steyerberg, E. W. (2019). Clinical prediction models: A practical approach to development, validation, and updating (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-16399-0
35. Swinckels, L., Bennis, F. C., Ziesemer, K. A., Scheerman, J. F. M., Bijwaard, H., de Keijzer, A., & Bruers, J. J. (2024). The use of deep learning and machine learning on longitudinal electronic health records for the early detection and prevention of diseases: Scoping review. Journal of Medical Internet Research, 26, e48320. https://doi.org/10.2196/48320
36. Tadesse, G., Rtbey, G., Tinsae, T., Andualem, F., Kelebie, M., Kibralew, G., Geremew, G. W., Abate, A. T., Wassie, Y. A., Alemayehu, T. T., Nakie, G., Fentahun, S., & Takelle, G. M. (2025). Depressive symptoms and their determinants among people living with HIV in Africa: Systematic review and meta-analysis. BMC Psychiatry, 25, 325. https://doi.org/10.1186/s12888-025-06766-8

37. Tibshirani, R., Wainwright, M., & Hastie, T. (2017). Statistical learning with sparsity: The lasso and generalizations. CRC Press.
38. Van Calster, B., Steyerberg, E. W., Wynants, L., & van Smeden, M. (2023). There is no such thing as a validated prediction model. BMC Medicine, 21, 70. https://doi.org/10.1186/s12916-023-02779-w
39. Vu, T., Dawadi, R., Yamamoto, M., Tay, J. T., Watanabe, N., Kuriya, Y., Oya, A., Tran, P. N. H., & Araki, M. (2025). Prediction of depressive disorder using machine learning approaches: Findings from the NHANES. BMC Medical Informatics and Decision Making, 25, 83. https://doi.org/10.1186/s12911-025-02903-1
40. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6
41. World Health Organization & Joint United Nations Programme on HIV/AIDS. (2022). Integration of mental health and HIV interventions: Key considerations. https://iris.who.int/handle/10665/353571
42. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200
43. Xiang, Y., Du, J., Fujimoto, K., Li, F., Schneider, J., & Tao, C. (2022). Application of artificial intelligence and machine learning for HIV prevention interventions. The Lancet HIV, 9(1), e54–e62. https://doi.org/10.1016/S2352-3018(21)00247-2
44. Yang, K., Tu, J., & Chen, T. (2019). Homoscedasticity: An overlooked critical assumption for linear regression. General Psychiatry, 32(5), e100148. https://doi.org/10.1136/gpsych-2019-100148
45. Ye, Y., Zhang, Z., Tao, Z., Cui, L., Wang, Y., Chen, H., Li, S., Chen, X., & Zhou, J. (2025). Academic pressure and psychological imbalance in high school students: Predictors of depression via polynomial regression and response surface analysis. Psychology Research and Behavior Management, 18, 15–23. https://doi.org/10.2147/PRBM.S498606
46. Yotebieng, M., Zotova, N., Bernard, C., Goodrich, S., Awoh, A. R., Watnick, D., Nsonde, D. M., Moungang, E. F. T., Noumedem, J. L. N., Mbongo'o, G. C. N., Minga, A., Seydi, M., Gandou, P., Kwobah, E. K., Atwoli, L., Jaquet, A., Wools-Kaloustian, K., & Anastos, K. (2024). Accuracy of the nine-item Patient Health Questionnaire against psychiatric diagnosis for depression among people with HIV. AIDS, 38(12), 1765–1773. https://doi.org/10.1097/QAD.0000000000003963
47. Zaongo, S. D., Wu, W., & Chen, Y. (2025). Pathogenesis of HIV-associated depression: Contributing factors and underlying mechanisms. Frontiers in Psychiatry, 16, 1557816. https://doi.org/10.3389/fpsyt.2025.1557816
48. Zeleke, T. A., Alemu, K., Ayele, T. A., Denu, Z. A., Mwanri, L., & Azale, T. (2024). Systematic review and meta-analysis on the effect of depression on antiretroviral therapy adherence among women living with HIV. PLOS ONE, 19(6), e0300106. https://doi.org/10.1371/journal.pone.0300106
49. Zeleke, T. A., Ayele, T. A., Denu, Z. A., Mwanri, L., & Azale, T. (2025). The effect of depression on antiretroviral drug non-adherence among women living with HIV in Gondar health facilities, northwest Ethiopia: A prospective cohort study. Frontiers in Psychiatry, 16, 1488183. https://doi.org/10.3389/fpsyt.2025.1488183