PRECISE4Q

PREDICTIVE MODELLING IN STROKE

DELIVERABLE

Project Acronym: **Precise4Q**

Grant Agreement number: **777107**

Project Title: **Personalised Medicine by Predictive Modelling in Stroke for better Quality of Life**

D5.5 – Stroke rehabilitation/reintegration study performed

Revision: 1.0

Authors and Contributors	Alejandro García (GUT); Helard Becerra (UCD); Katryna Cisek (TUD); Thi Nguyet Que Nguyen (TUD), Eloy Opisso (GUT), Josep Maria Tormos (GUT), Joan Sauri (GUT), Montserrat Bernabeu (GUT).			
Responsible Author	Alejandro García		**Email**	agarciar@guttmann.com
	Beneficiary	GUT	**Phone**	+34 934 97 77 00- 2265

Project co-funded by the European Commission within **H2020-SC1-2016-2017/SC1-PM-17-2017**		
Dissemination Level		
PU	Public, fully open	
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Revision History, Status, Abstract, Keywords, Statement of Originality

Revision History

Revision	Date	Author	Organisation	Description
v0.1	18/07/22	Alejandro García (GUT) Katryna Cisek (TUD)	GUT TUD	Initial outline of the document. Key figures and tables for the cognitive models.
v0.2	19/07/22	All contributors	UCD, TUD, GUT	Comments on v0.1
v0.3	18/08/22	Alejandro García (GUT)	GUT	First draft of sections 1, 2, 3, 4 related to the cognitive models.
v0.4	21/08/22	Alejandro García (GUT)	GUT	Updated sections 2, 3, 4 related to the motor models. First draft of sections 5, 6.
v0.5	23/08/22	Helard Becerra (UCD)	UCD	Updated sections 5, 6. Added executive summary.
v1.0	25/08/22	All contributors	UCD, TUD, GUT, UM	Comments on v0.8
v1.0	29/08/22	All contributors	GUT, UCD, TUD,	Final version

Date of delivery	Contractual:	31.08.2022	Actual:	31.08.2022
Status	final ☒ /draft ☐			

Abstract (for dissemination)	This deliverable describes the main results of the validation activities performed in WP5 of the predictive models developed for the personalised rehabilitation and reintegration stages from WP4. The efforts were focused using different datasets than those used for the derivation models of the four main clinical use cases. The followed approach was initially described in D5.4. Therefore, two use cases were validated in the context of cognitive and motor inpatient rehabilitation and the other two on social risk and community integration trajectories of community-dwelling stroke survivors. These four validation cohorts showed reliable predictions both when considering unadjusted models and also when adjusted using previously reported confounders. Such models therefore could reliably support clinical professionals at developing and deploying personalised rehabilitation and reintegration programs. This document is structured in 4 core chapters, each one presenting the validation activities and results for each of the main developed models: cognitive rehabilitation (Chapter 2), motor rehabilitation (Chapter 3), community integration social risk (Chapter 4) and community integration long term trajectories (Chapter 5).
Keywords	Validation, cognitive, motor,rehabilitation, community integration,

	personalised medicine

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Content

List of Figures

Executive Summary

This deliverable describes the validation (using different datasets) of the predictive models developed for the personalised rehabilitation and reintegration stages from WP4. The efforts were focused on the four main clinical use cases, (using the approach described in D5.4) two of them in the context of cognitive and motor inpatient rehabilitation and the other two on social risk and community integration trajectories of community-dwelling stroke survivors. These four validation cohorts showed reliable predictions that could help clinicians at developing personalised rehabilitation and reintegration programs.

Specifically, when addressing cognitive inpatient rehabilitation, two models were trained to predict: the cognitive improvement after therapy, and therapy compliance. Predictions were accompanied by complementary reports to contextualise this information and allow clinicians to evaluate the inner workings of the model. Performance results showed a drop in performance when the models were tested against the validation cohort. However, re-trained versions of both models reported similar, and for some cases, better results when compared to the models presented in D4.8. A comparison of the features' impact showed how certain variables (e.g., admission compliance) had a strong influence across base, validation and re-trained models.

In relation to motor inpatient rehabilitation the validation cohort clearly confirmed the results obtained with the derivation cohort, both when considering all 33 individual FMA-UE items using unadjusted models and when considering the top 3 items with adjusted models. Besides, as presented in Annex I the total number of included participants (287 in the derivation cohort + 109 in the validation cohort) is clearly larger than most of FMA-UE predictive models presented in previous research.

When addressing social risk, results confirm the utility of the models in the real-world clinical scenario, as well as the contribution of not only the EVSF predictors, such as SocialSupport, but also LengthofStay, highlighting that social risk is a complex and multifactorial phenomenon that can vary significantly for patients over the course of stroke rehabilitation and reintegration.

Finally, in relation to Community Integration trajectories the features describing the three classes identified using the derivation cohort were confirmed with the validation cohort. Individuals in Class 2 were the youngest, with the lowest NIHSS, the lowest proportion of hypertension, aphasia, the shorter LOS, the largest proportion of high educational level. Similarly as presented in D4.9 participants in Class 3 present intermediate demographic and clinical results when compared to Class 2 and Class 1. Participants in Class 1 clearly show the lower levels of Community Integration with highest functional dependence at rehabilitation discharge.

1 Introduction

Within PRECISE4Q, Work Package 4 (WP4) covers the development of predictive models building on the outcomes from previous deliverables and other WP efforts. For instance, the set of clinical use cases for each stroke phase is presented in D1.3 and then updated in D4.1. These two deliverables summarise the most relevant scenarios and use cases where the resulting predictive models can be deployed. In D4.2, the target outputs for such predictive models are presented. These output targets guided the development of each predictive model. Then, in D4.3, relevant factors associated with each model and the features derived from them are presented. Such features represent the input for the predictive models developed throughout WP4. Finally, modelling approaches applied in this project in the prevention and acute phases are presented in D4.4, cognitive and motor inpatient rehabilitation models are presented in D4.8 meanwhile reintegration models are presented in D4.9.

The validation activities of such models were also considered for a validation study, first described in D5.4 and to be reported in the deliverable D5.5. This document is structured in 4 core chapters, each one presenting the validation activities and results for each of the main developed models: cognitive rehabilitation (Chapter 2), motor rehabilitation (Chapter 3), community integration social risk (Chapter 4) and community integration long term trajectories (Chapter 5).

.

2 Cognitive Rehabilitation

2.1 Introduction and Background

For the cognitive rehabilitation domain, D4.8 presented two ML models targeting two clinical use cases. The first use case targeted the cognitive deficit reduction of patients after completing the corresponding therapy program. Meanwhile, the second use case covered the prediction of the level of compliance during a patient's therapy. Both models utilised three types of variables as input features: demographic variables, cognitive assessments, and therapy records. The models were trained to solve a binary classification problem using the Extreme Gradient Boosting (XGBoost) algorithm. In addition to standard classification metrics (e.g., Recall, F1, Precision and AUC), a feature impact analysis was conducted to analyse and understand the models' performance and its inner decision process. The implementation of these two models followed the architecture presented in Figure 1. As it was reported in D4.8, the output schema utilised for these two models seeks to provide clinicians with contextualised predictions to help them make informed decisions during patients' cognitive rehabilitation process.

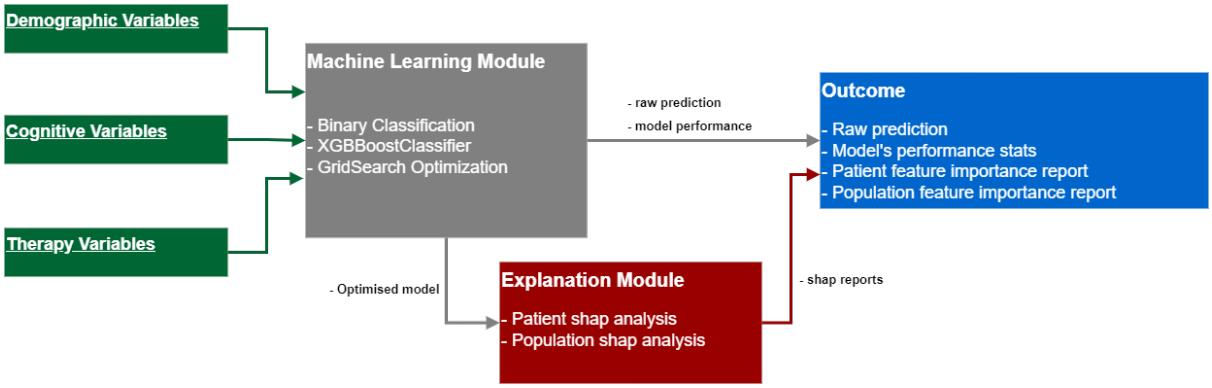


Figure 1 .Diagram for the Cognitive Rehabilitation Models

For the first cognitive model, the overall cognitive status after therapy was represented by the *global improvement* variable. This variable was derived from the sum of cognitive improvement marks across all cognitive assessments at admission and discharge divided by the total number of assessments. This variable was adapted to fit a binary classification problem using the following threshold: Class 0, *global improvement <=0*; Class 1, *global improvement >0*. The optimised version of the model reported Recall scores of 0.71 and F1 scores of 0.61, showing the ability of the model to identify 'at risk' patients belonging to the Class 0 (poor cognitive improvement). The impact feature analysis showed that *time since injury* was the most influential feature for predicting a patient's cognitive improvement. The *admission compliance*, determined by the number of assessments a patient is able to complete at admission, was the second most important feature for this model. Standardised cognitive assessments like *TMT-A* (attention), *Cubes* (visual-construction), and *Ravlt learning* (memory) had also a strong impact for the cognitive improvement prediction. Finally, therapy variables like memory and orientation task proportion contribute to the final prediction of the model.

For the second cognitive model, the therapy compliance of a patient was determined by the *non-executed proportion* variable. This variable was generated by aggregating the number of therapy tasks with performance scores of 0. This variable seeks to report the proportion of therapy tasks that a patients wasn't able to complete, thus, reflecting its therapy compliance. Similarly to the cognitive

improvement model, this variable was adapted to fit a binary classification problem using a threshold defined as: Class 0, *non-executed proportion* <=0.1, Class 1, *non-executed proportion* >0.1. Performance metrics reported Recall scores of 0.67 and F1 scores of 0.69. The impact feature analysis reported that the visual-construction assessment *Cubes* was the most determining feature for predicting the therapy compliance. As observed in the cognitive improvement model, the admission compliance was reported as being the second most influential feature for the model's prediction. The visual-perception assessment *Image* was also reported as having an important influence on the model's prediction. Finally, demographic variables like *age* and *educational level* were also identified as strong feature predictors.

Despite relying on specific clinical settings to build some of the predictor and target variables (e.g., *global improvement, non-executed proportion, admission compliance*), the approach utilised to generate these features is simple enough to be adapted over similar computerized settings. Moreover, the feature importance analysis demonstrated that some of the hand-crafted variables had a strong contribution to the model's final prediction (e.g., memory and orientation task proportion). In addition, it was observed how the trained models replicated some of the clinicians' reasoning process regarding certain variables (e.g., importance of *time since injury* for cognitive improvement). Following the project's plan of activities, a validation of the cognitive rehabilitation models presented in D4.8 is reported in this section. For this validation, unseen data from the Guttmann Rehab Centre is used to test the trained models. In addition, the models are re-trained using this new data. The performance across the D4.8 trained models, the validation and the re-trained version of the models is analysed.

2.2 Validation activities

Validation Patient Cohort

For the validation of the models, input data was gathered from the electronic health records from the Guttmann Rehabilitation Centre (Barcelona, Spain). Following the procedure described in D4.8, variables were organised as demographic, cognitive assessments, cognitive indicators, and therapy variables (see Table 1 from D4.8). These records were collected at the Guttmann Rehabilitation Centre from October 2019 through July 2022. Prior applying the eligibility criteria, a total of 791 registries containing demographic and cognitive variables were gathered from the bateria[1] records; meanwhile, 143811 registries containing the patient's therapy performance were collected from the GNPT[2] platform.

The inclusion criteria, also applied over the training dataset cohort, followed 5 main conditions: 1) being admitted at the rehabilitation centre during the first six months since the stroke episode; 2) having therapy performance records in the GNPT; 3) having a maximum therapy duration of six months, 4) having a minimum therapy duration of 14 days, and 5) having completed both admission and discharge cognitive assessments. For the validation cohort, the fifth inclusion criteria, applied over the training cohort, was not considered due to the low number of registries resulting from it (25 registries). The application of these four conditions resulted in 225 merged registries containing demographic, cognitive, and therapy variables. For the validation cohort, imputation techniques were utilised to deal with missing values. Table 1 below shows the validation cohort information of the total 225 patients' registries.

[1] Standardised cognitive assessments administrated within Guttmann Institut

[2] Guttmann's Neuro Personal Trainer for treatment systematization

Table 1. Validation cohort information including demographics, cognitive and therapy variables. CI: Cognitive Improvement, TC: Therapy Compliance, c: categorical variable, d: derived variable through data aggregation, a/d: administrated at admission and discharge, TB: Test Barcelona, TMT: Trail Making Test, WAIS-III: Wechsler Adult Intelligence Scale 3rd version, RAVLT: Rey Auditory Verbal Learning Test, WCST: Wisconsin Card Sorting, NIHSS: National Institutes of Health Stroke Scale, FIM: Functional Independence Measure, mRS: modified Ranking Scale, BI: Barthel Index, bateria: battery assessment records, gnpt: Guttmann Neuro Personal Trainer records, r: electronic records.

	CI>=0.5 (N=43)	CI<0.5 (N=182)	CI (N=225)	TC>=0.1 (N=153)	TC<0.1 (N=72)	TC (N=225)
Age						
Mean (SD)	49.3 (11.3)	52.8 (11.5)	52.1 (11.5)	50.1 (11.1)	56.4 (11.2)	52.1 (11.5)
Median [Min, Max]	49.4 [19.7,78.9]	54.2 [19.8,86.6]	53.8 [19.7,86.6]	52.5 [19.7,80.6]	56.1 [34,86.6]	53.8 [19.7,86.6]
Time since injury in days						
Mean (SD)	71.63 (39.46)	67.79 (41.22)	68.52 (40.83)	68.86 (40.98)	67.81 (40.79)	68.52 (40.83)
Median [Min, Max]	70 [16,171]	57 [13,176]	58 [13,176]	60 [13,170]	55.5 [13,176]	58 [13,176]
Sex (c)						
Male	26 (60.46%)	127 (69.78%)	153 (68%)	100 (65.35%)	53 (73.6%)	153 (68%)
Female	17 (39.53%)	55 (30.21%)	72 (32%)	53 (34.64%)	19 (26.4%)	72 (32%)
Marital status (c)						
Married	26 (60.46%)	109 (59.89%)	135 (60%)	84 (54.9%)	51 (70.83%)	135 (60%)
Single	10 (23.25%)	38 (20.87%)	48 (21.33%)	37 (24.18%)	11 (15.27%)	48 (21.33%)
Divorce	4 (9.3%)	18 (9.89%)	22 (9.77%)	17 (11.11%)	5 (6.94%)	22 (9.77%)
Separate	2 (4.65%)	13 (7.14%)	15 (6.66%)	10 (3.26%)	5 (6.94%)	15 (6.66%)
Widowed	1 (2.32%)	4 (2.19%)	5 (2.22%)	5 (3.26%)	-	5 (2.22%)
Education level (c)						
Illiterate	-	1 (0.54%)	1 (0.44%)		1 (1.38%)	1 (0.44%)
Read & write	-	5 (2.74%)	5 (2.22%)		3 (4.16%)	5 (2.22%)
Elementary school	12 (27.9%)	59 (32.41%)	71 (31.55%)		29 (40.27%)	71 (31.55%)
High school	21 (48.83%)	78 (42.85%)	99 (44%)		23 (31.94%)	99 (44%)
Superior	10 (23.25%)	39 (21.42%)	49 (21.77%)		16 (22.22%)	49 (21.77%)
NIHSS						
Mean (SD)	13.58 (5.5)	13.36 (5.15)	13.4 (5.21)	13.47 (5.5)	13.25 (4.57)	13.4 (5.21)
Median [Min, Max]	14 [4,25]	13 [2,25]	13 [2,25]	13 [2,25]	13 [5,24]	13 [2,25]
Cognitive FIM						
Mean (SD)	-			24.81 (8.32)	26.67 (6.99)	25.4 (7.95)
Median [Min, Max]	-			27 [5,35]	28.5 [9,35]	27 [5,35]
mRS						
Mean (SD)	-			3.41 (1.07)	3.79 (0.93)	3.53 (1.04)
Median [Min, Max]	-			4 [2,5]	4 [2,5]	4 [2,5]
Barthel Index						
Mean (SD)	-			47.16 (27.83)	38.65 (22.31)	44.31 (26.36)
Median [Min, Max]	-			50 [0,100]	35 [5,90]	45 [0,100]
Admission compliance						
Mean (SD)	0.47 (0.35)	0.46 (0.28)	0.46 (0.3)	0.42 (0.32)	0.55 (0.2)	0.46 (0.3)
Median [Min, Max]	0.58 [0,1]	0.54 [0,1]	0.54 [0,1]	0.46 [0,1]	0.58 [0,1]	0.54 [0,1]
Global improvement						
Mean (SD)	-0.12 (0.13)	0.25 (0.17)	0.18 (0.22)	0.14 (0.2)	0.25 (0.23)	0.18 (0.22)
Median [Min, Max]	-0.06 [-0.44,0]	0.25 [0.06,0.75]	0.12 [-0.44,0.75]	0.06 [-0.44,0.75]	0.31 [-0.44,0.69]	0.12 [-0.44,0.75]
TB Personal Orientation						
Mean (SD)	6.93 (0.26)	6.88 (0.62)	6.89 (0.57)	6.9 (0.62)	6.86 (0.45)	6.89 (0.57)
Median [Min, Max]	7 [6,7]	7 [0,7]	7 [0,7]	7 [0,7]	7 [4,7]	7 [0,7]
TB Spatial Orientation						
Mean (SD)	4.91 (0.48)	4.84 (0.6)	4.85 (0.58)	4.88 (0.52)	4.79 (0.67)	4.85 (0.58)
Median [Min, Max]	5 [2,5]	5 [0,5]	5 [0,5]	5 [0,5]	5 [2,5]	5 [0,5]
TB Temporal Orientation						
Mean (SD)	22.4 (2.06)	22.14 (2.68)	22.19 (2.57)	22.42 (2.54)	21.69 (2.6)	22.19 (2.57)
Median [Min, Max]	23 [13,23]	23 [0,23]	23 [0,23]	23 [0,23]	23 [11,23]	23 [0,23]
Digits Span						
Mean (SD)	5.7 (0.91)	5.76 (1.11)	5.75 (1.07)	5.84 (1.06)	5.54 (1.07)	5.75 (1.07)
Median [Min, Max]	6 [3,8]	6 [0,9]	6 [0,9]	6 [0,9]	6 [3,8]	6 [0,9]
TMT-A						
Mean (SD)	55.95 (26.21)	65.29 (47.87)	63.51 (44.65)	59.5 (38.15)	72.03 (55.37)	63.51 (44.65)
Median [Min, Max]	45 [21,141]	45 [16,325]	45 [16,325]	45 [16,259]	45 [27,325]	45 [16,325]
TB Language Repetition						
Mean (SD)	9.98 (0.15)	9.74 (1.3)	9.78 (1.18)	9.68 (1.42)	10 (0)	9.78 (1.18)
Median [Min, Max]	10 [9,10]	10 [1,10]	10 [1,10]	10 [1,10]	10 [10,10]	10 [1,10]
TB Language Denomination						
Mean (SD)	13.77 (1.15)	13.43 (2.25)	13.49 (2.08)	13.41 (2.42)	13.68 (1.07)	13.49 (2.08)
Median [Min, Max]	14 [7,14]	14 [1,14]	14 [1,14]	14 [1,14]	14 [7,14]	14 [1,14]
TB Language Comprehension						
Mean (SD)	15.91 (0.37)	14.97 (2.75)	15.15 (2.5)	15.12 (2.82)	15.21 (1.64)	15.15 (2.5)
Median [Min, Max]	16 [14,16]	16 [1,16]	16 [1,16]	16 [1,16]	16 [8,16]	16 [1,16]
Digit Span Backwards WAIS-III						
Mean (SD)	3.91 (0.78)	3.91 (0.89)	3.91 (0.87)	4.06 (0.87)	3.6 (0.78)	3.91 (0.87)
Median [Min, Max]	4 [3,7]	4 [2,7]	4 [2,7]	4 [2,7]	4 [2,5]	4 [2,7]
Numbers and Leters WAIS-III						
Mean (SD)	6.51 (1.97)	6.26 (1.51)	6.31 (1.6)	6.4 (1.52)	6.11 (1.76)	6.31 (1.6)

Median [Min, Max]	6 [3,14]	6 [1,12]	6 [1,14]	6 [3,14]	6 [1,11]	6 [1,14]
RAVLT Learning						
Mean (SD)	40.12 (8.4)	37.75 (9.51)	38.2 (9.34)	39.29 (8.98)	35.89 (9.73)	38.2 (9.34)
Median [Min, Max]	42 [11,58]	42 [11,63]	42 [11,63]	42 [11,63]	36.5 [14,55]	42 [11,63]
RAVLT Free Recall						
Mean (SD)	12.86 (3.08)	11.7 (4.48)	11.92 (4.27)	12.44 (3.95)	10.82 (4.72)	11.92 (4.27)
Median [Min, Max]	15 [2,15]	14 [0,15]	14 [0,15]	15 [0,15]	12.5 [0,15]	14 [0,15]
RAVLT Recognition						
Mean (SD)	6.12 (3.2)	5.93 (3.16)	5.96 (3.16)	5.9 (3.19)	6.1 (3.11)	5.96 (3.16)
Median [Min, Max]	4 [2,13]	4 [0,15]	4 [0,15]	4 [0,15]	6 [0,13]	4 [0,15]
PMR						
Mean (SD)	30.28 (9.99)	27.91 (9.99)	28.36 (10.01)	29.51 (9.46)	25.92 (10.74)	28.36 (10.01)
Median [Min, Max]	29 [8,69]	29 [1,68]	29 [1,69]	29 [2,69]	27 [1,56]	29 [1,69]
Images WAIS-III						
Mean (SD)	19.72 (1.26)	18.91 (2.68)	19.07 (2.49)	19.5 (1.61)	18.14 (3.56)	19.07 (2.49)
Median [Min, Max]	20 [12,20]	20 [3,20]	20 [3,20]	20 [10,20]	20 [3,20]	20 [3,20]
Cubes WAIS-III						
Mean (SD)	16.58 (8.79)	12.72 (7.34)	13.46 (7.77)	13.97 (8.52)	12.38 (5.79)	13.46 (7.77)
Median [Min, Max]	10 [10,40]	10 [4,60]	10 [4,60]	10 [7,60]	10 [4,40]	10 [4,60]
Daily sessions						
Mean (SD)	17.56 (17.24)	17.22 (14.39)	17.28 (14.94)	19.35 (16.76)	12.9 (8.62)	17.28 (14.94)
Median [Min, Max]	12 [1,86]	14 [1,78]	14 [1,86]	14 [1,86]	12 [1,40]	14 [1,86]
Length of therapy						
Mean (SD)	71.93 (39.32)	72.54 (33.92)	72.42 (34.92)	73.52 (36.45)	70.08 (31.55)	72.42 (34.92)
Median [Min, Max]	74 [15,163]	70 [16,176]	70 [15,176]	71 [15,176]	66 [18,166]	70 [15,176]
Non executed proportion						
Mean (SD)	0.08 (0.08)	0.09 (0.11)	0.09 (0.11)	0.03 (0.03)	0.22 (0.1)	0.09 (0.11)
Median [Min, Max]	0.06 [0,0.36]	0.05 [0,0.57]	0.05 [0,0.57]	0.02 [0,0.1]	0.18 [0.1,0.57]	0.05 [0,0.57]
Attention task proportion						
Mean (SD)	0.11 (0.09)	0.15 (0.15)	0.14 (0.14)	-	-	-
Median [Min, Max]	0.11 [0,0.3]	0.14 [0,1]	0.14 [0,1]	-	-	-
Attention non executed tasks						
Mean (SD)	1.86 (3.56)	1.94 (4.08)	1.92 (3.98)	-	-	-
Median [Min, Max]	1 [0,17]	0 [0,30]	0 [0,30]	-	-	-
Attention execution gain						
Mean (SD)	11.83 (18.91)	12.99 (16.67)	12.76 (17.08)	-	-	-
Median [Min, Max]	5.5 [0,91.5]	7 [0,95]	6.5 [0,95]	-	-	-
Memory task proportion						
Mean (SD)	0.4 (0.23)	0.4 (0.24)	0.4 (0.24)	-	-	-
Median [Min, Max]	0.44 [0,1]	0.48 [0,1]	0.47 [0,1]	-	-	-
Memory non executed tasks						
Mean (SD)	2.63 (6.68)	4.68 (9.38)	4.28 (8.95)	-	-	-
Median [Min, Max]	0 [0,39]	1 [0,62]	1 [0,62]	-	-	-
Memory execution gain						
Mean (SD)	38.72 (59.28)	44.02 (59.68)	43.01 (59.51)	-	-	-
Median [Min, Max]	23.5 [0,336.5]	23 [0,353]	23.5 [0,353]	-	-	-
Ex. Functions task proportion						
Mean (SD)	0.22 (0.15)	0.2 (0.14)	0.2 (0.14)	-	-	-
Median [Min, Max]	0.25 [0,0.57]	0.21 [0,0.75]	0.22 [0,0.75]	-	-	-
Ex. Functions non executed tasks						
Mean (SD)	5.26 (6.67)	6.37 (9.57)	6.16 (9.09)	-	-	-
Median [Min, Max]	3 [0,26]	3 [0,65]	3 [0,65]	-	-	-
Ex. Functions execution gain						
Mean (SD)	14.91 (21.75)	18.78 (30.08)	18.04 (28.67)	-	-	-
Median [Min, Max]	9 [0,115.5]	8.25 [0,173.5]	8.5 [0,173.5]	-	-	-
Language task proportion						
Mean (SD)	0.22 (0.36)	0.19 (0.34)	0.19 (0.34)	-	-	-
Median [Min, Max]	0 [0,1]	0 [0,1]	0 [0,1]	-	-	-
Language non executed tasks						
Mean (SD)	0.4 (1.24)	0.32 (1.11)	0.33 (1.13)	-	-	-
Median [Min, Max]	0 [0,6]	0 [0,8]	0 [0,8]	-	-	-
Language execution gain						
Mean (SD)	35.52 (118.77)	17.93 (37.89)	21.29 (62.07)	-	-	-
Median [Min, Max]	0 [0,756]	0 [0,166]	0 [0,756]	-	-	-
Orientation task proportion						
Mean (SD)	0 (0.01)	0 (0.01)	0 (0.01)	-	-	-
Median [Min, Max]	0 [0,0.03]	0 [0,0.06]	0 [0,0.06]	-	-	-
Orientation non executed tasks						
Mean (SD)	0 (0)	0.01 (0.1)	0.01 (0.09)	-	-	-
Median [Min, Max]	0 [0,0]	0 [0,1]	0 [0,1]	-	-	-
Orientation execution gain						
Mean (SD)	0.22 (0.58)	0.44 (0.93)	0.4 (0.88)	-	-	-
Median [Min, Max]	0 [0,2.5]	0 [0,6]	0 [0,6]	-	-	-
Calculus task proportion						
Mean (SD)	0.03 (0.06)	0.05 (0.07)	0.05 (0.07)	-	-	-
Median [Min, Max]	0 [0,0.25]	0 [0,0.4]	0 [0,0.4]	-	-	-
Calculus non executed tasks						
Mean (SD)	0.3 (1.3)	0.63 (1.78)	0.57 (1.71)	-	-	-
Median [Min, Max]	0 [0,7]	0 [0,13]	0 [0,13]	-	-	-
Calculus execution gain						

Mean (SD)	2.6 (5.1)	5.86 (11.61)	5.24 (10.75)	-	-	-
Median [Min, Max]	0 [0,22.5]	0 [0,74]	0 [0,74]	-	-	-
Gnosias task proportion						
Mean (SD)	0.01 (0.04)	0.01 (0.02)	0.01 (0.03)	-	-	-
Median [Min, Max]	0 [0,0.23]	0 [0,0.14]	0 [0,0.23]	-	-	-
Gnosias non executed tasks						
Mean (SD)	0.02 (0.15)	0.02 (0.18)	0.02 (0.18)	-	-	-
Median [Min, Max]	0 [0,1]	0 [0,2]	0 [0,2]	-	-	-
Gnosias execution gain						
Mean (SD)	1.06 (3.08)	1.46 (3.52)	1.38 (3.43)	-	-	-
Median [Min, Max]	0 [0,17]	0 [0,33]	0 [0,33]	-	-	-

Validation Results

The performance metrics corresponding to Recall, F1, Precision, and the Area Under the Receiver Operator Characteristic Curve (AUC-ROC) are presented in Table 2. The performance is reported for three instances of the cognitive improvement and therapy compliance models: 1) the base model performance as reported in D4.8, 2) the performance of base model tested over the validation dataset, and 3) the performance of a re-trained model using the original dataset plus the validation dataset. For the re-trained model, the same modelling approach described in D4.8 (Section 3.2.1 Modelling Approach) was applied. As commented before, in order to have a good amount of registries for validation, missing values were allowed to be later treated using simple imputation techniques. The *SimpleImputer* function from the *sklearn Python 3* library was used to treat missing values in the validation cohort.

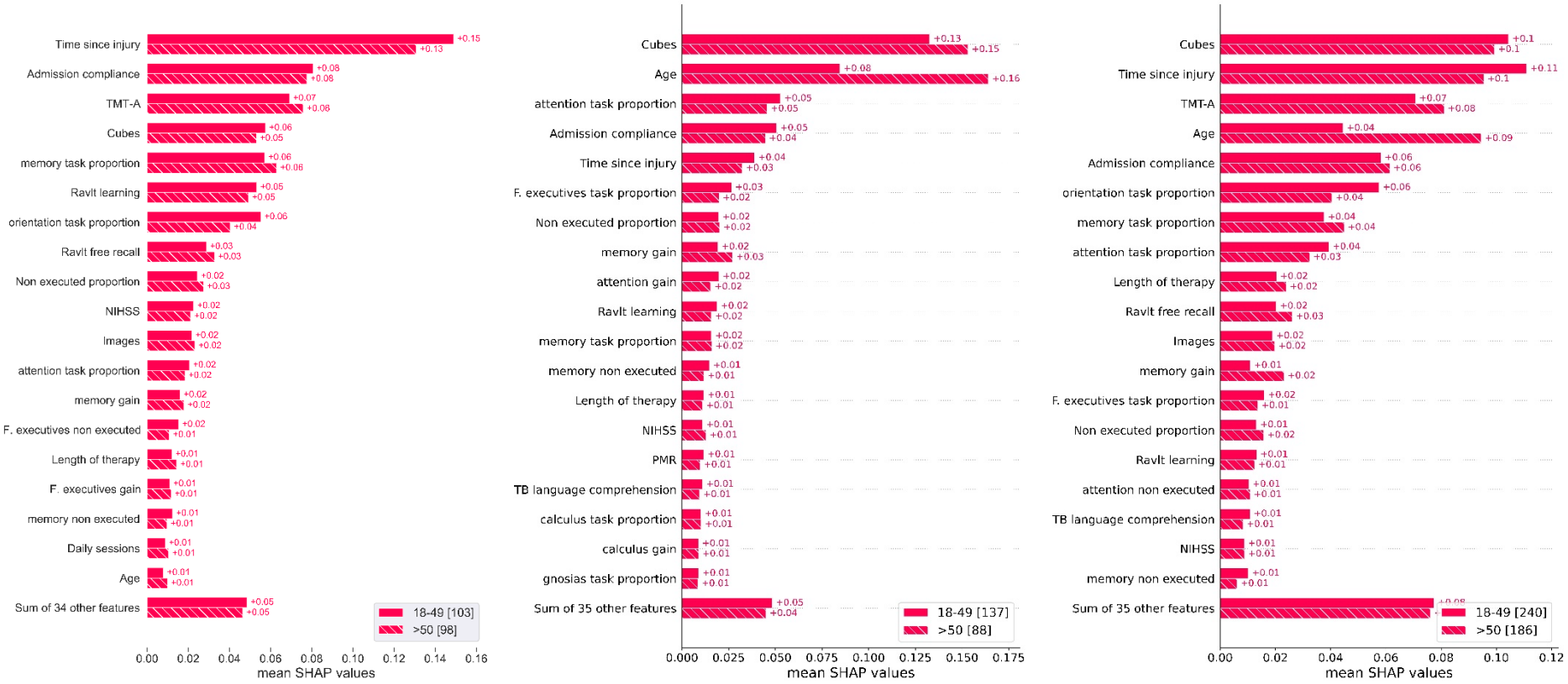
Table 2. Performance evaluation for the base model, validation test, and retrained models. Re-sampling at k-fold (k=5) cross-validation with 5 repetitions. CI: Cognitive Improvement, TC: Therapy Compliance.

Model	F1	Recall	Precision	AUC	Params
CI base	0.617 (0.09)	0.713 (0.07)	0.624 (0.16)	0.517 (0.04)	colsample_bytree: 0.6, eta: 0.01, gamma: 1, min_child_weight: 5, reg_lambda: 0.5, subsample: 0.8
CI test	0.47	0.51	0.91	0.51	
CI re-trained	0.688 (0.05)	0.761 (0.03)	0.724 (0.11)	0.526 (0.02)	colsample_bytree: 0.6, eta: 0.01, gamma: 0.05, min_child_weight: 5, reg_lambda: 0.5, subsample: 0.8
TC base	0.694 (0.04)	0.673 (0.04)	0.682 (0.04)	0.621 (0.04)	colsample_bytree: 0.8, eta: 0.01, gamma: 0.05, min_child_weight: 5, reg_lambda: 0.5, subsample: 0.8
TC test	0.35	0.51	0.55	0.51	
TC re-trained	0.682 (0.04)	0.694 (0.04)	0.689 (0.04)	0.678 (0.04)	colsample_bytree: 0.6, eta: 0.01, gamma: 0.05, min_child_weight: 5, reg_lambda: 1, subsample: 1

For the Cognitive Improvement model (CI), results showed a drop in performance when the base model (CI base) and the validation test (CI test) are compared, more specifically Recall (0.713 to 0.51) and F1 scores (0.617 to 0.47). On the other hand, a comparison of the base model and the re-trained version of the model reported an improvement in performance for Recall (0.71 to 0.76), F1 (0.61 to 0.68) and Precision (0.62 to 0.72). A similar behaviour was observed for the Therapy Compliance model (TC). For this model, performance dropped when the model was tested over

unseen data from the validation cohort. This was observed for Recall (0.67 to 0.51), F1 (0.69 to 0.35), Precision (0.68 to 0.55) and AUC (0.62 to 0.51) metrics. Re-training the model showed an improvement in performance for Recall (0.67 to 0.69) and AUC (0.62 to 0.67).

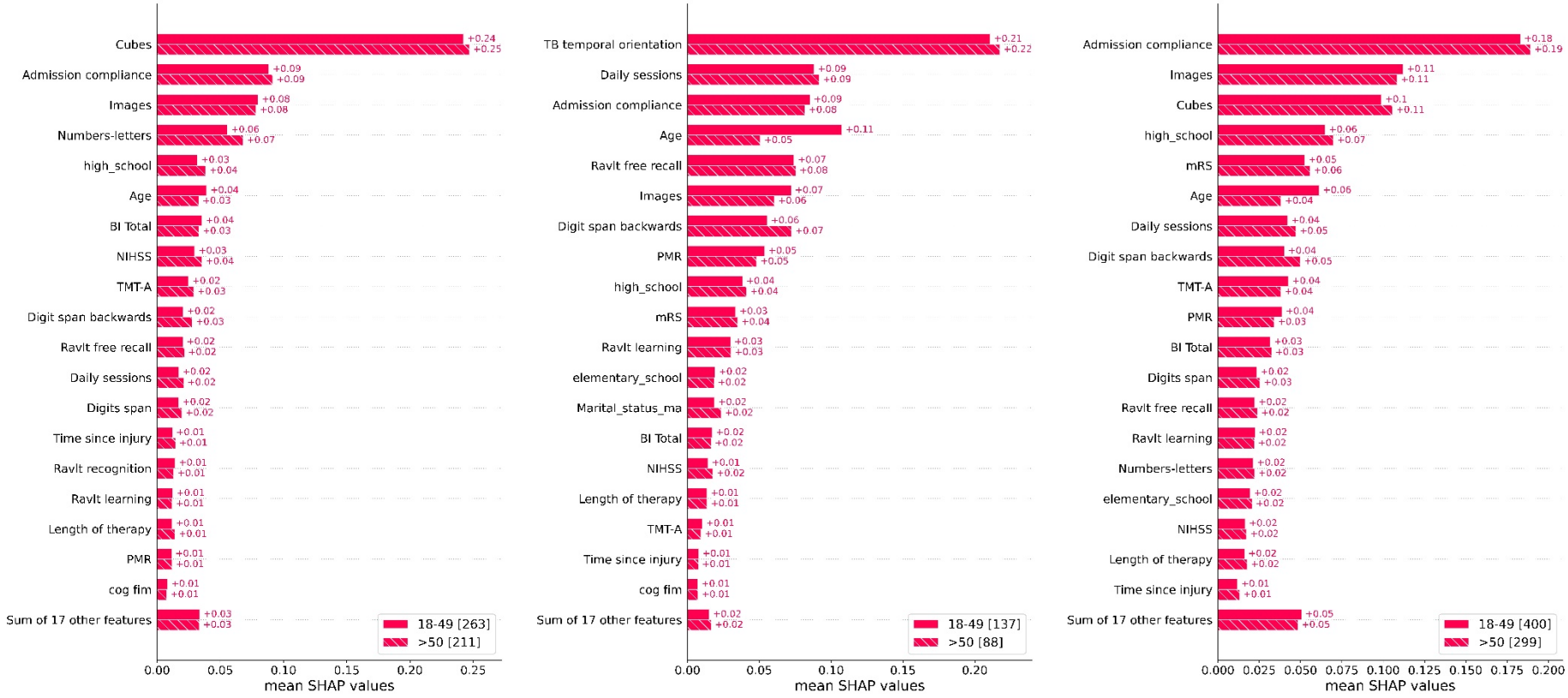
As reported in D4.8, a population prediction analysis was carried out to compute the absolute impact of each feature on the model's prediction. The *Python* implementation of the Shapley Additive exPlanations (SHAP) method was used to compute the feature importance values and generate bar plots to depict them. Figure 2 and Figure 3 show the feature importance plots for the base model, the validation test, and the re-trained model for the cognitive improvement and therapy compliance models respectively. These plots are organised according to age (below and above 50 years of age).



(a) Base **(b) Test** **(c) Re-trained**

Figure 2. SHAP bar plots of feature importance for the Cognitive Improvement Model. Age cohort: >50

.



(a) Base **(b) Test** **(c) Re-trained**

Figure 3. SHAP bar plots of feature importance for the Therapy Compliance Model. Age cohort: >50

For the cognitive improvement model, the standardised assessment *Cubes* (visual-construction) remained as the strongest and common influential factor (at different levels) across the base model, the validation test, and the retrained model with slightly different effects regarding age. *Admission compliance* and *time since injury* also showed a strong influence across three model reports with no significant differences regarding the age of patients. When comparing the base and the retrained models, the standardised assessment *TMT-A* (memory) was identified as a common influential factor, showing a similar effect for both models with no difference in terms of age. The comparison of the validation test and re-train plots showed that the variable *Age* was a common influential factor with a stronger effect for the validation test.

For the therapy compliance model, the *admission compliance* was identified as a common strong influential factor across the three feature importance reports. A stronger effect was observed for the re-trained model with no differences in terms of patient's age. The standardised assessment *Images* (visual-perception) showed also a relative important influence across all reports, but with slightly lower impact for the validation test. When comparing the base and re-train models, the standardised assessment *Cubes* (visual-construction) was identified as a common influential factor, with a stronger effect for the base model. At a lower influential effect, the demographic variable *high school* (educational level) was identified as a common factor for the base and the re-trained models. Finally, the comparison of the validation test and the re-trained models reported the variable *Age* as a common influential factor with a stronger effect for the validation test and patients below the 50 years of age. At a lower impact level, the *daily sessions* variable was identified as a common factor between the validation test and re-trained models.

The feature influential report showed that some variables remained as strong influential features for the three instances of the model (base, test validation, re-trained). More specifically, the variable *admission compliance* maintained a strong influence for the three model instances, but also across the two cognitive models (cognitive improvement and therapy compliance). As it was reported in D4.8, hand-crafted variables like *admission compliance*, which resulted from the close interaction of clinicians and data-science partners, were key for the model's development.

2.3 Conclusions

The trained models presented in D4.8 were tested against a set of unseen cohort of patients from the Guttmann Rehabilitation centre. In addition, both models were re-trained merging the original dataset and the unseen validation cohort following the same modelling approach as presented in D4.8. Performance results showed a drop in performance when the models were tested against the validation cohort. However, re-trained versions of both models reported similar, and for some cases, better results when compared to the models presented in D4.8. A comparison of the features' impact showed how certain variables (e.g., *admission compliance*) had a strong influence across base, validation and re-trained models.

3 Motor Rehabilitation

3.1 Introduction and Background

As presented in D4.8 motor rehabilitation models aimed to (1) use each of the 33 individual FMA-UE items to predict total FMA-UE score at discharge of patients with ischemic stroke admitted to rehabilitation ≤ 90 days since stroke onset, (2) select eight FMA-UE items (seven from the Scandinavian study plus the item with the highest predictive power from objective 1) and use each of them to predict mild impairment at discharge and (3) taking as starting point the 3 models with the highest predictive power identified in objective 2, adjust them with previously reported confounders to identify the most relevant predictors of mild impairment

Therefore, the derivation patient cohort, presented in D4.8 included n=287 patients from two data-sources (Institut Guttmann hospital and ICARE). In relation to the first one, from March 2018 to December 2020 a total of 134 patients with first-ever stroke were admitted to the rehabilitation unit of Institut Guttmann hospital and assessed at admission and discharge using the FMA-UE, 71 of them with ischemic stroke. After excluding 18 with more than 90 days since stroke onset to rehabilitation admission, 3 with more than 7 days since admission to assessment, 2 younger than 18 years old at the moment of admission, 1 with injury/condition prior to the stroke that limited the use of the affected arm and 1 with severe multi-impairment or diminished physical condition prior to stroke; 46 patients were included in the study.

The second data-source involves participants recruited during inpatient rehabilitation, from 7 sites in the United States metropolitan areas of Los Angeles, Atlanta and Washington D.C from June 2009 to March 2014. This dataset is registered at the ClinicalTrials.gov under the title Arm Rehabilitation Study After Stroke (ICARE) with NCT identifier 00871715. The ICARE investigators tested 3 different arm therapy interventions: Accelerated Skill Acquisition Program (ASAP), Behavioral: Dose-Equivalent Usual & Customary Care (DEUCC) and Behavioral: Usual and Customary Care (UCC).

The ICARE dataset was provided by the National Institute of Neurological Disorders and Stroke (NINDS). In supplementary material we present the protocol for accessing NINDS data and NINDS Data Request Form.

The ICARE dataset included 361 recruited between 5 and 106 days post-stroke [27], 288 of them with ischemic stroke, after excluding 39 with more than 90 days since stroke onset to rehabilitation admission and 2 younger than 18 years old at the moment of admission, leaves 241 patients to be included in the study.

Therefore, the total number of included patients for model derivation was 287.

3.2 Validation activities

Validation patient cohort

From January 2021 to July 2022 a total of 114 patients with first-ever ischemic stroke were admitted to the rehabilitation unit of Institut Guttmann hospital and assessed at admission and discharge using the FMA-UE. After excluding 5 with more than 90 days since stroke onset to rehabilitation admission; 109 patients were included in the validation study.

In this work we applied the same stratification of patients as in D4.8 using FMA-UE at discharge: 48-66 for mild impairment and ≤ 47 for severe and moderately-severe impairment. Table 3 presents patients' characteristics at admission, for each group.

Table 3. Clinical and demographic characteristics for all (n=109) validation patients stratified in two groups (Severe-moderately severe and Mild) according to their FMA-UA at rehabilitation discharge

	< 48 **Severe-moderately-severe** **(*n*=61)**	**≥ 48** **Mild** **(*n*=48)**	**TOTAL** **(*n*=109)**	***p***
Male, n(%)	40 (65.6)	30 (62.5)	70 (64.2)	0.740
Age at admission, median (Q1-Q3)	57 (50- 65)	53 (45- 59)	56 (48- 62)	0.017
Younger than 65 years, n(%)	45 (73.8)	41 (85.4)	86 (78.9)	0.139
NIHSS at admission, median(Q1-Q3)	10 (8- 13)	9 (6- 10)	9 (8- 13)	0.048
NIHSS categorization, n(%)				0.372
Mild	9 (14.8)	12 (25.0)	21 (19.3)	
Moderately severe	40 (65.6)	29 (60.4)	69 (63.3)	
Severe	12 (19.7)	7 (14.6)	19 (17.4)	
Affected side (Left), n(%)	42 (68.9%)	18 (37.5%)	60 (55.0%)	0.001
Dominance (Right), n(%)	61 (100.0%)	45 (93.8%)	106 (97.2%)	0.048
Aphasia, n(%)	14 (23.0%)	10 (20.8%)	24 (22.0%)	0.791
Smoking habits (at admission), n(%)	14 (23.0%)	13 (27.1%)	27 (24.8%)	0.620
COGNITVE-FIM at admission	27 (20- 31)	26 (19- 32)	27 (19- 32)	0.809
MOTOR-FIM at admission	36 (30- 52)	42 (33- 61)	41 (31- 55)	0.191
TOTAL-FIM at admission	67 (49- 84)	70 (53- 86)	68 (49- 85)	0.316
Days since injury to admission, median (Q1-Q3)	77 (49- 104)	44 (33- 86)	66 (41- 96)	0.002
FMA-UE at admission, median (Q1-Q3)	21 (6- 34)	46 (42- 51)	37 (12- 46)	< 0.001
FMA-UE at discharge, median (Q1-Q3)	27 (8- 43)	58 (53- 62)	45 (23- 57)	< 0.001
LOS, days, median (Q1-Q3)	96 (54- 130)	120 (58- 132)	114 (54- 131)	0.199
Living arrangement, n(%)				0.920
Alone	6 (9.8%)	5 (10.4%)	11 (10.1%)	
With a spouse, other relative or friend	55 (90.2%)	43 (89.6%)	98 (89.9%)	

FMA-UE: Fugl-Meyer Assessment – Upper Extremity; LOS: length of stay in rehabilitation; NIHSS: National Institutes of Health Stroke Scale; FIM: Functional Independence Measure;

Validation Results

To address the first validation objective we used as candidate predictors all 33 items from FMA-UE protocol to predict total FMA-UE score at rehabilitation discharge. Each item is presented in Table 4 as used in D4.8. Items superscripted with an * were those most commonly used in previous upper limb research.

Table 4. The 33-individual items that constitute the FMA-UE

A. UPPER EXTREMITY	***I. Reflex activity***	
	Flexors: biceps and finger flexors (at least one)	FM_BICEP
	Extensors: triceps	FM_TRICEP
	Subtotal I (max 4)	
	II. Volitional movement within synergies, without gravitational help	
	Flexor synergy – Shoulder - retraction	FM_FS_RET
	Flexor synergy – Shoulder - elevation	FM_FS_ELV
	Flexor synergy – Shoulder - abduction (90°)	FM_FS_ABD*
	Flexor synergy – Shoulder – external rotation	FM_FS_EXT
	Flexor synergy – Elbow – flexion	FM_FS_ELF
	Flexor synergy – Forearm – supination	FM_FS_SUP
	Extensor synergy – Shoulder - adduction/internal rotation	FM_ES_SHAD
	Extensor synergy – Elbow - extension	FM_ES_EXT*
	Extensor synergy – Forearm - pronation	FM_ES_FPR
	Subtotal II (max 18)	
	III. Volitional movement mixing synergies, without compensation	
	Hand to lumbar spine	FM_MS_HAND
	Shoulder flexion 0°- 90°	FM_MS_SHF
	Pronation-supination	FM_MS_PSUP*
	Subtotal III (max 6)	
	IV. Volitional movement with little or no synergy	
	Shoulder abduction 0 - 90°	FM_MOS_SAB
	Shoulder flexion 90° - 180°	FM_MOS_SFL
	Pronation/supination	FM_MOS_PRO
	Subtotal IV (max 6)	
	V. Normal reflex activity	
	Biceps, triceps,finger flexors	FM_NR
	Subtotal V (max 2)	
	Total A (max 36)	

B. WRIST	Stability at 15° dorsiflexion (Elbow at 90°, Shoulder at 0°)	FM_W_SE9*
	Repeated dorsifexion / volar flexion (Elbow at 90°, Shoulder at 0°)	FM_W_FE9
	Stability at 15° dorsiflexion (Elbow at 0°, Shoulder at 30°)	FM_W_SE3
	Repeated dorsifexion / volar flexion	FM_W_FE3
	Circumduction	FM_W_CIR
	Total B (max 10)	
C. HAND	Finger Mass flexion	FM_H_FMF
	Finger Mass extension	FM_H_FME*
	Hook grasp	FM_H_GRASP1
	Thumb adduction	FM_H_GRASP2
	Pincer grasp, opposition	FM_H_GRASP3*
	Cylinder grasp	FM_H_GRASP4*
	Spherical grasp	FM_H_GRASP5
	Total C (max 14)	
D. COORDINATION/SPEED	Tremor	FM_CS_TRE
	Dysmetria	FM_CS_DYS
	Time	FM_CS_SPE
	Total D (max 6)	

Table 5 and 6 present the results obtained using the validation cohort, for each item from subscale A (Table 5) and for subscales B, C and D (Table 6), the obtained coefficients (95%CI), level of significance and adjusted R^2

Figure 4 (top) ranks all 33 FMA-UE items obtained using the derivation cohort visually showing for each of them the obtained adjusted R^2, highest values are shown to the right of the Figure. Therefore, the overall top predictor was finger mass extension (FM_H_FME*) followed by finger mass flexion (FM_H_FMF).

Figure 4 (bottom) ranks all 33 FMA-UE items obtained using the validation cohort visually showing for each of them the obtained adjusted R^2, using the same order as presented for the derivation cohort.

We fitted a linear regression model to all items in Figure 4 bottom, in order to confirm that the left to right increasing order obtained using the derivation cohort was kept when using the validation cohort.

Table 5. Predictive models for FMA-UE score at discharge for each candidate predictor item (FMA – subscale A). Linear regression models

Item	Category	*n*	Coefficient (95% CI)	*p*	Adjusted R^2
FM_BICEP	0:none	6 (5.5%)			0.4
	2:elicited	103 (94.5%)	0.5 (-3.5; 4.6)	0.801	
FM_TRICEP	0:none	16 (14.7%)			0.3
	2:elicited	93 (85.3%)	4.1 (-6.7;15.8)	0.453	
FM_FS_RET	0: none	28 (25.7%)			33.7
	1:partial	47 (43.1%)	20.5(12.6;28.3)	<0.001	
	2: full	34 (31.2%)	31.6(23.2;40.0)	<0.001	
FM_FS_ELV	0: none	27 (24.8%)			51.1
	1:partial	40 (36.7%)	28.8(21.8;35.8)	<0.001	
	2: full	42 (38.5%)	36.7(29.8;43.7)	<0.001	
FM_FS_ABD*	0: none	25 (22.9%)			51.7
	1:partial	43 (39.4%)	28.6(21.5;35.6)	<0.001	
	2: full	41 (37.6%)	38.4(31.3;45.5)	<0.001	
FM_FS_EXT	0: none	38 (34.9%)			56.6
	1:partial	39 (35.8%)	29.3(23.2;35.3)	<0.001	
	2: full	32 (29.4%)	34.9(28.5;41.3)	<0.001	
FM_FS_ELF	0: none	22 (20.2%)			49.4
	1:partial	23 (21.1%)	23.7(15.1;32.2)	<0.001	
	2: full	64 (58.7%)	36.8(29.8;43.9)	<0.001	
FM_FS_SUP	0: none	35 (32.1%)			53.6
	1:partial	53 (48.6%)	27.9(21.9;33.9)	<0.001	
	2: full	21 (19.3%)	38.3(30.7;45.9)	<0.001	
FM_ES_SHAD	0: none	28 (25.7%)			64.8
	1:partial	34 (31.2%)	28.4(22.3;34.5)	<0.001	
	2: full	47 (43.1%)	40.6(34.9;46.3)	<0.001	
FM_ES_EXT*	0: none	30 (27.5%)			83.5
	1:partial	40 (36.7%)	32.3(28.3;36.2)	<0.001	
	2: full	39 (35.8%)	46.4(42.4;50.3)	<0.001	
FM_ES_FPR	0: none	34 (31.2%)			66.8
	1:partial	34 (31.2%)	28.2(22.6;33.8)	<0.001	
	2: full	41 (37.6%)	39.5(34.1;44.9)	<0.001	
FM_MS_HAND	0: none	34 (31.2%)			67.5

	1:partial	51 (46.8%)	32.4(27.3;37.5)	<0.001	
	2: full	24 (22.0%)	41.3(35.1;47.4)	<0.001	
FM_MS_SHF	0: none	49 (45.0%)			57.1
	1:partial	31 (28.4%)	27.7(21.6;33.7)	<0.001	
	2: full	29 (26.6%)	33.7(27.5;39.9)	<0.001	
FM_MS_PSUP*	0: none	35 (32.1%)			59.5
	1:partial	49 (45.0%)	29.9(24.2;35.6)	<0.001	
	2: full	25 (22.9%)	38.6(31.9;45.3)	<0.001	
FM_MOS_SAB	0: none	44 (40.4%)			46.1
	1:partial	38 (34.9%)	23.0(16.4;29.5)	<0.001	
	2: full	27 (24.8%)	33.1(25.8;40.3)	<0.001	
FM_MOS_SFL	0: none	65 (59.6%)			45.2
	1:partial	35 (32.1%)	25.6(19.4;31.9)	<0.001	
	2: full	9 (8.3%)	35.1(24.5;45.7)	<0.001	
FM_MOS_PRO	0: none	53 (48.6%)			49.5
	1:partial	41 (37.6%)	26.0(20.1;32.0)	<0.001	
	2: full	15 (13.8%)	34.3(25.9;42.7)	<0.001	
FM_NR	0: none	105 (96.3%)			1.9
	1:partial	1 (0.9%)	5.1(-35.0;45.1)	0.804	
	2: full	3 (2.8%)	24.1(0.6;47.4)	0.044	

Table 6. Predictive models for FMA-UE score at discharge for each candidate predictor item (FMA – subscales B, C, D). Linear regression models.

Item	Category	*n*	Coefficient (95%CI)	*p*	Adjusted R^2
FM_W_SE9*	0: none	41 (37.6%)			59.1
	1:partial	48 (44.0%)	30.6(25.1;36.1)	<0.001	
	2: full	20 (18.3%)	35.5(28.5;42.6)	<0.001	
FM_W_FE9	0: none	36 (33.0%)			55.5
	1:partial	52 (47.7%)	30.2(24.4;36.0)	<0.001	
	2: full	21 (19.3%)	36.2(28.8;43.6)	<0.001	
FM_W_SE3	0: none	47 (43.1%)			52.3
	1:partial	48 (44.0%)	27.7(22.0;33.5)	<0.001	
	2: full	14 (12.8%)	35.1(26.6;43.6)	<0.001	
FM_W_FE3	0: none	46 (42.2%)			43.6

	1:partial	50 (45.9%)	25.3(19.1;31.4)	<0.001	
	2: full	13 (11.9%)	33.2(23.7;42.8)	<0.001	
FM_W_CIR	0: none	36 (33.0%)			59.5
	1:partial	69 (63.3%)	33.0(27.7;38.2)	<0.001	
	2: full	4 (3.7%)	39.1(25.6;52.6)	<0.001	
FM_H_FMF	0: none	28 (25.7%)			58.5
	1:partial	36 (33.0%)	28.4(21.9;34.9)	<0.001	
	2: full	45 (41.3%)	38.8(32.5;45.0)	<0.001	
FM_H_FME*	0: none	28 (25.7%)			63.9
	1:partial	37 (33.9%)	33.0(26.9;39.0)	<0.001	
	2: full	44 (40.4%)	39.7(33.8;45.5)	<0.001	
FM_H_GRASP1	0: none	45 (41.3%)			41.1
	1:partial	33 (30.3%)	22.8(15.7;29.9)	<0.001	
	2: full	31 (28.4%)	29.7(22.5;36.9)	<0.001	
FM_H_GRASP2	0: none	34 (31.2%)			65.5
	1:partial	54 (49.5%)	32.4(27.2;37.6)	<0.001	
	2: full	21 (19.3%)	41.0(34.5;47.6)	<0.001	
FM_H_GRASP3*	0: none	46 (42.2%)			44.8
	1:partial	42 (38.5%)	27.7(21.4;34.1)	<0.001	
	2: full	21 (19.3%)	27.6(19.7;35.5)	<0.001	
FM_H_GRASP4*	0: none	41 (37.6%)			55.2
	1:partial	25 (22.9%)	27.4(20.6;34.3)	<0.001	
	2: full	43 (39.4%)	33.0(27.1;38.9)	<0.001	
FM_H_GRASP5	0: none	38 (34.9%)			58.5
	1:partial	33 (30.3%)	28.6(22.4;34.7)	<0.001	
	2: full	38 (34.9%)	35.3(29.3;41.3)	<0.001	
FM_CS_TRE	0: none	36 (33.0%)			56.6
	1:partial	25 (22.9%)	34.9(28.0;41.8)	<0.001	
	2: full	48 (44.0%)	31.2(25.3;37.0)	<0.001	
FM_CS_DYS	0: none	39 (35.8%)			52.0
	1:partial	28 (25.7%)	30.0(23.1;37.0)	<0.001	
	2: full	42 (38.5%)	31.1(24.9;37.3)	<0.001	
FM_CS_SPE	0: none	61 (56.0%)			32.3
	1:partial	31 (28.4%)	21.5(14.2;28.8)	<0.001	
	2: full	17 (15.6%)	26.8(17.7;35.9)	<0.001	

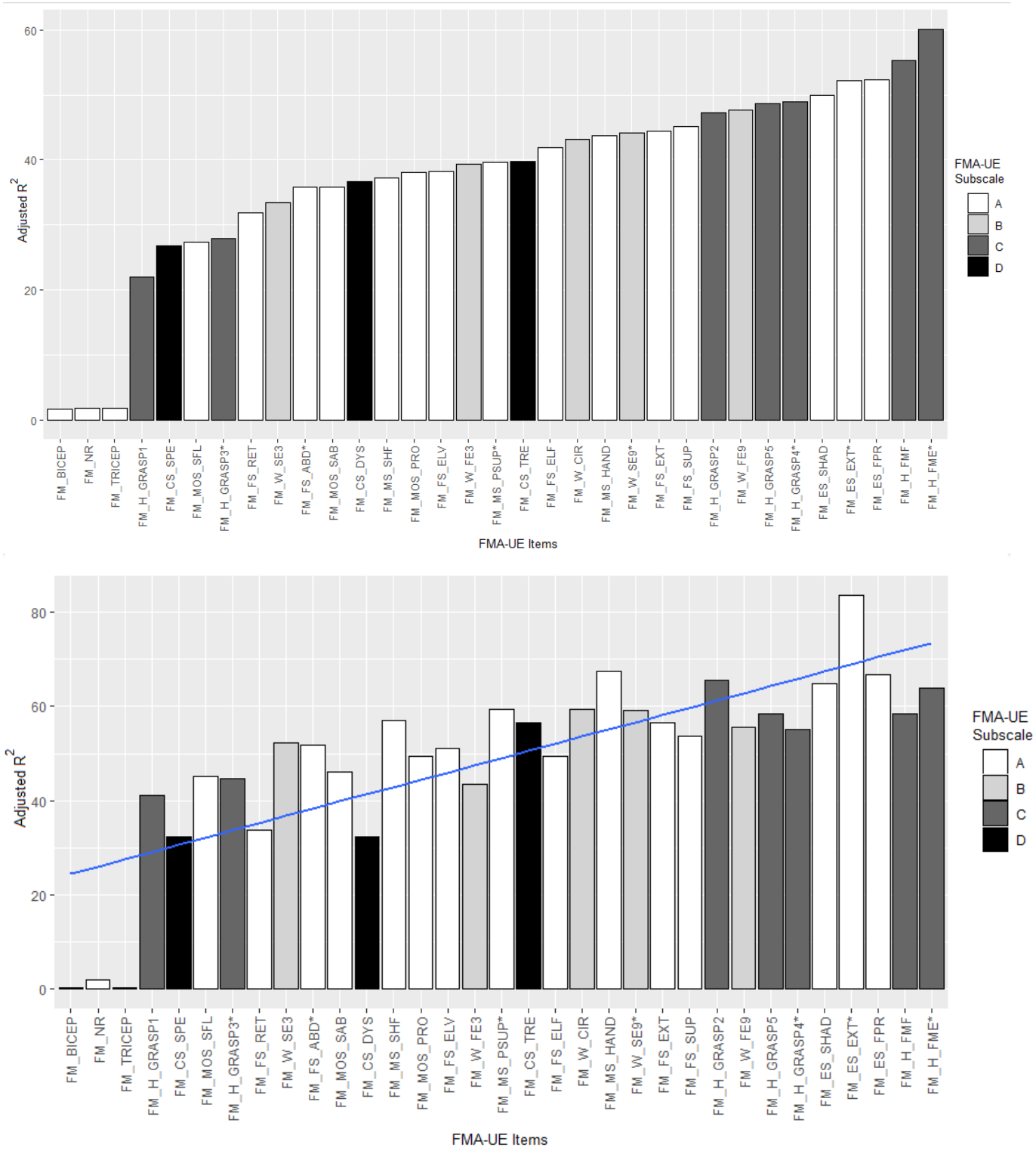


Figure 4 Obtained adjusted R^2 for each of the FMA-UE items for derivation cohort (top) and validation cohort (bottom)

Figure 4 (top) presents the obtained adjusted R^2 for each of the FMA-UE items using the derivation cohort presented in D4.8 ordered from left to right with the highest adjusted R^2 to the right and the adjusted R^2 obtained using validation cohort (bottom) with the same order of the FAM-UE items.

We then addressed the validation of the adjusted models. We included in Table 7 the results reported in D4.8. Finger mass flexion (FM_H_FMF) item yielded the highest AUC= 0.88 (0.82-0.94) with sensitivity and specificity = 0.83. The only other significant independent variable was the time since stroke onset to rehabilitation admission with an OR=0.9 indicating that each additional increase of one day in time to admission is associated with a 10% decrease in the odds of achieving mild motor impairment at discharge. We showed good predictive power in both cases with AUC: 0.70-0.82 for the unadjusted models and AUC: 0.85-0.88 for the adjusted models. We used two items frequently applied in previous related research for upper limb predictive models (FM_H_FME and FM_ES_EXT). Besides, we identified an additional FMA-UE item, finger mass flexion (FM_H_FMF) which yielded the highest unadjusted AUC = 0.82 as well as the highest adjusted AUC=0.88. Finger mass extension (FM_H_FME*) previously reported in the Scandinavian study also yielded an adjusted AUC=0.88, but to our best knowledge finger mass flexion was never proposed before as individual predictor of mild motor impairment.

Table 8 presents the results for the validation cohort, with no confounders reported as significant predictors, though results present a lower AUC when compared to the derivation cohort they range from 0.71 to 0.74, still showing good predictive power.

Table 7. Adjusted models of the dichotomized FMA-UE at discharge for each of the top 3 candidate predictor items (derivation cohort)

	OR (95%CI)	*P*	AUC (95%CI)	Sensitivity	Specificity
FM_ES_EXT_01	28.7(6.6;207.1)	<0.001	0.85(0.78-0.93)	0.71	0.89
FM_ES_EXT_02	107.9(24.2;801.4)	<0.001			
Sex.Female	1.21(0.6;2.2)	0.522			
Age	0.9(0.9;1.1)	0.339			
TSI	0.9 (0.9;1.0)	<0.001			
LOS	0.9(0.9;1.0)	0.319			
Aphasia	1.0(0.4;2.8)	0.900			
Smoke.	2.2(0.8;6.2)	0.123			
Living with a spouse, other relative or friend	2.0(1.0;4.1)	0.042			
FM_H_FME_01	37.5(6.1;741.3)	<0.001	0.88(0.81-0.94)	0.71	0.90
FM_H_FME_02	231.3(36.3;4662.5)	<0.001			
Sex.Female	1.2(0.6;2.2)	0.483			
Age	0.9(0.9;1.1)	0.038			
TSI	0.9(0.9;1.0)	<0.001			
LOS	0.9(0.9;1.0)	0.485			
Aphasia	0.8(0.2;2.2)	0.692			

Smoke.	1.2(0.4;3.7)	0.647			
Living with a spouse, other relative or friend	1.8(0.9;3.6)	0.092			
FM_H_FMF_01	21.5(3.5;424.8)	0.005	088(0.82-0.94)	0.83	0.83
FM_H_FMF_02	143.2(22.6;2881)	<0.001			
Sex.Female	1.3(0.7;2.5)	0.315			
Age	0.9(0.9;1.0)	0.460			
TSI	0.9(0.9;1.0)	<0.001			
LOS	0.9(0.9;1.0)	0.480			
Aphasia	0.8(0.3;2.3)	0.795			
Smoke.	1.1(0.4;3.1)	0.786			
Living with a spouse, other relative or friend	1.8(0.9;3.6)	0.099			

TSI: time since stroke onset to rehabilitation admission; LOS: length of stay

Table 8. Adjusted models of the dichotomized FMA-UE at discharge for each of the top 3 candidate predictor items (validation cohort)

	OR (95%CI)	*p*	AUC (95%CI)	Sensitivity	Specificity
FM_ES_EXT_01	18.6(2.6;104.7)	<0.001	0.71(0.62-0.75)	0.63	0.71
FM_ES_EXT_02	97.9(44.2;304.9)	<0.001			
Sex.Female	1.31(0.6;2.7)	0.522			
Age	0.9(0.9;1.1)	0.339			
TSI	0.9 (0.9;1.0)	0.142			
LOS	0.9(0.9;1.0)	0.319			
Aphasia	1.0(0.4;2.8)	0.900			
Smoke.	2.7(0.8;6.2)	0.123			
Living alone	2.3(1.0;4.1)	0.142			
FM_H_FME_01	68.5(6.1;521.3)	<0.001	0.74(0.61-0.78)	0.61	0.70
FM_H_FME_02	131.3(36.3;542.5)	<0.001			
Sex.Female	1.2(0.6;2.2)	0.483			
Age	0.9(0.9;1.1)	0.115			
TSI	0.9(0.9;1.0)	0.234			
LOS	0.9(0.9;1.0)	0.485			
Aphasia	0.8(0.2;2.2)	0.692			

Smoke.	1.2(0.4;3.7)	0.647			
Living alone	1.6(0.9;4.6)	0.112			
FM_H_FMF_01	25.5(3.5;324.8)	0.005			
FM_H_FMF_02	113.2(12.6;981)	<0.001			
Sex.Female	1.3(0.7;2.5)	0.315			
Age	0.9(0.9;1.0)	0.460			
TSI	0.9(0.9;1.0)	0.321	074(0.62-0.77)	0.63	0.69
LOS	0.9(0.9;1.0)	0.480			
Aphasia	0.8(0.3;2.3)	0.795			
Smoke.	1.1(0.4;3.1)	0.786			
Living alone	1.6(0.9;3.6)	0.127			

TSI: time since stroke onset to rehabilitation admission; LOS: length of stay

3.3 Conclusions

The validation cohort clearly confirmed the results obtained with the derivation cohort, both when considering all 33 individual FMA-UE items using unadjusted models and when considering the top 3 items with adjusted models. Besides, as presented in Annex I the total number of included participants (287 in the derivation cohort + 109 in the validation cohort) is clearly larger than most of FMA-UE predictive models presented in previous research.

4 Community integration: Social risk

4.1 Introduction and Background

The Reintegration deliverable D4.9 presented that the quality of life of post-ischemic stroke patients is affected by a range of factors, including the risk of insufficient social and family support, as well as socio-economic status (i.e., access to home health care, day center or private carer). Therefore, social risk predictive models were generated in D4.9 to meet the key goals of post-stroke reintegration to inform reintegration decisions and design personalized interventions for patients with social risk. This section presents the validation of the five GBM social risk models developed in D4.9 section 3.

Social risk modeling in D4.9 was based on Institut Guttmann's social risk assessment, "Escala de Valoracion Socio Familiar" (EVSF), which considers five dimensions of social risk: cohabitation, economic status (indicating income sufficiency), home status (indicating home accessibility in case of mobility problems), family support and social support, where patients were categorized in risk groups based on assessment scores (D4.9 section 3). Patients in the no social risk and mild social risk categories were considered as having negligible social risk (GREEN), whereas patients in the important and severe social risk categories were considered as having significant social risk (RED) (D4.9 Table 7).

In summary, demographic, diagnostic and EVSF assessment data (16 predictors) of 217 patients were used for training models (D4.9 Table 8). In the training cohort there were twice as many male patients as female patients as there was no way to control for this sex ratio in the admitted patients or any gender bias in the referral from acute treatment units. There was also a similar imbalance for the social risk classification; nearly twice as many patients with negligible social risk (GREEN) than significant social risk (RED) at discharge from the hospital. To account for this imbalance, five Generalized Boosted Regression (GBM) models were trained: original model (not correcting for class imbalance), weighted method (giving equal weight to both classes), up-sampling model, down-sampling model and smote-method model (D4.9 section 3.1.3).

The GBM social risk models performance metrics (D4.9 Table 10) indicated that all models performed similarly (as based on AUC) despite prediction target class imbalance, however, there were marked differences in other metrics, especially specificity. All models predominantly misclassified negligible social risk patients (misclassify GREEN patients as RED - False Positive) rather than significant social risk patients (misclassify RED patients as GREEN - False Negative). Variable importance as well as predictor contribution to GREEN and RED class prediction using approximate Shapley values were also calculated, and mostly indicated FamilySupport and Economic status, rather than demographic variables such as Sex, Educational Level or Civil Status contributing to social risk prediction.

4.2 Validation activities

Validation patient cohort

Demographic, diagnostic and assessment data utilizing the EVSF questionnaire during the rehabilitation and reintegration of patients were recorded and collected at the Institut Guttmann (Barcelona, Spain) from 2020 through 2021 during the prospective study. Inclusion criteria for this cohort consisted of adult patients 18 to 85 years of age at the time of stroke with an ischemic stroke diagnosis who were admitted within 3 weeks of the onset of symptoms, without any previous

comorbidities leading to disability, and whose data was recorded within a week of admission and discharge. Exclusion criteria were any of the following: diagnosis of stroke in the context of another concomitant comorbidity (e.g., traumatic brain injury), a previous history of another disabling condition, patients with EVSF assessment performed more than 5 months post injury, as well as more than 5 months stay at the rehabilitation hospital.

For this validation study we included 25 new patients meeting inclusion criteria and an additional dataset of 92 patients, that were filtered out for model training due to exclusion criteria (n=117). The benefit of including patients with exclusion criteria for validation is that it validates the utility and robustness of the models in a real-world clinical use case where patients at social risk may not meet inclusion criteria (specifically patients older than 85 years old, patients assessed more than 5 months post stroke and patients with a longer length of stay at the rehabilitation hospital). Similarly to the model training cohort, there is an imbalance in the validation dataset of the negligible (GREEN) and significant (RED) social risk patients, with twice as many GREEN than RED class patients. Table 9 below shows the validation cohort information of the total 117 patients (25 prospective study with inclusion criteria plus 92 patients not used for training due to exclusion criteria)

-

Table 9. Validation cohort information including social risk and demographics. Statistics of patients with negligible social risk (GREEN) and significant social risk (RED) including counts and percentages, the Mean (average value), Median (middle value with minimum and maximum value ranges) and Standard deviation (SD).

	GREEN (N=84)	RED (N=33)	Overall (N=117)
Sex			
Female	21 (25.0%)	10 (30.3%)	31 (26.5%)
Male	63 (75.0%)	23 (69.7%)	86 (73.5%)
AgeatStroke (years)			
Mean (SD)	51.9 (11.3)	51.3 (6.21)	51.8 (10.1)
Median [Min, Max]	50.8 [14.1, 85.8]	52.3 [39.4, 67.1]	51.0 [14.1, 85.8]
DaysSinceStroke			
Mean (SD)	78.2 (96.7)	105 (101)	85.7 (98.2)
Median [Min, Max]	46.5 [14.0, 605]	68.0 [26.0, 419]	58.0 [14.0, 605]
StrokeType			
Embolic	19 (22.6%)	12 (36.4%)	31 (26.5%)
Others	30 (35.7%)	9 (27.3%)	39 (33.3%)
Thrombolic	35 (41.7%)	12 (36.4%)	47 (40.2%)
LengthofStay (days)			
Mean (SD)	139 (83.0)	124 (73.1)	135 (80.4)
Median [Min, Max]	153 [11.0, 468]	117 [27.0, 341]	151 [11.0, 468]
EducationLevel			
High	50 (59.5%)	18 (54.5%)	68 (58.1%)
Low	34 (40.5%)	15 (45.5%)	49 (41.9%)
CivilStatus			
Married	52 (61.9%)	19 (57.6%)	71 (60.7%)
notMarried	32 (38.1%)	14 (42.4%)	46 (39.3%)
NIHSS			
Mean (SD)	12.8 (6.17)	15.3 (5.28)	13.5 (6.02)
Median [Min, Max]	13.0 [0, 26.0]	17.0 [5.00, 27.0]	14.0 [0, 27.0]
cognitive_FIM			
Mean (SD)	23.3 (9.32)	20.0 (8.36)	22.4 (9.15)
Median [Min, Max]	25.0 [5.00, 35.0]	22.0 [5.00, 35.0]	23.0 [5.00, 35.0]
motor_FIM			
Mean (SD)	42.8 (24.6)	35.8 (20.7)	40.8 (23.7)
Median [Min, Max]	37.5 [13.0, 89.0]	31.0 [13.0, 78.0]	36.0 [13.0, 89.0]
total_FIM			
Mean (SD)	66.1 (31.1)	55.8 (25.4)	63.2 (29.8)
Median [Min, Max]	62.5 [19.0, 124]	53.0 [18.0, 103]	59.0 [18.0, 124]
Cohabitation			
Median [Min, Max]	1.00 [1.00, 5.00]	1.00 [1.00, 5.00]	1.00 [1.00, 5.00]
EconomicStatus			

	GREEN (N=84)	RED (N=33)	Overall (N=117)
Median [Min, Max]	1.00 [1.00, 5.00]	2.00 [1.00, 5.00]	1.00 [1.00, 5.00]
HomeAccess			
Median [Min, Max]	2.00 [1.00, 5.00]	3.00 [1.00, 5.00]	2.00 [1.00, 5.00]
FamilySupport			
Median [Min, Max]	2.00 [1.00, 5.00]	3.00 [2.00, 5.00]	3.00 [1.00, 5.00]
SocialSupport			
Median [Min, Max]	2.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]

Validation Results

Confusion matrices as well as standard classification model metrics including AUC, Accuracy, Sensitivity and Specificity, were generated for the validation, for the 25 prospective study meeting inclusion criteria, 92 patients with exclusion criteria, as well as the total 117 set.

Table 10 Confusion matrices for validation utilizing new patients meeting inclusion criteria (25) and additional patients excluded from the dataset not meeting inclusion criteria (92) and combined (117)

	totals		**original**	GREEN	RED
	GREEN	**19**	GREEN	16	4
	RED	**6**	RED	3	2
weighted_method	GREEN	RED	**up_sampling**	GREEN	RED
GREEN	13	1	GREEN	15	1
RED	6	5	RED	4	5
down_sampling	GREEN	RED	**smote_method**	GREEN	RED
GREEN	12	1	GREEN	15	3
RED	7	5	RED	4	3

	totals		**original**	GREEN	RED
	GREEN	**65**	GREEN	59	3
	RED	**27**	RED	6	24
weighted_method	GREEN	RED	**up_sampling**	GREEN	RED
GREEN	47	1	GREEN	46	2
RED	18	26	RED	19	25
down_sampling	GREEN	RED	**smote_method**	GREEN	RED
GREEN	42	1	GREEN	51	4
RED	23	26	RED	14	23

	totals		original	GREEN	RED
	GREEN	84	GREEN	60	2
	RED	33	RED	24	31
weighted_method	GREEN	RED	up_sampling	GREEN	RED
GREEN	60	2	GREEN	61	3
RED	24	31	RED	23	30
down_sampling	GREEN	RED	smote_method	GREEN	RED
GREEN	54	2	GREEN	66	7
RED	30	31	RED	18	26

While the model training confusion matrices (D4.9 Table 9) did not reveal any discernible differences between the model performance, the validation results indicate robust performance by the original model (not corrected for prediction target class imbalance) for the 92 patients with exclusion criteria, and similar performance for the total validation set. Moreover, similarly to model training (D4.9 section 3), all models predominantly misclassify negligible social risk patients (misclassify GREEN patients as RED - False Positive) rather than significant social risk patients (misclassify RED patients as GREEN - False Negative).

However, what is of importance to clinicians is to identify patients with significant social risk (RED class), therefore, model specificity is a critical performance metric in addition to accuracy and AUC. Table 11 below presents these metrics for each of the validation sets.

Table 11. Model validation performance metrics for all validation datasets

GBM validation statistics	original_model	weighted_method	up_sampling	down_sampling	smote_method
AUC (25)	0.798	0.798	0.781	0.789	0.798
AUC (92)	0.925	0.931	0.904	0.923	0.932
AUC (117)	**0.891**	**0.909**	**0.881**	**0.899**	**0.904**
Accuracy (25)	0.72	0.72	0.8	0.68	0.72
Accuracy (92)	0.902	0.794	0.772	0.739	0.804
Accuracy (117)	**0.778**	**0.778**	**0.778**	**0.727**	**0.786**
Sensitivity (25)	0.842	0.684	0.79	0.632	0.79
Sensitivity (92)	0.908	0.723	0.708	0.646	0.785
Sensitivity (117)	**0.714**	**0.714**	**0.726**	**0.643**	**0.786**
Specificity (25)	0.333	0.833	0.833	0.833	0.5
Specificity (92)	0.889	0.963	0.926	0.963	0.852
Specificity (117)	**0.939**	**0.939**	**0.909**	**0.939**	**0.788**

Specificity was higher than sensitivity for model validation for the 92 and 117 patient sets, in contrast to the model training (D4.9 Table 10: model training specificity [0.597 – 0.855], sensitivity [0.800 – 0.916]). Although validation AUC values were generally higher than for model training (D4.9 Table 10: model training AUC [0.827 – 0.843]), validation accuracies were very comparable to the model training indicating that all models had a similarly stable and robust validation (D4.9 Table 10: training model Accuracy [0.811 - 0.880]), especially for the 92 and 117 validation sets. Due to the small number of the RED class in the 25 patient set, the validation outcome of this set is not as stable.

Class Prediction Contributions in Validation Data

Approximate Shapley values were calculated for the 117 patient validation set in the same method as the training for the model, using the R package fastshap (D4.9 Model Explainability section), to indicate the contribution of each of the predictors to the negligible risk prediction (GREEN class) and the significant risk prediction (RED class). Figure 5 to Figure 9 below present the distribution of the prediction results (classification) among the predictors.

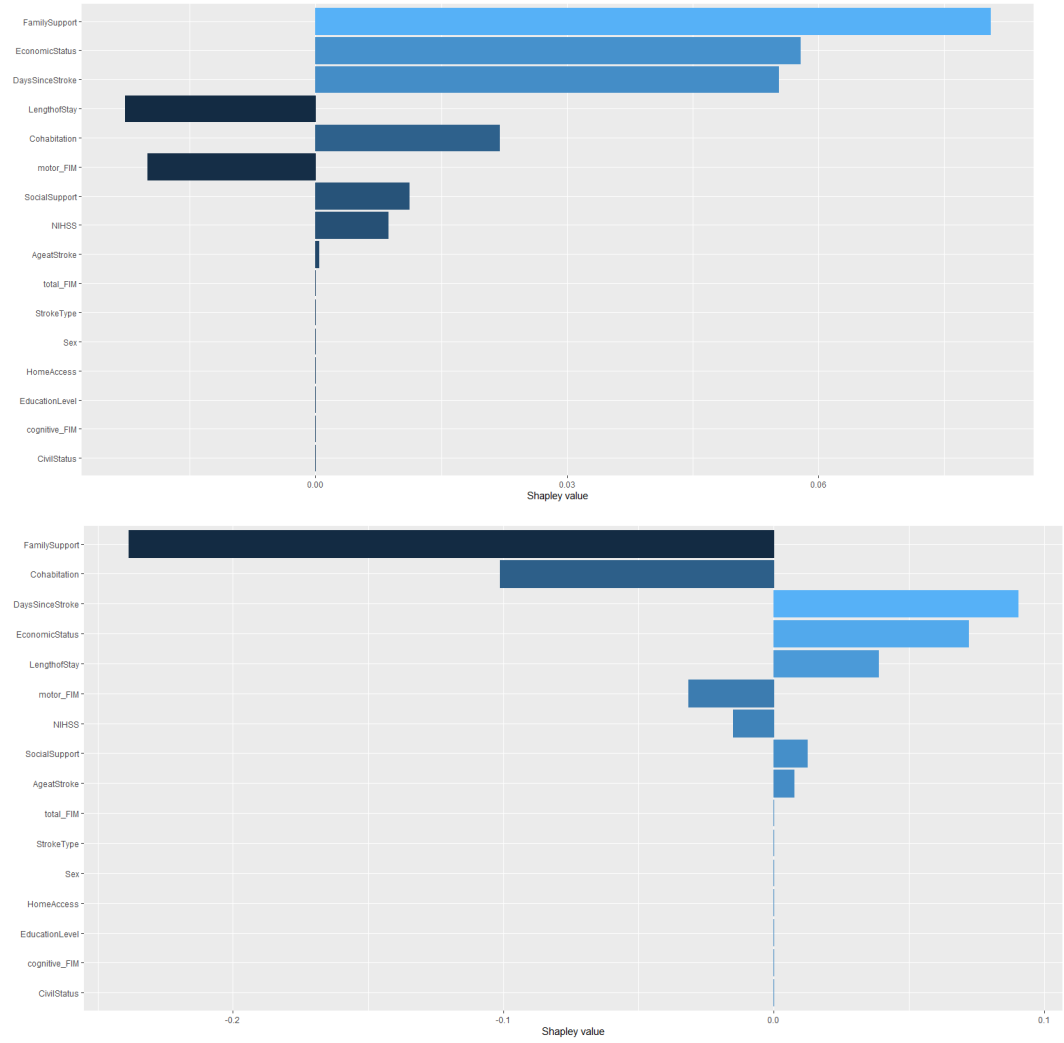


Figure 5. (a) Original model validation approximate Shapley values for GREEN class prediction. (b) Original model validation approximate Shapley values for RED class prediction.

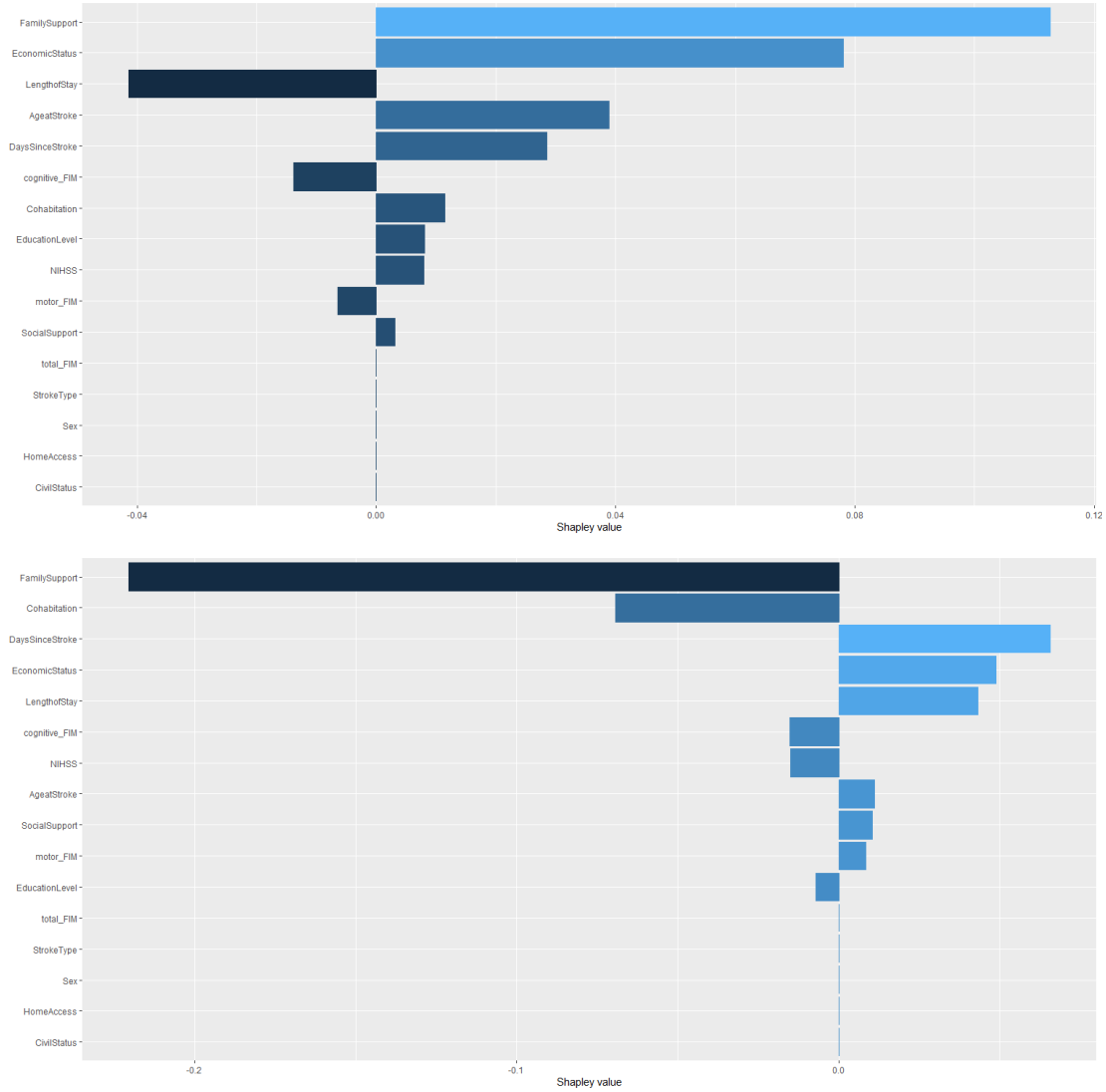


Figure 6.(a) Weighted model validation approximate Shapley values for GREEN class prediction. (b) Weighted model validation approximate Shapley values for RED class prediction.

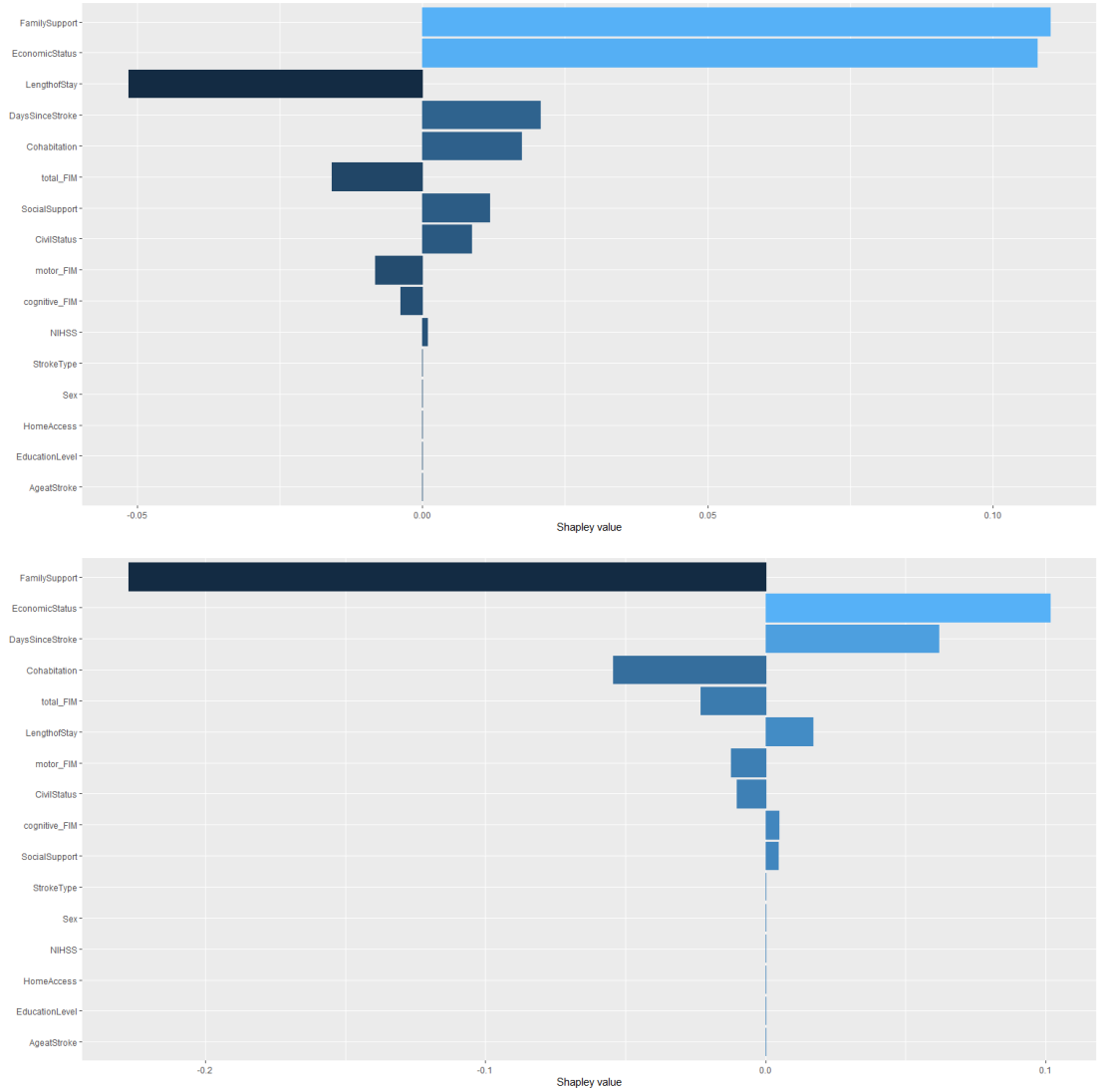
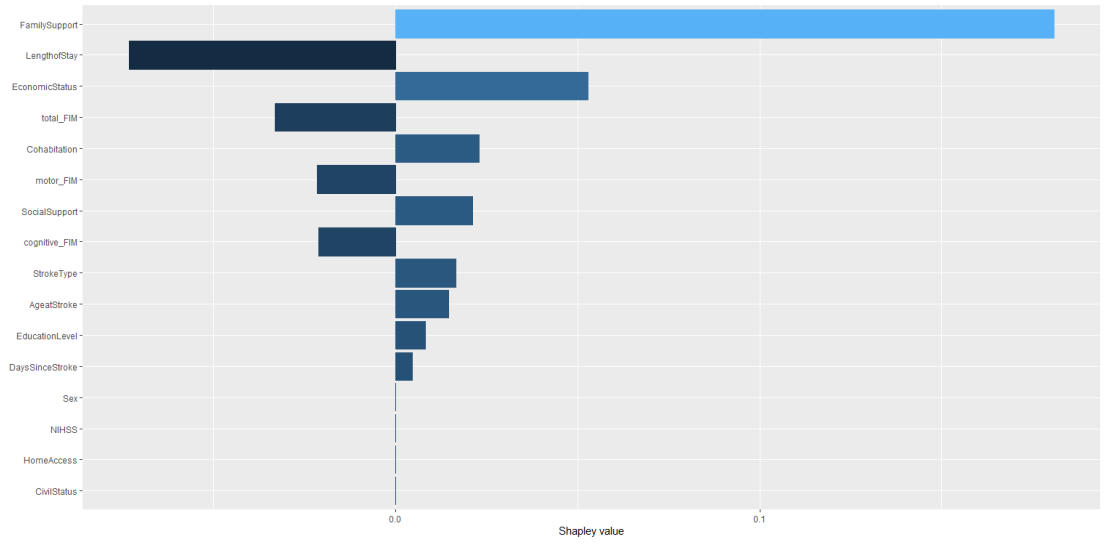


Figure 7. (a) Up-sampling model validation approximate Shapley values for GREEN class prediction. (b) Up-sampling model validation approximate Shapley values for RED class prediction.



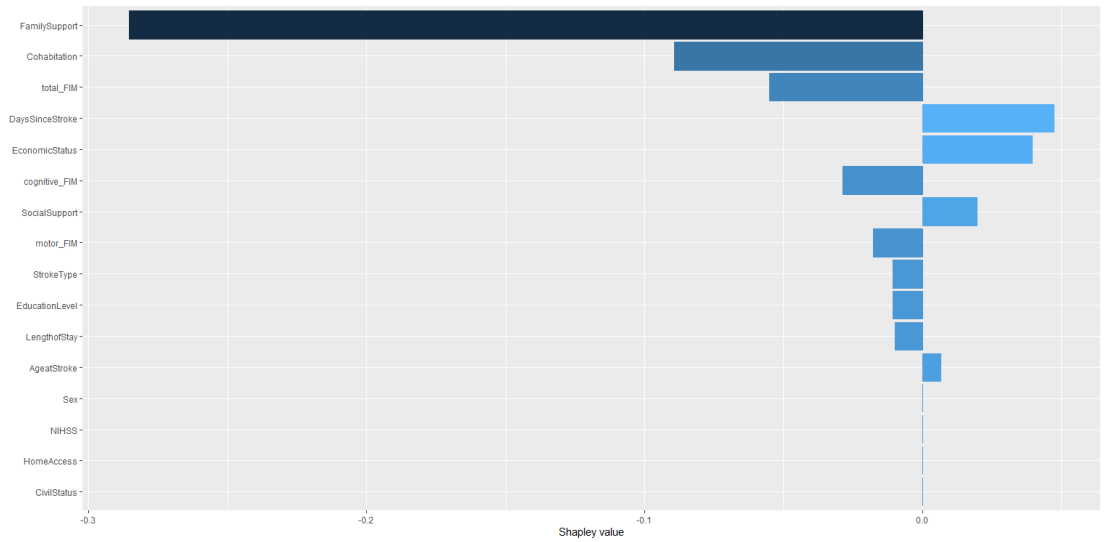


Figure 8. (a) Down-sampling model validation approximate Shapley values for GREEN class prediction. (b) Down-sampling model validation approximate Shapley values for RED class prediction.

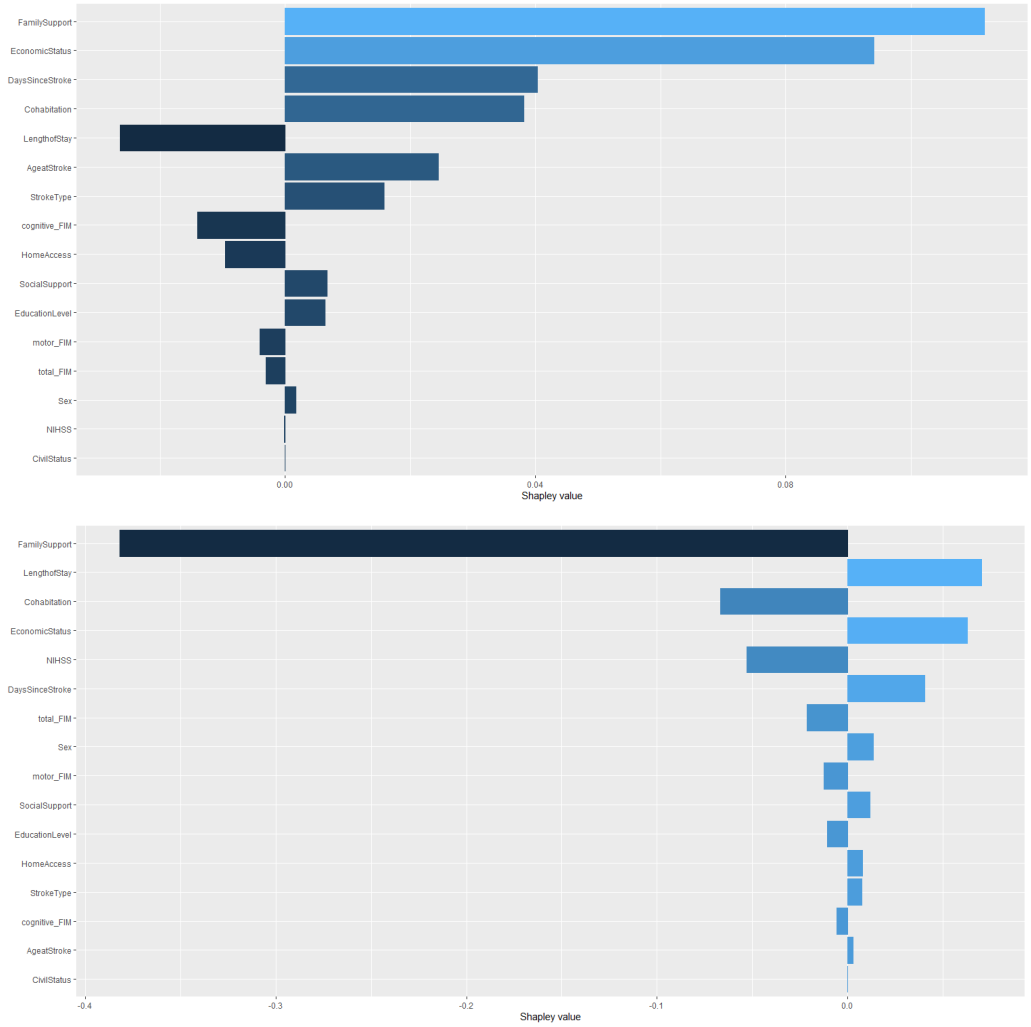


Figure 9. (a) Smote model validation approximate Shapley values for GREEN class prediction. (b) Smote model validation approximate Shapley values for RED class prediction.

For the training set cohort of patients of the Catalonia region of Spain, consisting of mostly male, young ischemic stroke patients, despite the prevalence of individuals in negligible social risk class upon discharge from the hospital, machine learning modeling of this data revealed that predictors contributing to significant social risk were primarily family support and economic status, as well as cohabitation and days since stroke, with lesser contribution of other predictors and specifically no contribution from the sex of the patient (D4.9 Model Explainability section). Similarly for the validation set, FamilySupport is the top predictor for all models, as well as Economic Status and Cohabitation, however, LengthofStay emerges as a highly ranked contributor especially to RED class prediction; this is not unexpected as for many of the validation set patients the LengthofStay meets exclusion criteria. Nevertheless, as in the training sets, the demographic variables, such as Sex and CivilStatus, have negligible contribution to class prediction.

4.3 Conclusions

The robust validation yielding good performance metrics (AUC, accuracy, sensitivity and specificity), consisting of a set of patients that was not part of model training, recapitulates the actual usage of the models by clinicians for patients that may not meet inclusion criteria. This confirms the utility of the models in the real-world clinical scenario, as well as the contribution of not only the EVSF predictors, such as SocialSupport, but also LengthofStay, highlighting that social risk is a complex and multifactorial phenomenon that can vary significantly for patients over the course of stroke rehabilitation and reintegration.

5 Community integration: Long-term trajectories

5.1 Introduction and Background

As presented in D4.9 the aims of this study were to use generalized mixed models (GMM) analysis to identify classes of community dwelling individuals after stroke with different trajectories of community integration up to 13 years after injury, to characterize the identified classes with baseline clinical factors (e.g. functional independence) and identify predictors of the obtained trajectories, using data from Institut Guttmann Neurorehabilitation hospital.

Eligible participants were adult patients (≥ 18 years at the moment of admission, no other age restriction was imposed to participants) with the diagnosis of first-time ischemic or hemorrhagic stroke, receiving inpatient rehabilitation between March 2002 and December 2021. Patients were excluded for the following reasons: less than 3 community integration assessments performed after discharge, more than 90 days since injury onset to inpatient rehabilitation admission as in related research, cases of transient ischemic attack, traumatic brain injury or spinal cord injury diagnosis in the context of first-time ischemic or hemorrhagic stroke, or a previous history of another disabling condition (e.g. cerebral palsy)

A total of 864 patients composed such initial derivation cohort population. After excluding 357 of them with less than three CIQ assessments, 118 with more than 90 days since injury to inpatient rehabilitation admission, 98 without a complete FIM assessment performed 10 days within discharge, 4 with a previous story of disability or another concomitant comorbidity (e.g. cerebral palsy), 287 individuals were included in the derivation study.

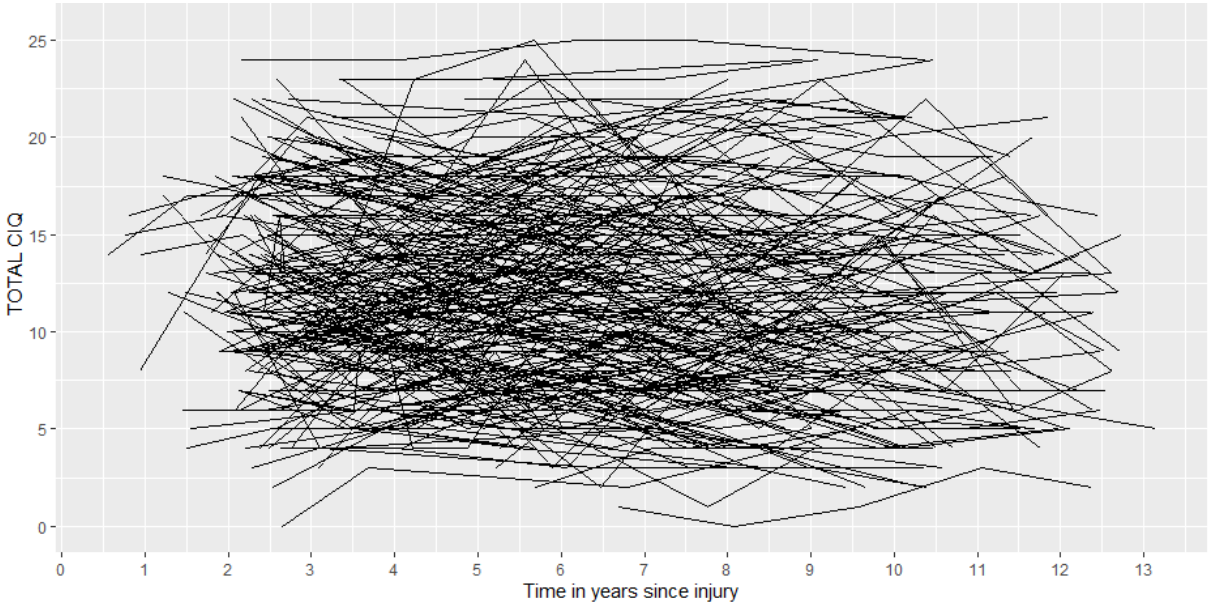


Figure 10. Spaghetti plot for all CIQ assessments included in the derivation cohort

In relation to the derivation cohort, a total of 1264 CIQ assessments were performed between 0.5 and 13 years after injury by the 287 participants, between January 2006 and April 2022. Figure 10 presents the spaghetti plot for them. Each line in the plot represents the trajectory of assessments followed by a participant. No clear pattern of trajectories can be observed.

5.2 Validation activities

Validation patient cohort

The validation cohort included all patients (n=51) with ischemic stroke admitted to inpatient rehabilitation at Institut Guttmann hospital between March 2002 and June 2022 who completed 3 CIQ assessments after rehabilitation discharge between January 2002 and April 2022 and who were not included in the derivation cohort. Patients were excluded for the following reasons: less than 3 community integration assessments performed after discharge, more than 90 days since injury onset to inpatient rehabilitation admission, cases of transient ischemic attack, traumatic brain injury or spinal cord injury diagnosis in the context of first-time ischemic stroke, or a previous history of another disabling condition (e.g. cerebral palsy).

A total of 220 CIQ assessments were performed between 0.5 and 13 years after injury by the 51 participants included in the validation cohort. Figure 11 presents the spaghetti plot for them. Each line in the plot represents the trajectory of assessments followed by a participant. No clear pattern of trajectories can be observed.

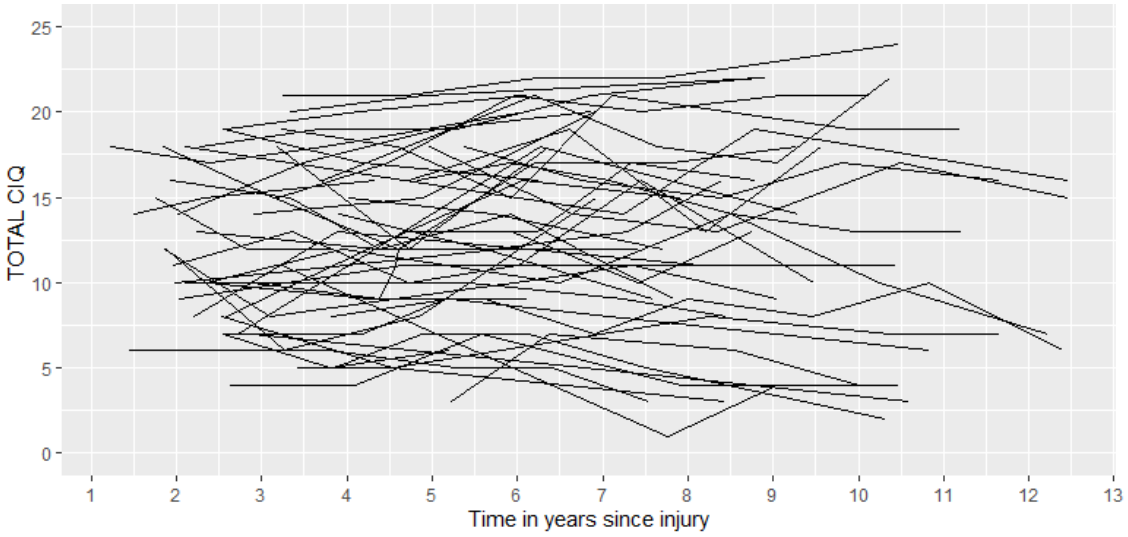


Figure 11. Spaghetti plot for all CIQ assessments included in the validation cohort

Table 12 presents the number of participants assessed at each time point, the time since injury to CIQ assessment, the mean age at the moment of assessment.

Table 12. Validation cohort: time since injury to each assessment point, age at each assessment point.

	1st	2nd	3rd	4th	5th	6th
n	51	51	51	34	19	14
Male, n(%)	33 (64.7)	33 (64.7)	33 (64.7)	22 (64.7)	12 (63.2)	7 (50.0)
Age at CIQ assessment, mean (SD)	52.8 (10.4)	54.3 (10.4)	55.1 (10.4)	56.6 (10.3)	58.7 (9.7)	59.0 (9.7)
Time since injury to CIQ assessment, mean(SD)	3.0 (1.1)	4.7 (1.3)	6.5 (1.7)	8.0 (1.9)	9.3 (1.7)	10.4 (1.5)

TOTAL CIQ, mean (SD)	12.5 (5.0)	12.1 (5.0)	12.2 (4.8)	12.5 (5.7)	13.2 (6.5)	12.4 (6.6)

Table 13 describes the main clinical and demographic characteristics of participants included in the validation cohort.

Table 13. Baseline characteristics for patients included in the validation cohort

Clinical and demographic features	**Included patients (N=51)**
Male, %	33 (64.7%)
Age at injury, mean (SD)	48.3 (11.0)
Younger than 65 years old at injury, %	48 (94.1%)
NIHSS, mean (SD)	10.4 (5.8)
Time since stroke onset to admission, days, mean (SD)	47.7 (30.6)
Hypertension	22 (43.1%)
Dysphagia	18 (35.3%)
Diabetes	2 (3.9%)
Dyslipidemia	13 (25.5%)
Aphasia	17 (33.3%)
Atrial Fibrillation	0 (0.0%)
Neglect	14 (27.5%)
Affected side	
Bilateral	9 (17.6%)
Left	17 (33.3%)
Right	25 (49.0%)
Dominance	
Left	1 (2.0%)
Right	50 (98.0%)
Dominant affected	24 (47.1%)
Smoking habits, %	
Current smoker at admission	14 (27.5%)
Never smoked	28 (54.9%)
Former smoker	9 (17.6%)
Educational level, %	
Read and write	3 (5.9%)
Primary	22 (43.1%)
Secondary	17 (33.3%)

University	9 (17.6%)
Marital status, %	
Married	35 (74.5%)
Single	9 (19.1%)
Separated	2 (4.3%)
Divorced	1 (2.1%)
Widow	0 (0.0%)
LOS in days, mean (SD)	69.8 (31.5)
FIM-at discharge, mean (SD)	
Cognitive FIM	25.255 (9.398)
Motor FIM	59.176 (27.918)
Total FIM	84.431 (36.164)
Motor FIM at discharge categorization, %	
good	32 (62.7%)
fair	2 (3.9%)
poor	17 (33.3%)

NIHSS: National Institutes of Health Stroke Scale; FIM: Functional Independence Measure; LOS: Length of Stay

Validation Results

To fit FMM models we followed the same procedure as in D4.9. We fitted GMMs to total CIQ as outcomes with time since injury (years) having both fixed and random effects, while also allowing for class-specific effects of time. Single class models were fitted with models selected according to the lowest Bayesian information criterion value as presented in Table 14. A cubic term for time gave the lowest Bayesian information criterion values. Models with two to five classes were subsequently fitted. The optimal number of classes was determined by selecting the model with the lowest Bayesian information criterion and highest entropy. Entropy is an indication of how well individuals have been allocated to each class on a scale of 0–1, with 1 meaning perfect classification, and is calculated using the mean posterior probability for each class.

Therefore, the optimal number of classes (three classes) was determined by selecting the model with the lowest Bayesian information criterion and highest entropy. Table 14 presents the obtained BIC and entropy values for the different number of classes.

Table 14. Information criteria for the different fitted models by number of tested classes

Transformation	G	BIC	Entropy	Class1 %	Class2 %	Class3 %	Class4 %	Class5 %
Linear outcome	1	5248	1					
Linear outcome Quadratic term for	1	5112	1					

time								
Linear outcome Cubic term for time	1	5043	1					
Linear outcome Cubic term for time Random effects	2	2471	0.63	59.7	40.3			
Linear outcome **Cubic term for time** **Random effects**	**3**	**2342**	**0.82**	**33.3**	**25.5**	**41.2**		
Linear outcome Cubic term for time Random effects	4	2354	0.71	1.7	16.8	38.6	42.9	
Linear outcome Cubic term for time Random effects	5	2411	0.54	3.9	5.1	42.6	31.3	17.1

G: Number of groups;

Table 15 describes the baseline characteristics for the 3 identified classes: Class 1 (n=17, 33.3%), Class 2 (n= 13, 25.5%) and Class 3 (n=21, 41.2%).

The features describing the three classes identified using the derivation cohort were confirmed with the validation cohort. Individuals in Class 2 were the youngest, with the lowest NIHSS, the lowest proportion of hypertension, aphasia, the shorter LOS, the largest proportion of high educational level. Similarly as presented in D4.9 participants in Class 3 present intermediate demographic and clinical results when compared to Class 2 and Class 1. Participants in Class 1 clearly show the lower levels of Community Integration with highest functional dependence at rehabilitation discharge.

Table 15. Baseline characteristics for participants by class

Features	**Class 1** **(N = 17)**	**Class 2** **(N = 13)**	**Class 3** **(N=21)**	**p**
Male, %	13 (76.5%)	8 (61.5%)	12 (57.1%)	0.446
Age at injury, mean (SD)	55.1 (8.8)	40.3 (13.2)	47.6 (7.3)	0.003
NIHSS, mean (SD)	14.4 (6.9)	4.6 (2.5)	10.9 (4.3)	0.010
Time since stroke onset to admission, days, mean (SD)	62.1 (33.4)	42.0 (31.3)	39.6 (24.5)	0.054
Hypertension, %	7 (41.2%)	5 (38.5%)	10 (47.6%)	0.854
Dysphagia, %	11 (64.7%)	5 (38.5%)	2 (9.5%)	0.002
Diabetes, %	2 (11.8%)	0 (0.0%)	0 (0.0%)	0.125
Dyslipidemia, %	4 (23.5%)	3 (23.1%)	6 (28.6%)	0.914
Aphasia, %	8 (47.1%)	2 (15.4%)	7 (33.3%)	0.190

Atrial Fibrillation, %	0	0	0	
Neglect, %	7 (41.2%)	3 (23.1%)	4 (19.0%)	0.290
Affected side, %				
Bilateral	5 (29.4%)	1 (7.7%)	3 (14.3%)	0.307
Left	6 (35.3%)	6 (46.2%)	5 (23.8%)	
Right	6 (35.3%)	6 (46.2%)	13 (61.9%)	
Dominance, %				
Left	0 (0.0%)	0 (0.0%)	1 (4.8%)	0.483
Right	17 (100.0%)	13 (100.0%)	20 (95.2%)	
Dominant affected, %	6 (35.3%)	6 (46.2%)	12 (57.1%)	0.405
Smoking habits, %				
Current smoker at admission	3 (17.6%)	5 (38.5%)	6 (28.6%)	0.444
Never smoked	10 (58.8%)	7 (53.8%)	11 (52.4%)	0.921
Former smoker	4 (23.5%)	1 (7.7%)	4 (19.0%)	0.517
Educational level, %				
Read and write	1 (5.9%)	1 (7.7%)	1 (4.8%)	0.467
Primary	8 (47.1%)	4 (30.8%)	10 (47.6%)	
Secondary	6 (35.3%)	3 (23.1%)	8 (38.1%)	
University	2 (11.8%)	5 (38.5%)	2 (9.5%)	
Marital status, %				
Married	14 (87.5%)	6 (50.0%)	15 (78.9%)	0.041
Single	0 (0.0%)	6 (50.0%)	3 (15.8%)	
Separated	1 (6.2%)	0 (0.0%)	1 (5.3%)	
Divorced	1 (6.2%)	0 (0.0%)	0 (0.0%)	
Widow	-	-		
LOS in days	79.5 (39.3)	64.4 (29.1)	65.3 (24.9)	0.469
FIM-at discharge				
Cognitive FIM	13.9 (5.8)	33.3 (1.7)	29.4 (4.7)	< 0.001
Motor FIM	22.0 (7.9)	88.0 (2.6)	71.3 (4.9)	< 0.001
Total FIM	36.0 (10.5)	121.3 (3.5)	100.7 (3.4)	< 0.001
Motor FIM at discharge categorization, %				
good	0 (0.0%)	13 (100.0%)	19 (90.5%)	<0.001
fair	0 (0.0%)	0 (0.0%)	2 (9.5%)	
poor	17 (100.0%)	0 (0.0%)	0 (0.0%)	
CIQ total	7.8 (2.7)	17.7 (2.4)	13.1 (4.0)	<0.001
CIQ home	1.9 (2.2)	8.9 (1.3)	5.1 (2.7)	<0.001

CIQ social	5.7 (1.9)	8.1 (1.4)	7.4 (1.8)	0.003
CIQ productivity	0.1 (0.7)	0.6 (1.3)	0.6 (1.2)	0.374

NIHSS: National Institutes of Health Stroke Scale; FIM: Functional Independence Measure; LOS: Length of Stay

5.3 Conclusions

The features describing the three classes identified using the derivation cohort were confirmed with the validation cohort. Individuals in Class 2 were the youngest, with the lowest NIHSS, the lowest proportion of hypertension, aphasia, the shorter LOS, the largest proportion of high educational level. Similarly as presented in D4.9 participants in Class 3 present intermediate demographic and clinical results when compared to Class 2 and Class 1. Participants in Class 1 clearly show the lower levels of Community Integration with highest functional dependence at rehabilitation discharge.

6 Conclusions

This deliverable presented the validation using different datasets of the predictive models developed for the personalised rehabilitation and reintegration stages from WP4. The efforts were focused on four main clinical use cases, two of them in the context of cognitive and motor inpatient rehabilitation and the other two on social risk and community integration trajectories of community dwelling stroke survivors. These validation cohorts showed reliable predictions that could help clinicians at developing personalised rehabilitation and reintegration programs.

Specifically, for the cognitive inpatient rehabilitation, two models were trained to predict: the cognitive improvement after therapy, and therapy compliance. Predictions were accompanied by complementary reports to contextualise this information and allow clinicians to evaluate the inner workings of the model. Performance results showed a drop in performance when the models were tested against the validation cohort. However, re-trained versions of both models reported similar, and for some cases, better results when compared to the models presented in D4.8. A comparison of the features' impact showed how certain variables (e.g., admission compliance) had a strong influence across base, validation and re-trained models.

In relation to motor inpatient rehabilitation the validation cohort clearly confirmed the results obtained with the derivation cohort, both when considering all 33 individual FMA-UE items using unadjusted models and when considering the top 3 items with adjusted models. Besides, as presented in Annex I the total number of included participants (287 in the derivation cohort + 109 in the validation cohort) is clearly larger than most of FMA-UE predictive models presented in previous research.

When addressing social risk, we analysed an outpatient dataset that was not part of model training. This confirms the utility of the models in the real-world clinical scenario, as well as the contribution of not only the EVSF predictors, such as SocialSupport, but also LengthofStay, highlighting that social risk is a complex and multifactorial phenomenon that can vary significantly for patients over the course of stroke rehabilitation and reintegration.

Finally, in relation to Community Integration trajectories the features describing the three classes identified using the derivation cohort were confirmed with the validation cohort. Individuals in Class 2 were the youngest, with the lowest NIHSS, the lowest proportion of hypertension, aphasia, the shorter LOS, the largest proportion of high educational level. Similarly as presented in D4.9 participants in Class 3 present intermediate demographic and clinical results when compared to Class 2 and Class 1. Participants in Class 1 clearly show the lower levels of Community Integration with highest functional dependence at rehabilitation discharge.

7 References

Arya KN, Verma R, Garg RK. Estimating the minimal clinically important difference of an upper extremity recovery measure in subacute stroke patients. Top Stroke Rehabil. 2011 Oct;18 Suppl 1:599-610. doi: 10.1310/tsr18s01-599. PMID: 22120029.

Azen, R., Traxel, N., 2009. Using Dominance Analysis to Determine Predictor Importance in Logistic Regression. J. Educ. Behav. Stat. 34, 319-347. DOI:10.3102/1076998609332754

Béjot, Y., Bailly, H., Durier, J., & Giroud, M. (2016). Epidemiology of stroke in Europe and trends for the 21st century. La Presse Médicale, 45(12), e391-e398.

Bleyenheuft Y, Gordon AM. Precision grip in congenital and acquired hemiparesis: similarities in impairments and implications for neurorehabilitation. Front Hum Neurosci. 2014 Jun 30;8:459. doi: 10.3389/fnhum.2014.00459. PMID: 25071502; PMCID: PMC4074995.

Chen CM, Tsai CC, Chung CY, Chen CL, Wu KP, Chen HC. Potential predictors for health-related quality of life in stroke patients undergoing inpatient rehabilitation. Health Qual Life Outcomes. (2015) 13:118.doi: 10.1186/s12955-015-0314-5

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Duncan Millar J, van Wijck F, Pollock A, Ali M. Outcome measures in post-stroke arm rehabilitation trials: do existing measures capture outcomes that are important to stroke survivors, carers, and clinicians? Clin Rehabil. 2019 Apr;33(4):737-749. doi: 10.1177/0269215518823248. Epub 2019 Jan 15. PMID: 30646750; PMCID: PMC6429625.

Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. Scand J Rehabil Med. 1975;7(1):13-31. PMID: 1135616.

FMA-UE PROTOCOL Rehabilitation Medicine, University of Gothenburg [Last accessed 12-04-2022]https://www.gu.se/sites/default/files/2021-01/FM-UE%20eng%20190303%20PROTOCOL.pdf

García-Molina, A., Rajo, P. R. R., Sánchez-Carrión, R., Pulido, A. G. G., Ensenyat, A., Rudolph, A. G. G., ... & Roig, T. (2010, October). Clinical program of cognitive tele-rehabilitation for traumatic brain injury. In eChallenges e-2010 Conference (pp. 1-10). IEEE.

García-Rudolph, A., & Gibert, K. (2014). A data mining approach to identify cognitive NeuroRehabilitation Range in Traumatic Brain Injury patients. Expert systems with applications, 41(11), 5238-5251.

Ghaziani E, Couppé C, Siersma V, Søndergaard M, Christensen H, Magnusson SP. Electrical Somatosensory Stimulation in Early Rehabilitation of Arm Paresis After Stroke: A Randomized Controlled Trial. Neurorehabil Neural Repair. 2018 Oct;32(10):899-912. doi: 10.1177/1545968318799496. Epub 2018 Sep 25. PMID: 30251591.

Ghaziani E, Couppé C, Siersma V, Christensen H, Magnusson SP, Sunnerhagen KS, Persson HC, Alt Murphy M. Easily Conducted Tests During the First Week Post-stroke Can Aid the Prediction of Arm Functioning at 6 Months. Front Neurol. 2020 Jan 9;10:1371. doi: 10.3389/fneur.2019.01371. PMID: 31993016; PMCID: PMC6962352.

Harari, Y., O'Brien, M. K., Lieber, R. L., & Jayaraman, A. (2020). Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. Journal of neuroengineering and rehabilitation, 17(1), 1-10.

Hatem SM, Saussez G, Della Faille M, Prist V, Zhang X, Dispa D, Bleyenheuft Y. Rehabilitation of Motor Function after Stroke: A Multiple Systematic Review Focused on Techniques to Stimulate Upper Extremity Recovery. Front Hum Neurosci. 2016 Sep 13;10:442. doi: 10.3389/fnhum.2016.00442. PMID: 27679565; PMCID: PMC5020059.

Hiragami S, Inoue Y, Harada K. Minimal clinically important difference for the Fugl-Meyer assessment of the upper extremity in convalescent stroke patients with moderate to severe hemiparesis. J Phys Ther Sci. 2019 Nov;31(11):917-921. doi: 10.1589/jpts.31.917. Epub 2019 Nov 26. PMID: 31871377; PMCID: PMC6879402.

Hoonhorst MH, Nijland RH, van den Berg JS, Emmelot CH, Kollen BJ, Kwakkel G. How Do Fugl-Meyer Arm Motor Scores Relate to Dexterity According to the Action Research Arm Test at 6 Months Poststroke? Arch Phys Med Rehabil. 2015 Oct;96(10):1845-9. doi: 10.1016/j.apmr.2015.06.009. Epub 2015 Jul 2. PMID: 26143054.

Kuhn, M, 2008. Building Predictive Models in R Using the caret Package," Journal of Statistical Software, Foundation for Open Access Statistics, vol. 28(i05). DOI: http://hdl.handle.net/10.18637/jss.v028.i05

Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, Saposnik G, Winstein C, van Wegen EE, Wolf SL, Krakauer JW, Bernhardt J. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2017 Jul;12(5):451-461. doi: 10.1177/1747493017711813. PMID: 28697709.

Kreutzer, J. S., DeLuca, J., & Caplan, B. (Eds.). (2011). Encyclopedia of clinical neuropsychology (p. 103). New York: Springer.

Lindsay, L. R., Thompson, D. A., & O'Dell, M. W. (2020). Updated approach to stroke rehabilitation. Medical Clinics, 104(2), 199-211.

Liu, Z., Xin, H., & Chopp, M. (2021). Axonal remodeling of the corticospinal tract during neurological recovery after stroke. Neural regeneration research, 16(5), 939–943. https://doi.org/10.4103/1673-5374.297060

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Miller, E. L., Murray, L., Richards, L., Zorowitz, R. D., Bakas, T., Clark, P., & Billinger, S. A. (2010). Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. Stroke, 41(10), 2402-2448.

Persson HC, Murphy M, Danielsson A, Lundgren-Nilsson A, Sunnerhagen KS. A cohort study investigating a simple, early assessment to predict upper extremity function after stroke - a part of the SALGOT study. BMC Neurol.(2015) 15:92. doi: 10.1186/s12883-015-0349-6

Plantin J, Verneau M, Godbolt AK, Pennati GV, Laurencikas E, Johansson B, Krumlinde-Sundholm L, Baron JC, Borg J, Lindberg PG. Recovery and Prediction of Bimanual Hand Use After Stroke. Neurology Aug 2021, 97 (7) e706-e719; DOI: 10.1212/WNL.0000000000012366

Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, van Wijck F. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev. 2014 Nov 12;2014(11):CD010820. doi: 10.1002/14651858.CD010820.pub2. PMID: 25387001; PMCID: PMC6469541.

Scrutinio D, Monitillo V, Guida P, Nardulli R, Multari V, Monitillo F, Calabrese G, Fiore P. Functional Gain After Inpatient Stroke Rehabilitation: Correlates and Impact on Long-Term Survival. Stroke. 2015 Oct;46(10):2976-80. doi: 10.1161/STROKEAHA.115.010440. Epub 2015 Sep 3. PMID: 26337968.

Scrutinio D, Lanzillo B, Guida P, Mastropasqua F, Monitillo V, Pusineri M, Formica R, Russo G, Guarnaschelli C, Ferretti C, Calabrese G. Development and Validation of a Predictive Model for Functional Outcome After Stroke Rehabilitation: The Maugeri Model. Stroke. 2017 Dec;48(12):3308-3315. doi: 10.1161/STROKEAHA.117.018058. Epub 2017 Oct 19. PMID: 29051222.

Snickars J, Persson HC, Sunnerhagen KS. Early clinical predictors of motor function in the upper extremity one month post-stroke. J RehabilMed. (2017) 49:216–22. doi: 10.2340/16501977-2205

Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381. doi: 10.1371/journal.pmed.1001381. Epub 2013 Feb 5. PMID: 23393430; PMCID: PMC3564751.

Stinear, Cathy M., Marie-Claire Smith, and Winston D. Byblow. "Prediction tools for stroke rehabilitation." Stroke 50.11 (2019): 3314-3322.

STROBE [Last accessed 12-04-2022] https://strobe-statement.org/index.php?id=available-checklists

Tauchi Y, Kyougoku M, Takahashi K, Okita Y, Takebayashi T. Dimensionality and item-difficulty hierarchy of the Fugl-Meyer assessment of the upper extremity among Japanese patients who have experienced stroke. Top Stroke Rehabil. 2021 Aug 20:1-9. doi: 10.1080/10749357.2021.1965797. Epub ahead of print. PMID: 34414858.

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019, October). What clinicians want: contextualizing explainable machine learning for clinical end use. In Machine learning for healthcare conference (pp. 359-380). PMLR.

Woytowicz EJ, Rietschel JC, Goodman RN, Conroy SS, Sorkin JD, Whitall J, McCombe Waller S. Determining Levels of Upper Extremity Movement Impairment by Applying a Cluster Analysis to the Fugl-Meyer Assessment of the Upper Extremity in Chronic Stroke. Arch Phys Med Rehabil. 2017 Mar;98(3):456-462. doi: 10.1016/j.apmr.2016.06.023. Epub 2016 Aug 9. PMID: 27519928; PMCID: PMC5299057.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87(4), 954-959.

8 Annex I

Table SM1. Literature on FMA-UE predictive models

Study	n	phase
Boissy et al 1997	26	Subacute two months
Feys et al 2000	100	two, six and 12 months after stroke
Shelton et al 2001	171	17 ± 12 days of an initial stroke
Luft et al 2004	21	Chronic: 1 year
Pang et al 2006	63	Chronic ≥ 1 year
Nijland et al 2010	188	72 hours and at 5 and 9 days after stroke
Stinear et al 2012	40	72 h after stroke
Hoonhorst et al 2015	460	at 6 months poststroke
Persson et al 2015	112	10 days and 1 and 12 months
Woytowicz et al 2017	247	chronic at 6 months poststroke
Snickars et al 2017	117	Within 3 days post-stroke
Ghaziani et al 2020	223	at 6 months poststroke
Plantin et al 2021	89	25 ± 7 days from stroke onset