

Nursing Assessment–Derived Risk Profiles and VALENF Instrument–Based Classification in Adult Inpatients: A Two-Stage Machine Learning Study

María Sánchez-Galán

Universitat Jaume I

Óscar Belmonte-Fernández

Universitat Jaume I

María Jesús Valero-Chillerón

chillero@uji.es

Universitat Jaume I

Carlos Cano-Genovés

Universidad de Zaragoza

Irene Llagostera-Reverter

Universitat Jaume I

Juan Carlos Amengual

Universitat Jaume I

David Luna-Aleixos

Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana

Reyes Grangel

Universitat Jaume I

Víctor M. González-Chordá

Universitat Jaume I

Research Article

Keywords: Nursing Assessment, Inpatients, Risk Assessment, Machine Learning, Cluster Analysis, Classification Algorithms, Electronic Health Records

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10680195/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Nursing assessments capture functional status and nursing-sensitive risks that are incompletely represented by diagnostic and administrative data. Integrating these dimensions may support multidimensional risk stratification. This study aimed to identify nursing assessment–derived risk profiles in adult inpatients, develop a VALENF Instrument–based classifier for automated profile assignment, and assess the profile-specific clinical equivalence of classifier-assigned profiles in an independent cohort.

Methods

This retrospective single-centre study included a development cohort of 2666 episodes and an independent cohort of 8218 episodes. Unsupervised clustering using age, sex, Charlson Comorbidity Index, and complete Barthel, Braden, and Downton assessments generated the reference profiles. Logistic regression, random forest, support vector machine, k-nearest neighbours, and ensemble models were evaluated across 1000 repeated 80/20 train–test splits. The final model was retrained on the complete development cohort and applied unchanged to the independent cohort. Profile distribution, ordinal gradients, and profile-specific clinical equivalence were assessed using Spearman correlations, bootstrap confidence intervals, and clinical margins.

Results

Five ordered profiles were identified (1: low-risk, 2: low-risk clinical-burden, 3: moderate-risk, 4: high-risk, 5: very-high-risk). Mean macro-F1 ranged from 0.819 to 0.842 across individual classifiers. Random forest was selected because it combined adequate performance (macro-F1 0.825; 82.43% correct assignments) with the lowest net imbalance towards lower-risk assignments. In the independent cohort, redistribution occurred between the two lower-risk profiles, whereas the remaining profiles had similar proportions. The ordinal gradient was preserved, and 25 of 30 profile-specific comparisons (83.3%) met the clinical-equivalence criterion.

Conclusions

Nursing assessment data supported five clinically interpretable, multidimensional risk profiles. A VALENF-based Random Forest classifier enabled automated assignment with adequate internal performance, while the independent cohort supported preservation of the aggregate clinical profile structure. Multicentre studies with complete reference assessments and prospective evaluation are required before clinical implementation.

1. BACKGROUND

Adult inpatient units care for an increasingly ageing and multimorbid population, whose clinical complexity is not fully captured by the number of diagnoses. The interplay between comorbidity, functional status and care needs is associated with greater resource utilisation, adverse events and poorer outcomes [1–3]. However, the characterisation of multimorbidity among hospitalised patients is heterogeneous and tends to take only limited account of functional and care-related dimensions [2].

Nursing assessments conducted on admission generate structured data on dimensions only partially captured by diagnostic and administrative systems [3]. Diagnosis-related groups (DRGs) may align poorly with patients' care requirements [4], while care complexity is associated with mortality [5] and adverse events [6]. Functional status, pressure injuries and falls are likewise associated with poorer outcomes and longer hospital stays [7, 8]. Yet these dimensions are typically assessed using separate instruments, and few integrated frameworks support systematic assessment of hospitalised patients [9].

The VALENF Instrument integrates functional status and pressure injury and fall risks into seven items and estimates Barthel Index, Braden Scale and Downton Fall Risk Index scores [10, 11]. Integration into routine clinical documentation provides these scores, albeit separately. The relationship among the three constructs raises the possibility of identifying multidimensional profiles to support nursing decision-making.

Clustering techniques can identify clinically interpretable subgroups based on patterns across variables, although the resulting solutions depend on the variables included, the algorithm used and how the clusters are interpreted [12]. Moreover, identifying profiles does not guarantee their operational use: a procedure is needed to assign new episodes to these profiles using variables routinely available in clinical practice [13, 14]. Accordingly, this study aimed to identify nursing assessment-derived risk profiles in adult inpatients, to develop a VALENF Instrument-based classifier for automated profile assignment, and to assess the profile-specific clinical equivalence of classifier-assigned profiles in an independent cohort

2. METHODS

2.1 Study design

A retrospective observational study was conducted using a development cohort and an independent cohort, both drawn from Hospital Universitario La Plana (Spain). The development cohort was used to derive the profiles and to develop and internally evaluate several supervised learning models. The final model was applied without modification to the independent cohort to assess the distribution of the assigned profiles and their clinical equivalence to the original profiles. Study reporting was guided by the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence statement (TRIPOD + AI), where applicable [15]. This analysis was conducted as

part of the registered VALENF-FTR project (ISRCTN77057974) and addresses one of its prespecified objectives.

2.2 Setting and study cohorts

The development cohort, derived from the VALENF Instrument project, included 2,666 hospital episodes involving adult patients admitted to Hospital Universitario La Plana (Spain) between October and December 2020 and during the corresponding period in 2021 [16]. The independent cohort, derived from the VALENF-FTR project, comprised a sample of 8,218 episodes recorded between March 2023 and February 2024. VALENF Instrument data were routinely recorded as part of standard nursing care and retrospectively extracted for analysis.

In both cohorts, the inclusion criteria were age ≥ 18 years, an expected hospital stay of ≥ 48 h and a nursing assessment conducted within the first 24 hours after admission. No additional exclusion criteria were applied. The hospital episode was the unit of analysis. Both datasets are available in the Universitat Jaume I institutional repository [17, 18]. Both projects were approved by the hospital's Research Ethics Committee (VALENF Instrument: approval code 12/01/2021; VALENF-FTR: approval code CEIC-PLANA-13/2024-01), and the requirement for informed consent was waived for both because of their retrospective designs and use of registry data.

2.3 Variables

Data on age, sex, admission type, clinical process type, the Charlson Comorbidity Index [19] and length of hospital stay were available for both cohorts. The development cohort additionally included the total score for each instrument and all 21 individual items across the Barthel Index, Braden Scale and Downton Fall Risk Index [20–22]. Both cohorts included the VALENF Instrument, a seven-item meta-instrument that assesses functional status, pressure injury risk and fall risk, and estimates scores for all three instruments [10, 11].

2.4 Data analysis

Continuous variables were summarised using the mean (m) and standard deviation (SD), or the median (Me) and interquartile range (IQR); categorical variables were summarised using frequencies (n) and percentages (%). The cohorts were compared using absolute standardised differences (ASDs), with values < 0.10 considered negligible [23]. The analytical sequence was then: i) Unsupervised identification of clinical risk profiles; ii) Development and internal evaluation of the VALENF-based classifier; iii) Evaluation of the profiles in an independent cohort (Fig. 1).

[INSERT Fig. 1 HERE]

Figure 1. Analytical pipeline of the study.

2.4.1 Unsupervised identification of clinical risk profiles.

In the development cohort, solutions with four to six profiles were explored using age, sex, Charlson Comorbidity Index, and the 21 items and total scores of the Barthel Index, Braden Scale, and Downton Scale. Variables were standardised using z-scores.

Fuzzy c-means and hierarchical clustering with Ward's method and squared Euclidean distance were applied. Solutions with four to six profiles were compared considering the separation, homogeneity, and size of the groups, the clinical risk gradient, and interpretability.

2.4.2 Development and internal evaluation of the VALENF-based classifier.

The resulting profiles served as the reference labels. The following algorithms were evaluated: logistic regression (LR), random forest (RF) (10 trees trained with 70% random subsamples), support vector machine (SVM), k-nearest neighbours with six neighbours and Euclidean (KNNE) and Manhattan (KNNM) distances, and six ensemble models. The same 1,000 random train–test splits were used for each algorithm. In each split, 80% of the development cohort was used for training and 20% for testing.

The final selection combined predictive performance and clinical direction of errors. For each model, macro-F1 values, correct assignments, and assignments to profiles above and below the reference were calculated. Excess lower assignments were defined as the difference, in percentage points, between lower and higher assignments; values close to zero indicated less imbalance between the two directions. Error patterns were represented using column-normalised confusion matrices. The selected model was retrained on the full cohort and applied without modification to the independent cohort.

2.4.3 Evaluation of the profiles in an independent cohort.

The ordinal gradient was assessed using Spearman's rank correlation coefficient (ρ) between the profile and age, Charlson Comorbidity Index, Barthel Index, Braden Scale score, Downton Scale score, and length of stay. For each variable and profile, the mean difference (independent cohort minus development cohort) and the standardised difference were calculated, using the standard deviation (SD) of the variable in the development cohort as the denominator. 95% confidence intervals (CIs) were estimated using 5,000 independent, non-parametric bootstrap resamples [24]. Based on the interpretation ranges of the instruments and the characterisation of the profiles, the clinical equivalence margins were pre-specified: age ± 5 years; Charlson Comorbidity Index ± 1 point; Barthel Index ± 10 ; Braden Scale score ± 3 ; Downton Scale score ± 1 ; and length of stay ± 2 days. The criterion was met when the entire 95% CI fell within the corresponding margin [25]. Statistical significance was not used to determine equivalence. For categorical variables, differences in proportions and bootstrap 95% confidence intervals were calculated. Analyses were performed using Julia 1.12, with two-tailed tests and $p < 0.05$ where appropriate.

2.5 Use of generative artificial intelligence

During manuscript preparation, the authors used OpenAI ChatGPT to support editorial restructuring, language refinement, literature identification and the preparation of reporting checklists. AI-generated outputs were critically reviewed and independently verified by the authors, including all scientific statements and cited sources. All methodological decisions, analyses and interpretation of the results remained under the responsibility of the authors, who take full responsibility for the content of the manuscript.

3. RESULTS

3.1 Characteristics of the study cohorts

The development and independent cohorts comprised 2,666 and 8,218 episodes, respectively. The independent cohort had a higher mean age ($m = 69.87$ vs 68.09 ; $ASD = 0.103$) and a higher mean Charlson Comorbidity Index score ($m = 3.97$ vs 3.48 ; $ASD = 0.195$). The distributions by sex and clinical process type were similar ($ASD = 0.03$), while emergency admissions were less frequent in the independent cohort (82.04% vs 85.18% ; $ASD = 0.085$). The mean length of stay was 6.53 (development cohort) and 6.89 (independent cohort) days (Table 1).

Table 1
Characteristics of the Development and Independent Cohorts

Variable	Cohort		Difference ^a	ASD ^b
	Development (n = 2666)	Independent (n = 8218)		
	m (95% CI)	m (95% CI)		
Age	68.09 (67.41–68.76)	69.87 (69.50–70.23)	+ 1.78	0.103
Charlson Index	3.48 (3.39–3.57)	3.97 (3.91–4.03)	+ 0.49	0.195
Length of stay	6.53 (6.27–6.78)	6.89 (6.74–7.04)	+ 0.37	0.054
	n (%)	n (%)	Difference ^a	ASD ^b
Male sex	1390 (52.14)	4404 (53.59)	+ 1.45 p.p.	0.029
Urgent admission	2271 (85.18)	6742 (82.04)	–3.14 p.p.	0.085
Medical process	1843 (69.13)	5568 (67.75)	–1.38 p.p.	0.030

a: Independent cohort minus development cohort. For categorical variables, the difference is expressed in percentage points; b: Absolute Standardised Difference.

3.2. Nursing assessment–derived risk profiles

The five-profile hierarchical solution was selected. The low-risk clinical-burden profile was the largest group ($n = 1,026$; 38.5%), whereas the high-risk profile was the smallest ($n = 318$; 11.9%). A progressive

gradient was observed across the five groups, characterised by increasing age and Charlson Comorbidity Index scores, poorer functional status and higher pressure injury and fall risks (Table 2).

The two lower-risk profiles (low-risk and low-risk clinical-burden, respectively) showed near-complete functional independence (Barthel: 99.45 and 98.16), low pressure injury risk (Braden: 21.64 and 21.08) and low fall risk (Downton: 0.09 and 0.49). However, patients in the latter were older (63.74 vs 49.61 years) and had greater comorbidity (Charlson: 2.87 vs 1.26). The moderate-risk profile combined greater comorbidity with an intermediate level of functional impairment ($m = 88.38$) and higher fall risk ($m = 1.71$). The high-risk profile was characterised by marked functional impairment ($m = 53.76$), a lower Braden Scale score ($m = 16.46$) and predominance of female patients (61.32%). Finally, the very-high-risk profile comprised older patients ($m = 82.28$) with greater comorbidity ($m = 5.28$), high functional dependence ($m = 9.09$) and the highest pressure injury ($m = 11.64$) and fall risks ($m = 2.86$). Urgent admissions (97.15%) and medical processes (90.50%) predominated in this group. Item-level descriptive analyses for the Barthel Index, Braden Scale and Downton Fall Risk Index by profile are presented in Additional file 1: Table S1.

Table 2

Descriptive characteristics of the five-cluster solution obtained from hierarchical cluster analysis.

	Profiles (n; %)				
	Low-risk	Low-risk clinical- burden	Moderate- risk	High-risk	Very-high- risk
Hospitalization episodes^a	401 (15.04)	1026 (38.48)	500 (18.75)	318 (11.93)	421 (15.79)
Sex^b					
Men	242 (60.35)	554 (54.00)	282 (56.40)	123 (38.68)	189 (44.89)
Women	159 (39.65)	472 (46.00)	218 (43.60)	195 (61.32)	232 (55.11)
Process type^b					
Medical	215 (53.62)	697 (67.93)	337 (67.40)	213 (66.98)	381 (90.50)
Surgical	186 (46.38)	329 (32.07)	163 (32.60)	105 (33.02)	40 (9.50)
Admission type^b					
Urgent	316 (78.80)	802 (78.17)	447 (89.40)	297 (93.40)	409 (97.15)
Scheduled	85 (21.20)	224 (21.83)	53 (10.60)	21 (6.60)	12 (2.85)
	m (95% CI)	m (95% CI)	m (95% CI)	m (95% CI)	m (95% CI)
Age	49.61 (48.05– 51.16)	63.74 (62.82– 64.65)	74.47 (73.29– 75.63)	76.58 (74.86– 78.30)	82.28 (81.11– 83.45)
Charlson Comorbidity Index	1.26 (1.09– 1.43)	2.87 (2.74– 3.00)	4.29 (4.09– 4.48)	4.58 (4.34– 4.82)	5.28 (5.10– 5.45)
Functional capacity (Barthel Index)	99.45 (99.17– 99.73)	98.16 (97.83– 98.49)	88.38 (87.15– 89.60)	53.76 (51.68– 55.83)	9.09 (7.77– 10.39)
Pressure injury risk (Braden Scale)	21.64 (21.51– 21.76)	21.08 (20.98– 21.17)	19.28 (19.08– 19.47)	16.46 (16.17– 16.62)	11.64 (11.39– 11.88)
Fall risk (Downton)	0.09 (0.06– 0.126)	0.49 (0.45– 0.51)	1.71 (1.64– 1.77)	2.03 (1.91– 2.15)	2.86 (2.75– 2.96)
a: Percentages per row; b: Percentages per column					

3.3. VALENF-Based Classifier model selection and internal performance

Mean macro-F1 ranged from 0.819 to 0.842 for the individual classifiers and from 0.820 to 0.838 for the ensemble models. SVM achieved the highest macro-F1 (0.842) and proportion of correct assignments (84.10%); SVM + LR+RF was the best-performing ensemble (macro-F1 = 0.838; 83.69% correct assignments) and had the lowest proportion of assignments below the reference profile (10.02%) (Table 3).

Table 3

Internal performance and directional assignment patterns across 1000 repeated train–test splits

Algorithm ^a	Mean macro- F1 (95% CI)	Assignment patterns (%)			
		Correct	To higher-risk	To lower-risk	Excess assignment ^b
Individual classifiers					
SVM	0.842 (0.841–0.843)	84.10	5.66	10.24	4.58
LR	0.833 (0.832–0.833)	83.06	6.45	10.49	4.04
RF	0.825 (0.824–0.826)	82.43	7.40	10.17	2.77
KNNE	0.821 (0.820–0.822)	82.03	6.51	11.46	4.95
KNNM	0.819 (0.818–0.819)	81.80	6.92	11.29	4.37
Ensemble classifiers					
SVM + LR + RF	0.838 (0.837–0.839)	83.69	6.29	10.02	3.73
SVM + KNNE	0.831 (0.830–0.832)	83.07	6.09	10.84	4.75
SVM + KNNM	0.830 (0.829–0.831)	82.93	6.29	10.78	4.49
LR + KNNM	0.825 (0.824–0.826)	82.45	6.68	10.87	4.19
SVM + KNNE + KNNM	0.823 (0.822–0.824)	82.24	6.43	11.32	4.89
KNNE + KNNM	0.820 (0.819–0.820)	81.89	6.74	11.37	4.63

a: SVM: support vector machine; RF: Random Forest; KNNE = K-nearest neighbours using Euclidean distance; KNNM: K-nearest neighbours using Manhattan distance; LR: logistic regression.

b: Excess assignment was calculated as lower-risk assignment minus higher-risk assignment. Values closer to zero indicate less imbalance between the two directions of misclassification; positive values indicate a net excess of lower-risk assignments.

Note

Risk profiles were ordered from 1 —lower risk— to 5 —highest risk—. Assignment to a higher-risk profile refers to cases classified into a profile with a higher numerical value than the reference profile. Assignment to a lower-risk profile refers to cases classified into a profile with a lower numerical value than the reference profile.

RF achieved a macro-F1 of 0.825 and 82.43% correct assignments, with the highest proportion of assignments above the reference profile (7.40%) and the smallest excess of below-reference over above-reference assignments (2.77 percentage points). Based on these criteria, RF was selected as the final model, retrained using the full development cohort and applied without modification to the independent cohort. Figure 2 shows the confusion matrices for the five individual classifiers and the best-performing ensemble. Across all models, profile 4 had the lowest proportion of correct assignments and was most frequently confused with profiles 3 and 5. The RF-assigned profiles showed a clinical structure similar to that derived from the clustering analysis, with higher-risk profiles characterised by increasing age and comorbidity, poorer functional status and higher pressure injury and fall risks (Fig. 3).

[INSERT Fig. 2 HERE]

Figure 2. Mean confusion matrices for individual classifiers and the best-performing ensemble across 1,000 train–test splits.

Note

Columns represent the clustering-derived observed profiles and rows represent the classifier-assigned profiles. Cell values are mean column-normalised percentages calculated within each observed profile. Risk profiles were ordered from 1 (lowest risk) to 5 (highest risk). Random Forest was the selected final classifier.

[INSERT Fig. 3 HERE]

Figure 3. Clinical structure of clustering-derived and Random Forest–assigned risk profiles in the development cohort.

Note

Left: shows profiles derived from the complete nursing assessment using hierarchical clustering. Right: shows profiles assigned using the VALENF-based Random Forest classifier. Variables were rescaled to a common metric for visualisation; functional capacity and pressure injury risk scores were directionally transformed so that higher values consistently represented greater clinical risk.

3.4. Evaluation of the profiles in the independent cohort

In the independent cohort, the proportion of episodes assigned to the low-risk profile increased from 15.04% to 27.35% (difference: +12.31 percentage points; 95% CI: 10.71–13.98), whereas the proportion assigned to the low-risk clinical-burden profile decreased from 38.48% to 25.74% (– 12.75 percentage points; 95% CI: –14.86 to – 10.71). The proportions assigned to the moderate-, high- and very-high-risk profiles were very similar between cohorts (Table 4).

The ordinal gradient was maintained in the independent cohort. Profile level was positively correlated with age ($p = 0.572$), Charlson Comorbidity Index ($p = 0.521$), Downton Fall Risk index ($p = 0.933$) and length of stay ($p = 0.193$) and inversely correlated with Barthel Index ($p = -0.853$) and Braden Scale ($p = -0.868$); all correlations were statistically significant ($p < 0.001$).

Table 4

Distribution and profile-specific clinical equivalence of classifier-assigned profiles in the independent cohort

Distribution of risk profiles				
Profile ^a	Cohort n (%)		Difference	
	Development	Independent	Percentage points (95% CI)	Standardised
P1	401 (15.04)	2248 (27.35)	+ 12.31 (10.71–13.98)	+ 0.305
P2	1026 (38.48)	2115 (25.74)	–12.75 (– 14.86 – 10.71)	–0.276
P3	500 (18.75)	1618 (19.69)	+ 0.93 (– 0.80–2.71)	+ 0.024
P4	318 (11.93)	934 (11.37)	–0.56 (– 1.95 – 0.88)	–0.018
P5	421 (15.79)	1303 (15.86)	+ 0.06 (– 1.54–1.66)	+ 0.002
Summary of profile-specific clinical equivalence				
Variable	Clinical margin	Profiles meeting criterion ^a	Largest mean difference	
			Raw	Standardised
Age	± 5 years	3/5 (P3, P4, P5)	+ 6.96 (P1)	0.392 (P1)
Charlson	± 1 point	3/5 (P3, P4, P5)	+ 0.88 (P1)	0.366 (P1)
Barthel	± 10 points	5/5	–5.60 (P4)	0.163 (P4)
Braden	± 3 points	5/5	+ 0.77 (P3)	0.195 (P3)
Downton	± 1 point	5/5	+ 0.57 (P2)	0.461 (P2)
Length of stay	± 2 days	4/5 (P1, P2, P3, P5)	+ 1.50 (P4)	0.223 (P4)
a: P1: Low-risk profile; P2: Low-risk clinical-burden profile; P3: Moderate-risk profile; P4: High-risk profile; P5: Very-high-risk profile				
Note: Differences are independent cohort minus development cohort. The profile-specific criterion was met when the complete 95% bootstrap confidence interval for the mean difference was contained within the clinical margin. Standardised differences were calculated using the standard deviation of the corresponding variable in the complete development cohort.				

Of the 30 comparisons, 25 (83.3%) met the clinical equivalence criterion. The criterion was met for the Barthel Index, Braden Scale and Downton Fall Risk Index across all five profiles and for length of stay in

four. The moderate and very-high-risk profiles met the criterion for all six variables. The criterion was not met for age and the Charlson Comorbidity Index in the low-risk and low-risk clinical-burden profiles, or for length of stay in the high-risk profile. Compared with the development cohort, patients assigned to the low-risk and low-risk clinical-burden profiles in the independent cohort were older (+ 6.96 and + 4.18 years) and had higher Charlson Comorbidity Index values (+ 0.88 and + 0.84 points). Length of stay in the high-risk profile was 1.50 days longer in the independent cohort. Detailed results are presented in Additional file 1: Table S2.

4. DISCUSSION

This study not only identified five risk groups but also translated nursing assessment into a multidimensional, ordered representation of vulnerability on admission. The two lower-risk profiles showed preserved functional independence and low pressure injury and fall risks but differed in age and comorbidity. Greater disease burden may coexist with preserved function and does not alone imply greater needs for functional support or preventive care. This non-linear pattern is consistent with evidence that the relationship between multimorbidity and functional status varies by disease combinations rather than their number [26]. The profiles represent multidimensional configurations on admission, not prognostic categories.

The transition across the moderate, high and very-high-risk profiles was gradual. The high-risk profile showed the greatest classification uncertainty. The underlying dimensions vary continuously, whereas clustering introduces discrete operational boundaries. Ocagli et al. [13] also identified an intermediate zone of assignment uncertainty in care-intensity classification. Clustering solutions depend on variables, scales, metrics and algorithms and may emerge without clear natural boundaries [12]. The five profiles therefore represent a clinically interpretable, data-driven structure rather than fixed or universal patient phenotypes; episodes near profile boundaries could be accompanied by membership probabilities or review alerts.

The supervised phase translated this classification into a VALENF-based classifier using the hospital's routine nursing assessment. Although SVM achieved the highest macro-F1 (0.842) and the highest proportion of correct assignments (84.10%), its absolute advantage over RF was limited (0.017 in macro-F1; 1.67 percentage points in correct assignments). By contrast, RF showed the highest proportion of assignments above the reference profile (7.40%) and the smallest net imbalance towards below-reference assignments (2.77 percentage points). Given that the classifier is intended to support risk stratification and prioritisation of preventive interventions, this pattern was considered clinically more cautious because it reduced the net imbalance towards lower-risk assignment. RF selection therefore balanced overall performance against the clinical direction of errors, rather than indicating predictive superiority. However, higher-risk assignment could prompt unnecessary assessments or preventive interventions. Longitudinal studies should compare the potential clinical and economic consequences of higher- and lower-risk assignments [27]. The independent cohort provided evidence regarding the clinical stability of the profiles in a different case mix but did not constitute a conventional external validation of

classification accuracy. Changes in profile distribution were concentrated in the two lower-risk profiles, whereas the proportions of the moderate-, high- and very-high-risk profiles differed little between cohorts. The ordinal gradient was maintained in the independent cohort and 25 of the 30 comparisons met the pre-specified criteria for clinical equivalence. The profile structure thus appears relatively robust yet sensitive to population change. The literature on dataset shift cautions that temporal changes in patient characteristics, clinical practice or documentation may alter model behaviour within one organisation [28, 29]. The clustering-derived reference profiles were unavailable for the independent cohort because the VALENF Instrument, rather than the complete Barthel, Braden and Downton assessments used for clustering, was routinely recorded during the study period.

Furthermore, Barthel, Braden and Downton scores in the development cohort were obtained from the original instruments, whereas those in the independent cohort were estimated using the VALENF Instrument. The preservation of the risk gradients is a favourable finding, but it may partly reflect the relationship between the seven VALENF items and the estimated scores for the three instruments. Accordingly, differences in age and Charlson Comorbidity Index values in the two lower-risk profiles, and in length of stay in the high-risk profile, are particularly informative because these variables are not directly derived from the VALENF Instrument. External validation should involve concurrent application of the classifier and the full nursing assessment to generate classifier-independent reference profiles and examine associations between profile assignments and outcomes such as functional decline, pressure injuries, falls or in-hospital mortality. Another avenue would be to examine whether changes in risk profile occur in relation to predefined clinical events, such as surgical procedures, transfers to or from the intensive care unit (ICU) or failure-to-rescue events.

From a clinical informatics perspective, the study's contribution lies in linking profiles derived from a comprehensive nursing assessment with profile assignment based on seven routinely documented items. This approach may facilitate the synthesis and visualisation of multidimensional information, but the present study does not yet demonstrate improvements in decision-making, resource use or outcomes. Evidence on nursing classification systems remains limited [30]. Before implementation, profile-specific interventions, procedures for managing uncertain assignments and arrangements for professional oversight need to be defined, and the model's comprehensibility and integration into clinical workflow need to be evaluated. FUTURE-AI [27], DECIDE-AI [31] and the implementation literature emphasise prospective evaluation and consideration of human and organisational factors to determine whether retrospective performance translates into clinical utility [32].

The study's strengths include routinely collected structured nursing data, the combination of unsupervised and supervised learning, the use of the same 1,000 data splits across algorithms, explicit consideration of the direction of classification errors, pre-specified clinical equivalence margins and application of the final model without modification to an independent cohort. The limitations include the retrospective, single-centre design; dependence on variable and algorithm selection; limited hyperparameter optimisation; the absence of a complete, independent reference assessment; and the lack of clinical outcome data and conventional external evaluation. The findings support the clinical

coherence and technical feasibility of the approach, but do not establish its effectiveness. The next step should combine multicentre validation using a complete reference assessment with prospective evaluation of clinical utility and safety.

5. CONCLUSIONS

Five clinical risk profiles derived from nursing assessments conducted on admission were identified, and a classifier based on the VALENF Instrument was developed for profile assignment, with adequate internal performance and a lower imbalance towards assignments to lower-risk profiles. The ordinal gradient was preserved, and most profile-specific comparisons met the clinical-equivalence criterion in the independent cohort, although between-cohort differences were observed in the distribution of the lower-risk profiles, the age and comorbidity of patients assigned to them and the length of stay of patients assigned to the high-risk profile. These findings support the feasibility of using nursing assessment data for automated, clinically interpretable risk stratification. Multicentre studies are needed to assess the accuracy, stability and impact on care delivery of this approach and to determine whether it has clinical utility.

Abbreviations

ASD

Absolute standardised difference

CI

Confidence interval

DRG

Diagnosis-related group

ICU

Intensive care unit

IQR

Interquartile range

KNNE

k-nearest neighbours using Euclidean distance

KNNM

k-nearest neighbours using Manhattan distance

LR

Logistic regression

RF

Random forest

SD

Standard deviation

SVM

Support vector machine

TRIPOD + AI

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence

Declarations

Ethics approval and consent to participate:

This study analysed two datasets derived from the VALENF Instrument and VALENF-FTR projects. Both source projects were approved by the Research Ethics Committee of Hospital Universitario La Plana (VALENF Instrument: approval reference 12/01/2021; VALENF-FTR: approval reference CEIC-PLANA-13/2024-01). The requirement for informed consent was waived by the Research Ethics Committee for both projects because of their retrospective designs and use of registry data. All methods were conducted in accordance with the Declaration of Helsinki and applicable institutional and national ethical requirements.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

The VALENF Instrument project was supported by an internal research grant from Universitat Jaume I [grant number UJI-A2020-08]. The VALENF-FTR project was supported by the Generalitat Valenciana, Conselleria de Educación, Cultura, Universidades y Empleo [grant number CIGE/2024/10]. The funders had no role in the design of the present study; data analysis or interpretation; preparation of the manuscript; or the decision to submit the article for publication.

Author Contribution

[MSG]: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Writing – review and editing. [OBF]: Conceptualization, Methodology, Data curation, Software, Formal analysis, Writing – original draft, Writing – review and editing. [MJVC]: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review and editing, Funding acquisition. [CCG]: Data curation, Software, Validation, Writing – review and editing. [ILR]: Conceptualization, Writing – review and editing. [JCAA]: Data curation, Software, Validation, Writing –

review and editing. [DLA]: Writing – review and editing. [RGS]: Writing – review and editing. [VMGC]: Conceptualization, Methodology, Formal analysis, Visualization, Resources, Project administration, Funding, Supervision, Writing – review and editing. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank the Information Technology Department of Hospital Universitario La Plana for its support in data extraction and provision, and Rafael Martínez-González in particular for his assistance with data retrieval and preparation.

Data Availability

Availability of data and materials: The datasets analysed during the current study are publicly available in the institutional repository of Universitat Jaume I [17: <http://hdl.handle.net/10234/734861>, 18: <http://hdl.handle.net/10234/814342>]. The analysis code, software environment, model specifications, hyperparameters, random seeds and reproducibility documentation are publicly available in the VALENF risk-profile classifier repository [33]

References

1. Georgiev K, McPeake J, Shenkin SD, Fleuriot J, Lone N, Guthrie B, et al. Understanding hospital activity and outcomes for people with multimorbidity using electronic health records. *Sci Rep Nat Res.* 2025;15. <https://doi.org/10.1038/s41598-025-92940-7>.
2. Bunn JG, Steell L, Hillman SJ, Witham MD, Sayer AA, Cooper R. Approaches to characterising multimorbidity in older people accessing hospital care: a scoping review. *Eur Geriatr Med [Internet]* Springer Sci Bus Media Deutschland GmbH. 2025;16:1099–113. <https://doi.org/10.1007/s41999-025-01166-3>.
3. Huber E, Kleinknecht-Dolf M, Kugler C, Spirig R. Patient-related complexity of nursing care in acute care hospitals – an updated concept. *Scand J Caring Sci [Internet]* Blackwell Publishing Ltd. 2021;35:178–95. <https://doi.org/10.1111/scs.12833>.
4. Fabrizi D, Bernasconi DP, Locatelli G, Luciani M, Beretta G, Di Mauro S et al. How Are Diagnosis-Related Groups and Staffing Allocation Systems Associated with the Complexity of Nursing Care? An Observational Study. *Healthcare [Internet]*. Multidisciplinary Digital Publishing Institute (MDPI); 2024;12:1988. <https://doi.org/10.3390/healthcare12191988>
5. Ausili D, Bernasconi DP, Rebora P, Prestini L, Beretta G, Ferraioli L, et al. Complexity of nursing care at 24 h from admission predicts in-hospital mortality in medical units: A cohort study. *BMC Health Serv Res BioMed Cent Ltd.* 2020;20. <https://doi.org/10.1186/s12913-020-5038-5>.
6. Adamuz J, Juvé-Udina M-E, González-Samartino M, Jiménez-Martínez E, Tapia-Pérez M, López-Jiménez M-M et al. Care complexity individual factors associated with adverse events and in-

- hospital mortality. Ramagopalan S V., editor. PLoS One [Internet]. Public Library of Science; 2020;15:e0236370. <https://doi.org/10.1371/journal.pone.0236370>
7. Koch D, Kutz A, Haubitz S, Baechli C, Gregoriano C, Conca A, et al. Association of functional status and hospital-acquired functional decline with 30-day outcomes in medical inpatients: A prospective cohort study. *Applied Nursing Research* [Internet]. Volume 54. W.B. Saunders; 2020. p. 151274. <https://doi.org/10.1016/j.apnr.2020.151274>.
 8. Hasan B, Bechenati D, Bethel HM, Cho S, Rajjoub NS, Murad ST, et al. A Systematic Review of Length of Stay Linked to Hospital-Acquired Falls, Pressure Ulcers, Central Line–Associated Bloodstream Infections, and Surgical Site Infections. *Mayo Clin Proc Innov Qual Outcomes* [Internet]. Elsevier B V. 2025;9:100607. <https://doi.org/10.1016/j.mayocpiqo.2025.100607>.
 9. Wiseman T, Kourouche S, Jones T, Kennedy B, Curtis K. The impact of whole of patient nursing assessment frameworks on hospital inpatients: A scoping literature review. *J. Adv. Nurs.* John Wiley and Sons Inc; 2024. pp. 3448–63. <https://doi.org/10.1111/jan.16025>
 10. Luna-Aleixos D, Llagostera-Reverter I, Castelló-Benavent X, Aquilué-Ballarín M, Mecho-Montoliu G, Cervera-Gasch Á, et al. Development and Validation of a Meta-Instrument for Nursing Assessment in Adult Hospitalization Units (VALENF Instrument) (Part I). *Int J Environ Res Public Health* [Internet] MDPI. 2022;19:14622. <https://doi.org/10.3390/ijerph192214622>.
 11. Luna-Aleixos D, Llagostera-Reverter I, Castelló-Benavent X, Aquilué-Ballarín M, Mecho-Montoliu G, Cervera-Gasch Á, et al. Development and Validation of a Meta-Instrument for the Assessment of Functional Capacity, the Risk of Falls and Pressure Injuries in Adult Hospitalization Units (VALENF Instrument) (Part II). *Int J Environ Res Public Health* [Internet]. 2023;20:5003. <https://doi.org/10.3390/ijerph20065003>. Multidisciplinary Digital Publishing Institute (MDPI).
 12. Loftus TJ, Shickel B, Balch JA, Tighe PJ, Abbott KL, Fazzone B et al. Phenotype clustering in health care: A narrative review for clinicians. *Front Artif Intell* [Internet]. 2022;5. <https://doi.org/10.3389/frai.2022.842306>
 13. Ocagli H, Lanera C, Lorenzoni G, Prosepe I, Azzolina D, Bortolotto S, et al. Profiling Patients by Intensity of Nursing Care: An Operative Approach Using Machine Learning. *J Pers Med* [Internet] MDPI AG. 2020;10:279. <https://doi.org/10.3390/jpm10040279>.
 14. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, et al. Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach*. *Crit Care Med* [Internet] Lippincott Williams Wilkins. 2017;45:1607–15. <https://doi.org/10.1097/CCM.0000000000002548>.
 15. 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* [Internet] BMJ Publishing Group. 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>.
 16. Llagostera-Reverter I, Ortiz-Mallasén V, Valero-Chillerón MJ, Luna-Aleixos D, Cervera-Gasch Á, Orts-Cortés MI et al. Risk Profiles in Hospitalized Adults Based on Functional Capacity, Pressure Injury

- Risk, and Fall Risk: A Cluster and Multiple Correspondence Analysis. *Journal of Nursing Scholarship* [Internet]. John Wiley and Sons Inc; 2026;58. <https://doi.org/10.1111/jnu.70096>
17. Llagostera-Reverter I, González-Chordá V, Luna-Aleixos D, Valero-Chillerón MJ. Risk Profiles in Hospitalized Adults [Dataset]. <http://hdl.handle.net/10234/734861>. 2025.
 18. Valero-Chillerón MJ, Cano Genovés C, Llagostera Reverter I, Amengual J-C, Sánchez Galán M, Belmonte-Fernández Ó et al. VALENF Risk Profiles in Hospitalized Adults: Independent Cohort [Dataset]. <http://hdl.handle.net/10234/814342>. 2026.
 19. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. [New York, NY]: Elsevier Science Ltd.; 1994;47:1245–51. [https://doi.org/10.1016/0895-4356\(94\)90129-5](https://doi.org/10.1016/0895-4356(94)90129-5)
 20. González N, Bilbao A, Forjaz MJ, Ayala A, Orive M, Garcia-Gutierrez S, et al. Psychometric characteristics of the Spanish version of the Barthel Index. *Aging Clin Exp Res* [Internet] Springer Int Publishing. 2018;30:489–97. <https://doi.org/10.1007/s40520-017-0809-5>.
 21. Huang C, Ma Y, Wang C, Jiang M, Yuet Foon L, Lv L et al. Predictive validity of the braden scale for pressure injury risk assessment in adults: A systematic review and meta-analysis. *Nurs. Open*. John Wiley and Sons Inc; 2021. pp. 2194–207. <https://doi.org/10.1002/nop2.792>
 22. Bueno-García MJ, Roldán-Chicano MT, Rodríguez-Tello J, Meroño-Rivera MD, Dávila-Martínez R, Berenguer-García N. Características de la escala Downton en la valoración del riesgo de caídas en pacientes hospitalizados. *Enferm Clin Elsevier Doyma*. 2017;27:227–34. <https://doi.org/10.1016/j.enfcli.2017.02.008>.
 23. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* [Internet]. 2009;28:3083–107. <https://doi.org/10.1002/sim.3697>.
 24. Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* [Internet]. 1979;7. <https://doi.org/10.1214/aos/1176344552>
 25. Walker E, Nowacki AS. Understanding Equivalence and Noninferiority Testing. *J Gen Intern Med* [Internet]. 2011;26:192–6. <https://doi.org/10.1007/s11606-010-1513-8>.
 26. Piacenza F, Di Rosa M, Soraci L, Montesanto A, Corsonello A, Cherubini A, et al. Interactions between patterns of multimorbidity and functional status among hospitalized older patients: a novel approach using cluster analysis and association rule mining. *J Transl Med* [Internet] BioMed Cent Ltd. 2024;22:669. <https://doi.org/10.1186/s12967-024-05444-9>.
 27. Lekadir K, Feragen A, Fofanah AJ, Frangi AF, Buyx A, Emelie A, et al. FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* [Internet]. BMJ Publishing Group; 2024. <http://arxiv.org/abs/2309.12325>.
 28. Silva GFdosS, Barcellos Filho FN, Wichmann RM, da Silva Junior FC, Chiavegatto Filho ADP. Strategies for detecting and mitigating dataset shift in machine learning for health predictions: A systematic review. *J Biomed Inf* [Internet] Acad Press Inc. 2025;170:104902. <https://doi.org/10.1016/j.jbi.2025.104902>.

29. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The Clinician and Dataset Shift in Artificial Intelligence. *New Engl J Med* [Internet]. 2021;385:283–6. <https://doi.org/10.1056/NEJMc2104626>.
30. Artuso I, Da Rin Della Mora R, Dasso N, Rebuffi C, Ottonello G, Calza S, et al. Do Nursing Patient Classification Systems to Classify Inpatients Impact Outcomes? A Scoping Review. *West J Nurs Res* [Internet]. Volume 47. SAGE Publications Inc.; 2025. pp. 1214–24. <https://doi.org/10.1177/01939459251367529>.
31. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* [Internet] Nat Res. 2022;28:924–33. <https://doi.org/10.1038/s41591-022-01772-9>.
32. Ardito V, Cappellaro G, Compagni A, Petracca F, Preti LM. Implementation of Machine Learning Applications in Health Care Organizations: Protocol for a Systematic Review of Empirical Studies. *JMIR Res Protoc* [Internet]. Volume 12. JMIR Publications Inc.; 2023. p. e47971. <https://doi.org/10.2196/47971>.
33. Óscar Belmonte-Fernández VíctorM, González-Chordá JC, Amengual. Cano-Genovés. gh repo clone giant-uji/VALENF_Groupper_Classifier [Internet]. Nursing Assessment–Derived Risk Profiles and VALENF Instrument–Based Classification in Adult Inpatients: A Two-Stage Machine Learning Study. 2026. https://github.com/giant-uji/VALENF_Groupper_Classifier.git

Figures

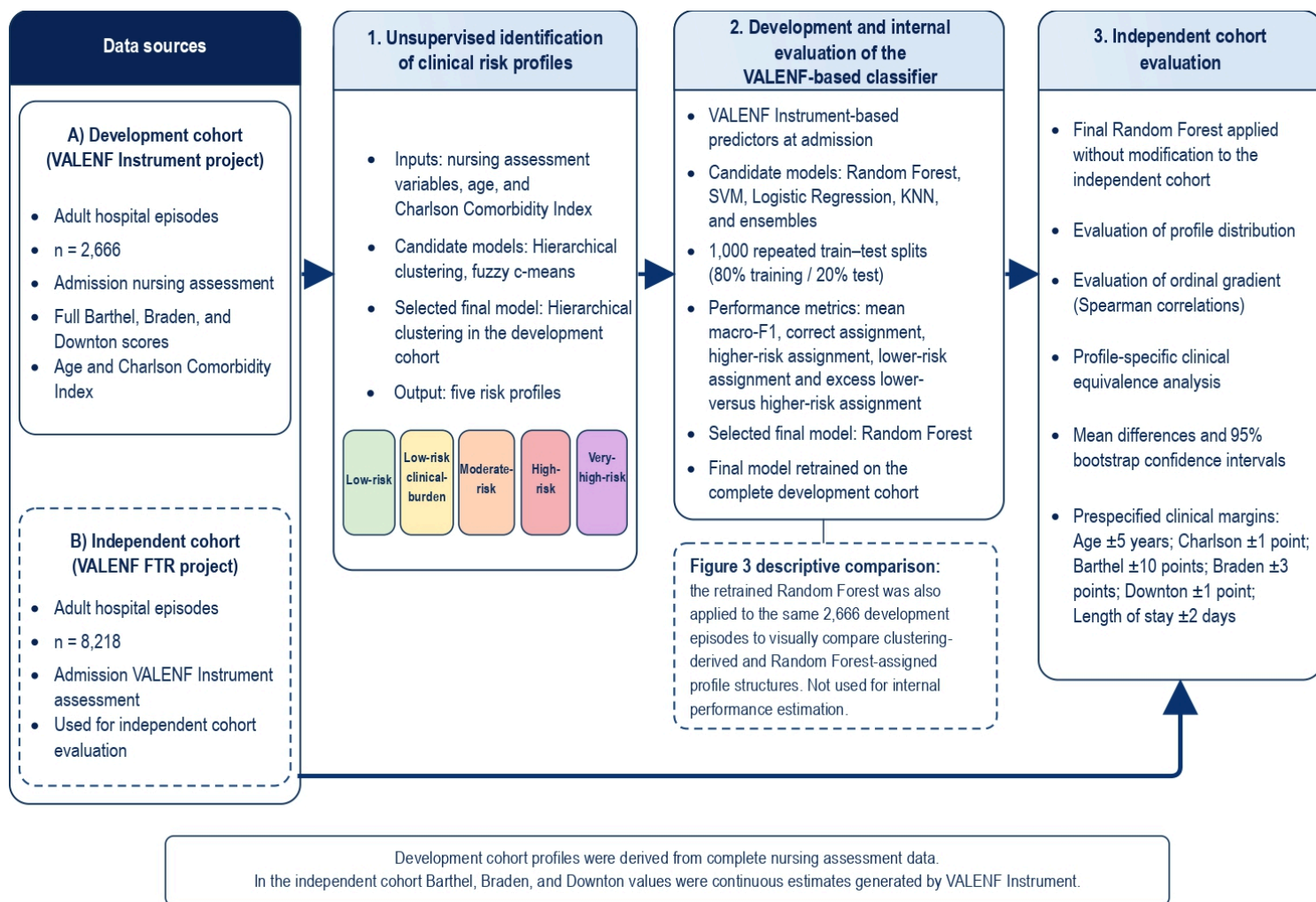


Figure 1

Analytical pipeline of the study.

Clustering-derived reference profile

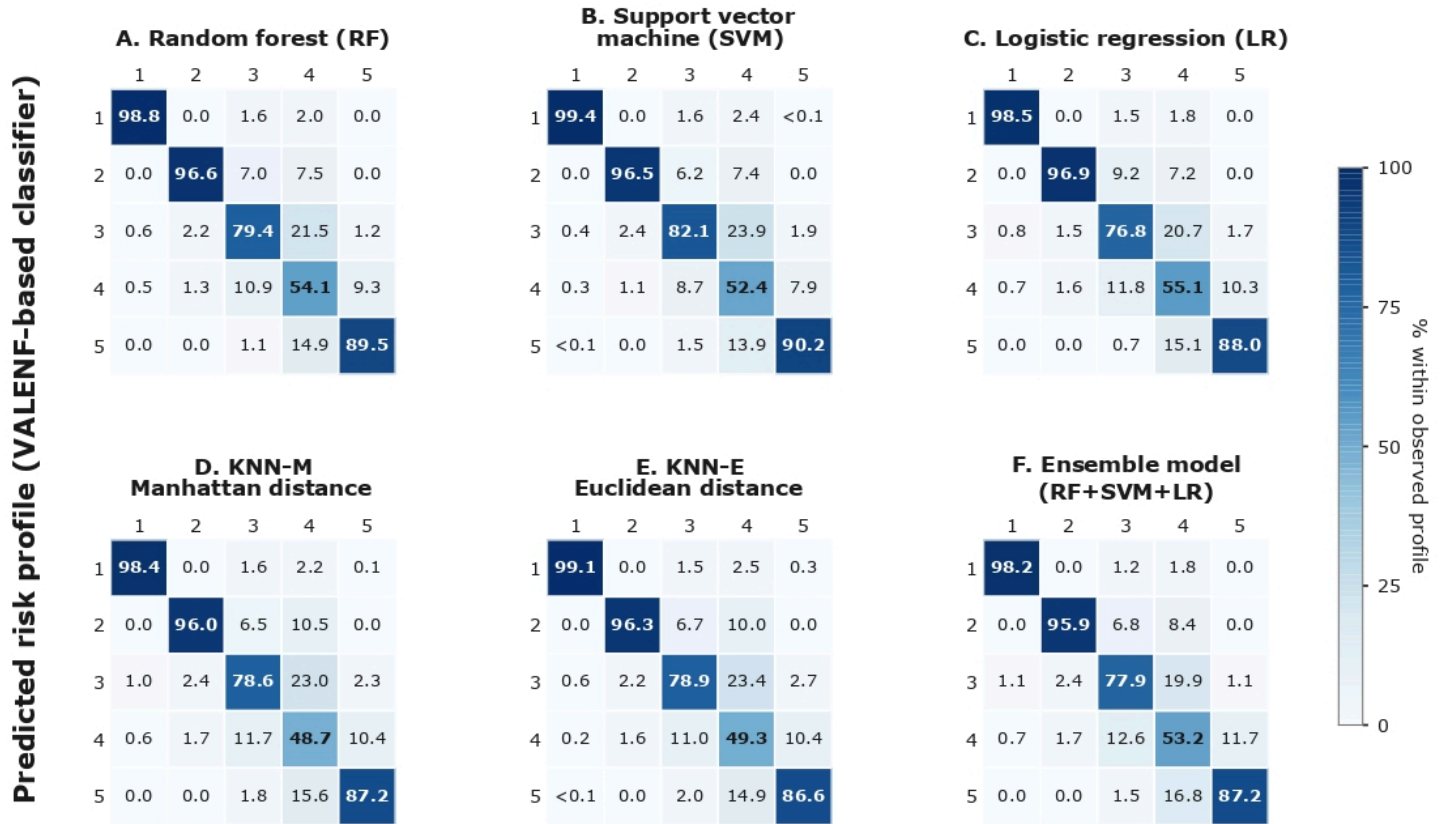


Figure 2

Mean confusion matrices for individual classifiers and the best-performing ensemble across 1,000 train-test splits.

Note: Columns represent the clustering-derived observed profiles and rows represent the classifier-assigned profiles. Cell values are mean column-normalised percentages calculated within each observed profile. Risk profiles were ordered from 1 (lowest risk) to 5 (highest risk). Random Forest was the selected final classifier.

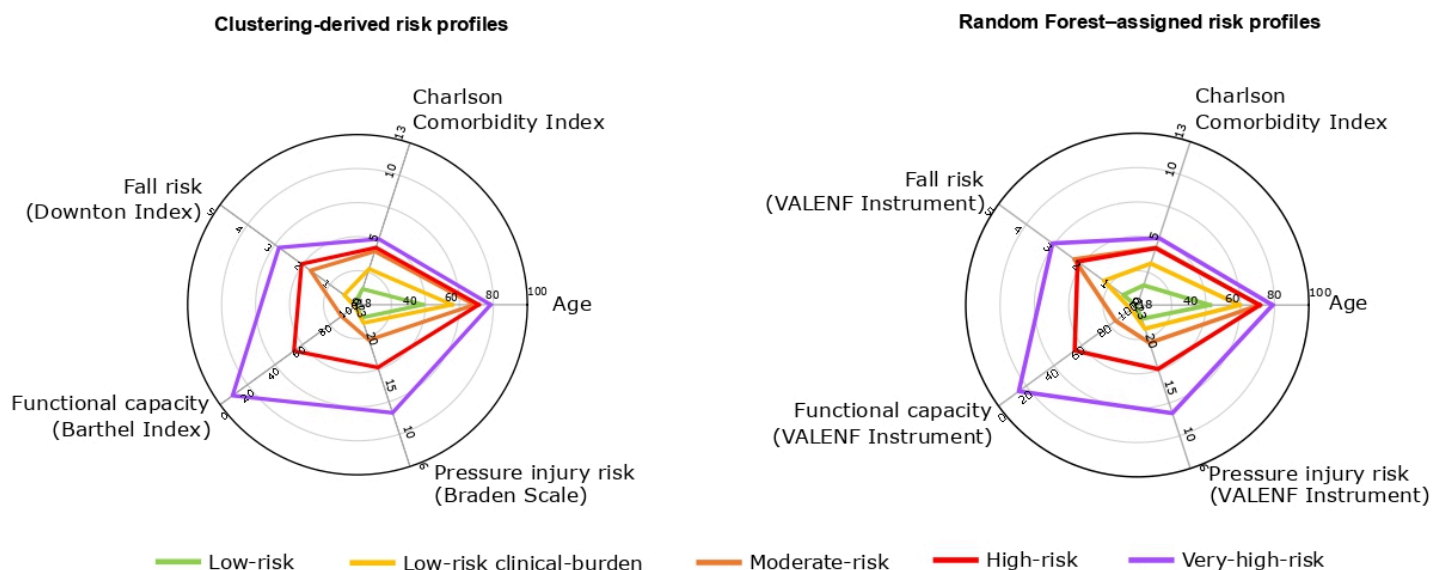


Figure 3

Clinical structure of clustering-derived and Random Forest-assigned risk profiles in the development cohort.

Note: Left: shows profiles derived from the complete nursing assessment using hierarchical clustering. Right: shows profiles assigned using the VALENF-based Random Forest classifier. Variables were rescaled to a common metric for visualisation; functional capacity and pressure injury risk scores were directionally transformed so that higher values consistently represented greater clinical risk.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [5Additionalfile1.xlsx](#)