scientific reports

Check for updates

OPEN

Machine learning using PROFUND components for 30-day readmission prediction in multimorbid patients: a prospective multicentre study

Amaia Pikatza-Huerga[1,2]✉, Aitor Almeida[1,2], Raúl Quirós[2,3], María José Legarreta[2,4,5], Unai Zulaika[1], Daniela Mestre[2,5,6] & Susana García-Gutiérrez[2,4,5,7]

Early hospital readmission in multimorbid patients remains a major clinical challenge. Although risk stratification tools are widely used, predictive performance is often limited. The PROFUND index captures frailty, functional dependence, and social vulnerability, but its role in predicting 30-day readmission is unclear. In this prospective multicentre cohort study, multimorbid patients admitted to Internal Medicine and Geriatrics departments were followed after discharge. The primary outcome was unplanned 30-day readmission among patients surviving to 30 days. Models based on PROFUND components were developed using logistic regression and gradient boosting, including a calibrated ensemble model, and compared with LACE and HOSPITAL scores. Performance was assessed in an external validation cohort. Among 435 patients included in the readmission analysis, 14% were readmitted within 30 days. In external validation, discrimination remained modest (AUC 0.52–0.59). The ensemble XGBoost model achieved the highest AUC (0.59), followed by XGBoost (0.58), HOSPITAL (0.54), and LACE (0.52). Differences were incremental. SHAP analysis identified cognitive impairment, anaemia, advanced age, heart failure severity, functional dependence, and limited caregiver support as key contributors. Incorporating frailty, functional, and social vulnerability domains through PROFUND components resulted in only modest improvements in 30-day readmission prediction. Even with machine learning, discrimination remained limited. The observed performance likely reflects both the intrinsic complexity of short-term readmission and the constraints imposed by sample size and available predictors.

Multimorbidity, also referred to as multiple chronic conditions (MCC), is a widely used term to describe patients with multiple coexisting chronic diseases. In Spain, these patients are often classified as pluripathological according to Ollero's criteria. It affects a significant proportion of adults and increases markedly with age. It drives greater use of resources, hospital admissions, and referrals to specialists[1]. Beyond clinical complexity, the social context is also a determining factor; the social determinants of health (SDOH), such as socioeconomic status, education, social support and access to effective care, influence disease control, continuity of care and the risk of avoidable readmissions[2]. In a scenario of population ageing, identifying which patients are at risk of early readmission and directing care transition interventions towards them is a clinical and economic priority.

The poor performance of readmission models is a significant concern in healthcare, as these models are intended to predict which patients are at risk of being readmitted shortly after discharge. Despite numerous attempts to enhance predictive accuracy, many readmission models have demonstrated limited effectiveness, particularly in external validation settings, often failing to account for critical social and behavioural factors[2].

[1]Faculty of Engineering, University of Deusto, Av. de las Universidades, 24, Deusto, Bilbao, Bizkaia E-48007, Spain. [2]Network for Research on Chronicity, Primary Care, and Health Prevention and Promotion (RICAPPS), Galdakao, Bilbao, Bizkaia, Spain. [3]Internal Medicine Department, Hospital Costa del Sol, Marbella, Málaga 29603, Spain. [4]Research Unit, Osakidetza Basque Health Service, Galdakao-Usansolo University Hospital, Galdakao, Spain. [5]Biosistemak Institute for Health Systems Research, Bilbao, Bizkaia, Spain. [6]Osakidetza Basque Health Service, Research and Innovation Unit, Basurto University Hospital, Bilbao, Spain. [7]Faculty of Health Sciences, University of Deusto, Av. de las Universidades, 24, Deusto, Bilbao, Bizkaia E-48007, Spain. ✉email: a.pikatza@deusto.es

Apart from that, many readmission models exhibit poor discriminative ability, indicating limited effectiveness in accurately predicting readmissions[2,3]. Reported discrimination varies widely depending on the population and setting, with c-statistics typically ranging from modest values (~0.55–0.70) in external validation studies, while higher values (up to ~0.80) are occasionally reported in development cohorts. Often rely on retrospective administrative data, which may not capture the full spectrum of patient health, including social determinants and functional status[2]. The performance of readmission models can degrade over time, necessitating continuous training and adjustment to reflect current patient populations and trends[4]. These limitations suggest that current models may not adequately capture the multidimensional vulnerability of multimorbid patients, particularly domains related to frailty, functional dependence, and social context, highlighting the need for alternative approaches.

Traditionally, prediction models for 30-day readmission in patients with multimorbidity have been based on classical statistical approaches, such as logistic regression or Cox models. The LACE index and HOSPITAL score are among the most widely used tools for predicting 30-day hospital readmissions in clinical practice. The LACE index incorporates length of stay, acuity of admission, comorbidity, and emergency department use, while the HOSPITAL score includes laboratory parameters and prior healthcare utilisation. Both have been externally validated in multiple populations, although their predictive performance varies considerably depending on patient characteristics and healthcare settings. Despite their utility, the predictive accuracy of both indices can vary significantly based on patient demographics and clinical settings, suggesting a need for further refinement and validation in diverse healthcare environments. The PROFUND index is a predictive tool designed to assess mortality risk in polypathological patients, demonstrating strong predictive validity for both short-term and long-term outcomes[5]. In Spain, the PROFUND index is routinely used in selected clinical settings, particularly in Internal Medicine, to stratify mortality risk in multimorbid patients. It is calculated using a set of clinical, functional, and social variables collected at discharge and has been incorporated into clinical decision-making processes in several hospitals.

In recent years, machine learning (ML) techniques have been increasingly applied to readmission prediction, offering the potential to capture non-linear relationships and complex interactions between clinical, functional, and social variables. The literature in multimorbid populations reflects a maturing field, with tree-based methods and neural networks playing central roles, alongside growing attention to interpretability and real-world implementation[6,7]. Importantly, there is increasing recognition that dimensions such as frailty, functional status, and social vulnerability can enhance predictive accuracy and support more targeted interventions. Evidence across diverse settings—including heart failure, postoperative populations, and geriatric cohorts—consistently shows that multidimensional frailty constructs add incremental value beyond traditional comorbidity measures, while functional status and social determinants of health further refine risk stratification in older, multimorbid patients[8–18].

However, despite these advances, the added value of ML over traditional statistical models remains uncertain, particularly in external validation settings, where improvements in discrimination are often modest and calibration issues may arise[19]. Moreover, many existing models fail to incorporate variables capturing the intensity, continuity, or quality of care during the index hospitalisation—factors that may significantly influence subsequent readmission risk. This omission may partly explain the modest predictive performance observed across modelling approaches, as current models often rely heavily on structured clinical variables while overlooking critical aspects of care delivery and continuity[20–22].

The primary objective of this study was to evaluate whether the individual components of the PROFUND index, capturing frailty, functional dependence, and social vulnerability, improve the prediction of 30-day hospital readmission in multimorbid patients. In addition, we aimed to assess whether machine learning approaches meaningfully enhance predictive performance compared with established clinical risk scores. A preliminary conceptual reflection on the clinical implications of readmission prediction in multimorbid patients was previously discussed in a brief correspondence[23]. The present study provides the full methodological development and external validation of the proposed approach.

Methods

Data collection and participants

1) The RePluris project (NCT05574790) was a prospective multicentre study conducted in five public hospitals in Spain (the Basque Country, Catalonia and Andalusia), whose objective is to identify the factors associated with hospital readmission within 30 days in patients with multimorbidity (multiple pathologies). Adult patients hospitalised for decompensation of chronic diseases and classified as multimorbid according to Ollero's criteria[24], which define multimorbidity based on the coexistence of specific chronic disease categories with clinical complexity. Exclusion criteria were operationalised prospectively at the time of inclusion by trained clinical staff. Patients considered to be at the end of life were identified based on clinical judgement supported by established criteria (e.g., advanced disease with limited life expectancy and palliative care orientation). Scheduled admissions were defined as pre-planned hospitalisations for elective procedures, and admissions unrelated to chronic disease were excluded based on primary diagnosis at admission. These criteria were applied consistently across participating centres. Data were collected prospectively between 2023 and 2024 in the Internal Medicine and Geriatrics departments of the participating centres.
2) External validation data were collected from the electronic health records of Galdakao- Usansolo Hospital. Data were internally stored and managed by the Osakidetza-Basque Public Health System. After anonymisation and removal of protected health information, the data were released in a text-delimited format for research purposes. Patient-level data were collected for the initial analyses in our study. We identified patients

admitted to the Internal Medicine department and defined as multimorbid from 2022 to 2023 with the same eligibility criteria as the RePluris project.

Outcome

The primary outcome was unplanned hospital readmission within 30 days after discharge. As death precludes the occurrence of readmission, the readmission analyses were performed in patients who were alive at 30 days (i.e., conditional on 30-day survival). We report the number of 30-day deaths in the study flow diagram and acknowledge this competing event as a limitation.

Variables

Predictor variables were grouped into the following domains:

- Clinical characteristics: including comorbidities, laboratory parameters, and clinical conditions (e.g., anaemia, heart failure, oncological disease), as well as indices such as the Charlson Comorbidity Index.
- Functional status: assessed using the Barthel Index, which was recorded at discharge and operationalised as a categorical variable reflecting levels of dependency.
- Social and caregiver-related factors: including social support (Gijón scale) and caregiver burden assessed using the Zarit Burden Interview (ZBI).
- Patient-reported outcomes: including quality of life measured by the EQ-5D-5 L questionnaire and treatment adherence assessed using the ARMS scale.
- Frailty and prognostic indices: including the PROFUND index and Frail-VIG index.

All variables were collected using validated instruments and were encoded numerically for modelling. Continuous variables were standardised within the preprocessing pipeline as described below.

Where applicable, variables were used according to their validated scoring systems. For example, the Barthel Index was recorded using its standard scale and subsequently categorised for analysis, and questionnaire-based instruments (e.g., EQ-5D-5 L, ZBI) were processed according to their established scoring guidelines.

All predictor data were collected prospectively at or shortly before discharge by trained healthcare professionals involved in the study (physicians and/or research nurses), using standardised data collection protocols and validated instruments. Data collection procedures were harmonised across centres to ensure consistency[25].

Sample size

This study was not prospectively designed using a formal sample size calculation for prediction model development or external validation. We acknowledge that recent methodological guidance recommends sample size planning based on the expected precision of model performance, calibration, and clinical utility, rather than rule-of-thumb criteria such as events-per-variable[26]. Therefore, the present analyses should be interpreted as an evaluation of model performance in a prospective multicentre cohort rather than definitive model development.

Analytical methods

Data preprocessing

The cohort was divided into a training set (80%) and a validation set (20%), maintaining the proportion of patients with and without readmission. All preprocessing steps, including missing data imputation, feature scaling, and any feature selection procedures, were implemented within a unified pipeline.

To avoid information leakage, these transformations were fitted exclusively on the training data within each cross-validation fold and subsequently applied to the corresponding validation fold. The same fitted transformations were then applied to the test set and external validation cohort without re-fitting.

All numerical variables were scaled using a robust scaling approach, which is less sensitive to outliers and skewed distributions compared to standard normalisation methods. This choice was made to improve model stability given the characteristics of clinical data.

Feature selection and reduction

The large number of clinical, functional, and social variables makes it necessary to discriminate which ones provide predictive value. To do this, different approaches were compared:

- Principal Component Analysis (PCA): transforms the original variables into uncorrelated components that capture as much variance as possible, reducing noise and redundancy.
- SelectKBest: identifies the variables with the strongest statistical association with the outcome, prioritising those that are most informative.
- Clinical subsets: selected by experts based on clinical relevance and pathophysiological plausibility.
- No reduction: control scenario to assess whether the elimination of variables compromises predictive performance.

All feature selection and dimensionality reduction techniques (including PCA and SelectKBest) were incorporated within the cross-validation framework to ensure that variable selection was performed exclusively on training folds, thereby preventing optimisation bias.

These approaches were selected to explore different strategies for balancing dimensionality reduction, interpretability, and predictive performance, given the high dimensionality and heterogeneity of clinical, functional, and social variables. No single method was assumed to be optimal a priori.

Statistical and machine learning models

Three modelling approaches were selected for primary evaluation: (a) logistic regression as an interpretable baseline model, (b) gradient boosting (XGBoost) to account for potential non-linear relationships and complex interactions, and (c) a calibrated ensemble based on XGBoost to improve probability estimation. Model performance was compared with established clinical risk scores (LACE and HOSPITAL), which are widely used for readmission prediction. The PROFUND index itself was not used as a comparator model, as it was originally developed for mortality prediction; instead, its individual components were used as predictors in the proposed models.

Additional exploratory analyses using alternative feature selection methods, dimensionality reduction techniques, resampling strategies, and machine learning algorithms were conducted as part of an initial model screening phase and are reported in the Supplementary Methods.

Hyperparameter optimisation

Each algorithm requires the configuration of internal parameters. GridSearchCV was used to identify the optimal combinations to maximise performance and minimise overfitting. For XGBoost-based models, tuning focused on parameters such as the number of trees (n_estimators), tree depth (max_depth), and learning rate (learning_rate). For logistic regression, tuning focused on regularisation-related parameters, including the regularisation strength (C) and penalty type where applicable.

For the ensemble track, hyperparameters were optimised using RandomizedSearchCV with stratified five-fold cross-validation and ROC AUC as the scoring metric. The best estimator was subsequently calibrated via isotonic regression. The full hyperparameter search space, optimisation procedure, and selected parameter values for the final models are reported in the Supplementary Methods and Supplementary Table S2.

Handling missing data and class imbalance

Missing data were handled using K-nearest neighbours imputation (KNNImputer), which demonstrated improved performance in preliminary analyses compared to simpler imputation and iterative strategies. Imputation was performed within the cross-validation folds as part of the preprocessing pipeline to prevent information leakage.

Class imbalance was addressed either through internal class weighting or resampling strategies, depending on the modelling approach.

Model evaluation

Performance was evaluated using standard binary classification metrics, including sensitivity, specificity, precision, F1 score, area under the receiver operating characteristic curve (AUC-ROC), and area under the precision–recall curve (PR-AUC), all reported with 95% confidence intervals. Confidence intervals for performance metrics were estimated using bootstrapping techniques, providing an empirical measure of variability across resampled datasets. Calibration was assessed using calibration plots and comparison of predicted versus observed risk across probability ranges. All performance metrics were computed on data not used during model fitting or preprocessing, ensuring an unbiased estimation of model performance.

Competing risk analysis

Given the non-negligible proportion of patients who died within 30 days, additional analyses were performed to account for competing risk. A composite outcome combining 30-day readmission and all-cause mortality was defined. All models were retrained using this composite endpoint, applying the same preprocessing pipeline, feature selection strategies, hyperparameter optimisation procedures, and validation framework as in the primary analysis.

This approach was used as a pragmatic strategy to assess the potential impact of competing risk on model performance, while maintaining consistency with the original modelling framework.

Fairness and calibration

In addition to discrimination, model calibration was analysed, verifying that the predicted probabilities accurately reflected the observed readmission rates. Optimal decision thresholds were adjusted beyond the conventional value of 0.5; they were determined from out-of-fold predictions in the training set (to avoid data leakage) using Youden's index, maximum F1 score, and sensitivity-oriented optimisation; the Youden threshold was pre-specified for external evaluation. The objective was to prioritise the detection of patients at risk of readmission, reducing clinically relevant false negatives.

External validation

Variables in the external validation cohort were defined and operationalised using the same criteria and coding as in the development cohort to ensure comparability of predictors across datasets. The final model was first categorised using the optimal decision threshold to be applied in an external cohort of patients. LACE and Hospital were also applied in our sample. The available bibliography was used to determine the optimal cut-off points[27,28]. Precision–recall and calibration plots were also assessed in the external cohort to evaluate discrimination and reliability. For LACE and HOSPITAL, scores were min–max scaled to the [0,1] range to allow comparison with probabilistic model outputs; therefore, their calibration should be interpreted with caution.

Ethical approval

The research was conducted in accordance with the principles outlined in the Declaration of Helsinki (Br Med J 1964; ii: 177). The study was approved by the Basque Country Ethics Committee (Reference Number: PI2018191). Written informed consent was obtained from patients or their legal representatives.

Software and libraries

The remaining analyses were conducted using Python within a virtual environment configured for the project. The core ML tasks were implemented using scikit-learn (version 1.0.2). To address class imbalance in the dataset, the imbalanced-learn library (version 0.8.1) was employed. Additionally, XGBoost (version 1.6.2) was included for gradient boosting-based models. All packages were installed and managed in a local virtual environment, and dependencies such as NumPy, SciPy, and joblib were also utilised as required by the main libraries. All analyses were conducted using reproducible scripts with fixed random seeds to ensure deterministic results across runs.

Results

Participants

A total of 1,003 patients were initially evaluated for inclusion in the study. Of these, 468 patients were excluded due to not meeting the inclusion criteria, refusal to participate, or incomplete data (e.g., missing key variables required for model development). A detailed breakdown of exclusions is provided in Fig. 1. Therefore, the final sample included 485 patients. Because the study outcome was 30-day readmission, analyses were restricted to patients who were alive at 30 days after discharge. Therefore, 50 patients who died within 30 days were excluded from the readmission analyses, and the final analytical sample consisted of 435 patients.

The cohort was divided as follows: 388 patients were assigned to the training cohort, of whom 56 experienced readmissions (14%); and 97 patients formed the test cohort, with 14 readmissions recorded(14%). The external cohort was composed of 173 patients; 22% were readmitted.

There were no clinically important differences between the main predictors in the train and test sets, except for the previous diagnosis of heart failure, which was found to be more frequent in the train set (168 cases, 43.3%) compared to the test set (29 cases, 29.9%)This reflects a moderate imbalance in heart failure history between the two sets, which should be considered in interpretation and potential model recalibration steps. On the other hand, patients in the external validation cohort present a greater number of comorbidities (see Table 1).

Internal validation

During internal validation, multiple modelling strategies and predictor sets were evaluated (see Table 2). Models based on PROFUND components showed modest performance across different algorithms. The inclusion of additional frailty measures (PROFUND + Frail-VIG) did not result in a substantial or consistent improvement in predictive performance across metrics. As a result, and to ensure consistency with the external validation cohort where Frail-VIG was not available, subsequent analyses focused on models based on PROFUND components.

Overall, differences in performance between models were modest, with no single approach clearly outperforming the others across all metrics.

XGBoost and the calibrated ensemble XGBoost were selected for external validation as they showed stable performance across metrics, good calibration properties, and compatibility with SHAP-based interpretability. This choice reflects a pragmatic selection based on overall model behaviour rather than clear superiority in discrimination.

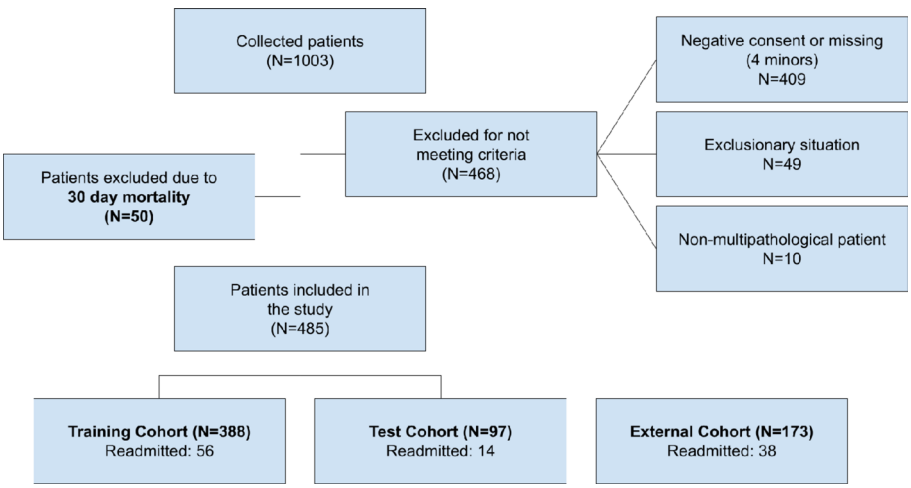

Fig. 1. Flow chart of the participants.

Characteristic	Total ($N=485$)	Train ($N=388$)	Test ($N=97$)	p-value	External Validation ($N=173$)	p-value
Demographic Characteristics						
Age, years (Mean ± SD)	81.9 ± 9.9	81.8 ± 9.7	82.6 ± 10.8	0.4812	86.1 ± 6.5	< 0.0001
Male sex, n (%)	255 (52.6)	208 (53.6)	47 (48.5)	0.3632	96 (55.5)	0.5095
Main Comorbidities						
Heart failure, n (%)	253 (52.2)	207 (53.4)	46 (47.4)	0.2959	40 (23.1)	< 0.0001
COPD, n (%)	124 (25.6)	101 (26.0)	23 (23.7)	0.6395	119 (68.8)	< 0.0001
Diabetes mellitus, n (%)	234 (48.2)	190 (49.0)	44 (45.4)	0.5247	102 (58.9)	0.0155
Chronic kidney disease, n (%)	239 (49.3)	187 (48.2)	52 (53.6)	0.3403	113 (65.3)	0.0003
Dementia/cognitive impairment, n (%)	125 (25.8)	95 (24.5)	30 (30.9)	0.1944	56 (32.4)	0.0953
Active neoplasia, n (%)	126 (26.0)	103 (26.5)	23 (23.7)	0.5690	43 (24.9)	0.7714
Number of chronic conditions Mean ± SD	2.3 ± 1.2	2.3 ± 1.1	2.2 ± 1.3	0.8309	2.7 ± 1.2	< 0.0001
Functional Status						
Barthel Index				0.2748		< 0.0001
No dependency (> = 95)	138 (28.45)	113 (29.12)	25 (25.77)		19 (11.0)	
· Mild-moderate dependency(90 – 65)	121 (24.95)	102 (26.29)	19 (19.59)		76 (43.9)	
· Moderate-severe dependency(60 – 25)	143 (29.48)	107 (27.58)	36 (37.11)		48 (27.8)	
· Severe dependency (< = 20)	66 (13.61)	51 (13.14)	15 (15.46)		21 (12.1)	

Table 1. Characteristics of the patient in different datasets:.

Model	Predictors	ROC AUC (95% CI)	Precision (95% CI)	Recall (95% CI)	F1 Score (95% CI)
KNN	PROFUND	0.67 (0.58–0.76)	0.21 (0.13–0.29)	0.93 (0.88–0.98)	0.34 (0.25–0.44)
Decision Tree	PROFUND	0.63 (0.53–0.73)	0.38 (0.29–0.48)	0.36 (0.26–0.45)	0.37 (0.27–0.47)
SVC	PROFUND	0.66 (0.57–0.76)	0.21 (0.13–0.30)	0.86 (0.79–0.93)	0.34 (0.25–0.44)
XGBoost	PROFUND	0.68 (0.58–0.77)	0.25 (0.16–0.34)	0.71 (0.62–0.80)	0.37 (0.27–0.47)
Ensemble XGBoost	PROFUND	0.67 (0.49–0.82)	0.15 (0.09–0.24)	0.93 (0.78–1.00)	0.26 (0.15–0.38)
KNN	PROFUND + Frail-VIG	0.69 (0.60–0.78)	0.24 (0.16–0.33)	0.79 (0.70–0.87)	0.37 (0.28–0.47)
Decision Tree	PROFUND + Frail-VIG	0.66 (0.57–0.76)	0.26 (0.17–0.34)	0.64 (0.55–0.74)	0.37 (0.27–0.46)
SVC	PROFUND + Frail-VIG	0.61 (0.51–0.71)	0.20 (0.12–0.28)	0.64 (0.55–0.74)	0.31 (0.22–0.40)
XGBoost	PROFUND + Frail-VIG	0.69 (0.60–0.78)	0.24 (0.16–0.33)	0.79 (0.70–0.87)	0.37 (0.28–0.47)

Table 2. Performance of different models assessed in the test set. Abbreviations. ROC AUC: Area under the receiver operating characteristic curve; Prec.: Precision; F1 Score: Harmonic mean of precision and recall.

Model	ROC AUC (95% CI)	Precision (95% CI)	Recall (95% CI)	F1 Score (95% CI)
LACE (> 10)	0.52 (0.47–0.56)	0.22 (0.16–0.29)	**0.95 (0.87–1.00)**	0.36 (0.27–0.44)
HOSPITAL (> 4)	0.54 (0.45–0.63)	0.24 (0.16–0.34)	0.54 (0.38–0.71)	0.34 (0.23–0.44)
XGB (threshold = 0.51)	0.58 (0.47–0.68)	**0.32 (0.19–0.45)**	0.43 (0.27–0.59)	0.37 (0.23–0.49)
Ensemble XGB (threshold = 0.11)	**0.59 (0.48–0.70)**	0.27 (0.18–0.36)	0.68 (0.52–0.82)	**0.38 (0.27–0.48)**

Table 3. Performance of final models in validation cohort. Abbreviations. AUC: Area under the receiver operating characteristic curve; XGB: XGBoost; Threshold: Probability cut-off used for classification.

External validation

In the external validation cohort, discrimination remained modest across all approaches (AUC range 0.52–0.59). The ensemble XGBoost model achieved the highest AUC (0.59), followed by XGBoost (0.58), HOSPITAL (0.54), and LACE (0.52). Differences between models were incremental rather than substantial (see Table 3).

Figure 2 summarises the discrimination ability of the LACE, HOSPITAL, XGBoost, and Ensemble XGBoost models in the external validation cohort through receiver operating characteristic (ROC) and precision–recall (PR) curves. Machine learning–based models demonstrated numerically higher discrimination compared with traditional scores, although absolute differences were modest. The Ensemble XGBoost achieved the highest area under the ROC curve (AUC = 0.59), followed by XGBoost (AUC = 0.58), HOSPITAL (AUC = 0.54), and LACE (AUC = 0.52). The confidence interval for the AUC of the best-performing model included 0.5, indicating that discrimination was only marginally better than chance in the external validation cohort. A similar pattern was

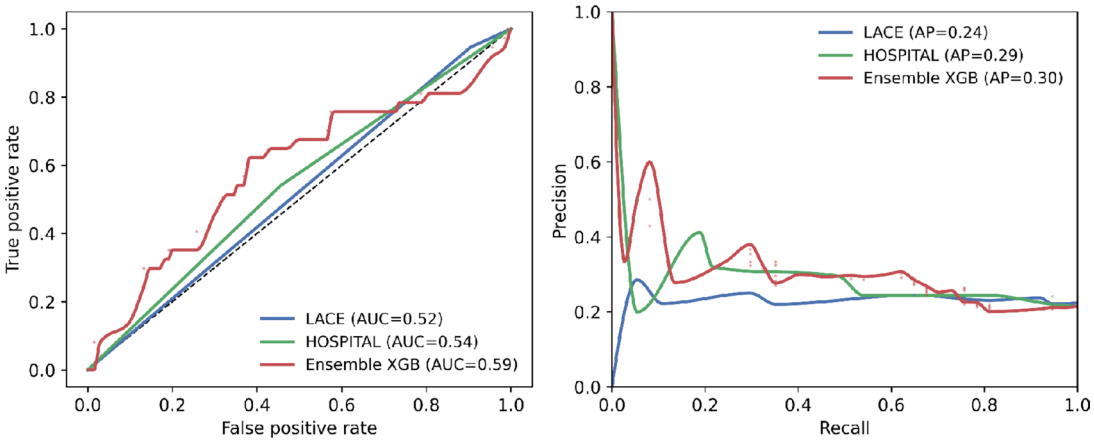

Fig. 2. ROC and PR curves for LACE, HOSPITAL, and the XGBoost ensemble in the validation cohort.

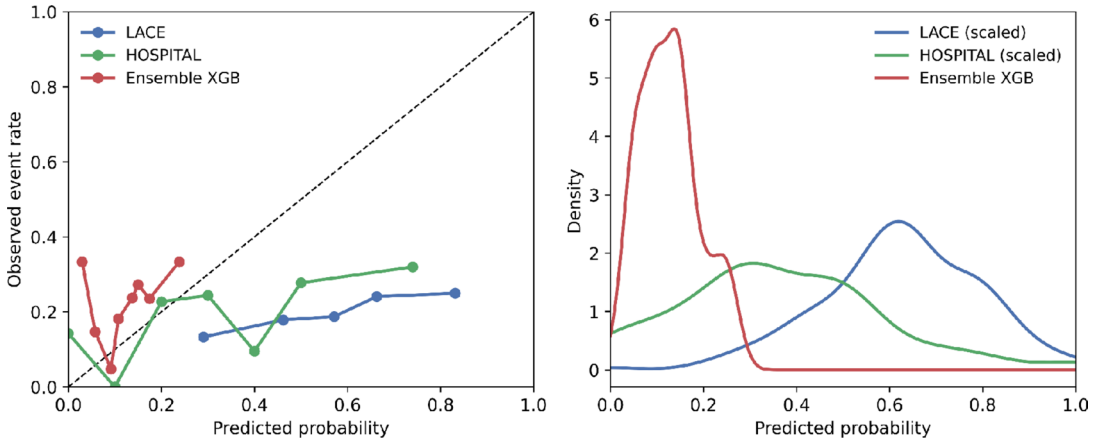

Fig. 3. Calibration curves and distribution of predicted probabilities for LACE, HOSPITAL, and Ensemble XGBoost models in the external validation cohort.

observed in the precision–recall space, where the Ensemble XGBoost obtained the largest area under the PR curve (AP = 0.30), followed by HOSPITAL (AP = 0.29), XGBoost (AP = 0.26), and LACE (AP = 0.24). Together, these results indicate numerically higher discrimination when using the ensemble model compared with the traditional indices.

Figure 3 depicts the calibration performance and the distribution of predicted probabilities across models. Calibration curves for LACE, HOSPITAL, and Ensemble XGBoost are shown, allowing direct comparison of predicted and observed event rates. The Ensemble XGBoost curve is closest to the ideal diagonal, indicating better agreement between predicted probabilities and observed readmission rates. In contrast, LACE and HOSPITAL show greater variability, reflecting less consistent calibration, which may be partly explained by the use of rescaled scores rather than true predicted probabilities. The histogram of predicted probabilities reveals that the Ensemble XGBoost concentrates most predictions in lower probability ranges (approximately 0.05–0.25), which is typical of well-calibrated models when the positive event is infrequent. This distribution suggests that the ensemble model effectively captures the class imbalance present in the dataset and provides more realistic probability estimates compared with traditional and single-model approaches.

When models were retrained using the composite outcome of 30-day readmission or all-cause mortality, overall performance remained modest and comparable to the primary analysis. Discrimination metrics showed no meaningful improvement across models, and differences between modelling approaches remained minimal. These findings suggest that including mortality as a competing event does not substantially improve predictive performance in this setting. Importantly, the relative ranking of models remained unchanged, with ensemble-based approaches showing only incremental improvements over logistic regression.

To better understand the contribution of individual PROFUND items to the model's predictions, SHAP (SHapley Additive exPlanations) values were computed for the Ensemble XGBoost model trained on the final model based on PROFUND components. The global SHAP summary plot (see Fig. 4) illustrates the distribution and direction of impact of each variable on the predicted risk of 30-day readmission. To facilitate

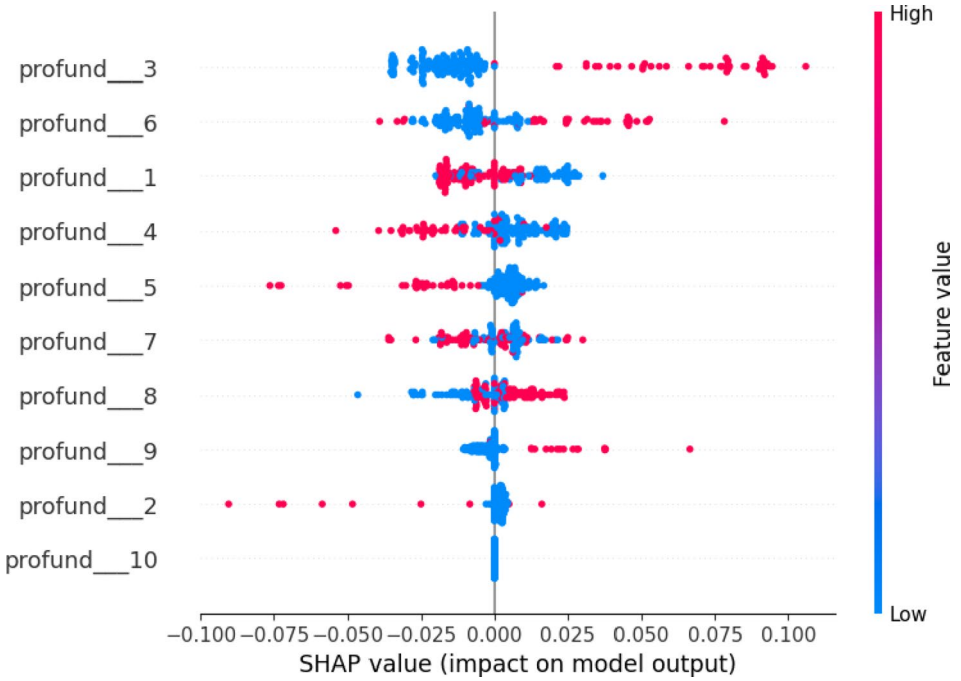

Fig. 4. Shap values for the final model.

the interpretation of SHAP values, the variables included in the Profund-based model and their corresponding clinical meanings are detailed next.

The variables used in the SHAP analysis correspond to the individual components of the PROFUND index as operationalised in this study. Although the original PROFUND score is typically used as a composite index, we modelled its individual components separately to assess their independent contribution to readmission risk.

- **profund___1**: Age > 85 years.
- **profund___2**: Active neoplasia.
- **profund___3**: Dementia.
- **profund___4**: NYHA class III–IV heart failure.
- **profund___5**: Delirium.
- **profund___6**: Hemoglobin < 10 g/dL (latest laboratory value at admission).
- **profund___7**: Barthel Index < 60 (at hospital discharge).
- **profund___8**: No caregiver or caregiver other than spouse.
- **profund___9**: ≥ 4 hospital admissions in the past 12 months (including current admission).
- **profund___10**: Indicator of non-computable PROFUND score due to missing data.

Figure 4 shows the SHAP (SHapley Additive Explanations) values, which reveal how much each variable in the PROFUND index contributes to the model's predictions. The SHAP summary plot revealed that the most influential predictors of the model were dementia (profund___3), hemoglobin < 10 g/dL at admission (profund___6), and age > 85 years (profund___1). For these variables, the presence of the condition ("Yes") was associated with positive SHAP values, indicating an increased probability of the adverse outcome, whereas the absence ("No") was linked to negative SHAP values, reducing the predicted risk. Other factors, such as NYHA class III–IV heart failure (profund___4), delirium (profund___5), Barthel Index < 60 at discharge (profund___7), lack of a caregiver or caregiver other than spouse (profund___8), and ≥ 4 hospital admissions in the past 12 months (profund___9), showed smaller or more variable impacts on the model output. The variable profund___10 does not represent a clinical feature but rather a flag indicating that the PROFUND score could not be calculated due to missing data. It was therefore not used as a predictor in the modelling process and does not contribute to model interpretation.

Discussion

In this prospective multicentre study of multimorbid patients, we evaluated whether variables capturing key domains of vulnerability, specifically frailty (e.g., advanced age, cognitive impairment, heart failure severity), functional dependence (e.g., Barthel Index), and social vulnerability (e.g., caregiver availability), as represented by the individual components of the PROFUND index, could improve the prediction of 30-day hospital readmission. Although these variables are clinically meaningful and conceptually linked to post-discharge vulnerability, overall discrimination remained modest, even when combined with machine learning techniques and externally validated. These findings underscore the intrinsic difficulty of predicting early readmission in

multimorbid older adults, a population in whom rehospitalisation likely reflects a complex interaction between baseline vulnerability, acute decompensation, and post-discharge care processes.

The SHAP analysis provided clinically coherent insights. Cognitive impairment, severe anaemia at admission, advanced age, and advanced heart failure emerged as the most influential contributors to readmission risk. Rather than representing isolated risk factors, these variables reflect multidimensional vulnerability: impaired self-management capacity, reduced physiological reserve, and unstable chronic disease trajectories.

These findings are consistent with previous studies highlighting the role of frailty, cognitive impairment, anaemia, and functional decline as key determinants of adverse post-discharge outcomes in older multimorbid populations[29–32]. In particular, frailty and functional dependence have been repeatedly associated with increased healthcare utilisation and readmission risk, while social factors such as caregiver availability and social support are increasingly recognised as critical determinants of transitional care success.

Functional dependence at discharge and absence of a spousal caregiver further emphasise the importance of post-discharge support structures[33–36]. These findings align with evidence highlighting the role of frailty progression, social isolation, and unmet care needs in adverse post-discharge outcomes. However, despite the conceptual relevance of these domains, their predictive contribution remained insufficient to achieve strong discrimination at the individual level.

Lack of a caregiver or having a non-spousal caregiver also exerted positive effects on model output when present[35]. This social determinant of health may serve as an indicator of social isolation or loneliness, which has been consistently demonstrated as a strong predictor of adverse health outcomes[36]. The prominence of this variable highlights that readmission risk extends beyond purely medical factors to encompass critical social support structures.

In contrast, frequent hospital use, defined as four or more admissions in the past 12 months[37] and active neoplasia[38], showed comparatively smaller and more heterogeneous contributions, with SHAP values clustering closer to zero. This suggests that while these conditions are clinically significant, their marginal impact on discriminating between patients at risk of 30-day readmission was less pronounced in this dataset compared to functional, cognitive, and social vulnerability indicators. This finding may reflect the fact that oncological patients often have specialised follow-up pathways that mitigate readmission risk, while frequent prior hospitalisations may be less predictive when the current episode represents an acute decompensation rather than a continuation of chronic instability[39].

Some associations observed in the SHAP analysis may appear counterintuitive, such as certain severe clinical conditions (e.g., delirium or advanced heart failure) being associated with lower predicted readmission risk. This pattern may be explained by the presence of competing risks, particularly early mortality or transitions to palliative care pathways. In this context, patients with greater clinical severity may be less likely to be readmitted because they are at increased risk of death or are managed through non-hospital-based care strategies.

Similar paradoxical associations have been reported in the literature, such as the "obesity paradox" in heart failure[40], where factors traditionally associated with worse prognosis appear protective in selected patient populations. These observations are generally attributed to competing risks, survival bias, and selection effects, reinforcing the need for cautious interpretation of model-derived associations.

This interpretation is consistent with our additional analyses that incorporated a composite outcome of readmission or death, which did not materially improve model performance, suggesting that competing risks alone do not fully account for the observed limited predictive ability.

We cannot compare our findings regarding model performance to those of other studies because the PROFUND index has never been evaluated for predicting readmissions, leaving no reference metrics for this outcome. In our evaluation, neither the recommended baseline tools[41] nor the additional classical and machine-learning models achieved acceptable predictive properties, underscoring the inherent limitations of current artificial intelligence methods in complex clinical prediction tasks.

When comparing the PROFUND-based XGBoost and Ensemble XGBoost models with LACE and HOSPITAL, it became clear that all approaches struggled to achieve strong performance. Discrimination in the external cohort remained modest across the board (AUC 0.52–0.59), consistent with the persistent difficulty of forecasting 30-day readmissions in multimorbid older adults. Notably, none of the evaluated approaches reached conventional thresholds for strong discrimination (AUC $\geq$ 0.70), underscoring the structural difficulty of this prediction task. The wide confidence intervals observed in the external validation, including overlap with 0.5 for AUC, further highlight the uncertainty in model performance and reinforce the limitations imposed by sample size. These findings reinforce a recurring conclusion in the literature: even sophisticated or widely used models often fail to account for the multidimensional drivers of early rehospitalisation.

Notably, the external validation cohort included older patients with a higher burden of comorbidities and worse functional status, which may partly explain the observed decrease in performance and highlights the challenges of model transportability across heterogeneous clinical populations.

From a clinical and organisational perspective, these findings suggest that reliance on static risk scores at discharge may be insufficient to optimise transitional care strategies in multimorbid patients. Early readmission in this population likely depends not only on baseline vulnerability but also on dynamic factors such as medication reconciliation, timely primary care follow-up, caregiver engagement, and access to community resources. Risk stratification tools may therefore be more useful as components of comprehensive discharge planning rather than as standalone decision triggers for resource allocation.

The presence of competing risk due to early mortality is an important consideration in this population. In our cohort, a substantial proportion of patients died within 30 days, which may influence readmission patterns and bias traditional prediction approaches.

To address this, we performed additional analyses using a composite outcome of readmission or death. However, incorporating mortality into the outcome did not meaningfully improve model performance, suggesting that competing risk alone does not fully explain the limited discrimination observed.

These findings indicate that the modest predictive performance is likely driven by a combination of factors rather than a single underlying limitation. First, the available sample size and number of outcome events may be insufficient for stable machine learning model development and precise external validation, potentially contributing to model instability and wide confidence intervals. Second, the models rely exclusively on PROFUND components, which were originally designed to predict mortality rather than readmission, and may therefore not fully capture the determinants of early rehospitalisation. Third, the external validation cohort differed from the development cohort in clinically relevant aspects, including older age, higher comorbidity burden, and worse functional status, which may have affected model transportability and performance.

As expected, LACE prioritised sensitivity at the expense of precision, classifying nearly all readmitted patients but generating many false positives. HOSPITAL offered a slightly better balance, though its overall discrimination and calibration remained limited. These classical scores rely heavily on administrative variables and appear insufficient for the complexity of the multimorbid population studied.

The PROFUND-based XGBoost models showed performance comparable to other modelling approaches, with only small differences across metrics. Importantly, no model demonstrated clear superiority in discrimination.

The choice of XGBoost-based models for final evaluation was therefore based on their stable behaviour across performance metrics, flexibility in probability estimation, and compatibility with interpretability techniques such as SHAP, rather than on meaningful gains in predictive accuracy.

Despite using clinically rich and pathophysiologically plausible variables, such as frailty, functional dependence, cognition, anaemia, and caregiver availability, the models still fell short of a performance threshold that would justify routine clinical use. This highlights both the relevance of the PROFUND domains and the potential limitations of using baseline vulnerability measures alone for predicting short-term readmission. The observed performance should therefore be interpreted as reflecting both the intrinsic complexity of the outcome and the constraints imposed by the study design, including sample size, predictor selection, and cohort differences.

This apparent disconnect between conceptual relevance and predictive performance has also been described in previous work, where variables strongly associated with vulnerability, such as frailty, comorbidity burden, and functional status, do not necessarily translate into accurate prediction of short-term readmission at the individual level[2,42–44].

Several limitations should be acknowledged. A key limitation of this study is the sample size. Although the study was prospective and multicentre, the number of readmission events was modest for prediction modelling, particularly for machine learning approaches and for precise external validation. Recent guidance recommends formal sample size planning for both development and validation studies, with emphasis on calibration, observed-to-expected ratios, and clinical utility. The absence of such planning, together with the limited number of outcome events, may have contributed to model instability, wide confidence intervals, and limited transportability across cohorts. Another limitation of our study is that we use a specific cohort composed of patients in Internal Medicine; generalisability to other populations (e.g., surgical or speciality patients) has not been demonstrated. Given these considerations, the use of PROFUND components for readmission risk stratification should be interpreted with caution. While the domains captured by the index reflect clinically meaningful vulnerability, their predictive performance for early readmission was limited in this setting. Further prospective validation studies, ideally incorporating dynamic post-discharge variables and care-process indicators, are warranted before routine implementation for clinical decision-making.

In conclusion, in a prospective multicentre cohort of multimorbid patients, the use of PROFUND index components, reflecting frailty, functional dependence, and social vulnerability, resulted in only modest improvements in the prediction of 30-day readmission. Even when combined with machine learning techniques, discrimination remained limited in external validation. These findings highlight the complexity of early readmission in multimorbid older adults and suggest that predictive performance is influenced by both the intrinsic characteristics of the outcome and the constraints of the study design. In particular, the use of mortality-oriented predictors, limited sample size, and differences between cohorts may have contributed to the modest discrimination observed. Risk stratification tools based solely on baseline vulnerability may therefore be insufficient to guide post-discharge interventions without incorporating dynamic and care-process–related factors.

Data availability

The datasets generated and analysed during the current study are not publicly available due to institutional ethical restrictions and the presence of sensitive patient-level information. De-identified data may be made available from the corresponding author upon reasonable request and subject to approval by the relevant institutional review board.

Received: 16 February 2026; Accepted: 27 April 2026
Published online: 14 May 2026

References

1. Impact of multimorbidity and complex multimorbidity on healthcare utilisation in. older Australian adults aged 45 years or more: a large population-based cross-sectional data linkage study | BMJ Open [Internet]. [citado el 24 de octubre de 2025]. Disponible en: https://bmjopen.bmj.com/content/14/1/e078762

2. Kansagara, D. et al. Risk prediction models for hospital readmission: A systematic review. *JAMA* **306**(15), 1688–98 (2011).
3. Acute Myocardial Infarction Readmission Risk Prediction Models | Circulation. Cardiovascular Quality and Outcomes [Internet]. [citado el 27 de octubre de 2025]. Disponible en: https://www.ahajournals.org/doi/https://doi.org/10.1161/CIRCOUTCOMES.117.003885
4. Courtney Hebert, C. et al. Diagnosis-specific readmission risk prediction using electronic health data: a retrospective cohort study. *BMC Med. Inf. Decis. Mak.* **14** (1), 65–65 (2014).
5. Martín-Escalante, M. D. et al. Validation of the PROFUND index to predict early post-hospital discharge mortality. *QJM Int. J. Med.* 1 de noviembre de. **112**(11), 854–860. (2019).
6. Anthonimuthu, D. J., Hejlesen, O., Zwisler, A. & Udsen, F. W. Application of machine learning in multimorbidity research: Protocol for a scoping review. *JMIR Res. Protoc.* **13**, e53761. https://doi.org/10.2196/53761 (2024).
7. Sharda, M. et al. The Role of Machine Learning in Predicting Hospital Readmissions Among General Internal Medicine Patients: A Systematic Review. *Cureus* https://doi.org/10.7759/cureus.84761 (2025).
8. Marcus, G. et al. Limited predictive value of traditional comorbidities for readmission in acute decompensated heart failure. *PLoS One* **20**(8), e0329829. https://doi.org/10.1371/journal.pone.0329829 (2025).
9. Péan, C. A. et al. Use of a Six-Item Modified Frailty Index to Predict 30-day Adverse Events, Readmission, and Mortality in Older Patients Undergoing Surgical Fixation of Lower Extremity, Pelvic, and Acetabular Fractures. *Jaaos Global Res. Reviews*. **7** (1). https://doi.org/10.5435/jaaosglobal-d-22-00286 (2023).
10. Pritchard, K. et al. Rates and predictors of opioid deprescribing after fracture: A retrospective study of Medicare fee-for-service claims. *J. Am. Geriatr. Soc.* **73**(3), 737–749. https://doi.org/10.1111/jgs.19290 (2024).
11. Boyi, T., Williams, L. C., Kafle, S., Roche, A. M. & Judson, B. L. Association of age and frailty with 30-day outcomes among patients undergoing oral cavity cancer surgery. *Otolaryngology* **169**(6), 1523–1532. https://doi.org/10.1002/ohn.476 (2023).
12. Desai, A. et al. Evaluation of Modified Frailty Index for predicting postoperative outcomes after upper extremity replantation and revascularization procedures. *J. Reconstr. Microsurg.* **41**(07), 557–565. https://doi.org/10.1055/a-2460-4706 (2024).
13. Lekan, D. et al. Comparison of a Frailty Risk Score and comorbidity indices for hospital readmission using Electronic Health Record data. *Res. Gerontol. Nurs.* **14**(2), 91–103. https://doi.org/10.3928/19404921-20210115-03 (2021).
14. Sheets, K. M. et al. Life-space mobility and healthcare costs and utilization in older men. *J. Am. Geriatr. Soc.* **69**(8), 2262–2272. https://doi.org/10.1111/jgs.17187 (2021).
15. Kang, Y., Fang, X., Wang, D. & Wang, X. Activity of daily living upon admission is an independent predictor of in-hospital mortality in older patients with community-acquired pneumonia. *BMC Infect. Dis.* https://doi.org/10.1186/s12879-021-06006-w (2021).
16. Gay, P. et al. Modifiable factors associated with postoperative atrial fibrillation in older patients with hip fracture in an orthogeriatric care pathway: A nested case–control study. *BMC Geriatr.* https://doi.org/10.1186/s12877-022-03556-9 (2022).
17. Stutsrim, A. E. et al. Automated Electronic Frailty Index is associated with non-home discharge in patients undergoing open revascularization for peripheral vascular disease. *Am. Surg.* **89**(11), 4501–4507. https://doi.org/10.1177/00031348221121547 (2022).
18. Pinto, A. Artificial intelligence in the management of infectious diseases in older adults: Diagnostic, prognostic, and therapeutic applications. *Biomedicines* **13**(10), 2525. https://doi.org/10.3390/biomedicines13102525 (2025).
19. Obuobi, S., Chua, R. F. M., Besser, S. A. & Tabit, C. E. Social determinants of health and hospital readmissions: can the HOSPITAL risk score be improved by the inclusion of social factors? *BMC Health Serv. Res. 4 de enero de.* **21** (1), 5 (2021).
20. Sharma, Y., Thompson, C., Mangoni, A. A., Horwood, C. & Woodman, R. Comparative evaluation of machine learning models using structured and unstructured clinical data for predicting unplanned general medicine readmissions in a tertiary hospital in Australia. *Computers* **15**(3), 138. https://doi.org/10.3390/computers15030138 (2026).
21. Sharda, M. et al. The role of machine learning in predicting hospital readmissions among general internal medicine patients: A systematic review. *Cureus* **17**(5), e84761. https://doi.org/10.7759/cureus.84761 (2025).
22. Cara, R. et al. An Artificial Intelligence model for predicting hospital readmission using Electronic Health Records data. *Mesopotam. J. Artif. Intell. Healthc.* **2025**, 116–123. https://doi.org/10.58496/MJAIH/2025/012 (2025).
23. Garcia-Gutierrez, S. et al. Rethinking readmission prediction in multiple chronic conditions: Clinical utility over accuracy, European Journal of Internal Medicine, 106754, ISSN 0953–6205. https://doi.org/10.1016/j.ejim.2026.106754. (2026). https://www.sciencedirect.com/science/article/pii/S0953620526000592).
24. Ramírez-Duque N., et al. Clinical, functional, mental and sociofamiliar features in pluripathological patients. One-year prospective study in Primary Health Care. Revista Clínica Española, **208**, Issue 1, pages 4–11. (https://www.sciencedirect.com/science/article/abs/pii/S0014256508716606).
25. Garcia-Gutierrez, S. RePluris project: Identification of multimorbid patients with a higher probability of readmission and mortality. Zenodo.
26. Riley R, Ensor J, Snell K et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. The Lancet Digital Health, ; 7, (2025). https://www.thelancet.com/journals/landig/article/PIIS2589-7500(25)00021-4/fulltext).
27. Fonseca, J. et al. Identification of high-risk patients for early death or unplanned readmission using the LACE index in an older Portuguese population [Internet]. F1000Research; [citado 31 de octubre de 2025]. Disponible en: (2017). https://f1000research.com/articles/6-1798
28. The HOSPITAL Score Predicts Potentially Preventable 30-Day Readmissions in Conditions Targeted by the Hospital Readmissions Reduction Program. *Med. Care 1 de marzo de*. **55** (3), 285–290 (2017).
29. Hughes, L. D. & Witham, M. D. Causes and correlates of 30 day and 180 day readmission following discharge from a Medicine for the Elderly Rehabilitation unit. *BMC Geriatr.* 28 de agosto de. **18**(1), 197. (2018).
30. Costa, MLM. et al. Development and validation of predictive model for long-term hospitalization, readmission, and in-hospital death of patients over 60 years old. Dev Valid Predict Model Long-Term Hosp Readmission -Hosp Death Patients *60 Years Old* ;**20**. (2022).
31. Rizzi, MA. et al. Factors associated with in-hospital mortality and adverse outcomes during the vulnerable post-discharge phase after the first episode of acute heart failure: results of the NOVICA-2 study. Clin Res Cardiol. 1 de julio de. ;110(7):993–1005. (2021).
32. Hospitalized Patients with Delirium. and 28-Day Unplanned Hospital Readmissions: A Longitudinal Retrospective Cohort Study - ClinicalKey [Internet]. [citado 3 de noviembre de 2025]. Disponible en: https://www.clinicalkey.es/#!/content/playContent/1-s2.0-S1525861024003475?returnurl=https:%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1525861024003475%3Fshowall%3Dtrue&referrer=https:%2F%2Fscispace.com%2F
33. Samuel, S. V., Viggeswarpu, S., Wilson, B. P. & Ganesan, M. P. Readmission rates and predictors of avoidable readmissions in older adults in a tertiary care centre. *J. Fam Med. Prim. Care septiembre de*. **11** (9), 5246 (2022).
34. Hoogendijk, E., Dent, E. & Recent developments in frailty identification treatment, and prevention in older adults: a review. *Innov Aging.* 1 de diciembre de 2023; **7**(Supplement_1), 868.
35. Care Configurations and Unmet Care Needs in Older Men and Women - Andrew J. Potter. [citado 3 de noviembre de 2025]. Disponible en: https://journals.sagepub.com/doi/ (2019). [Internet] https://doi.org/10.1177/0733464817733239
36. Loneliness, social isolation, and living alone: a comprehensive systematic review, meta-analysis, and meta-regression of mortality risks in older adults | Aging Clinical and Experimental Research [Internet]. [citado 3 de noviembre de 2025]. Disponible en: https://link.springer.com/article/https://doi.org/10.1007/s40520-024-02925-1

37. Relationship between Anemia and Readmission among Older Patients in Rural Community Hospitals: A Retrospective Cohort Study [Internet]. [citado 3 de noviembre de 2025]. Disponible en: https://www.mdpi.com/2077-0383/13/2/539
38. Kaya, S., Guven, G. S., Aydan, S. & Toka, O. A comprehensive framework identifying readmission risk factors using the CHAID algorithm: a prospective cohort study. *Int. J. Qual. Health Care 1 de junio de.* **30** (5), 366–374 (2018).
39. Garrigues, S. K. et al. The transition tightrope: Optimizing post-discharge care for acute leukemia patients. *JCO Oncol. Pract. noviembre de.* **19** (11_suppl), 323–323 (2023).
40. El-Khoury, R. et al. Obesity is associated with a lower risk of mortality and readmission in heart failure patients with diabetes. *Biomedicines* **13**(12), 3086. https://doi.org/10.3390/biomedicines13123086 (2025).
41. Scarlat, A. & Campion, F. X. Predicting 30-Day Mortality and Readmission Using Hospital Discharge Summaries: A Comparative Analysis of Machine Learning Models, Large Language Models, and Physicians [Internet]. medRxiv; 2025 [citado 3 de noviembre de 2025]. p. 2025.03.26.25324714. Disponible en: https://www.medrxiv.org/content/https://doi.org/10.1101/2025.03.26.25324714v1
42. Scarlat, A. et al. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: An updated systematic review. *BMJ. Open* **6**(6), e011060 (2016).
43. Clegg, A., Young, J., Iliffe, S., Rikkert, M. O. & Rockwood, K. Frailty in elderly people. *Lancet* **381**(9868), 752–762 (2013).
44. Hippisley-Cox, J. & Coupland, C. Development and validation of risk prediction algorithms to estimate future risk of common adverse outcomes in people with multimorbidity. *BMJ* **350**, h1251 (2015).

Acknowledgements

The authors acknowledge the Re-Pluris Working Group for their contribution to patient recruitment and data collection. The authors also acknowledge the Deustek Research Group (University of Deusto) for institutional and technical support.

Author contributions

A.P.-H. conceived the study and led the machine learning modelling and computational analyses. A.A. and U.Z. contributed to model development and technical implementation. M.J.L. and D.M. performed the statistical analyses and contributed to data interpretation. S.G.-G. and R.Q. provided clinical expertise, contributed to study design, and interpreted the results from a clinical perspective. A.P.-H. and S.G.-G. drafted the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final version.

Funding

Funding for this research was provided by the Carlos III Health Institute (Instituto de Salud Carlos III, ISCIII), project reference PI18/01438 "Identification of multimorbid patients with a higher probability of re-admission and mortality: Role of the patient-related variables," co-funded by the European Union through the European Regional Development Fund (ERDF) "A way to make Europe."

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-51343-y.

Correspondence and requests for materials should be addressed to A.P.-H.