Buscarini et al. Journal of NeuroEngineering and Rehabilitation (2025) 22:117
https://doi.org/10.1186/s12984-025-01654-4

Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

Enhancing patient rehabilitation outcomes: artificial intelligence-driven predictive modeling for home discharge in neurological and orthopedic conditions

Leonardo Buscarini[1], Paola Romano[1*], Elena Sofia Cocco[2], Carlo Damiani[3], Sanaz Pournajaf[2,4], Marco Franceschini[2] and Francesco Infarinato[1]

Abstract

In recent years, the fusion of the medical and computer science domains has gained significant traction in the scientific research landscape. Progress in both fields has enabled the generation of a vast amount of data used for making predictions and identifying interesting clusters and pathways. The Machine Learning (ML) model's application in the medical domain is one of the most compelling and challenging topics to explore, bridging the gap between Artificial Intelligence (AI) and healthcare. The combination of AI and medical information offers the possibility to create tools that can benefit both healthcare providers and physicians. This enables the enhancement of rehabilitation therapy and patient care. In the rehabilitation context, this work provides an alternative perspective: prediction of patients' home discharge upon completing the rehabilitation protocol. Demographic and clinical data were collected on 7282 inpatients from electronic Medical Record, each record was categorized into Neurological Patients (NP, $N=3222$) or Orthopedic Patients (OP, $N=4060$). To identify the most suitable machine learning model, an extensive data preprocessing phase was conducted. This process involved variables recoding, scaling, and the evaluation of different dataset balancing methods to optimize model performance. Following a thorough review and comparison of algorithms commonly employed in the clinical-rehabilitative field, the Random Over Sampling (ROS) technique, in combination with the Random Forest (RF) machine learning model, was selected. Subsequently, a comprehensive hyperparameter tuning phase was performed using a grid search approach. The optimized model achieved an average accuracy of 98% for OP and 96% for NP, based on 10-fold cross-validation applied to the balanced training set (unrealistic scenario). When tested on the unbalanced dataset (real-world condition), the RF model maintained strong generalization performance, achieving 90% accuracy for OP and 83% for NP. This work points out the increasing importance of AI in medicine, especially in the realm of personalized rehabilitation. The use of such approaches could signify a transformative shift in healthcare. The integration of machine learning not only enhances the precision of treatment but also opens new possibilities for patient-centered care, improving outcomes and quality of care for individuals undergoing rehabilitation.

*Correspondence:
Paola Romano
paola.romano@sanraffaele.it

Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords Personalized rehabilitation, Machine learning, Classification algorithm, Random forest, Unbalance, Hyperparameters optimization, Accuracy

Introduction

In the global population, approximately one in seven individuals is forced to live with disabilities every day. This rate is on the verge of increasing, primarily due to the high incidence of chronic diseases and the demographic shift in Western countries [1]. For this reason, rehabilitation is considered a priority of the 21st century for healthcare systems, as well as disease treatment and prevention [2]. From a patient's perspective, it's very important to know whether he or she will return to independently perform Activities of Daily Living (ADL) such as dressing, eating, or walking. In recent years, the concept of personalized rehabilitation has incredible development; nowadays, it's considered a personalized multimodal process aimed at improving patient autonomy [3].

Prediction in rehabilitation outcomes using AI is a promising area of research. Several studies discuss the potential of AI in rehabilitation, including its use in assisting rehabilitation sessions, evaluating treatment progress, and providing prognosis regarding the risk of complications or treatment success. These advancements in AI have the potential to improve decision-making, develop precision medicine tools, and optimize rehabilitation programs [4–9]. Determining the rehabilitative protocol to use, depending on the patient's medical history and clinical features, is a fundamental purpose. Consequently, recent clinical studies have focused on identifying specific targets that allow us to understand whether the chosen protocol is the most suitable for the patient. Positive outcomes have been achieved using parameters like the modified Barthel Index (mBI) at discharge or the Length of Stay (LOS) [10–15]. The leitmotifs of this work are personalized rehabilitation and management of human and non-human resources in healthcare facilities. It integrates clinical information with the remarkable potential of Artificial Intelligence (AI), particularly within its branch known as Machine Learning (ML). The latter is primarily portrayed as the field that allows computers to learn how to make predictions without the need for explicit programming [16–20]. The concept of a learning machine capable of making predictions and drawing conclusions that are difficult to reach with conventional statistical methods is an old idea. However, statistics and ML are not completely distinct; ML algorithms are built upon common statistical methods, and they continue to advance alongside AI development. Statistics focuses on the verification of a hypothesis, whether null (H_0) or alternative (H_1), to assess how well the data distribution fits other known models, such as the Gaussian one. On the other hand, the aim of ML is prediction through algorithms that attempt to be as generalizable and applicable as possible for new validation datasets [21].

In this context, our work goal is to provide clinicians and healthcare professionals with a tool capable of estimating the likelihood of patient home discharge at the end of rehabilitation based on individual characteristics. So far, literature provides recent studies that demonstrate the efficacy and the reliability of AI algorithms in the rehabilitation outcomes analysis. Santilli et al. [6] leveraged a 3-year dataset (4050 patients) comprising Acceptance and Discharge Report of Rehabilitation (ADR-r) data to predict mBI at discharge, exploiting an in-built function for importance ranking. Concerning the Intensive Care Unit (ICU), Safaei et al. [22] proposed a CatBoost model to predict patient mortality using the discharge status variable and ten other impactful characteristics. Rufo et al. [23] employed a Light Gradient Boosting Machine (LightGBM) algorithm and other significant classification models, including Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), to predict the condition of diabetes mellitus with reliable results. In recent years, heightened stress on healthcare management, especially during the COVID-19 pandemic, has led to the publication of several works on prediction analysis. Yan et al. [24] utilized 485 patient blood samples to assess three important biomarkers for quickly predicting mortality risk in COVID-19 individuals. Furthermore, Park et al. [25] developed an XGBoost model to identify optimal candidates for robotic-assisted gait rehabilitation in patients with subacute stroke, highlighting the critical role of functional clinical features, e.g. the mBI, in predicting motor recovery following robotic rehabilitation interventions. Similarly, using an XGBoost model, Fan et al. [26] analyzed patients' preferences to optimize rehabilitation management, demonstrating how clinical and logistical factors (e.g. distance to facilities and hospital stay duration) influence treatment decisions.

Drawing on the classification and regression ML models employed in these studies, our contribution was to increase the literature knowledge about the impact of rehabilitation therapies on orthopedic and neurological patients in terms of home discharge rate after the treatment period. Our investigation aimed to conduct an intensive and in-depth preprocessing phase on 8468 ADR-r data, addressing the issue of unbalanced distribution in the home discharging target variable. We confirmed previous research findings regarding the importance of certain characteristics in patient autonomy recovery [6] and provided an accurate clinical model built on a massive input dataset to predict discharging status that could help healthcare management systems. The implemented prototyping tool could be used to explore new and interesting targets for further analysis. For better readability and clarity, we have structured the paper as follows: in Sect. Materials and methods, we describe the materials and methods used in this study, including data collection, a detailed preprocessing phase, and a brief overview of the applied machine learning techniques. Section Results presents the results, highlighting the predictive performance of various ML models in both balanced and real-world scenarios, as well as the significant impact of different resampling methods. In Sect. Discussion, we discuss the findings, their implications, and the potential limitations of the study. Finally, Sect. Conclusions concludes by summarizing the key contributions and suggesting future directions to validate our findings and guide further work.

Materials and methods

The study protocol was approved by the ethical committee of the IRCCS San Raffaele Pisana of Rome on 18/07/2018 (code number 07/18).

Original dataset and initial cleaning steps

The original analysis dataset includes clinical and demographic data of adults admitted to the neurology and orthopedic departments of a rehabilitation hospital in Italy from January 2015 to August 2022. The completion of patients' ADR-r form [27] resulted in the collection of data for 10,520 individuals, whose information is distributed across 120 initial features. The dataset size reduced to 8468 after the duplicate removal (2052). We anonymized dataset rows by assigning a unique ID to each patient and collected personal information about them, such as gender, age, marital status, date, place of birth, and more. Clinical data, including the primary reason for rehabilitation, any associated medical conditions, impairments, and admission/discharge mBI scores were collected.

Dataset cleaning and encoding

Underlying medical conditions were defined establishing 13 general macro-categories. Through categorization based on *Basic Pathology*, the dataset has been divided into two subsets: Neurologic Patients (NP), related to individuals whose basic pathology is primarily neurological, and Orthopedic Patients (OP), which pertains to orthopedic basic pathology. An additional macro-category called *Not Attributable* didn't belong to any of the identified 13 macro-categories and wasn't considered in further analysis. Details are shown in Table 1.

The *Basic Pathology* consists of alphanumeric codes established by ICD-9-CM, which is the ninth revision of this classification system with clinical modifications [27].

Most of the used algorithms are unable to interpret a variable with values represented as text strings. This is what happens in the *Basic Pathology* variable; for this reason, a numeric encoding was necessary to ensure accurate predictions [28], the one-hot encoding technique, commonly used in ML prediction analyses that involve categorical features, was applied; its application converts text strings into n binary numeric vectors, where n is the macro-categories amount: (i) 1 if the categorical feature corresponds to the vector label, and (ii) 0 in the other cases [29].

The same encoding method was applied to the variables related to the presence of any patient comorbidity in 17 macro-categories, as shown in Table 1.

Regarding the impairment's variables, following the ADR-r form the possible values range from 0 to 9. The smaller the number, the lower the degree of impairment shown by the patient. We applied a different encoding technique called label encoding, resulting in modified categorical impairment variables transformed into numerical ones [28]. The labels used in this work are: (i) 0, (ii) 1, and (iii) 2, which are respectively associated with the absence, presence, and not-evaluability of patient impairment.

A fundamental step in the data preprocessing phase was the definition of the target variable. Keeping in mind that our analysis aimed to create a model that allows the prediction of patient-specific home discharge, we defined the Discharge Type Category based on the Discharge Type Feature in the ADR-r form. To this purpose we used four labels to encode the categorization of patient discharge, assigning to each label a specific condition: (i) 1 - the patient is home discharged at the end of rehabilitation, (ii) 2– the patient is not home discharged at the end of rehabilitation, (iii) Voluntary Discharge– the patient voluntarily leaves the clinic, (iv) Death– the patient deceased during rehabilitation protocol.

Finally, we created the final target variable for our analysis, referred to as Home Discharge (HD). Applying one-hot encoding, we assigned the value 1 when the patient falls under label 1 of the Discharge Type Category, indicating the patient is home discharged. On the other hand, 0 is assigned to patients corresponding to the label 2, indicating that they will not home discharge at the end of rehabilitation. Cases related to other labels were not considered in the further analysis (voluntary discharge and death).

The mBI is a well-established patient-centered clinical outcome measure that quantifies a patient's disability, ranging from 0 to 100; this value derives from the sum of scores assigned by the clinician to each of the scale

Table 1 Summary of input and target variables for ML models with corresponding encoded labels in round brackets

Variable Name	Variable Type	Value Range	Variable Group
Patient ID	Q	[1-8468]	NP/OP
Gender	C	[female (0), male (1)]	NP/OP
Stroke	C	[A (0), P (1)]	NP
Parkinson	C	[A (0), P (1)]	NP
Multiple Sclerosis	C	[A (0), P (1)]	NP
Brain Tumors	C	[A (0), P (1)]	NP
Post mild/moderate trauma	C	[A (0), P (1)]	NP
Other neurological pathologies	C	[A (0), P (1)]	NP
Non-traumatic myeloradiculopathies	C	[A (0), P (1)]	NP
Hip arthroplasty	C	[A (0), P (1)]	OP
Knee arthroplasty	C	[A (0), P (1)]	OP
Femur osteosynthesis	C	[A (0), P (1)]	OP
Amputation	C	[A (0), P (1)]	OP
Spinal pathologies	C	[A (0), P (1)]	OP
Other orthopedic pathologies	C	[A (0), P (1)]	OP
Other types of comorbidities	C	[A (0), P (1)]	NP/OP
Hematological	C	[A (0), P (1)]	NP/OP
Dysmetabolic	C	[A (0), P (1)]	NP/OP
Hepatic	C	[A (0), P (1)]	NP/OP
Neurological	C	[A (0), P (1)]	NP/OP
Respiratory	C	[A (0), P (1)]	NP/OP
Diabetics	C	[A (0), P (1)]	NP/OP
Hypertension	C	[A (0), P (1)]	NP/OP
Heart diseases	C	[A (0), P (1)]	NP/OP
Cardiac arrhythmias	C	[A (0), P (1)]	NP/OP
Dyslipidemic	C	[A (0), P (1)]	NP/OP
Tumors	C	[A (0), P (1)]	NP/OP
Rheumatology/Orthopedics	C	[A (0), P (1)]	NP/OP
Circulatory complications	C	[A (0), P (1)]	NP/OP
Intestinal complications	C	[A (0), P (1)]	NP/OP
Kidney and urinary tract complications	C	[A (0), P (1)]	NP/OP
Different types of comorbidities	C	[A (0), P (1)]	NP/OP
Cognitive impairment	C	[A (0), P (1), NE (2)]	NP/OP
Behavior impairment	C	[A (0), P (1), NE (2)]	NP/OP
Communication/language impairment	C	[A (0), P (1), NE (2)]	NP/OP
Sensory impairment	C	[A (0), P (1), NE (2)]	NP/OP
Manipulation impairment	C	[A (0), P (1), NE (2)]	NP/OP
Balance impairment	C	[A (0), P (1), NE (2)]	NP/OP
Locomotion impairment	C	[A (0), P (1), NE (2)]	NP/OP
Cardiovascular impairment	C	[A (0), P (1), NE (2)]	NP/OP
Respiratory system impairment	C	[A (0), P (1), NE (2)]	NP/OP
Ulcers	C	[A (0), P (1)]	NP/OP
Sphincter control impairment	C	[A (0), P (1), NE (2)]	NP/OP
Urinary system impairment	C	[A (0), P (1), NE (2)]	NP/OP
Nutrition impairment	C	[A (0), P (1), NE (2)]	NP/OP
Categorized mBI admission	C	[1 (CD) – 6 (I)]	NP/OP
Age	Q	[18–97]	NP/OP
Target variable: Home discharge	C	[NHD (0), HD (1)]	NP/OP

Q: quantitative variables; C: categorical variables; A: absence of pathology/impairment; P: presence of pathology/impairment; NE: not-evaluability; CD: complete dependence; I: independence; NHD: not home discharge; HD: home discharge subfields [13]. It is commonly administrated in rehabilitation settings to evaluate the functional status of patients at admission and discharge [30]. In particular, it assesses a patient's independence in performing ADL [31, 32]. Label encoding technique divided the collected continuous variable into classes to better perform the model training. Based on previous studies [33, 34], we created six classes corresponding to the degree of dependency exhibited by the patient (Table 2). The mBI henceforth will be referred to as "Categorized mBI" in the analysis.

Before the final cleaning step, the dataset consisted of 8468 patients (rows) and the new 50 categorical and non-categorical variables (columns) relevant to our study to schedule the demographic and clinical characteristics of patients admitted to the clinic. Table 1 provides an overview of the encoded input features used for prediction models.

Analysis dataset

To address the presence of missing values and the dataset integrity, we didn't consider patients for further analysis when: the admission and/or discharge mBI was not recorded in the ADR-r form; Discharge Type Category variable returned Voluntary Discharge or Death; ICD-9-CM code in ADR-r form wasn't registered; patient pathology was not attributable into identified categories for the *Basic Pathology*. Figure 1 provides an overview of the analysis dataset building.

The definitive analysis dataset consists of 7282 patients (rows): 3222 were admitted for a neurological basic pathology and 4060 for orthopedic conditions. Table 3 shows substantial dataset details for NP and OP groups.

Table 4 provides a detailed overview of the datasets used in this study alongside those reported in recent literature. It emphasizes the strengths of our dataset, particularly in terms of data granularity and population-level detail, when compared to key features of datasets commonly employed in clinical machine learning research. By highlighting often underreported aspects, such as the number of duplicate records, the count of excluded patients, and detailed age distributions, the table reinforces our study's contribution and commitment to transparency in addressing current limitations within the clinical ML landscape.

Table 2 Categories and labels of mBI at admission and discharge

ADR-*r* value	Label	Criteria	Level of dependence
mBI on admission or mBI at discharge	1	mBI ∈ [0, 24]	Complete
	2	mBI ∈ [25, 49]	Serious
	3	mBI ∈ [50, 74]	Moderate
	4	mBI ∈ [75, 90]	Mild
	5	mBI ∈ [91, 99]	Minimal
	6	mBI = 100	Indipendent

Supervised learning algorithms

In this study, we tested the supervised learning models. The goal was to implement a predictive function that can map the features exhibited at the time of a patient's admission (X) to a target variable (y). This is achieved using a training set consisting of tuples formed by the output variable and the i-th patient features [35]. Generally, the supervised ML domain is employed to solve classification and regression problems [36]. When the target variable is categorical, we refer to it as classification; on the other hand, a quantitative target variable is used in regression algorithms. The principal aim was to predict label 1 of the home discharge variable relied on the individual's characteristics observed at admission. We trained and tested five of the most widely used classification algorithms in the rehabilitation field: some of them relied on decision trees, such as Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), as well as others like Gradient Boosting (GB), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost). The dataset was partitioned into NP and OP, which were further divided into training (70%) and test (30%) sets. The preprocessing phase, discussed in Sect. Feature scaling and Balancing methods, as well as the prediction step, was carried out using Python 3.10.12 and proper libraries, including SciKit-Learn (for model learning) [37], Imbalanced-Learn (for dataset imbalance issues), Matplotlib (for graphical visualization), SciPy (for statistical tests), Numpy (for mathematical operations) [38] and Pandas (for data manipulation and analysis). We improved the efficiency of the model and, consequently, prediction accuracy through an optimization phase of hyper-parameters using a grid search technique. Hyperparameters are specific variables of models that can be configured during the training step; their tuning (HPT) influences model performance [39]. We chose to use the grid search because it exhaustively investigates all hyperparameter combinations, returning their best values. However, it's worth noting that this search process is computationally and temporally demanding [40].

During the splitting step, there is a risk of making a significant mistake. It is common to train the model with a dataset that may not allow it to take full advantage of the maximum available information. This can potentially hinder the achievement of the most accurate prediction. Randomly, a patient may be assigned to either the training or the test set; however, if we are in the latter condition, their information will not be used for model learning. This approach can lead to a high risk of overfitting: a model exhibits high accuracy in predictions on the training set but performs poorly on the test set. A

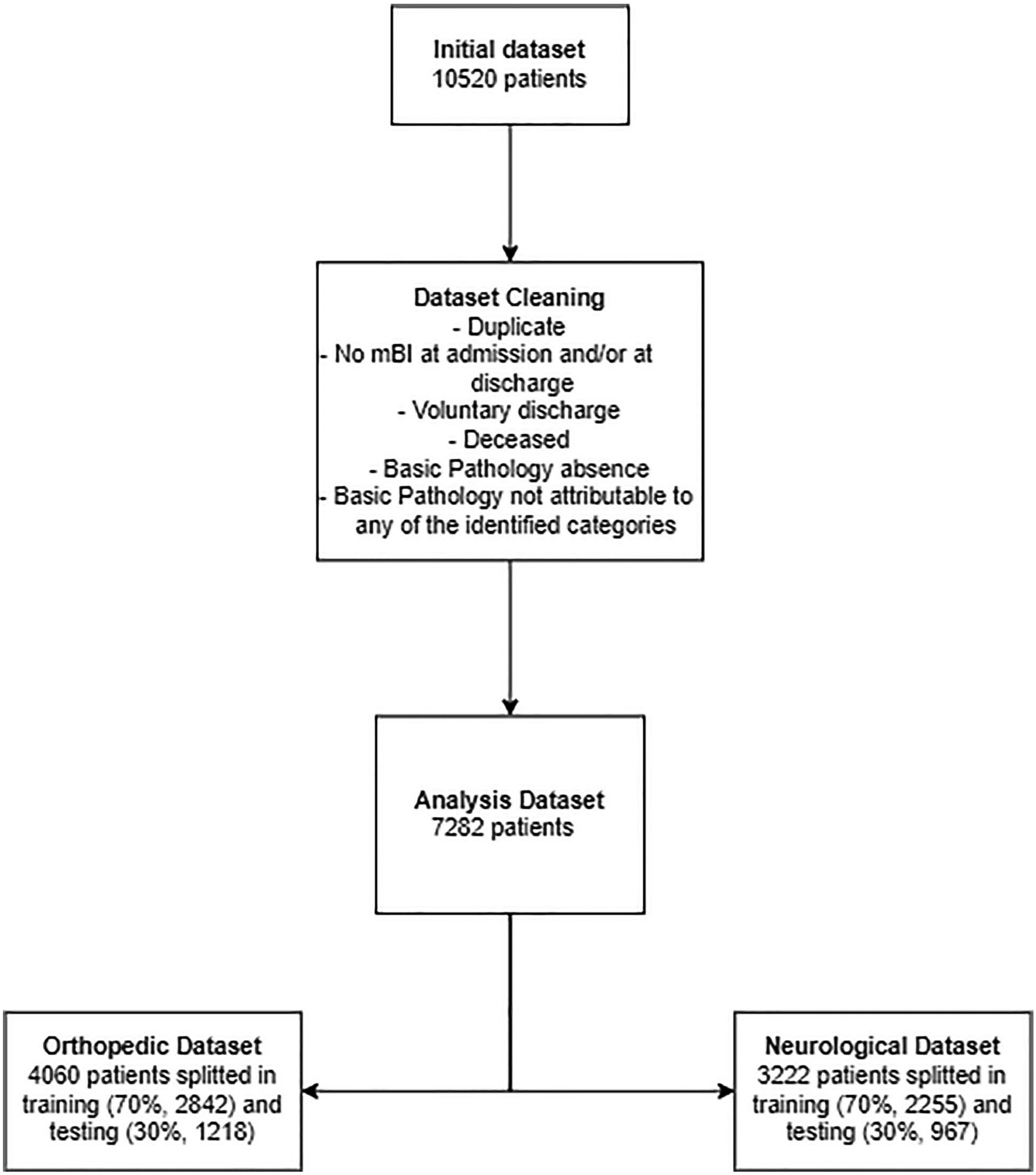

Fig. 1 Flowchart detailing the data cleaning and splitting process

solution to address this issue is the k-fold Cross-Validation (CV) method. It involves dividing the dataset into k folds, where k is a positive integer: k-1 folds are used for model training, and one-fold is reserved for testing. In this analysis, we chose k = 10, creating 10 distinct subsets from the balanced training set, each used as the test set in one iteration. The final accuracy of the model was then obtained by averaging the accuracies from these 10 iterations [40]. This process also makes it possible to provide approximate model validation when it is not possible to give input to a dataset never considered in the training and testing phase in a tight time frame.

Feature scaling

The feature scaling step has a significant impact on the quality of predictions. Its application ensures that

Table 3 Demographic insights on model input patient

Demographic Data	NP	OP	Total
Sample size	3222	4060	7282 (44% NP; 56% OP)
Mean age (± St. Dev.) (years)	73 (± 13)	74 (± 11)	74 (± 12)
Median mBI at admission (± St. Dev.)	30 (± 13)	40 (± 9)	37 (± 12)
Median mBI at discharge (± St. Dev.)	73 (± 26)	91 (± 18)	86 (± 23)
Median mBI change (± St. Dev.)	41 (± 20)	48 (± 14)	46 (± 17)
Gender: Male (%)	1715 (53%)	1399 (34%)	3114 (43%)
Gender: Female (%)	1507 (47%)	2661 (66%)	4168 (57%)
Home discharge ratio (%)	2742 (85%)	3800 (94%)	6542 (90%)

variables distributed in large value ranges do not dominate features distributed in smaller intervals [41], to guarantee the latter mentioned be examined by the model [28]. There are two main techniques for feature scaling: Standard Scaling and Min-Max Scaling.

In this work, we used the Min-Max Scaling technique, which involves applying a formula to the variables' values to be scaled:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X is the original value of the variable to be scaled, X_{min} is the minimum value, X_{max} is the maximum value and X_{norm} is the new value obtained after the application of this scaling method. This technique ensures that the distribution range of the scaled variable is reduced to [0, +1] if it shows only positive values or to [-1, +1] if both positive and negative values are present. This technique is powerful when Standard Scaling is not suitable, such as when the variable to be scaled doesn't fit a Gaussian curve [42], as in the case of the variable Age in our study. Section Non-normality verification shows the results of non-normality tests applied to this variable.

Balancing methods

The last preprocessing step involves the use of balancing techniques to address the unbalanced distribution of the target variable: 0 (15%) vs. 1 (85%) for the NP set; 0 (6%) vs. 1 (94%) for the OP set. This study shows a binary classification analysis; in this case, it's common to observe a high unbalance of target variable distribution. This condition adversely affects the predictive performance of the model [28]. Not solving this problem results in an algorithm that tends to accurately predict the most common classes, at the expense of the rare ones [43, 44]. Proposed solutions in the scientific landscape are mainly based on resampling techniques. Usually, in ML analysis we observe the low number of minority classes rather than the majority class abundance. For this reason, in medical domain, oversampling (synthesis) is preferred to undersampling (reduction) [45, 46]. This study compared the main methodologies of both approaches, including Random Over Sampling (ROS) [44], Synthetic Minority Oversampling TEchnique (SMOTE) [47], ADAptive SYNthetic (ADASYN) [48], as well as Random Under Sampling (RUS) [49] and Cluster Centroids (CC) [50]. In addition, the SMOTE-Tomek technique, which combines both methodologies simultaneously, was also considered [51].

Evaluation metrics

Once the model is implemented, we must establish the effectiveness of its performance. Since this is a classification task, the main evaluation metrics used to compare predictions are Accuracy, Precision, Recall and F1-score [52]. The possible range of these values is [0, 1]: the closer to the right extreme, the better the model's ability to correctly predict instances belonging to both *Home discharge* classes.

Results

Figure 1 is a graphical representation of ML input datasets construction, along with their absolute and percentage frequencies. Duplicate entries, absence of admission and/or discharge mBI scores, and other criteria were considered in constructing the analysis dataset.

Non-normality verification

We needed to conduct some statistical checks to determine the appropriate technique for scaling the *Age* quantitative variable. These checks were performed to confirm statistically the non-normal distribution, justifying the subsequent application of the Min-Max Scaling technique. In addition to visual inspection (Fig. 2) and a comparison of nominal percentiles (Fig. 3), two statistical tests were also applied for a quantitative assessment.

The Shapiro-Wilk test is properly employed for small sample sizes but remains effective for larger datasets. The Kolmogorov-Smirnov test is primarily applicable to larger datasets. In both checks, we chose a significance level (alpha) of 0.05. Table 5 shows the results for the *Age* variable in NP and OP. For this reason, we accepted the alternative hypothesis: *Age* variable for both datasets doesn't fit a normal distribution.

Quantitative assessments justify the application of the Min-Max Scaler to this variable, helping to avoid problems that could lead to less accurate predictions.

Table 4 Dataset key characteristics used in research studies employing ML in the clinical field in recent years

Research Study	Geographical coverage	Diagnosis Category	Main application	Type of data	Coverage Year(s)	Sample size	Duplicates	Excluded Patients	Number of features	Age range
Current	Italy	Neurologic Orthopedic	Rehabilitation	Continuous Categorical	2015–2022	7282 (10520 Original Dataset)	2052	1186	50	73 ± 13(NP) 74 ± 11(OP)
Fan et al. (2025), *Front Artif Intell* [26]	China	Neurologic Orthopedic	Rehabilitation	Continuous Categorical	-	4244	-	Not numerically specified	80	-
Park et al. (2024), *NeuroRehabilitation* [25]	South Korea	Neurologic	Stroke Rehabilitation	Continuous Categorical	2022–2023	187	-	Not numerically specified	18	64 ± 13
Santilli et al. (2023), *Int J Environ Res Public Health* [6]	Italy	Neurologic Orthopedic	Rehabilitation	Continuous Categorical	2015–2018	3421 (4050 Original dataset)	-	629	19	69 ± 13 (NP) 72 ± 11 (OP)
Safaei et al. (2022), *PLoS One* [22]	U.S. Hospitals (208 Units)	Cardiovascular, Pulmonary, Neurologic and others	Intensive Care Unit Management	Continuous Categorical	2014–2015	over 200,000	-	Not numerically specified	55	-
Rufo et al. (2021), *Diagnostics* [23]	Ethiopia	Endocrine	Diabetes diagnosis	Continuous Categorical	-	2109	-	Not numerically specified	23	-
Yan et al. (2020), *Nat Mach Intell* [24]	Wuhan, China	Infectious Diseases	COVID-19 Mortality	Continuous Categorical	2020	375	-	Not numerically specified	75	59 ± 16

OP Orthopedic Patients; NP Neurologic Patients

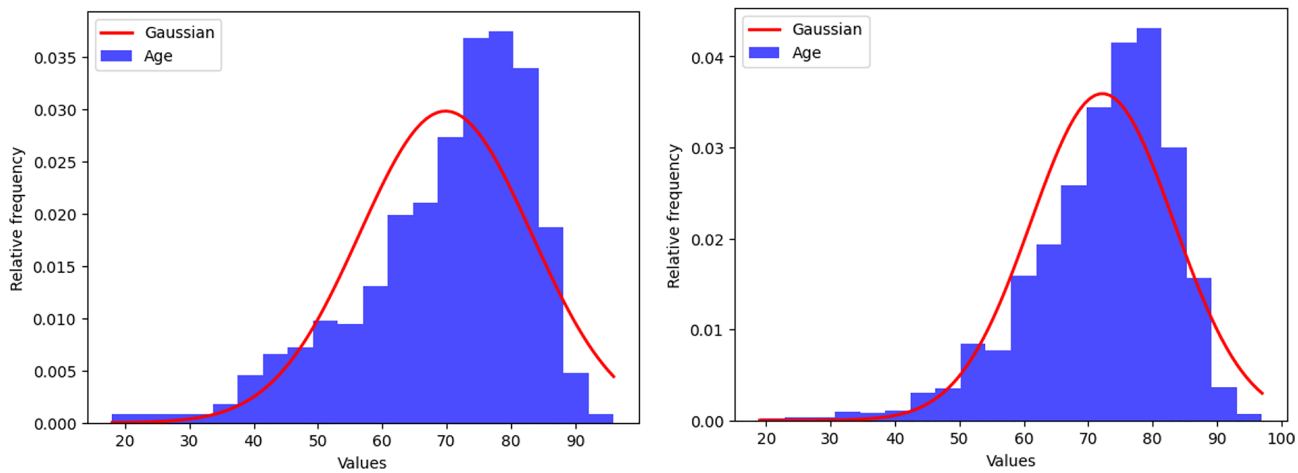

Fig. 2 Clinical examination of the non-normal Age distribution in NP and OP datasets. The red line shows the ideal Gaussian distribution, while the blue histograms depict the Age distribution in NP (left) and OP sets (right). Both exhibit a leptokurtic distribution: the histograms are asymmetric, leaning towards the right extreme, and reach higher levels, deviating from the normal symmetric red curve

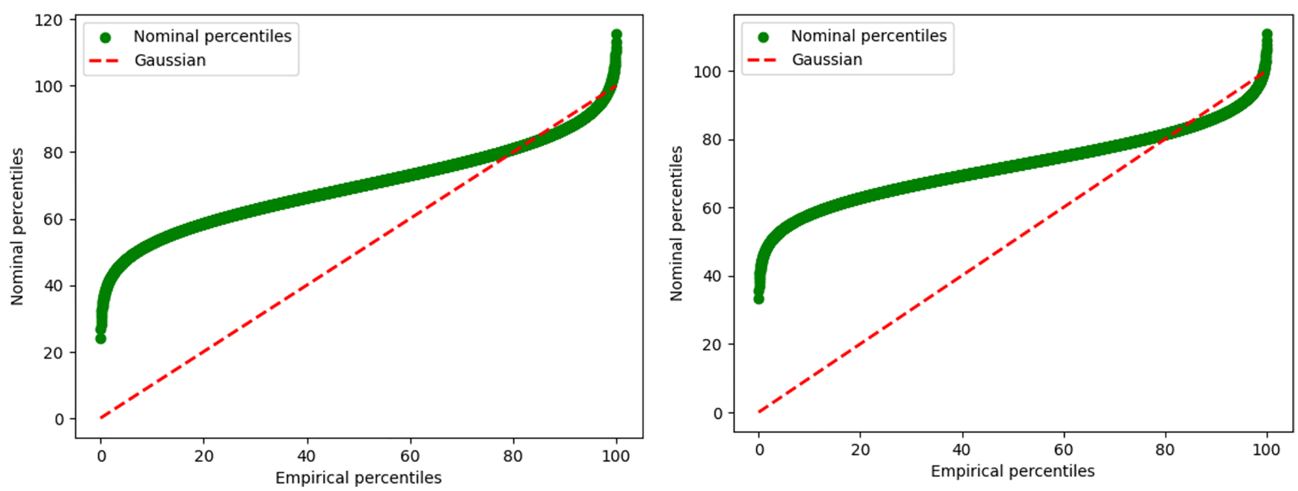

Fig. 3 Graphical representation of nominal percentiles to assess the non-normality of the Age distribution in NP (left) and OP (right) datasets. The green line corresponds to the normal percentiles of the Age variable, while the red dashed line represents the Gaussian distribution. The noticeable gap between the lines in each graph serves as a visual confirmation of the non-normal Age distribution

Table 5 Statistical confirmation of the non-normal distribution of the *Age* variable in NP and OP datasets

Test	*p*-value NP	*p*-value OP
Shapiro-Wilk	7.76E-35	1.08E-34
Kolmogorov-Smirnov	4.34E-34	8.68E-30

Best combination between model and balancing techniques

We address class imbalance by applying various resampling techniques, including ROS, RUS, SMOTE, SMOTE-Tomek, ADASYN, and Cluster Centroids (CC). Before resampling, the training set contained 2255 neurological patients (NP) and 2842 orthopedic patients (OP), i.e. the 70% of the NP and OP set sample size. After resampling, the training dataset size varied significantly, balancing the dataset by oversampling the minority class and ensuring a more representative distribution. The test set, comprising the remaining 30% of the NP (967) and OP (1218) dataset, was not exposed to any resampling procedure. This decision was made to preserve a realistic imbalanced distribution in the test data, better reflecting real-world conditions. By avoiding the introduction of synthetic data in the evaluation phase, we prevent potential biases related to overfitting and inflated performance metrics, thereby ensuring a more robust assessment of the models' generalization capability, as well as their reliability, reproducibility, and credibility.

We used accuracy values to select the optimal combination of balancing techniques and classifiers for achieving accurate predictions on unbalanced test set. Values for NP and OP are shown in Table 6.

Table 6 Accuracy values for different combinations of balancing techniques and classifier in the NP and OP datasets

Accuracy	ROS		RUS		SMOTE		SMOTE Tomek		ADASYN		CC	
	NP	*OP*	*NP*	*OP*	*NP*	*OP*	*NP*	*OP*	*NP*	*OP*	*NP*	*OP*
GB	0.7053	0.7463	0.6112	0.6617	0.7249	0.7898	0.7187	0.7824	0.7053	0.7800	0.4292	0.2422
RF	**0.8418**	**0.9171**	0.6039	0.6273	0.7704	0.8588	0.7684	0.8612	0.7663	0.8563	0.4261	0.2611
XGBoost	0.7932	0.8785	0.5832	0.6494	0.7870	0.8842	0.7839	0.8793	0.7901	0.8842	0.3847	0.2463
LightGBM	0.7684	0.8555	0.5988	0.6650	0.7787	0.8859	0.7839	0.8916	0.7808	0.8966	0.3857	0.2323
CatBoost	0.8014	0.8678	0.6163	0.6741	0.7849	0.8826	0.7808	0.8851	0.7777	0.8851	0.4012	0.2430

Table 7 Top 20 most critical features identified through random forest (RF) feature ranking in the NP and OP datasets

Rank	NP	OP
1	Age	Age
2	Categorized mBI on admission	Hip arthroplasty
3	Balance impairment	Urinary system condition
4	Cognitive impairment	Knee arthroplasty
5	Neurological	Balance impairment
6	Gender	Hypertension
7	Nutrition impairment	Gender
8	Hypertension	Amputation
9	Sensory impairment	Other orthopedic pathologies
10	Ulcers	Different types of comorbidities
11	Communication/language impairment	Diabetes
12	Manipulation impairment	Dysmetabolic
13	Sphincter control impairment	Rheumatology/Orthopedics
14	Urinary system condition	Tumors
15	Different types of comorbidities	Dyslipidemic
16	Behavior impairment	Ulcers
17	Diabetes	Cardiovascular impairment
18	Cardiovascular impairment	Neurological
19	Stroke	Categorized mBI on admission
20	Heart diseases	Locomotion impairment

The highest accuracy was obtained using ROS as the balancing technique in combination with the Random Forest Classifier. This result holds for both NP and OP groups. ROS expanded the training set to 3814 NP and 5316 OP instances.

Feature importance ranking

Using built-in functions of various models in the Python SK-Learn library, it's possible to determine a feature importance ranking in predicting the outcome. As already shown in Table 6, the model with the highest accuracy is the RF classifier. Table 7 shows the ranking obtained with this model for NP and OP, respectively.

Both RF rankings confirm that Age holds the greatest influence in forecasting patient discharge destination at the end of the rehabilitation program. Urinary and cardiac systems hold a place among the top twenty most influential features, along with hypertension diagnosis. In addition, by logical expectations, substantial weight is also associated with balance impairment, ranking third for NP and fifth for OP. The mBI proves to be a significant variable for NP; its importance decreases among orthopedic ones but still ranks among the top 20 most influential predictors. Finally, it can be observed how the presence of comorbidities influences the prediction of the outcome in both subsets [53] and how the gender isn't a top five important feature.

Hyperparameter optimization

This step consists of the optimization of the classifier hyperparameters. We conducted a grid search on the most common values of RF classifier specific parameters, such as n_estimators, max_depth, min_samples_split, min_samples_leaf, and max_features. The aim is to find the best combination of those hyper-parameters to optimize the model performance. Table 8 shows the best values found for NP and OP sets.

Table 8 Best combination of RF hyperparameters in NP and OP implemented model

Best values		
HPT	**NP**	**OP**
n estimators	500	100
max depth	20	20
min samples split	2	2
min samples leaf	1	1
max features	log2	sqrt

Model evaluation

The evaluation metrics were used to analyze model performance in predicting the outcome of our study. In both groups, values were calculated using SK-learn specific functions (*accuracy_score*, *precision_score*, *recall_score*, *f1_score*). Figures 4 and 5 present four Whisker-Box plots illustrating the distribution of model performance across 10-fold cross-validation, applied exclusively to the ROS-balanced training dataset. For reference, the mean values obtained during cross-validation were: neurological model (NP)– accuracy: 96%, precision: 99%, recall: 93%, F1-score: 96%; orthopedic model (OP)– accuracy: 98%, precision: 99%, recall: 96%, F1-score: 98%. However, it is important to note that the 10-fold cross-validation process, as described in Sect. Supervised learning algorithms, introduce synthetic samples into the test fold during each iteration. This condition creates an idealized evaluation scenario, since both the training and test sets come from a synthetically balanced dataset. As a result, there is a risk of data leakage and an overestimation of the model's performance. Although this setup is valuable for assessing model stability and estimating upper-bound performance in a controlled environment, it does not accurately reflect generalization under real-world conditions.

We further evaluated model performance on the imbalanced test set that preserves the original class distributions of the NP and OP datasets. These real-world metrics are overlaid in Figs. 4 and 5 as red dots, representing the model's predictions after training on the ROS-balanced data. The results demonstrate the robustness of both models under realistic conditions, with the

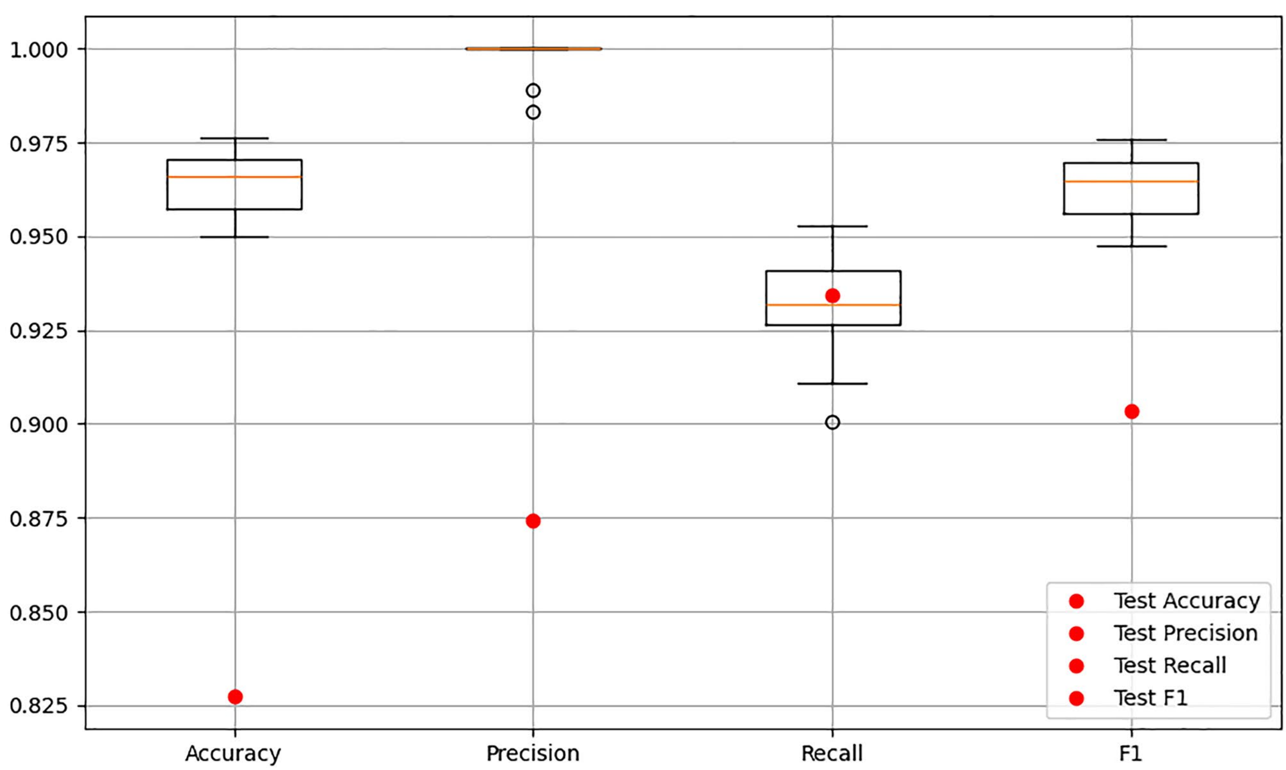

Fig. 4 Graphical comparison of the NP model's performance under balanced and real conditions, showing accuracy, precision, recall, and F1-score. Boxplots represent 10-fold cross-validation results on the balanced training set (e.g. the orange line identifies the median value), while red dots indicate performance on the imbalanced test set

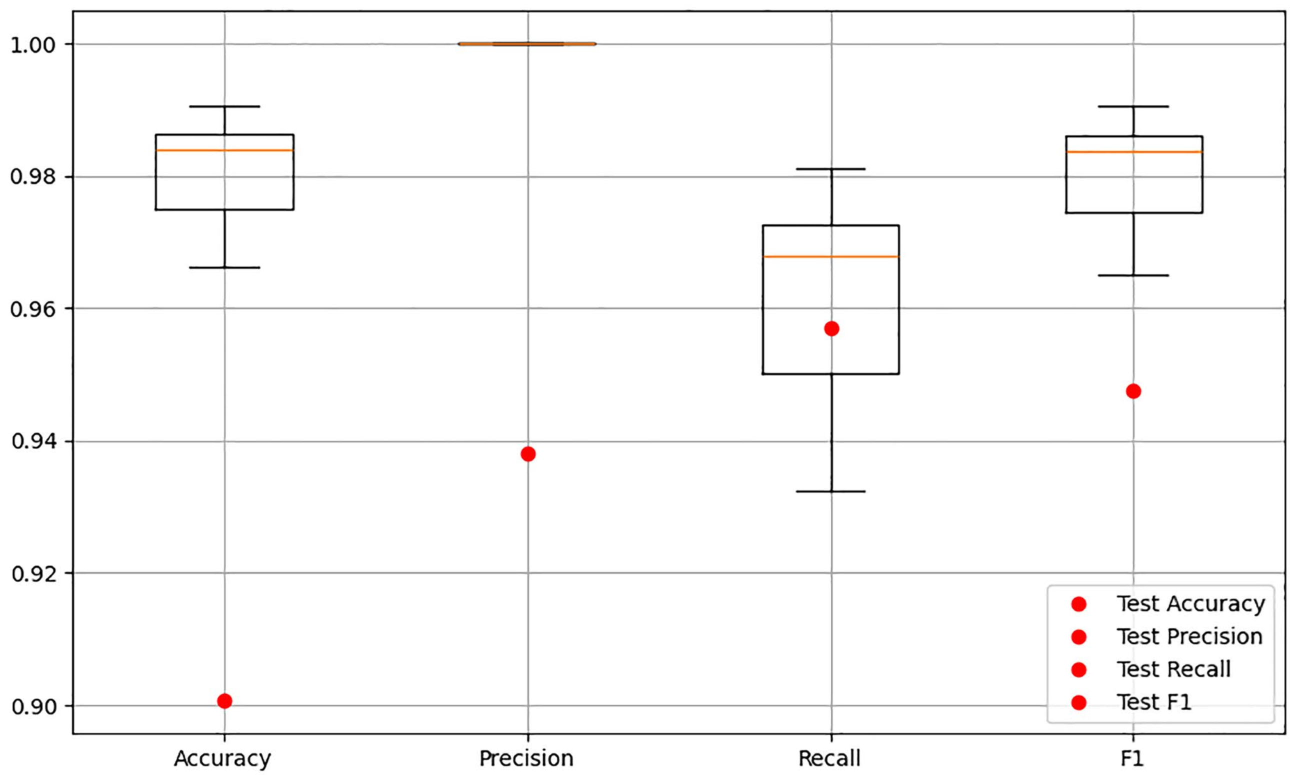

Fig. 5 Graphical comparison of the OP model's performance under balanced and real conditions, showing accuracy, precision, recall, and F1-score. Boxplots represent 10-fold cross-validation results on the balanced training set (e.g. the orange line identifies the median value), while red dots indicate performance on the imbalanced test set

NP model achieving an accuracy of 83% and the OP model reaching 90%. Additional metrics confirm this generalization capacity: NP– precision: 87%, recall: 93%, F1-score: 90%; OP– precision: 94%, recall: 96%, F1-score: 95%.

Regarding performance on the imbalanced test set, the OP model exhibited a slight decrease in precision, which is expected due to the skewed class distribution. Nevertheless, recall and F1-score remained high, indicating that the model effectively identifies true positives while minimizing false positives, as evidenced by a precision above 93%. The small performance gap between cross-validation ideal scenario and real-world evaluation further supports the model's stability and generalization capability. Similarly, the NP model showed a drop in precision, due to the higher proportion of negative instances in the imbalanced test set, which increases the likelihood of false positives. However, recall remained nearly unchanged compared to cross-validation ideal result, suggesting that the model maintains its ability to accurately predict positive cases even under realistic data conditions.

Overall, the findings confirm that both models are reliable, with strong generalization performance when applied to a real-world patient home discharge prediction goal.

Discussion

This study aimed at advancing the patient-centered approach to rehabilitation by integrating medical knowledge with supervised learning methods to predict probability of patients' home discharge considering modified Barthel Index and ADR-r's data. Utilizing a substantial dataset for robust model training and employing machine learning techniques represent significant strengths of our analysis. However, a considerable challenge arose due to the unbalanced distribution of the target variable, which we addressed through extensive data preprocessing, i.e. recoding, scaling, resampling. This phase significantly contributed to the good generalization observed in the implemented classifiers, represented in Figs. 4 and 5.

Within the field of rehabilitation, selecting appropriate protocols for each patient remains a formidable challenge due to the lack of uniformity in their application among individuals showing similar demographic and clinical conditions. Despite this challenge, our work aims to simplify the selection process, narrowing down potentially applicable protocols to promote a more consistent approach to rehabilitation program choice and enhance overall personalization.

A common limitation in machine learning applications is the lack of an independent validation set. While both training and testing are essential steps to assess model learning and performance in supervised ML workflows, a thorough evaluation of model generalizability requires testing on an entirely unseen dataset. In our study, this limitation is expected to be addressed through the prospective collection of a substantial number of new ADR-r forms from patients admitted to the clinic over the coming years. This new data will provide a true external validation set, further improving the reliability and real-world applicability of the proposed tool. In the meantime, an approximate assessment of generalization has been conducted using a test set that was not subjected to any resampling procedures. This approach avoids the introduction of synthetic data, thereby reducing potential bias and allowing for a more realistic evaluation of model reproducibility and robustness.

Moving to the specific findings of our analysis, we aim to provide a predictive model for patient home discharge at the end of the rehabilitation period. The dataset was divided into two subsets based on the nature of the basic pathology encountered: Neurological (NP) and Orthopedic (OP). Our analysis identified the top 20 characteristics with the most significant impact on outcomes in both subsets, detailed in Table 7.

The results align with previous studies, confirming Age as the leading predictive feature for both subsets, consistent with recent pioneering work targeting mBI at the end of the rehabilitation program [6, 54–57]. Notably, patient independence (mBI) has a more substantial impact on NP predictions than on OP, highlighting the lower autonomy levels in performing ADLs among individuals undergoing rehabilitation for basic neurological pathologies. Despite its importance decreases among orthopedic ones, it still ranks among the top 20 most influential predictors, thus confirming the findings in Masaru Uragami's study on readmission of patients with hip fracture [57].

In the case of OP, the most influential variable ranking reveals that the level of urinary system impairment is the third most important feature in predicting home discharge. This suggests that elderly patients facing urinary system issues are likely to have a lower probability of home discharge after rehabilitation, as specific mBI subitems relate to intestinal and urinary continence, as well as toilet use.

The consistent identification of age as the leading predictive feature for both NP and OP subsets reinforces the importance of age in assessing patient outcomes and discharge disposition [6, 54]. Moreover, the observation that patient independence exerts a more substantial impact on predictions for NP compared to OP highlights the complexity of rehabilitation needs among individuals with neurological conditions.

The significant influence of urinary system condition as the third most important feature in predicting discharge destination among OP underscores the importance of comprehensive assessment and targeted interventions to address patient-specific needs [55, 56]. Elderly patients with urinary system issues may require multidisciplinary care and support to facilitate their transition to home settings post-rehabilitation.

These implications emphasize the necessity for tailored interventions aimed at improving patient outcomes and reducing healthcare utilization post-rehabilitation. Regarding potential demographic bias within the dataset, gender was identified as one of the top eight predictors for both models, though it was not among the top five. However, in the orthopedic (OP) group, females were predominant (66% vs. 34%), which may represent a limitation in this field of ML application. Another limitation is the exclusion of the length of stay (LOS) parameter [15] and disease onset time, which reflects the temporal distance between the acute event (e.g., a stroke episode or a hip fracture) and hospitalization. Including these factors could provide additional insights and improve the predictive reliability of both models in real-world settings. Furthermore, balancing techniques, such as Random Over Sampling (ROS), were applied only to the training set: although the model's effective generalization was verified on a test set not influenced by these techniques, the use of balancing methods still constitutes a limitation. Future research should explore stratified analyses or alternative ML models to mitigate such biases.

Although the implemented NP and OP models demonstrate optimal generalization metrics, some clinical implications must be considered. While the dataset used for the analysis comprises a sufficiently large sample, only a limited number of rehabilitation centers were included, which may affect the generalizability of these findings to different clinical settings with varied patient populations and therapeutic protocols. Additionally, psychosocial and environmental factors, which could play a crucial role in discharge probability and functional recovery, were not considered. Furthermore, the model does not account for potential complications or subsequent hospitalizations, which could be addressed by incorporating additional predictors.

Future research should focus on expanding the dataset by including more rehabilitation centers, incorporating psychosocial [26] and environmental factors, and refining predictive models to enhance their generalizability. By addressing these aspects, we can further bridge the gap between AI research and clinical practice, ensuring that ML models offer effective decision-making tools in rehabilitation planning. Additionally, future studies may explore other factors [25] that influence discharge destination after rehabilitation and implement targeted interventions to optimize patient care and reduce healthcare costs.

The extensive data pre-processing phase, comprising ROS and hyperparameter optimization of the RF classifier, significantly improved the predictive performance of our models under real-world conditions. Evaluation metrics, including accuracy, precision, recall, and F1-score, demonstrated promising generalization capabilities for both the OP and NP datasets, as illustrated by the red dots in Figs. 4 and 5 and detailed in Sect. Model evaluation. The observed performance differences between the two models in a real-world context can be attributed to underlying clinical factors. Neurological conditions tend to affect a broader spectrum of physiological systems and functions compared to orthopedic pathologies. This results in greater heterogeneity among NP patient data, leading to slightly lower evaluation metrics for the NP model (e.g., 83% accuracy) compared to the OP model (e.g., 90% accuracy). These findings reflect the inherent complexity and multifactorial nature of neurological disorders, emphasizing the challenges faced by AI practitioners in accurately predicting rehabilitation outcomes for individual neurological patients.

Conclusions

This study introduces a medical tool leveraging ADR-r data to predict patient clinic home discharge post-rehabilitation, facilitating the personalization of treatment plans. The RF output could be used to align rehabilitation programs with patient needs, enabling the full recovery of ADLs. Additionally, it provides timely alerts to family members and caregivers, optimizing home care management. Successful home discharge predictions could be linked to in-home health assistance, while the other case predictions support preparing patients, physically and mentally, for a not home discharge. The model could also be used for clinical resource management, improving bed allocation, medical equipment, and staffing. Future hopes spin around gathering new ADR-r forms to validate more robustly the implemented tool, which will persist in being refined and enhanced along with the significant advancements in AI and ML. This ongoing process will simplify the exploration of new clusters, pathways, and targets, thereby improving the personalization of rehabilitation.

The successful integration of machine learning techniques into clinical practice offers promising avenues for enhancing patient-centered care and improving rehabilitation outcomes. By accurately predicting patient's home discharge, healthcare providers can tailor treatment plans to individual patient needs, optimize resource allocation, and provide timely support to patients and their caregivers. This study represents a significant step forward in employing artificial intelligence and machine learning to optimize patient care and improve healthcare delivery.

Abbreviations

OP	Orthopedic patients
NP	Neurologic patients
ML	Machine learning
AI	Artificial intelligence
ADL	Activities of daily living
mBI	Modified barthel index
RF	Random forest
XGBoost	eXtreme gradient boosting
GB	Gradient boosting
CatBoost	Categorical boosting
LightGBM	Light gradient boosting model
ADR-r	Acceptance and discharge report of rehabilitation
ICD-9-CM	International classification of diseases, revision number 9 with clinical modifications
k-CV k-fold	Cross validation
ROS	Random over sampling
SMOTE	Synthetic minority oversampling technique
ADASYN	ADAptive SYNthetic
RUS	Random under sampling
CC	Cluster centroids
HPT	Hyper-parameter tuning

Acknowledgements

Not applicable.

Author contributions

MF has made substantial contributions to conception and design of the study. SP and CD carried out the data collection. FI, PR and LB designed the algorithm for data analysis. SP and FI participated in the study design and coordination. LB, PR, ESC performed the statistical analysis. LB has elaborated the original draft of the manuscript. MF, FI, SP, and PR and ESC participated in the manuscript revisions. MF gave the final approval of the version. All authors contributed to the article and approved the submitted version.

Funding

This study has been funded by the Italian Ministry of Health (Ricerca Corrente).

Data availability

The dataset supporting the conclusion of this article is available in Zenodo repository with this DOI: https://doi.org/10.5281/zenodo.10991206 and this link: https://zenodo.org/records/10991206.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the ethical committee of the IRCCS San Raffaele Pisana of Rome on 18/07/2018 (code number 07/18).

Competing interests

The authors declare no competing interests.

Author details

[1]Rehabilitation Bioengineering Laboratory, IRCCS San Raffaele Roma, 00166 Rome, Italy
[2]Neurorehabilitation and Robotic Rehabilitation, Department of Neurological and Rehabilitative Sciences, IRCCS San Raffaele Roma, 00166 Rome, Italy
[3]Department of Neurological and Rehabilitation Science, IRCCS San Raffaele Roma, Rome, Italy
[4]Department of Mental and Physical Health and Preventive Medicine, University of Campania 'Luigi Vanvitelli', Naples, Italy

Received: 7 May 2024 / Accepted: 15 May 2025
Published online: 26 May 2025

References

1. World Health Organization. WHO global disability action plan 2014–2021: better health for all people with disability [Internet]. Geneva: World Health

Organization. 2015 [cited 2025 Mar 25]. Available from: https://iris.who.int/handle/10665/199544
2. Gimigliano F, Negrini S, The World Health Organization. Rehabilitation 2030: a call for action. Eur J Phys Rehabil Med. 2017;53:155–68.
3. Negrini S, Meyer T, Arienti C, Kiekens C, Pollock A, Selb M, et al. The 3rd Cochrane rehabilitation methodology meeting: rehabilitation definition for scientific research purposes. Eur J Phys Rehabil Med. 2020;56:658–60.
4. Rahgozar P. AIM in rehabilitation. In: Lidströmer N, Ashrafian H, editors. Artificial Intelligence in Medicine [Internet]. Cham: Springer International Publishing; 2022 [cited 2025 Mar 25]. pp. 1809–17. Available from: https://doi.org/10.1007/978-3-030-64573-1_177
5. Mennella C, Maniscalco U, De Pietro G, Esposito M. The role of artificial intelligence in future rehabilitation services: A systematic literature review. IEEE Access. 2023;11:11024–43.
6. Santilli V, Mangone M, Diko A, Alviti F, Bernetti A, Agostini F, et al. The use of machine learning for inferencing the effectiveness of a rehabilitation program for orthopedic and neurological patients. Int J Environ Res Public Health. 2023;20:5575.
7. Zhang X, Rong X, Luo H. Optimizing lower limb rehabilitation: the intersection of machine learning and rehabilitative robotics. Front Rehabil Sci. 2024;5:1246773.
8. Mazza O, Haim A, Ofir-Geva S. [Applications of artificial intelligence and machine learning in rehabilitation]. Harefuah. 2024;163:600–8.
9. Zaher M, Ghoneim,Amr S, Abdelhamid, Laila, and, Atia A. Rehabilitation monitoring and assessment: a comparative analysis of feature engineering and machine learning algorithms on the UI-PRMD and KIMORE benchmark datasets. Journal of Information and Telecommunication. 0:1–21.
10. Koh GC-H, Chen C, Cheong A, Choo TB, Pui CK, Phoon FN, et al. Trade-offs between effectiveness and efficiency in stroke rehabilitation. Int J Stroke. 2012;7:606–14.
11. Chow P, Chen C, Cheong A, Fong NP, Chan KM, Tan BY, et al. Factors and trade-offs with rehabilitation effectiveness and efficiency in newly disabled older persons. Arch Phys Med Rehabil. 2014;95:1510–e15204.
12. Damiani C, Mangone M, Paoloni M, Goffredo M, Franceschini M, Servidio M, et al. Trade-Offs with rehabilitation effectiveness (REs) and efficiency (REy) in a sample of Italian disabled persons in a in post-acuity rehabilitation unit. Ann Ig. 2020;32:327–35.
13. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL index: a reliability study. Int Disabil Stud. 1988;10:61–3.
14. Hsueh I-P, Lin J-H, Jeng J-S, Hsieh C-L. Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. J Neurol Neurosurg Psychiatry. 2002;73:188–90.
15. Liu H, Xing F, Jiang J, Chen Z, Xiang Z, Duan X. Random forest predictive modeling of prolonged hospital length of stay in elderly hip fracture patients. Front Med (Lausanne). 2024;11:1362153.
16. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol. 2017;69:2657–64.
17. Aha DW, editor. Exploring synergies of knowledge management & case-based reasoning: papers from the AAAI workshop, [19 July 1999 in Orlando, Florida]. Menlo Park, Calif: AAAI; 1999.
18. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380:1347–58.
19. Saria S, Butte A, Sheikh A. Better medicine through machine learning: what's real, and what's artificial? PLoS Med. 2018;15:e1002721.
20. Machine Learning textbook [Internet]. [cited 2025 Mar 25]. Available from: https://www.cs.cmu.edu/~tom/mlbook.html
21. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med. 2018;284:603–19.
22. Safaei N, Safaei B, Seyedekrami S, Talafidaryani M, Masoud A, Wang S, et al. E-CatBoost: an efficient machine learning framework for predicting ICU mortality using the eICU collaborative research database. PLoS ONE. 2022;17:e0262895.
23. Rufo DD, Debelee TG, Ibenthal A, Negera WG. Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). Diagnostics (Basel). 2021;11:1714.
24. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020;2:283–8.
25. Park S, Choi J, Kim Y, You JSH. Clinical machine learning predicting best stroke rehabilitation responders to exoskeletal robotic gait rehabilitation. NeuroRehabilitation. 2024;54:619–28.
26. Fan X, Ye R, Gao Y, Xue K, Zhang Z, Xu J, et al. Prediction of outpatient rehabilitation patient preferences and optimization of graded diagnosis and treatment based on XGBoost machine learning algorithm. Front Artif Intell. 2024;7:1473837.
27. Seccia R, Boresta M, Fusco F, Tronci E, Di Gemma E, Palagi L, et al. Data of patients undergoing rehabilitation programs. Data Brief. 2020;30:105419.
28. Al-Shehari T, Alsowail RA. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. Entropy. 2021;23:1258.
29. El Affendi MA, Al Rajhi KHS. Text encoding for deep learning neural networks: A reversible base 64 (Tetrasexagesimal) Integer Transformation (RIT64) alternative to one hot encoding with applications to Arabic morphology. 2018 Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC) [Internet]. 2018 [cited 2025 Mar 25]; pp. 70–4. Available from: https://ieeexplore.ieee.org/document/8356998
30. Pournajaf S, Pellicciari L, Proietti S, Agostini F, Gabbani D, Goffredo M, et al. Which items of the modified Barthel index can predict functional independence at discharge from inpatient rehabilitation? A secondary analysis retrospective cohort study. Int J Rehabil Res. 2023;46:230–7.
31. Liu F, Tsang RC, Zhou J, Zhou M, Zha F, Long J, et al. Relationship of Barthel index and its short form with the modified Rankin scale in acute stroke patients. J Stroke Cerebrovasc Dis. 2020;29:105033.
32. Ocagli H, Cella N, Stivanello L, Degan M, Canova C. The Barthel index as an indicator of hospital outcomes: A retrospective cross-sectional study with healthcare data from older people. J Adv Nurs. 2021;77:1751–61.
33. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel index for stroke rehabilitation. J Clin Epidemiol. 1989;42:703–9.
34. Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. Stroke. 1999;30:1538–41.
35. Witten I, Frank IH. Data mining-practical machine learning tools and techniques with JAVA implementations. Morgan Kaufmann. 2002;31.
36. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. Artif Intell Rev. 2006;26:159–90.
37. Pedregosa F, Pedregosa F, Varoquaux G, Varoquaux G, Org N, Gramfort A et al. Scikit-learn: machine learning in Python. Machine learning in python.
38. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–7.
39. Jalaeian Zaferani E, Teshnehlab M, Khodadadian A, Heitzinger C, Vali M, Noii N, et al. Hyper-Parameter optimization of stacked asymmetric Auto-Encoders for automatic personality traits perception. Sens (Basel). 2022;22:6206.
40. Younis EM, Zaki SM, Kanjo E, Houssein EH. Evaluating ensemble learning methods for multi-modal emotion recognition using sensor data fusion. Sensors. 2022;22:5611.
41. Tangadpalliwar SR, Vishwakarma S, Nimbalkar R, Garg P. ChemSuite: A package for chemoinformatics calculations and machine learning. Chem Biol Drug Des. 2019;93:960–4.
42. Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics. 2016;17:359.
43. Fernández A, García S, Luengo J, Bernadó-Mansilla E, Herrera F. Genetics-based machine learning for rule induction: state of the art, taxonomy, and comparative study. IEEE Trans Evol Comput. 2010;14:913–41.
44. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl. 2017;73:220–39.
45. Laurikkala J. Improving identification of difficult small classes by balancing class distribution. Artificial Intelligence in Medicine: 8th Conference on Artificial Intelligence in Medicine in Europe, AIME 2001 Cascais, Portugal, July 1–4, 2001, Proceedings 8. Springer. 2001; pp. 63–6.
46. Huang M-W, Tsai C-F, Lin W-C, Lin J-Y. Interaction effect between data discretization and data resampling for class-imbalanced medical datasets. Technol Health Care. 2025;33:1000–13.
47. An oversampling method of unbalanced data for mechanical fault. Diagnosis Based on MeanRadius-SMOTE [Internet]. [cited 2025 Mar 25]; Available from: https://www.mdpi.com/1424-8220/22/14/5166
48. He H, Bai Y, Garcia EA, Li S, ADASYN. Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) [Internet]. 2008 [cited 2025 Mar 25]; pp. 1322–8. Available from: https://ieeexplore.ieee.org/document/4633969

49. Wu Y, Fang Y. Stroke prediction with machine learning methods among older Chinese. Int J Environ Res Public Health. 2020;17:1828.
50. Ali Z, Hayat MF, Shaukat K, Alam TM, Hameed IA, Luo S, et al. A proposed framework for early prediction of schistosomiasis. Diagnostics (Basel). 2022;12:3138.
51. Swana EF, Doorsamy W, Bokoro P. Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. Sens (Basel). 2022;22:3246.
52. Kirk D, Catal C, Tekinerdogan B. Precision nutrition: A systematic literature review. Comput Biol Med. 2021;133:104365.
53. Shah AA, Devana SK, Lee C, Bugarin A, Lord EL, Shamie AN, et al. Machine learning-driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. Eur Spine J. 2022;31:1952–9.
54. Adhiya J, Barghi B, Azadeh-Fard N. Predicting the risk of hospital readmissions using a machine learning approach: a case study on patients undergoing skin procedures. Front Artif Intell. 2024;6:1213378.
55. Lin Y, Wu JY, Lin K, Hu YH, Kong GL. [Prediction of intensive care unit readmission for critically ill patients based on ensemble learning]. Beijing Da Xue Xue Bao Yi Xue Ban. 2021;53:566–72.
56. Zhu Y, Venugopalan J, Zhang Z, Chanani NK, Maher KO, Wang MD. Domain adaptation using convolutional autoencoder and gradient boosting for adverse events prediction in the intensive care unit. Front Artif Intell. 2022;5:640926.
57. Uragami M, Matsushita K, Shibata Y, Takata S, Karasugi T, Sueyoshi T, et al. A machine learning-based scoring system and ten factors associated with hip fracture occurrence in the elderly. Bone. 2023;176:116865.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.