Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting In-Hospital Falls Using Routine Care Data: A Precision Nursing Evaluation of Resampling Strategies for Imbalanced Classification

Greici Capellari Fabrizzio
Federal University of Santa Catarina
Lincoln Moura Oliveira
Federal University of Ceará
Cecilia Biasibetti Soster
cecilia.soster@gmail.com
Federal University of Santa Catarina
Alacoque Lorenzini Erdmann
Federal University of Santa Catarina
Elisiane Lorenzini
Federal University of Santa Catarina

Research Article

Keywords: precision nursing, in-hospital falls, machine learning, resampling techniques, risk prediction, hospitals

Posted Date: September 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10639456/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. Cecilia Biasibetti Soster is a public health nurse at the institution where data collection occurred; however, no institutional interference occurred in the development or analysis of the study. All other authors declare no potential conflicts of interest.

Predicting In-Hospital Falls Using Routine Care Data: A Precision Nursing Evaluation of Resampling Strategies for Imbalanced Classification

Greici Capellari Fabrizzio[1]

Lincoln Moura de Oliveira[2]

Cecilia Biasibetti Soster[3]

Alacoque Lorenzini Erdmann[1]

Elisiane Lorenzini[1]*

[1] Department of Nursing, Health Sciences Center, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

[2] Department of Electrical Engineering, Federal University of Ceará, Fortaleza, Ceará, Brazil.

[3] Grupo Hospitalar Conceição, Porto Alegre, Rio Grande do Sul, Brazil.

***Corresponding author:**

Cecilia Biasibetti Soster

Email: cecilia.soster@gmail.com

Email addresses:

Greici Capellari Fabrizzio: greicicapellari@gmail.com

Lincoln Moura de Oliveira: lincolnmoura@petrobras.com.br

Cecilia Biasibetti Soster: cecilia.soster@posgrad.ufsc.br

Alacoque Lorenzini Erdmann: alacoque.erdmann@ufsc.br

Elisiane Lorenzini: elisiane.lorenzini@ufsc.br

Abstract

Background: This study aimed to evaluate whether routinely collected hospital data can support interpretable prediction of in-hospital falls and to determine how class-imbalance handling strategies affect model performance within a precision nursing framework.

Methods: This retrospective study used fall notification forms from three public hospitals and included sociodemographic, disability-related, and hospitalization variables not specifically designed for fall risk assessment, such as age, sex, race/ethnicity, marital status, education level, disability status, and hospital length of stay. The modeling framework integrated four resampling techniques—random undersampling, random oversampling, Synthetic Minority Over-sampling Technique for Nominal and Continuous features, and Conditional Tabular Generative Adversarial Network—alongside a non-resampled baseline. Performance was evaluated using stratified 5-fold cross-validation, balanced accuracy, recall, accuracy, precision, and F1-score. Feature importance was analyzed using Shapley Additive Explanations values applied to the best-performing model.

Results: The highest performance was achieved by the Random Forest model combined with RandomOverSampler and hyperparameter tuning, reaching a balanced accuracy of 77.00%. This represented a 4.2-percentage-point gain over the baseline Random Forest model, which achieved 72.8%, indicating the importance of addressing class imbalance. While Conditional Tabular Generative Adversarial Network–based synthetic data offered privacy and reproducibility benefits, it did not improve performance in this dataset, with the best configuration underperforming at 70.05%. Simpler resampling techniques proved more effective, especially given the dataset's structured and low-complexity nature. The most influential variables

according to Shapley Additive Explanations values were age, followed by hospital length of stay and sex.

Conclusions: Among the evaluated resampling strategies, RandomOverSampler combined with a Random Forest model showed the best predictive performance. These findings support the potential use of routinely collected hospital data, including age, hospital length of stay, and sex, to develop interpretable and low-cost predictive models for in-hospital fall-risk assessment within precision nursing.

Keywords: precision nursing; in-hospital falls; machine learning; resampling techniques; risk prediction; hospitals.

1 Introduction

A fall was defined as an event that causes the individual to involuntarily stop on the ground or another lower level unintentionally[1]. The pathophysiology of a fall results from an imbalance between the individual's ability to maintain postural stability and the demands placed on them by their physical environment and can vary according to individual circumstances. Thus, fall can be the result of an imbalance between cognitive, sensory and motor information to maintain postural stability[2].

According to the Joint Commission International (JCI) falls are included among the six International Patient Safety Goals aimed at enhancing global healthcare safety[3]. Falls are the second leading cause of accidental death in the world. In the hospital environment, they range from 1.3 to 16.9 per 1,000 patients per day and have an impact on increased mortality rates, hospitalization time and costs, reducing patients' quality of life[4-5], as well as causing injuries, fractures, traumatic brain injury and fear of falling again, which are linked to immobility and loss of progressive independence[4].

Falls are considered rare events, appearing in an unbalanced way in a data set, and can be considered an anomaly[6]. Imbalanced datasets are commonly encountered in healthcare applications, posing significant challenges to effective modeling, particularly when the minority class, such as instances of falls, is substantially underrepresented compared to the majority class (e.g., non-falls). Data imbalance is characterized by a considerable disparity in the number of samples across classes, which can adversely affect model performance, often resulting in low sensitivity and high false negative rates for the minority class. In these cases, resampling methods are useful in identifying rare and significant health events in a data set[7].

To address this issue, resampling techniques have been widely investigated to rebalance datasets and enhance model effectiveness. These methods aim to improve class representation, especially for the minority class, thereby allowing development of more robust predictive models. Resampling strategies include undersampling methods, which reduce the number of samples from the majority class, and oversampling methods, which increase the representation of the minority class by generating synthetic instances. Hybrid approaches, often referred to as "smoothing methods," combine elements of both undersampling and oversampling to improve balance while preserving data integrity[8-11].

The sociodemographic, disability-related, and hospitalization variables of patients who had fallen were analyzed using the Precision Nursing framework, aiming to stratify fall risk considering the individual characteristics that may influence this outcome. Precision Nursing proposes the integrated use of clinical, sociodemographic and contextual data to support clinical decision-making and promote personalized, safe and evidence-based care[12,13]. In this context, the identification of variables that

predict falls benefits from the incorporation of real-world data and use of advanced analysis technologies, which contributes to recognize specific risk profiles and to plan care interventions more assertively. The Precision Nursing approach contributes to increasing effectiveness of in-hospital falls prevention strategies, reinforcing the importance of individualized, data-supported care.

The main objective of this work is to perform a systematic evaluation of resampling strategies and machine learning models for in-hospital falls risk prediction within imbalanced hospital data based on the precision nursing framework.

This study contributes to nursing research by demonstrating how routinely available administrative and sociodemographic variables can be transformed into interpretable fall-risk information. Rather than replacing clinical judgement or validated fall-risk scales, the proposed model is intended to support nurses' situational awareness, prioritization of reassessment, and early identification of patients who may require intensified fall-prevention measures.

2 Methods

2.1 Design

This is a retrospective study that used fall notification forms sent through the institutional system. This study was conducted according to Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD+AI statement) (Appendix S1). No study protocol was prospectively registered.

2.2 Study Setting and Sampling

The hospitals in the study are three public hospitals that are part of the Conceição Hospital Group, located in southern Brazil. The hospital has an occupancy

rate of over 83.03%, with 1,179 beds available for patients requiring acute care. To define the sample, the total number of hospitalizations during the study period (n = 738,647) was considered, with a margin of error of 1%, a population proportion of 1%, and a 99% confidence level. A sample consisting of 16,213 records was deemed statistically significant.

2.3 Data collection

Data was collected from January to April 2024 in an automated way by the IT department and comprises falls notifications forms from 1st January 2013 to 31st December 2023. Records of patients under 18 or incomplete were excluded.

The analysis was conducted on an initial dataset comprising 73,977 patient records. Figure 1 shows the main stages of the modeling pipeline, organized into three integrated phases. The first phase involves data preprocessing, including data cleaning, dataset partitioning, categorical encoding, feature scaling, and the application of resampling techniques to address class imbalance. During this step, records with missing values were excluded, and no imputation techniques were applied. To prevent data leakage, encoding and scaling procedures were fitted exclusively on the training set and then applied to the test set. After preprocessing, the dataset was reduced to 45,554 complete entries.

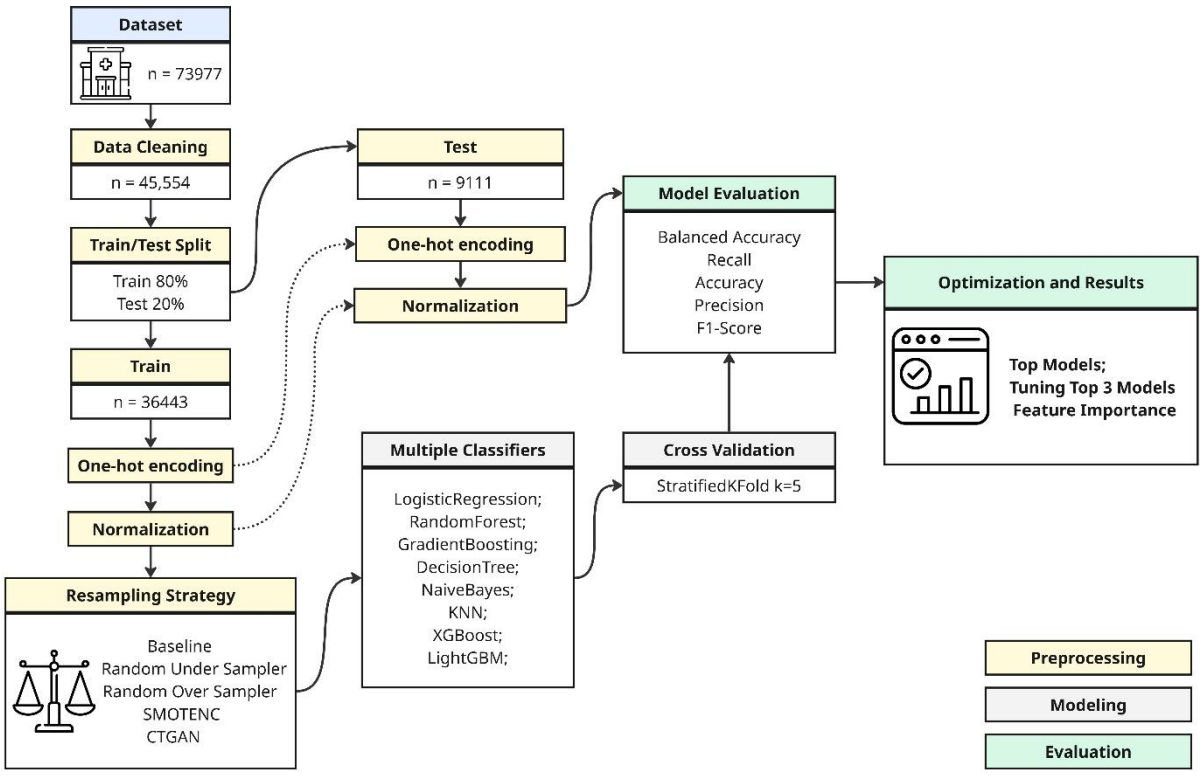


Figure 1 - Workflow of the Predictive Modeling Pipeline.

Categorical variables were encoded using the one-hot encoding technique, while the numerical variables, age and hospital length of stay, were retained in their continuous format. Consequently, the resulting dataset was predominantly composed of categorical features. The dataset was then divided into training (80%) and testing (20%) subsets using stratified sampling to preserve the original class distribution.

Numerical features were normalized using the Min-Max scaling technique, ensuring that all values were transformed to range [0, 1], as expressed in Equation (1):

$$x_{\text{scaled}} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where, x is the original variable value, x_{min} and x_{max} are the minimum and maximum variable values, respectively. x_{scaled} is the value resulting scaled to [0, 1] interval.

A significant class imbalance was identified, with 76% of data corresponding to non-fall events and 24% representing fall occurrences, thus defining the positive class as the minority. To mitigate this issue and evaluate its impact on model performance, four resampling strategies were applied exclusively to training data: random undersampling, random oversampling, Synthetic Minority Over-sampling Technique for Nominal and Continuous features (SMOTE-NC), and Conditional Tabular Generative Adversarial Network (CTGAN).

The second phase of the pipeline consists of training multiple classification algorithms and evaluating their performance using cross-validation. Balanced accuracy was employed as primary evaluation metric, while recall, accuracy, precision, and F1-score served as complementary performance indicators. Models were trained on both the resampled datasets and the original imbalanced training set, referred to as the baseline data. Evaluation was conducted on the original, unaltered test set to ensure a consistent and unbiased comparison. The performance of resampled models was then compared against baseline model.

The third and final phase encompasses the selection of the three best-performing models, followed by hyperparameter optimization and post hoc interpretability analysis. Shapley Additive Explanations were used to evaluate the relative importance of input features to predict decisions on the best-performing model and its associated resampling technique.

The processed dataset comprised variables context relevant to Precision Nursing. Sociodemographic variables considered were: age a numerical variable representing patient's age in years; sex (SEX_F, SEX_M), encoded as two binary indicators to capture patient's sex indicated on birth; race/ethnicity categorized into

five groups, white (ETH_W), black or brown (ETH_BB), mixed race (ETH_Y), indigenous (ETH_I), and other (ETH_O), each represented as separate one-hot encoded variables; marital status encoded as categorical indicators for single (MS_S), married or in a stable union (MS_MU), divorced (MS_D), widowed (MS_W), and other (MS_O); educational level was grouped into four ordinal categories: no formal education or illiterate (EDU_0), primary education (EDU_1), secondary education (EDU_2), and higher education (EDU_3), each represented through one-hot encoding.

Regarding clinical and hospitalization-related variables, disability status included binary indicators reflecting the presence of specific disability types: physical (DIS_PHY), visual (DIS_VIS), hearing (DIS_HEA), autism spectrum disorder (DIS_AUT), intellectual or mental disability (DIS_INT), multiple disabilities (DIS_MUL), other types (DIS_OTH), or no reported disability (DIS_NONE). The target variable (FALL_EVENT) indicated a fall event during hospitalization, with a binary outcome: 1 for fall occurrence and 0 for no fall. The database did not contain variables related to lifestyle habits or genetic information. Hospital length of stay (HOSP_DAYS) was a numerical variable representing the total length of the hospitalization episode recorded in the administrative database. In this retrospective study, HOSP_DAYS was interpreted as an indirect marker of cumulative exposure to the hospital environment and care trajectory duration. Therefore, its predictive contribution should be understood as an association between fall occurrence and hospitalization-duration patterns, rather than as a fixed baseline patient characteristic. In future prospective applications, HOSP_DAYS should be updated at each risk-assessment point as the number of days elapsed since admission.

2.4 Class Imbalance Handling Strategies and Synthetic Data

The problem of data imbalance can be defined in terms of a dataset X, composed of a minority class subset X_{min} and a majority class subset X_{maj}. The class balance ratio is expressed as expressed in Equation (2):

$$R_X = \frac{|X_{min}|}{|X_{maj}|} \quad (2)$$

Where $|\cdot|$ denotes the absolute number of instances. Resampling techniques aim to adjust this ratio by creating a new dataset Y, such that $R_x < R_Y \approx 1$. Achieving $R_Y \approx 1$ can be done in three main ways: by reducing $|X_{maj}|$ through undersampling; by increasing $|X_{min}|$ through oversampling; or by applying hybrid methods that combine both strategies. These approaches are essential to improve model performance in imbalanced classification tasks and are effectively implemented in tools such as the imbalanced-learn Python library[14].

It is important to note that while random oversampling increases number instances in the minority class by replicating existing samples, it does not increase diversity or uniqueness data. In contrast, more advanced techniques, such as SMOTE-NC or CTGAN, aim to increase minority class cardinality effective by generating new, synthetic instances that introduce additional variation.

To address class imbalance, all resampling strategies were applied exclusively to the training set, thereby maintaining original class distribution in the test set as a reference for evaluating model performance under realistic conditions. The resampling techniques were selected with particular attention to categorical variables present in the dataset, which determined our choice to adopt widely used methods such as SMOTE and ADASYN. These methods, due to their interpolation-based extrapolation processes, may generate values that are inconsistent with semantics categorical variables.

Four resampling strategies were considered to address class imbalance, along with a baseline model used for comparison. The **baseline approach** involved training classification models on the original, imbalanced dataset without any resampling. It served as a reference to evaluate the effectiveness of balancing techniques. To mitigate class imbalance, the imbalanced-learn library[14] was used to apply three resampling methods. **Random undersampling** reduces the size of the majority class by randomly removing instances, while **random oversampling** increases the representation of the minority class by duplicating existing samples. These methods are straightforward but may lead to trade-offs such as loss of information or overfitting. In contrast, **SMOTE-NC** generates synthetic samples for the minority class and is specifically designed for datasets containing both categorical and continuous features, by correctly handling categorical variables during the interpolation process.

In addition to classical resampling techniques, a deep learning-based generative model was employed to synthesize new training samples: Conditional Tabular Generative Adversarial Network (CTGAN)[15]. CTGAN is specifically designed to generate synthetic tabular data and is particularly effective in datasets containing a mixture of numerical and categorical features. Unlike traditional Generative Adversarial Networks (GANs), CTGAN conditions the generation process on discrete column values, allowing it to better capture complex dependencies and preserve underlying data distributions especially in imbalanced classification problems, where augmenting the minority class is crucial.

CTGAN is built upon the broader GANs, framework originally introduced by Ian Goodfellow et al.[16]. GANs are a type of generative model that learns probability distributions of how high-dimensional data is likely to be distributed[17]. GANs are composed of two neural networks trained in opposition: a generator (G), which learns

to produce synthetic data from a random noise vector, and a discriminator (D), which attempts to distinguish between real and generated samples. The training process is formulated as a two-player minimax game, in which both networks are simultaneously optimized through adversarial feedback[16].

The generator aims to approximate the true data distribution by transforming latent noise vectors \\\($z \sim p_z(z)$\\\) into synthetic samples \\\($G(z)$\\\) that resemble real data. Simultaneously, the discriminator is trained to maximize its ability to distinguish between real samples \\\($x \sim p_{\text{data}}(x)$\\\) and generated instances. This adversarial process is governed by the following objective in (3) function:

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[logD(x)] + E_{z\sim p_z(z)}\left[log\left(1 - D\big(G(z)\big)\right)\right] \quad (3)$$

where p_{data} is the true data distribution and p_z is the noise distribution. This formulation allows GANs to generate high-fidelity, context-aware synthetic data that preserves statistical dependencies and original dataset structural properties.

2.5 Classification algorithms evaluated in predictive pipeline

The classification pipeline evaluated eight distinct algorithms, commonly used in supervised binary learning tasks. **Logistic Regression**: A linear classification model that estimates class probabilities using logistic (sigmoid) function. It is interpretable and effective for linearly separable problems. **Decision Tree**: A tree-based model that recursively partitions the feature space to create decision rules. It is intuitive and capable of capturing nonlinear relationships. **Random Forest**: An ensemble method that constructs multiple decision trees and aggregates their outputs to improve generalization and reduce overfitting. **Gradient Boosting**: A boosting technique that builds decision trees sequentially, each one corrects the errors of its predecessor, offering strong performance on structured data. **Gaussian Naive**

Bayes: A probabilistic classifier based on Bayes' theorem, assuming feature independence and Gaussian distribution. It is fast and performs well on small datasets. **K-Nearest Neighbors**: A non-parametric algorithm that classifies instances based on majority class of their nearest neighbors in feature space. **XGBoost**: An optimized gradient boosting framework known for its high performance and efficiency. It incorporates regularization to control model complexity. **LightGBM**: A gradient boosting framework designed for speed and scalability, using histogram-based algorithms and leaf-wise tree growth for improved accuracy.

2.6 Stratified cross-validation

Each dataset variant comprising the baseline and four distinct resampling techniques was combined with eight classification algorithms, resulting in a total of 40 unique model configurations (resampling method × classifier). A 5-fold stratified cross-validation was conducted exclusively on a training set to ensure that class proportions were preserved within each fold. During each iteration, performance metrics were computed to provide a fair assessment of model behavior under imbalanced class distributions. Balanced accuracy across the five folds were recorded by calculating mean and standard deviation for each configuration.

Following cross-validation, each model was retrained on the entire training set using its optimized hyperparameters and subsequently evaluated on the held-out test set. This final evaluation aimed to assess generalization performance and detect potential signs of overfitting.

2.7 Test set evaluation

297 Following cross-validation, each model was retrained using a complete training
298 dataset and then evaluated on a held-out test set. To quantify model's performance,
299 the evaluation metrics in equations 4, 5, 6, 7 and 8 were computed:

300 $$\text{Balanced Accuracy (BA)} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \quad (4)$$

301 $$\text{Recall (REC)} = \frac{TP}{TP+FN} \quad (5)$$

302 $$\text{Accuracy (ACC)} = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

303 $$\text{Precision (PRE)} = \frac{TP}{TP+FP} \quad (7)$$

304 $$\text{F1-Score (F1)} = \frac{2\cdot(\text{Precision}\cdot\text{Recall})}{\text{Precision}+\text{Recall}} \quad (8)$$

305 Where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False
306 Negatives) represent the entries of the confusion matrix.

307 Balanced Accuracy is particularly important in the context of imbalanced
308 datasets, as it ensures that both classes contribute equally to the final evaluation
309 score, regardless of their prevalence. In this study, Balanced Accuracy was adopted
310 as primary metric for model selection, offering a fair and reliable criterion for identifying
311 the most robust and generalizable model configurations. Mathematically, Balanced
312 Accuracy is defined as the arithmetic mean of recall (true positive rate) and specificity
313 (true negative rate).

314 **2.8 Hyperparameter Fine-Tuning**

315 The three top-performing models, selected based on their highest Balanced
316 Accuracy scores on the test set, were further refined through hyperparameter
317 optimization using GridSearchCV with 5-fold cross-validation. Each classifier was

assigned a custom hyperparameter grid tailored to its architecture and implementation. For example, the Decision Tree model was tuned by varying the maximum tree depth (with values of 5, 10, 15, and no depth limit) and the minimum number of samples required to split a node (using values of 2, 5, and 10). Similarly, the Random Forest model was evaluated using different numbers of decision trees (100, 200, and 300 estimators), maximum depths (5, 10, 20, and no limit), and minimum samples per split (2, 5, and 10).

2.9 Shapley Additive exPlanations

Shapley Additive explanations (SHAP) is a comprehensive and model-agnostic framework for interpreting machine learning predictions, grounded in cooperative game theory. It assigns to each feature an importance value for a particular prediction by computing the average marginal contribution of that feature across all possible feature combinations. SHAP values ensure consistency and local accuracy, offering both global interpretability and instance-level explanations. This analysis supports a deeper understanding of the model's decision-making process across both majority and minority classes[18].

2.10 Experimental Setup

All experiments were conducted using Python 3.8 on a computer equipped with an Intel Core i7 CPU, 16 GB of RAM, and an NVIDIA RTX 2060 GPU. Data preprocessing, modeling, and evaluation were primarily implemented using scikit-learn (v1.3.2), with imbalanced-learn (v0.12.4) employed for resampling techniques, except for generative approach, which utilized SDV (v1.17.3). Gradient boosting models were developed using XGBoost (v2.1.1) and LightGBM (v4.5.0). Data manipulation and visualization were performed using pandas (v1.5.3), numpy

342 (v1.24.4), matplotlib (v3.6.0), and seaborn (v0.13.2), while feature importance
343 analysis based on SHAP values was conducted using the shap library (v0.47.2).

344 **3 Results**

345 The sample for this study comprised 45,554 records, including 10,868 patients
346 who experienced falls and 34,686 patients who did not. Table 1 presents the
347 characteristics of the sample.

348

Table 1 - Characteristics of patients included in the study. Comparison of groups of patients who fell and those who did not fall.

		Fall (n=10,868)		No fall (n=34,686)		
Variable		**n**	**%**	**n**	**%**	**P[a]**
Sex						<0.001
	Female	4,439	40.84	18,404	53.06	
	Male	6,429	59.16	16,282	46.94	
Age						<0.001
	Average (SD)	66.39 (18.27)		61.09 (18.27)		
	Min/max	18/100		18/100		
Length of Stay (days)						0.269
	Average (SD)	30.65 (29.44)		30.31 (29.44)		
	Min/max	1/199		1/199		
Race						0.415
	Asian	6	0.06	37	0.11	
	White	8,895	81.85	28,400	81.88	

	Indigenous	0	0	5	0.01
	Black/brown	1,446	13.31	4,595	13.25
	Other	521	4.79	1,649	4.75
Marital status					<0.001
	Single	4,144	38.13	14,841	42.79
	Married	3,883	35.73	12,097	34.88
	Unmarried	1,199	11.03	3,019	8.70
	Widowed	1,430	13.16	4,058	11.70
	Other	212	1.95	671	1.93
Education					<0.001
	No education	646	5.94	2,077	5.99
	1 to 4 years	7,835	72.09	23,170	66.80
	5 to 9 years	2,148	19.76	8,224	23.71
	Over 9 years	239	2.20	1,215	3.50
Disability					<0.001
	Auditory	5	0.05	40	0.12
	Autism	0	0.00	2	0.01
	Physical	66	0.61	342	0.99
	Mental	26	0.24	82	0.24
	Multiple	14	0.13	24	0.07
	Visual	11	0.10	31	0.09
	Other	6	0.06	25	0.07
	No disability	10.740	98.82	34,140	98.43

a) For categorical variables, the Chi-square test was used, with statistical significance if $p<0.05$. For continuous variables, the Student's t-test was used, with statistical significance if $p<0.05$.

Table 2 presents the top 10 experimental results, ranked by their Balanced Accuracy scores. This ranking highlights the most effective model configurations according to class balancing strategy type employed. The columns provide detailed performance metrics and contextual information for each setup. Sampling Method indicates resampling technique used to address class imbalance. Model specifies algorithm classification applied; entries marked with the suffix "(tuned)" indicate that hyperparameter optimization was performed for that configuration. The remaining columns report the evaluation metrics computed for each corresponding model.

According to the results shown in Table 2, it was found that for the Balanced Accuracy metric, the Random over sampling: random forest (tuned), random forest, decision tree (tuned) and decision tree. Methods had significantly higher estimates compared to the other methods. In relation to the models for Recall performance, the Random over-sampling: random forest method [0.776 - 95%CI: 0.760 - 0.792] stood out as a significantly higher estimator compared to all the other methods. The lowest estimates for Recall were for the Baseline: forest and random forest (tuned) models, as well as the SMOTE-NC: decision tree model.

Regarding the metric estimated by Accuracy, the methods with significantly lower estimates were the ones that stood out: random over-sampling: random forest and CTGAN: random forest. In terms of Precision, the highest estimate came from the Baseline: random forest method, compared to all the other models. Significantly lower estimates were found in the Random over-sampling: random forest and CTGAN: random forest methods.

Regarding the F1 Score metric, most of the methods showed statistically similar estimates. It is worth noting that the statistically lowest values occurred in the Random over-sampling: random forest; SMOTE-NC: decision tree; and CTGAN: random forest methods.

376 Table 2 - Performance Metrics of Binary Classification Models under Different Class Balancing Strategies.

Sampling Method	**BA**[1]	**REC**[2]	**ACC**[3]	**PRE**[4]	**F1**[5]
Model	(IC 95%)	(IC 95%)	(IC 95%)	(IC 95%)	(IC 95%)
Random Over Sampling					
Random forest (tuned)	0.770a (0.758-0.782)	0.731b (0.719-0.743)	0.791a (0.776-0.806)	0.547b (0.522-0.572)	0.626[a] (0.600-0.652)
Random forest	0.769a (0.756-0.782)	0.731b (0.718-0.744)	0.790a (0.775-0.805)	0.544b (0.521-0.567)	0.624a (0.602-0.646)
Decision tree (tuned)	0.767a (0.754-0.780)	0.738b (0.724-0.752)	0.782a (0.766-0.798)	0.530b (0.505-0.555)	0.617a (0.593-0.641)
Decision tree	0.763a (0.748-0.778)	0.728b (0.712-0.744)	0.781a (0.765-0.797)	0.530b (0.507-0.553)	0.613a (0.591-0.635)
Baseline					
Random forest	0.728b (0.703-0.753)	0.575e (0.552-0.598)	0.809a (0.771-0.847)	0.605a (0.577-0.633)	0.589a (0.557-0.621)
Random forest (tuned)	0.727b (0.701-0.754)	0.569e (0.547-0.591)	0.810a (0.788-0.832)	0.509 (0.477-0.541)	0.588a (0.552-0.624)
SMOTE-NC					
Random forest	0.726b (0.708-0.744)	0.623d (0.611-0.635)	0.780a (0.766-0.794)	0.534b (0.511-0.557)	0.575b (0.553-0.597)
Random Over Sampling					
Random forest	0.714b (0.697-0.731)	0.776a (0.760-0.792)	0.682d (0.665-0.699)	0.441c (0.415-0.468)	0.538c (0.515-0.561)

SMOTE-NC					
Decision tree	0.710b (0.692-0.728)	0.589e (0.573-0.605)	0.773b (0.757-0.789)	0.522b (0.499-0.545)	0.554c (0.530-0.578)
CTGAN					
Random forest	0.705b (0.688-0.722)	0.691c (0.672-0.710)	0.711c (0.695-0.727)	0.434c (0.408-0.460)	0.534c (0.515-0.553)

1) BA: Balanced Accuracy; 2) REC: Recall; 3) ACC: Accuracy; 4) PRE: Precision; 5) F1: F1 score.

Notation: significant differences presented by the letter system where performance metrics followed by equal letters (in the same col-umn) do not differ at a significance level of 5%.

Figure 2 illustrates two fundamental components that comprise Balanced Accuracy metric. The x-axis represents Recall (True Positive Rate), while the y-axis corresponds to Specificity (True Negative Rate). Each point on the plot denotes a classification model, where the color encodes the algorithm used and the marker shape indicates the resampling strategy applied. Models located near the top-right corner demonstrate superior performance, as they simultaneously achieve high sensitivity and specificity, reflecting a well-balanced classification across both classes.

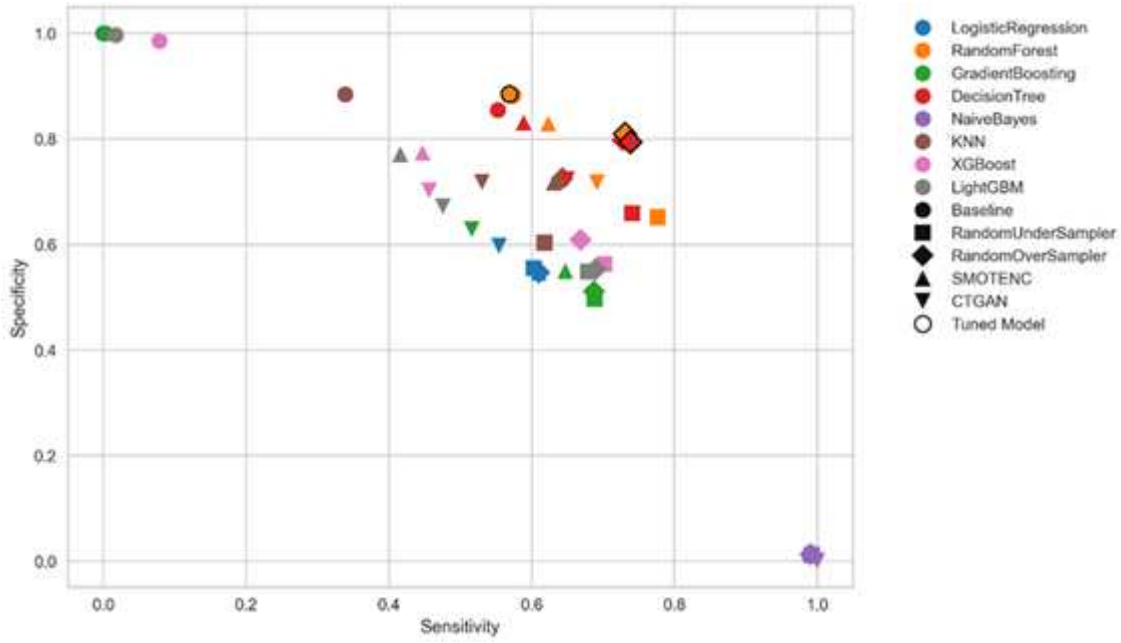


Figure 2 - Balanced Accuracy Component Analysis.

Progression training of the CTGAN is demonstrated in Figure 3, highlighting the adversarial min-max dynamics between generator and discriminator, as defined in Equation 3. Over successive training epochs, generator gradually improves its ability to produce realistic synthetic samples, while discriminator simultaneously refines its capacity to distinguish between real and generated data. This interaction reflects the expected behavior of generative adversarial networks, in which both models iteratively compete and learn from one another.

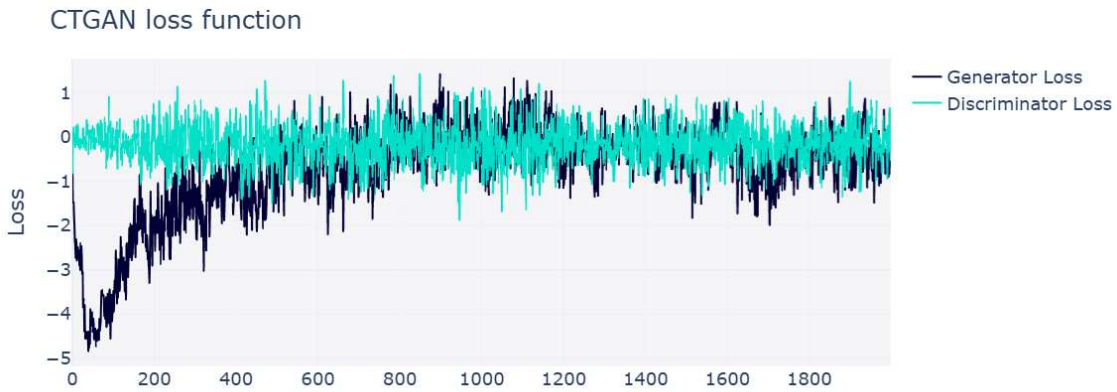


Figure 3 - Training Dynamics of the Generative Adversarial Network (GAN).

Initially, the generator exhibited high loss values (approximately -5), which progressively approached zero over course of 2,000 training epochs, an indication of improved generative performance. In contrast, the discriminator-maintained a near-zero loss throughout training, suggesting a consistent ability to identify synthetic samples. Although loss values alone do not guarantee full convergence, their stabilization over time supports the conclusion that the model has reached a sufficiently trained state for downstream evaluation. It is important to note that both loss curves exhibit stable behavior, a desirable trait in adversarial training processes.

Figure 4 presents a SHAP beeswarm summary plot, where y-axis lists independent variables ranked by their overall contribution to the model's decision-making process. Each point represents the SHAP value of a single observation, with its horizontal position indicating magnitude and direction of variable's impact on predicted outcome. Color gradient, ranging from blue (low feature values) to red (high feature values), provides insight into how original feature value relates to its model influence. Variables are ordered by their mean absolute SHAP value, emphasizing features with a more generalized influence across dataset, while reducing visual impact of outlier contributions.

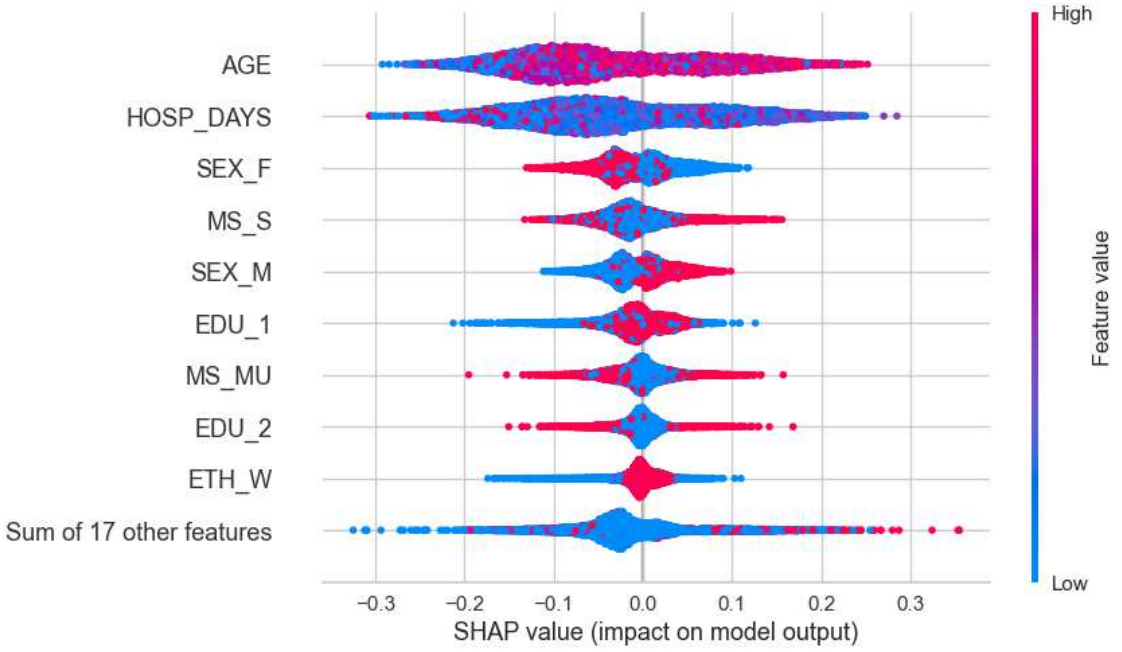


Figure 4 - Global Feature Importance via SHAP Summary Plot.

From Figure 4, it is evident that higher age values tend to be associated with fall predictions, whereas lower age values are more frequently linked to non-fall classifications. A similar trend is observed for SEX_F variable, suggesting that female patients are more strongly associated with reduced fall risk. In contrast, HOSP_DAYS variable, which ranges from 1 to 200 days, exhibits a more complex behavior that cannot be fully interpreted through the SHAP beeswarm plot alone. Although red-colored points (indicating hospital length of stay higher values) show wide dispersion in SHAP values, their direction and magnitude substantially vary, depending on the context, suggesting possible interactions with other features.

To better elucidate this relationship, Figure 5 presents a combined boxplot and swarmplot analysis of HOSP_DAYS, grouped into 10-day intervals according to the total hospital length of stay recorded in the retrospective database. In this representation, the boxplot conveys statistical summaries, including median, quartiles, and outliers, while the swarmplot illustrates the distribution and density of

individual SHAP values. This dual representation shows a higher concentration of patients with total hospitalization durations up to 100 days. More importantly, patients whose total length of stay fell within the 10–59 day range tended to present more positive SHAP values, indicating that these hospitalization-duration profiles contributed more strongly to fall classification by the model.

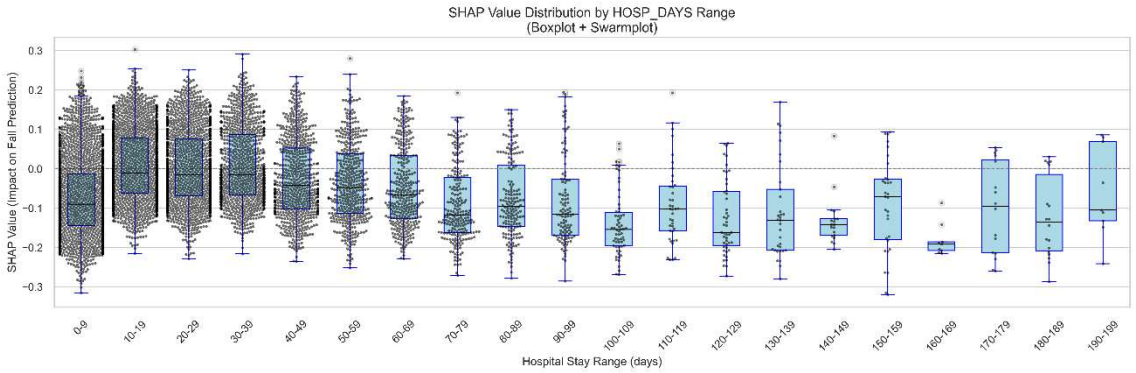


Figure 5. Distribution of SHAP values for HOSP_DAYS across 10-day intervals of total hospital length of stay, using boxplot and swarmplot. The y-axis represents SHAP values for the positive class, corresponding to fall occurrence. Higher SHAP values indicate stronger contributions to fall prediction. The visualization highlights increased SHAP impact among patients whose total hospitalization duration was within the 10–59 day range.

4 Discussion

4.1 Evaluation of classification performance under resampling strategies

In this study, we examine the impact of resampling techniques on enhancing model performance across eight distinct algorithms classification. Given that the dataset contains only two numerical variables and is predominantly composed of categorical variables, commonly used methods such as SMOTE and ADASYN are not suitable, as they can generate inconsistent synthetic values for categorical

features[19,20]. Therefore, we implement resampling strategies that are compatible with this data structure.

As shown in Table 2, Random Forest and Decision Tree classifiers achieved superior performance in terms of balanced accuracy, with marginal gains observed in their tuned versions. The best result was obtained by the Random Forest model combined with random oversampling and hyperparameter tuning, reaching a balanced accuracy of 77%. The baseline Random Forest model, trained without any resampling strategy, achieved a balanced accuracy of 72.8%, which, although lower than that of the best-performing models, still outperformed several other configurations with suboptimal resampling. These results underscore the importance of properly addressing class imbalance when developing predictive models for imbalanced clinical datasets[21-22].

As observed in Figure 2, the Random Forest and Decision Tree models exhibited superior performance both with and without resampling techniques, indicating a lower sensitivity to class imbalance. At the same time, these models benefited from random oversampling, achieving performance gains beyond the baseline configuration. More sophisticated methods, such as SMOTE-NC and CTGAN, demonstrated intermediate performance, with balanced accuracy scores ranging from 71% to 70% among the top 10 evaluated configurations. Despite their algorithmic complexity, these techniques were consistently outperformed by the baseline model, suggesting that, in this specific context, advanced resampling strategies may have introduced noise or overfitting, ultimately leading to a negative impact on model performance.

Unlike traditional resampling techniques, CTGAN is computationally intensive, requiring approximately 10 hours of training to complete 2,000 epochs with the training dataset. This represents a significant increase in processing time and hardware requirements, potentially limiting its applicability in environments with constrained computational resources. In contrast, other resampling methods such as RandomOverSampler and SMOTE-NC generated balanced datasets in less than one minute. This approach also demands careful hyperparameter tuning, further increasing the methodological complexity.

This discrepancy is due to the nature of GAN-based models, which involve training an adversarial network through a minimax optimization game, and demand careful tuning of hyperparameters, a process that adds further complexity and increases computational burden. Despite the stable loss progression observed in Figure 3, CTGAN did not yield superior classification results. This suggests that, while the model was successfully trained in terms of convergence, the synthetic data generated may not have effectively reinforced the decision boundary between classes. Consequently, the inclusion of these samples may have increased class overlap and model confusion, ultimately leading to suboptimal predictive performance.

4.2 Synthetic data generation and evaluation

The use of synthetic data has gained prominence as a practical solution to address the scarcity of real-world clinical datasets, particularly in contexts constrained by privacy and ethical concerns[17]. Rather than being artificial imitations, synthetic datasets are designed to replicate the statistical distributions and structural properties of real data[23].

In the present study, however, CTGAN-based synthetic data did not translate into improved predictive performance. Although CTGAN was able to generate synthetic samples and showed stable adversarial training dynamics, its use as a resampling strategy resulted in lower balanced accuracy than simpler approaches, particularly random oversampling. This finding suggests that the utility of synthetic data should not be assumed from successful model convergence alone, but should be evaluated according to downstream predictive performance, class separability, synthetic data fidelity, and clinical relevance.

Beyond data availability, synthetic data offers notable advantages in terms of cost-effectiveness and mitigation of privacy-related risks. However, complex generative models, such as Generative Adversarial Networks, require continuous evaluation in production settings, as data distribution shifts (data drift) can compromise the representativeness and utility of the generated samples[24].

While synthetic data enhances reproducibility and adaptability in AI-driven healthcare applications, its integration into regulatory and ethical frameworks remains limited. Future research should focus on clearly defining its acceptable use cases, whether for training, simulation, or validation, ensuring that synthetic data contributes to privacy protection and the development of resilient, clinically aligned decision-support systems[17].

Precision medicine, which tailors' treatment to individual genetic and clinical profiles, faces ongoing challenges related to data scarcity, cost, and privacy. Generative AI offers a promising solution by synthesizing realistic, privacy-preserving patient data, with the potential to transform patient-centered care. However, its

519 practical application still requires careful validation to ensure predictive utility in clinical
520 contexts[25].

521 **4.3 Global feature importance and contribution patterns**

522 Figure 4 illustrates that the variables with the greatest influence on the output
523 of the best-performing model were age and hospital length of stay, followed by sex.
524 Within the precision nursing framework, these variables represent routinely available
525 sociodemographic and hospitalization-related information that may support
526 individualized fall-risk assessment and nursing decision-making. The strong influence
527 of age aligns with findings in the literature, which supports that age was a significant
528 factor in the fall[26], patients ≥ 60 years had a significantly higher risk of falling ($p <$
529 0.001)[27] and 78% of the cases of falls with injuries occurred in elderly people ≥ 65
530 years (mean 73 ± 11.6 years)[28]. This consistency reinforces the model's validity,
531 indicating that older individuals are more likely to be classified in the positive class
532 (i.e., fall occurrence).

533 However, when hospitalization time is analyzed in greater detail (Figure 5), a
534 more nuanced pattern emerges. Specifically, patients with hospital stays between 10
535 and 59 days exhibited higher SHAP values, indicating a contribution to fall predictions
536 within this range. This suggests that falls can occur late in the hospitalization period,
537 after the 10th day[29]. This suggests a weakness in the continuity and prevention of
538 nursing care in reassessing the risk of falls throughout hospitalization once extended
539 lengths of stay are more systematically linked to increased fall risk. These findings
540 reinforce the relevance of HOSP_DAYS as a discriminative and context-sensitive
541 feature in fall risk stratification models.

542 It is also worth noting that variables such as marital status and education level
543 appeared among the top 10 most influential features in the SHAP analysis. However,
544 this influence should be interpreted with caution, as SHAP values reflect the
545 magnitude of a variable’s contribution to the model's prediction, not causal
546 relationships. These variables may act as contextual factors influencing fall risk and
547 further investigation is required.

548 Finally, the variable sex revealed a trend indicating that female patients may
549 have a lower likelihood of falling. One possible interpretation is that women could be
550 more adherent to hospital safety protocols and recommendations, thereby
551 contributing to a reduced fall risk. However, further investigation is warranted to
552 explore the behavioral, cultural, or systemic factors that may underline this
553 association. The SHAP interpretability framework not only confirms known clinical
554 associations, such as the role of age, but also enables a more granular understanding
555 of how each variable contributes to individual predictions, including the detection of
556 non-linear or context-dependent effects.

557 In this context, constructing a model based on variables commonly used in
558 everyday nursing practice may support nurses in stratifying patients at higher risk of
559 falling and implementing more targeted measures for older patients, those
560 hospitalized for more than 10 days, and male patients, in order to prevent the
561 occurrence of falls. Although nurses cannot modify these factors, they can enhance
562 situational awareness, as it influences their decisions and actions to prevent falls [30].

563 **4.4 Real-world healthcare data collection issues**

564 The dataset used in this study was not specifically designed for fall risk
565 assessment and did not include clinically targeted variables such as mobility

566 impairment and physiological factors (eg, co-morbidities, low muscular strength,
567 visual impairment, poor reaction time, and movement disorders)[4], nocturia[31-32],
568 medication use[4], validated fall risk scales (e.g., Morse, St Thomas Risk Assessment
569 Tool in Falling Elderly Inpatients - STRATIFY, Medication Fall Risk Score - MFRS)[33-
570 34].

571 The absence of these domain-specific features may have limited the model's
572 ability to capture more nuanced or clinically grounded predictors of in-hospital falls
573 risk. Despite these limitations, it was possible to identify relevant predictors such as
574 age, hospital length of stay, and sex which demonstrated meaningful predictive value
575 from a generic database with data commonly generated in day-to-day clinical nursing
576 practice. This result highlights the potential for developing low-cost and interpretable
577 predictive models using variables that are routinely collected in electronic health
578 records or hospital administrative systems. Such approach is especially
579 advantageous in nursing practice, where the implementation and maintenance of
580 models based on accessible data can facilitate integration into clinical workflows,
581 support early risk identification, and enhance decision-making in resource-
582 constrained environments[35].

583 The literature provides strong evidence of a causal relationship between
584 advanced age and the occurrence of falls in hospital settings, particularly among
585 elderly patients. However, the demographic variables available in the dataset, such
586 as educational level and marital status, although statistically associated with the
587 outcome, may reflect merely correlational rather than causal relationships. This
588 methodological limitation underscores the importance of interpreting results with
589 caution, acknowledging that predictive models based on observational data are
590 effective for forecasting, but do not support direct causal inference.

Because HOSP_DAYS was extracted as the total hospital length of stay in this retrospective dataset, its temporal interpretation is limited. Future prospective models should calculate this variable dynamically at each risk-assessment time point, as the elapsed number of days since admission. This would reduce temporal ambiguity, better reflect real-time nursing decision-making, and clarify how fall risk evolves across hospitalization stages.

5 Conclusion and future work

This study conducted a systematic evaluation of binary classification models combined with resampling strategies to predict in-hospital falls within a data setting characterized by significant class imbalance. Using a general-purpose dataset not specifically designed for fall risk assessment, the models demonstrated a balanced accuracy of 77%, outperforming the baseline model (72.8%), particularly Random Forest combined with random oversampling.

One contribution of this work is to demonstrate that routinely recorded variables, such as age, sex, and hospital length of stay, can provide meaningful predictive information in retrospective fall-risk modeling. However, because hospital length of stay was available as the total duration of the hospitalization episode, its use in prospective clinical decision support should rely on dynamic recalculation at each risk-assessment point. This distinction is important to ensure that future implementations reflect real-time nursing workflows and avoid temporal ambiguity.

In this dataset, CTGAN did not improve predictive performance compared with simpler resampling strategies, particularly random oversampling. Its main value lies in its potential for privacy-preserving simulation and methodological reproducibility. However, its use in nursing decision-support systems requires further validation,

careful assessment of synthetic data quality, evaluation of clinical utility, and consideration of ethical and regulatory requirements for clinical deployment.

The incorporation of explainability through SHAP represents a significant contribution of this study, as it enables identification of specific patterns in the most impactful features. Detailed analysis revealed a pattern, especially in the length of hospital stay variable, and confirmed the relevance of age and sex as the main predictors. Furthermore, it supported a nuanced discussion on the distinction between predictive importance and causal inference, emphasizing that not all influential variables imply a direct causal relationship with the outcome. Transparency is especially critical in clinical contexts, where explainable machine learning models play a key role in fostering trust among healthcare professionals, improving interpretability, and supporting validation and adoption of predictive models in real-world settings. Future work should explore datasets enriched with clinically specific variables, such as medication use, prior falls, and mobility assessments, to further enhance the precision and clinical relevance of predictive models. Additionally, efforts are needed to define regulatory standards and best practices for the responsible use of synthetic data in healthcare, ensuring the development of robust, ethical, and patient-centered AI systems.

List of abbreviations

ACC: Accuracy.

ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning.

AI: Artificial intelligence.

BA: Balanced accuracy.

CI: Confidence interval.

CTGAN: Conditional Tabular Generative Adversarial Network.

DIS_AUT: Dataset variable code for autism spectrum disorder.

DIS_HEA: Dataset variable code for hearing/auditory disability.

DIS_INT: Dataset variable code for intellectual or mental disability.

DIS_MUL: Dataset variable code for multiple disabilities.

DIS_NONE: Dataset variable code for no reported disability.

DIS_OTH: Dataset variable code for other disability types.

DIS_PHY: Dataset variable code for physical disability.

DIS_VIS: Dataset variable code for visual disability.

EDU_0: Dataset variable code for no formal education or illiteracy.

EDU_1: Dataset variable code for primary education.

EDU_2: Dataset variable code for secondary education.

EDU_3: Dataset variable code for higher education.

ETH_BB: Dataset variable code for black or brown race/ethnicity.

ETH_I: Dataset variable code for Indigenous race/ethnicity.

ETH_O: Dataset variable code for other race/ethnicity.

ETH_W: Dataset variable code for white race/ethnicity.

ETH_Y: Dataset variable code for Asian/yellow race/ethnicity.

F1: F1-score.

FALL_EVENT: Dataset target variable code for fall occurrence.

GAN: Generative Adversarial Network.

HOSP_DAYS: Dataset variable code for hospital length of stay in days.

JCI: Joint Commission International.

LightGBM: Light Gradient Boosting Machine.

MFRS: Medication Fall Risk Score.

MS_D: Dataset variable code for divorced marital status.

MS_MU: Dataset variable code for married or stable-union marital status.

MS_O: Dataset variable code for other marital status.

MS_S: Dataset variable code for single marital status.

MS_W: Dataset variable code for widowed marital status.

PRE: Precision.

REC: Recall.

SD: Standard deviation.

SDV: Synthetic Data Vault.

SEX_F: Dataset variable code for female sex.

SEX_M: Dataset variable code for male sex.

SHAP: Shapley Additive Explanations.

SMOTE: Synthetic Minority Over-sampling Technique.

SMOTE-NC: Synthetic Minority Over-sampling Technique for Nominal and Continuous features.

STRATIFY: St Thomas' Risk Assessment Tool in Falling Elderly Inpatients.

TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis – Artificial Intelligence.

XGBoost: eXtreme Gradient Boosting.

Declarations

Ethics approval and consent to participate:

The research received institutional authorization from the setting in which it was conducted and was approved by the Research Ethics Committee of the Conceição Hospital Group (approval No. 6,523,073). The study was conducted in accordance with the ethical principles established in Brazilian National Health Council Resolution No. 466/2012, which provides the regulatory guidelines and standards for research involving human participants, as well as all applicable complementary ethical regulations. The requirement for informed consent was waived by the Research Ethics Committee because the study used previously anonymized secondary data. No formal study protocol was prospectively registered for this prediction model development.

Consent for publication:

Not applicable.

Availability of data and materials:

The patient-level data used and/or analyzed during the current study are not publicly available due to institutional privacy and confidentiality agreements. Data may be available from the corresponding author upon reasonable request and with permission from the institutional data holder and the relevant ethics authority, where applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was supported by the National Council for Scientific and Technological Development – CNPq, the Ministry of Health – MS, and the National Program for Genomics and Precision Health – Genomas Brasil, under Grant/Process No. 444274/2023-5. The funders had no role in the study design, data collection, data analysis, interpretation of results, manuscript preparation, or decision to submit the manuscript for publication.

Authors' contributions:

GCF, LMO, CBS, ALE, and EL contributed to the conception and design of the study. GCF, CBS, ALE, and EL contributed to the clinical and nursing interpretation of the data. LMO contributed to data preprocessing, machine learning modeling, resampling analysis, and model interpretability. GCF and LMO drafted the initial version of the manuscript. All authors contributed to data interpretation, critically revised the manuscript for important intellectual content, read and approved the final manuscript.

Acknowledgements:

Not applicable.

REFERENCES

1. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med. 1988;319(26):1701-7. doi:10.1056/NEJM198812293192604
2. Appeadu MK, Bordoni B. Falls and Fall Prevention in Older Adults. 2023 Jun 4. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan. PMID: 32809596. https://www.ncbi.nlm.nih.gov/books/NBK560761/
3. Joint Commission International. International patient safety goals [Internet]. 2024 [cited 2025 Jul 26]. Available from: https://www.jointcommissioninternational.org/standards/international-patient-safety-goals
4. Locklear T, Kontos J, Brock CA, Holland AB, Hemsath R, Deal A, et al. Inpatient falls: epidemiology, risk assessment, and prevention measures. A narrative review. HCA Healthc J Med. 2024;5(5):517-25. doi:10.36518/2689-0216.1982. PMID:39524960
5. Souza AB, Maestri RN, Rohsig V, Lorenzini E, Alves BM, Oliveira D, et al. In-hospital falls in a large hospital in the south of Brazil: a 6-year retrospective study. Appl Nurs Res. 2019;48:100770. doi:10.1016/j.apnr.2019.05.017
6. Galvão YM, Portela L, Barros P, Fagundes RAA, Fernandes BJT. OneFall-GAN: A one-class GAN framework applied to fall detection. Eng Sci Technol Int J. 2022;35:101227. doi:10.1016/j.jestch.2022.101227
7. Gurcan F, Soylu A. Learning from imbalanced data: Integration of advanced resampling techniques and machine learning models for enhanced cancer diagnosis and prognosis. Cancers (Basel). 2024;16(19):3417. doi:10.3390/cancers16193417

8. Yuan X, Xie L, Abouelenien M. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recognit. 2018;77:160-72. doi:10.1016/j.patcog.2017.12.017
9. Liu H, Liu Z, Jia W, Zhang D, Tan J. A novel imbalanced data classification method based on weakly supervised learning for fault diagnosis. IEEE Trans Ind Inform. 2022;18(3):1583-92. doi:10.1109/TII.2021.3084132
10. Walsh R, Tardy M. A comparison of techniques for class imbalance in deep learning classification of breast cancer. Diagnostics (Basel). 2023;13(1):67. doi:10.3390/diagnostics13010067
11. Khan AA, Chaudhari O, Chandra R. A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. Expert Syst Appl. 2024;244:122778. doi:10.1016/j.eswa.2023.122778
12. Fabrizzio GC, de Oliveira LM, Da Costa DG, Santos JLGD, Jensen R, Lorenzini E, et al. Comparison of machine learning algorithms for predicting admission in intensive care units based on the precision nursing framework. J Med Artif Intell. 2025;8:44. doi:10.21037/jmai-24-296
13. Ghebrehiwet I, Zaki N, Damseh R, et al. Revolucionando a medicina personalizada com IA generativa: uma revisão sistemática. Artif Intell Rev. 2024;57:128. doi:10.1007/s10462-024-10768-5
14. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res. 2017;18(17):1-5. Available from: https://jmlr.org/papers/v18/16-365.html

15. Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. Modeling tabular data using conditional GAN. In: Adv Neural Inf Process Syst. 2019;32:1-11. Available from: https://papers.nips.cc/paper_files/paper/2019/file/254ed7d2de3b23ab10936522dd547b78-Paper.pdf
16. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, editors. Adv Neural Inf Process Syst. 2014;27. Available from: https://proceedings.neurips.cc/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf
17. Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng. 2021;5(6):493-7. doi:10.1038/s41551-021-00751-8
18. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Adv Neural Inf Process Syst. 2017;30:4765–74. Available from: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
19. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-57. doi:10.1613/jair.953
20. Husain G, Nasef D, Jose R, Mayer J, Bekbolatova M, Devine T, et al. SMOTE vs. SMOTEENN: A study on the performance of resampling algorithms for addressing class imbalance in regression models. Algorithms. 2025;18(1):37. doi:10.3390/a18010037

21. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE Int Joint Conf Neural Netw (IJCNN); 2008. p.1322-8.
22. Kang CW, Yan ZK, Tian JL, Pu XB, Wu LX. Constructing a fall risk prediction model for hospitalized patients using machine learning. BMC Public Health. 2025;25:242. doi:10.1186/s12889-025-21284-8
23. Patki N, Wedge R, Veeramachaneni K. The synthetic data vault. In: IEEE Int Conf Data Sci Adv Anal (DSAA); 2016. p.399–410. doi:10.1109/DSAA.2016.49
24. ellermann J, Lessmann S. Leveraging image-based generative adversarial networks for time series generation. 2023. doi:10.48550/arXiv.2112.08060
25. Goyal M, Mahmound QH. A systematic review on synthetic tabular data generation using generative artificial intelligence. Electronics 2024; 13:3509. doi:10.3390/electronics13173509.
26. Kim J, Lee E, Lee S. Multilevel factors influencing falls of older patients in hospital: the risk of hospital-induced delirium and the impact of nurse staffing. Geriatr Nurs. 2025;61(1):13–9. doi:10.1016/j.gerinurse.2024.10.077
27. Falcão RMM, Costa KNFM, Fernandes MGM, Pontes MLF, Vasconcelos JMB, Oliveira JS. Risk of falls in hospitalized elderly people. Rev Gaúcha Enferm. 2019;40(esp):e20180266. doi:10.1590/1983-1447.2019.20180266
28. Luzia MF, Prates CG, Bombardelli CF, Adorna JB, Moura GMSS. Characteristics of damage falls in hospitalized patients. Rev Gaúcha Enferm. 2019;40(esp):e20180307. doi:10.1590/1983-1447.2019.20180307
29. Satoh M, Miura T, Shimada T, Hamazaki T. Risk stratification for early and late falls in acute care settings. J Clin Nurs. 2023;32(3–4):494–505. doi:10.1111/jocn.16267. PMID:35224808

30. Takase M. Falls as the result of interplay between nurses, patient and the environment: using text-mining to uncover how and why falls happen. Int J Nurs Sci. 2022;10(1):30–7. doi:10.1016/j.ijnss.2022.12.003. PMID:36860705
31. Pesonen JS, Vernooij RWM, Cartwright R, Aoki Y, Agarwal A, Mangera A, et al. The impact of nocturia on falls and fractures: a systematic review and meta-analysis. J Urol. 2020;203(4):674–83. doi:10.1097/JU.0000000000000459
32. Kim SY, Bang W, Kim MS, Park B, Kim JH, Choi HG. Nocturia is associated with slipping and falling. PLoS One. 2017;12(1):e0169690. doi:10.1371/journal.pone.0169690
33. Morse JM, Black C, Oberle K, Donahue P. A prospective study to identify the fall-prone patient. Soc Sci Med. 1989;28(1):81–6. doi:10.1016/0277-9536(89)90309-2
34. Souza LM, Ramos SM, Isoppo CS, Diniz AM, Ferreira CI, Sousa P. Aplicação da Medication Fall Risk Score em doentes com notificação de incidente de queda: análise retrospectiva. Rev Enferm Referência. 2023;6(2):e22087. doi:10.12707/RVI22087
35. Parsons R, Blythe R, Cramb S, Abdel-Hafez A, McPhail S. An electronic medical record–based prognostic model for inpatient falls: development and internal-external cross-validation. J Med Internet Res. 2024;26:e59634. doi:10.2196/59634