BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Development of an explainable machine learning model for predicting device-related pressure injuries in clinical settings

Yijie Qian[1], Hongying Pan[1*], Jun Chen[1], Hongyang Hu[1], Mei Fang[1], Chen Huang[1], Yihong Xu[1] and Yang Gao[1]

Abstract

Background Device-related pressure injury (DRPI) is a prevalent and severe problem for patients using medical devices. Timely identification of patients at high risk of DRPI is crucial for healthcare providers to make informed decisions and prevent DRPI quickly. Given the rapid advancements in computer technology, we aimed to develop an interpretable artificial intelligence (AI) model for predicting DRPI, utilizing SHAP (SHapley Additive exPlanations) to enhance the model's transparency and provide insights into feature importance.

Methods We enrolled 675 study subjects (225 in the DRPI group and 450 in the non-DRPI group) from a single medical center between January 2019 and December 2020. Python was used to perform classification models, including extreme gradient boosting (XGBoost), random forest (RF), decision tree (DT), Logistic Regression (LR), support vector machine (SVM), and K-Nearest neighbors (KNN). We evaluated the performance of the six models using area under the ROC curve (AUC), specificity, accuracy, and sensitivity, with the dataset split into a 80% training set and a 20% testing set. We utilized several analyses, such as SHAP and Uniform Manifold Approximation and Projection (UMAP), to explore the potential contribution of different characteristics in our risk prediction models.

Results In the test set, XGBoost model outperformed the other models (AUC = 0.964). The interpretation of the model using SHAPscores revealed that the length of stay, instrument type, emergency admissions, instrument material, and instrument duration of use are the top five most important features in predicting DRPI.

Conclusion Our study demonstrated that the development of DRPI in patients can be accurately predicted using the machine learning (ML) model. The findings not only provide clinical caregivers with a valuable framework to identify patients at high risk of DRPI, but also lay the groundwork for developing targeted preventive strategies and personalized interventions.

Keywords Device-Related Pressure Injury, Pressure Ulcer, Machine learning, Explainable artificial intelligence, Forecasting

*Correspondence:
Hongying Pan
3191016@zju.edu.cn
[1]Nursing Department, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

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

Background

With the advancement of science, technology, and modern medicine, medical devices have become indispensable tools in clinical practice. Any patient with a medical device in contact with their skin or mucous membranes is at risk of developing a device-related pressure injury (DRPI). DRPIs not only cause additional pain to patients and prolong their hospital stay but also increase the workload of medical personnel, making it challenging for patients to achieve the desired outcome and increasing the cost of treatment [1, 2]. In the field of nursing, the Braden, Norton, and Waterlow scales were proposed to assess the risk of pressure injury (PI) and to adopt preventive measures to prevent its progression [3–5].

According to the European Pressure Ulcer Advisory Panel (EPUAP), National Pressure Injury Advisory Panel (NPIAP) and Pan Pacific Pressure Injury Alliance (PPPIA) guidelines, PI is defined as localized damage to the skin and/or underlying tissue, often occurring over bony prominences or under medical or other equipment [6–8]. Previous studies have suggested that the etiology of PI is complex and that the development of PI involves interactions between various factors, including ischemia, impaired lymphatic drainage, reperfusion injury, and deformation [9–12]. Diabetes, sensory perception, low albumin, and malnutrition are considered crucial indirect causal factors because they alter direct causal factors, which can influence outcomes [13].Confined damage to the skin resulting from the use of a medical device for diagnostic or therapeutic purposes is known as DRPI, and the shape of the local injury generally corresponds to the shape of the medical device.

Despite growing concern and active prevention, there remains a high prevalence of PI. In recent years, with the increased clinical use of medical devices, the likelihood of close contact between the device and the skin has increased, resulting in more frequent contact site injuries. A study that included 13,254 patients in 1,117 Intensive Care Units across 90 countries showed an overall prevalence of 26.6% [14]. A meta-analysis reported a 14% incidence and 11% prevalence of DRPI in adults [15]. Adult inpatients with DRPI account for as much as 29%-35% of in-hospital acquired pressure injuries and are 2.4 times more likely to develop a pressure injury with a medical device than without [16].

Advanced computer algorithms efficiently analyze vast healthcare datasets to improve patient assessment accuracy [17]. Numerous studies have highlighted the effectiveness of machine learning (ML) algorithms in developing predictive models for the occurrence of pressure injuries [18–20]. Furthermore, it is recommended that predictive models for medical DRPI be developed independently [21]. To prepare prediction models for DRPI, we expect artificial intelligence (AI) models constructed using ML algorithms. However, models built using powerful ML methods limit their use in medical decision-making due to a lack of interpretability [22]. Thus, interpretable AI is an essential branch of research in AI that has brought about fundamental changes in AI and continues to transform it. The SHapley Additive exPlanations (SHAP) approach is critical in interpretable AI, as it assigns a value to each feature, with higher absolute SHAP values indicating higher importance and values closer to zero indicating lower importance [23].

Pressure injuries are preventable and cost-effective to prevent in today's modern healthcare environment [24, 25]. Currently, nursing staff mainly identifies high-risk factors for DRPI based on clinical experience and the Braden Scale, which are not specific for identifying high-risk factors for DRPI. Risk identification is the first step in PI prevention. However, few studies have focused on risk prediction models for DRPI, and those that do tend to focus on patients in intensive care units, limiting the generalizability of the findings [26]. Therefore, the aim of this study is to construct a universal DRPI risk prediction model for inpatients using ML methods to provide clinical caregivers with a comprehensive screening tool for patients at high risk of DRPI.

Methods

Study design and setting

This is a retrospective, single-center, and cross-sectional study, which predicted the DRPI risk in hospitalized patients by using ML algorithms. In this study, the workflow(Fig. 1) includes data preparation, model development, and evaluation. The features are selected, labels are assigned. The Random Over-Sampling Examples (ROSE) technique is used to handle imbalanced datasets. Hyperparameters are chosen, and cross-validation is performed to identify the best model. The finalized model is used for predictions on unseen data. Our study design ensures a comprehensive approach to ML, addressing imbalanced datasets and achieving robust model performance.

Study subjects

We conducted a retrospective nested case-control study at Sir Run Run Shaw Hospital Affiliated to School of Medicine, Zhejiang University in Hangzhou from January 2019 to December 2020. Patients who used medical devices and developed DRPI were selected as the case group, while patients who used medical devices but did not develop DRPI during the same period were randomly selected as the control group in a 1:2 ratio. The inclusion criteria were age ≥ 18 years, length of stay ≥ 24 hours, use of at least one medical device, and complete medical history. Patients with pre-existing DRPI at the site of device use upon admission were excluded. The final dataset

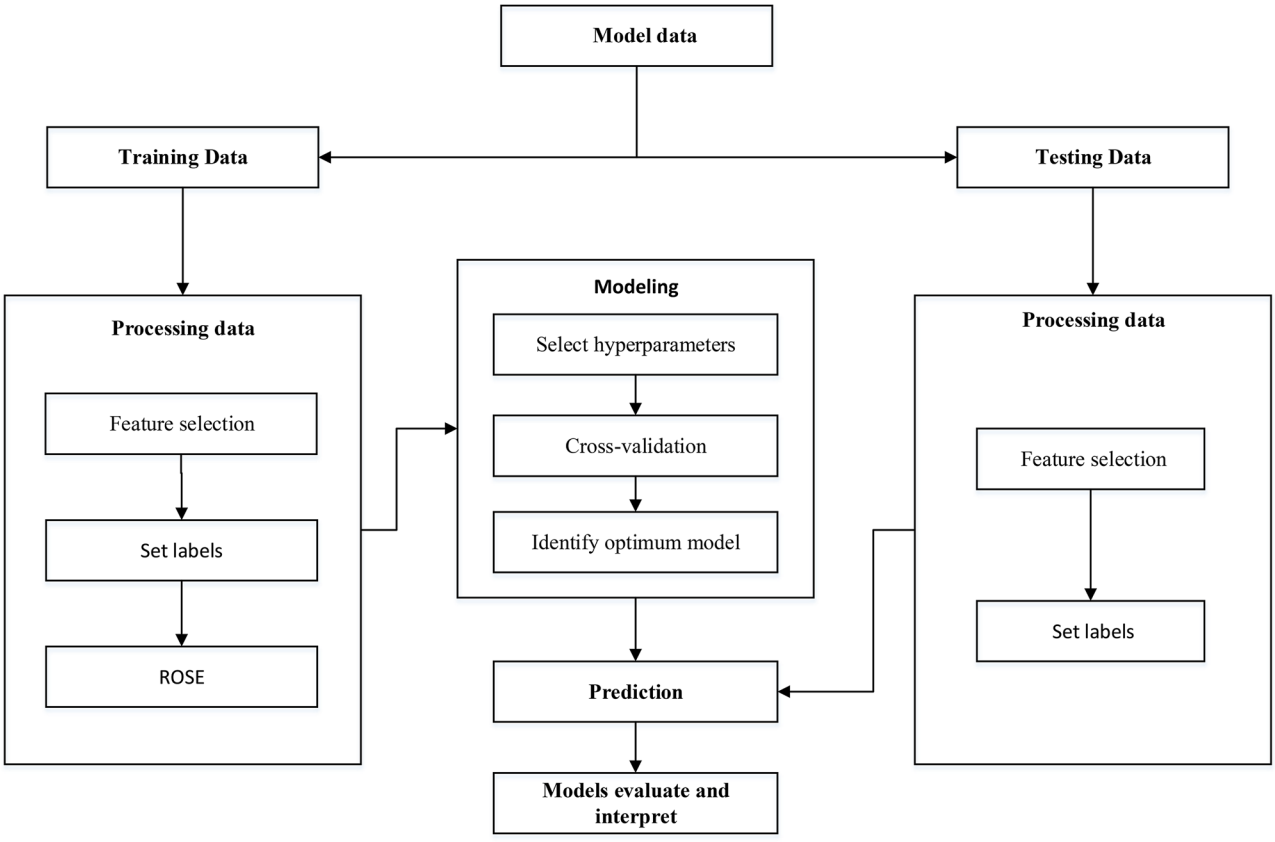


Fig. 1 Workflow for data processing and model building. ROSE, Random Over-Sampling Examples

Table 1 Summary of the included risk factors

Categories	Specific entries
General information	Age, gender, BMI, admission, length of stay, NRS2002, temperature
lab tests	Albumin, C-reactive protein, hypoxia
nursing scale	Braden score (sensory perception subscore, mobility subscore)
medical history	History of circulatory diseases, diabetes mellitus, neurological diseases, disturbance of consciousness
treatment	History of sedatives and myorelaxants, vasoactive agent, enteral nutrition, parenteral nutrition, surgery, ICU
status of instrument	Instrument material, type, duration of use

BMI, Body mass index; NRS2002, Nutrition Risk Screening 2002; hypoxia, SPO_2<90%; circulatory diseases include various types of heart disease such as essential hypertension, coronary heart disease and heart failure; neurological diseases include stroke, cerebral infarction, and cerebral haemorrhage, etc. BMI, NRS2002, enteral nutrition, parenteral nutrition, sedatives and myorelaxants, vasoactive agent are all last values before device use. Braden scores and laboratory indicators are collected for the first time after the use of the device. If multiple instruments are used, the main instrument is selected to collect relevant information

included 675 patients, with 225 in the DRPI group and 450 in the non-DRPI group. The diagnosis of DRPI was based on the Prevention and Treatment of Pressure Ulcer/Injuries Clinical Practice Guideline[8], and DRPI assessment was carried out by a professionally trained member of the pressure injuries team.

Feature selection

Patient data were obtained from the hospital's electronic medical record system. Based on literature reviews, expert panel discussions, and clinical practice, we selected 25 risk factors that were known to influence DRPI as features [2, 6, 27–30]. These risk factors include demographics, medical history, nursing scale, laboratory tests, treatment, and status of instruments (Table 1).

Data preprocessing

Samples from patients diagnosed with DRPI were labeled as positive (1), while samples from patients diagnosed with non-DRPI were labeled as negative (0). As previously mentioned, DRPI accounted for 33% of our study population, and the imbalance in the data could significantly impact the efficacy of ML classifiers. While various oversampling methods have been proposed to address this issue, we opted to utilize the classical ROSE method, which replicates the minority class samples to balance the training set [31–33]. By using real samples instead of generating virtual samples, we ensured that our model was clinically relevant. ROSE's simple strategy aligns with

our goal of considering only the effects of actual samples when training the model [34].

Model building

We constructed six ML models using potential features, including Extreme Gradient Boosting (XGBoost), Random Forest Classifier (RFC), Decision Tree Classifier (DTC), Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). We reviewed the relevant literature on PI to guide our selection of suitable ML algorithms. In order to construct a predictive model, we employed six different ML algorithms. The selection of these algorithms is based on several key considerations, including the nature of the classification problem, the characteristics of the dataset, and the research objectives. Tree models, such as XGBoost, RFC and DTC, are prominent ML algorithms used for predictive modeling. These models are based on the concept of constructing hierarchical tree-like structures to facilitate prediction and classification tasks. LR is a well-known interpretable binary classification algorithm that provides information about how each predictor variable influences the outcome. The KNN algorithm is a flexible algorithm used for regression and classification tasks by comparing the features of neighboring instances. SVM have the ability to identify complex relationships within a dataset and create distinct decision boundaries, thereby enabling the capture of nonlinear interactions among predictor variables. In a predictive classification model, all samples can be classified into four categories based on comparing predicted and true values: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). We calculated evaluation metrics for the classification model, such as sensitivity, specificity, accuracy, and the ROC-AUC as another evaluation of model accuracy. We conducted a series of manually tuned experiments with different model training algorithms and corresponding hyperparameters to achieve better model performance, and we determined the optimum hyperparameters by cross-validation in the training set. To rigorously evaluate the calibration quality of our optimal model, we calculated the Brier score using bootstrap resampling with 1000 iterations, which confirmed the high reliability of our probability estimates. We created Norman diagrams using the Logistic Regression algorithm for a more intuitive and accessible presentation of our research results.

Interpretable ML model explanation

The SHAP was applied to interpret the model output by calculating the contribution of each input characteristic for all samples in the dataset and investigate the impact of different variables [23, 35]. A positive SHAP value indicates an increase in the likelihood of a particular outcome, while a negative SHAP value indicates a decrease. We used global SHAP values to show the importance of features and their interactions in our study, while local visualizations were utilized to show the impact of features on patients suffering from PI outcomes. After a series of processes, such as data scaling and principal component analysis (PCA), the Uniform Manifold Approximation and Projection (UMAP) analysis was used with the raw data as input to provide a 2D representation of the total patient population.

Statistical analysis

All samples were randomly divided into training and test sets in an 80:20 proportion. Numerical variables that followed a normal distribution were expressed as means with standard deviations, and the Student t-test was used for comparing the means. For non-normally distributed measurement data, median with quartiles was used, and the Mann-Whitney U test was applied. Enumeration data was expressed as the number of cases and percentages, and the chi-square test was used for comparing groups. We performed these tests and analyses using SPSS 25.0 software (IBM, Armonk, NY) and Python software (version 3.7.3). We considered a P-value of < 0.05 as statistically significant.

The ML classification models were constructed using python software (version 3.7.3) and statistical packages R (version 4.1.3).

Results

Characteristics of the study population

The study included 675 participants, among which 225 were diagnosed with DRPI. The participants were randomly divided into training and test datasets. We selected 25 potential features, including demographic and clinically relevant variables, to predict DRPI. Table 2 summarizes the characteristics of the study population. To assess sample size adequacy, we considered the Events Per Variable (EPV) criterion. With 225 events and 25 predictors, the resulting EPV is 9.0 (225/25). While slightly below the commonly recommended threshold of 10, this was deemed acceptable for this study[36]. Furthermore, there were no missing data in our study. The completeness of the data is supported by procedures at our hospital, where detailed medical records and necessary examinations are mandatory upon admission, and where the use of medical devices requires thorough documentation, including physician orders and clinical notes.

Performance classification

During the model-building process, we performed continuous parameter tuning to obtain optimal parameter settings and prevent overfitting or underfitting of the model. The final parameter settings are presented

Table 2 Clinical characteristic information of the study objects

	Non DRPI (N = 450)	DRPI (N = 225)	P-value	Test (N = 135)	Training (N = 540)	P-value
CRP (mg/L)						
Median [25%, 75%]	1.80 [0.700, 11.3]	37.3 [8.50, 104]	<0.001	5.00 [1.25, 40.2]	4.75 [1.00, 35.2]	0.37
Age (years)						
Median [25%, 75%]	59.0 [47.0, 67.0]	66.0 [55.0, 74.0]	<0.001	63.0 [51.0, 71.5]	61.0 [50.0, 69.0]	0.25
Length of stay (days)						
Median [25%, 75%]	6.00 [3.00, 8.75]	29.0 [17.0, 48.0]	<0.001	9.00 [5.00, 22.0]	8.00 [4.00, 19.0]	0.325
Sex (%)						
Male	254 (56.4%)	169 (75.1%)	<0.001	88 (65.2%)	335 (62.0%)	0.564
Female	196 (43.6%)	56 (24.9%)		47 (34.8%)	205 (38.0%)	
Emergency admissions (%)						
No	388 (86.2%)	69 (30.7%)	<0.001	88 (65.2%)	369 (68.3%)	0.551
Yes	62 (13.8%)	156 (69.3%)		47 (34.8%)	171 (31.7%)	
ICU (%)						
No	433 (96.2%)	103 (45.8%)	<0.001	101 (74.8%)	435 (80.6%)	0.175
Yes	17 (3.8%)	122 (54.2%)		34 (25.2%)	105 (19.4%)	
Surgery (%)						
No	267 (59.3%)	152 (67.6%)	0.046	88 (65.2%)	331 (61.3%)	0.463
Yes	183 (40.7%)	73 (32.4%)		47 (34.8%)	209 (38.7%)	
BMI (%)						
≥24	165 (36.7%)	71 (31.6%)	0.379	52 (38.5%)	184 (34.1%)	0.436
18.5 ~ 23.9	247 (54.9%)	131 (58.2%)		69 (51.1%)	309 (57.2%)	
<18.5	38 (8.4%)	23 (10.2%)		14 (10.4%)	47 (8.7%)	
NRS2002 (%)						
<3分	422 (93.8%)	116 (51.6%)	<0.001	99 (73.3%)	439 (81.3%)	0.053
≥3分	28 (6.2%)	109 (48.4%)		36 (26.7%)	101 (18.7%)	
Diabetes (%)						
No	376 (83.6%)	177 (78.7%)	0.147	105 (77.8%)	448 (83.0%)	0.202
Yes	74 (16.4%)	48 (21.3%)		30 (22.2%)	92 (17.0%)	
Circulatory diseases (%)						
No	284 (63.1%)	99 (44.0%)	<0.001	74 (54.8%)	309 (57.2%)	0.683
Yes	166 (36.9%)	126 (56.0%)		61 (45.2%)	231 (42.8%)	
Neurological diseases (%)						
No	416 (92.4%)	150 (66.7%)	<0.001	110 (81.5%)	456 (84.4%)	0.48
Yes	34 (7.6%)	75 (33.3%)		25 (18.5%)	84 (15.6%)	
Sensory perception subscore (%)						
>2分	447 (99.3%)	151 (67.1%)	<0.001	117 (86.7%)	481 (89.1%)	0.525
≤2分	3 (0.7%)	74 (32.9%)		18 (13.3%)	59 (10.9%)	
Mobility subscore (%)						
>2分	433 (96.2%)	120 (53.3%)	<0.001	106 (78.5%)	447 (82.8%)	0.305
≤2分	17 (3.8%)	105 (46.7%)		29 (21.5%)	93 (17.2%)	
Instrument material (%)						
Metal	0 (0%)	6 (2.7%)	<0.001	2 (1.5%)	4 (0.7%)	0.965
Gypsum	2 (0.4%)	2 (0.9%)		1 (0.7%)	3 (0.6%)	
Plastic	370 (82.2%)	93 (41.3%)		90 (66.7%)	373 (69.1%)	
silicone or rubber	78 (17.3%)	68 (30.2%)		31 (23.0%)	115 (21.3%)	
Nonwoven	0 (0%)	10 (4.4%)		2 (1.5%)	8 (1.5%)	
Cotton	0 (0%)	46 (20.4%)		9 (6.7%)	37 (6.9%)	
Instrument type (%)						
1	170 (37.8%)	11 (4.9%)	<0.001	33 (24.4%)	148 (27.4%)	0.714
2	89 (19.8%)	14 (6.2%)		20 (14.8%)	83 (15.4%)	
3	113 (25.1%)	12 (5.3%)		23 (17.0%)	102 (18.9%)	
>3	78 (17.3%)	188 (83.6%)		59 (43.7%)	207 (38.3%)	

Table 2 (continued)

	Non DRPI (N = 450)	DRPI (N = 225)	P-value	Test (N = 135)	Training (N = 540)	P-value
Instrument duration of use (%)						
Continuous 24-hour use for > 3 days	111 (24.7%)	153 (68.0%)	<0.001	58 (43.0%)	206 (38.1%)	0.707
Continuous 24-hour use for ≤ 3 days	121 (26.9%)	30 (13.3%)		28 (20.7%)	123 (22.8%)	
Intermittent use for > 30 min each	87 (19.3%)	41 (18.2%)		26 (19.3%)	102 (18.9%)	
Intermittent use for ≤ 30 min each	131 (29.1%)	1 (0.4%)		23 (17.0%)	109 (20.2%)	
Albumin (%)						
>35 g/L	345 (76.7%)	73 (32.4%)	<0.001	79 (58.5%)	339 (62.8%)	0.417
≤35 g/L	105 (23.3%)	152 (67.6%)		56 (41.5%)	201 (37.2%)	
Temperature (%)						
>37.4	42 (9.3%)	90 (40.0%)	<0.001	30 (22.2%)	102 (18.9%)	0.568
35 ~ 37.4	408 (90.7%)	132 (58.7%)		104 (77.0%)	436 (80.7%)	
<35	0 (0%)	3 (1.3%)		1 (0.7%)	2 (0.4%)	
Consciousness (%)						
Yes	444 (98.7%)	157 (69.8%)	<0.001	117 (86.7%)	484 (89.6%)	0.406
No	6 (1.3%)	68 (30.2%)		18 (13.3%)	56 (10.4%)	
Hypoxia (%)						
No	444 (98.7%)	184 (81.8%)	<0.001	127 (94.1%)	501 (92.8%)	0.734
Yes	6 (1.3%)	41 (18.2%)		8 (5.9%)	39 (7.2%)	
Enteral nutrition (%)						
No	443 (98.4%)	155 (68.9%)	<0.001	119 (88.1%)	479 (88.7%)	0.976
Yes	7 (1.6%)	70 (31.1%)		16 (11.9%)	61 (11.3%)	
Parenteral nutrition (%)						
No	438 (97.3%)	188 (83.6%)	<0.001	122 (90.4%)	504 (93.3%)	0.317
Yes	12 (2.7%)	37 (16.4%)		13 (9.6%)	36 (6.7%)	
History of vasoactive agent (%)						
No	447 (99.3%)	191 (84.9%)	<0.001	125 (92.6%)	513 (95.0%)	0.375
Yes	3 (0.7%)	34 (15.1%)		10 (7.4%)	27 (5.0%)	
History of sedatives and myorelaxants (%)						
No	300 (66.7%)	90 (40.0%)	<0.001	74 (54.8%)	316 (58.5%)	0.495
Yes	150 (33.3%)	135 (60.0%)		61 (45.2%)	224 (41.5%)	

CRP, C-Reactive Protein; BMI, Body mass index; NRS2002, Nutrition Risk Screening 2002

Table 3 Performance classification with 25 features

Model	AUC	Accuracy	Sensitivity	Specificity
XGBoost	0.964	0.919	0.957	0.898
SVM	0.962	0.904	0.915	0.898
LogisticRegression	0.959	0.919	0.936	0.909
RandomForestClassifier	0.957	0.919	0.957	0.898
DecisionTreeClassifier	0.945	0.904	0.915	0.898
KNeighborsClassifier	0.911	0.889	0.915	0.875

AUC, Area under the ROC Curve; XGBoost, Extreme Gradient Boosting

in Supplementary Table 1. Default values were used for all parameters not listed. The performance of all six ML models for predicting DRPI risk is summarized in Table 3 and visualized in Fig. 2. Among the models, the XGBoost model exhibited superior performance in the test set. The calibration quality of this optimal model was assessed using the Brier score, yielding a value of 0.075 (95% CI: 0.036–0.115) calculated through bootstrap resampling with 1000 iterations, further confirming the model's excellent probability estimation capabilities.

Model interpretation

The XGBoost model was used to interpret the ML model, and a summary plot of the average absolute SHAP values was created to demonstrate the importance of features in descending order (Fig. 3A, B). The top five most important predictors were length of stay, instrument type, emergency admissions, instrument material, and instrument duration in predicting DRPI.

Figure 3C-E display the relationship between the SHAP values of the three most crucial predictors and their measured values. Cut-off thresholds can be established from the graph to differentiate between high risk (SHAP-value > 0) and low risk (SHAP-value < 0). The SHAP values indicated that patients with a length of stay exceeding 16 days, instrument type greater than three, or admissions as emergencies have a higher risk of DRPI.

The force plot of SHAP sought to illustrate the direction and magnitude of the effect of each feature on the predictive model of DRPI in individual patients. The blue force pushes the forecast to the left, and the red pushes

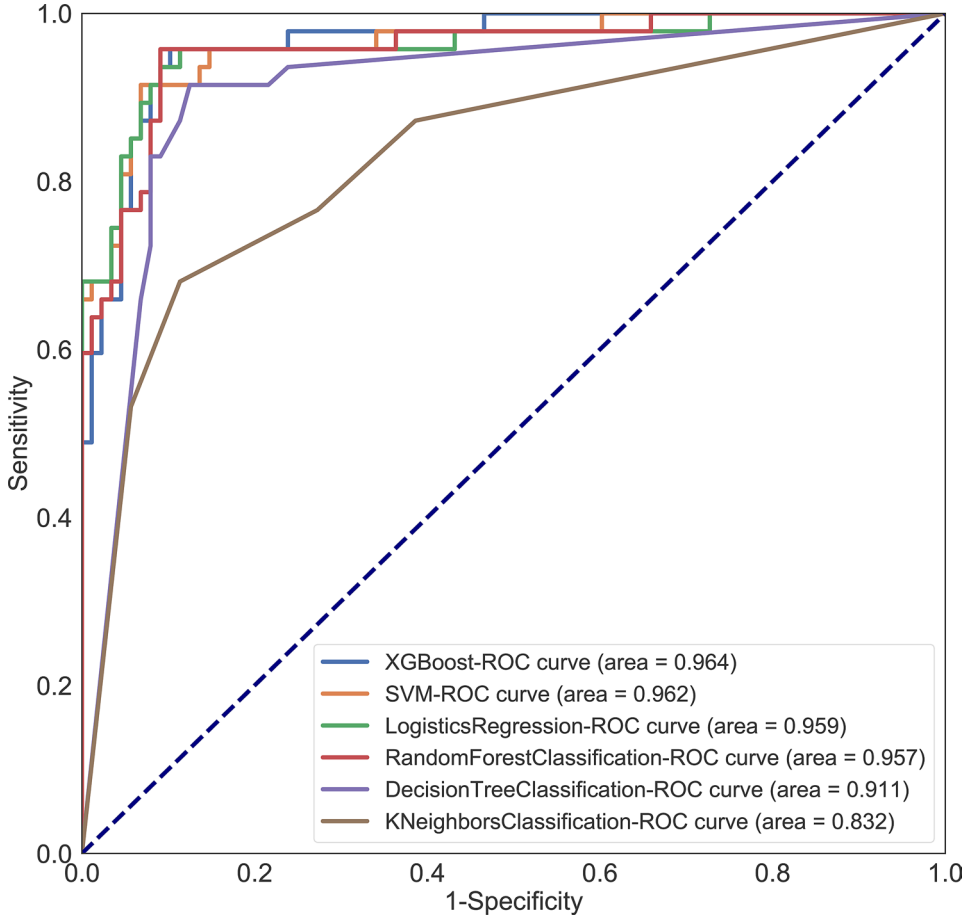


Fig. 2 Receiver operating characteristic curves showing the performance of the six models

it to the right; the wider the force, the more pronounced the push. There were individual characteristics attributed to 2 representative patients who did and did not have DRPI in Fig. 3F. For example, high values of length of stay (51), instrument material (Cotton), instrument type (4 types), age (82), and history of heart disease had a boosting effect on risk prediction while admission as a non-emergency depressed it.

A nomogram was created to depict the logistic regression algorithm model using the seven most important features (Fig. 4). Since the exact length of stay is unpredictable, it was transformed into a binary variable with a cutoff value of 16 in LR model. Although it was less accurate than XGBoost (Table 4), the logistic regression algorithm model was simpler to interpret and use.

Cluster analysis of DRPI patients

To evaluate the significant variables further, we performed reduction and clustering analyses on the training and test sets using PCA and UMAP. We divided the patients included in the analysis into five clusters (Fig. 5A), with Clusters 0 and 1 having a higher percentage of DRPI compared to Clusters 2, 3, and 4 (Fig. 5B). Figure 5C displays the distribution of features, instrument types, and emergency admissions. The most important feature was the length of stay, and the distribution of the groups obtained from the SHAP analysis after grouping according to different cut-off values was most similar to the distribution of DRPI.

Discussion

Traditional scales have significantly assisted doctors and nurses in medical decision-making and patient care [37, 38]. However, unlike previous studies, our study leverages ML algorithms to predict the risk of DRPI. ML algorithms have significant benefits in developing predictive models, including the development of cancer predictors [39]. Our results indicate that the XGBoost algorithm outperforms the other five algorithms in predicting DRPI. XGBoost is particularly suitable for handling complex classification problems [40]. Additionally, SHAP analysis can reflect feature importance and facilitate model interpretation in clinical practice. Although previous studies have identified several risk factors [27, 28], they only provide static results, which are difficult to analyze on a whole-system basis. The interpretable models we create

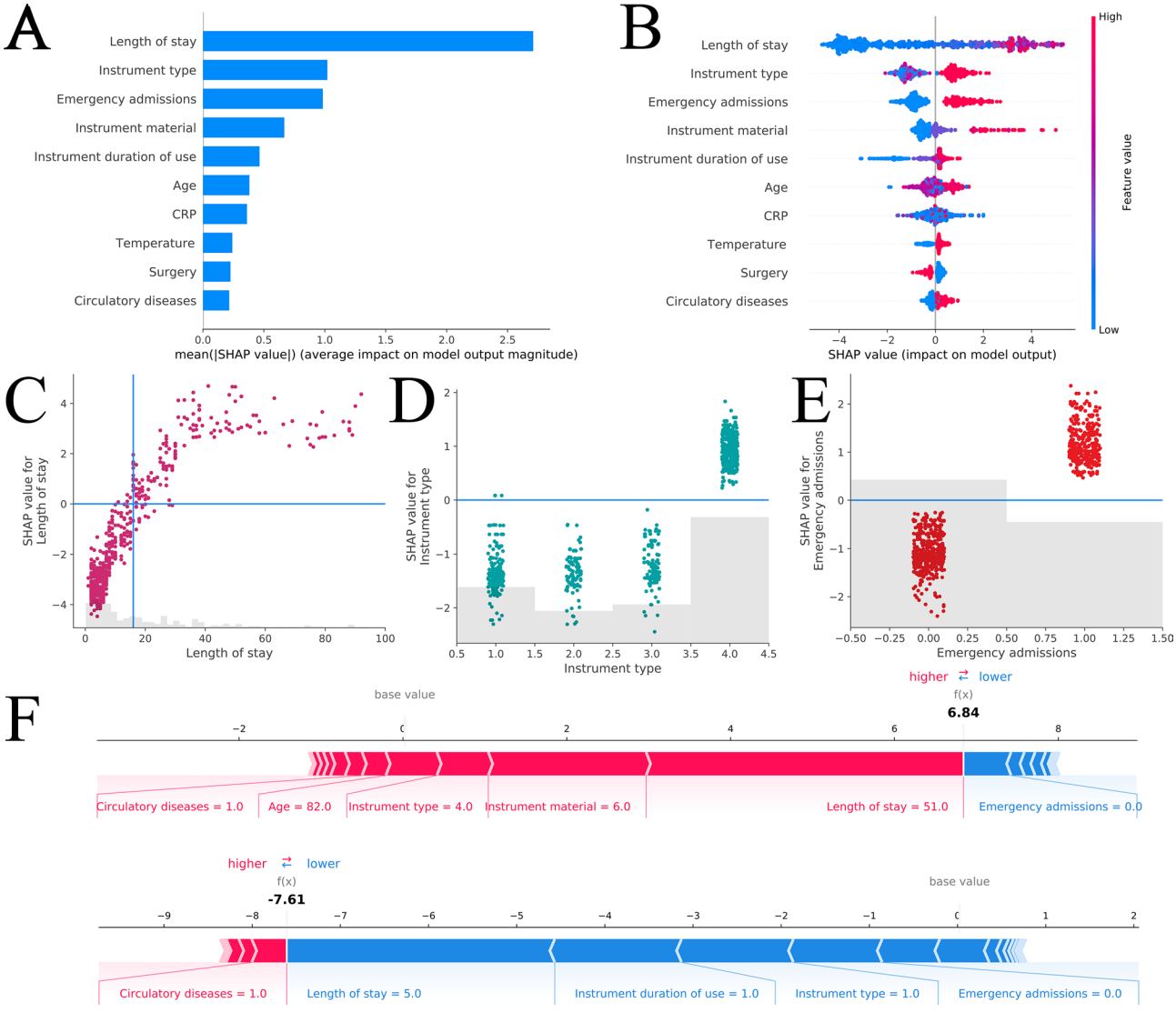


Fig. 3 The analysis of SHapley Additive exPlanations. (**A**) Histogram of mean SHAP value for each predictor. (**B**) bee-swarm plots. The X-axis is the SHAP value, representing the impact of the outcome predictors on the Y-axis. One dot represents one individual patient. The higher the SHAP value, the higher the likelihood of DRPI. (**C**) The dependence plot of the **length of stay** indicates the characteristic's effect on the predictions produced by the model. (**D**) The dependence plot of **the instrument type**. (**E**) The dependence plot of **the emergency admissions**. (**F**) Force Plots for 2 Patients. Red arrows indicate pushing the prediction higher, and blue arrows indicate pushing the prediction lower

allow for dynamic analysis, combining known and uncertain risk factors. As a result, we can identify risk factors and provide personalized and targeted advice to individual patients. This is a merit of SHAP values in the clinical interpretability of ML-based models, although not as clear as the traditional statistical model concepts of HR, relative risk, nomogram, etc. Our research further establishes the high applicability of ML in predicting DRPI, and SHAP analysis can provide valuable information to guide clinical practice.

In this study, a prediction model was constructed using a total of 25 features, which can be generally accessed and made available, presenting a significant advantage. Therefore, it is crucial to screen for features associated with DRPI development. Our SHAP analysis revealed that length of stay, instrument type, emergency admissions, instrument material, and instrument duration of use were the five most critical variables in Fig. 3. Length of stay, instrument type, and emergency admissions had a positive association with the risk of DRPI.

A previous study reported that the risk of DRPI increases by seven times in patients on day 11 of hospitalization compared to day 1 of hospitalization, which is consistent with our findings that longer hospital stays increase the risk of DRPI [41]. Longer hospital stays are associated with increased complexity of patient conditions and dependence on medical devices for extended periods, leading to sustained pressure on the skin at the

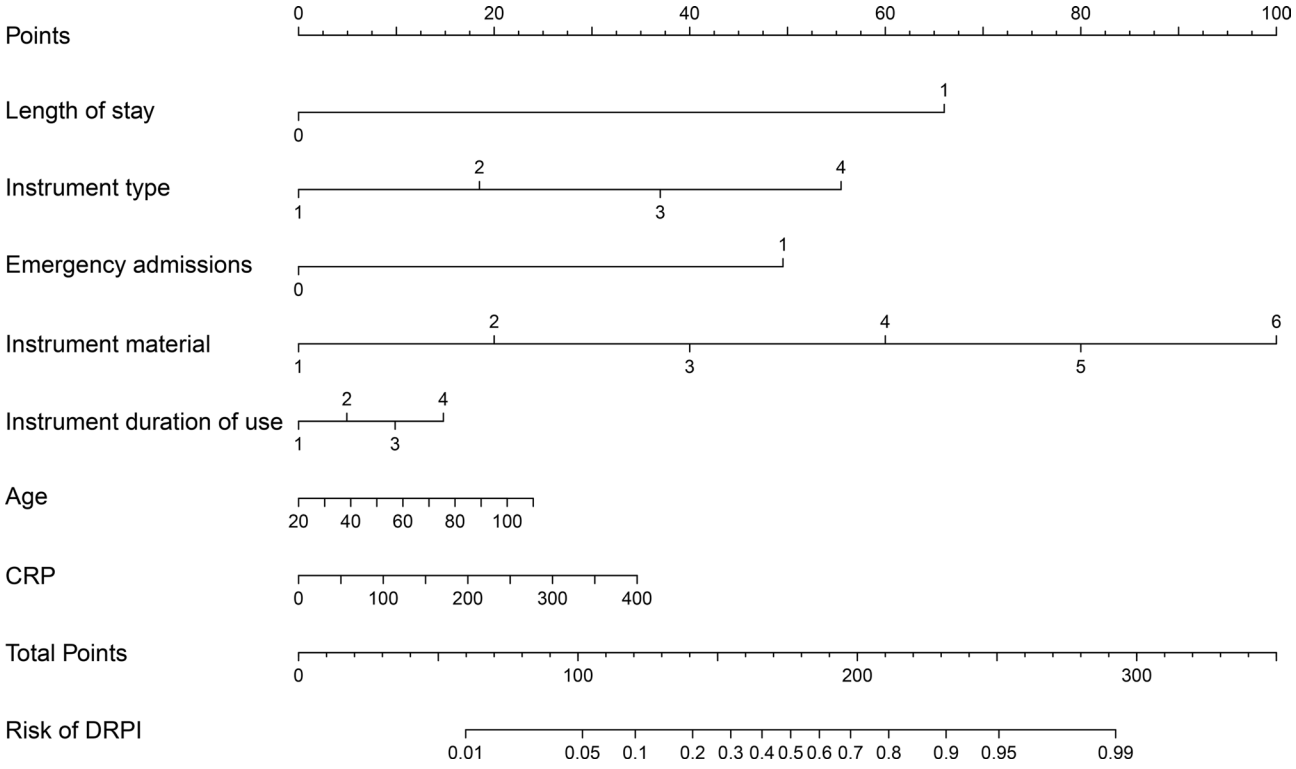


Fig. 4 Nomogram

Table 4 Performance classification with seven most important features

Model	AUC	Accuracy	Sensitivity	Specificity
RandomForestClassifier	0.958	0.933	0.957	0.920
XGBoost	0.950	0.904	0.851	0.932
DecisionTreeClassifier	0.935	0.904	0.915	0.898
SVM	0.930	0.896	0.872	0.909
LogisticRegression	0.929	0.896	0.872	0.909
KNeighborsClassifier	0.813	0.763	0.511	0.898

These models are constructed by the seven most important features. AUC, Area under the ROC Curve; XGBoost, Extreme Gradient Boosting

device site, causing localized skin ischaemia, hypoxia, microvascular congestion, impaired microcirculation, tissue cell deformation, and eventually necrosis, which increase the risk of DRPI [42, 43]. Furthermore, Black J found a 66% increase in the risk of ulcers for each additional day of wearing the cervical collar [44], and a prospective study reported that the number of hospital days was an essential determinant of risk [45]. Therefore, it is crucial to provide reference values for the length of stay to enable medical staff to intervene earlier to prevent DRPI. In our study, SHAP analysis revealed an increased risk of DRPI when the length of stay exceeded 16 days. PCA analysis reduced high-dimensional data to two dimensions, and the distribution of DRPI and length of stay was plotted, showing a high degree of overlap compared to other lengths of stay. In recent studies, they assessed the impact of the synthetic minority-sample oversampling technique (SMOTE) method, which can generate new data, and the Braden score on modeling [18]. In our study, we utilized the ROSE technique to address the problem of imbalanced samples. Our model, constructed without directly incorporating the Braden score, achieved outstanding results on the test data [18, 46].

In our study, the incidence of DRPI increased significantly when the instrument type exceeded three. This is because the larger area of skin in contact with multiple instruments causes more friction and movement between them, which mainly affects the skin surface, leading to stratum corneum damage and a higher risk of skin damage [47]. The study also found that nonwoven and cotton-based devices had a higher risk of DRPI than other materials. Previous studies have identified several devices associated with the occurrence of DRPI, such as plaster/splint/cervical collar, nasogastric tubes, endotracheal tubes, masks, urinary catheters, and electrodes [45]. Our findings showed that the medical device causing DRPI was different from other studies, which may be due to the characteristics and number of participants in our study.

To enhance the practicality of our findings, we built a logistic regression model that incorporates only seven selected features and presented it as a nomogram for the convenience of medical personnel. This model is simpler

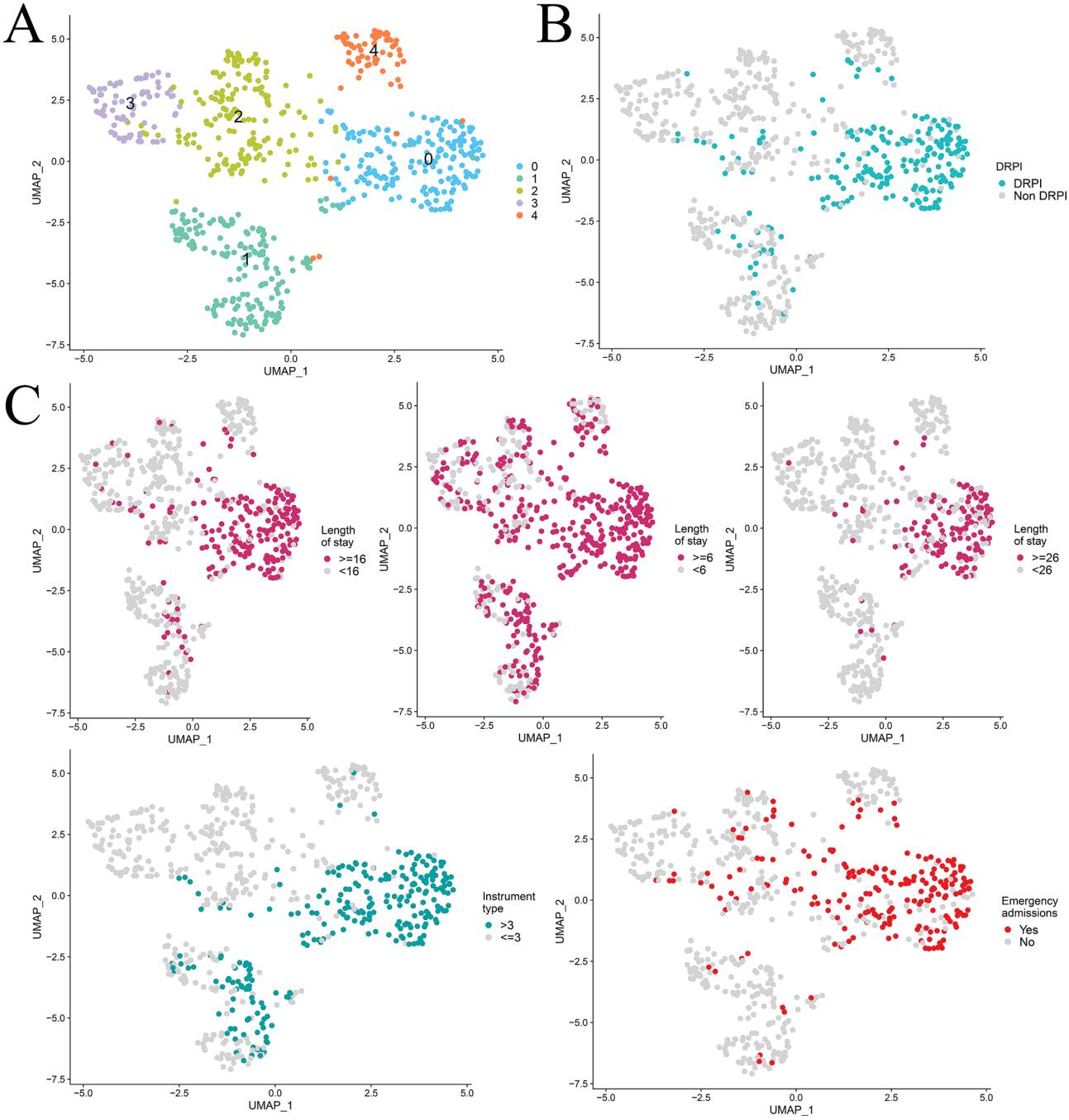


Fig. 5 Cluster analysis of DRPI patients using Uniform Manifold Approximation and Projection (UMAP). (**A**) The UMAP plot of the five identified main patient types in DRPI. (**B**) The scatter plot showing the distribution of DRPI patients. (**C**) The first three variables (length of stay, instrument type, and emergency admissions) help predict the risk of DRPI. Length of stay is grouped according to different cut-off values

to use than the prediction model that uses 25 features and has good performance in accurately predicting the risk of DRPI. Notably, although the logistic regression model has fewer features, the Random Forest Classifier algorithms exhibit even higher AUROC, which may be attributed to the elimination of confounders. Thus, filtering predictor variables is a critical step in constructing predictive models.

However, our study has some limitations. First, the retrospective case-control study design may limit the generalizability of our findings. Second, the study was conducted in a single center, which may have impacted the external validity of our results. Lastly, we have not validated the performance of our model on a large scale, and it is unclear whether it can be widely used.

Conclusion

In summary, this study establishes an interpretable model for predicting the risk of device-related pressure injury in patients. The XGBoost algorithm demonstrated the best predictive performance among the six ML algorithms evaluated. The interpretable models highlight length of stay as the most critical predictor, enabling personalized recommendations for patients through SHAP value analysis. This study's significance lies in the development of a reliable tool that can aid medical staff in identifying patients at high risk of DRPI, allowing for timely intervention to prevent or mitigate the condition's impact.

Abbreviations

DRPI	Device-related pressure injury
AI	Artificial intelligence
SHAP	SHapley Additive exPlanations
XGBoost	EXtreme Gradient Boosting
RF	Random forest
DTC	Decision tree classifier
SVM	Support vector machine
KNN	K-Nearest neighbors
LR	Logistic Regression
AUC	Area under the ROC curve
UMAP	Uniform Manifold Approximation and Projection
ML	Machine learning
PI	Pressure injury
ROSE	Random Over-Sampling Examples
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
PCA	Principal component analysis
EPV	Events Per Variable
SMOTE	Synthetic minority-sample oversampling technique

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-03090-9.

Supplementary Material 1

Acknowledgements

I would like to extend my sincere appreciation to Xiaolin Liu, who have generously shared her knowledge and expertise during the course of this project. Her valuable insights and discussions have greatly enriched this study and have helped shape its outcomes.

Author Contributions

Study design: HP, YQ. Data extraction: JC. Data clean: HH, MF. Data analysis and interpretation: YQ, YX. Data interpretation: YQ, YG. Manuscript preparation: YQ, JC, CH. Manuscript review: HP. Manuscript editing: YG, CH. All authors approved the final manuscript for submission.

Funding

This work was supported by Zhejiang Medical Health Science and Technology Project (2021KY180). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript

Data availability

The datasets used and analysed during the current study are available from the corresponding author on reasonable request. The analysis scripts are accessible at Github: https://github.com/phongying/DRPI_ML.

Declarations

Ethics approval and consent to participate

This study conformed to the ethical guidelines of the Declaration of Helsinki. The Sir Run Run Shaw Hospital Affiliated to Zhejiang University Ethics Review Committee approved this study (Ethics No. 20210907-32). All participant names and personal information have been removed. The need for Informed Consent was waived by the Sir Run Run Shaw Hospital Affiliated to Zhejiang University Ethics Review Committee because the study was retrospective in nature.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 June 2023 / Accepted: 24 June 2025
Published online: 09 July 2025

References

1. Demarré L, Van Lancker A, Van Hecke A, Verhaeghe S, Grypdonck M, Lemey J, et al. The cost of prevention and treatment of pressure ulcers: a systematic review. Int J Nurs Stud. 2015;52:1754–74.
2. Dang W, Liu Y, Zhou Q, Duan Y, Gan H, Wang L, et al. Risk factors of medical device-related pressure injury in intensive care units. J Clin Nurs. 2022;31:1174–83.
3. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden Scale for Predicting Pressure Sore Risk. Nurs Res. 1987;36:205–10.
4. Norton D, McLaren R, Exton-Smith AN. An investigation of geriatric nursing problems in hospital. Churchill Livingstone. 1975.
5. Waterlow J. Pressure sores: a risk assessment card. Nurs Times. 1985;81:49–55.
6. Gefen A, Alves P, Ciprandi G, Coyer F, Milne CT, Ousey K, et al. Device-related pressure ulcers: SECURE prevention. J Wound Care. 2020;29:S1–52.
7. Edsberg LE, Black JM, Goldberg M, McNichol L, Moore L, Sieggreen M. Revised National Pressure Ulcer Advisory Panel Pressure Injury Staging System. J Wound Ostomy Cont Nurs. 2016;43:585–97.
8. Haesler E. European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel, and Pan Pacific Injury Alliance. Prevention and Treatment of Pressure Ulcers/injuries: Clinical Practice Guideline the International Guideline EPUAP/NPIAP/PPPIA. 2019.
9. Ceelen KK, Stekelenburg A, Loerakker S, Strijkers GJ, Bader DL, Nicolay K, et al. Compression-induced damage and internal tissue strains are related. J Biomech. 2008;41:3399–404.
10. Gawlitta D, Oomens CWJ, Bader DL, Baaijens FPT, Bouten CVC. Temporal differences in the influence of ischemic factors and deformation on the metabolism of engineered skeletal muscle. J Appl Physiol. 2007;103:464–73.
11. Bullkich E, Kimmel E, Golan S. A novel ischemia reperfusion injury hereditary tissue model for pressure ulcers progression. Biomech Model Mechanobiol. 2019;18:1847–66.
12. Kasuya A, Sakabe J, Tokura Y. Potential application of in vivo imaging of impaired lymphatic duct to evaluate the severity of pressure ulcer in mouse model. Sci Rep. 2014;4:4173.
13. Coleman S, Nixon J, Keen J, Wilson L, McGinnis E, Dealey C, et al. A new pressure ulcer conceptual framework. J Adv Nurs. 2014;70:2222–34.
14. Labeau SO, Afonso E, Benbenishty J, Blackwood B, Boulanger C, Brett SJ, et al. Prevalence, associated factors and outcomes of pressure injuries in adult intensive care unit patients: the DecubICUs study. Intensive Care Med. 2021;47:160–69.
15. Jackson D, Sarki AM, Betteridge R, Brooke J. Medical device-related pressure ulcers: a systematic review and meta-analysis. Int J Nurs Stud. 2019;92:109–20.
16. Black JM, Cuddigan JE, Walko MA, Didier LA, Lander MJ, Kelpe MR. Medical device related pressure ulcers in hospitalized patients. Int Wound J. 2010;7:358–65.
17. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20:e262–73.
18. Anderson C, Bekele Z, Qiu Y, Tschannen D, Dinov ID. Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. BMC Med Inform Decis Mak. 2021;21:253.

19. Xu J, Chen D, Deng X, Pan X, Chen Y, Zhuang X, et al. Development and validation of a machine learning algorithm–based risk prediction model of pressure injury in the intensive care unit. Int Wound J. 2022;19:1637–49.
20. Nakagami G, Yokota S, Kitamura A, Takahashi T, Morita K, Noguchi H, et al. Supervised machine learning-based prediction for in-hospital pressure injury development using electronic health records: a retrospective observational cohort study in a university hospital in Japan. Int J Nurs Stud. 2021;119:103932.
21. Pei J, Guo X, Tao H, Wei Y, Zhang H, Ma Y, et al. Machine learning-based prediction models for pressure injury: a systematic review and meta-analysis. Int Wound J. 2023;20:4328–39.
22. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018;2:749–60.
23. Lundberg S, Lee S-I A Unified Approach to Interpreting Model Predictions. arXiv:170507874 [cs, stat]. 2017.
24. Chou R, Dana T, Bougatsos C, Blazina I, Starmer AJ, Reitel K, et al. Pressure ulcer risk assessment and prevention: a systematic comparative effectiveness review. Ann Internal Med. 2013;159:28–38.
25. Padula WV, Pronovost PJ, Makic MBF, Wald HL, Moran D, Mishra MK, et al. Value of hospital resources for effective pressure injury prevention: a cost-effectiveness analysis. BMJ Qual Saf. 2019;28:132–41.
26. Choi BK, Kim MS, Kim SH. Risk prediction models for the development of oral-mucosal pressure injuries in intubated patients in intensive care units: a prospective observational study. Journal of Tissue Viability. 2020;29:252–57.
27. Alderden J, Rondinelli J, Pepper G, Cummins M, Whitney J. Risk factors for pressure injuries among critical care patients: a systematic review. Int J Nurs Stud. 2017;71:97–114.
28. Raju D, Su X, Patrician PA, Loan LA, McCarthy MS. Exploring factors associated with pressure ulcers: a data mining approach. Int J Nurs Stud. 2015;52:102–11.
29. Dallı Ö E, İ C, Kelebek Girgin N. Incidence, characteristics and risk factors of medical device-related pressure injuries: an observational cohort study. Intensive Crit Care Nurs. 2022;69:103180.
30. Weber P, Weaver L, Miller C. Risk factors associated with medical device-related pressure injuries in the adult intensive care patient: a scoping review. Wound Pract Res. 2021;29:219–25.
31. Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: from Theory to Algorithms. Cambridge: Cambridge University Press. 2014.
32. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic Minority Over-sampling Technique. J Artificial Intelligence Res. 2002;16:321–57.
33. Haibo H, Bai Y, Garcia EA, Shutao L. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). Hong Kong, China: IEEE, 2008:1322–28.
34. Menardi G, Torelli N. Training and assessing classification rules with imbalanced data. Data Min Knowl Discov. 2014;28:92–122.
35. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery| nature Biomedical Engineering. https://www.nature.com/articles/s41551-018-0304-0. 12 Apr 2022.
36. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165:710–18.
37. Šateková L, Žiaková K, Zeleníková R. Predictive validity of the Braden Scale, Norton Scale, and Waterlow Scale in the Czech Republic. Int J Nurs Pract. 2017;23.
38. Anthony D, Parboteeah S, Saleh M, Norton PP. Waterlow and Braden scores: a review of the literature and a comparison between the scores and clinical judgement. J Clin Nurs. 2008;17:646–53.
39. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022;601:623–29.
40. Chen T, Guestrin C XGBoost: a Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 785–94.
41. Prospective A. Descriptive Study to Determine the Rate and Characteristics of and Risk Factors for the Development of Medical Device-related Pressure Ulcers in Intensive Care Units. Wound Manage Prevention. 2016;62.
42. Hajhosseini B, Longaker MT, Gurtner GC. Pressure Injury. Ann Surg. 2020;271:671–79.
43. Boyko TV, Longaker MT, Yang GP. Review of the Current Management of Pressure Ulcers. Adv Wound Care. 2018;7:57–67.
44. Black J, Alves P, Brindle CT, Dealey C, Santamaria N, Call E, et al. Use of wound dressings to enhance prevention of pressure ulcers caused by medical devices. Int Wound J. 2013;12:322-27.
45. Celik S, Taskin Yilmaz F, Altas G. Medical Device-Related Pressure injuries in adult intensive care units. J Clin Nurs. n/a n/a.
46. Levy JJ, Lima JF, Miller MW, Freed GL, O'Malley AJ, Emeny RT. Machine Learning Approaches for Hospital Acquired Pressure Injuries: a Retrospective Study of Electronic Medical Records. Frontiers in Medical Technology. 2022;4.
47. Mimura M, Ohura T, Takahashi M, Kajiwara R, Ohura Jr. N. Mechanism leading to the development of pressure ulcers based on shear force and pressures during a bed operation: influence of body types, body positions, and knee positions. Wound Repair Regen. 2009;17:789–96.

Publisher's Note

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