Improving machine learning algorithm for risk of early pressure injury prediction in admission patients using probability feature aggregation

Shu-Chen Chang[1,2], Shu-Mei Lai[3], Mei-Wen Wu[3], Shou-Chuan Sun[3], Mei-Chu Chen[3] and Chiao-Min Chen[4,5] (iD)

Abstract

Objective: Pressure injuries (PIs) pose a significant concern in hospital care, necessitating early and accurate prediction to mitigate adverse outcomes.

Methods: The proposed approach receives multiple patients records, selects key features of discrete numerical based on their relevance to PIs, and trains a random forest (RF) machine learning (ML) algorithm to build a predictive model. Pairs of significant categorical features with high contributions to the prediction results are grouped, and the PI risk probability for each group is calculated. High-risk group probabilities are then added as new features to the original feature subset, generating a new feature subset to replace the original one, which is then used to retrain the RF model.

Results: The proposed method achieved an accuracy of 83.44%, sensitivity of 84.59%, specificity of 83.42%, and an area under the curve of 0.84.

Conclusion: The ML-based approach, coupled with feature aggregation, enhances predictive performance, aiding clinical teams in understanding crucial features and the model's decision-making process.

Keywords

Artificial intelligence, quality of care, electronic health record, pressure injury, machine learning

Introduction

Pressure injury (PI) was common clinical medical problems while was an evaluation indicator of critical clinical monitoring and the quality of nursing care.[1] PI was an area of injured skin when force was applied on the surface of the skin.[2] If the wound advanced to a deep or infected that could be life-threatening for patients.[3,4] For the intensive care unit and general ward, PI prevalence probability of 16.9–23.8% and 8.5% in the United States, respectively.[5–8] PIs increased the average hospital length of stay by 4–10 days with a total accounting of approximately $10,708 per patient for treatment.[9–11] Obviously, PIs contributed adversely to patient mortality, hospital length of stay, and healthcare costs, thereby considered severe adverse events.[12–14] Most

PIs were reasonably preventable.[15] In 2019,[16] the clinical practice guideline with prevention and treatment of pressure

[1]Department of Healthcare System Operation Center, Changhua Christian Hospital, Changhua, Taiwan
[2]Department of Nursing, Hungkuang University, Taichung, Taiwan
[3]Nursing Department, Changhua Christian Hospital, Changhua, Taiwan
[4]Department of Mathematics, National Changhua University of Education, Changhua, Taiwan
[5]Department of Artificial Intelligence Development Center, Changhua Christian Hospital, Changhua, Taiwan

Corresponding author:
Chiao-Min Chen, Department of Mathematics, National Changhua University of Education, Changhua 500, Taiwan.
Email: cmchen@cc.ncue.edu.tw

ulcers development protocol updated, except that compiling various PIs risk factors, while suggested that clinical guidelines focus on preventive nursing interventions thereafter to identify individuals at risk of developing PIs. Hence, identifying PIs as early as possible was the best way to reduce healthcare resources utilization and advance the patient's safety. Despite accompanying the precise treatment and professional assessment, there still were two challenges with early identification risk of PI. First, existing evaluation of PI risk tools had limited accuracy, including the inability to incorporate diverse types of data, such as unstructured clinical notes and environmental factors. Second, several existing tools rely on manually selected features, which may introduce bias and fail to capture relationships among variables. Moreover, these tools often lack generalizability across different healthcare settings due to patient populations and clinical practice variations, leading to inconsistent performance.[17–20] Subsequently, the tools had not been shown to reduce PIs.[21]

Artificial intelligence (AI) approaches such as random forest (RF),[22] gradient boosting machine,[23] linear regression,[24] support vector machine (SVM),[25] and deep neural network[26] played an important role in disease prediction. Meanwhile, previous studies suggested that the AI approaches were meaningful for predicting of high risk events.[27] Jin et al.[15] proposed an automatic scheme to assess fall risk and PI based on generally machine learning (ML). In this work, the advantage was proving the degree of agreement between the automated risk assessment systems and the scales commonly used by nurses. Song et al.[21] developed a framework for PI risk assessment for in-hospital and non-hospital patients. In the prediction method, the PI risk assessment used three ML models that used lab tests, medical diagnoses, and nursing flowsheet data in the electronic health record (EHR) for building an automatic early assessment tool to facilitate early prevention and treatment. However, the clinical dataset contained a data imbalance problem likely resulting in model overfitting and miss prediction of positive cases. In addition, laboratory test results and medication information were often incomplete or inaccurately recorded, further complicating the data preprocessing and feature extraction processes. To overcome these challenges, this study proposes an effective automated approach that initially employs expert knowledge to define and extract feature variables during the data preparation. Subsequently, statistical algorithms are utilized in feature engineering to enable automated feature selection. To further reduce false predictions and address imbalanced datasets, a proposed feature aggregation method using feature scaling followed by feature interaction is introduced to generate combined features. Following this, ML techniques are applied to develop a predictive model for PI risk. Moreover, by recalibrating the model's internal weights, the effectiveness of these combined features is improved, ultimately improving the model's early prediction performance for PI risk.

Materials and methods

We developed an AI-prediction approach combining ML and expert knowledge to predict generating PI risk of admission patients. In data preparation, expert knowledge was used to define and extract feature variables. The proposed model was composed of feature engineering and predictive model development. First, for feature engineering, the one-hot encoding was performed on feature normalization, and then the recursive feature elimination (RFE) technology[28] was performed on the initial input feature set to remove non-informative features. Then, for the predictive model development, the RF[29] was applied to classify the admission patient as risk of PI, while using feature aggregation to combine the top five of discrete numerical features and 1 of clinically significant feature to improve predictive model performance. Moreover, the 10-fold cross-validation scheme[30] was used for the stability and effectiveness validation in the training phase of the proposed method. Figure 1 illustrates the flowchart of the proposed ML-based method.

Materials

As the dataset, we retrospectively reviewed the EHR of 95,027 adult patients (aged ≥20 years) and without including the PI was determined before admission who admission nursing assessment from August 2018 to June 2020. All personal information was identified at the point of analysis, while the patients were assigned a unique study identification number in the dataset. In addition, each admission of a patient at a different time was considered as a single data point. As a quality control measure, the patients who had hospice care were excluded ($n = 363$). The abnormal value of height and weight ($n = 28$), the record of nursing care which had missing value ($n = 28$), and abnormal assessment data ($n = 11$). Ultimately, total 94,589 patients which included the 1852 patients were defined as the admission PI and 92,737 patients were defined as the control group (Figure 2). This study was approved by the Institutional Review Board of the Changhua Christian Hospital (CCH) and granted a waiver of informed consent. We random split the data into 80% training/validation dataset ($n = 75,672$, positive case = 1482, negative case = 74,190) and 20% testing dataset ($n = 18,917$, positive case = 370, negative case = 18,847) to evaluate the performance of the proposed method. In the training/validation dataset, we applied a 10-fold cross-validation scheme using stratified sampling to ensure that the class distribution remained consistent across each fold. Stratified sampling was employed to maintain the original ratio of positive and negative cases in each fold, thereby preserving the balance of class distribution. This approach allows for a more accurate estimation of the model's performance, especially when dealing with an imbalanced dataset, as it ensures that each fold has a representative sample of both positive and negative cases.
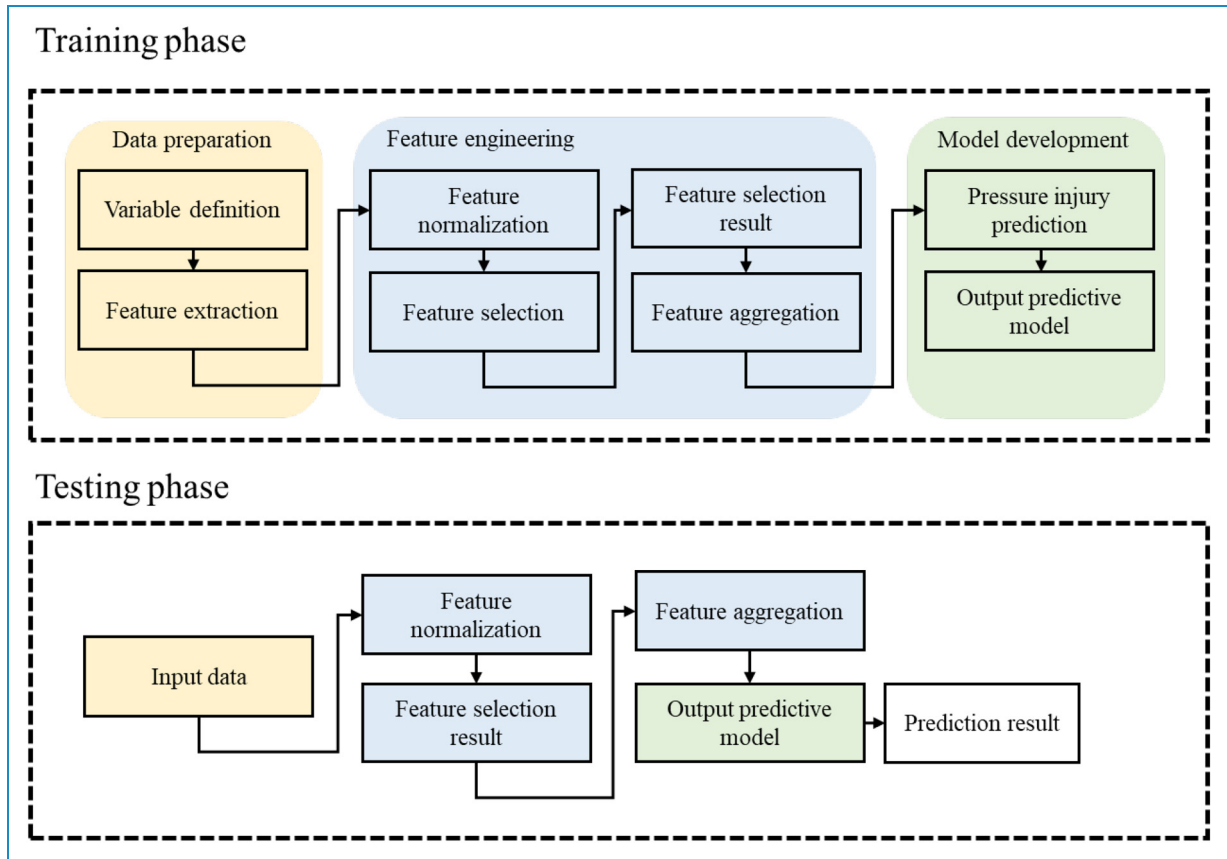
Figure 1. Flowchart of our proposed ML-based method.

Data preparation

Before developing the predictive model, we conducted data preparation, including variable definition and feature extraction. For feature extraction, we performed a literature review to identify relevant features from our defined variables based on previous studies. Clinical experts then reviewed all features obtained from the CCH clinical research database, a compilation of data from CCH EHR systems. We selected 46 features, encompassing demographics, clinical characteristics, nursing assessments, and medical history. The decision to exclude other types of features, such as laboratory results and medication records, was based on several considerations. First, nursing assessments are conducted during the initial admission process, where nursing staff perform a comprehensive evaluation of the patient's condition. These initial assessments are critical as they provide the foundation for patient care, especially in identifying the risk of PIs. The tools used during this process are specifically designed to evaluate PI risk, and the results serve as essential information for nurses working in shifts to ensure continuity of care. Second, laboratory results and medication records are typically generated after admission and thus are not available during the

initial evaluation. Since the proposed method aims to predict PI risk immediately following the first nursing assessment, such features would not align with the intended clinical application. Third, while existing nursing assessment tools provide valuable insights, they are often subjective and lack comprehensiveness. To address this limitation, we incorporated features related to demographics, clinical characteristics, and medical history to develop a more objective and robust predictive framework. This integration ensures that the model not only complements existing nursing assessments but also establishes a standardized approach to PI risk prediction.

Feature engineering

Feature normalization and selection. In total, 46 features were defined, and one-hot encoding was applied to normalize all features into initial feature sets. For feature selection, we employed RFE technology to choose the initial optimal subset of features for predicting PI occurrence risk, resulting in 23 selected features. The RFE algorithm used a proposed predictive model as its estimator. RF was chosen due to its ability to handle both categorical and numerical data
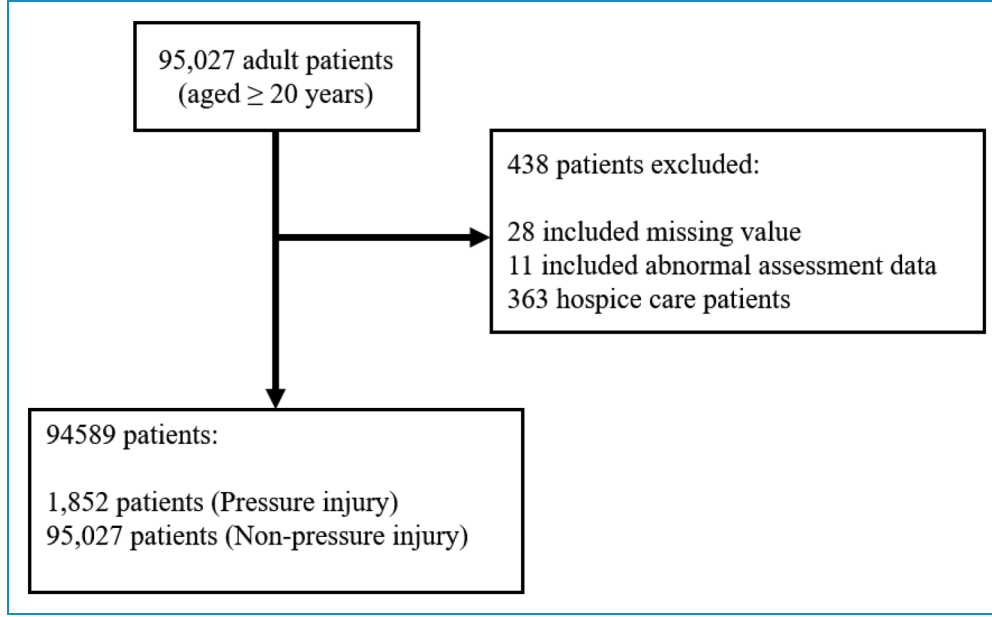
Figure 2. Flowchart of the study population.

effectively and its robustness in assessing feature importance by calculating Gini index.[31] During each iteration of RFE, the RF model ranked features based on their contribution to the model's predictive performance, using sensitivity as the selection criterion, which allowed us to systematically eliminate less relevant features while retaining the most informative ones. This process improved model interpretability and enhanced computational efficiency by reducing dimensionality. After evaluating different subsets of features, we determined that selecting 23 features provided the best balance between predictive power and computational efficiency. Descriptive statistics of these selected features for the training and test sets were presented in Table 1.

Subsequently, the Gini index was applied to rank features based on feature importance. The Gini index calculates node impurity for each feature and sample, representing the impact of each feature on the trained model. For example, consider the feature of PressureInjuryRiskGroup for 0 (low risk), there were 10 cases, with 2 developing PIs and 8 not. The Gini index for 0 is calculated as $0.2 \times 2 + 0.8 \times 2 = 0.68$. For 1 (high risk), there were 20 cases, with 13 developing PIs and 7 not, resulting in a Gini coefficient of $0.65 \times 2 + 0.35 \times 2 = 0.5$. The weighted Gini index for PressureInjuryRiskGroup is then calculated as $(10/30) \times 0.68 + (20/30) \times 0.55 = 0.59$.

Feature aggregation. To improve the accuracy and robustness of the predictive model, the proposed feature aggregation method was developed to extract potential interactions between discrete numerical features and the clinically significant feature PressureInjuryBefore. This feature, which reflects whether a patient had a prior history of PIs before admission, is recognized as a key factor in predicting future PI risks. By examining the relationships between PressureInjuryBefore and other discrete numerical features, the method identifies hidden correlations that are not easily captured through conventional feature selection processes. These aggregated features contribute additional predictive value by incorporating clinically relevant interactions, allowing the model to provide more precise and meaningful predictions while addressing the limitations of traditional approaches.

Discrete numerical features with a Gini index of the top five were prioritized and selected for feature aggregation to improve prediction performance and identify potentially correlated features. For example, a pair of features such as feature BarthelIndexTotalGroup (categorized into 1 to 5) and feature PressureInjuryRiskGroup (categorized into 0, 1) might be selected. After determining key features, pairs of these features were created, and the conditional probabilities of PI risk for each feature pair were calculated. These probabilities were defined as new feature values and added to the feature subset, forming a new feature set.

During the feature aggregation, a set of the top five discrete numerical features $\{X_1, X_2, \ldots, X_m\}$, each containing n samples, had been selected. Subsequently, the mean μ_j and standard deviation σ_j had been computed for each feature X_j across all samples, as follows:

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} X_j^{(i)}, \quad \sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_j^{(i)} - \mu_j)^2}$$

Table 1. Proposed features with RFE feature selection in overall cohort and split datasets.

Feature	Patients Data (n=94,589)		Split Data (n=94,589)				p
			Training Data (n=75,672)		Testing Data (n=18,917)		
	Non-PI (n=92,737)	PI (n=1852)	Non-PI (n=74,190)	PI (n=1482)	Non-PI (n=18,547)	PI (n=370)	
Continuous numerical (mean ± SD)							
Age	59.34 ± 17.84	71.46 ± 15.22	58.78 ± 17.42	71.34 ± 15.11	61.55 ± 19.24	71.94 ± 15.64	<.001
Barthel Index Total Number	79.43 ± 30.91	42.90 ± 37.94	80.99 ± 29.64	43.74 ± 38.34	73.22 ± 34.86	39.55 ± 36.17	<.001
Braden Scale	20.83 ± 3.33	16.76 ± 4.48	21.02 ± 3.16	16.83 ± 4.49	20.07 ± 3.83	16.46 ± 4.44	<.001
BMI	24.35 ± 4.66	22.88 ± 4.89	25.72 ± 4.15	23.66 ± 4.79	18.88 ± 1.36	19.76 ± 3.94	<.001
Binary (%)							
Sex (female)	45,695 (49.3)	711 (38.4)	36,617 (49.4)	559 (37.7)	9078 (48.9)	152 (41.1)	<.001
Past history of assessment (yes)	25,679 (27.7)	241 (13.0)	20,706 (27.9)	195 (13.2)	4973 (26.8)	46 (12.4)	<.001
Hypertension (yes)	35,272 (38.0)	977 (52.8)	29,945 (40.4)	800 (54.0)	5327 (28.7)	177 (47.8)	<.001
Diabetes (yes)	20,630 (22.2)	650 (35.1)	17,625 (23.8)	531 (35.8)	3005 (16.2)	119 (32.2)	<.001
Heart (yes)	12,037 (13.0)	443 (23.9)	9886 (13.3)	366 (24.7)	2151 (11.6)	77 (20.8)	<.001
Pressure injury risk group (yes)	8985 (9.7)	769 (41.5)	6176 (8.3)	618 (41.7)	2809 (15.1)	151 (40.8)	<.001
Pressure injury before (yes)	9535 (10.3)	1668 (90.1)	6474 (8.7)	1330 (89.7)	3061 (16.5)	338 (91.4)	<.001
Discrete numerical (%)							
DiagnosticType							<.001
ICD-10:A00-B99	1410 (1.5)	65 (3.5)	1042 (1.4)	49 (3.3)	368 (2.0)	16 (4.3)	
ICD-10:C00-D49	27,552 (29.7)	387 (20.9)	20,865 (28.1)	306 (20.6)	6687 (36.1)	81 (21.9)	

(continued)

Table 1. Continued.

Feature	Patients Data (n=94,589)		Split Data (n=94,589)				p
			Training Data (n=75,672)		Testing Data (n=18,917)		
	Non-PI (n=92,737)	PI (n=1852)	Non-PI (n=74,190)	PI (n=1482)	Non-PI (n=18,547)	PI (n=370)	
ICD-10:E00-E90	1921 (2.1)	29 (1.6)	1532 (2.1)	23 (1.6)	389 (2.1)	6 (1.6)	
ICD-10:F00-F99	839 (0.9)	9 (0.5)	676 (0.9)	6 (0.4)	163 (0.9)	3 (0.8)	
ICD-10:G00-G99	1468 (1.6)	27 (1.5)	1206 (1.6)	24 (1.6)	262 (1.4)	3 (0.8)	
ICD-10:H00-H59	553 (0.6)	1 (0.1)	477 (0.6)	1 (0.1)	76 (0.4)	0 (0.0)	
ICD-10:I00-I99	487 (0.5)	0 (0.0)	431 (0.6)	0 (0.0)	56 (0.3)	0 (0.0)	
ICD-10:J00-J99	8195 (8.8)	318 (17.2)	7124 (9.6)	266 (17.9)	1071 (5.8)	52 (14.1)	
ICD-10:K00-K95	8903 (9.6)	437 (23.6)	6429 (8.7)	349 (23.5)	2474 (13.3)	88 (23.8)	
ICD-10:L00-L99	9375 (10.1)	107 (5.8)	7493 (10.1)	81 (5.5)	1882 (10.1)	26 (7.0)	
ICD-10:M00-M99	1404 (1.5)	17 (0.9)	1209 (1.6)	11 (0.7)	195 (1.1)	6 (1.6)	
ICD-10:N00-N99	4201 (4.5)	58 (3.1)	3621 (4.9)	50 (3.4)	580 (3.1)	8 (2.2)	
ICD-10:O00-O99	7200 (7.8)	89 (4.8)	5809 (7.8)	67 (4.5)	1391 (7.5)	22 (5.9)	
ICD-10:P00-P96	4505 (4.9)	3 (0.2)	4323 (5.8)	2 (0.1)	182 (1.0)	1 (0.3)	
ICD-10:Q00-Q99	6 (0.0)	0 (0.0)	5 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	
ICD-10:R00-R99	133 (0.1)	4 (0.2)	102 (0.1)	4 (0.3)	31 (0.2)	0 (0.0)	
ICD-10:S00-T98	5204 (5.6)	121 (6.5)	4062 (5.5)	89 (6.0)	1142 (6.2)	32 (8.6)	
ICD-10:V01-Y98	7510 (8.1)	166 (9.0)	6133 (8.3)	141 (9.5)	1377 (7.4)	25 (6.8)	
ICD-10:Z00-Z99	14 (0.0)	0 (0.0)	11 (0.0)	0 (0.0)	3 (0.0)	0 (0.0)	

(continued)

Table 1. Continued.

Feature	Patients Data (n = 94,589)		Split Data (n = 94,589)				p
			Training Data (n = 75,672)		Testing Data (n = 18,917)		
	Non-PI (n = 92,737)	PI (n = 1852)	Non-PI (n = 74,190)	PI (n = 1482)	Non-PI (n = 18,547)	PI (n = 370)	
ICD-10:U00–U99	1857 (2.0)	14 (0.8)	1640 (2.2)	13 (0.9)	217 (1.2)	1 (0.3)	
Barthel Index Total Group (completely dependent)	8895 (9.6)	750 (40.5)	6164 (8.3)	594 (40.1)	2731 (14.7)	156 (42.2)	<.001
Sensory perception (completely impaired)	1870 (2.0)	202 (10.9)	1325 (1.8)	159 (10.7)	545 (2.9)	43 (11.6)	<.001
Moisture (severely moist)	542 (0.6)	65 (3.5)	363 (0.5)	49 (3.3)	179 (1.0)	16 (4.3)	<.001
Activity (bedridden, confined to bed)	8870 (9.6)	699 (37.7)	6562 (8.8)	572 (38.6)	2308 (12.4)	127 (34.3)	<.001
Mobility (completely immobile)	3245 (3.5)	332 (17.9)	2186 (2.9)	262 (17.7)	1059 (5.7)	70 (18.9)	<.001
Nutrition (poor nutrition)	3681 (4.0)	178 (9.6)	2813 (3.8)	135 (9.1)	868 (4.7)	43 (11.6)	<.001
Friction/shear (significant friction/shear)	2570 (2.8)	222 (12.0)	1589 (2.1)	164 (11.1)	981 (5.3)	58 (15.7)	<.001
Nutritional total number (high risk groups ≥ 4)	4123 (4.45)	180 (9.71)	2400 (3.23)	130 (8.78)	1723 (9.29)	50 (13.51)	<.001
Smoke (yes)	12,429 (13.4)	227 (12.3)	9770 (13.2)	187 (12.6)	2659 (14.3)	40 (10.8)	<.001
BMI group (underweight)	7738 (8.3)	349 (18.8)	1190 (1.6)	274 (18.5)	6548 (35.3)	75 (20.3)	<.001
Admission nursing station type							<.001
Intensive care nursing unit	7393 (8.0)	476 (25.7)	5958 (8.0)	406 (27.4)	1435 (7.7)	70 (18.9)	
Chronic care nursing unit	1147 (1.2)	33 (1.8)	792 (1.1)	23 (1.6)	355 (1.9)	10 (2.7)	
Internal medicine nursing unit	37,371 (40.3)	811 (43.8)	28,678 (38.7)	619 (41.8)	8693 (46.9)	192 (51.9)	
Surgical nursing unit	39,943 (43.1)	505 (27.3)	32,514 (43.8)	411 (27.7)	7429 (40.1)	94 (25.4)	
Respiratory care center	192 (0.2)	18 (1.0)	128 (0.2)	16 (1.1)	64 (0.3)	2 (0.5)	

(continued)

Table 1. Continued.

Feature	Patients Data (n=94,589)		Split Data (n=94,589)				p
			Training Data (n=75,672)		Testing Data (n=18,917)		
	Non-PI (n=92,737)	PI (n=1852)	Non-PI (n=74,190)	PI (n=1482)	Non-PI (n=18,547)	PI (n=370)	
Other nursing units	6691 (7.2)	9 (0.5)	6120 (8.2)	7 (0.5)	571 (3.1)	2 (0.5)	

Note. RFE: recursive feature elimination; PI: pressure injury; BMI: body mass index.

After μ_j and σ_j had been derived, each feature was standardized to obtain $\tilde{X}_j$. The standardized value for each sample i in feature j was calculated as:

$$\tilde{X}_j^{(i)} = \frac{X_j^{(i)} - \mu_j}{\sigma_j}, \quad j = 1, 2, \ldots, m, \quad i = 1, 2, \ldots, n.$$

After the standardization had been completed, a set S of selected feature pairs (a, b) had been defined, where a and b indicate indices of standardized features $\tilde{X}_a$ and $\tilde{X}_b$. Each pair $(a, b) \in S$ represented a combination of features intended for interaction. To generate the interaction features, a function f had been introduced, which mapped the standardized feature set $\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_m\}$ to a set of interaction features $\{\tilde{X}_{a\_b} \mid (a, b) \in S\}$. Within each pair $(a, b) \in S$, the interaction feature $\tilde{X}_{a\_b}$ had been defined as the element-wise product of $\tilde{X}_a$ and $\tilde{X}_b$: $\tilde{X}_{a\_b}^{(i)} = \tilde{X}_a^{(i)} \times \tilde{X}_b^{(i)}, \quad (a, b) \in S, \quad i = 1, 2, \ldots, n.$

By applying feature aggregation, the feature space was enriched with interaction terms that captured nonlinear relationships among the standardized features, while improving the model's predictive performance. For example, consider two discrete numerical features, X_1 and X_2, each containing three samples ($n=3$), $X_1 = \{10, 12, 14\}$ and $X_2 = \{30, 28, 32\}$. First, for X_1, the mean (μ_1) was $(10 + 12 + 14)/3 = 12$, and the standard deviation (σ_1) was calculated as $\sqrt{[((10 - 12)^2 + (12 - 12)^2 + (14 - 12)^2)/3]} = \sqrt{(8/3)} \approx 1.6329$. For X_2, the mean (μ_2) was $(30 + 28 + 32)/3 = 30$, and the standard deviation (σ_2) was $\sqrt{[((30 - 30)^2 + (28 - 30)^2 + (32 - 30)^2)/3]} = \sqrt{(8/3)} \approx 1.6329$. Then, standardize each feature using the computed mean and standard deviation.

Algorithm 1. Feature selection and feature aggregation

Input: RFE features (*RFEFeature*) with binary class (0/1), X_{clin} is PressureInjuryBefore

Output: A modified feature dataset with newly added interaction-based features.

1: **Load** *RFEFeature*
2: **For** each feature X_j, $j=1,\ldots,m$:
3: **Compute** GINI index (X_j)
4: **Sort** features by GiniIndex in descending order
5: **Select** top 5 discrete numerical features $\{X_{a1},\ldots,X_{a5}\}$
6: **Combine** these 5 features with $\{X_{a1},\ldots,X_{a5}, X_{\text{clin}}\} \leftarrow X_{\text{clin}}$
7: **For** each selected feature $X_{a1},\ldots,X_{a5}, X_{\text{clin}}$
8: **Compute** mean μ_j and std σ_j for each of the 6 selected features
9: **Standardize** each feature
8: **For** each X_j and sample i: $X_j^{\sim (i)} = (X_j^{\sim (i)} - \mu_j)/\sigma_j$
10: **Define** set S of all feature pairs (a,b) from the 6 standardized features
11: **For** each pair $(a,b) \in S$:
12: **For** each sample i: $X_{a\_b}^{\sim (i)} = X_a^{\sim (i)} \times X_b^{\sim (i)}$
13: **Append** all $X_{a\_b}^{\sim}$ to the modified feature dataset

Table 2. Fifteen features generated by the feature aggregation of the proposed method.

Barthel Index Total Group and Pressure Injury Before	Barthel Index Total Group and BMI Group	Barthel Index Total Group and Diagnostic Type
Barthel Index Total Group and Activity	Barthel Index Total Group and Mobility	Pressure Injury Before and BMI Group
Pressure Injury Before and Diagnostic Type	Pressure Injury Before and Activity	Pressure Injury Before and Mobility
Diagnostic Type and Activity	Diagnostic Type and Mobility	Diagnostic Type and BMI Group
BMIGroup and Activity	BMI Group and Mobility	Mobility and Activity

For $\tilde{X}_1(1) = (10-12)/1.6329 \approx -1.2247$. For $\tilde{X}_2(1) = (30-30)/1.6329 = 0$. Finally, generate an interaction feature by multiplying the standardized values of X_1 and X_2 element-wise, producing the new interaction feature $\tilde{X}_{1\_2}$. For $\tilde{X}_{1\_2}(1) = \tilde{X}_1(1) \times \tilde{X}_2(1) = (-1.2247) \times 0 = 0$. Algorithm 1 presents details of the feature aggregation and Table 2 shows the 15 features generated by the feature aggregation of the proposed method.

Model development

In the development of the predictive model, the proposed model was built upon the RF ML algorithm. RF is an ensemble learning method for classification that constructs numerous decision trees. It addresses overfitting to the training dataset and alleviates the issue of data imbalance. To mitigate data imbalance, class weights were employed by assigning different weights to both the majority and minority classes in the proposed model. The formula for calculating $weight_j$ is as follows: $weight_j = n\_samples/(n\_classes \times n\_samples_j)$, where the j is the signified class, $weight_j$ is the weight for each class, $n\_samples$ is the total number of samples, $n\_classes$ is the total number of classes, and $n\_samples_j$ is the total number of the respective class. To optimize the performance of the RF model and reduce issues such as underfitting or overfitting, hyperparameters were adjusted using exhaustive grid search algorithm. In this approach, a specific range of parameter values was manually defined based on prior experience, and step sizes were set to iteratively adjust parameters. For instance, the $n\_estimators$ parameter started at $n\_estimators = 40$, and with a step size of 2, the next value was set to $n\_estimators = 40+2$ until the entire range was

Table 3. Parameter settings, step sizes, and optimal values for the proposed method.

Parameters	Setting Range	Step Size	Optimal Value
n_estimators	40–70	2	60
max_depth	0–30	2	13
min_samples_split	100–150	2	120
min_samples_leaf	10–30	2	20
oob_score	–	–	True
random_state	–	–	10

evaluated. The grid search algorithm computed the model performance for all parameter combinations and identified the hyperparameter settings that achieved the highest accuracy on the testing dataset. The best hyperparameter settings obtained in this study are listed in Table 3. The best hyperparameters that included *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, *oob_score*, and *random_state* were set at 60, 13, 120, 20, 7, True, and 10, respectively.

Comparison

In this study, performance in total, risk of PI was performed by the proposed predictive model and four well-known ML algorithms to estimate the model capability. Additionally, ensemble learning methods were included in the comparison. Four well-known ML algorithms, the RF, light gradient boosting (LGBM), eXtreme gradient boosting (XGBoost), and SVM, were executed for risk prediction after mode training. All parameters and settings were the same as the default for each algorithm. Moreover, each well-known ML algorithm trained with synthetic minority oversampling technique (SMOTE) and SMOTE edited nearest neighbor (SMOTEENN) were also added to our model comparison. All ML models used in this study were trained and tested with the same training, validation, and test sets.

Statistical analysis

There were 10 performance indices, including accuracy (ACC), sensitivity (SEN), specificity (SPEC), precision (macro average), recall (macro average), positive-predictive value (PPV), negative-predictive value (NPV), Matthews correlation coefficient (MCC), *F*1-score (macro average), and the area under the curve (AUC). In addition, the receiver operating characteristic (ROC) curve was used to compare the trade-offs between sensitivity and specificity

by applying ROCKIT software (C. Metz; University of Chicago, Chicago, IL, USA). All performance indices also computed and compared based on the chi-square test and Kolmogorov–Smirnov test whether as normal distribution and 95% confidence interval for odds ratio. Besides, the *t*-test was used for normal distribution; the Mann–Whitney *U*-test was applied for non-normal distribution. The *p*-value was considered to indicate a statistically significant difference if less than 0.05.

Results

All our experiments were conducted using the Python programming language with Python open libraries such as pandas, scikit-learn, SHAP, lightGBM, SVC, etc. We used a machine with the Microsoft Windows 11 operating system, Intel Core i5-9600 3.10 GHz processor, and 32 GB of random access memory with 2400 MHz.

The confusion matrix results are presented in Table 4, where the full names of true negative (TN), true positive (TP), false negative (FN), and false positive (FP), respectively. The estimated experimental results are shown in Table 5, with ACC, SEN, SPEC, Precision, Recall, PPV, NPV, MCC, *F*1-score, and AUC being 83.44, 84.59, 83.42, 54.44, 84.01, 9.24, 99.63, 24.56, 53.73, and 0.838, respectively. The results demonstrate the proposed method was an efficient method that can improve a relatively accurate result and reduce miss predictions in positive cases. Moreover, the ROC curves of all models are displayed in Figure 3(a) to (c). On the other hand, compared to the proposed method without (W/O) feature aggregation, the performance of the proposed method was better. The ROC curve of the proposed model and W/O feature aggregation performances are illustrated in Figure 3(d). The odds ratios between the other models and the proposed method performance in the predicting risk of PI are listed in Table 6, and the *p*-values are listed in Table 7.

Figure 4 exhibits the results of two pair features ranking with Gini index by using the proposed method and RFE feature selection stage. Figure 4(a) shows the result of the RFE feature selection stage and Figure 4(b) shows the result by executing the proposed method. To demonstrate the effectiveness and interpretability in the clinical relationship of significant features, the SHapley Additive exPlanations (SHAP) algorithm is used in revealing each feature contribution to an individual prediction. All features were ranked depending on their importance for the classification in decreasing order.[32,33] Figure 5 provides beeswarm plot to exhibit an overview of the effects of individual features on the prediction of PI risk.

All the features on the left side of the beeswarm plot indicate to exert a negative effect on PI; on the other hand, features on the right side of the beeswarm plot indicate to exert a positive effect on PI. In the beeswarm plot, all dots represent the SHAP values of each feature for all individual patients, whereas the colors range from blue (low feature value) to red (high feature value). The blue dots are negatively

Table 4. TN, TP, FN, and FP of the proposed method and other different models.

Models	TN	TP	FN	FP
Classical ML-method				
RF	12,549	151	219	5998
XGBoost	4851	198	172	13,696
SVM	18,545	2	368	2
LGBM	10,237	199	171	8310
Ensemble learning (all classical ML-method)	11,408	199	171	7139
SMOTE				
RF	1387	232	138	17,160
XGBoost	10	268	102	18,537
SVM	15,893	254	116	2654
LGBM	246	261	109	18,301
Ensemble learning (all SMOTE method)	193	264	106	18,354
SMOTEENN				
RF	1016	260	110	17,531
XGBoost	3	255	115	18,544
SVM	12,534	161	209	6013
LGBM	121	278	92	18,426
Ensemble learning (all SMOTEENN method)	141	262	108	18,406
Proposed				
W/O data aggregation	7377	293	77	11,170
Proposed method	15,472	313	57	3075

Note. TN: true negative; TP: true positive; FN: false negative; FP: false positive; ML: machine learning; RF: random forest; SVM: support vector machine; LGBM: light gradient boosting; XGBoost: eXtreme gradient boosting; SMOTE: synthetic minority oversampling technique; SMOTEENN: SMOTE edited nearest neighbor.

correlated with PI, whereas red dots are positively correlated with PI for all features on the right side. Figure 5(a) shows the result of the RFE feature selection stage and Figure 5(b) shows the result by executing the proposed method.

Table 5. Performance of the proposed method and other different models.

Models	ACC (%)	SEN (%)	SPEC (%)	Precision (Macro average)	Recall (Macro Average)	PPV (%)	NPV (%)	MCC (%)	F1-Score (Macro Average)	AUC
Classical ML-method										
RF	67.14	40.81	67.66	50.37	54.24	2.46	98.28	2.50	42.39	0.5116
XGBoost	26.69	53.51	26.16	49.00	39.83	1.43	96.58	−6.38	21.97	0.4849
SVM	98.04	0.54	99.99	74.03	50.26	50.00	98.05	5.05	50.04	0.5628
LGBM	55.17	53.78	55.19	50.35	54.49	2.34	98.36	2.50	37.60	0.5778
Ensemble learning (all classical ML-method)	61.36	53.78	61.51	50.62	57.65	2.71	98.52	4.35	40.45	0.4832
SMOTE										
RF	8.56	62.70	7.48	46.14	35.09	1.33	90.95	−15	8.22	0.2469
XGBoost	1.47	72.43	0.05	5.18	36.24	1.43	8.93	−50	1.45	0.2577
SVM	85.36	68.65	85.69	54.00	77.17	8.73	99.28	21	53.74	0.8340
LGBM	2.68	70.54	1.33	35.35	35.93	1.41	69.30	−29	2.68	0.2633
Ensemble learning (all SMOTE method)	2.42	71.35	1.04	32.98	36.20	1.42	64.55	−31	2.41	0.5123
SMOTEENN										
RF	6.75	70.27	5.48	45.85	37.87	1.461413	90.23091	−14	6.60	0.272
XGBoost	1.36	68.92	0.02	1.95	34.47	1.356455	2.542373	−55	1.35	0.230
SVM	67.11	43.51	67.58	50.48	55.55	2.60771	98.35988	3	42.52	0.604
LGBM	2.11	75.14	0.65	29.15	37.89	1.486313	56.80751	−32	2.10	0.241
Ensemble learning (all SMOTEENN method)	2.13	70.81	0.76	29.01	35.79	1.403471	56.62651	−35	2.13	0.342
Proposed										
W/O data aggregation	40.55	79.19	39.77	50.76	59.48	2.56	98.97	5.37	30.85	0.708
Proposed method	83.44	84.59	83.42	54.44	84.01	9.24	99.63	24.56	53.73	0.838

Note. ACC: accuracy; SEN: sensitivity; SPEC: specificity; PPV: positive-predictive value; NPV: negative-predictive value; MCC: Matthews correlation coefficient; AUC: area under the curve; ML: machine learning; RF: random forest; SVM: support vector machine; LGBM: light gradient boosting; XGBoost: eXtreme gradient boosting; SMOTE: synthetic minority oversampling technique; SMOTEENN: SMOTE edited nearest neighbor.
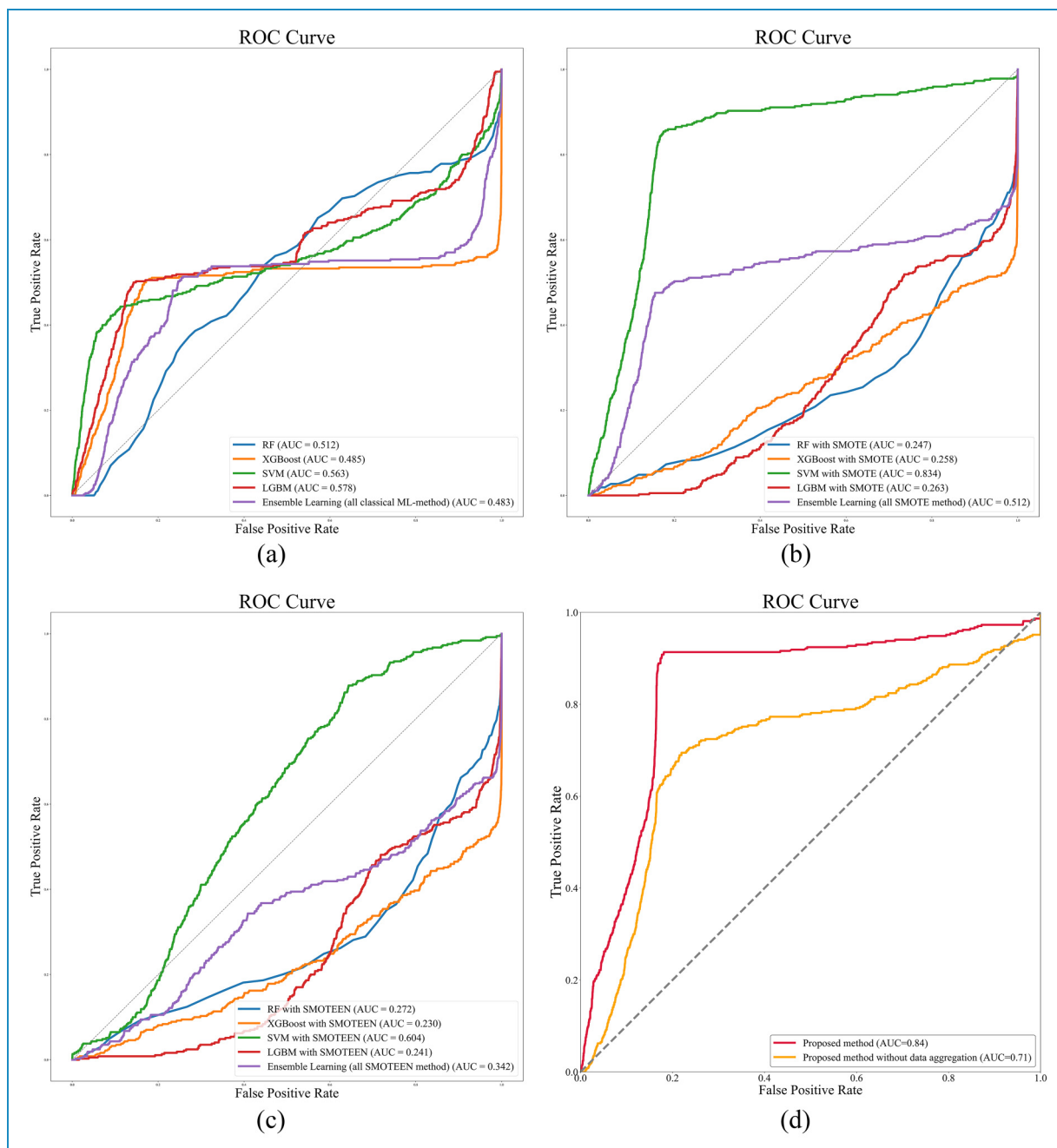
Figure 3. Receiver operating characteristic (ROC) curve of the proposed method with other different machine learning methods.

Discussion

Early detection of hospitalized patients at risk for PI was an essential step that led to future earlier discussions regarding patient goals of nursing care and clinical treatments procedures, accordingly improving clinical outcomes.[34] Therefore, a reliable approach that can early identify and alarm risk of PI will effectively advance nursing care quality and control healthcare resources utilization for clinical units.[35] The prediction model that was developed based on the ML architecture has been proven to be robust in clinical medicine analysis.[36] However, for the prediction of clinical state, the reduction of false prediction was critical factors affecting the model performance. Hence, we developed an effective risk assessment method that combined ML-based and feature aggregation for predicting occurrence PI in admission patients. Our results indicated that the proposed method exhibit high discrimination performance, with ACC, SEN, SPEC, Precision, Recall, PPV, NPV, MCC, F1-score, and AUC being 83.44, 84.59,

Table 6. Comparison odd ratio of between other method and the proposed method.

Models	ACC	SEN	SPEC	PPV	NPV
Classical ML-method					
RF	2.467 (2.349–2.591)	7.964 (5.611–11.304)	2.405 (2.289–2.527)	4.043 (3.314–4.934)	4.737 (3.536–6.346)
XGBoost	13.843 (13.167–14.554)	4.770 (3.367–6.759)	14.206 (13.504–14.944)	7.041 (5.868–8.448)	9.624 (7.121–13.008)
SVM	0.101 (0.090–0.112)	1010.386 (244.697–4172.009)	0.001 (0.000–0.002)	0.102 (0.014–0.725)	5.386 (4.071–7.125)
LGBM	4.096 (3.904–4.297)	4.719 (3.330–6.686)	4.084 (3.892–4.287)	4.251 (3.542–5.101)	4.534 (3.356–6.125)
Ensemble learning (all classical ML-method)	3.174 (3.025–3.331)	4.719 (3.330–6.686)	3.149 (2.999–3.306)	3.652 (3.042–4.383)	4.069 (3.012–5.496)
SMOTE					
RF	53.848 (50.522–57.393)	3.266 (2.297–4.645)	62.250 (58.215–66.565)	7.529 (6.326–8.961)	27.007 (19.740–36.949)
XGBoost	337.910 (298.360–382.702)	2.090 (1.454–3.004)	9326.974 (5011.476–17,358.649)	7.041 (5.955–8.325)	2768.674 (1375.406–5573.303)
SVM	0.865 (0.818–0.914)	2.508 (1.754–3.586)	0.840 (0.794–0.889)	1.064 (0.894–1.265)	1.981 (1.442–2.722)
LGBM	183.007 (166.221–201.489)	2.293 (1.600–3.287)	374.318 (328.155–426.975)	7.137 (6.029–8.449)	120.272 (85.243–169.694)
Ensemble learning (all SMOTE method)	203.582 (184.131–225.087)	2.205 (1.536–3.164)	478.492 (413.076–554.268)	7.077 (5.981–8.373)	149.080 (104.862–211.944)
SMOTEENN					
RF	69.678 (65.062–74.621)	2.323 (1.621–3.329)	86.819 (80.614–93.501)	6.863 (5.797–8.126)	29.388 (21.210–40.719)
XGBoost	364.495 (320.475–414.561)	2.476 (1.731–3.542)	31,101.655 (10,023.191–96,507.484)	7.402 (6.247–8.771)	10,405.146 (3211.997–33,707.091)
SVM	2.470 (2.352–2.594)	7.128 (5.028–10.107)	2.414 (2.297–2.536)	3.802 (3.128–4.620)	4.526 (3.374–6.072)
LGBM	233.907 (210.313–260.148)	1.817 (1.258–2.625)	766.209 (638.134–919.988)	6.747 (5.715–7.965)	206.383 (141.746–300.496)
	231.536 (208.275–257.395)	2.264 (1.579–3.246)	656.813 (554.046–778.642)	7.151 (6.042–8.463)	207.910 (144.881–298.360)

Table 6. Continued.

Models	ACC	SEN	SPEC	PPV	NPV
Ensemble learning (all SMOTEENN method)					
Proposed					
W/O data aggregation	7.390 (7.043–7.754)	1.443 (0.989–2.105)	7.619 (7.257–7.998)	3.880 (3.293–4.573)	2.833 (2.009–3.995)

Note. ACC: accuracy; SEN: sensitivity; SPEC: specificity; PPV: positive-predictive value; NPV: negative-predictive value; MCC: Matthews correlation coefficient; AUC: area under the curve; ML: machine learning; RF: random forest; SVM: support vector machine; LGBM: light gradient boosting; XGBoost: eXtreme gradient boosting; SMOTE: synthetic minority oversampling technique; SMOTEENN: SMOTE edited nearest neighbor.

83.42, 54.44, 84.01, 9.24, 99.63, 24.56, 53.73, and 0.838, respectively. This indicates that the proposed ML method has the potential for clinical implementation and prediction of PI risk in admission patients. Furthermore, such a method is less complex to develop and is simply to be implanted in the clinical units. Generally, most ML methods had inherent defects and were rarely applied in the prediction of imbalanced datasets. Besides, it could severely interfere with the model performance to result in overfitting.[37,38] In previous research, the sampling method was adapted for solving modeling within an imbalanced dataset.[39,40] However, the sampling method could lead to a rise in miss predictions. Hence, the proposed method used class weights to reduce model overfitting. The result indicated the RF with class weights algorithm performed total performance better than other comparison methods (Table 5). In addition, the results demonstrated that our proposed approach, integrating feature aggregation with expert-defined features and automated selection, achieved a more balanced predictive performance compared to conventional methods. Although the method exhibited improved accuracy, sensitivity, and specificity, its PPV remained relatively modest, indicating a need for further refinement to reduce false positives. In Table 5, the proposed method demonstrated a relatively low PPV (9.24%) and a high NPV (99.63%). A low PPV indicates that a number of predicted high-risk patients may be false positives, creating additional demands on clinical resources and potentially causing unnecessary interventions. In contrast, a high NPV suggests that most patients classified as low risk are indeed unlikely to develop PIs, helping to avoid missed diagnoses. In clinical decision-making, a priority to reduce false negatives or missed cases enhances the importance of a high NPV. However, when PPV is low, the frequency of false alarms may lead to resource inefficiencies or undue anxiety among patients. Under those circumstances, adjusting the model parameters or employing a secondary evaluation step could mitigate the impact of false positives.

Nevertheless, this improved balance in performance suggests that the proposed strategy holds promise for supporting clinicians in the early identification and management of high-risk patients, ultimately contributing to better patient outcomes. Moreover, the results showed the effectiveness of the feature aggregation with our proposed method. The performance of the RF ML method with feature aggregation was better than that of the method without feature aggregation. This indicates that the proposed feature aggregation improved the classification performance. There are two major reasons for this performance improvement: the first is that the feature aggregation improves between informative features more the correlation and could enable the model to use suitable association features during the training stage. Another is that the miss prediction of positive cases is reduced to classify the imbalanced data

Table 7. Comparison of the *p*-value between the other method and the proposed method performance.

Models	ACC	SEN	SPEC	PPV	NPV
Classical ML-method					
RF	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
XGBoost	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SVM	<0.0001	<0.0001	<0.0001	0.005	<0.0001
LGBM	<0.0001	<0.0001	<0.0001	<0.05	<0.0001
Ensemble learning (all classical ML-method)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SMOTE					
RF	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
XGBoost	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SVM	<0.0001	<0.0001	<0.0001	0.486	<0.0001
LGBM	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Ensemble learning (all SMOTE method)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SMOTEENN					
RF	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
XGBoost	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SVM	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
LGBM	<0.0001	0.001	<0.0001	<0.0001	<0.0001
Ensemble learning (all SMOTEENN method)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Proposed					
W/O feature aggregation	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Note. Except for PPV in SVM, SEN and NPV in LGBM, and NPV in W/O feature aggregation, each *p*-value that is less than 0.05 indicates a statistically significant difference (Kolmogorov–Smirnov test and chi-square test) within all the comparison classification methods. ACC: accuracy; SEN: sensitivity; SPEC: specificity; PPV: positive-predictive value; NPV: negative-predictive value; MCC: Matthews correlation coefficient; AUC: area under the curve; ML: machine learning; RF: random forest; SVM: support vector machine; LGBM: light gradient boosting; XGBoost: eXtreme gradient boosting; SMOTE: synthetic minority oversampling technique; SMOTEENN: SMOTE edited nearest neighbor.

successfully. This showed that the RF ML method, as a basic prediction model for the proposed model, significantly enhances the classification performance, which is of great interest.

For the clinical feature interpretability, we visually illustrated our proposed method with RFE feature selection by plotting rank of feature by using Gini index (Figures 3 and 4). Our model indicated that the top five significant discrete numerical features were BartheIindexTotalGroup, Activity, BMIGroup, Mobility, and DiagnosticType, respectively.

The single clinically significant feature, PIs before, served as a fundamental indicator of a patient's prior susceptibility to PI. Building upon this, the proposed feature aggregation approach identified the top five discrete features BarthelIndexTotalGroup, Activity, BMIGroup, Mobility, and DiagnosticType each reflecting different facets of a patient's functional status, body composition, and clinical profile. While DiagnosticType is not traditionally considered a direct PI assessment factor, previous studies have suggested its association with PI occurrence.[16]
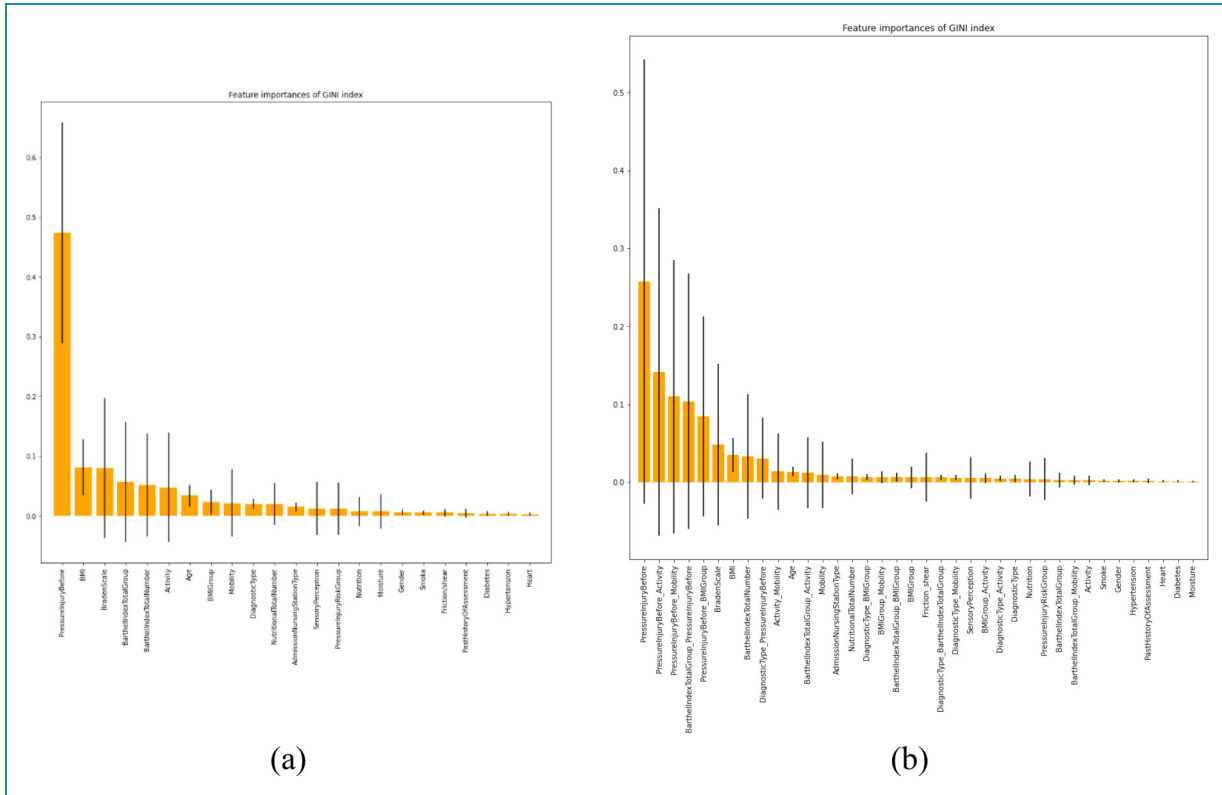
Figure 4. Results of features ranking with Gini index by using the proposed method: (a) the result of the recursive feature elimination (RFE) feature selection stage and (b) the result by executing the proposed method.

By incorporating these features through probability aggregation, the method was able to reveal how their combined interactions contribute to PI risk. As illustrated in Figure 4, the resulting combination features provided enhanced insight into underlying risk associations. Notably, PressureInjuryBefore_BMIGroup exhibited a particularly high contribution to PI risk, underscoring how a patient's prior history of PI interacts with their body mass index (BMI) status. Similar patterns emerged from other combination features, such as PressureInjuryBefore_BarthelIndex TotalGroup, PressureInjuryBefore_Mobility, and Diagnostic Type_BMIGroup. The consistency in feature importance, observed through both the Gini index and SHAP analyses (Figures 4 and 5), validated the robustness and interpretability of these integrated features. From a clinical perspective, feature aggregation offers a more intuitive and actionable representation of complex data interactions. By providing probability-driven visualizations or dashboards that highlight how these combination features shape a patient's PI risk, clinicians can more readily identify individuals requiring closer monitoring or preventive measures. In this way, the approach not only refines the predictive model's accuracy but also transforms data-driven insights into tangible strategies for improving patient outcomes.

In addition to the clinically significant feature of PIs before, the integration of BMI, Braden Scale, and Barthel index is relevant for guiding clinical decision-making. Each variable addresses a distinct yet interrelated dimension of patient status. BMI indicates nutritional status and body composition, the Braden Scale encompasses skin integrity, moisture, and mobility, and the Barthel index measures functional independence in daily tasks. These variables are integral to identifying patients at elevated risk of PIs. When aggregated, PressureInjuryBefore _BMIGroup emerges as an influential combined factor, suggesting that prior PI history and body composition together may increase susceptibility to future PIs. Recognizing this joint effect can inform resource allocation and targeted preventive measures for high-risk patients. Identifying the synergy among BMIGroup, BarthelIndex TotalGroup, Activity, Mobility, and DiagnosticType further clarifies how multiple factors concurrently influence PI risk. Integrating these combined indicators into clinical systems may improve predictive accuracy and enable timely preventive actions, contributing to better patient outcomes.

Although the proposed method displayed outstanding performance, it still had some drawbacks and challenges. First, only structured data were used for this study;
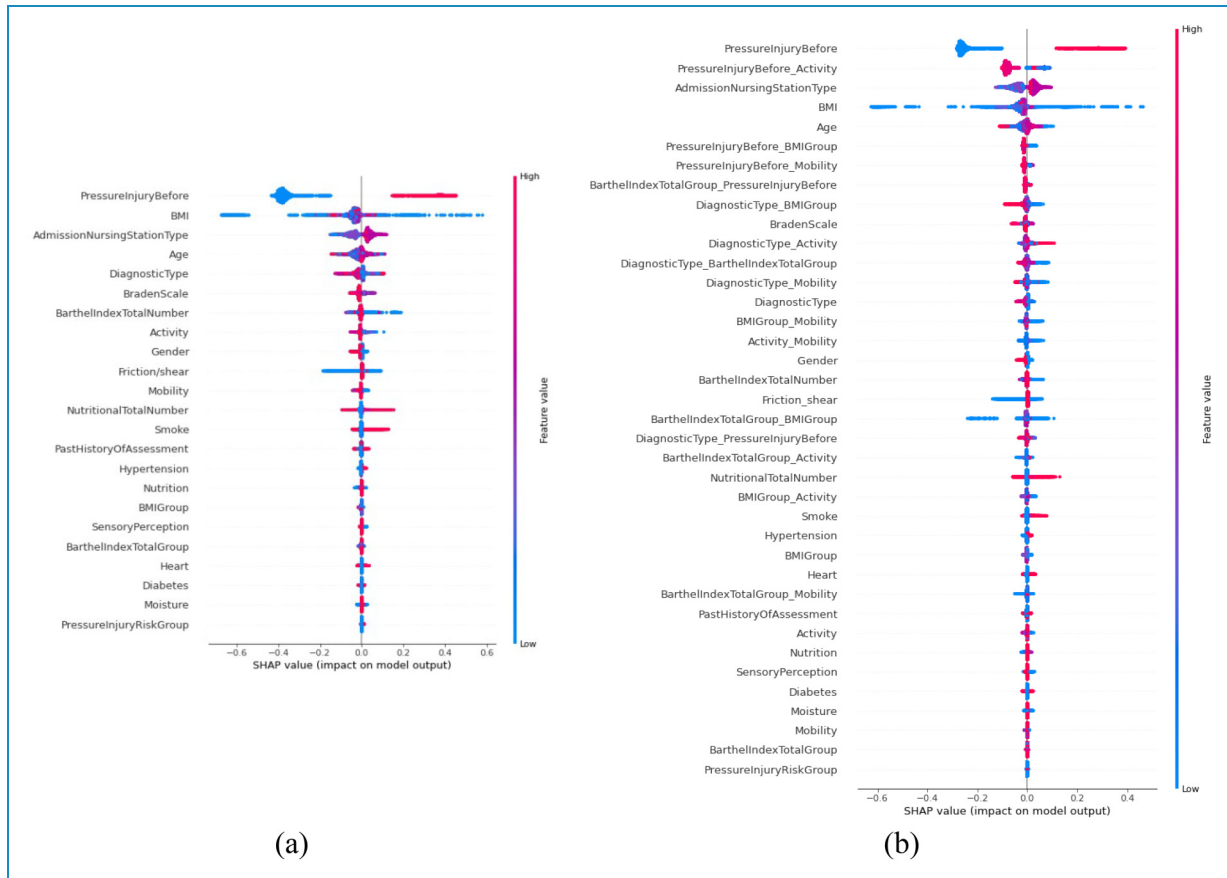
Figure 5. Feature analysis: (a) the result of the recursive feature elimination (RFE) feature selection stage and (b) the result by executing the proposed method.

besides, patient comorbidity, acuity, and values of laboratory tests were not considered, which may be associated with the risk factors in our developed prediction method. Furthermore, the integration of multi-dimensional data that includes narrative nursing notes, individual habits, environmental factors, and image reports is suggested to improve the accuracy of the prediction method. Advanced analytics methods, including natural language processing, can extract context not evident through conventional numeric variables, while environmental measurements may clarify how room conditions affect PI development. However, this approach involves several practical obstacles. Ensuring compatible data formats across different systems is important for consistent information exchange, safeguarding patient confidentiality and following institutional regulations requires meticulous oversight, and establishing a technical infrastructure capable of handling large, diverse datasets necessitates significant computational resources and expertise. Moreover, clinical staff may require training to interpret and act on the insights produced by these additional sources. Addressing these considerations in future research could lead to a more comprehensive framework

that seamlessly integrates both structured and unstructured data in routine clinical practice.

The feature engineering process also has limitations, such as the potential challenges in capturing complex feature interactions and the risk of overfitting. We used RFE for feature selection, which systematically ranks and removes less relevant features while retaining the most informative ones. However, RFE also has limitations, including sensitivity to the dataset and the potential for overfitting when the dataset size is limited. This limitation might have led to less robust features and increased the risk of overfitting, especially when dealing with a limited dataset size. Second, we used a dataset extracted from a single medical center. Although the results may be useful for local patients, the method may not be generalizable to other healthcare systems, as performance may vary due to different population characteristics and clinical practices. Multi-healthcare system validation is required to optimize the predictive performance of the proposed method, ensuring its applicability across different healthcare settings. Third, integrating this method into clinical workflows presents several challenges, such as ensuring interoperability with existing EHR systems and maintaining real-time data

processing capabilities. Moreover, clinical staff training and workflow adaptation would be necessary to maximize the effectiveness of the model in everyday clinical practice. Extending the model to different healthcare institutions requires additional validation and potential adaptations to account for variations in patient populations, healthcare resources, and information system infrastructures. Conducting multi-center studies may help assess the model's robustness across diverse clinical settings and guide parameter adjustments to accommodate data from facilities of varying sizes and regions. Such efforts can involve techniques designed to handle heterogeneous data, including model calibration or transfer learning, which allow the model to maintain predictive performance despite differences in patient demographics and clinical practices. By confirming the model's stability in multiple contexts, a broader implementation in routine clinical workflows may become feasible.

Finally, we only used an RF ML model for constructing the prediction model. Compared with more complex deep learning (DL) models, our ML-based method is relatively simple. Our future work suggests the establishment of a DL model for predicting PIs, which would enable the use of both structured and unstructured data, such as narrative notes and imaging data. Multi-center validation and incorporating unstructured data will help create a more robust and generalizable predictive system. The proposed feature aggregation method could provide effective correlation information between different features of the DL network, which can be useful for improving and modifying our prediction architecture. Compared with a more complex DL model, our ML-based method is relatively simple. Our future work suggests the establishment of a DL model for predicting PIs to train the clinical predictive system and create a unified DL network. The proposed feature aggregation method could provide effective correlation information between different feature for the DL network which can be useful for improving and modifying our prediction architecture.

Conclusion

In this study, a ML-based method was proposed for the predicting risk of PI. The performance of the proposed method was good in evaluating the risk with PI in patients. The proposed feature aggregation was critical in predicting risk of PI because feature aggregation could reduce false prediction in positive cases and make the higher predicting ability. Hence, our method provides risk predictions to assist clinical nursing staff in developing or adjusting care guidelines. Furthermore, the key contributions of this study lie in the integration of expert knowledge, automated feature selection, and feature aggregation, enabling improved predictive accuracy. However, several limitations, such as the reliance on the available dataset and the need for further validation

in diverse clinical settings, remain and highlight areas for future improvement. In conclusion, the proposed method provides an objective and effective prediction performance that accounts for correlation with different key features and can help clinical care team tailor their treatment of patients in PI.

ORCID iD: Chiao-Min Chen ⬦ https://orcid.org/0000-0001-7553-6326

References

1. Edsberg LE, Langemo D, Baharestani MM, et al. Unavoidable pressure injury: state of the science and consensus outcomes. *J Wound Ostomy Continence Nurs* 2014; 41: 313–334.
2. Pacific P. Prevention and treatment of pressure ulcers: quick reference guide, 2014, https://www.epuap.org/wp-content/uploads/2010/10/Quick-Reference-Guide-DIGITAL-NPUAP-EPUAP-PPPIA-16Oct2014.pdf
3. Jaul E. Assessment and management of pressure ulcers in the elderly. *Drugs Aging* 2010; 27: 311–325.
4. Kayser SA, VanGilder CA and Lachenbruch C. Predictors of superficial and severe hospital-acquired pressure injuries: a cross-sectional study using the International Pressure Ulcer Prevalence™ survey. *Int J Nurs Stud* 2019; 89: 46–52.

5. Sen CK, Gordillo GM, Roy S, et al. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair Regen* 2009; 17: 763–771.

6. Chaboyer WP, Thalib L, Harbeck EL, et al. Incidence and prevalence of pressure injuries in adult intensive care patients: a systematic review and meta-analysis. *Crit Care Med* 2018; 46: e1074–e1081.

7. Tubaishat A, Papanikolaou P, Anthony D, et al. Pressure ulcers prevalence in the acute care setting: a systematic review, 2000–2015. *Clin Nurs Res* 2018; 27: 643–659.

8. Igarashi A, Yamamoto-Mitani N, Gushiken Y, et al. Prevalence and incidence of pressure ulcers in Japanese long-term-care hospitals. *Arch Gerontol Geriatr* 2013; 56: 220–226.

9. Padula WV and Delarmente BA. The national cost of hospital-acquired pressure injuries in the United States. *Int Wound J* 2019; 16: 634–640.

10. Bauer K, Rock K, Nazzal M, et al. Pressure ulcers in the United States' inpatient population from 2008 to 2012: results of a retrospective nationwide study. *Ostomy Wound Manage* 2016; 62: 30–38.

11. Theisen S, Drabik A and Stock S. Pressure ulcers in older hospitalised patients and its impact on length of stay: a retrospective observational study. *J Clin Nurs* 2012; 21: 380–387.

12. Hong H-J, Kim N-c, Jin Y, et al. Trigger factors and outcomes of falls among Korean hospitalized patients: analysis of electronic medical records. *Clin Nurs Res* 2015; 24: 51–72.

13. Han Y, Jin Y, Jin T, et al. Impact of pressure injuries on patient outcomes in a Korean hospital: a case-control study. *J Wound Ostomy Continence Nurs* 2019; 46: 194–200.

14. Serrano ML, Méndez MG, Cebollero FC, et al. Risk factors for pressure ulcer development in intensive care units: a systematic review. *Medicina Intensiva (English Edition)* 2017; 41: 339–346.

15. Jin Y, Kim H, Jin T, et al. Automated fall and pressure injury risk assessment systems: nurses' experiences, perspectives, and lessons learned. *CIN: Comput, Inf, Nurs* 2021; 39: 321–328.

16. Kottner J, Cuddigan J, Carville K, et al. Prevention and treatment of pressure ulcers/injuries: the protocol for the second update of the International Clinical Practice Guideline 2019. *J Tissue Viability* 2019; 28: 51–58.

17. Shi C, Dumville JC and Cullum N. Evaluating the development and validation of empirically-derived prognostic models for pressure ulcer risk assessment: a systematic review. *Int J Nurs Stud* 2019; 89: 88–103.

18. Raju D, Su X, Patrician PA, et al. Exploring factors associated with pressure ulcers: a data mining approach. *Int J Nurs Stud* 2015; 52: 102–111.

19. Hyun S, Vermillion B, Newton C, et al. Predictive validity of the Braden scale for patients in intensive care units. *Am J Crit Care* 2013; 22: 514–520.

20. Seong-Hi P and Lee HS. Assessing predictive validity of pressure ulcer risk scales—a systematic review and meta-analysis. *Iran J Public Health* 2016; 45: 122.

21. Song W, Kang M-J, Zhang L, et al. Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *J Am Med Inform Assoc* 2021; 28: 759–765.

22. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 1998; 20: 832–844.

23. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002; 38: 367–378.

24. Montgomery DC, Peck EA and Vining GG. *Introduction to linear regression analysis.* Hoboken, New Jersey, USA: John Wiley & Sons, 2021.

25. Suykens JA and Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett* 1999; 9: 293–300.

26. Catling FJ and Wolff AH. Temporal convolutional networks allow early prediction of events in critical care. *J Am Med Inform Assoc* 2020; 27: 355–365.

27. Wu L, Hu Y, Zhang X, et al. Development of a knowledge mining approach to uncover heterogeneous risk predictors of acute kidney injury across age groups. *Int J Med Inf* 2022; 158: 104661.

28. Yan K and Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens Actuators, B* 2015; 212: 353–363.

29. Belgiu M and Drăguţ L. Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 2016; 114: 24–31.

30. Refaeilzadeh P, Tang L and Liu H. Cross-validation. *Encyclop Datab Syst* 2009; 5: 532–538.

31. Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens* 2005; 26: 217–222.

32. Hu Y-H, Lee Y-L, Kang M-F, et al. Constructing inpatient pressure injury prediction models using machine learning techniques. *CIN: Comput, Inf, Nurs* 2020; 38: 415–423.

33. Jaul E, Barron J, Rosenzweig JP, et al. An overview of co-morbidities and the development of pressure ulcers among older adults. *BMC Geriatr* 2018; 18: 1–11.

34. Anderson C, Bekele Z, Qiu Y, et al. Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. *BMC Med Inform Decis Mak* 2021; 21: 1–13.

35. Jiang M, Ma Y, Guo S, et al. Using machine learning technologies in pressure injury management: systematic review. *JMIR Med Inform* 2021; 9: e25704.

36. Kor C-T, Li Y-R, Lin P-R, et al. Explainable machine learning model for predicting first-time acute exacerbation in patients with chronic obstructive pulmonary disease. *J Pers Med* 2022; 12: 228.

37. Anand R, Mehrotra KG, Mohan CK, et al. An improved algorithm for neural network classification of imbalanced training sets. *IEEE Trans Neural Netw* 1993; 4: 962–969.

38. Branco P, Torgo L and Ribeiro RP. A survey of predictive modeling on imbalanced domains. *ACM Comput Surv (CSUR)* 2016; 49: 1–50.

39. Soltanzadeh P and Hashemzadeh M. RCSMOTE: range-controlled synthetic minority over-sampling technique for handling the class imbalance problem. *Inf Sci (NY)* 2021; 542: 92–111.

40. Khushi M, Shaukat K, Alam TM, et al. A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Access* 2021; 9: 109960–75.