ENHANCING PATIENT-CENTERED CARE THROUGH PREDICTIVE MODELING OF PATIENT-REPORTED OUTCOMES IN HOSPITAL PHARMACY SETTING

[1,2]S. HERRERA-PÉREZ, [3,4]G. MERCADAL, [5]P. VENTAYOL, [6]J. SERRANO, [7]M.A. MAESTRE, [8]F. FERNANDEZ, [9]L. ANOZ, [10]F. MATEU (Spain)

1Universidad Internacional de Valencia. Facultad de Ciencias de la Salud, Valencia, Spain,2CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain,3Pharmacy Department, Hospital Mateu Orfila, Menorca, Spain ,4Department of Biochemistry and Molecular Biology, Universitat de les Illes Balears (UIB), Palma de Mallorca, Spain ,5Pharmacy Department, Hospital Universitari Son Espases, Palma de Mallorca, Spain ,6Pharmacy Department, Hospital Universitari Son Llàtzer, Palma de Mallorca, Spain ,7Pharmacy Department, Hospital de Manacor, Mallorca, Spain ,8Pharmacy Department, Hospital Comarcal d' Inca, Spain,9Farmacia Hospital Can Misses(Eivissa), 10Principal, Healthcare Industry Solutions at MongoDB - Digital Health & Innovation, Barcelona, Spain

Nº:6ER-013

Background and objectives

Background and importance:
In the era of value-based medicine, it is essential to focus on outcomes that matter to patients. In this context, Patient-reported Outcomes (PROs) have established themselves as key tools for measuring the real impact of medical interventions from the patient's perspective. However, to maximize their usefulness, it is crucial to anticipate and understand these outcomes. Machine learning is emerging as a powerful solution to accurately predict PROs and, consequently, optimize healthcare.

Aim and objectives:
This study presents a novel predictive model based on the Random Forest algorithm for the prediction of PRO scores from sociodemographic variables and medication knowledge obtained in hospital pharmacy practice.

Materials and Methods

Utilizing Python's robust libraries, our study's methodology entailed a rigorous, five-step process, beginning with pre-processing to cleanse and standardize the dataset. We integrated the refined data from the NAVETA dataset, which comprised data from more than 34000 questionnaire responses (from more than 3000 patients), and performed a multiple regression analysis to identify key variable relationships. Dimensionality reduction through PCA was conducted to focus on the most significant predictors, followed by the training of a Random Forest Regressor, optimized with Cross-Validation Grid Search. The model's success was ultimately quantified by calculating the probability that the standardized difference between the observed and predicted values was less than 0.25, a metric that underscores the accuracy and predictive quality of our model.

Results

Cohort of patients included in the ALGOPROMIA study

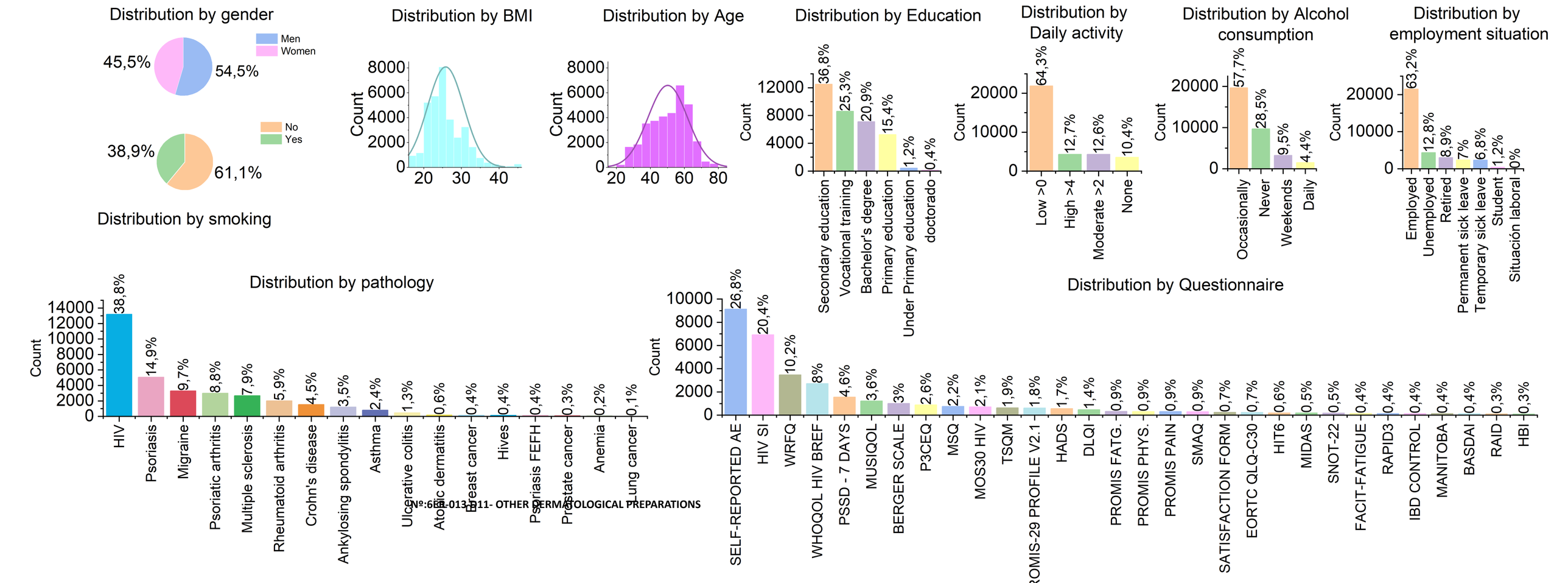

Fig. 1: Comprehensive Health and Lifestyle Data (questionnaire responses) Distributions. This figure presents a multi-faceted view of health-related data categorized by demographic and lifestyle factors. The upper panels display distributions by gender, body mass index (BMI), age, educational background, daily activity levels and alcohol consumption. The lower left panel illustrates the prevalence of various pathologies within the population sampled. The lower right panel offers insight into the distribution of responses to different PROM questionnaires (only those questionnaires with 100 or more responses have been considered). Together, these graphs provide a detailed view of the interplay between lifestyle choices and health outcomes in the studied cohort. Note: comma represents decimals.

Variable reduction using the Principal Component Analysis (PCA) method

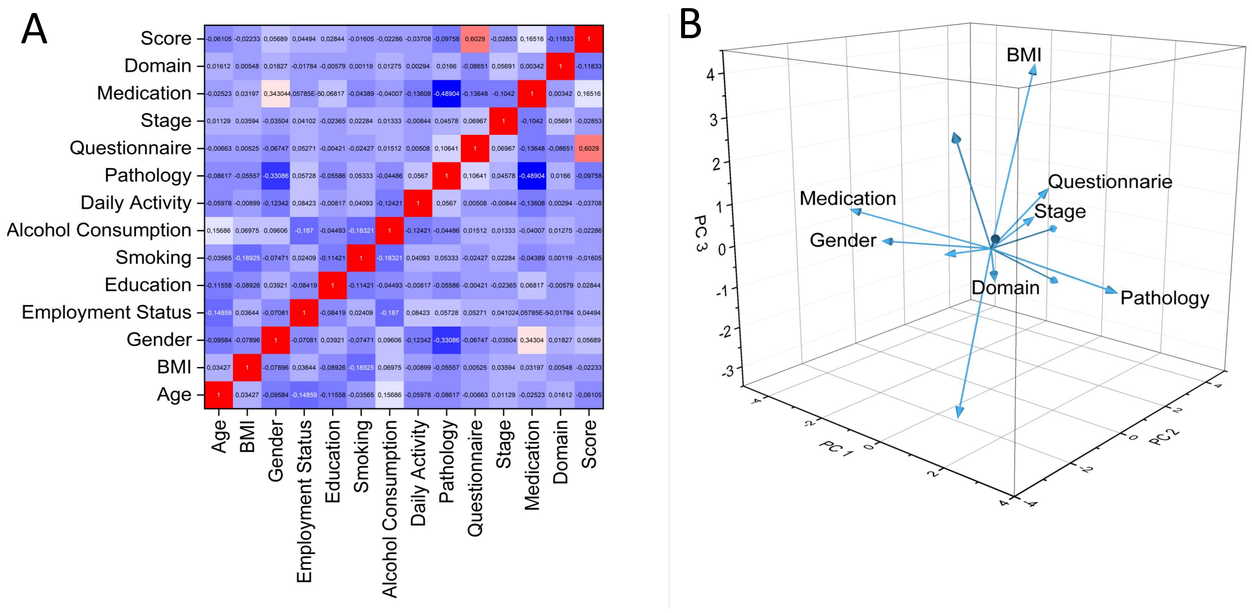

Fig. 2: Correlation Matrix of Lifestyle Factors and Health Indicators. **(A)** Heatmap illustrates the statistical correlations between various health indicators and lifestyle factors. Positive correlations are shown in red, indicating a direct relationship, while negative correlations are displayed in blue, indicating an inverse relationship. **(B)** Three-dimensional biplot of a PCA where only those variables with significant ($p<0.05$) weight in the principal component analysis are highlighted. The vectors show the relative contribution of each significant variable (such as BMI, Medication, Gender) to the principal components (PC1, PC2, PC3). The direction and magnitude of the vectors suggest the influence of these variables on the principal components, illustrating the key factors characterizing the data structure. Note: comma represents decimals.

Model training and validation

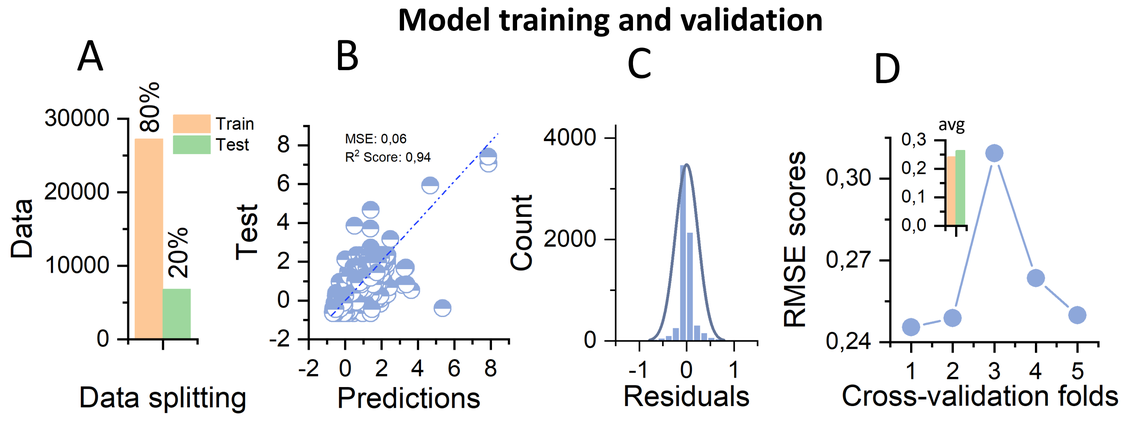

Fig. 3: Comprehensive evaluation of a Random Forest model's performance. **(A)** Data Splitting: The data is divided into training (80%) and testing (20%) sets using balanced method. **(B)** A scatterplot displays the model's predictions compared to the actual test data. The Mean Squared Error (MSE) is 0.06, and the R^2 score is 0.94, indicating high accuracy. **(C)** A histogram illustrates the distribution of residuals, which appear to be normally distributed around zero, suggesting unbiased predictions. **(D)** A plot shows the Root Mean Squared Error (RMSE) scores across five cross-validation folds, giving insights into the model's stability across different data subsets. Note: comma represents decimals.

Success rate in questionnaire prediction

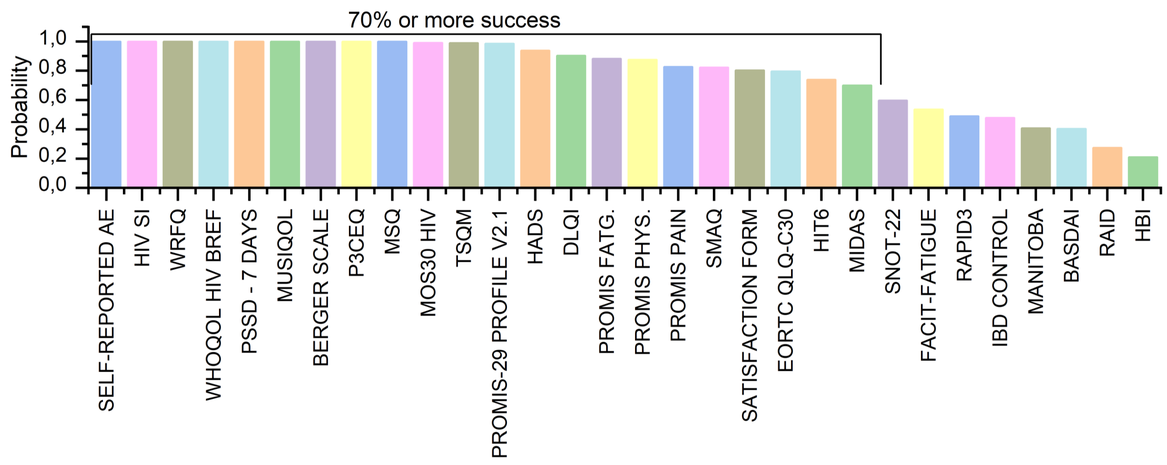

Fig. 4: Predictive success rates of questionnaire values using a Random Forest algorithm. The graph shows the probability that the standardized difference is less than or equal to 0.25, indicating a 70% or more success rate in prediction across various health-related quality of life measures. The scores of 22 out of 30 questions can be predicted with our model. Note: comma represents decimals.

Conclusion and relevance

The results indicate that the Hospital Pharmacy records obtained from the NAVETA cohort significantly predict patient health outcomes. The use of this predictive model in telemedicine systems such as NAVETA would improve patient care by helping to quickly identify needs and tailor treatments, leading to accurate, patient-centered care in the context of hospital pharmacy. Scores from over 73% of the questionnaires are susceptible to being predicted using our model.

Acknowledgments

We extend our gratitude to all the patients who have graciously allowed the inclusion of their anonymous data, as well as to the healthcare professionals who have contributed to the ALGOPROMIA project