RESEARCH ARTICLE

Data-driven clinical decision trees and generative topographic mapping to guide stroke rehabilitation program allocation

Dimitrios Sokratis Komaris[1☯*], Panagiotis Anagnostou[1☯], Spiros Nikolopoulos[1], Dimitrios Tsaopoulos[1], Anna Guillen-Sola[2,3], Esther Duarte[2,3], Ryan A. A. Bellfield[4,5], Sandra Ortega-Martorell[4,5], Ivan Olier[4,5], Ioannis Kompatsiaris[1]

1 Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece, **2** Physical Medicine and rehabilitation department, Hospital del Mar Medical Research Institute, Barcelona, Catalunya, Spain, **3** Department of Physical Medicine and Rehabilitation, Hospital del Mar, Barcelona, Spain, **4** Artificial Intelligence and Digital Technologies Research Institute, Liverpool John Moores University, Liverpool, United Kingdom, **5** Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart and Chest Hospital, Liverpool, United Kingdom

☯ These authors contributted equally to this work.
* s.komaris@iti.gr

Check for updates

OPEN ACCESS

Citation: Komaris DS, Anagnostou P, Nikolopoulos S, Tsaopoulos D, Guillen-Sola A, Duarte E, et al. (2026) Data-driven clinical decision trees and generative topographic mapping to guide stroke rehabilitation program allocation. PLoS One 21(11): e0363251. https://doi.org/10.1371/journal.pone.0363251

Editor: Hiroki Annaka, Niigata University of Health and Welfare: Niigata Iryo Fukushi Daigaku, JAPAN

Received: April 22, 2026

Accepted: November 5, 2026

Published: November 19, 2026

Copyright: © 2026 Komaris et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The datasets generated and/or analyzed during the current study are available in the figshare repository, doi: 10.6084/m9.figshare.30588821.v1.

Abstract

Background

Early and timely admission to post-stroke rehabilitation remains a key determinant of optimal functional recovery. In this study, we present a structured framework for the development and evaluation of multivariable prediction models and Classification and Regression Tree (CART) decision-support tools to guide patient allocation across diverse subacute rehabilitation settings.

Methods

We have constructed a heterogeneous synthetic dataset of 700 stroke patients characterized by 49 demographic, clinical, and functional variables derived from multiple, literature-reported, statistical distributions of patient populations. Subsequently, the dataset was independently annotated by an expert panel of clinicians, who assigned each stroke patient to one of seven rehabilitation programs ranging from home-based to high-intensity hospital-based care (target outcome). Feature importance analysis identified cognitive function status, National Institutes of Health Stroke Scale (NIHSS) score, age, presence of early seizures, and alcohol consumption as the most influential predictors of rehabilitation program allocation. Six supervised classifiers were then trained to predict rehabilitation program allocation, along with shallow CARTs of different depths. Internal evaluation used 200 repetitions of stratified ten-fold cross-validation, with metrics calculated from pooled out-of-fold predictions and summarized as mean and standard deviation.

Funding: The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by TARGET that is co-funded by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101136244. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Abbreviations: ML: Machine Learning; CART: Classification and Regression Tree; NIHSS: National Institutes of Health Stroke Scale; GTM: Generative Topographic Mapping; SNF: skilled nursing facilities; IRF: in inpatient rehabilitation facilities; FIM: Functional Independence Measure; mRS: Modified Rankin Scale; MoCA: Montreal Cognitive Assessment; PMF: probability mass function; CCI: Charlson Comorbidity Index; MLR: logistic regression; KNN: k-nearest neighbors; SVM: support vector machine; RF: random forest; XGBoost: extreme gradient boosting; LDA: linear discriminant analysis; RBFs: radial basis functions

Results

In internal evaluation against the expert-assigned labels, the XGBoost and the best-performing CART model achieved mean accuracies of 0.832 and 0.847 and balanced accuracies of 0.381 and 0.393, respectively. Both exceeded the institutional benchmark, which achieved accuracy of 0.753 and balanced accuracy of 0.325. Key predictors included cognitive function status, NIHSS score, age, presence of early seizures, and alcohol consumption. In addition, Generative Topographic Mapping (GTM) facilitated the exploration of patient relationships within different prediction models that may help distinguish clusters with similar phenotypes, i.e., patient groups with shared clinical characteristics. It further enabled the visual comparison of stroke rehabilitation program predictions across models (e.g., baseline versus CART model). This allowed for assessment of program recommendations by the developed models in transparent and meaningful ways, and the identification of patient characteristics influencing divergent model decisions.

Conclusions

This study demonstrates the feasibility of developing and evaluating multivariable prediction models for rehabilitation program allocation using an openly shared and expert-annotated dataset. This framework offers a transparent and adaptable foundation for developing bespoke, data-driven decision-support systems in stroke rehabilitation. Ultimately, the proposed approach promotes interpretable, context-sensitive, and individualized rehabilitation planning adaptable to varying clinical resources and practices.

Background

Early and accurate triage of stroke patients to the most appropriate setting during their subacute rehabilitation persists as a challenge in stroke care. Specifically, identifying hospitalized patients who are better suited for high-intensity, short-term, hospital-based programs versus those most suited for low-intensity, prolonged, home-based pathways will critically affect both recovery time and functional outcomes, while optimizing healthcare resource utilization. Such rehabilitation options are diverse and heterogeneous, varying from country to country and the clinical facility itself. For example, in the United States [1] and Canada [2] (and in a high-level overview in Fig 1), key rehabilitation options typically include skilled nursing facilities and nursing homes (SNF for short), home-based healthcare programs, outpatient rehabilitation, and care provided in inpatient rehabilitation facilities (IRF). The timing and intensity of care used is most critical, with guidelines suggesting the initiation of rehabilitation as soon as the patient can tolerate it and to the highest bearable intensity [4], with higher intensity being directly associated with improved functional benefit and balance, overall rehabilitation outcomes, and lower or upper limb ability [5]. For example, SNFs offer rehabilitation services to those requiring daily skilled nursing or

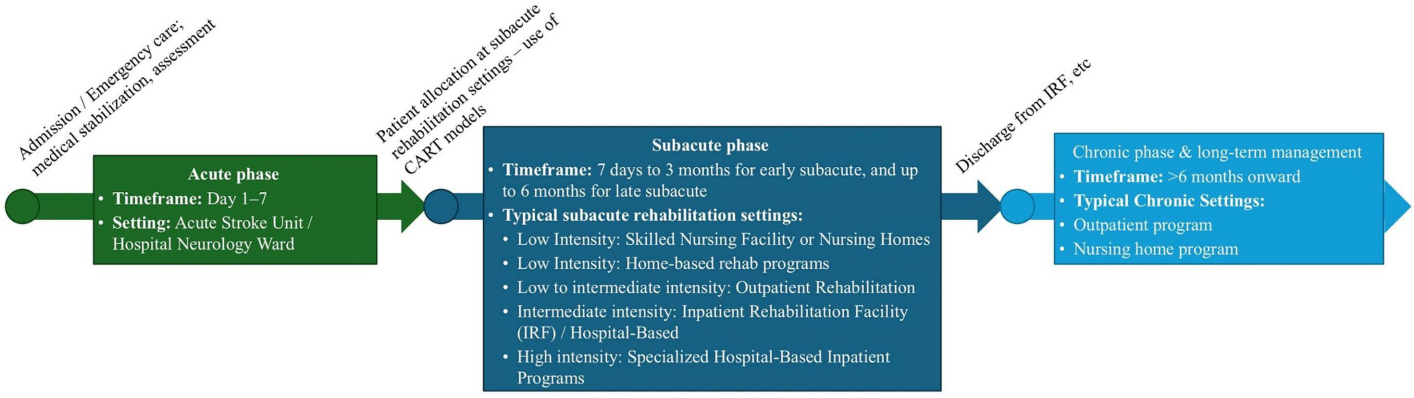

Fig 1. Demonstrative pathway of stroke rehabilitation care. Timeframe examples as per [3].

https://doi.org/10.1371/journal.pone.0363251.g001

therapy, typically do not have a caretaker present, and they are unlikely to achieve full or partial recovery after stroke. For stroke patients discharged home after acute hospitalization, rehabilitation is typically delivered through outpatient clinics or at home by a physician, focusing on supporting the patient's ability to live independently in their own home. On the other hand, IRFs are opted for stroke survivors who require intensive, around-the-clock interdisciplinary rehabilitation, provided under the direct supervision of a physician, and are expected to make meaningful progress within a reasonable timeframe and soon return to a community or work setting. Other key determinants varying between rehabilitation pathways include duration (e.g., from 5 days in inpatient, to more than 30 days in outpatient programs [4]); type of intervention (e.g., occupational, physiotherapy or speech therapy); and the degree of involvement of the medical staff (e.g., medical doctor or only specialist nurse needed).

Such rehabilitation allocation and/or discharge decisions from healthcare professionals often rely on empirical judgment, along with interpretations of national or international guidelines and, more often than not, considerations of the resources of the respective clinical setting. Clinical facilities will typically standardize the type and duration of rehabilitation services that they offer, and develop bespoke clinical tools (e.g., decision trees) to support discharge planning that can be easily and rapidly interpreted by all clinical staff. Such tools will typically leverage a measure of stroke severity (as numerically quantified, for example, by the National Institutes of Health Stroke Scale – NIHSS), functional status (e.g., using the Functional Independence Measure and Modified Rankin Scale; FIM and mRS, respectively), age, cognitive status (e.g., by means of the Montreal Cognitive Assessment – MoCA), and social support (usually, caregiver presence). A small number of items to make a decision and their ease of use is critical, since tools that expect clinicians to resolve complex multivariable equations, consider many parameters, or use multiple cutoff scores are unlikely to gain traction in everyday clinical practice. As an example, Wasti, et al. [6] or the American Heart Association [1] offer simple pathways for level-of-care patient allocation to IRFs, SNFs, outpatient, and other home health options, based on broad and qualitative criteria without cutoffs, such as patient is "able to tolerate intensive rehab" or for those with "daily skilled nursing and/or rehab needs."

Yet, such clinical rules or decision tools are often not standardized, are largely empirical, and they are typically not validated on diverse patient profiles, which, by extent, limits their generalizability across outlier or uncommon cases. For example, thresholds such as "NIHSS ≤ X" or "age ≥ Y" are often arbitrarily used, reflecting historical practice and empirical observation rather than statistically optimized outcomes. Moreover, guideline-based decision tools rarely capture complex interactions (e.g., between age and severity or cognitive status and caregiver support) and tend to underrepresent patient heterogeneity, including comorbidities and variable patient needs (e.g., patient on dialysis). Binary qualitative classifications such as "stable/unstable" or "can/cannot tolerate intensive therapy" are also commonly used, however, these are

subjective and may reduce reproducibility and consistency across clinicians. On this basis, alternative approaches have been explored for the development of classification and regression trees (CARTs for short) or other Machine learning (ML) models that use clear and robust discriminating factors. Such approaches can potentially identify nonlinear or complex interactions between predictors and better portray the variability of patient populations, while incorporating a broader range of parameters (e.g., comorbidities and social factors) that are often overlooked in today's everyday clinical practice. For instance, Imura, et al. [7] developed a CART that uses FIM subscales and patient age to binary allocate stroke survivors to home and no-home discharge destinations; Nakaizumi, et al. [8] used a similar approach with FIM scores and the duration of acute hospitalization as the two classifying variables needed for home or hospital discharge. Veerbeek, et al. [9] reported that the Short-LIMOS score and age were the best splitting criteria for discharge home CART models, with a cut-off of 39 out of 50 points and 55.5 years, respectively. Pérez, et al. [10] classified patients based on caregiver presence, NIHSS, Barthel index, and the Rancho Los Amigos scale to multiple clusters: low, moderate, or high impact rehabilitation, either with or without the presence of a caregiver. Cui, et al. [11] finally trained ML models to again dichotomize the discharge planning of patients into home and non-home options. Evidently, current decision models are still largely simplistic, relying on only a few splitting criteria (typically one or two, for example, FIM and age, see [8]) and restricting classification to two rehabilitation settings (i.e., home versus no-home discharge), thereby overlooking the complexity and diversity of both stroke patient characteristics and the range of (subacute) rehabilitation pathways that today's clinical practice offers.

Another theme investigated in parallel in this work is Generative Topographic Mapping (GTM), which further enables the identification of patient subgroups with distinct rehabilitation allocations (e.g., those allocated to intensive versus outpatient programs), better supporting precision rehabilitation and more efficient resource allocation. The use of GTM to visualize and extract meaningful relationships from data has been adopted in several fields, from assessing e-commerce market segmentation [12], understanding global censorship [13], and even develop clinical relevant phenotypes [14,15]. In the context of this paper, GTM enables data visualization to interpret data trends in a meaningful way and reduce dimensionality without losing important data relationships that link or differentiate patient subgroups (i.e., nodes). Furthermore, the process supports explainability through comparative visualization of predicted stroke rehabilitation programs suggested by different models (e.g., baseline versus CART models). In doing so, we enhance decision transparency by assessing similarities and discrepancies between subacute rehabilitation program assignments and examining the patient-related variables driving these differences.

This work is part of a major European project, TARGET (Grant Agreement No 101136244), aiming to advance personalized management of stroke by means of health virtual twins, artificial intelligence, and other technologies [16–19]. Here, we introduce a structured paradigm for developing ML and CART-based data-driven decision tools to support patient allocation along multiple, diverse, subacute, stroke rehabilitation options. By creating and openly sharing an expert-annotated dataset built from synthetic stroke patient profiles characterized by multiple clinical parameters, we enable transparent and reproducible modeling of rehabilitation decisions across seven distinct care programs, including IRFs, SNFs, outpatient programs, and home-based options. We aim to identify the most influential decision-making parameters for subacute stroke rehabilitation program allocation, and propose exemplary decision-support trees with clear discriminative thresholds, benchmarked against a comparable tool currently used in clinical practice, aiming to balance clinical usability with modeling robustness. While such CARTs must ultimately be tailored to the specific resources and practices of each clinical setting or region, this work provides a flexible foundation for bespoke decision-support systems that can be further validated or adapted to alternative contexts. Finally, GTM was applied to project multidimensional data from the expert-annotated dataset into a low-dimensional latent space, facilitating comparative visualization of model predictions, and enabling data visualization and interpretation of trends in a meaningful way. The integration of GTM thus offers a complementary perspective for understanding patient profiles allocated to each scheme and supports more individualized and refined rehabilitation planning.

Methods

Dataset

A synthetic and expert-annotated dataset of 700 stroke patients was generated to model real-world clinical heterogeneity, using literature-reported distributions for demographic, clinical, and functional parameters. Each patient on the dataset was profiled by forty-nine predictor variables in total, that were randomly sampled from statistical distributions of stroke patient populations in the literature [10,20–31]. In detail, stroke severity was modeled using a mixture of four normal distributions based on the clinical profiles of 104 stroke patients reported in [20], with each patient's NIHSS score sampled from one of the following Gaussian components (μ, σ): 20% of the dataset from N (7.70, 3.21); 43% from N (10.40, 3.23); 31% from N (18.76, 4.19); and 6% from N (27.00, 6.00). Resulting scores were rounded and clipped, as necessary, to a restricted simulation range of 1–42 rather than the standard clinical NIHSS range of 0–42 [21]; a score of 0, denoting no stroke-related neurological deficit, was intentionally excluded from the simulation. Based on NIHSS, patients were then categorized into the following ordinal categorical stroke severity groups: Minor (NIHSS ≤4), Moderate (5–15), Moderate-Severe (16–20), and Severe (>20). Stroke type was ischemic in 81% of cases [10], hemorrhagic in 19%, and laterality was assigned as right (50%), left (42.3%), or bilateral (7.7%) [20]. The study size was pragmatically set at 700 patients to provide sufficient representation across rehabilitation allocation outcomes while maintaining a tractable dataset for model development and internal evaluation. No formal *a priori* sample size calculation was performed, as the study was methodological and simulation-based rather than based on recruitment of real-world participants. No real-world patient data, identifiable participant information, or public participants were involved in this study, and therefore, informed consent, ethics committee approval, and pre-hoc registration were not applicable. Development and internal evaluation of multivariable prediction models was performed in accordance with the TRIPOD Statement [32].

Demographic variables were generated using either normal distributions or stratified sampling according to population-level proportions. Age assumed to follow a normal distribution with a mean of 79.6 years and a standard deviation of 7.9 [10]. Sex was assigned with 50.80% probability for females and 49.20% for males [10]. Caregiver presence was modeled at 73.30% [10], smoking consumption at 38.50% [10], alcohol use at 13.35% [10], and obesity at 7.70% [20] of the population.

Cognitive and functional outcomes were determined conditionally based on NIHSS. The MoCA [22] score was uniformly sampled from progressively lower ranges as severity increased. Similarly, the mRS [23], Barthel Index [24], and FIM (both motor and cognitive subscales [25],) were assigned using a combination of uniform random sampling from severity-specific intervals, normal distributions (μ, σ), and probability mass functions (PMF, see **Table 1**). For example, motor FIM values in Moderate-Severe patients (i.e., NIHSS of 16–20) were sampled from N (28.00, 7.74), while cognitive FIM followed N (25.92, 6.55) based on the FIM distributions reported in [20]. Total FIM was calculated as the sum

Table 1. Distributions and probabilities of the MoCA, mRS, Barthel, FIM and seizure outcomes.

NIHSS Range	MoCA distribution	mRS distribution	Barthel Index distribution	mFIM / cFIM distribution	Seizures probability
1–4	Uniform: 20–25	PMF: [1 = 50%, 2 = 40%, 3 = 10%]	Uniform: 80–99	mFIM ~ N (75.00, 7.00), cFIM ~ N (29.00, 5.00)	Early: 0.0%, Late: 0.0%
5–15	Uniform: 15–25	PMF: [1 = 20%, 2 = 40%, 3 = 30%, 4 = 10%]	Uniform: 61–90	mFIM ~ N (59.42, 12.24), cFIM ~ N (26.96, 4.69)	Early: 0.3%, Late: 0.8%
16–20	Uniform: 10–20	PMF: [2 = 10%, 3 = 30%, 4 = 40%, 5 = 20%]	Uniform: 21–60	mFIM ~ N (28.00, 7.74), cFIM ~ N (25.92, 6.55)	Early: 1.4%, Late: 3.3%
21–42	Uniform: 1–12	PMF: [3 = 20%, 4 = 40%, 5 = 40%]	Uniform: 5–25	mFIM ~ N (16.12, 6.69), cFIM ~ N (10.58, 4.14)	Early: 1.3%, Late: 3.2%

https://doi.org/10.1371/journal.pone.0363251.t001

of motor and cognitive sub scores. Seizure incidence was modeled separately for early seizures (within the first week of admission) and late seizures (after the first week), with probabilities conditioned, likewise, on stroke severity [26]. For example, in moderate strokes (NIHSS of 5–15), early and late seizures were modeled with probabilities of 0.3% and 0.8%, respectively. These probabilities increased to 1.4% and 3.3% for moderate-severe strokes, and 1.3% and 3.2% for severe strokes (NIHSS of 21–42). No seizures were assigned to minor stroke patients (NIHSS of 1–4). The probability of exhibiting symptoms of aphasia and dysphagia were also modelled with increased probabilities as the NIHSS increased and as per the distributions reported in [27]: 2% for the 0–3 NIHSS range; 12% for 4–6; 26% for 7–9; 40% for 10–12; 50% for 13–15; 48% for 16–18; and 73% for 19–42. Cognitive stability was defined from MoCA scores, with MoCA > 26 indicating normal cognitive function, scores 19–25 linked to mild cognitive impairment, and values lower than 18 were associated with moderate to severe impairment.

Binary comorbidity variables and nosocomial complications were generated independently with predefined probabilities from [28,29] as shown in **Table 2**. Urinary incontinence was assigned at a 48% prevalence rate, and among these, urinary catheterization occurred in 74% of cases [30]. Dialysis was assigned only to patients with chronic kidney disease, using a conditional probability of 2.6% [31]. The Charlson Comorbidity Index (CCI) was computed for each patient by summing age-weighted points (1–4 based on age ≥ 50) with weighted scores assigned to each comorbidity per standard definitions [33]. For example, mild liver disease added 1-point, severe liver disease 3 points, and metastatic tumor 6 points. The dataset contains no missing values by construction, and robustness to missing data was therefore outside the scope of this simulation-based study. Aside from the NIHSS-based conditioning applied to cognitive, functional, seizure, aphasia, and dysphagia variables, all variables in this dataset were sampled independently of one another, with two exceptions: dialysis probability was conditioned on chronic kidney disease status, and catheterization was conditioned on urinary incontinence. The CCI and total FIM are deterministic functions of previously generated variables rather than independently sampled. No broader multivariate dependency structure, e.g., between age and comorbidity burden, was modeled.

Table 2. Comorbidity prevalence and nosocomial complications rates from [28,29].

Comorbidity	Prevalence	Comorbidity	Prevalence
Coronary heart disease	21.1%	Connective tissue disease	2.1%
Diabetes	12.1%	Leukemia	0.7%
End-organ diabetes	2.7%	AIDS	0.1%
Chronic kidney disease	10.4%	Hemiplegia	40.1%
Congestive heart failure	9.0%	Hypertension	66.9%
Pulmonary disease	7.8%	Atrial fibrillation	48.4%
Vascular disease	6.7%	Dyslipidemia	12.1%
Dementia	5.0%	Mild liver disease	1.0%
Solid tumor	6.1%	Severe liver disease	2.4%
Metastatic tumor	4.0%	Peptic ulcer	2.8%
Lymphoma	2.4%		
Nosocomial complications	Rate	Nosocomial complications	Rate
Hospital-acquired Pneumonia	9.0%	Other hospital-acquired infections	9.0%
Urinary tract infections	10.0%	Urinary incontinence	48.0%
Urinary catheterization	74.0% of urinary incontinence cases		
Dialysis	2.6% of chronic kidney disease cases		

https://doi.org/10.1371/journal.pone.0363251.t002

Clinical annotation

The dataset was subsequently reviewed and annotated in a structured, multi-stage approach, by an expert panel of four attending physicians from the Rehabilitation Department at University Hospital del Mar (neurorehabilitation unit, Barcelona, Spain), each with substantial clinical experience: one senior physician with over 30 years of experience, two physicians with over 20 years, and one physician with more than 10 years of experience. An initial review of 250 patients was performed to establish inter-rater reliability and consensus methodology. During this stage, agreement on rehabilitation pathway assignment was reached for 155 cases (62.0% raw agreement), while discrepancies among physicians were observed in 95 cases (38.0%). These cases were resolved through a consensus process, in which 63 cases were resolved by majority agreement, and the remaining 32 cases were assigned to the most intensive rehabilitation program suitable for each patient. Following this initial calibration, the evaluation was extended to the entire cohort of 700 patients, with the annotation distributed equally among the four clinicians. The general allocation strategy favored assigning patients to the highest intensity rehabilitation program compatible with their profile, reflecting a conservative approach aimed at providing the greatest opportunity for functional recovery. Seven rehabilitation options were considered here: 1) no need for hospitalization (usually for patients with very low NIHSS scores and good motor function); 2) low intensity: nursing home; 3) low intensity: home-based; 4) low/intermediate intensity: outpatient; 5) intermediate intensity: hospital-based; 6) high intensity: hospital-based; 7) high intensity: complex programs (due to patients' severe cognitive alteration or other necessities, e.g., dialysis). This annotated dataset subsequently served as the ground truth for the development of ML tools and data-driven CART models, and ML-based phenotypic clustering of shared characteristics using GTM. The fully annotated dataset is publicly available in figshare [34].

Development of ML tools and CART models

Supervised learning was conducted to predict program allocation for subacute rehabilitation. Out of the forty-nine considered parameters per subject, nosocomial complications (see bottom part of Table 3: hospital-acquired pneumonia, urinary tract infection, urinary catheterization, urinary incontinence and other hospital-acquired infections) were summed and merged into a single index expressing the cumulative number of complications, and the individual variables were then removed from the predictor set to prevent redundancy. Similar approaches have been adopted in rehabilitation research to consolidate correlated clinical complications into composite indices for modelling efficiency and interpretability [35]. All variables were preprocessed before model training. Numerical predictors were standardized to zero mean and unit variance using a scaler fitted exclusively on each training fold. At the same time, categorical variables were one-hot encoded to produce entirely numeric design matrices. These steps, widely recommended in clinical ML workflows, prevent information leakage and maintain consistency across resampling folds [36]. Predictor associations were assessed using absolute Pearson correlation for numerical pairs, Cramér's V for categorical pairs, and the correlation ratio η for categorical–numerical pairs [37]. The correlation ratio was defined as $\eta = \sqrt{(SS_between / SS_total)}$ for categorical–numerical pairs, where $SS_between = \Sigma_g n_g(\bar{x}_g - \bar{x})^2$ and $SS_total = \Sigma_i(x_i - \bar{x})^2$ where n_g and $\bar{x}_g$ are the size and numerical mean of category g and $\bar{x}$ is the overall numerical mean. This variance-based measure is not derived from a Kruskal–Wallis test. Association filtering was fitted within each training fold. Pairs with a constant variable or a single category were assigned zero association. When association exceeded 0.70, one predictor of the pair was removed. The retained predictors were used for the corresponding fold's model development [38].

Feature importance analysis with the XGBoost classifier [see 39,40] was performed to identify the set of the ten most influential parameters for predictions. Feature importance was quantified using the gain metric, which represents the improvement in the model objective attributable to splits involving each feature. XGBoost was selected as a pragmatic, embedded feature-ranking approach, as gradient-boosted decision trees perform strongly on structured clinical data with potentially nonlinear relationships and interactions among predictors, and a benchmark across 45 tabular datasets found

Table 3. Demographics and clinical characteristics of the modelled sample, expressed in mean ± std, mean [95% confidence intervals], and prevalence rates, where applicable.

		Stroke severity			
Variable	**Total (n = 700)**	**Minor (n = 31)**	**Moderate (n = 453)**	**Moderate-Severe (n = 111)**	**Severe (n = 105)**
Age	79.66 ± 7.77	78.16 ± 7.69	79.92 ± 7.92	79.26 ± 7.21	79.38 ± 7.69
Sex (No of Females)	343	15	224	59	45
NIHSS	13.24 ± 6.50	3.10 ± 0.91	10.17 ± 2.89	17.73 ± 1.36	24.78 ± 4.23
Barthel	61.47 ± 25.51	88.90 ± 5.29	75.74 ± 8.62	40.50 ± 11.91	14.01 ± 6.08
MoCA	17.41 ± 6.06	22.71 ± 1.75	20.23 ± 3.13	15.28 ± 3.32	5.97 ± 3.33
mRS	2.76 [2.00 4.00]	1.77 [1.00 2.75]	2.29 [2.00 3.00]	3.65 [3.00 4.00]	4.15 [3.00 5.00]
mFIM	48.74 ± 20.61	73.35 ± 6.47	59.04 ± 13.02	28.96 ± 7.61	17.97 ± 5.53
cFIM	24.13 ± 7.19	27.23 ± 3.90	26.73 ± 4.49	24.74 ± 6.43	11.36 ± 3.80
tFIM	72.87 ± 25.16	100.58 ± 8.13	85.76 ± 13.85	53.70 ± 10.10	29.33 ± 6.12
Charlson Index	5.62 ± 2.21	5.94 ± 3.07	5.64 ± 2.07	5.53 ± 2.37	5.57 ± 2.35
Cognition: normal or mild impairment	50.70%	100.00%	66.00%	22.50%	0.00%
Laterality (Left)	40.4%	32.5%	41.9%	37.4%	39.4%
Caregiver	71.70%	71.00%	71.50%	72.10%	72.40%
Smoke Consumption	40.00%	51.60%	40.20%	37.80%	38.10%
Alcohol Consumption	12.70%	3.20%	12.60%	12.60%	16.20%
Obesity	7.60%	9.70%	6.80%	9.00%	8.60%
Stroke Type (Ischemic)	80.10%	87.10%	79.20%	83.80%	78.10%
Aphasia	42.30%	3.20%	34.70%	53.20%	75.20%
Dysphagia	45.00%	3.20%	37.50%	61.30%	72.40%
Early Seizure (1st Week)	1.40%	0.00%	2.00%	0.90%	0.00%
Late Seizures (after 1st Week)	1.10%	0.00%	0.70%	2.70%	1.90%
Pneumonia	9.00%	9.70%	9.30%	4.50%	12.40%
Urinary Tract infection	11.90%	12.90%	10.80%	16.20%	11.40%
Other Hospital Infections	10.00%	12.90%	9.10%	9.00%	14.30%
Coronary heart disease	20.70%	19.40%	21.00%	18.90%	21.90%
Diabetes	13.90%	16.10%	14.30%	9.00%	16.20%
End-organ diabetes	1.90%	0.00%	2.00%	1.80%	1.90%
Chronic kidney disease	11.40%	22.60%	9.10%	18.00%	11.40%
Congestive heart failure	11.60%	12.90%	13.70%	7.20%	6.70%
Pulmonary Disease	8.10%	12.90%	9.30%	5.40%	4.80%
Vascular Disease	7.10%	12.90%	7.30%	6.30%	5.70%
Dementia	5.00%	0.00%	5.50%	4.50%	4.80%
Solid Tumor	6.70%	0.00%	6.20%	9.90%	7.60%
Metastatic Tumor	4.30%	6.50%	4.20%	3.60%	4.80%
Severe Liver Disease	3.30%	9.70%	3.30%	2.70%	1.90%
Mild Liver Disease	1.40%	0.00%	1.50%	0.90%	1.90%
Peptic Ulcer	3.00%	0.00%	3.30%	4.50%	1.00%
Lymphoma	1.30%	0.00%	1.80%	0.00%	1.00%
Connective Tissue Disease	2.30%	0.00%	1.80%	5.40%	1.90%
Leukemia	0.60%	0.00%	0.90%	0.00%	0.00%
AIDS	0.00%	0.00%	0.00%	0.00%	0.00%
Hemiplegia	39.60%	45.20%	39.10%	36.90%	42.90%
Hypertension	64.40%	64.50%	65.10%	61.30%	64.80%

(Continued)

Table 3. (Continued)

		Stroke severity			
Variable	**Total (n = 700)**	**Minor (n = 31)**	**Moderate (n = 453)**	**Moderate-Severe (n = 111)**	**Severe (n = 105)**
Atrial fibrillation	48.10%	58.10%	48.60%	54.10%	37.10%
Dyslipidemia	12.30%	16.10%	12.60%	10.80%	11.40%
On Dialysis	0.40%	3.20%	0.00%	0.90%	1.00%
Urinary incontinence	48.30%	35.50%	50.10%	51.40%	41.00%
Urinary catheterization	34.60%	16.10%	36.40%	39.60%	26.70%

https://doi.org/10.1371/journal.pone.0363251.t003

that tree-based methods, including XGBoost, remain state-of-the-art on medium-sized datasets [41]. The resulting ranking was used to reduce the number of input variables, maintaining a small and clinically interpretable set of predictors while preserving predictive performance. Moreover, XGBoost-based feature ranking and selection of the ten most influential predictors were performed independently within each cross-validation fold, using only the corresponding training data; the held-out fold was not involved in feature ranking or selection and was used exclusively for model evaluation, thereby avoiding information leakage. We then trained six commonly used classifiers from distinct algorithmic families and bias–variance trade-offs: multinomial logistic regression (MLR), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), and linear discriminant analysis (LDA). The algorithms were executed using their default parameter settings with no further hyperparameter tuning. In detail, for the MLR, the solver was set to Limited-memory Broyden–Fletcher–Goldfarb–Shanno (lbfgs), with a maximum of 100 iterations, a regularization strength of 1.0, and an L2 penalty. KNN used a five-neighbor search space with uniform weights, the Minkowski (Euclidean) distance metric, and a Minkowski power parameter of 2. SVM used a radial basis function (*rbf*) kernel, and a regularization parameter C of 1.0. The kernel coefficient gamma (γ) was set to the "scale" option, computed as $\gamma = \frac{1}{d \times Var(X)}$, where $X \in (x_1, x_2, \ldots, x_n) \in R^d$ is the training data matrix, n is the number of samples, and d is the number of features; the variance term $Var(X)$ was computed over the features of X to represent typical feature variance. The RF used 100 trees, the Gini impurity as the splitting criterion, no restriction on maximum tree depth, a minimum of two samples required to split an internal node, one sample required at each leaf node, and bootstrap sampling enabled. The parameters for the XGBoost algorithm were set as follows: 100 boosting rounds, a learning rate of 0.3, a maximum tree depth of 6, a subsampling ratio of 1, a column subsampling ratio per tree of 1, uniform random sampling, and a multiclass classification objective using a *softmax logistic* loss that outputs a probability distribution over all classes. Finally, LDA used the solver set to singular value decomposition (SVD), the shrinkage option left unspecified (None), and the convergence tolerance set to $1e-4$.

To assess model performance and its variability across random data partitions, we conducted 200 repetitions of stratified ten-fold cross-validation [42]. Within each repetition, preprocessing was fitted exclusively on the training split (90% of the dataset), and the fitted transformations were applied to the held-out fold (10%). Models were trained and evaluated across all ten folds, and the held-out predictions were assembled into a complete out-of-fold prediction vector aligned with the true outcome vector, ensuring that each observation was predicted once per repetition. Evaluation metrics, including macro-averaged metrics, were then calculated using these complete vectors rather than by averaging fold-level metrics. This procedure yielded 200 repetition-level estimates for each model-metric pair. Final performance was summarized by the mean of these 200 estimates, while their standard deviation quantified variability associated with random partitioning, conditional on the fixed dataset. This variability does not represent uncertainty about performance in independent real-world stroke populations. All pooled evaluations used the same seven rehabilitation categories. Precision, recall and F1 were macro-averaged across these categories, with undefined precision or F1 assigned zero. Balanced accuracy was the

arithmetic mean of the seven class-specific recalls and therefore equaled macro recall. Because four rehabilitation categories contained fewer than ten observations, stratified partitioning could not represent every category in every validation fold; the observations of such categories were distributed across as many folds as they contained observations. Consequently, whenever the single observation of the no need for hospitalization category was held out, that category was absent from the corresponding training data and could not be predicted by the developed models. Pooling the out-of-fold predictions ensured that all seven categories contributed to each repetition-level metric, but it did not enable the models to learn a category that was absent from a training partition.

Subsequently, shallow CARTs of different depths were exported [43], to derive human-readable decision rules and facilitate qualitative comparison with an empirical decision tree currently used in clinical practice (see subsequent section: **Comparative Benchmarking of CART Models**). We used the DecisionTreeClassifier provided by the Scikit-learn library to produce our CART models, with the splitting criterion set to Gini impurity, and the splitter strategy set to "best." The maximum tree depth ranged from 3 to 7; the minimum number of samples required to split an internal node was set to 2; the minimum number of samples required at a leaf node was set to 1; and the minimum weighted fraction of input samples at a leaf node was set to 0. The ten most influential parameters determined from the feature importance analysis were considered at each split, and the minimum impurity decrease was set to 0. Class weights were left unspecified, and the cost-complexity pruning strength (alpha) was set to 0. An alternative CART model with comparable length of the longest path and set of input parameters (i.e., the NIHSS, patient age, caregiver presence, MoCA score, mRS and Charlson Index) to the currently used in-house decision tree was also produced. The aim of this was to generate a CART with comparable inputs and complexity —and not driven by feature importance analysis or the optimization of the number of layers— to the currently used one, while only improving its predictive accuracy and reevaluating the discriminative thresholds assigned to each decision node.

Computational details

All analyses were performed using Python 3.13.13 with scikit-learn 1.8.0, XGBoost 3.2.0, NumPy 2.4.6, and pandas 3.0.3. Computations were performed on Ubuntu 24.04.5 LTS using an Intel Core i5-12400F CPU (6 cores/12 threads) and 32 GB RAM. Each iteration of the 200 × 10-fold cross-validation, comprising training and evaluation of all six classifiers across one complete 10-fold partition, required approximately 2.1 seconds.

Comparative benchmarking of CART models

To benchmark the data-driven CART models developed from the expert-annotated cohort, we compared their agreement with the in-house decision tree that is currently used at the institution (Hospital del Mar, Barcelona, Spain), and it is based on national stroke rehabilitation guidelines and current practices. The in-house decision tree was included as a clinically meaningful institutional benchmark rather than as an equivalently optimized predictive model. This in-house decision tree was also made publicly available for review along with the dataset [34]. The in-house decision tree utilizes as predictor variables the NIHSS, MoCA score, mRS, patient age, caregiver presence, and the Charlson Index. The possible outputs (leaf nodes) correspond to the same seven rehabilitation pathway categories that were considered for the annotation of the cohort (refer to subsection, **Clinical Annotation**). The in-house tree has a total of 22 decision nodes with an equal number of total branches and a maximum depth of 5 layers. Each split is performed either using a threshold inequality (e.g., "NIHSS ≤ 2") for continuous variables or a categorical feature (e.g., "caregiver present: yes/no") for binary predictors. The fixed institutional rule was evaluated once on the same 700 labelled profiles using accuracy, macro precision, macro recall, macro F1 and seven-class balanced accuracy. Unlike the developed ML/CART models, these metrics were computed without resampling, and are reported descriptively without standard deviations across partitions, as the empirical tree was not subjected to the 200 × 10 stratified cross-validation protocol.

Generative topographic mapping

Categorical variables in the dataset (e.g., alcohol consumption) were first treated as binomial and represented by a Gaussian if both $n(p) \geq 5$ and $n(1-p) \geq 5$, where n denotes the number of samples and p the probability of a successful event. Log transformations were applied to highly skewed data, followed by standard scaling to ensure appropriate normalization for model training. GTM is a probabilistic dimensionality reduction technique that assumes all observations exist in a R^D space, generated through a nonlinear, topology-preserving mapping from a lower-dimensional latent space R^L to a manifold embedded in R^D. L is usually set to 2 or 3 to allow for a latent space that can be visualized and interpreted from a human perspective (here, L = 2). The latent space is pre-defined to contain a uniform grid of N nodes generated using a Dirac delta function, with each of these centers responsible for generating a radially symmetric, isotropic Gaussian in the data space. This formulation yields a soft clustering approach in which each data point is probabilistically associated with every cluster (referred to as *responsibilities* [15,44]). Hard clustering can subsequently be obtained using a method known as the mode assignment, where each data point is assigned to the Gaussian with the highest associated probability. Cluster assignments in data space are directly mapped onto latent nodes, which can be visualized as a 2-D *membership map* (given L = 2). The size of each node is determined by the number of data points (here, stroke patients) assigned to it. In addition to this, we can also utilize the vectors that define the Gaussian center locations and can be considered as data prototypes. By looking at the vector components that correspond to each (clinical) variable, we can generate visualizations henceforth referred to as *reference maps* to help us understand how each value corresponds to the distribution of data within the latent space. The GTM model was trained using 64 latent nodes arranged in an 8 × 8 grid and 16 radial basis functions (RBFs), ensuring at least four latent points per RBF, as described in Olier, et al. [45]. We used 10-fold-cross validation and a grid-search approach to optimize the RBF regularization parameter which regulates the mapping process to avoid overfitting (with values between 0.001 and 2 being tested in 30 steps) and the width of the Gaussians used in the radial basis functions (values from 0 and 2, in 30 steps). We also generated additional visualizations for variables not included in the development of the GTM model to understand, through a post-hoc analysis, how variables are distributed amongst the clusters, as described in [15].

Once the GTM membership map was trained, we visualized the rehabilitation program assignments derived from three sources: (a) outcomes annotated by the panel of medical doctors (MDs), (b) outcomes predicted by the currently used in-house CART model, and (c) predictions generated by the developed XGBoost model. Each latent node was assigned to the rehabilitation program with the highest prevalence among its associated data points (i.e., patients clustered in the node). These node-level program assignments were then superimposed onto the membership map to enable comparison of the distributions of the clinical variables within and across the three prediction methods. Finally, to visualize the agreements and disagreements between the labels provided by the clinicians (see section **Clinical Annotation**), the in-house decision tree and best-performing ML predictions (XGBoost model, see section **Development of** ML **tools and** CART **models**), Sankey diagrams were constructed to illustrate label flows (as derived from latent nodes generated from the fitted GTM models) and whether a model systematically favors assigning patients to a particular program (i.e., to an intermediate intensity hospital-based node).

Results

Dataset

Demographics and clinical characteristics of the synthetic stroke cohort (n = 700), along with the corresponding prevalence and nosocomial complications rates, were grouped based on patients' stroke severity (minor, moderate, moderate-severe, and severe) and are detailed in Table 3. Most of the cohort was characterized by a stroke of moderate severity (n = 453), with an average NIHSS of 10.17 units, followed by 111 cases of moderate-severe stroke with an average NIHSS of 17, and 105 severe stroke incidences with average NIHSS equal to 24.78. Only 31 cases are present with a minor stroke

in the dataset, with an NIHSS score of 3.10 on average. The expert-annotated rehabilitation outcome was imbalanced across the seven categories: no need for hospitalization was represented with 1 patient (0.1%); low intensity: nursing home with 29 patients (4.1%); low intensity: home-based with 3 patients (0.4%); low/intermediate intensity: outpatient with 8 patients (1.1%); intermediate intensity: hospital-based with 395 patients (56.4%); high intensity: hospital-based with 261 patients (37.3%); high intensity: complex programs with 3 patients (0.4%). This uneven class distribution was considered when interpreting model performance, through balanced accuracy and macro-averaged metrics.

ML tools and CART models: within each training fold, predictors whose pairwise association exceeded 0.70 (see Methods) were removed from those listed in Table 3, and the ten most influential of the remaining predictors were retained for our predictions based on a feature importance analysis [40]. These were the patients' cognitive status, NIHSS, age, presence of early seizures during the 1st week of hospitalization, alcohol consumption, type of stroke, presence of caregiver, smoke consumption, Barthel index, and the mRS (see Fig 2 for the full ranked list). The ten predictors listed above and the ranking in Fig 2 were obtained using the complete dataset of 700 profiles, as were the final CART models displayed in Figs 3 and 4; internal evaluation separately repeated all data-driven preprocessing, feature-selection and model-fitting steps within each training fold. **Table 4** reports the final internal evaluation using 200 pooled out-of-fold estimates per model. XGBoost achieved the highest balanced accuracy among the six non-CART classifiers, with mean accuracy of 0.832, macro precision of 0.390, macro recall of 0.381 and macro f1 of 0.378. The CART model in Fig 3 (the best-performing CART by overall accuracy) achieved mean accuracy of 0.847, macro precision of 0.400, macro recall of 0.393 and macro f1 of 0.390. Balanced accuracy equalled macro recall for every model. XGBoost, Random Forest and both CART models exceeded the institutional benchmark in overall accuracy and balanced accuracy; the benchmark achieved 0.753 and 0.325, respectively. Random Forest had the highest overall accuracy among the six non-CART classifiers (0.848) and a balanced accuracy of 0.358. The remaining non-CART classifiers exceeded the benchmark in overall

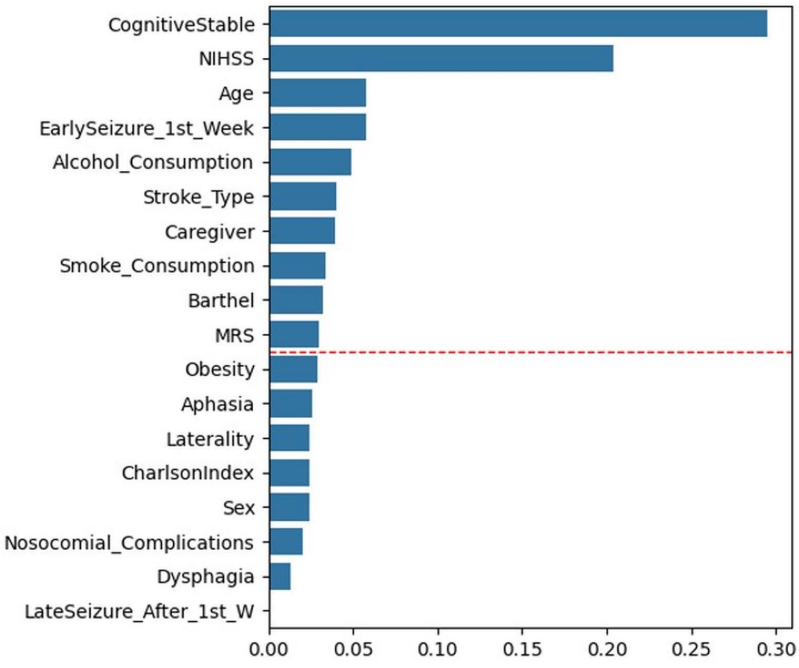

Fig 2. The 10 most influential predictors identified using XGBoost feature importance analysis [40]. The red dashed line indicates the threshold used to retain the top 10 predictors for subsequent model development.

https://doi.org/10.1371/journal.pone.0363251.g002

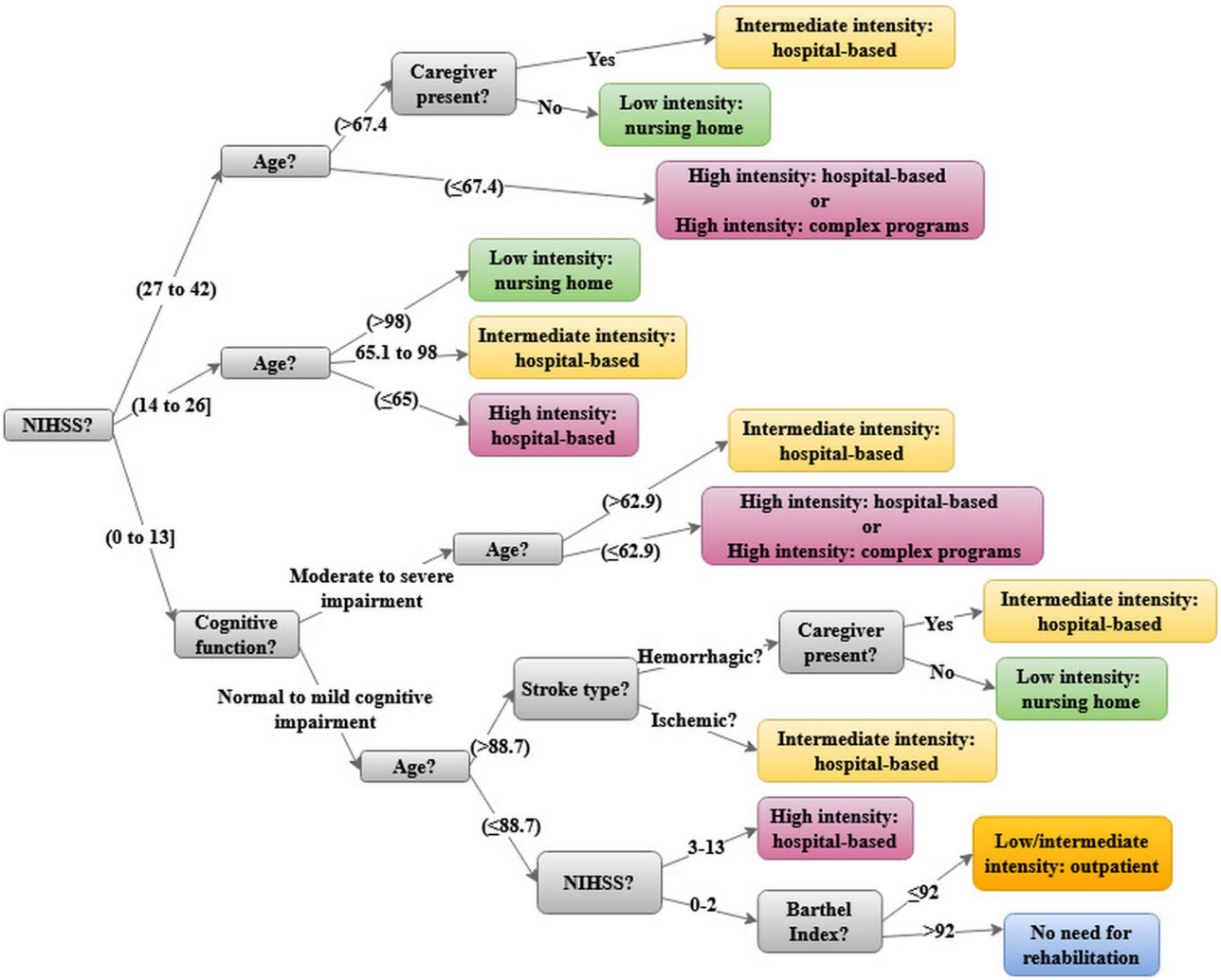

Fig 3. The developed CART model has a maximum depth of five layers and includes as decision parameters the NIH Stroke Scale/Score (NIHSS; 0–42-point scale), cognitive function, age, stroke type, caregiver presence, and Barthel Index. Values displayed alongside the arrows indicate the corresponding decision criteria, including threshold-based splits (e.g., age > 67.4 years) and binary conditions (e.g., presence of ischemic versus hemorrhagic stroke).

https://doi.org/10.1371/journal.pone.0363251.g003

accuracy but had lower balanced accuracy. The reader can also refer to S1 Table in the **Supporting Information** for class-specific precision, recall, and F1-score values for each of the seven rehabilitation categories.

Subsequently, we have produced an alternative CART version (Fig 4) with a fixed depth of 5 layers and predetermined set of input parameters, comparable to the structure and composition of the currently used in-house decision tree. Likewise, this CART version returned similar performance quality metrics to the standard CART model in Fig 3 (as shown on Table 4). This is not surprising, as both versions have the same depth level and use the same or similar input features: NIHSS, patient age and caregiver presence are shared decision variables among both trees; cognitive function (model in Fig 3) and MoCA (model in Fig 4) are both direct indicators of cognitive status; similarly, Barthel (Fig 3) and mRS (Fig 4) are measures of functional independence; the only notable difference between the two trees is the substitution of the stroke type with the Charlson Index, and the presence of stroke type (hemorrhagic or ischemic) as the sixth most critical classification variable in our analysis (Fig 2).

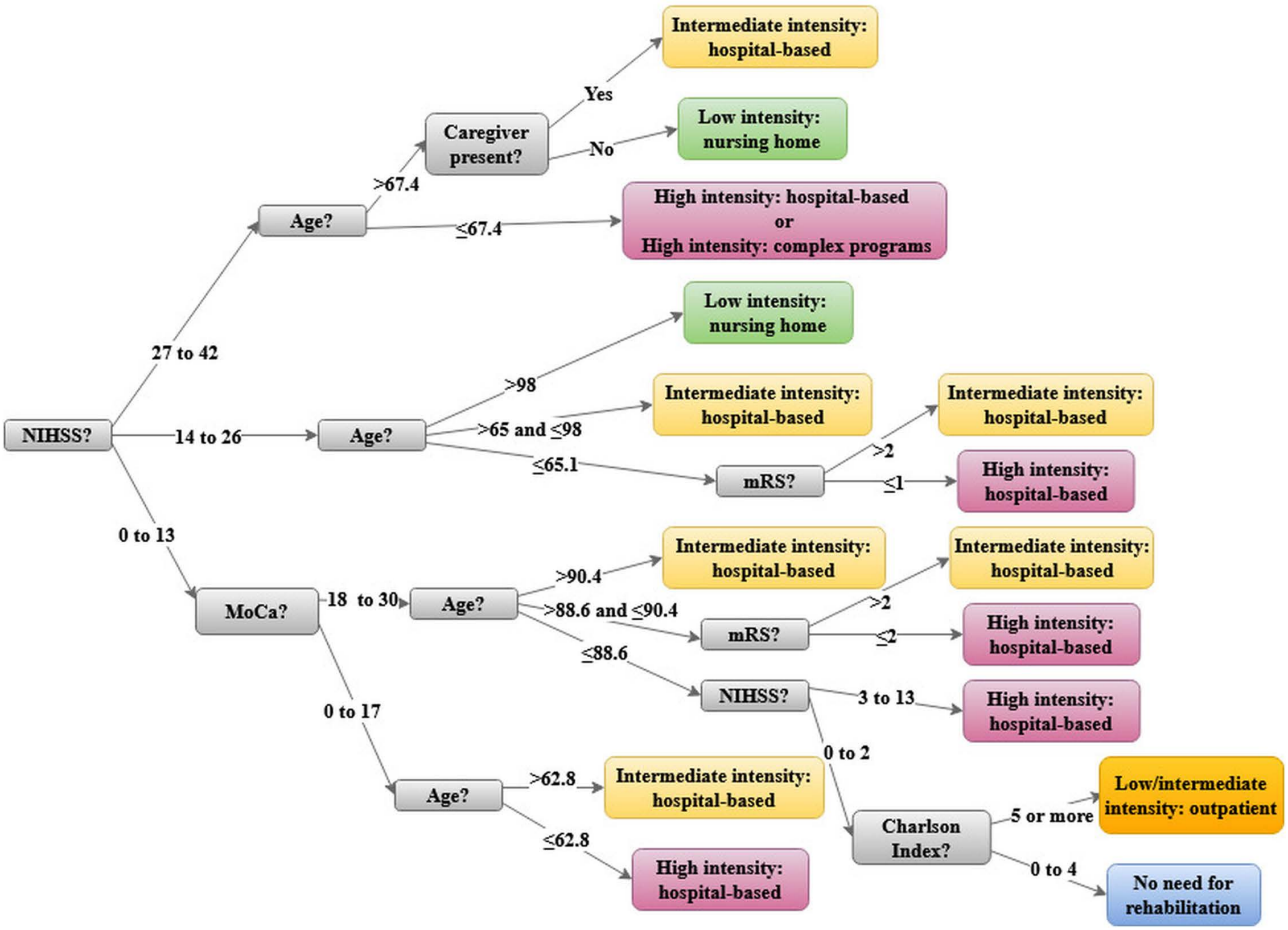

Fig 4. Alternative CART model with a predefined depth of 5 layers, and a pre-selected set of parameters, same as in the currently used in-house decision tree, including the NIHSS, patient age, caregiver presence, Montreal Cognitive Assessment score (MoCA; 30-point scale), Modified Rankin Scale (mRS; 0-6) and Charlson Index (0-29 scale). Values displayed alongside the arrows indicate the decision criteria, including threshold-based splits and binary conditions.

https://doi.org/10.1371/journal.pone.0363251.g004

Table 4. Model performance expressed as mean (standard deviation) across 200 repetitions of stratified ten-fold cross-validation. Metrics were calculated from complete pooled out-of-fold predictions within each repetition using all seven rehabilitation categories. Precision, recall and F1 are macro-averaged; balanced accuracy equals macro recall. Parenthetical values describe variability across repetitions, not standard errors or confidence intervals. The top row shows the performance of the currently used in-house decision tree.

ML / Cart Model	Accuracy	Precision	Recall	f1	Balanced accuracy
In-house CART	0.752857	0.344419	0.325230	0.327859	0.325230
CART model: Fig 3	0.84691 (0.00435)	0.40009 (0.01450)	0.39326 (0.00964)	0.38993 (0.01017)	0.39326 (0.00964)
Alt. CART model: Fig 4	0.84584 (0.00587)	0.39692 (0.01618)	0.39711 (0.01217)	0.39090 (0.01253)	0.39711 (0.01217)
XGBoost	0.83210 (0.00653)	0.38956 (0.01593)	0.38138 (0.01147)	0.37781 (0.01236)	0.38138 (0.01147)
Random Forest	0.84756 (0.00570)	0.37832 (0.02094)	0.35782 (0.01650)	0.35991 (0.01680)	0.35782 (0.01650)
Mult. Logistic Regression	0.82630 (0.00372)	0.33489 (0.01043)	0.31552 (0.00447)	0.31812 (0.00534)	0.31552 (0.00447)
K-Nearest Neighbors	0.77899 (0.00617)	0.28598 (0.01266)	0.26580 (0.00521)	0.26737 (0.00650)	0.26580 (0.00521)
Support Vector Machine	0.81341 (0.00469)	0.29916 (0.01188)	0.27260 (0.00455)	0.27665 (0.00599)	0.27260 (0.00455)
Linear Discriminant Analysis	0.81203 (0.00377)	0.33201 (0.00916)	0.32061 (0.00477)	0.31973 (0.00532)	0.32061 (0.00477)

https://doi.org/10.1371/journal.pone.0363251.t004

The first decision of both CART models for subacute treatment is made based on the NIHSS of the patient, splitting into three branches: the bottom branch encapsulates all NIHSS scores below 13 (i.e., for minor and moderate stroke events), the middle one includes cases with scores from 14 to 26 (i.e., mostly moderately severe and some severe events), while the top branch corresponds to patients with scores of over 27 (severe cases only). From there, and for both trees (Fig 3 and Fig 4), patient age and caregiver presence are the two parameters in the internal nodes of the top branch (NIHSS from 27 and 42). As a general overview, older patients (over 67.4 years of age with NIHSS higher than 27) will be allocated to nursing homes, while younger patients will follow high intensity regimes. Please note that patients with a NIHSS score greater than 27 typically present with very severe neurological deficits. In these cases, rehabilitation tolerance is usually low, and it is rarely feasible to initiate an intensive rehabilitation program. Instead, treatment generally begins with lower-intensity interventions while awaiting improvement in cognitive and motor functions that will allow the patient to participate effectively in more intensive therapies. Likewise, age (Fig 3), and age and mRS (Fig 4), are incorporated as decision nodes for the middle branches (NIHSS from 14 to 26). Again, older individuals (98 years or older, approximately 1% of the patients in this study, with an NIHSS between 14 and 26) will be directly assigned to nursing homes, younger patients (between 65.1 and 98) to intermediate care, while very young persons (≤65) to high intensity schemes. The bottom branches (NIHSS≤13) are the most complex for decision making, with options varying from "no need for rehabilitation" for younger cases (approx. ≤ 88.6 years) with good cognitive function, very low NIHSS scores (≤2), and either very high Barthel Index scores (≥92) or low Charlson Comorbidity Index scores (≤4), to intermediate intensity allocation for older stroke survivors with declined cognitive function or high MoCA scores.

Generative topographic mapping

To better understand where differences in rehabilitation program assignment arise between models, we employed GTM, a process that results in the clustering of patients with similar clinical characteristics into nodes within a membership map, providing a clear visualization of shared phenotypic patterns. By examining these nodes, we can observe how each model assigns patients to rehabilitation programs, identify areas of agreement and divergence, gain insight into the underlying factors influencing model decisions, and explore, as a hypothesis-generating exercise, which model may suit a given setting—for example, a conservative classifier that preferentially assigns patients to intermediate-level rather than intensive programs. First, hyperparameter tuning was used to derive optimal values for the RBF regularization term of 0.001 and a width of 0.2759 for the Gaussian functions. Using these parameters, we then trained a GTM model and extracted a fitted membership map (Fig 5, panel **a**). Each node in blue in the membership map contains at least one patient, with larger nodes indicating a higher number of assigned patients (approximately from 1 to 25 patients per node in this analysis). Likewise, Fig 5b presents the corresponding reference maps, which illustrate how the most important clinical variables influence the distribution of patient data within the latent space. These maps employ a grey–red color scheme, where grey areas represent lower variable values and increasingly red tones indicate higher values. For example, patients in nodes on the right side of the membership map (panel **a**) had moderate to severe cognitive impairments (see *CognitiveStable* variable) and had low Barthel and high NIHSS indices.

Similarly, Fig 6 shows the distribution of clinical program predictions (e.g., high intensity hospital based) superimposed on the membership maps, with panels (a), (b), and (c) corresponding to the program assignment provided by the annotation of the medical team (see Clinical Annotation), the in-house CART model, and our best-performing model, XGBoost, respectively. For each panel in Fig 6, the schematic on the left-hand side shows the main clinical program node assignment; the visualizations on the right-hand side show the prevalence of each program per node without aggregation, with intensive rehabilitation options colored in red, intermediate and intermediate/low pathways colored in orange, and low intensity or no rehabilitation program needed on green. For example, patients belonging to the top upper-left node in panel

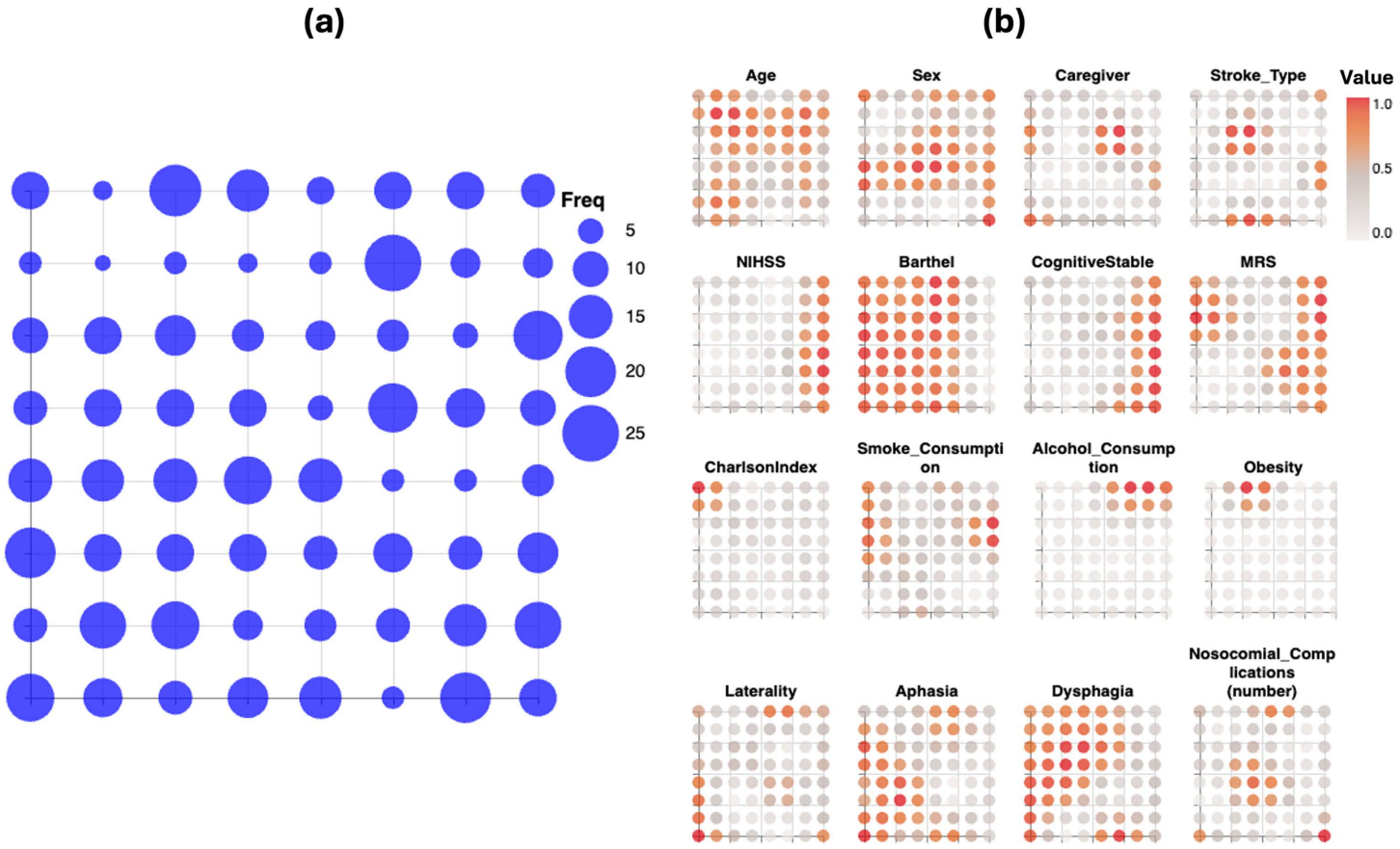

Fig 5. Membership and reference GTM maps. (a) Membership map demonstrating the latent patient node assignments, with the larger the node the more patients that have been assigned to that cluster. (b) The reference maps that indicate how the values of each clinical variable affect the distribution of node assignments. The red color coding denotes greater numeric variables, and the following patient characteristics: male sex; absence of aphasia/dysphagia; hemorrhagic stroke type; smoking/alcohol consumption; obesity; right-sided stroke laterality; absence of caregiver. The maps can be read in three steps: first, a node or region of interest is located on the membership map (Fig 5a); second, the clinical profile of the patients in that region is read from the reference maps (Fig 5b); third, the rehabilitation program assigned to that region is compared across the three sources in Fig 6, showing where the clinicians, the in-house decision tree and XGBoost agree or diverge.

https://doi.org/10.1371/journal.pone.0363251.g005

(**a**) are more likely to exhibit the characteristics indicated by the membership map in Fig 5 (e.g., high Charlson index) and will be almost certainly assigned to an intensive or intermediate hospital-based program according to the clinical team's annotation.

Exemplary Sankey diagrams in Fig 7a also demonstrate the program allocation agreements and disagreements between the annotated labels of the medical team and the in-house decision tree, and panel (**b**) depicts the relationships between the medical team and XGBoost labels. These labels were derived from latent nodes generated from the fitted GTM using the process outlined in the methodology. For example, Fig 7b shows that patient allocations made by the clinical team for subacute rehabilitation programs are, on visual inspection, largely consistent with the XGBoost model predictions. However, this decision-making process is less consistent when comparing the clinical team's annotations with the in-house CART assignments (panel **a**), where several patients classified by the clinicians as requiring intensive hospital-based programs (in red, left-hand side) were instead allocated to intermediate options by the in-house CART model (in orange, right-hand side).



Fig 6. Outcome labels superimposed on to the latent membership maps. Panel (a) contains the distribution of the clinical program predictions (e.g., high intensity hospital based) as annotated by the MDs; (b) contains the distribution of the labels defined by the in-house CART model; and (c) contains the distribution of labels assigned by XGBoost.

https://doi.org/10.1371/journal.pone.0363251.g006



Fig 7. Sankey diagrams demonstrate the similarities and differences between the labels assigned to patients via different methods. Panel (a) shows the assignment by the annotation of the clinical team on the left, and the in-house CART assignment on the right; similarly, panel (b) shows the program allocation assigned by clinical team on the left, and the XGBoost model predictions on the right.

https://doi.org/10.1371/journal.pone.0363251.g007

Discussion

Early and appropriate admission to a rehabilitation setting following a stroke event has been identified as a key predictor of optimal functional recovery [30,31]. We conducted here a supervised learning study, with the patient's subacute rehabilitation program being assigned as the target outcome, in an expert-annotated dataset of 700 stroke patients. The experiment was framed as a multiclass classification problem at a multi-level rehabilitation taxonomy (e.g., from no need for hospitalization to high intensity hospital-based schemes). Considered candidate predictors included an array of patients' demographic, neurological, functional, and in-hospital complication variables. Pairwise correlation coefficients and feature importance analysis dictated the selection of the most appropriate variables for decision making. We have tested a number of different ML models, such as the XGBoost, that are well established in clinical prediction research for their balance between accuracy and interpretability [46,47]. We followed a stratified cross-validation approach which is well supported in recent literature for handling multiclass imbalance and reducing variance in performance estimates [26,27]. Model performance was quantified with standard multiclass indicators, such as the f1-score and balanced accuracy, that capture both global and class-specific performance and are widely recommended for clinical classification tasks [27]. An empirical

in-house decision tree was included as a transparent baseline for comparison with the developed ML and data-driven CART models. Generative Topographic Mapping (GTM) was applied to reduce dimensionality without losing important data relationships and interpret patient data trends in a meaningful and transparent way, identifying key predictor variables shared across patient clusters and revealing underlying patterns in the predictions of our models for rehabilitation program allocation. Sankey diagrams were constructed to highlight similarities and differences between predictions coming from the different models developed, as aggregated via the fitted GTM model. XGBoost, Random Forest and both CART models exceeded the institutional tree in overall accuracy and balanced accuracy when evaluated against the expert annotations. This comparison does not establish improved clinical outcomes. From a clinical perspective, these findings highlight the importance of combining neurological severity, cognitive status, functional independence, and social support variables during early rehabilitation planning, while also demonstrating the potential utility of transparent ML-assisted decision-support systems for improving consistency, personalization, and explainability in rehabilitation program allocation.

ML tools and CART models

The XGBoost and the best-performing data-driven CART model (Fig 3) demonstrated moderate predictive performance, with overall accuracies of 0.832 and 0.847, and balanced accuracies of 0.381 and 0.393, respectively. The gap between overall and balanced accuracy indicates uneven performance across the seven categories. As a point of comparison, in published decision tree models to predict home discharge as a binary outcome, the reported accuracies were of the order of 0.788 [7], 0.867 [8], and 0.878 [48]. In our study, the ML and CART models showed broadly comparable predictive performance under the evaluated configurations. Because the ML algorithms were applied using default hyperparameter settings, these results should not be interpreted as a definitive comparison of the maximum achievable performance of the different algorithms. A formal benchmarking analysis would require systematic hyperparameter optimization within an appropriately nested cross-validation framework [49]. We further investigated which of the 49 features considered carried the greatest weight in determining prediction outcomes (Fig 2). These findings may enable the generation of CART models to be tailored to the customary practices of different clinical pathways, depending on which parameters are routinely assessed during admission or early rehabilitation. For example, if a facility routinely logs the FIM score instead of the Barthel Index, the CART models can be adjusted accordingly before they are generated while also considering the relative contribution of each variable or the complexity of any existing decision-making tools. Such an alternative CART structure is illustrated in Fig 4, which also demonstrated comparable performance. XGBoost, Random Forest and both CART models exceeded the institutional benchmark in both overall accuracy and balanced accuracy. This comparison with the institutional tree should not be interpreted as evidence that the data-driven CART provides superior clinical decisions. Rather, it illustrates the potential value of using contemporary, locally derived expert decisions to inform the development or future updating of transparent institutional decision-support tools. In this sense, the proposed framework may serve as a data-driven approach for refining existing tools while preserving the interpretability required for routine clinical use.

In terms of discriminatory capacity, rather than categorizing patients into two classes (i.e., home or non-home options) based on a limited set of clinical variables, our modeling approach allows the identification of patient clusters using admission data, and the allocation of patients across a range of distinct rehabilitation options. This was achieved by identifying and applying clear clinical predictors (see Fig 2 for the most influential parameters) and their corresponding thresholds, that include stroke severity (NIHSS), patient age, social support (i.e., caregiver presence), cognitive function (MoCA), stroke type (hemorrhagic or ischemic), and functional independence (Barthel Index or mRS). The NIHSS score and cognitive function emerged as the two key determinants for program allocation. Consistent with Pérez, et al. [10], who established a threshold of 16 based on cluster analysis, we trichotomized our cohort using NIHSS cutoffs of 13 and 26, followed by a MoCA cutoff score of 18 to further split patients with mild cognitive impairment from those with moderate to severe status (see Fig 3). In parallel, previous studies have also reported age as a negative predictor of functional recovery and rehabilitation effectiveness after stroke [50–53], and as a key determinant for rehabilitation setting allocation

[7,9,53]. In this work, younger patients (on average, under 65.1 years of age) were frequently allocated to high-intensity rehabilitation programs. This age threshold numerically resembles the clinical convention of treating stroke patients over roughly 65 years as a distinct, higher-risk group with reduced rehabilitation tolerance [52]; however, as no rehabilitation outcomes were observed in this study, this resemblance reflects the allocation practice captured in the expert annotations and should not be interpreted as independent clinical validation. Similarly, the NIHSS cutoffs of 13 and 26 fall close to the Moderate/Moderate-Severe and Severe boundaries already used to categorize stroke severity in this cohort (see Methods), and the MoCA cutoff of 18 sits within the range conventionally used to flag moderate-to-severe cognitive impairment, but such proximity does not establish that these splits reflect clinical relationships rather than features of the annotation and tree-fitting process. Pre-stroke living arrangements are also reported as the second most critical predictive factor for home discharge [10,53], with caregiver presence having also been highlighted in this work as a differentiator between SNF and IRF options. Similarly to Pérez, et al. [10] who used the Barthel Index with a cutoff score of 10 to assign patients to lower- and higher-impact pathways, we defined a threshold of 92 to differentiate the "low/intermediate intensity" and "no need for rehabilitation" leaf nodes (Fig 3), and an mRS score of 2 to distinguish between "intermediate" and "high-intensity hospital-based" outcomes (Fig 4). Stroke type, however, is a variable often overlooked in studies employing decision trees or patient clustering approaches, whether for predicting functional outcomes or discharge destinations. In our analysis, the distinction between hemorrhagic and ischemic events emerged as the sixth most important prognostic feature. In fact, patients who experienced hemorrhagic strokes were more frequently allocated to SNFs, particularly when a caregiver is not present, possibly reflecting the annotating clinicians' expectation of a lower likelihood of full recovery (Fig 3). This allocation pattern is consistent with the documented tendency of hemorrhagic strokes to present with greater severity, higher mortality rates, and poorer long-term neurological outcomes [54,55]. However, the present data contain no rehabilitation outcomes with which to test such a prognostic link. Finally, other key predictors included the presence of early seizures during the early rehabilitation of patients, along with the Charlson index, nosocomial complications, aphasia and dysphagia as factors with a moderate level of importance (see Fig 2). This is well aligned with current decision-making practices, as patients with severe stroke events that have suffered from nosocomial complications or other infections tend to have poorer outcomes, and reduced or interrupted participation in rehabilitation [52]. In this regard, the presence of aphasia and dysphagia is likewise associated with nasogastric tube feeding or gastrostomy, which negatively impact recovery, contribute to long-term disability, and shift rehabilitation goals toward lower-intensity programs aiming at restoring a patient's ability to live and feed independently [20]. In every case, the decision thresholds in this study are dependent on the annotation practices of a single institution, they should not be interpreted as universal clinical decision rules, and require validation on independent, real-world cohorts before clinical use.

Generative topographic mapping

As an exploratory and hypothesis-generating component of this study producing descriptive observations, GTM aided in the identification of key patterns within the data, as evidenced by the distribution of stroke rehabilitation strategies across the latent space (Fig 6). Patients requiring intensive hospital-based rehabilitation were predominantly grouped in the lower-left and upper-central regions of the map (as indicated by the red nodes in Fig 6**, panel a**), while the remaining nodes primarily contained patients assigned to intermediate hospital-based programs. Moreover, Fig 6**, panel b,** shows that the in-house tool leads to a more conservative approach in assigning patients to intensive programs, with most latent nodes corresponding instead to intermediate hospital-based rehabilitation (i.e., greater number of orange nodes when compared to panel **a**). A key utility of this approach is demonstrated by comparing the different regions in the membership maps. For example, see two distinct clusters containing patients most frequently assigned to intensive hospital-based rehabilitation programs, better highlighted in Fig 8 (i.e., a modified version of Fig 6). Following the reference maps in Fig 5, even though patients in these two clusters were repeatedly assigned to the same rehabilitation label, they may share dissimilar clinical characteristics per cluster: the bottom cluster contains patients without aphasia, with high Barthel Index

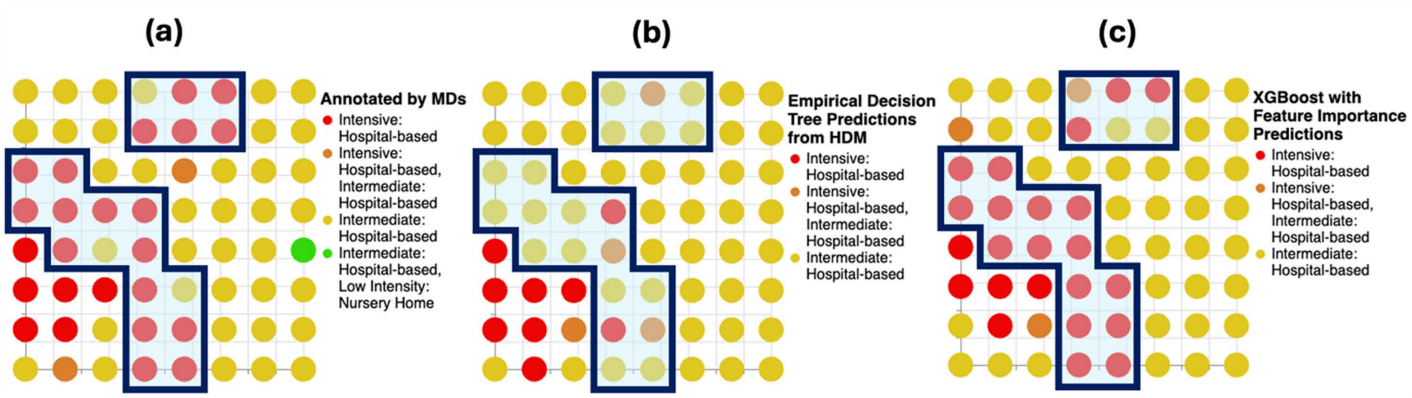

Fig 8. Demonstration of the key areas/cluster with differences in program allocation distributions in the latent space based on the source of the prediction: (a) labels provided by the MDs; (b) labels determined by the in-house decision tree (c) labels assigned by the XGBoost model.

https://doi.org/10.1371/journal.pone.0363251.g008

scores, only mild cognitive impairment, and low NIHSS scores; however, those allocated in the upper-central cluster were additionally more likely to present with obesity and report alcohol consumption.

This analysis can also aid in examining differences in patient characteristics between nodes and assessing how these variations influence rehabilitation program assignments across different predictive models. For example, as illustrated now in Fig 9 (bottom panels), a cluster of four latent nodes was identified, showing consistent rehabilitation

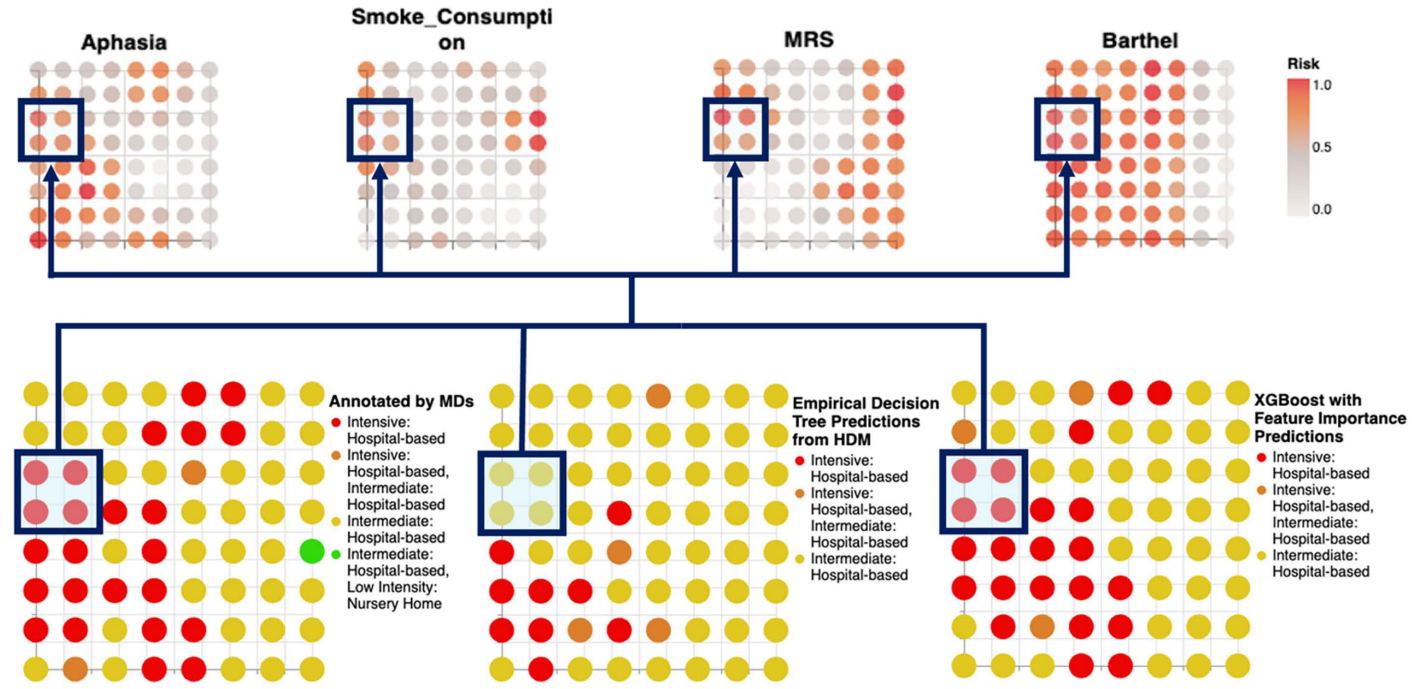

Fig 9. A comparison between a selected cluster of 4 nodes in the latent space and how the rehabilitation program assignments differ depending on how they were generated. Key reference maps of variables linked to these differences are also highlighted (top schematics). The red color coding denotes greater values in the numeric variables, the absence of aphasia, and smoking consumption.

https://doi.org/10.1371/journal.pone.0363251.g009

program assignments between the medical team and the XGBoost model (red nodes, intensive hospital-based allocation), but differing from those assigned by the in-house decision tree (orange nodes, middle panel). The examination of the reference maps (Fig 9, top panels) reveals that the in-house CART does not allocate patients with high Barthel and mRS scores—who are also likely to have no aphasia and to be smokers—to intensive hospital-based programs. Instead, it often assigns them to the intermediate hospital-based category. This contrasts with the labels produced by the annotation from the medical team and the prediction from the XGBoost model, further supporting the observation that the current in-house CART model adopts a more conservative approach when assigning patients to intensive rehabilitation programs (also refer to Sankey diagrams in Fig 7). In practice, a new patient could be projected onto the trained map to identify the node that best matches their clinical profile. The programs assigned to patients in that node, and whether the clinicians and models agree there, would indicate how similar patients were allocated; regions where the sources disagree (e.g., Fig 9) could be flagged for closer clinical review rather than relying on a single model's recommendation.

Limitations and general suggestions

Certainly, limitations exist within this line of work. Firstly, although the cohort of 700 patients provided adequate representation of the most frequently allocated programs, the rarest programs contained very few cases, which limited the models' ability to learn and be evaluated on these categories (for example, low intensity: home-based rehabilitation was represented by 3 cases, and no need for hospitalization by a single case). Because a category represented by a single case is absent from the training data whenever that case is held out, its recall was zero for all developed models by construction, whereas the fixed institutional tree was not subject to this constraint; this asymmetry should be considered when comparing balanced accuracy between the developed models and the institutional benchmark. Future work should expand the dataset, particularly with profiles corresponding to imbalanced programs. Secondly, the list of clinically relevant parameters and metrics is extensive and can be expanded as needed. In clinical practice, easily obtainable metrics are needed to make agile decisions about the type of rehabilitation program to follow. As an example, the duration of acute hospitalization has also been reported as a factor influencing discharge pathways, with shorter hospital stays being associated with an increased likelihood of home discharge [8]. In general, we aimed to represent each patient's clinical status using quantitative metrics that are robust, reproducible, and easy to interpret. Nonetheless, integrating qualitative variables—such as the patient's ability to walk independently or perform self-grooming activities (e.g., washing hands, brushing teeth, combing hair)—may refine or produce alternative predictive models.

Most importantly, all models were developed and tested only on the synthetic dataset described above, built from literature-reported distributions rather than real patients, and it cannot capture the complex correlations, missing data, and clinical heterogeneity seen in actual stroke cohorts. Our analysis was tailored to the specific rehabilitation modalities and resources available within a single clinical center, and the dataset was annotated based on the expertise and judgment of practitioners from a single institution. As a result, the findings are optimized for a certain clinical audience and may not be directly generalized to other practices. Thus, the accuracy reported here therefore reflects how well the models reproduce the panel's labels, not how they would perform in broad clinical use, and any claim that they improve rehabilitation allocation needs confirmation on real-world cohorts before it can be taken at face value. Nonetheless, our objective was to develop a stroke patient dataset encompassing multiple realistic and clinically relevant variables—one that can be appraised, adapted, and expanded by other research groups. Rehabilitation pathways may be likewise added or excluded according to local capacity, while each clinical team may re-evaluate and assign patients in accordance with their regional guidelines, healthcare-system structure, insurance-driven resource availability, and institutional protocols and practices—a process that can be further optimized through GTM analysis. In keeping with the framework proposed in this study, customized CART models can thereby be generated with predefined depth, complexity, and feature composition, supporting reproducibility and contextual adaptation across diverse rehabilitation settings.

Conclusions

Predictions for the subacute rehabilitation pathway using ML and CART models were generated from an extended, expert-annotated dataset and benchmarked against an institutional decision tree currently used in clinical practice. These findings should be interpreted in light of two principal constraints: all data are synthetic, generated from literature-reported distributions rather than real-world observations, and all expert annotations originate from a single institution's clinical practice. Given the class imbalance across the seven rehabilitation programs, our models demonstrated moderate predictive performance, with balanced accuracies of 0.381 and 0.393 for the ML and CART models, respectively, with overall accuracies of 0.832 and 0.847. Both exceeded the institutional benchmark on these global metrics, although classification performance remained uneven across pathways. The most explanatory variables based on a feature importance analysis for determining the program allocation during subacute rehabilitation were metrics of cognitive function, the NIHSS score, patient age, and the presence of early seizures during acute rehabilitation. These findings further highlight the importance of neurological severity, cognitive status, and functional independence in supporting rehabilitation allocation decisions. As a methodological proof-of-concept, this framework illustrates how such models could eventually support optimization of patient discharge pathways, assist in planning rehabilitation programs and determining their duration or intensity (e.g., weekly physiotherapy or speech therapy sessions during home-based rehabilitation programs), and help define the most appropriate care environment. Realizing these potential benefits, including any reduction in costs for clinical centers and families, will require validation in real clinical populations before routine implementation can be considered. This work aims to support the development of transparent and insightful predictive models that can be recalibrated to reflect local practices, resource availability, and rehabilitation protocols. Importantly, the proposed framework was designed to remain adaptable and explainable across different rehabilitation settings and institutional workflows. Testing and validating the produced CART models on an independent cohort is an essential next step before they can be considered for clinical use, not an optional extension. The methodological framework, analyses, and findings presented in this study are accompanied by an expert-annotated dataset that can be further optimized, adapted, and generalized across diverse clinical settings. Ultimately, such an approach may contribute to more data-driven, individualized, highly explainable, and context-sensitive decision-support in post-stroke rehabilitation planning.

Supporting information

S1 Table. Class-specific performance of the evaluated classification models across the seven rehabilitation programs. Precision, recall (sensitivity), and F1-score are reported for the institutional in-house CART, Random Forest, multinomial logistic regression, k-nearest neighbors (KNN), support vector machine (SVM), linear discriminant analysis (LDA), XGBoost, and the two data-driven CART models shown in Fig 3 and Fig 4 (maximum depth = 5). For the developed machine-learning and CART models, class-specific metrics were aggregated across the repeated stratified 10-fold cross-validation procedure. The in-house CART was evaluated descriptively against the expert-annotated outcomes on the complete dataset, consistent with the primary analysis. Values closer to 1 indicate better class-specific classification performance.
(DOCX)

Acknowledgments

The authors wish to acknowledge the contributions of **Dr. Roser Boza-Gomez** and **Dr. Cindry Ramirez-Fuentes** from the Rehabilitation Department at University Hospital del Mar (Barcelona, Spain) that participated in the review and annotation of the dataset in this study.

Author contributions

Conceptualization: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Dimitrios Tsaopoulos, Anna Guillen-Sola, Esther Duarte, Sandra Ortega-Martorell, Ivan Olier.

Data curation: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Anna Guillen-Sola, Esther Duarte, Ryan A.A. Bellfield, Sandra Ortega-Martorell.

Formal analysis: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Anna Guillen-Sola, Esther Duarte, Ryan A.A. Bellfield, Sandra Ortega-Martorell.

Funding acquisition: Spiros Nikolopoulos, Dimitrios Tsaopoulos, Anna Guillen-Sola, Esther Duarte, Sandra Ortega-Martorell, Ivan Olier.

Investigation: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Anna Guillen-Sola, Esther Duarte, Ryan A.A. Bellfield, Ivan Olier.

Methodology: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Spiros Nikolopoulos, Dimitrios Tsaopoulos, Anna Guillen-Sola, Ryan A.A. Bellfield, Sandra Ortega-Martorell, Ivan Olier.

Project administration: Dimitrios Sokratis Komaris, Spiros Nikolopoulos, Dimitrios Tsaopoulos, Ioannis Kompatsiaris.

Resources: Spiros Nikolopoulos, Dimitrios Tsaopoulos, Anna Guillen-Sola, Esther Duarte, Ioannis Kompatsiaris.

Software: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Ryan A.A. Bellfield, Sandra Ortega-Martorell.

Supervision: Dimitrios Sokratis Komaris, Spiros Nikolopoulos, Dimitrios Tsaopoulos, Sandra Ortega-Martorell, Ivan Olier, Ioannis Kompatsiaris.

Validation: Dimitrios Sokratis Komaris, Anna Guillen-Sola, Esther Duarte.

Visualization: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Ryan A.A. Bellfield.

Writing – original draft: Dimitrios Sokratis Komaris, Panagiotis Anagnostou, Ryan A.A. Bellfield.

Writing – review & editing: Dimitrios Sokratis Komaris, Spiros Nikolopoulos, Dimitrios Tsaopoulos, Anna Guillen-Sola, Esther Duarte, Sandra Ortega-Martorell, Ivan Olier, Ioannis Kompatsiaris.

References

1. Winstein CJ. Guidelines for adult stroke rehabilitation and recovery. Stroke. 2016;47(6):e98–169. https://doi.org/10.1161/STR.0000000000000098
2. Heran M, Lindsay P, Gubitz G, Yu A, Ganesh A, Lund R, et al. Canadian stroke best practice recommendations: Acute stroke management, 7th edition practice guidelines update, 2022. Can J Neurol Sci. 2024;51(1):1–31. https://doi.org/10.1017/cjn.2022.344 PMID: 36529857
3. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: The stroke recovery and rehabilitation roundtable taskforce. Int J Stroke. 2017;12(5):444–50. https://doi.org/10.1177/1747493017711816 PMID: 28697708
4. Miller EL, Murray L, Richards L, Zorowitz RD, Bakas T, Clark P, et al. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient. Stroke. 2010;41(10):2402–48. https://doi.org/10.1161/str.0b013e3181e7512b
5. MacDonald SL, Linkewich E, Bayley M, Jeong IJ, Fang J, Fleet JL. The association between inpatient rehabilitation intensity and outcomes after stroke in Ontario, Canada. Int J Stroke. 2024;19(4):431–41. https://doi.org/10.1177/17474930231215005 PMID: 38078378
6. Wasti SA, Surya N, Stephan KM, Owolabi M. Healthcare settings for rehabilitation after stroke. Clin Pathways in Stroke Rehabilit. 2021. p. 261–82.
7. Imura T, Iwamoto Y, Inagawa T, Imada N, Tanaka R, Toda H, et al. Decision tree algorithm identifies stroke patients likely discharge home after rehabilitation using functional and environmental predictors. J Stroke Cerebrovasc Dis. 2021;30(4):105636. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105636 PMID: 33545520
8. Nakaizumi D, Miyata S, Uchiyama K, Takahashi I. Development and validation of a decision tree analysis model for predicting home discharge in a convalescent ward: A single institution study. Phys Ther Res. 2024;27(1):14–20. https://doi.org/10.1298/ptr.E10267 PMID: 38690531
9. Veerbeek JM, Ottiger B, Cazzoli D, Vanbellingen T, Nyffeler T. Speed up discharge planning at the acute stroke unit: A development and external validation study for the early prediction of discharge home. Front Neurol. 2022;13. https://doi.org/10.3389/fneur.2022.999595

10. Pérez LM, Inzitari M, Quinn TJ, Montaner J, Gavaldà R, Duarte E, et al. Rehabilitation profiles of older adult stroke survivors admitted to intermediate care units: A multi-centre study. PLoS One. 2016;11(11):e0166304. https://doi.org/10.1371/journal.pone.0166304 PMID: 27829011
11. Cui Y, Xiang L, Zhao P, Chen J, Cheng L, Liao L, et al. Machine learning decision support model for discharge planning in stroke patients. J Clin Nurs. 2024;33(8):3145–60. https://doi.org/10.1111/jocn.16999 PMID: 38358023
12. Vellido A, Lisboa PJ, Meehan K. Segmenting the e-commerce market using the generative topographic mapping. In: Mexican International Conference on Artificial Intelligence. 2000. 470–81.
13. Ortega-Martorell S, Bellfield RAA, Harrison S, Dyke D, Williams N, Olier I. Mapping the global free expression landscape using machine learning. SN Appl Sci. 2023;5(12). https://doi.org/10.1007/s42452-023-05554-x
14. Bellfield RAA, Hormiga PR, Olier I, Lotto R, Jones I, Lip GYH, et al. AI-driven clustering and visualization of electrocardiogram signals to enhance screening for atrial fibrillation: The supermarket/hypermarket opportunistic screening for atrial fibrillation study. Heart Rhythm O2. 2025;6(10):1601–12. https://doi.org/10.1016/j.hroo.2025.07.003 PMID: 41169967
15. Bellfield RAA, Olier I, Lotto R, Jones I, Dawson EA, Li G, et al. AI-based derivation of atrial fibrillation phenotypes in the general and critical care populations. EBioMedicine. 2024;107:105280. https://doi.org/10.1016/j.ebiom.2024.105280 PMID: 39153412
16. Ortega-Martorell S, Olier I, Ohlsson M, Lip GYH, TARGET Consortium. TARGET: A major european project aiming to advance the personalised management of atrial fibrillation-related stroke via the development of health virtual twins technology and artificial intelligence. Thromb Haemost. 2025;125(1):7–11. https://doi.org/10.1055/a-2438-5671 PMID: 39401519
17. Komaris DS. Comparative analysis of muscle synergies in gait of stroke patients and healthy controls. Front Human Neurosci. 2025;19:1601147.
18. Ortega-Martorell S, Olier I, Lip GYH, TARGET consortium. A European network to develop virtual twin technology for personalized stroke management in atrial fibrillation: The TARGET consortium. Eur Heart J. 2025;46(3):229–32. https://doi.org/10.1093/eurheartj/ehae673 PMID: 39565293
19. Ortega-Martorell S, Olier I, Ohlsson M, Lip GYH, TARGET Consortium. Advancing personalised care in atrial fibrillation and stroke: The potential impact of AI from prevention to rehabilitation. Trends Cardiovasc Med. 2025;35(4):205–11. https://doi.org/10.1016/j.tcm.2024.12.003 PMID: 39653093
20. Albu S, Izcara López de Murillas E, Secanell Espluga M, Jimenez Crespo A, Kumru H. Clinical profiles and functional outcomes in elderly stroke survivors undergoing neurorehabilitation: A retrospective cohort study. Egypt J Neurol Psych Neurosurg. 2024;60(1). https://doi.org/10.1186/s41983-024-00877-x
21. Ortiz GA, Sacco RL. National Institutes of Health Stroke Scale (NIHSS). Wiley StatsRef: Statistics reference online.
22. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9. https://doi.org/10.1111/j.1532-5415.2005.53221.x PMID: 15817019
23. Haggag H, Hodgson C. Clinimetrics: Modified Rankin Scale (mRS). J Physiother. 2022;68(4):281. https://doi.org/10.1016/j.jphys.2022.05.017 PMID: 35715375
24. Collin C, Wade D, Davies S, Horne V. The Barthel ADL Index: A reliability study. Int Disability Stud. 1988;10(2):61–3.
25. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: A quantitative review. Arch Phy Med Rehab. 1996;77(12):1226–32.
26. Shariff E, Nazish S, Zafar A, Shahid R, Aljaafari D, Soltan NM, et al. Characteristics of stroke-related seizures and their predictive factors: A tertiary care center experience. Ann Afr Med. 2024;23(3):285–90. https://doi.org/10.4103/aam.aam_15_23 PMID: 39034547
27. Jeyaseelan RD, Vargo MM, Chae J. National Institutes of Health Stroke Scale (NIHSS) as an early predictor of poststroke dysphagia. PM R. 2015;7(6):593–8. https://doi.org/10.1016/j.pmrj.2014.12.007 PMID: 25556573
28. Barow E, Probst A-C, Pinnschmidt H, Heinze M, Jensen M, Rimmele DL, et al. Effect of comorbidity burden and polypharmacy on poor functional outcome in acute ischemic stroke. Clin Neuroradiol. 2023;33(1):147–54. https://doi.org/10.1007/s00062-022-01193-8 PMID: 35831611
29. Chen W, Li D. Comorbidity and outcomes among hospitalized patients with stroke: A nationwide inpatient analysis. Front Neurol. 2023;14:1217404. https://doi.org/10.3389/fneur.2023.1217404 PMID: 37915378
30. Lawrence ES, Coshall C, Dundas R, Stewart J, Rudd AG, Howard R, et al. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke. 2001;32(6):1279–84. https://doi.org/10.1161/01.str.32.6.1279 PMID: 11387487
31. Filipska A, Bohdan B, Wieczorek PP, Hudz N. Chronic kidney disease and dialysis therapy: Incidence and prevalence in the world. PHAR. 2021;68(2):463–70. https://doi.org/10.3897/pharmacia.68.e65501
32. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378 PMID: 38626948
33. Roffman CE, Buchanan J, Allison GT. Charlson comorbidities index. J Physiother. 2016;62(3):171. https://doi.org/10.1016/j.jphys.2016.05.008 PMID: 27298055
34. Komaris D. Data-driven clinical decision trees and generative topographic mapping to guide stroke rehabilitation program allocation. https://doi.org/10.6084/m9.figshare.30588821.v1
35. Plummer P, Feld JA, Mercer VS, Ni P. Brief composite mobility index predicts post-stroke fallers after hospital discharge. Front Rehabil Sci. 2022;3:979824. https://doi.org/10.3389/fresc.2022.979824 PMID: 36275923

36. Büyükkeçeci M, Okur MC. A comprehensive review of feature selection and feature selection stability in machine learning. Gazi Univ J Sci. 2023;36(4):1506–20. https://doi.org/10.35378/gujs.993763
37. Agresti A, Kateri M. Categorical data analysis. Int Encycl Stat Sci. Springer; 2011. 206–8.
38. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Front Bioinform. 2022;2(Review). https://doi.org/10.3389/fbinf.2022.927312
39. Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. 2018. https://arxiv.org/abs/1802.03888
40. Muflikhah L, Fatyanosa TN, Widodo N, Perdana RS, Solimun, Ratnawati H. Feature selection for hypertension risk prediction using XGBoost on single nucleotide polymorphism data. Healthc Inform Res. 2025;31(1):16–22. https://doi.org/10.4258/hir.2025.31.1.16 PMID: 39973033
41. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining. San Francisco, California, USA; 2016. https://doi.org/10.1145/2939672.2939785
42. Allgaier J, Pryss R. Cross-validation visualized: A narrative guide to advanced methods. Machine learning and knowledge extraction. 2024;6(2):1378–88. https://doi.org/10.3390/make6020065
43. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and regression trees. Chapman and Hall/CRC; 2017.
44. Bishop CM, Svensén M, Williams CKI. GTM: The generative topographic mapping. Neural Comp. 1998;10(1):215–34. https://doi.org/10.1162/089976698300017953
45. Olier I, Ortega-Martorell S, Margereson G, Bellfield RA, Welters ID, Lip GYH. The integrated multiple event representation framework (IMERF): A case study on critically-ill patients. Comput Biol Med. 2025;198(Pt A):111196. https://doi.org/10.1016/j.compbiomed.2025.111196 PMID: 41075478
46. Huang J-C, Tsai Y-C, Wu P-Y, Lien Y-H, Chien C-Y, Kuo C-F, et al. Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput Methods Programs Biomed. 2020;195:105536. https://doi.org/10.1016/j.cmpb.2020.105536 PMID: 32485511
47. Hong W, Zhou X, Jin S, Lu Y, Pan J, Lin Q, et al. A comparison of XGBoost, random forest, and nomograph for the prediction of disease severity in patients with COVID-19 pneumonia: Implications of cytokine and immune cell profile. Front Cell Infect Microbiol. 2022;12:819267. https://doi.org/10.3389/fcimb.2022.819267 PMID: 35493729
48. Iwai N, Aoyagi Y. Discharge index and prediction for stroke patients in the post-acute stage: Evaluation of the usefulness of Nichijo-seikatsu-kino-hyokahyo. Jap J Comprehen Rehab Sci. 2012;3(0):37–41. https://doi.org/10.11336/jjcrs.3.37
49. Ying X. An overview of overfitting and its solutions. J Phys: Conf Ser. 2019;1168:022022. https://doi.org/10.1088/1742-6596/1168/2/022022
50. Bagg S, Pombo AP, Hopman W. Effect of age on functional outcomes after stroke rehabilitation. Stroke. 2002;33(1):179–85. https://doi.org/10.1161/hs0102.101224
51. Li TK-T, Ng BH-P, Chan DY-L, Chung RS-F, Yu K-K. Factors predicting clinically significant functional gain and discharge to home in stroke inpatients after rehabilitation - A retrospective cohort study. Hong Kong J Occup Ther. 2020;33(2):63–72. https://doi.org/10.1177/1569186120979428 PMID: 33815025
52. Zucchella C, Consilvio M, Iacoviello L, Intiso D, Tamburin S, Casale R, et al. Rehabilitation in oldest-old stroke patients: A comparison within over 65 population. Eur J Phys Rehabil Med. 2019;55(2):148–55. https://doi.org/10.23736/S1973-9087.18.05297-8 PMID: 30160435
53. Tanwir S, Montgomery K, Chari V, Nesathurai S. Stroke rehabilitation: availability of a family member as caregiver and discharge destination. Eur J Phys Rehabil Med. 2014;50(3):355–62. PMID: 24518145
54. Chiu D, Peterson L, Elkind MSV, Rosand J, Gerber LM, Silverstein MD, et al. Comparison of outcomes after intracerebral hemorrhage and ischemic stroke. J Stroke Cerebrovasc Dis. 2010;19(3):225–9. https://doi.org/10.1016/j.jstrokecerebrovasdis.2009.06.002 PMID: 20434051
55. Andersen KK, Olsen TS, Dehlendorff C, Kammersgaard LP. Hemorrhagic and ischemic strokes compared. Stroke. 2009;40(6):2068–72. https://doi.org/10.1161/STROKEAHA.108.540112