Integrated Pharmacy Research and Practice

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Developing a Prototype Machine Learning Model to Predict Quality of Life Measures in People Living With HIV

Gabriel Mercadal-Orfila[1,2], Joaquin Serrano López de las Hazas[3], Melchor Riera-Jaume[4], Salvador Herrera-Perez[5]

[1]Pharmacy Department, Hospital Mateu Orfila, Maón, Spain; [2]Department of Biochemistry and Molecular Biology, Universitat de Les Illes Balears (UIB), Palma de Mallorca, Spain; [3]Pharmacy Department, Hospital Universitario Son Llàtzer, Palma de Mallorca, Spain; [4]Unidad de Enfermedades Infecciosas, Servicio de Medicina Interna, Hospital Universitario Son Espases, Palma de Mallorca, Spain; [5]Facultad de Ciencias de la Salud, Universidad Internacional de Valencia, Valencia, España

Correspondence: Salvador Herrera-Perez; Gabriel Mercadal-Orfila, Email salva.herrera@me.com; bielmercadal@gmail.com

Background: In the realm of Evidence-Based Medicine, introduced by Gordon Guyatt in the early 1990s, the integration of machine learning technologies marks a significant advancement towards more objective, evidence-driven healthcare. Evidence-Based Medicine principles focus on using the best available scientific evidence for clinical decision-making, enhancing healthcare quality and consistency by integrating this evidence with clinician expertise and patient values. Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs) have become essential in evaluating the broader impacts of treatments, especially for chronic conditions like HIV, reflecting patient health and well-being comprehensively.

Purpose: The study aims to leverage Machine Learning (ML) technologies to predict health outcomes from PROMs/PREMs data, focusing on people living with HIV.

Patients and Methods: Our research utilizes a ML Random Forest Regression to analyze PROMs/PREMs data collected from over 1200 people living with HIV through the NAVETA telemedicine system.

Results: The findings demonstrate the potential of ML algorithms to provide precise and consistent predictions of health outcomes, indicating high reliability and effectiveness in clinical settings. Notably, our ALGOPROMIA ML model achieved the highest predictive accuracy for questionnaires such as MOS30 VIH (Adj. $R^2 = 0.984$), ESTAR (Adj. $R^2 = 0.963$), and BERGER (Adj. $R^2 = 0.936$). Moderate performance was observed for the P3CEQ (Adj. $R^2 = 0.753$) and TSQM (Adj. $R^2 = 0.698$), reflecting variability in model accuracy across instruments. Additionally, the model demonstrated strong reliability in maintaining standardized prediction errors below 0.2 for most instruments, with probabilities of achieving this threshold being 96.43% for WHOQoL HIV Bref and 88.44% for ESTAR, while lower probabilities were observed for TSQM (44%) and WRFQ (51%).

Conclusion: The results from our machine learning algorithms are promising for predicting PROMs and PREMs in AIDS settings. This work highlights how integrating ML technologies can enhance clinical pharmaceutical decision-making and support personalized treatment strategies within a multidisciplinary integration framework. Furthermore, leveraging platforms like NAVETA for deploying these models presents a scalable approach to implementation, fostering patient-centered, value-based care.

Keywords: machine-learning in healthcare, PROMs-PREMs, chronic HIV management, evidence-based medicine, NAVETA

Introduction

Evidence-Based Medicine (EBM) is a relatively new approach to medical practice introduced by Gordon Guyatt in the early 1990s. It focuses on using the best available scientific evidence to guide clinical decision-making, rather than relying solely on intuition, unsystematic clinical experience, or pathophysiologic rationale.[1–3] EBM involves critically appraising the quality and applicability of research evidence to assess the confidence that can be placed in the findings, considering factors such as study design, risk of bias, and result precision. This evidence is then integrated with clinician

Received: 21 August 2024
Accepted: 14 January 2025
Published: 22 January 2025

expertise and patient values and preferences to make informed care decisions.[4,5] In today's digital era, the transformation of the healthcare sector has become a global priority, driven by technological advances that redefine how medical care is delivered and perceived. These changes aim to enhance both the accessibility and efficiency of healthcare services and to ensure personalized care based on the best available scientific evidence.[6–8] In this contemporary healthcare context Value-based healthcare (VBH) has emerged as the paradigm seeking to redefine how the best medical care is delivered and evaluated, emphasizing the need to focus on outcomes relative to costs.[9] EBM and VBH are complementary approaches that work together to improve the quality and cost-effectiveness of healthcare. While EBM provides the scientific evidence base, VBH focuses on enhancing care that maximizes value for patients. Thus, both approaches prioritize patient-centered care and shared decision-making between providers and patients.[10,11] Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs), along with their electronic versions (ePROMs and ePREMs), have become indispensable tools for capturing patients' perspectives on their health and the impact of treatment. These measures assess dimensions such as physical state, mental health, social functioning, and general health perception,[12–14] providing essential metrics for evaluating health-related quality of life (HRQoL) and informing VBH strategies.[15–20] Hospital pharmacies, given their significant role in healthcare budgets and the high costs associated with pharmaceuticals, have rapidly integrated VBH principles alongside ePROMs and ePREMs to ensure cost-effective, patient-centered care. ePROMs quantify changes in physical symptoms, mental health, functional status, and HRQoL, thereby evaluating the effectiveness of pharmaceutical interventions. Meanwhile, ePREMs capture patients' experiences with pharmacy services, such as timeliness, communication, and clarity, helping identify opportunities for improvement and enhancing the quality of care.[21–24]

ePROMs and ePREMs are particularly important for people living with chronic conditions like HIV. Due to significant advances in antiretroviral therapy (ART), HIV has shifted from being a fatal disease to a manageable chronic condition. As a result, treatment strategies for HIV have evolved accordingly. The effectiveness and widespread availability of ART have substantially reduced mortality rates, bringing them close to those of the general population. However, despite these positive developments, the long-term health of people living with HIV (PLHIV) still lags behind that of the general population. PLHIV continue to face more physical and mental health challenges, as well as social discrimination. Despite the remarkable progress, the persistent health disparities between PLHIV and the general population highlight the need for continued efforts to address these issues comprehensively. Addressing both medical and social factors is essential to improving the overall well-being of PLHIV.[25] Emphasizing the importance of equitable access to treatment, stigma reduction, and attention to psychosocial needs. This holistic approach underscores that interventions must go beyond clinical care to have a meaningful impact on the quality of life of individuals. In this scenario, hospital pharmacy plays a crucial role, not only in the dispensing and monitoring of ART but also in patient education, treatment adherence, and side effect management. This contributes significantly to monitoring and improving the HRQoL of PLHIV.[26–28]

The integration of telemedicine with ePROMs and ePREMs has enabled the application of artificial intelligence (AI) and Machine Learning (ML) techniques.[29] ML models using ePROMs and ePREMs data alongside clinical, analytical, sociodemographic parameters can improve follow up, diagnostic accuracy and personalize treatment plans tailored to individual needs. These data also support monitoring and predicting disease progression or treatment impact on HRQoL and symptoms over time. Additionally, ePROMs and ePREMs provide valuable outcome measures for evaluating the effectiveness of AI interventions, ensuring patient-centered care remains a priority in the digitization of healthcare.[30]

ML is increasingly used to predict ePROMs and ePREMs scores, leveraging electronic medical records (EMRs) and patient-specific data to enhance predictive accuracy and treatment outcomes. Studies have shown ML's effectiveness in analyzing HRQoL data, particularly in conditions like neurological disorders or cancer.[31–33] For HIV patients, ML models outperform traditional methods by predicting risks such as treatment interruptions and failure to achieve viral suppression, using demographic, clinical, and behavioral data.[34,35] Additionally, significant research has focused on predicting ART responses through genotypic sequences, employing either rule-based systems or ML models. Rule-based approaches use expert-curated tables mapping drug resistance to specific mutations, which are then assigned resistance scores for genotypes.[36] In contrast, ML models are trained directly on genotypic data to predict treatment outcomes.[37,38] Recent advancements combine drug resistance scores with genotypic data, further enhancing predictive capabilities and

informing ART strategies.[39] These advances promote shared decision-making and personalized care goals, though challenges remain in data quality, model validation, clinical integration, data interoperability and ethical issues like privacy and bias. Collaborative efforts among healthcare providers, data scientists, and patients are essential to fully realizing ML's potential for patient-centered, VBH.[30,40–42]

In this dynamic and transformative context, we have developed a prototype of a predictive model supported by ML algorithms within the ALGOPROMIA project. The model is designed to predict both subdomains and overall scores of ePROMs and ePREMs, based on the sociodemographic and clinical characteristics of PLHIV, with a focus on changes in patient states such as symptoms and HRQoL. This AI-powered model uses data from more than 1200 PLHIV monitored within the NAVETA system.

Materials and Methods

Database Usage

To conduct this research, we utilized a HRQoL database supplemented by sociodemographic and pharmacological data collected through various monitoring programs within the NAVETA system.[43] NAVETA is an innovative platform designed to facilitate remote patient monitoring and engagement by collecting both ePROMs and ePREMs data. The study used the NAVETA telepharmacy care model, developed in collaboration with FARUPEIB and BiblioPRO, to evaluate the treatment impact on patients' quality of life and disease symptoms. NAVETA operates independently of Electronic Medical Records (EMR), ensuring secure data storage in compliance with Spanish and European data protection standards.[43] The system enables standardized and continuous data collection, allowing healthcare providers to monitor patient-reported outcomes and experiences in real-time. Its robust infrastructure supports the collection of raw responses and aggregated scores, including subdomain-level and global scores for each ePROMs and ePREMs, making it an excellent tool for implementing predictive models like ALGOPROMIA.

The database comprised records from 1240 patients, encompassing around 240,000 responses to ePROMs and ePREMs. The dataset spanned from January 1, 2021, to June 1, 2024, across hospitals within the Spanish public health system. Prior to data extraction from NAVETA, an anonymization process was implemented to safeguard patient confidentiality, ensuring that the dataset used for both model training and validation purposes did not contain identifiable patient information. The dataset included 77.22% of male patients and 20.93% of female patients with chronic HIV who responded to the standard set of ePROMs and ePREMs designed specifically for the monitoring of PLHIV (see Table 1). Data collected were analyzed and plotted using Origin (Version 2021, OriginLab Corporation, Northampton, MA, USA).

Table 1 Frequency and Scope of Patient-Reported Measures (PROMs and PREMs) in the NAVETA HIV Standard Set

PROMs / PREMs	Dimensions Measured	Total responses
MOS30 VIH	Health-related quality of life, physical and mental well-being	76,629
HIV SI	HIV-specific symptoms, emotional well-being	22,232
Morisky-Green	Medication adherence	10,488
SMAQ	Medication adherence, treatment compliance	5676
P3CEQ	Patient care experience, quality of care	5184
WRFQ	Work-related functional status	2582
HADS	Anxiety and depression levels	2425
ESTAR	Antiretroviral treatment satisfaction scale	2244
Escala Berger	Burden of symptoms, mental distress	670

(*Continued*)

Table 1 (Continued).

PROMs / PREMs	Dimensions Measured	Total responses
WHOQoL HIV Bref	Overall quality of life, physical and psychological health	600
EESS	Self-reported side effects	135
TSQM	Treatment satisfaction, side effects, treatment efficacy	105
EQ5D	General health status, usual activities, pain/discomfort	56
PROMIS-29	Physical, mental, and social health	38
MMAS-8	Medication adherence scale	9

Notes: This table displays a list of PROMs and PREMs used in the study, each ranked by the total number of responses received. It highlights the dimensions each PROM measures, providing insights into the most engaged aspects of patient health and satisfaction. Frequencies indicate the extent of each PROM's application, reflecting its relevance and utility in capturing diverse health outcomes.

Abbreviations: *MOS SF-30 hIV*, Medical Outcomes Study Short Form 30 hIV; *HIV SI*, HIV Symptom Index; *Morisky-Green*, Morisky Medication Adherence Scale; *SMAQ*, Simplified Medication Adherence Questionnaire; *P3CEQ*, Patient Care Experience Questionnaire; *WRFQ*, Work Role Functioning Questionnaire; *HADS*, Hospital Anxiety and Depression Scale; *ESTAR*, Antiretroviral treatment satisfaction scale; *Escala Berger*, Berger's Symptom Checklist; *WHOQoL HIV Bref*, World Health Organization Quality of Life HIV Short Form; *EESS*, Self-reported side effects scale; *TSQM*, Treatment Satisfaction Questionnaire for Medication; *EQ5D*, EuroQol 5-Dimension Scale; *PROMIS-29*, Patient-Reported Outcomes Measurement Information System 29-Item Profile and *MMAS-8*, Morisky Medication Adherence Scale 8-Item Version.

Data Preprocessing and Modeling Preparation

In this study, all ML analyses were conducted using the Python programming language (Version 3.10). The ML pipeline used in this work is illustrated in Figure 1. To prepare the data for modeling, a preprocessing process was conducted. Briefly, key variables were selected from the dataset, including features such as disease stage (referring to the phase of the patient's treatment), age, body mass index (BMI), gender, employment status, education level, daily activity, physical activity, balanced diet, patient association membership, marital status, housing situation, alcohol consumption, tobacco use, caffeine intake, drug allergies, treatment line and the specific questionnaire used.

In this study, a Random Forest Regressor (RF) model was implemented. The RF algorithm is well-established in machine learning for its capability to build predictive models, first introduced in 2001.[44] This supervised learning method is composed of multiple tree predictors, where each tree is built using values from an independent random vector. All vectors are created using the same configuration. Random Forests are versatile and can effectively handle both classification and regression tasks.[45–47] To optimize the Random Forest model's performance, we performed a systematic hyperparameter search using GridSearchCV function from the sklearn.model selection library.

Key hyperparameters commonly optimized for Random Forest models include n_estimators, max_depth, bootstrap, max_features, min_samples_leaf, criterion, and min_samples_split.[48] For this study, we focused on three key hyperparameters: n_estimators, min_samples_leaf, and min_samples_split. The search grid included the following values: n_estimators: [100, 200, 300, 500], min_samples_leaf: [1, 2, 4], min_samples_split: [2, 5, 10]. The best-performing combination, selected based on mean cross-validated RMSE, was n_estimators = 300, min_samples_leaf = 1, and min_samples_split = 2. These optimized hyperparameters were subsequently used to train the final model.

Performance metrics such as Mean Squared Error (MSE) (equation 1) and R-squared (R^2) were computed to assess the model's predictive accuracy on scaled target values (equation 2). To ensure robust evaluation, a stratified 5-fold cross-validation was performed on the training data, yielding Root Mean Squared Error (RMSE) scores for each fold and allowing for comparison against the RMSE computed on the test set.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\gamma i - xi)^2$$

Where, n is the number of observations in the dataset and: The actual observed values for each data point,: The predicted values from the model for each data point. This formula thus provides a comprehensive measure of the model's

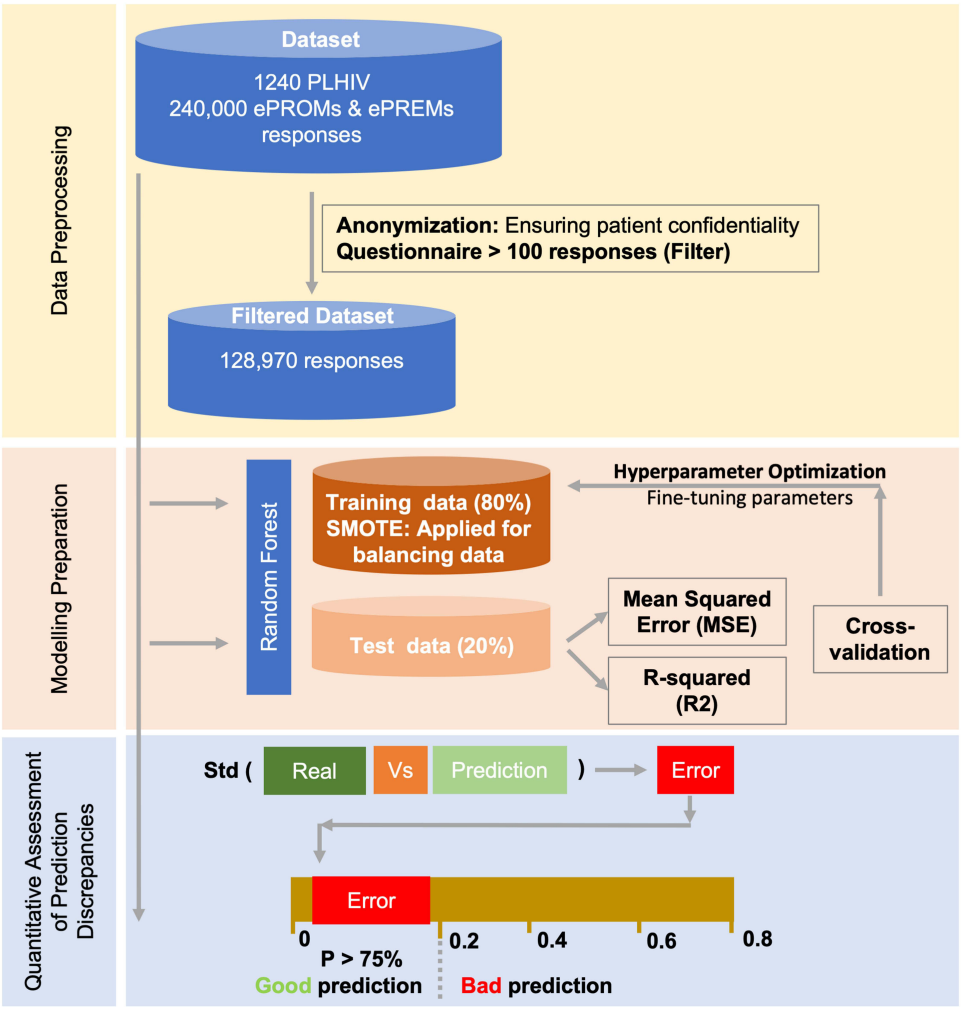

Figure 1 Data Processing and Validation Pipeline for ALGOPROMIA analysis. The workflow of the Machine Learning pipeline used in this study.

prediction accuracy by calculating the average of the squares of the errors, highlighting how closely the model's predictions match the actual data points.

$$R^2 = 1 - \frac{SS\,res}{SS\,tot}$$

Were *SSres* is the sum of the squared differences between the observed values and the values predicted by the decision tree and *SStot* is the sum of the squared differences between the observed values and the mean of the observed values. For decision tree regression models, a high R-squared value indicates that the model explains a significant portion of the variance in the data, suggesting a good fit. Conversely, a low R-squared might suggest underfitting, where the model fails to capture important explanatory information.

Model Training

We select only those questionnaire responses that received at least 100 answers, resulting in a final dataset of 128,970 responses to 12 questionnaires including MOS SF-30 hIV (Medical Outcomes Study Short Form 30 hIV), HIV SI (HIV Symptom Index), Morisky-Green (Morisky Medication Adherence Scale) 8-Item Version,[49,50] SMAQ (Simplified Medication Adherence Questionnaire), P3CEQ (Patient Care Experience Questionnaire), WRFQ (Work Role Functioning Questionnaire), HADS (Hospital Anxiety and Depression Scale), ESTAR (Antiretroviral treatment satisfaction scale), Escala Berger (Berger's Symptom Checklist), WHOQoL HIV Bref (World Health Organization Quality of

https://doi.org/10.2147/IPRP.S492422

Life HIV Short Form), EESS (Self-reported side effects scale), and TSQM (Treatment Satisfaction Questionnaire for Medication). The NAVETA platform collects both raw data and aggregated scores (calculated online based on patient responses), calculated for each subdomain of the ePROMs/ePREMs as well as their overall composite global scores. For the ALGOPROMIA model, both subdomain and global scores are included as prediction targets. For instance, the WHOQoL HIV Bref questionnaire is structured into six specific domains: Physical Health, Psychological Health, Level of Independence, Social Relationships, Environment, and Spirituality/Religious/Personal Beliefs. Additionally, it includes general scores for overall quality of life and health. In this study, the model predicts scores for each of these domains, as well as the overall quality of life score, demonstrating its ability to address both detailed and aggregated dimensions of patient well-being. To ensure consistency across the diverse scoring methods of the ePROMs and ePREMs included in this study, all target scores (subdomains and global scores) were normalized using the StandardScaler function.[51] This transformation scales the scores to have a mean of 0 and a standard deviation of 1, allowing the model to process the data without bias introduced by differences in scoring ranges (eg, scales of 0–10 vs 0–100). The scaling process was applied separately to the training and test sets. Specifically, the scaler was fitted on the training set and then applied to the test set to avoid information leakage. By normalizing the target variables, the model achieves better stability and consistency during training, accommodating the heterogeneity of the instruments included in the dataset.

Subsequently, we split the dataset into two groups, one for training and another for validation, with 80% and 20% of the total responses, respectively. To prevent data leakage, all responses from the same patient were grouped and assigned exclusively to either the training or the validation set, ensuring no overlap of patient-specific information between the two datasets. This strategy helps to mitigate the risk of artificially inflated performance metrics caused by patterns unique to individual patients.[52] Due to data imbalance across different ePROMs/ePREMs (see Table 2), we applied a SMOTE (Synthetic Minority Oversampling Technique) oversampling process to the training dataset. This process was applied to all variables in the training data except for the questionnaire variable, which was used as a stratification variable during the model training to preserve the proportional distribution of ePROMs/ePREMs categories.

Table 2 Demographic and Behavioral Characteristics of Study Participants

BMI Range	**n**	**%**	**Gender**	**n**	**%**	**Balanced diet**	**n**	**%**
Normal weight	112,933	48.24	Male	180,787	77.23	Regular	118,235	50.51
Overweight	77,036	32.91	Female	49,002	20.93	Habitual	70,354	30.05
Obesity	31,085	13.28	Other	528	0.23	Occasional	41,25	17.62
Underweight	9,131	3.90				Never	4,254	1.82
Unclassified	132	0.06						
Education Level	**n**	**%**	**Age Range**	**n**	**%**	**Membership in Patient Association**	**n**	**%**
Secondary	69,311	29.61	46–65	129,549	55.34	No	216,02	92.28
Degree	58,272	24.89	21–45	73,324	31.32	Yes	11,926	5.09
Professional	54,614	23.33	66–100	12,956	5.53	NR	6,147	2.63
Primary	40,937	17.49						
Doctorate	3,753	1.60						
None	3,43	1.47						

(*Continued*)

Table 2 (Continued).

Employment Status	**n**	**%**	**Alcohol Consumption**	**n**	**%**	**Smoking Status**	**n**	**%**
Employed	146,179	62.44	Occasionally	135,248	57.77	Yes	43,619	18.63
Unemployed	43,24	18.47	Never	57,386	24.51	Ex-smoker	37,168	15.88
Retired	17,732	7.57	Weekends	25,655	10.96	Never	35,071	14.98
Disability	12,338	5.27	Daily	10,383	4.44			
Temporary disability	9,791	4.18						
Student	1,037	0.44						
Treatment Line	**n**	**%**	**Caffeine Consumption**	**n**	**%**	**Drug Allergies**	**n**	**%**
First line	146,173	62.44	Regular	118,235	50.51	No	196,463	83.93
Second line	34,516	14.74	Occasional	65,426	27.95	Yes	33,389	14.26
Third line	29,386	12.55	Never	16,754	7.16	NR	4,241	1.81
Fourth line	9,421	4.02						
Fifth line	8,705	3.72						
Sixth line	3,632	1.55						
Seventh line	1,356	0.58						

Notes: This table presents a comprehensive breakdown of patient data collected in the study, including demographics such as BMI range, gender, and age, educational attainment, employment status, lifestyle choices like alcohol and caffeine consumption, and treatment line adherence. It also covers patient responses on dietary habits, smoking status, and allergies to medications. Percentages are calculated based on the total number of respondents.
Abbreviation: *NR*, No Response.

Quantitative Assessment of Prediction Discrepancies

Here, an analytical approach was employed to assess the discrepancy between predictions from a random forest regression model and the descaled actual values of a target variable. To achieve this, model predictions were descaled using an inverse transformation of the scaling applied during model training, as were the actual test values. Subsequently, these predictions and actual values were standardized to facilitate a uniform comparison of their differences. To ensure the robustness and accuracy of the regression fit in our analysis, we employed a linear regression model characterized by the equation 3.

$$y = a + b * x$$

The regression was performed using standardized actual values, which helps to eliminate bias and provides a more precise comparison. Additionally, no weighting was applied during the fitting process, ensuring that each data point contributed equally to the determination of the model parameters. This strategy allows for a clear and unbiased assessment of the relationship between the actual values and the predictions generated by the ALGOPROMIA model. The absolute standardized error was then computed between the standardized predictions and actual values for each group of a relevant categorical variable. Finally, the standard normal distribution was utilized to determine the probability that the standardized difference was equal to or less than 0.2. If the probability of keeping the error below the threshold remained above 75%, we considered the prediction to be sufficiently accurate, indicating good model performance (see Figure 1).

Results

Cohort of Patients Included in the ALGOPROMIA Study

Responses to HRQoL questionnaires from 1240 PLHIV from the Balearic Islands in Spain, totaling 128,970, were used for training and validating the predictive model. The data shows that normal weight (48.24%) and overweight (32.91%) are the most prevalent BMI categories among the participants. Male participants dominate (77.23%) compared to females (20.93%) and other genders (0.23%). In terms of dietary habits, around half (50.51%) adhere to a habitual balanced diet, while occasional adherence is noted in 30.05% of respondents. Educationally, secondary education (29.61%) and degrees (24.89%) are the most common levels achieved, followed by professional training (23.33%) and primary education (17.49%). Age distribution of participants ranges from 21 to 45 years (31.32%), 46 to 65 years (55.34%), and over 66 years (5.53%). These insights provide a picture of the demographic and lifestyle factors within the surveyed population (see Table 2).

Model Training and Validation

In the performance analysis of the Random Forest model, notable results were observed across several key metrics, underscoring its robustness and generalization capabilities. The analysis demonstrates a strong linear relationship between the real data and the model predictions, corroborated by a high R^2 value of 0.976. The close clustering of points along the identity line indicates that the model has high predictive accuracy (2-A), with minimal deviation between the observed values and predictions. This high R^2 suggests that approximately 97.6% of the variability in the real data is successfully explained by the model, confirming its effectiveness. The distribution of model residuals appears as a narrow, almost normal distribution centered around zero (2-B). This indicates a well-fitted model with symmetrically distributed errors, showing no obvious signs of bias or skewness. The peak around zero underscores the model's accuracy, where most predictions are very close to the true values, leading to small residuals.

The analysis also includes the Root Mean Squared Error (RMSE) for each of the five folds in the cross-validation process, along with the RMSE on a separate test dataset (2-C). The RMSE values from cross-validation range 0.2548 to 0.3197, with an average RMSE for the training data being approximately 0.2607. The RMSE on the test data is slightly lower at 0.1613, which is intriguing as it suggests the model may perform slightly better on unseen data compared to some of the training folds. This could be due to the model's ability to generalize well or variations in the difficulty of the test set compared to specific training folds. The difference in RMSE between the average training data (0.2607) and the test set (0.1613) further supports the idea that the model is not overfitting; rather, it demonstrates an improved performance on unseen data. This is an excellent sign of the model's robustness and its capability to generalize beyond the training dataset. While some variability is observed across different folds of cross-validation, the model demonstrates reasonable overall performance. This indicates that its reliability is sufficient for practical applications, although further refinement may be needed to improve consistency across data subsets. Potential factors contributing to this variability, as well as strategies for refinement, are explored in the discussion section.

Success Rate in Questionnaire Prediction

To ensure the accuracy and effectiveness of our model in making individual-level predictions, we implemented an analysis that involved a direct comparison between actual values and predictions generated by the model. First, we analyzed the results of contrasting the actual values with the predicted values (3-A). As shown in Table 3, the table presents the results of analyzing the relationship between actual values and those predicted by the ALGOPROMIA model for various questionnaires. Notable performance is observed with the MOS30 VIH questionnaire, which achieves an Adjusted R-Square (Adj. R^2) of 0.984, indicating that the model explains over 98% of the variance in actual values. The ESTAR questionnaire also demonstrates exceptional performance, with an Adj. R^2 of 0.963, reflecting its high accuracy. The BERGER questionnaire achieves an Adj. R^2 of 0.936 and a Pearson's r of 0.968, while the HIV SI questionnaire shows good performance with an Adj. R^2 of 0.828. In contrast, the TSQM questionnaire, with a Pearson's r of 0.836 and an Adjusted R-Square of 0.698, demonstrates a decent linear relationship but with room for improvement. Similarly, the P3CEQ questionnaire shows a moderately strong fit with an Adj.R2 of 0.753. On the other hand, questionnaires such

Table 3 Analysis of ALGOPROMIA Model Performance Across Various Questionnaires

Questionnaire	Intercept	Slope	Pearson's r	R-Square (COD)	Adj. R-Square
ESTAR	−0.003 ± 0.008	0.98816 ± 0.0076	0.98119	0.96274	0.96268
BERGER	−0.00578 ± 0.00268	0.98318 ± 0.00264	0.96763	0.93631	0.9363
HIV SI	−0.00996 ± 0.0011	0.97333 ± 0.00167	0.91022	0.8285	0.8285
WRFQ	−0.02968 ± 0.0279	0.99557 ± 0.00783	0.83702	0.7006	0.70056
MOS30 VIH	0.00847 ± 0.00166	0.99924 ± 0.00181	0.99217	0.9844	0.98439
WHOQoL HIV Bref	6.05556E-4 ± 4.02643E-4	0.98359 ± 0.00368	0.72835	0.53049	0.53048
HADS	−2.55271E-4 ± 0.003	0.98636 ± 0.00771	0.73426	0.53914	0.53911
P3CEQ	7.83007E-4 ± 0.00368	0.98639 ± 0.01144	0.86772	0.75293	0.75283
TSQM	−0.02447 ± 0.04505	0.99464 ± 0.01391	0.83568	0.69836	0.69822
SMAQ	9.34924E-4 ± 0.00787	0.98975 ± 0.01201	0.76898	0.59133	0.59124

Notes: The regression analysis results for various questionnaires compare actual values with those predicted by the ALGOPROMIA model.

as HADS, SAMQ and WHOQoL HIV Bref demonstrate lower Adjusted R-Square values of 0.539, 0.591 and 0.530, respectively, highlighting potential challenges in accurately capturing their variance. These results suggest that further refinement of the model or additional data preprocessing may be beneficial for these instruments. The metrics such as intercept, slope, Pearson's r and Adj.R^2 collectively demonstrate the model's robustness in predicting health-related quality of life measures. For instance, the intercept and slope values indicate a strong linear relationship. High Pearson's r values reflect strong correlations, and high R-Square and Adjusted R-Square values underscore the model's explanatory power. Overall, these results highlight the ALGOPROMIA model's potential for accurate and effective predictions in managing health outcomes for individuals living with HIV.

Next, we calculated the difference between each pair of standardized values, that is, between each actual value and its corresponding prediction. This standardized difference allowed us to measure the error accurately. We then assessed the probability that this error would be below a specific threshold, set at 0.2 (see Materials and Methods), to determine the precision of the predictions. If the probability of keeping the error below the threshold remained above 75%, we considered the prediction to be sufficiently accurate, indicating good model performance (see Figure 1). This strategy not only ensures the global validity of the model but also its applicability and precision at the most detailed and practical level of each individual prediction. Overall, we observed that the probability of keeping the error below 0.2 exceeded 80% in most instruments (Figure 3A, histogram). Specifically, the ALGOPROMIA model demonstrated excellent performance in predicting the scores, maintaining the error below the threshold, for the SMAQ (100%), Naveta Satisfaction Form (99.99%), HIV SI (99.24%), WHOQoL HIV Bref (96.43%), MOS30 VIH (93.22%), ESTAR (88.44%), HADS (86.90%), P3CEQ (79.73%), and (66.14%). However, in the Berger Scale, WRFQ and TSQM questionnaires, it showed lower performance with probabilities of maintaining the error of the predictions below 0.2 at 66%, 51% and 44%, respectively. These results suggest that, although ALGOPROMIA is capable of predicting 73.83% of the questionnaires from the standard NAVETA VIH set (3-B), there is still room for improvement.

Discussion

The incorporation of emerging technologies, such as IA and ML, should improve pharmaceutical drug validation, clinical decision making, drug adherence, drug-drug interaction identification or treatment personalization between others challenges. In recent years, the integration of ML techniques with ePROMs and ePREMs has significantly advanced the predictive capabilities in healthcare. Several studies have explored various supervised ML algorithms to effectively predict patient-specific outcomes from ePROMs data. For instance, these methods have demonstrated substantial success in predicting outcomes such as unplanned hospital utilization and mortality, highlighting the value of ePROMs in

enhancing model performance alongside objective clinical data. Furthermore, specific applications of ML algorithms have been examined for their potential to improve decision-making processes by accurately forecasting diverse ePROMs. A comprehensive literature review has also noted an increasing trend in leveraging ML approaches to analyze ePROMs datasets, emphasizing the importance of feature selection, model evaluation, and overall performance metrics. Additionally, studies have shown promising results in estimating the success of rehabilitation and surgical treatments based on a combination of clinical assessments and ePROMs inputs. This evolving landscape underscores the growing recognition of ML as a transformative tool for optimizing healthcare outcomes through detailed and personalized patient data analysis.[53–55]

ML models have shown significant potential in enhancing HRQoL measures for individuals living with HIV. For instance, predictive models have demonstrated robust performance, achieving AUC metrics as high as 0.76 in predicting viral load suppression—a critical determinant of quality of life—accurately classifying viral suppression in approximately three out of four patients.[35] Additionally, ML algorithms can precisely identify patients at risk for poor HRQoL outcomes based on their EMR data, enabling healthcare providers to proactively intervene and improve outcomes in HIV care.[34] However, to date, we are not aware of any system that utilizes an ML system to predict a comprehensive set of HRQoL questions for HIV patients. The ALGOPROMIA prototype focused on chronic HIV patients also delves into the goal proposed, of using ePROMs and ePREMs data as a result of predictive models, with the clear intention of improving patient-centered care.[30] In developing the prototype, we utilized data from more than 1200 PLHIV, a substantially larger dataset compared to those used in similar studies within the field.[55] This substantial patient dataset enhances the robustness and reliability of our predictive model, which benefits from the extensive and well-structured data available in NAVETA. Additionally, addressing a common criticism in the field, we ensure transparency in the development, training, and validation processes of our machine learning models by providing a detailed description of these steps (Figures 1–3). The results of our work demonstrated that the Random Forest model employed was able to predict with high accuracy ($R^2 = 0.976$) the responses to the HRQoL questionnaires ePROMs and ePREMs in AIDs patients. The performance analysis revealed the model's robustness and generalization capabilities, explaining approximately 97% of the variance in the target variable. Additionally, we compared Random Forest with other ML models, including Gradient Boosting, Neural Networks, and Linear Regression. Random Forest outperformed these approaches, achieving an MSE of 0.026 and an R^2 of 0.976, compared to Gradient Boosting (RMSE: 0.068, R^2: 0.937), Neural Network (RMSE: 0.167, R^2: 0.847), and Linear Regression (RMSE: 0.912, R^2: 0.166). While Gradient Boosting provided competitive results, the Random Forest model displayed slightly better predictive accuracy and consistent performance across different subsets of data, suggesting a stronger generalization ability. This supports its selection as the foundation for the ALGOPROMIA model, which aims to reliably predict HRQoL outcomes and support personalized care strategies for HIV patients.

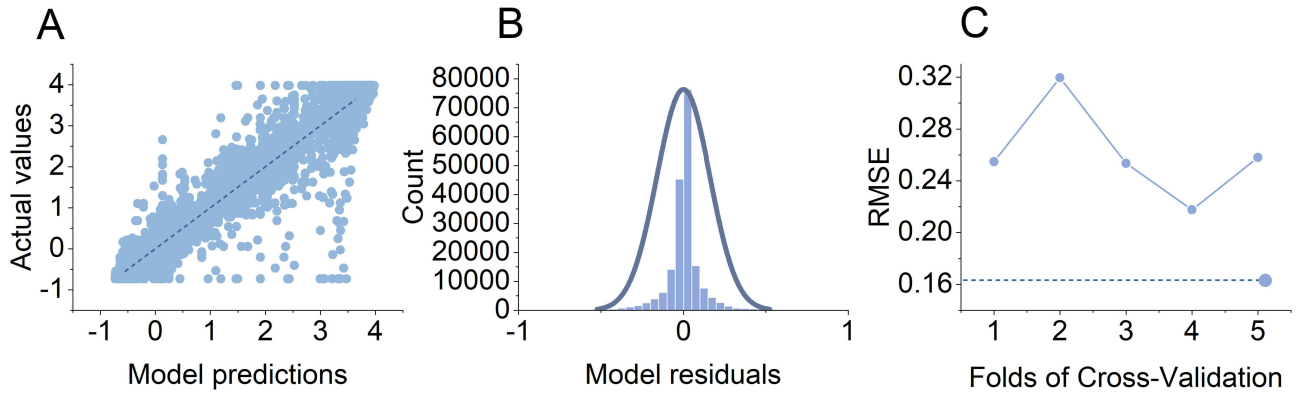

Figure 2 Analysis of Model Performance and Predictive Accuracy. Panel (**A**) Illustrates the strong linear relationship between actual data and model predictions, highlighted by a high R^2 value of 0.976, indicating that the model explains approximately 97.6% of the data variability. Panel (**B**) Shows the distribution of residuals, centered around zero with a nearly normal distribution, emphasizing the model's precision and the symmetry of prediction errors. Panel (**C**) Depicts the Root Mean Squared Error (RMSE) across five folds of cross-validation, ranging from 0.254 to 0.319, alongside a test dataset RMSE of 0.161, demonstrating the model's consistent performance and effective generalization from training to unseen data.

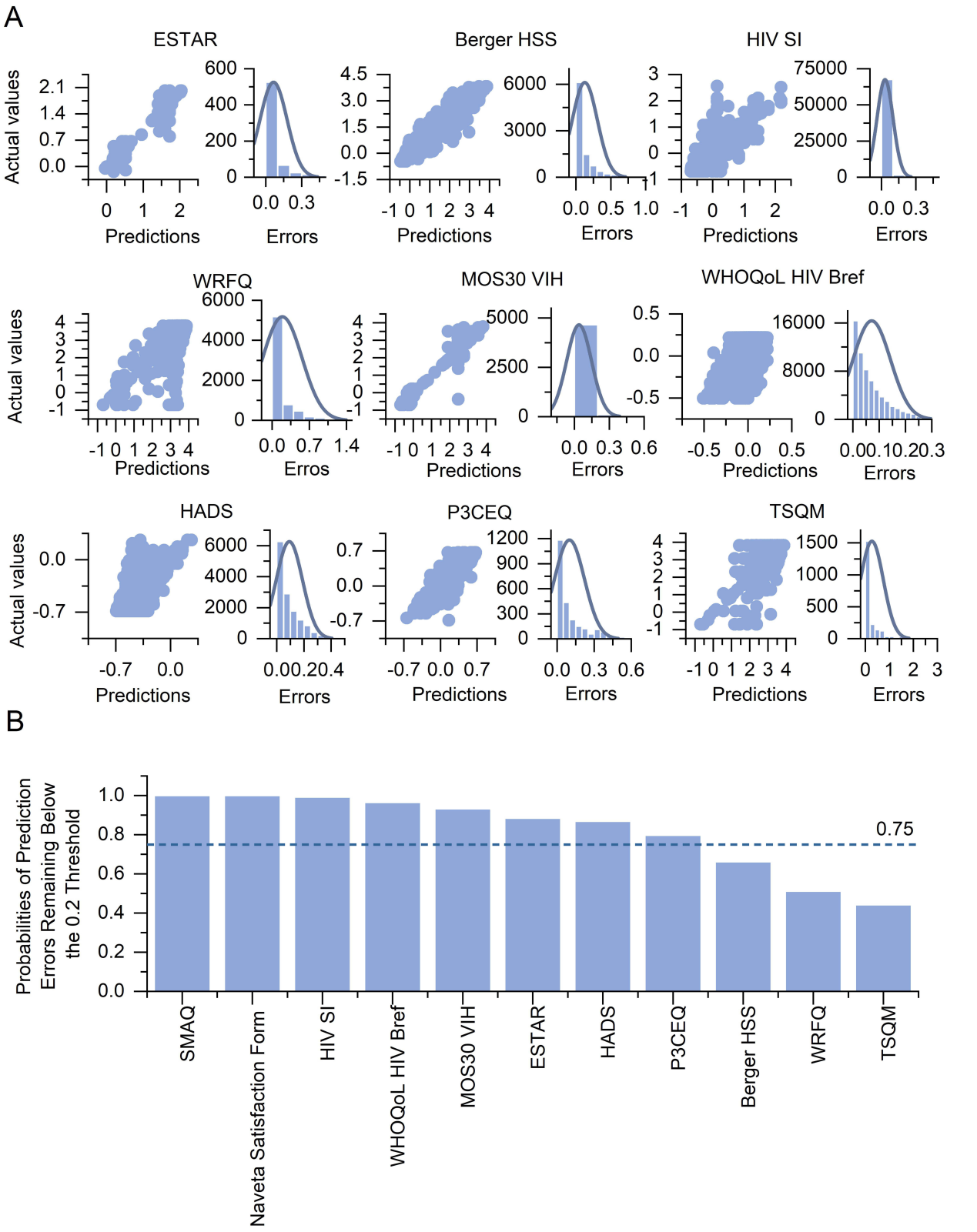

Figure 3 Assessment of Model Performance Across Various Health-Related Instruments. (**A**) Scatter plots and error distributions for each instrument, illustrating the alignment between actual values and model predictions (left, see Table 3) and the distribution of errors (right). Each plot demonstrates how closely predictions match the actual data and the frequency and extent of errors. (**B**) Bar chart showing the probability that the error remains below the threshold (0.2) in predictions across various instruments. This graph highlights the effectiveness of the ALGOPROMIA model in consistently producing errors within acceptable limits, with a detailed comparison of performance between different health-related metrics. The dotted line represents the threshold for good (>0.75) and poor (<0.75) predictions.

https://doi.org/10.2147/IPRP.S492422

The RMSE exhibited slight variability between training and test data but remained consistent during cross-validation, reflecting the model's stability across different data subsets. Moreover, the exceptionally high correlation between actual and predicted values, with an RMSE of 0.026, underscores the model's precision and capacity to generalize effectively to new data (Figure 2). The evaluation of the ALGOPROMIA model includes RMSE for each of the five folds in the cross-validation process, as well as the RMSE on a separate test dataset. RMSE values from cross-validation range from 0.2177 to 0.3196, with an average RMSE of approximately 0.2607 for the training data. The slightly lower RMSE on the test data (0.1613) suggests the model performs better on unseen data, likely due to effective generalization. The observed trends in RMSE, with variability across folds, reflect differences in data distribution and complexity within the dataset. Such trends are consistent with findings from prior studies, which highlight that fluctuations in cross-validation results are inherent in heterogeneous datasets and do not imply instability.[56–58] Despite these variations, the narrow RMSE range across folds highlights the model's reliability and robustness for practical applications.

In the detailed evaluation of the model's performance for each questionnaire, we observed that in most HRQoL questionnaires, the model achieves Pearson correlation coefficients above 0.8 in the fit between predicted values and actual values (Table 3). This indicates a very strong relationship between the two variables, demonstrating that the model is highly accurate and effective in its predictions. Innovatively, we included a new system for verifying the quality of the prediction by calculating the standardized difference between the actual value and the prediction. We then assessed the probability that such an error would remain below 0.2 for each of the questionnaires. The analysis showed that for most instruments evaluated, the probability of keeping the error below this threshold exceeded 85%, suggesting that the model is capable of making predictions with a high degree of precision and that deviations from real values are minimal for most cases. Interestingly, discrepancies arise in some cases, such as the WHOQoL HIV Bref questionnaire. Although this instrument exhibits a relatively low Adjusted R-Square (0.530), indicating limited variance explained by the model, the probability of maintaining the prediction error below 0.2 is remarkably high at 96.43%. This highlights an important nuance: while metrics like Adjusted R-Square evaluate the model's ability to capture variance, practical performance measures, such as error thresholds, provide complementary insights into its clinical applicability and precision. These results reinforce the value of assessing multiple performance metrics to gain a more comprehensive understanding of the model's capabilities.

The ALGOPROMIA model is remarkably precise in predicting aspects of HIV treatment adherence, mental health status, and stigma, demonstrating minimal deviations from actual values across most cases. For instance, in key questionnaires such as the SMAQ, WHOQoL HIV Bref, HADS, HIV SI and MOS30 VIH, the model achieved nearly 100% accuracy in its predictions. This remarkable precision aids healthcare providers in effectively identifying and addressing crucial factors that influence treatment outcomes, reduce transmission risks, and enhance the decision-making process in patient care what may result in a better and more efficient resources use in this setting. By accurately predicting adherence and mental health challenges, the ALGOPROMIA model enables targeted interventions and personalized care plans, ultimately leading to improved health outcomes and patient satisfaction within the HIV community. The insights provided by ALGOPROMIA add to the existing body of predictive models in HIV care, including those based on rule-based systems and ML approaches for ART response predictions. These models, which utilize genotypic data and drug resistance scores, have significantly advanced treatment strategies. ALGOPROMIA contributes further by focusing on patient-reported outcomes and experiences, enriching the decision-making process not only with insights into HRQoL but also with actionable information on treatment adherence and clinical efficacy. However, not all instruments demonstrated equally high predictive performance. While the model performed exceptionally well for the majority of the questionnaires, others, such as TSQM (Adj. $R^2 = 0.698$) and WRFQ (Adj. $R^2 = 0.701$), showed relatively lower predictive accuracy and exhibited probabilities of maintaining prediction errors below 0.2 at 44% and 51%, respectively, compared to the consistently high probability observed for instruments like WHOQoL HIV Bref (96.43%). The lower performance observed in questionnaires such as TSQM and WRFQ may be attributed to several factors. First, the limited number of responses for these instruments compared to others, such as WHOQoL HIV Bref, reduces their representation during model training. Second, while SMOTE oversampling partially addresses data imbalance, it may not fully capture the complexity of underrepresented instruments, potentially leading to less precise predictions. Additionally, during cross-validation, the smaller sample size for these questionnaires results in higher

variability and reduced reliability of the model's performance. Lastly, the dominance of highly represented questionnaires in the training dataset could skew the model's predictions towards more common patterns, further impacting the accuracy for less frequent instruments. Despite this, it is important to note that some questionnaires with lower representation in the dataset, such as the SMAQ (100% probability of maintaining the error below 0.2) and ESTAR (88.44%), demonstrate exceptionally high performance. This finding indicates that limited representation is not the sole factor determining the model's performance on specific instruments. Other elements, such as the simplicity of the questionnaire, the quality of its design, and the distribution of responses, may play a crucial role. For example, less complex questionnaires may facilitate clear predictive patterns even with fewer data points. Similarly, questionnaires that capture redundant or overlapping information with more represented ones, may benefit from patterns learned across the broader dataset. These observations suggest that intrinsic factors related to the design and structure of the questionnaires also influence predictive accuracy and warrant further exploration in future studies.

The integration of ML techniques into the management of chronic conditions, particularly HIV, is revolutionizing personalized clinical care. EMR-based clinical decision support tools and electronic alerts have significantly improved outcomes across the HIV care continuum, including diagnosis, retention in care, and viral suppression.[34] This enhances patient outcomes and increases the accessibility of healthcare services, ensuring that individuals living with HIV receive timely and personalized care, regardless of geographic location. The use of technologies like artificial intelligence and ML not only enhances clinical and pharmaceutical decision-making and treatment personalization but also optimizes telepharmacy services. This contributes to overcoming geographic barriers, stratify patients and increasing access to quality pharmaceutical care.[43,59] However, for models like ALGOPROMIA to be successfully expanded to larger and more heterogeneous populations, it is critical to address barriers and leverage enablers of ePROMs/ePREMs implementation.[7,17] Factors such as clinician support, patient involvement, and institutional commitment play a pivotal role in ensuring broader adoption and scalability, ultimately maximizing the model's potential to transform patient care across diverse settings.[60] In this sense, we have recently published how ML technology can be used to identify patients with low adherence to follow-up programs via ePROMs, demonstrating that more personalized strategies can enhance not only the adoption of ePROMs and ePREMs by clinical centers but also patient response rates.[61]

In summary, our research underscores the transformative potential of ML in delivering value-based healthcare for PLHIV through advanced telecare and telepharmacy services, marking the evolution of HIV into a manageable chronic condition. By leveraging these digital health services, care precision and accessibility are improved, allowing for more effective remote monitoring and treatment management. This not only elevates healthcare standards but also fosters the adoption of personalized medicine, leading to better patient outcomes and more efficient healthcare systems. The ALGOPROMIA project prototype, which utilizes telemedicine-generated data and advanced data analysis techniques to predict HRQoL measures, exemplifies this shift. It enables a deeper understanding of patient needs and supports a more tailored therapeutic approach. ML application within the project personalizes treatment plans and enhances the patient journey, improving clinical outcomes, treatment safety, and patient adherence which result in a more efficient health resources use. Moreover, this methodology streamlines healthcare professionals' schedules and boosts patient satisfaction through shared decision-making, Accurate predictions of ePROMs and ePREMs enable targeted interventions that meet individual patient expectations, ensuring personalized and effective care while also optimizing resource allocation proactively to meet anticipated needs.

Despite promising results, the study faces limitations, including dependence on data quality and sample representativeness, as the data comes exclusively from patients within the Balearic Spanish public health system, which may not be generalizable to other contexts or populations. Additionally, the sample has an overrepresentation of male participants, which could limit the applicability of the findings to more diverse populations or bias gender. Furthermore, while machine learning techniques are powerful, they require rigorous validation to ensure that the models are not only accurate but also free from biases and applicable in real-world clinical practice. Another limitation of the study is the imbalanced distribution of responses across the various ePROMs and ePREMs instruments included in the dataset. Although the application of SMOTE oversampling to the training data partially addressed this issue, this method has inherent constraints, particularly for instruments with very low representation. In this regard, the NAVETA initiative is actively working to increase response rates to ePROMs and ePREMs, aiming to achieve greater representation by

designing systems to detect patients at high risk of dropping out from follow-up programs via ePROMs and PREMs.[61] These constraints could impact the model's ability to produce robust and reliable predictions for such instruments. To further improve model performance and fairness, future research should explore alternative strategies to address data imbalance more effectively. Potential approaches include advanced resampling techniques, cost-sensitive learning, or integrating domain-specific insights to prioritize critical underrepresented data.[62]

Conclusions

The results of our ML algorithms demonstrate significant potential for predicting ePROMs and ePREMs in the context of AIDS care. This study highlights how integrating emerging technologies, such as machine learning, can enhance clinical and pharmaceutical decision-making and enable personalized treatment within a multidisciplinary framework. However, successful implementation requires addressing challenges such as handling diverse scoring systems, mitigating biases, and ensuring generalizability across different populations. Integrating such predictive models into independent platforms like NAVETA could serve as a crucial first step towards broader adoption and eventual integration with electronic health record systems. This approach not only facilitates early implementation but also allows for iterative improvements before scaling to more comprehensive clinical settings. Ultimately, our findings support the ongoing adoption of advanced ML technologies in healthcare to enhance value-based care and improve health outcomes.

Ethics Approval

This study is part of a wider project with approval by the Research Ethics Committee of the Balearic Islands (IB 4542/21 EOm).

Acknowledgments

We acknowledge the use of the Morisky Medication Adherence Scale (MMAS-4) in this study. ©MMAS 2006 www.adherence.cc.

Author Contributions

Conceptualization, S.H.-P. and G.M.-O.; data acquisition and interpretation, S.H.-P.; data analysis, S.H.-P.; drafting of the manuscript, S.H.-P.; critical revision and editing of the manuscript, G.M.-O. and S.H.-P.; supervision, S.H.-P. and G.M.-O.; funding acquisition, G.M.-O.; approval of the final version, all authors. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Guyatt G. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420. doi:10.1001/jama.1992.03490170092032
2. Sur R, Dahm P. History of evidence-based medicine. *Indian J Urol*. 2011;27(4):487. doi:10.4103/0970-1591.91438
3. Zimerman AL. Evidence-based medicine: a short history of a modern medical movement. *AMA J Ethics*. 2013;15(1):71–76. doi:10.1001/virtualmentor.2013.15.1.mhst1-1301
4. Guyatt G, Jaeschke R, Wilson MC, Montori VM, Richardson WS. Users' Guides to the Medical Literature: a Manual for Evidence-Based Clinical Practice. In: En Guyatt G, Rennie D, Meade MO, Cook DJ, editors. *What Is Evidence-Based Medicine?* 3rd ed ed. New York, NY: McGraw-Hill Education; 2015.
5. Akobeng AK. Principles of evidence based medicine. *Arch Dis Child*. 2005;90(8):837–840. doi:10.1136/adc.2005.071761
6. Sujan M. Integrating digital health technologies into complex clinical systems. *BMJ Health Care Inform*. 2023;30(1):e100885. doi:10.1136/bmjhci-2023-100885
7. Borges Do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. *Npj Digit Med*. 2023;6(1):161. doi:10.1038/s41746-023-00899-4

8. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform*. 2020;27(1):e100166. doi:10.1136/bmjhci-2020-100166
9. Souliotis K. Patient participation in contemporary health care: promoting a versatile patient role. *Health Expect*. 2016;19(2):175–178. doi:10.1111/hex.12456
10. Brown GC, Brown MM, Sharma S. Value-based medicine: evidence-based medicine and beyond. *Ocul Immunol Inflamm*. 2003;11(3):157–170. doi:10.1076/ocii.11.3.157.17355
11. Bae JM. Value-based medicine: concepts and application. *Epidemiol Health*. 2015;37:e2015014.
12. Lohiniva AL, Isosomppi S, Pasanen S, Sutinen J. A qualitative study to identify thematic areas for HIV related patient-reported outcome measures (PROM) and patient-reported experience measures (PREM). *J Patient-Rep Outcomes*. 2023;7(1):41. doi:10.1186/s41687-023-00582-y
13. Miners A, Phillips A, Kreif N, et al. Health-related quality-of-life of people with HIV in the era of combination antiretroviral treatment: a cross-sectional comparison with the general population. *Lancet HIV*. 2014;1(1):e32–40. doi:10.1016/S2352-3018(14)70018-9
14. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res*. 2013;22(9):2265–2278. doi:10.1007/s11136-013-0390-0
15. Bull C, Teede H, Watson D, Callander EJ. Selecting and implementing patient-reported outcome and experience measures to assess health system performance. *JAMA Health Forum*. 2022;3(4):e220326. doi:10.1001/jamahealthforum.2022.0326
16. Catt S, Starkings R, Shilling V, Fallowfield L. Patient-reported outcome measures of the impact of cancer on patients' everyday lives: a systematic review. *J Cancer Surviv*. 2017;11(2):211–232. doi:10.1007/s11764-016-0580-1
17. Fontaine G, Poitras ME, Sasseville M, et al. Barriers and enablers to the implementation of patient-reported outcome and experience measures (PROMs/PREMs): protocol for an umbrella review. *Syst Rev*. 2024;13(1):96. doi:10.1186/s13643-024-02512-5
18. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Educ*. 2017;17(4):137–144. doi:10.1093/bjaed/mkw060
19. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evid-Based Med*. 2022;27(3):153–155. doi:10.1136/bmjebm-2020-111573
20. Mercadal-Orfila G, Herrera-Pérez S, Piqué N, et al. Person-centered coordinated care experience of people with long-term conditions in the balearic islands measured by the P3CEQ. *Health Serv Insights*. 2024;17:11786329241258856. doi:10.1177/11786329241258856
21. Álvarez-Díaz AM, Mercadal-Orfila G, Ramírez-Herráiz E, Joaquín Borrás-Blasco J. Humanization guide of the Spanish Society of Hospital Pharmacy for patients with inflammatory immune-mediated diseases. *Farm Hosp*. 2022;46(6):340–345.
22. Benson T. Why it is hard to use PROMs and PREMs in routine health and care. *BMJ Open Qual*. 2023;12(4):e002516. doi:10.1136/bmjoq-2023-002516
23. Escudero-Vilaplana V, Bernal E, Casado G, et al. Defining a standard set of patient-reported outcomes for patients with advanced ovarian cancer. *Front Oncol*. 2022;12:885910. doi:10.3389/fonc.2022.885910
24. Cool L, Missiaen J, Vandijck D, et al. An observational pilot study to evaluate the feasibility and quality of oncological home-hospitalization. *Eur J Oncol Nurs*. 2019;40:44–52. doi:10.1016/j.ejon.2019.03.003
25. Lazarus JV, Safreed-Harmon K, Barton SE, et al. Beyond viral suppression of HIV – the new quality of life frontier. *BMC Med*. 2016;14(1):94, s12916–016–0640–4. doi:10.1186/s12916-016-0640-4
26. Scott JD, Abernathy KA, Diaz-Linares M, Graham KK, Lee JC. HIV clinical pharmacists – the US perspective. *Farm Hosp*. 2010;34(6):303–308. doi:10.1016/j.farma.2010.07.001
27. Saberi P, Dong Johnson M, Greenblatt Cocohoba JG, Cocohoba J. The impact of HIV clinical pharmacists on HIV treatment outcomes: a systematic review. *Patient Prefer Adherence*. 2012;297. doi:10.2147/PPA.S30244
28. Mercadal-Orfila G, Font J, Pons M, Blasco Mascaró I, Piqué N. HIV-positive women have higher prevalence of comorbidities and anticholinergic burden: assessment of the HIV-positive population from Menorca (Balearic Islands). *AIDS Care*. 2023;35(6):800–809. doi:10.1080/09540121.2022.2067518
29. Harris JD. Editorial commentary: artificial intelligence analysis of biomedical, large, clinical registry data using machine learning requires tens of thousands of subjects and a focus on substantial clinical benefit: minimal clinically important difference is too low a bar. *Arthrosc J Arthrosc Relat Surg*. 2024;40(4):1164–1167.
30. Cruz Rivera S, Liu X, Hughes SE, et al. Embedding patient-reported outcomes at the heart of artificial intelligence health-care technologies. *Lancet Digit Health*. 2023;5(3):e168–73. doi:10.1016/S2589-7500(22)00252-7
31. Karri R, Chen YPP, Drummond KJ. Using machine learning to predict health-related quality of life outcomes in patients with low grade glioma, meningioma, and acoustic neuroma. Vellido A, editor. *PLoS One*. 2022;17(5):e0267931. doi:10.1371/journal.pone.0267931
32. Savić M, Kurbalija V, Ilić M, et al.Analysis of machine learning models predicting quality of life for cancer patients. En: Proceedings of the 13th International Conference on Management of Digital EcoSystems [Internet]. Virtual Event Tunisia: ACM; 2021. p. 35–42. doi: 10.1145/3444757.3485103.
33. Kim J, Jeong K, Lee S, Baek Y. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a cross-sectional study. *BMC Public Health*. 2024;24(1):159. doi:10.1186/s12889-023-17457-y
34. Ridgway JP, Lee A, Devlin S, Kerman J, Mayampurath A. Machine learning and clinical informatics for improving HIV care continuum outcomes. *Curr HIV/AIDS Rep*. 2021;18(3):229–236. doi:10.1007/s11904-021-00552-3
35. Maskew M, Sharpey-Schafer K, De Voux L, et al. Applying machine learning and predictive modeling to retention and viral suppression in South African HIV treatment cohorts. *Sci Rep*. 2022;12(1):12715. doi:10.1038/s41598-022-16062-0
36. Wensing AM, Calvez V, Ceccherini-Silberstein F, et al. 2022 update of the drug resistance mutations in HIV-1. *Top Antivir Med*. 2022;30(4):559–574.
37. Beerenwinkel N. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res*. 2003;31(13):3850–3855. doi:10.1093/nar/gkg575
38. Pironti A, Pfeifer N, Walter H, et al. Using drug exposure for predicting drug resistance – a data-driven genotypic interpretation tool. Yamanishi Y, editor. *PLoS One*. 2017;12(4):e0174992. doi:10.1371/journal.pone.0174992
39. Di Teodoro G, Pirkl M, Incardona F, et al. Incorporating temporal dynamics of mutations to enhance the prediction capability of antiretroviral therapy's outcome for HIV-1. Martelli PL, editor. *Bioinformatics*. 2024;40(6):btae327. doi:10.1093/bioinformatics/btae327

40. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop*. 2019;477(6):1267–1279. doi:10.1097/CORR.0000000000000687
41. Jang SJ, Rosenstadt J, Lee E, Kunze KN. Artificial intelligence for clinically meaningful outcome prediction in orthopedic research: current applications and limitations. *Curr Rev Musculoskelet Med*. 2024;17(6):185–206. doi:10.1007/s12178-024-09893-z
42. Hunter J, Soleymani F, Viktor H, Michalowski W, Poitras S, Beaulé PE. Using unsupervised machine learning to predict quality of life after total knee arthroplasty. *J Arthroplasty*. 2023; 39:S0883540323009865.
43. Mercadal-Orfila G, Herrera-Pérez S, Piqué N, et al. Implementing systematic patient-reported measures for chronic conditions through the naveta value-based telemedicine initiative: observational retrospective multicenter study. *JMIR MHealth UHealth*. 2024;12:e56196. doi:10.2196/56196
44. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl*. 2019;134:93–101. doi:10.1016/j.eswa.2019.05.028
45. Cano G, Garcia-Rodriguez J, Garcia-Garcia A, et al. Automatic selection of molecular descriptors using random forest: application to drug discovery. *Expert Syst Appl*. 2017;72:151–159. doi:10.1016/j.eswa.2016.12.008
46. Chehreh Chelgani S, Matin SS, Hower JC. Explaining relationships between coke quality index and coal properties by Random Forest method. *Fuel*. 2016;182:754–760. doi:10.1016/j.fuel.2016.06.034
47. Naderi K, Kalami Yazdi MS, Jafarabadi H, Bahmanzadegan F, Ghaemi A, Mosavi MR. Modeling based on machine learning to investigate flue gas desulfurization performance by calcium silicate absorbent in a sand bed reactor. *Sci Rep*. 2024;14(1):954. doi:10.1038/s41598-024-51586-7
48. Abbas F, Zhang F, Ismail M, et al. Optimizing machine learning algorithms for landslide susceptibility mapping along the Karakoram Highway, Gilgit Baltistan, Pakistan: a comparative study of baseline, bayesian, and metaheuristic hyperparameter optimization techniques. *Sensors*. 2023;23 (15):6843. doi:10.3390/s23156843
49. Berlowitz DR, Foy CG, Kazis LE, et al. Effect of intensive blood-pressure treatment on patient-reported outcomes. *N Engl J Med*. 2017;377 (8):733–744. doi:10.1056/NEJMoa1611179
50. Bress AP, Bellows BK, King JB, et al. Cost-effectiveness of intensive versus standard blood-pressure control. *N Engl J Med*. 2017;377(8):745–755. doi:10.1056/NEJMsa1616035
51. Hendy A, Abdelaliem SMF, Sultan HM, et al. Unlocking insights: using machine learning to identify wasting and risk factors in Egyptian children under 5. *Nutrition*. 2025;131:112631. doi:10.1016/j.nut.2024.112631
52. Rouzrokh P, Khosravi B, Faghani S, et al. Mitigating bias in radiology machine learning: 1. *Data Handling Radiol Artif Intell*. 2022;4(5):e210290. doi:10.1148/ryai.210290
53. Verma D, Jansen D, Bach K, Poel M, Mork PJ, d'Hollosy WON. Exploratory application of machine learning methods on patient reported data in the development of supervised models for predicting outcomes. *BMC Med Inform Decis Mak*. 2022;22(1):227. doi:10.1186/s12911-022-01973-9
54. Wójcik Z, Dimitrova V, Warrington L, Velikova G, Absolom K. Using machine learning to predict unplanned hospital utilization and chemotherapy management from patient-reported outcome measures. *JCO Clin Cancer Inform*. 2024;8(8):e2300264. doi:10.1200/CCI.23.00264
55. Verma D, Bach K, Mork PJ. Application of machine learning methods on patient reported outcome measurements for predicting outcomes: a literature review. *Informatics*. 2021;8(3):56. doi:10.3390/informatics8030056
56. Roberts DR, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 2017;40(8):913–929. doi:10.1111/ecog.02881
57. Markatou M, Tian H, Biswas S, Hripcsak GM, Hripcsak G. Analysis of variance of cross-validation estimators of the generalization error. *AMIA. Annual Symposium Proceedings. AMIA Symposium*. 2005;2005:106–110.
58. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. *Adv Neural Inf Process Syst*. 2003;16.
59. Nic Giolla Easpaig B, Tran Y, Bierbaum M, et al. What are the attitudes of health professionals regarding patient reported outcome measures (PROMs) in oncology practice? A mixed-method synthesis of the qualitative evidence. *BMC Health Serv Res*. 2020;20(1):102. doi:10.1186/s12913-020-4939-7
60. Ziegenfuss JY, Grossman ES, Solberg LI, et al. Is the promise of proms being realized? Implementation experience in a large orthopedic practice. *Am J Med Qual*. 2022;37(6):489–494. doi:10.1097/JMQ.0000000000000079
61. Mercadal-Orfila G, López Sánchez P, Pou Alonso A, et al. TelePROM psoriasis: enhancing patient-centered care and patient reported outcomes measures follow-up in moderate to severe plaque psoriasis. *Front Med*. 2024;11:1465725. doi:10.3389/fmed.2024.1465725
62. Giolla Easpaig B N, Tran Y, Bierbaum M, Arnolda G, Delaney GP, Liauw W. Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowl Inf Syst*. 2023;65(1):31–57. doi:10.1007/s10115-022-01772-8