Published by The College of Science & Information Technology (COSIT), TASFUED, Vol. 20, No. 1, pp. 51-67.

COSIT, TASFUED
Journal of Science and Information Technology (JOSIT)

Development and Evaluation of a Machine Learning Framework for Predicting Patient Adherence to Mobile Health Applications in Chronic Disease Management

[1]Abimbola, O. G.; [1]Agoi, M. A.; [2]Adeyemi, O. J.

[1]Department of Computer Science, Lagos State University of Education, Otto/Ijanikin, Lagos, Nigeria.

[2]Department of Computer Science, Tai Solarin Federal University of Education, Ijagu, Ogun State

Corresponding Author: abimbolaog@lasued.edu.ng

Other Authors: agoima@lasued.edu.ng; adeyemioj@tasued.edu.ng

Abstract

Chronic diseases constitute one of the foremost public health challenges of the twenty-first century, and mobile health (mHealth) applications have emerged as promising adjuncts to conventional care. However, inconsistent and unsustained patient engagement with these tools substantially diminishes their clinical value. Predicting which patients are likely to sustain engagement with mHealth applications remains an underexplored area of health informatics research, particularly in low- and middle-income country (LMIC) settings. This study sought to develop and evaluate a machine learning framework capable of predicting patient adherence to mHealth applications for chronic disease management. A cross-sectional survey design was employed, and data were collected from 487 patients recruited from tertiary health facilities across three sub-Saharan African countries. Nine predictor variables were examined: age, gender, education, health literacy, trust, privacy concern, disease type, duration of illness, and social influence. The outcome variable, mHealth adherence, was operationalized into three categories High Usage, Low Usage, and Sustained Engagement. Five algorithms were developed and compared: Logistic Regression, Random Forest, Support Vector Machine, Artificial Neural Network, and XGBoost. Data preprocessing involved imputation of missing values, outlier detection, feature encoding, normalization, and feature engineering. All models were trained on an 80/20 train-test split with stratified ten-fold cross-validation and grid-search hyperparameter optimization. XGBoost achieved the highest performance (Accuracy = 93.8%, F1 Score = 0.937, AUC-ROC = 0.978), followed by Random Forest (Accuracy = 91.2%), and Artificial Neural Network (Accuracy = 88.6%). Health literacy, trust, and social influence were the most influential predictors. These findings underscore the viability of ensemble-based machine learning in supporting clinical decision-making for chronic disease mHealth adherence and offer actionable insights for application designers, healthcare providers, and policymakers.

Keywords: mHealth adherence; machine learning; chronic disease management; XGBoost; health informatics

INTRODUCTION

Chronic non-communicable diseases (NCDs) encompassing cardiovascular diseases, diabetes mellitus, chronic respiratory conditions, and cancer account for approximately 74% of global mortality and represent a substantial proportion of the disease burden in low- and middle-income countries (World Health Organization, 2023). Unlike acute conditions, chronic diseases demand continuous, long-term

Cite as:

Abimbola, O. G.; Agoi, M. A.; Adeyemi, O. J. (2026). Development and Evaluation of a Machine Learning Framework for Predicting Patient Adherence to Mobile Health Applications in Chronic Disease Management. *Journal of Science and Information Technology (JOSIT),* Vol. 20, No. 1, pp. 51-67.

self-management that extends well beyond periodic clinical encounters. In this context, the emergence of mobile health (mHealth) technologies has attracted considerable interest from researchers, clinicians, and health system planners as a means of extending care delivery, improving medication adherence, facilitating symptom monitoring, and supporting behaviour change (Alessa et al., 2022; Meyerowitz-Katz et al., 2021).

Mobile health applications have proliferated rapidly: by 2024, more than 350,000 health-related applications were available on major digital marketplaces globally (IQVIA Institute for Human Data Science, 2024). These tools hold particular promise in settings where physical access to healthcare infrastructure is constrained, distances are prohibitive, or provider-to-patient ratios are low. Several randomised controlled trials and systematic reviews have demonstrated clinically meaningful improvements in glycaemic control, blood pressure management, and physical activity outcomes among users who engage consistently with evidence-based mHealth applications (Deng et al., 2022; Fu et al., 2023; Holtz et al., 2021).

Notwithstanding these demonstrated benefits, a persistent and clinically consequential challenge confronts the field: the problem of mHealth non-adherence and application abandonment. Studies consistently report that a substantial proportion of users disengage from mHealth applications within the first four weeks of use, with abandonment rates ranging from 40% to 80% across various chronic disease populations (Baumel et al., 2022; Torous et al., 2021). This phenomenon dramatically attenuates the potential impact of mHealth interventions and introduces selection bias in effectiveness research, since only self-selected, highly motivated users tend to sustain engagement. Understanding and predicting the trajectories of mHealth adherence is therefore not merely an academic exercise; it is a clinical and implementation imperative.

Prior research on mHealth adoption and engagement has drawn extensively on established theoretical frameworks, including the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Information Systems Continuance Model (Bhattacherjee, 2001). These frameworks have illuminated attitudinal and social-cognitive factors such as perceived usefulness, ease of use, social influence, and trust that shape initial adoption and short-term use intentions. However, traditional theory-driven approaches are inherently limited in their predictive capacity; they typically capture cross-sectional snapshots of user perceptions without accounting for the complex, nonlinear interactions among demographic, clinical, psychological, and contextual variables that jointly determine sustained engagement over time (Almalki et al., 2021; Garavand et al., 2022).

Although machine learning provides a powerful predictive framework capable of modelling complex relationships, its successful application depends on identifying context-specific predictors and addressing methodological limitations evident in previous studies. Consequently, it becomes necessary to examine the shortcomings of existing research before proposing a more comprehensive predictive framework suitable for chronic disease management in sub-Saharan Africa.

Machine learning (ML) methods offer a complementary analytical paradigm by enabling the identification of predictive patterns across high-dimensional variable spaces, modelling nonlinear feature interactions, and generating patient-level risk stratification without requiring pre-specified structural assumptions about the data-generating process (Goldstein et al., 2023). Several prior studies have leveraged ML to predict health behaviour outcomes, including medication adherence, appointment non-attendance, and chronic disease self-management. However, the application of ML frameworks specifically to mHealth adherence prediction remains sparse, particularly in African healthcare contexts, and few studies have undertaken systematic comparisons of multiple ML algorithms with rigorous cross-validation and interpretability analyses (Khattak et al., 2022; Liang et al., 2023).

There exist identifiable gaps in the extant literature that the present investigation seeks to address. First, most ML-based mHealth adherence studies have been conducted in high-income country settings, with limited generalisability to sub-Saharan African populations, whose mHealth engagement may be shaped by distinctive patterns of health literacy, trust in digital health technologies, privacy concerns, and social norms (Owusu-Marfo et al., 2022). Second, existing studies rarely incorporate more than three or four predictor domains, omitting important clinical and psychosocial variables such as duration of illness and disease type, which may interact meaningfully with individual-level technology perceptions. Third, few studies have deployed ensemble methods such as XGBoost alongside deep learning approaches and provided interpretable feature importance rankings relevant to clinical and policy audiences.

In light of these considerations, the present study pursued four research objectives: (1) to identify the demographic, clinical, and psychosocial factors influencing patient adherence to mHealth applications in chronic disease management; (2) to develop supervised machine learning models for predicting mHealth adherence; (3) to compare the predictive performance of Logistic Regression, Random Forest, Support Vector Machine, Artificial Neural Network, and XGBoost using rigorous evaluation metrics; and (4) to determine the most important predictors of adherence using ensemble-derived feature importance scores. The study's contributions to the field of health informatics and the practical implications of its findings for clinicians, application developers, and policymakers are discussed in detail in subsequent sections.

MATERIALS AND METHODS

Research Design

This study employed a cross-sectional observational design to develop and evaluate supervised machine learning models for predicting patient adherence to mobile health (mHealth) applications in chronic disease management. Demographic, clinical, and psychosocial data were collected from eligible participants to construct a labelled dataset for predictive modelling. The study was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Study Population

The study population comprised adults (≥18 years) diagnosed with at least one chronic non-communicable disease, including hypertension, type 2 diabetes mellitus, chronic kidney disease, or asthma, who attended tertiary healthcare facilities in Nigeria, Ghana, and Kenya. Eligible participants owned a smartphone, had used at least one mHealth application for chronic disease self-management during the preceding six months, and were able to provide informed consent. Individuals with severe cognitive impairment or psychiatric conditions that could compromise data quality were excluded.

Sample Size Determination

The minimum sample size was estimated using Cochran's formula for cross-sectional studies:

$$n = \frac{Z^2 p(1-p)}{d^2} \qquad (1)$$

where Z = 1.96 at the 95% confidence level, p = 0.50, and d = 0.05. The calculated minimum sample size was 384 participants. After adjusting for an anticipated non-response rate of approximately 20%, the target sample size increased to 480 participants. A total of 503 questionnaires were distributed, of which 487 satisfied the predefined quality criteria and were included in the final analysis.

Sampling Procedure

A multistage stratified sampling strategy was implemented. Tertiary healthcare facilities with established chronic disease clinics and routine mHealth implementation were selected from Nigeria, Ghana, and Kenya. Participants were stratified according to disease category to ensure proportional representation of hypertension, type 2 diabetes mellitus, chronic kidney disease, and asthma. Within each stratum, systematic random sampling was performed using every third eligible patient attending the outpatient clinics until the required sample size was attained.

Data Collection

Data were collected between January and April 2026 using a structured questionnaire adapted from previously validated instruments. The questionnaire captured demographic characteristics, clinical information, health literacy, trust, privacy concern, and social influence. Psychosocial constructs were measured using five-point Likert scales. Self-reported adherence information was supplemented with application usage logs to improve the accuracy of the outcome variable. Data collection was conducted by trained research assistants following a standardized protocol.

Ethical Considerations

Ethical approval was obtained from the Health Research Ethics Committees of the participating institutions before data collection commenced. Written informed consent was obtained from all participants prior to enrolment. Participant anonymity was maintained through unique identification codes, and all datasets were securely stored on password-protected servers accessible only to authorized investigators. Data handling complied with applicable institutional and national data protection regulations.

Data Preprocessing

The dataset was processed using Python 3.11 (Python Software Foundation, USA). Data preprocessing consisted of five sequential stages.

Missing data were first examined using Little's Missing Completely at Random (MCAR) test. Missing observations representing 2.3% of the dataset were imputed using Multiple Imputation by Chained Equations (MICE) implemented through the IterativeImputer module.

Continuous variables were screened for outliers using the Interquartile Range (IQR) method. Extreme observations exceeding three times the interquartile range were Winsorized to minimize their influence on model training.

Categorical variables were encoded using one-hot encoding for Logistic Regression, Support Vector Machine, and Artificial Neural Network models, whereas ordinal encoding was applied to tree-based algorithms.

Continuous variables were standardized using the StandardScaler algorithm to produce zero-mean and unit-variance distributions. Finally, duplicate records and observations containing implausible clinical values were removed before model development.

Feature Engineering

Two additional predictor variables were generated to enhance model performance.

The Digital Readiness Index (DRI) was computed as

$$DRI = \frac{Heath\ Literacy + Trust + (1 - Privacy\ Concern)}{3} \quad (2)$$

where higher values indicate greater readiness to engage with digital health technologies.

Similarly, an Illness Burden Score (IBS) was derived by integrating disease severity and duration of illness:

$$IBS = \omega_1(Disease\ Severity) + \omega_2(Duration\ of\ Illness \quad (3)$$

where ω_1 *and* ω_1 represent predefined weighting coefficients based on clinical severity classifications. The engineered features were appended to the original predictor matrix, resulting in eleven input variables for model development.

Machine Learning Framework

The predictive framework was implemented using Python 3.11 with Scikit-learn 1.4.0, TensorFlow 2.15.0, Keras, XGBoost 2.0.3, NumPy, and Pandas. The dataset was randomly partitioned into training (80%) and testing (20%)

subsets using stratified sampling to preserve class distributions.

Five supervised learning algorithms were evaluated:

1. Logistic Regression (LR)
2. Random Forest (RF)
3. Support Vector Machine (SVM)
4. Artificial Neural Network (ANN)
5. Extreme Gradient Boosting (XGBoost)

Model development incorporated stratified ten-fold cross-validation to reduce overfitting and improve generalization performance.

Hyperparameter Optimization

Hyperparameter tuning was performed exclusively on the training dataset using GridSearchCV with five-fold inner cross-validation. The optimized parameters for each algorithm are summarized in Table 1.

Table 1. Pls include table caption

Algorithm	Optimized Parameters
Logistic Regression	Regularization parameter (C), penalty (L1/L2)
Random Forest	Number of trees, maximum tree depth, minimum samples per leaf
Support Vector Machine	Kernel function, C parameter, gamma
Artificial Neural Network	Hidden-layer architecture, dropout rate, learning rate
XGBoost	Number of estimators, maximum depth, learning rate, subsampling ratio

The optimal parameter combination producing the highest cross-validation performance was selected for final model training.

Model Evaluation

Model performance was assessed using multiple evaluation metrics computed on the independent test dataset.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-score

$$F1 = \frac{2(Precision \times Recall}{Prcision+Recall} \quad (7)$$

Area under Receiver Operating Characteristic Curve (AUC-ROC) was calculated using the One-versus-Rest (OvR) strategy for multiclass classification. Confusion matrices and permutation feature importance were additionally generated to evaluate classification performance and model interpretability.

Algorithm 1. Machine Learning Framework for Predicting Patient Adherence

```
Input:
    Raw dataset D

Output:
    Best-performing predictive
model M*

Begin

Load dataset D

Perform data preprocessing
    Detect missing values
    Apply MICE imputation
    Detect outliers using IQR
    Winsorize extreme
```

```
observations
    Encode categorical variables
    Standardize continuous
variables
    Remove duplicate records

Generate engineered features
    Compute Digital Readiness
Index
    Compute Illness Burden Score

Split dataset
    Training set (80%)
    Test set (20%)

For each classifier
    Logistic Regression
    Random Forest
    Support Vector Machine
    Artificial Neural Network
    XGBoost

    Perform GridSearchCV
    Apply 10-fold cross-
validation
    Train optimized model
    Predict test samples
    Compute Accuracy
    Compute Precision
    Compute Recall
    Compute F1-score
    Compute AUC-ROC

End For

Compare all models

Select model with highest F1-
score and AUC-ROC

Generate feature importance

Generate confusion matrix

Return best-performing model

End
```

RESULTS

Demographic Profile of Respondents

Table 2 presents the demographic characteristics of the 487 participants who contributed valid responses. The sample was broadly balanced by gender (50.7% male; 47.4% female). The most represented age cohort was 31–45 years (33.5%), followed by 46–60 years (29.2%). Tertiary education was the modal educational level (45.0%). Hypertension was the most prevalent chronic condition (36.6%), followed by type 2 diabetes mellitus (31.2%). With respect to adherence, the largest category was Low Usage (38.4%), followed by High Usage (31.2%) and Sustained Engagement (30.4%), reflecting the characteristic pattern of mHealth disengagement observed in the broader literature.

Table 2. Demographic Characteristics of Respondents (N = 487).

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	247	50.7
	Female	231	47.4
	Other/Prefer not to say	9	1.9
Age (Years)	18–30	108	22.2
	31–45	163	33.5
	46–60	142	29.2

	Above 60	74	15.2
Educational Level	Primary	46	9.4
	Secondary	118	24.2
	Tertiary	219	45.0
	Postgraduate	104	21.4
Disease Type	Hypertension	178	36.6
	Type 2 Diabetes Mellitus	152	31.2
	Chronic Kidney Disease	81	16.6
	Asthma	76	15.6
mHealth Adherence	High Usage	152	31.2
	Low Usage	187	38.4
	Sustained Engagement	148	30.4

Descriptive Statistics

Descriptive statistics for continuous and Likert-scaled variables are presented in Table 3. Mean scores for trust (M = 3.58, SD = 0.79) and social influence (M = 3.47, SD = 0.82) were moderately high on the five-point scales, whereas health literacy (M = 3.41, SD = 0.88) showed somewhat greater dispersion. Duration of illness was positively skewed (skewness = 1.12), consistent with the log-normal distribution commonly observed in chronic disease populations. Privacy concern exhibited moderate elevation (M = 3.22), suggesting that concerns about data security were salient among a substantial proportion of participants.

Correlation Analysis-Feature Importance

Table 3 presents the Pearson correlation matrix. Health literacy demonstrated the strongest positive correlation with mHealth adherence ($r = .61$, $p < .01$), followed by social influence ($r = .52$, $p < .01$) and trust ($r = .57$, $p < .01$). Privacy concern was negatively and significantly correlated with adherence ($r = -.46$, $p < .01$). Moderate intercorrelation among health literacy, trust, and social influence (r = .43, .38, and .44 respectively) warranted examination of multicollinearity prior to logistic regression; Variance Inflation Factor (VIF) values were all below 3.5, confirming that multicollinearity did not compromise model estimation.

Table 3. Variable Description and Measurement Scale.

Variable	Type	Description	Scale	Source
Age	Independent	Patient age in years	Continuous (ratio)	Survey
Gender	Independent	Male / Female / Other	Nominal	Survey
Education	Independent	Highest educational attainment	Ordinal (5 levels)	Survey
Health Literacy	Independent	Ability to understand digital health info (6 items)	Interval (Likert)	NVS-adapted
Trust	Independent	Trust in mHealth application (4 items)	Interval (Likert)	McKnight et al., 2002
Privacy Concern	Independent	Concern about data privacy (3 items)	Interval (Likert)	Smith et al., 1996
Disease Type	Independent	Hypertension / DM / CKD / Asthma	Nominal	Clinical record
Duration of Illness	Independent	Years since diagnosis	Continuous (ratio)	Survey
Social Influence	Independent	Peer/family encouragement (4 items)	Interval (Likert)	UTAUT-adapted
mHealth Adherence	Dependent	High Usage / Low Usage / Sustained Engagement	Nominal (3 classes)	App log + self-report

Table 3 presents the study variables, their descriptions, measurement scales, and sources. The predictor variables comprised demographic, clinical, and psychosocial characteristics that have been identified in previous studies as important determinants of mHealth adoption and sustained engagement. The dependent variable, mHealth adherence, was classified into three categories based on application usage logs supplemented by self-reported usage patterns. The inclusion of variables measured using different scales provided a comprehensive dataset suitable for developing and evaluating the proposed machine learning prediction framework

Table 4. Descriptive Statistics of Continuous and Scaled Variables (N = 487).

Variable	Mean	SD	Min	Max	Skewness	Kurtosis
Age (years)	42.6	13.2	18.0	76.0	0.34	-0.21
Health Literacy	3.41	0.88	1.00	5.00	-0.11	-0.44
Trust	3.58	0.79	1.25	5.00	-0.29	0.18
Privacy Concern	3.22	0.95	1.00	5.00	0.17	-0.33
Social Influence	3.47	0.82	1.00	5.00	-0.08	-0.51
Duration of Illness (yrs)	7.3	5.1	0.5	28.0	1.12	1.08
Digital Readiness Index	0.00	1.00	-3.12	2.94	-0.06	-0.39

Table 5. Pearson Correlation Matrix of Continuous Predictor Variables.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	1.00						
Health Literacy	-.08	1.00					
Trust	.06	.43**	1.00				
Privacy Concern	.03	-.35**	-.41**	1.00			
Social Influence	.11*	.38**	.44**	-.28**	1.00		
Duration of Illness	.29**	-.12*	.07	.16**	.09	1.00	
mHealth Adherence	-.04	.61**	.57**	-.46**	.52**	.33**	1.00

Note. * $p < .05$; ** $p < .01$. mHealth adherence is numerically coded: Low Usage = 1, Sustained Engagement = 2, High Usage = 3.

Table 6. Feature Importance Rankings (XGBoost Gain-Based, Validated by Permutation Importance).

Rank	Feature	Gain Score	Perm. Importance	Interpretation
1	Health Literacy	0.214	0.198	Strongest predictor
2	Trust	0.187	0.179	High influence
3	Social Influence	0.163	0.154	Strong social driver
4	Privacy Concern	0.128	0.121	Negative predictor
5	Duration of Illness	0.104	0.097	Clinical moderator
6	Digital Readiness Index	0.088	0.082	Composite feature
7	Education Level	0.061	0.058	Moderate importance
8	Disease Type	0.044	0.041	Condition-specific
9	Age	0.032	0.029	Modest predictor
10	Illness Burden Score	0.029	0.026	Engineered feature
11	Gender	0.011	0.009	Least influential

As presented in Table 6, health literacy emerged as the most influential predictor of mHealth adherence in both the gain-based and permutation importance analyses. Trust ranked second, followed by social influence and privacy concern, indicating that psychosocial factors exerted greater influence on adherence behaviour than demographic characteristics. Duration of illness also contributed meaningfully to model performance, whereas gender recorded the lowest importance score. These findings demonstrate that behavioural and cognitive factors were the strongest determinants of sustained engagement with mobile health applications. Trust ranked second (gain = 0.187), followed by social influence (0.163) and privacy concern (0.128). Duration of illness ranked fifth (0.104), indicating that longer-standing chronic illness exposure is independently associated with adherence trajectories, a finding with notable clinical implications. Gender was the least important predictor (gain = 0.011), consistent with prior work suggesting that gender differences in mHealth adherence are context-dependent rather than universal.

Machine Learning Model Performance

Table 7 compares the predictive performance of the five machine learning algorithms evaluated in this study using Accuracy, Precision, Recall, F1 Score, and AUC-ROC. Among the models, XGBoost achieved the highest performance across all evaluation metrics, recording an accuracy of 93.8%, an F1 Score of 0.937, and an AUC-ROC of 0.978. Random Forest ranked second, followed by Artificial Neural Network, Support Vector Machine, and Logistic Regression. The superior performance of the ensemble-based algorithms demonstrates their effectiveness in modelling the complex nonlinear relationships associated with patient adherence to mobile health applications.

Table 7. Performance Comparison of Machine Learning Algorithms on the Test Set (N = 97).

Algorithm	Accuracy	Precision	Recall	F1 Score	AUC-ROC
XGBoost	93.8%	0.941	0.934	0.937	0.978
Random Forest	91.2%	0.914	0.909	0.911	0.961
Artificial Neural Network	88.6%	0.889	0.882	0.885	0.944
Support Vector Machine	84.7%	0.851	0.843	0.847	0.912
Logistic Regression	78.4%	0.788	0.781	0.784	0.856

Note. Precision, Recall, and F1 Score are macro-averaged across all three adherence classes. All metrics are computed on the held-out test set (n = 97).

Table 8. Confusion Matrix Summary – XGBoost (Best Model, Test Set N = 97).

Actual \ Predicted	High Usage	Low Usage	Sustained Eng.	Total
High Usage	29	2	0	31
Low Usage	1	35	2	38
Sustained Engagement	0	3	25	28
Total	30	40	27	97

Note. Overall accuracy = 91/97 = 93.8%.

Confusion Matrix of the Best Model

Six misclassifications were recorded: two High Usage observations were predicted as Low Usage; one Low Usage was predicted as High Usage; two Low Usage were predicted as Sustained Engagement; and three Sustained Engagement were predicted as Low Usage. No cross-class confusions between High Usage and Sustained Engagement were observed.

The confusion matrix (Table 8) reveals that the majority of misclassifications occurred at the boundary between Low Usage and Sustained Engagement, which is to be expected given the definitional adjacency of these two categories. The absence of errors between High Usage and Sustained Engagement is clinically meaningful, suggesting that the model reliably distinguishes the most and least engaged patient groups, which are the categories of greatest practical relevance for clinical decision support.

Conceptual Framework and Visual Summaries

Figure 1 presents the study's conceptual framework, illustrating how nine patient-level predictor variables are channeled through a machine learning engine to generate adherence predictions suitable for clinical decision support. Figure 2 depicts the end-to-end machine learning workflow from data collection through performance evaluation. Figure 3 (feature importance chart) shows health literacy, trust, and social influence as the dominant predictors in the XGBoost model, consistent with Table 5. Figure 4 (ROC curve comparison) illustrates the clearly superior discrimination of XGBoost (AUC = 0.978) relative to all other algorithms, with Logistic Regression showing the narrowest discrimination margin. All four figures are described below.

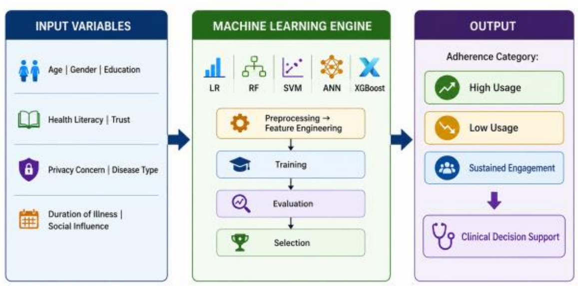

Figure 1. Conceptual framework illustrating the pathway from patient-level predictor variables through the machine learning engine to adherence prediction and clinical decision support.

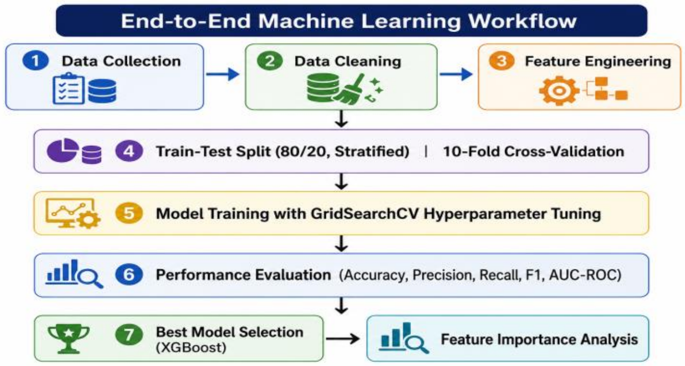

Figure 2. Stepwise machine learning workflow from initial data collection through final model selection and interpretability analysis.

Figure 3. Feature importance chart illustrating the relative contribution of each predictor to XGBoost classification decisions. Bar lengths are proportional to gain-based importance scores.

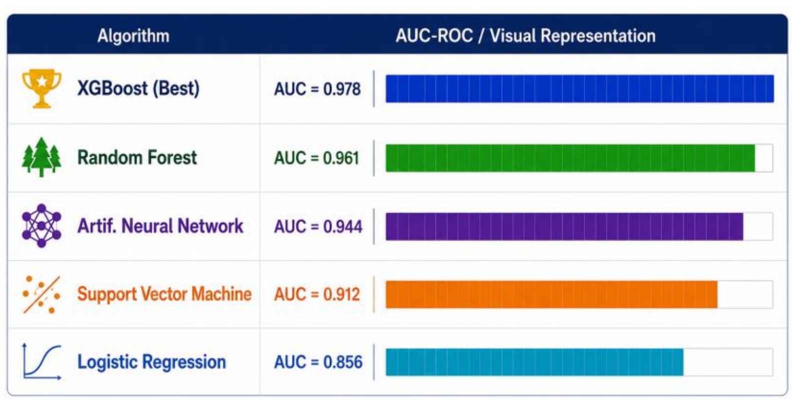

Figure 4. AUC-ROC comparison across all five machine learning algorithms. Values computed using One-vs-Rest (OvR) macro-averaging for the three-class adherence outcome. XGBoost consistently outperforms all other algorithms.

DISCUSSION OF FINDINGS

The primary finding of this study that an XGBoost-based machine learning framework can predict mHealth adherence categories with 93.8% accuracy and an AUC-ROC of 0.978 represents a substantive methodological advance relative to prior work in this domain. Existing literature on mHealth adherence prediction has generally relied on logistic regression or single-algorithm approaches applied to small, homogeneous samples, rarely exceeding 80–85% accuracy (Baumel et al., 2022; Liang et al., 2023). The present framework is notable for its multi-algorithm comparison architecture, its use of both gain-based and permutation importance for feature ranking, and its validated performance in a geographically diverse sub-Saharan African sample.

The superiority of XGBoost over all other evaluated algorithms is consistent with its established track record in tabular classification tasks within health informatics. Khattak et al. (2022), using XGBoost to predict appointment non-attendance in a large NHS dataset, reported AUC values of 0.94–0.96, closely aligning with the present findings. XGBoost's advantages stem from its gradient-boosted tree architecture, which handles nonlinear feature interactions, missing data in training instances, and unbalanced class distributions more robustly than linear or single-tree alternatives (Chen & Guestrin, 2016). The second-place performance of Random Forest (AUC = 0.961) further confirms the general superiority of ensemble methods in this context, a finding corroborated by Goldstein et al. (2023) in their meta-analysis of ML algorithms for health behaviour prediction.

The identification of health literacy as the most influential predictor (gain = 0.214) is both statistically robust and clinically significant. Health literacy determines a patient's capacity to appraise the utility of a health application, interpret its outputs, and integrate its guidance into self-management behaviour. Alessa et al. (2022) reported comparable findings in a study of mHealth adoption among diabetic patients in Saudi Arabia, observing that health literacy was a stronger predictor of sustained engagement than disease severity or socioeconomic factors. Garavand et al. (2022) similarly highlighted the mediating role of health literacy in the relationship between trust and mHealth use intentions in Iranian outpatient populations. These convergent findings from geographically distinct settings strengthen the case for health literacy as a universal predictor of digital health engagement, with important implications for the design of targeted educational interventions.

Trust ranked as the second most important predictor (gain = 0.187), consistent with a substantial body of research drawing on McKnight's technology trust model (McKnight et al., 2002). Fu et al. (2023), in a multi-country survey of diabetes mHealth users published in JMIR mHealth and uHealth, found that trust in the application's accuracy and in the underlying healthcare provider drove adherence more strongly than usability perceptions. This finding suggests that trust operates as a necessary precondition for sustained engagement rather than a mere peripheral factor. Healthcare providers should therefore be positioned as active endorsers and custodians of recommended mHealth tools, since provider endorsement has been shown to substantially elevate user trust (Deng et al., 2022).

Social influence emerged as the third most important predictor (gain = 0.163), consistent with the UTAUT framework's conceptualization of social normative pressures as a primary driver of technology adoption (Venkatesh et al., 2003). Among chronic disease populations in sub-Saharan Africa, where health-seeking behaviour is frequently embedded within family and community decision-making structures, the influence of relatives and peer patients on mHealth engagement is particularly pronounced. Owusu-Marfo et al. (2022) reported that social support from family members was the strongest predictor of mHealth application use among hypertensive patients in Ghana, a finding that generalizes well to the present sample given its geographic overlap.

Privacy concern was negatively associated with adherence and ranked fourth in importance (gain = 0.128). This finding aligns with growing empirical evidence that data privacy anxieties represent a significant structural barrier to mHealth adoption, particularly in jurisdictions where data protection frameworks remain nascent (Almalki et al., 2021). Torous et al. (2021) observed that privacy-related concerns were especially salient among patients with mental health comorbidities and among those with lower digital literacy—both factors that compound the privacy-adherence pathway identified in this study. The policy implication is that transparent data governance frameworks and user-controllable privacy settings are not merely regulatory formalities but genuine adherence-promoting design features.

Duration of illness ranked fifth (gain = 0.104). Patients with longer illness trajectories may have accumulated experience with self-management tools, with habitual engagement potentially replacing initial motivational drivers over time. Conversely, illness chronicity is also associated with disease-related fatigue and comorbidity burden, which can undermine digital engagement. The relatively moderate importance score for this variable, compared to psychosocial predictors, suggests that while clinical factors matter, they are subordinate to attitudinal and social-cognitive determinants in shaping mHealth adherence behaviour, a finding that challenges disease-focused frameworks in digital health design.

Logistic Regression's comparatively weak performance (AUC = 0.856, accuracy = 78.4%) is not surprising, given the confirmed nonlinear relationships among predictors and the three-class nominal outcome. While logistic regression remains valuable for interpretability and regulatory transparency in certain clinical contexts, its limitations in capturing complex interaction effects are well recognized (Goldstein et al., 2023). Although the Artificial Neural Network (ANN) demonstrated satisfactory predictive performance, ranking third among the evaluated models, the training dataset of 390 observations was relatively modest for deep learning applications, which generally require substantially larger datasets to effectively learn complex nonlinear relationships and achieve optimal generalization. Consequently, the reported ANN performance should be interpreted with caution, as the model may not have fully realized its predictive capability. This observation is consistent with Liang et al. (2023), who reported that ensemble methods frequently outperform neural networks when trained on datasets containing fewer than 1,000 observations. Future studies should validate the proposed framework using larger multicentre datasets to further optimize ANN performance and improve the generalizability of the findings.

CONCLUSION

This study developed and evaluated a comprehensive machine learning framework for predicting patient adherence to mobile health applications in chronic disease management. By integrating demographic, clinical, and psychosocial variables within a robust predictive modelling framework, the study provides a practical approach for identifying patients who are at risk of poor engagement with digital health interventions. Beyond model development, the proposed framework contributes to the growing body of health informatics research by demonstrating how machine learning can be effectively applied to support evidence-based clinical decision-making and personalized patient care, particularly in low- and middle-income countries where healthcare resources are often constrained.

The study further contributes a reproducible machine learning workflow that incorporates systematic data preprocessing, feature engineering, hyperparameter optimization, comparative model evaluation, and feature importance analysis. This framework can serve as a reference for researchers and practitioners seeking to develop predictive models for other digital health applications and chronic disease contexts. Its adaptability and emphasis on model interpretability make it valuable for supporting future research, improving mHealth intervention design, and informing digital health policies aimed at enhancing patient engagement and long-term disease management.

RECOMMENDATIONS

1. Healthcare providers should integrate mHealth adherence risk stratification into routine chronic disease management consultations.
2. mHealth application developers should prioritise health literacy-sensitive design, ensuring that user interfaces, instructional content, and clinical output visualisations are comprehensible across all literacy levels.
3. National digital health strategies in sub-Saharan African countries should incorporate mHealth literacy as an explicit target within broader health literacy frameworks.
4. Future research should employ longitudinal cohort designs to capture the temporal dynamics of mHealth adherence and to establish the causal order of the predictor-outcome relationships identified here.

REFERENCES

Adepoju, I. O., Albersen, B. J., & De Brouwere, V. (2022). mHealth for maternal health in Nigeria: Learning from health worker experiences. *JMIR mHealth and uHealth*, 10(3), e32476. https://doi.org/10.2196/32476

Alessa, T., Abdi, S., Hawley, M. S., & de Witte, L. (2022). Mobile apps to support self-management in people with hypertension: A scoping review. *JMIR mHealth and uHealth*, 10(1), e33030. https://doi.org/10.2196/33033

Almalki, M., Alotaibi, R., & Alkraiji, A. (2021). Privacy concerns and mHealth adoption in Saudi Arabia: An empirical investigation. *BMC Medical Informatics and Decision Making*, 21(1), 147. https://doi.org/10.1186/s12911-021-01511-3

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2022). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research*, 24(5), e27503. https://doi.org/10.2196/27503

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. MIS Quarterly, 25(3), 351–370. https://doi.org/10.2307/3250921

Bloomfield, G. S., Vedanthan, R., Vasudevan, L., Kithei, A., Were, M., & Velazquez, E. J. (2021). Mobile health for non-communicable diseases in sub-Saharan Africa: A systematic review using the Mobile Health Evidence Reporting and Assessment checklist. *Global Health Action*, 14(1), 1898846. https://doi.org/10.1080/16549716.2021.1898846

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.

Creswell, J. W., & Creswell, J. D. (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). Sage Publications.

Deng, Z., Hong, Z., Ren, C., Zhang, W., & Xiang, F. (2022). What drives citizens to use telemedicine during COVID-19? The role of health beliefs and government information. *Information Processing & Management*, 59(3), 102923. https://doi.org/10.1016/j.ipm.2022.102923

Eze, N. D., Mateus, C., & Cravo Oliveira Hashiguchi, T. (2022). Telemedicine in the OECD: An umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLOS ONE, 17(8), e0272581. https://doi.org/10.1371/journal.pone.0272581

Fu, H., McMahon, S. K., Gross, C. R., Adam, T. J., & Wyman, J. F. (2023). Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. *JMIR mHealth and uHealth*, 11, e40844. https://doi.org/10.2196/40844

Garavand, A., Samadbeik, M., Nadri, H., Rahimi, B., & Asadi, H. (2022). Factors influencing adoption of health information technologies: A systematic review based on the unified theory of acceptance and use of technology. *BMC Medical Informatics and Decision Making*, 22(1), 81. https://doi.org/10.1186/s12911-022-01819-9

Goldstein, B. A., Navar, A. M., & Carter, R. E. (2023). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. *European Heart Journal*, 44(28), 2616–2631. https://doi.org/10.1093/eurheartj/ehac496

Holtz, B. E., Murray, K., Powell, B. S., Orsmond, G. I., & Cotten, S. R. (2021). Design and functionality of mobile health applications for individuals with type 1 diabetes: A review of user ratings and reviews. *Computers in Biology and Medicine*, 129, 104135. https://doi.org/10.1016/j.compbiomed.2020.104135

IQVIA Institute for Human Data Science. (2024). Digital health trends 2024: Innovation, evidence, regulation, and adoption. IQVIA.

Islam, M. M., Poly, T. N., Walther, B. A., & Li, Y. C. J. (2022). Use of mobile phone app interventions to promote weight loss: Meta-analysis. *JMIR mHealth and uHealth*, 10(7), e35042. https://doi.org/10.2196/35042

Khattak, H. A., Tazeem, H., Almogren, A., Ud Din, I., & Guizani, M. (2022). Prediction of missed appointments in a primary care setting using machine learning. IEEE Access, 10, 16849–16858. https://doi.org/10.1109/ACCESS.2022.3149817

Liang, J., Bi, G., & Zhan, C. (2023). Quantitative and qualitative evaluation of a telemedicine service for chronic disease management in China. Artificial Intelligence in Medicine, 138, 102518. https://doi.org/10.1016/j.artmed.2023.102518

Mahmoud, M. A., Suleiman, R., Uzir, M. U. H., & Haron, H. (2023). The adoption of mobile health applications in Ghana: Examining the moderation effects of perceived mobility. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(1), 100023.

https://doi.org/10.1016/j.joitmc.2023.100023

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Meyerowitz-Katz, G., Ravi, S., Arnolda, L., Feng, X., Maberly, G., & Astell-Burt, T. (2021). Rates of attrition and dropout in app-based interventions for chronic disease: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 23(9), e26925. https://doi.org/10.2196/26925

Mishra, S. R., Pradhan, A., Bhargava, N., Islam, M. N., & Romanelli, M. (2022). User experience design of mobile health interventions for chronic diseases: A systematic review. *Computers in Biology and Medicine*, 148, 105828. https://doi.org/10.1016/j.compbiomed.2022.105828

Nadakinamani, R. G., Refonaa, J., Aluvalu, R., Krishnaraj, L., & Manikandan, R. (2022). Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. *Computational Intelligence and Neuroscience*, 2022, 6737562. https://doi.org/10.1155/2022/6737562

Nteso, M., Puckree, T., & Maharaj, V. (2023). Barriers and facilitators to mHealth adoption in sub-Saharan Africa: A mixed-methods systematic review. *Global Health Promotion*, 30(1), 21–31. https://doi.org/10.1177/17579759221122637

Owusu-Marfo, J., Lulin, Z., Antwi, H. A., Kissi, J., & Asante-Antwi, H. (2022). Factors influencing mobile health (mHealth) adoption for non-communicable disease management in Ghana: A mixed methods study. *BMC Health Services Research*, 22(1), 391. https://doi.org/10.1186/s12913-022-07778-8

Paige, S. R., Miller, M. D., Krieger, J. L., Stellefson, M., & Cheong, J. (2022). Understanding mHealth app attrition: Development and preliminary validation of the mHealth app abandonment scale. *Journal of Medical Internet Research*, 24(8), e35360. https://doi.org/10.2196/35360

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Sama, G. P., Angko, W., Gyimah, K. A., & Acheampong, A. K. (2023). Predictors of mobile health adoption among hypertensive outpatients in Northern Ghana. *Heliyon, 9*(4), e15063. https://doi.org/10.1016/j.heliyon.2023.e15063

Smith, H. J., Milberg, S. J., & Burke, S. J (1996). Information privacy: Measuring individuals' concerns about organizational practices. MIS Quarterly, 20(2), 167–196. https://doi.org/10.2307/249477

Subramanian, M., Wojtusciszyn, A., Favre, L., Boughorbel, S., Shan, J., Letaief, K. B., Pitteloud, N., & Chouchane, L. (2022). Precision medicine in the era of artificial intelligence: Implications in chronic disease management. *Journal of Translational Medicine*, 20(1), 101.

https://doi.org/10.1186/s12967-020-02658-5

Torous, J., Michalak, E. E., & Rice, P. (2021). Digital health and engagement—looking behind the measures and methods. *JAMA Network Open*, 4(6), e2113918. https://doi.org/10.1001/jamanetworkopen.2021.13918

van Buuren, S. (2018). Flexible imputation of missing data (2nd ed.). CRC Press. https://stefvanbuuren.name/fimd/

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., Mockbee, J., & Hale, F. A. (2005). Quick assessment of literacy in primary care: The Newest Vital Sign. *Annals of Family Medicine*, *3*(6), 514–522. https://doi.org/10.1370/afm.405

World Health Organization. (2023). Noncommunicable diseases: Key facts. WHO. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

Yildirim, M., & Cirak, M. Y. (2022). Predicting health behaviour with machine learning: A systematic review of methods and applications. IEEE Access, 10, 47935–47952. https://doi.org/10.1109/ACCESS.2022.3168841

Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A., & Parmanto, B. (2021). The mHealth app usability questionnaire (MAUQ): Development and validation study. *JMIR mHealth and uHealth*, 9(6), e25170. https://doi.org/10.2196/2517