Scientific Reports

Article in Press

Interpretable machine-learning risk prediction of unplanned extubation among cancer patients with peripherally inserted central catheters

Yanying Yang, Yuwei Yang, Qiong Liu, Chenxi Li, Yin Long, Xuanxuan Li, Hongmei Liu & Run He

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Interpretable machine-learning risk prediction of unplanned extubation among cancer patients with peripherally inserted central catheters

Yanying Yang[1][†][*], Yuwei Yang[1], Qiong Liu[1][†], Chenxi Li[1][†], Yin Long[2], Xuanxuan Li[2], Hongmei Liu[1],Run He[1]

[1] Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang 621000, PR China

[2] School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China

[†] These authors contributed equally to this work.

* Corresponding Author:

Yanying Yang, E-mail: 2729433618@qq.com;

Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang 621000, PR China

ABSTRACT

Purpose: Unplanned extubation of peripherally inserted central catheters (PICC-UE) in patients with cancer has been linked to factors including women, diabetes, thrombosis history, valved catheter, double-lumen catheter, and self-management. However, the effect of patient quality of life has not been explored. Therefore, this study aimed to analyze PICC-UE risk factors using a machine-learning algorithm, focusing on the role of patient quality of life.

Methods: A total of 212 cancer patients who underwent PICC catheterization were included in this study from February 2021 to June 2022. Patients were categorized into two groups based on PICC-UE occurrence: the PICC-UE group (n = 23) and the non-PICC-UE group (n = 189). Referring to previous reports and professional cognition, data of 30 potential risk factors within one week before extubation were collected, with a focus on incorporating health-related quality of life (HRQOL) and patient self-management scores. PICC-UE risk factors were examined using four machine-learning algorithms with three encoding methods and four data imbalance processing methods. Then, the key factors causing PICC-UE were interpreted using the SHapley Additive exPlanations (SHAP) tool.

Results: PICC-UE occurred in 23 of 212 patients (overall incidence: 10.8%). The HRQOL score, which has been underexplored in prior studies, demonstrated a statistically significant difference between the PICC-UE and non-PICC-UE groups (P < 0.001) and exhibited a strong association with patient self-management score, as evidenced by its concentration in the upper right quadrant of the planar scatter plot. As a novel derivative composite metric, weighted quality of life (WQOL), calculated as the product of HRQOL and self-management scores, was identified as the most influential risk factor for PICC-UE, surpassing both individual self-management and HRQOL scores (SHAP-value=1.02 vs. 0.22 and 0.00). Furthermore, increased upper arm circumference was also found to be a significant predictor (SHAP value=0.22).

Conclusion: This study demonstrates the synergistic effect of patients' quality of life and self-management capacity on the PICC-UE occurrence. The WQOL metric, which integrates both factors, serves as a significant predictor of PICC-UE occurrence and should therefore be considered an essential component in clinical assessment.

1. INTRODUCTION

Peripherally inserted central catheters (PICCs) are widely used for patients with cancer undergoing long-term chemotherapy. These catheters reduce the need for multiple venous punctures and protect peripheral blood vessels and surrounding tissues from the damaging effects of chemotherapeutic agent extravasation, thereby reducing the risk of chemical phlebitis and tissue necrosis [1-3]. Despite these significant advantages of PICC, unplanned catheter removal remains the most undesirable accident during the intended catheterization period [4-8]. PICC unplanned extubation (PICC-UE) refers to the premature removal of the catheter during the intended catheterization period due to severe complications or unintended dislodgement resulting from patient-related or operator-related factors [13]. The reported incidence of PICC-UE varies considerably across studies, with rates ranging from 2.5% to 40.7%, predominantly observed in cancer patients undergoing chemotherapy and requiring long-term catheterization [9]. PICC-UE not only exacerbates patient discomfort and imposes additional economic burdens but also disrupts post-extubation planned treatment, thus seriously

affecting patients' quality of life and, in some cases, posing a threat to their lives [10,11]. Therefore, identifying risk factors for PICC-UE is crucial to improving the safety and effectiveness of PICC use.

Previous studies have identified several risk factors for PICC-UE, including women, diabetes patients, a history of thrombosis, elevated D-dimer levels, undergoing targeted therapy, experiencing infusion failure, and using a valved or double-lumen catheter [12]. Self-management has also been recognized as a personal factor that may lead to PICC-UE after patients return home [12]. However, to date, no study has incorporated health-related quality of life (HRQOL) as a potential risk factor under investigation. Clinical observations indicate that discharged patients with a lower quality of life are more susceptible to PICC-UE events, primarily due to diminished capacity to adhere to recommended catheter care protocols. Therefore, PICC self-management ability and HRQOL should be considered complementary factors during the follow-up period. It is essential to conduct comprehensive studies on these risk factors to enable in-depth analysis and provide empirical

evidence for developing effective strategies for preventing PICC-UE.

To systematically evaluate the contributions of various risk factors, this study applied four machine learning (ML) algorithms, recognized as an advanced data-driven approach to intelligent analysis and prediction [13], to identify the most significant risk factors associated with PICC-UE. Given the low incidence of PICC-UE reported in the majority of existing studies and the resulting category imbalance in the dataset, the focal loss (FL) method was further employed to optimize the training loss function and improve algorithm performance. Finally, the SHapley Additive exPlanations (SHAP) tool was employed to provide a quantitative and visually interpretable assessment of the relationship between each risk factor and PICC-UE outcome.

2. METHODS

2.1 Patients and grouping

During February 2021 and June 2022, cancer patients who had undergone PICC catheterization were consecutively enrolled from the Intravenous Catheter Care Clinic of Mianyang Central

Hospital, School of Medicine, University of Electronic Science and Technology of China. Each patient was prospectively observed and followed up weekly until catheter removal. The inclusion criteria comprised patients: (i) who met the indications for and successfully underwent PICC placement, (ii) with placement sites in the upper arm, (iii) with a histopathological diagnosis of malignant tumor, and (iv) who participated voluntarily and provided written informed consent (patients or their family). The exclusion criteria comprised patients: (i) aged <18 years, (ii) with incomplete medical records, (iii) who did not receive maintenance care at our hospital after catheter placement, (iv) with mental or communication disorders, and (v) who discontinued treatment or died during the study period.

The primary outcome, PICC-UE, was defined as PICC-UE owing to any PICC-related complication or accidental dislodgement by the patient during healthcare, necessitating the insertion of new vascular access for treatment. According to the presence or absence of PICC-UE, patients were categorized into the PICC-UE group and the non-PICC-UE group.

2.2 Quality control of PICC catheterization

Specially trained intravenous infusion specialist nurses performed PICC placement and maintenance according to standard procedures. Following catheterization, the position of the catheter tip was verified by chest X-ray. Throughout the treatment period, catheter maintenance was performed by oncology nurses who had undergone standardized training. All procedures during chemotherapy intervals were carried out by specialized nurses from the intravenous therapy center. For skin disinfection during catheter insertion and routine maintenance, iodine tincture (manufactured by Shanghai Licon Disinfection High-Tech Co., Ltd.) was used, and 3M transparent dressings were applied as wound coverings. The researchers conducted prospective observations with weekly follow-ups for each patient until catheter removal, which occurred upon completion of treatment, unplanned removal, or death.

2.3 Compilation of data collection form

Based on literature review and clinical expertise, the potential influencing factors of PICC-UE were categorized and summarized to develop a data collection form. This form comprise four categories and 30 indicators: 1) sociodemographic information, including age, gender, height,

weight, body mass index (BMI), education level, marital status, hypertension, diabetes, drug allergy history, left-sided PICC placement history, right-sided PICC placement history, and clinical diagnosis; 2) PICC placement data, including patient handedness, diameter of punctured vein, arm circumference, insertion length, exposed catheter length, current PICC placement side (left or right arm), number of puncture attempts, number of catheter lumens, and punctured vein; 3) recent laboratory data one week prior to extubation, including albumin (ALB), white blood cell count (WBC), platelet count (PLT), international normalized ratio (INR), prothrombin time (PT), and activated partial thromboplastin time (APTT); 4) on-site investigation data collected at extubation, including the scores of Cancer Patients PICC Self-management Scale and Health-related Quality of Life (HRQOL) questionnaire.

2.4 Data acquisition

The sociodemographic data of cancer patients were collected by researchers from electronic medical records, as well as the recent laboratory data one week before extubation. The PICC placement data were measured or recorded by the specialist nurse in the Venous Catheter Care Outpatient Clinic. The PICC

self-management and HRQOL scores within one week before extubation were collected via face-to-face interviews conducted by the researchers at the time of extubation.

2.5 Questionnaires

Cancer Patients PICC Self-management Scale: This scale, developed by Ling Li et al. [14], was used to reflect each patient's PICC self-management ability through seven dimensions (each dimension included five items): daily life with PICC, exercise with PICC, adherence to PICC maintenance, daily catheter observation, handling of abnormalities, information acquisition for PICC maintenance, and PICC management confidence.

HRQOL questionnaire: The five-dimensional EuroQol scale (EQ-5D) [15] was used to reflect patient HRQOL through the following five dimensions (each dimension included three items): mobility, self-care, daily activities, pain/discomfort, and anxiety/depression. Patients selected one level for each dimension according to their condition that day. A time trade-off conversion table was used to calculate the composite score of the five dimensions, i.e., the EQ-5D index, with values ranging

between -.11 and 1.00, with scores closer to 1 indicating better health.

2.6 Data analysis

SPSS 26.0 and Python 1.2.1 software were used for statistical analysis. Measurement data were described using M (P25, P75) after a Shapiro–Wilk test indicated that the data were non-normally distributed, and a Mann–Whitney U test was used for difference analysis. Count data are described as n (%), and a Pearson's chi-square test was used for difference analysis. Point prevalence was employed, which refers to the proportion of individuals presenting a clinical condition at a specific time. This was calculated using the following formula:

Prevalence (%) = the number of patients with PICC-UE / total number of evaluated patients × 100.

To assess potential multicollinearity introduced by the interaction term (WQOL), the Variance Inflation Factor (VIF) was calculated. To mitigate structural multicollinearity between the product term and its individual components, we applied mean-centering to the HRQOL and self-management scores

before calculating WQOL. A VIF value of less than 5 was considered indicative of no severe multicollinearity.

A machine-learning algorithm was implemented in the sci-kit-learn Python library (version 1.2.1) and XGBoost (version 1.7.3) software. If $p<0.050$ was considered significant. Given the possible low incidence of PICC-UE and the extensive number of risk factor indicators collected, this study employed four machine learning algorithms with three variable encodings to improve the accuracy and robustness of risk factor identification. To address the challenge of data imbalance arising from the possible rare occurrence of PICC-UE, four sampling strategies for mitigating class imbalance were incorporated. Furthermore, to better explain the contribution of each identified risk factor, the SHAP (SHapley Additive exPlanations) tool was applied to provide both case-specific and global model interpretations.

Given the imbalance of the dataset (23 positive events), we employed Stratified 5-Fold Cross-Validation during the model evaluation phase to prevent overfitting and ensure the generalizability of the results. Furthermore, to assess the reliability of the predicted probabilities, we performed a

calibration analysis. The agreement between predicted probabilities and observed outcomes was evaluated using the Brier score and visualized via calibration plots.

3. RESULTS

3.1 Occurrence of PICC-UE in cancer patients

During the study period, 278 patients with cancer were admitted for PICC placement, and based on the inclusion and exclusion criteria, 212 patients were enrolled. The patient enrollment process is shown in Figure 1.

The PICC-UE occurred in 23 of 212 patients (overall incidence: 10.8%). The primary causes of PICC-UE were accidental catheter dislodgement (6.6%, 14/212), catheter occlusion (2.4%, 5/212), severe contact dermatitis (1.4%, 3/212), and suspected catheter-associated bloodstream infections (0.5%, 1/212).

3.2 Synergistic effect of HRQOL and self-management

In the planar scatter plot depicting self-management and HRQOL scores, non-PICC-UE patients were predominantly clustered in the upper right quadrant (Figure 2A), which is strongly suggestive of the ability to differentiate non-PICC-UE

from PICC-UE. Therefore, we constructed a novel feature metric using the product of the self-management and HRQOL scores, named weighted quality of life (WQOL). Collinearity diagnostics showed that after mean-centering, the VIF values for WQOL, HRQOL, and self-management were all reduced to below 2.5 (ranging from 1.11 to 2.37, see Table 1), indicating that multicollinearity was effectively eliminated and would not inflate the feature importance estimates in the subsequent machine learning models. Further covariance analysis found that PICC-UE had higher goodness of fit with respect to self-management, HRQOL, and their synergistic effect (that is, WQOL) compared to models involving self-management and HRQOL alone (R^2=0.93 vs 0.89, ΔR^2=0.04) (Figure 2B and 2C). The findings suggested that WQOL was a comprehensive indicator integrating self-management and HRQOL, and can effectively predict the occurrence of PICC-UE.

3.3 Univariate association of each factor with PICC-UE occurrence

Univariate analysis was performed on the 31 characteristic variables; 13 factors were significantly different between the

PICC-UE and non-PICC-UE groups, including marital status, PICC history of left arm, PICC history of right arm, drug allergy history, body mass index, diagnosis, puncture vein, puncture vein diameter, arm circumference, exposed length, self-management, HRQOL, and WQOL (all $p < 0.100$) (Table 2).

3.4 Optimal independent association based on four machine-learning algorithms

The corresponding comparative analysis excluding WQOL incorporation showed that: 1) Among three mainstream encoding methods, the WoE encoding was superior to the One-Hot and CE encodings (AUC=0.99 vs 0.97 and 0.94) (Figure 3A); 2) Among four imbalance-data processing methods, the FL method was superior to NearMiss, ENN, and SMOTE methods (AUC=0.99 vs 0.91–0.95) (Figure 3B); 3) Among four machine-learning algorithms (RF, LR SVM, and XGB), the XGB had the highest AUC and recall (0.99 and 0.87, respectively) (Figure 3C). Decision curve analysis also showed that the WoE encoding (Figure 3E), FL imbalance-data processing (Figure 3F), and XGB algorithm (Figure 3G) had broader net benefits among the corresponding comparisons. So, the XGB algorithm with WoE

encoding and FL method had a better ability to identify risk factors. Based on the above-mentioned discovery, further comparison between the XGB algorithm (or XGB+FL algorithm) and its combination incorporating WQOL (i.e., XGB+FC algorithm or XGB+FL+FC algorithm) indicated that the inclusion of WQOL enhanced the prediction of PICC-UE occurrence (AUC=0.993 vs. 0.985 or 0.994 vs. 0.976; recall=0.783 vs. 0.739 or 0.870 vs. 0.783) (Figure 3D and Table 3), with a broader net benefit (Figure 3H). It is evident that the predictive role of WQOL in relation to PICC-UE should not be overlooked. Besides，the calibration analysis yielded a **Brier score of 0.03**, indicating a low discrepancy between predicted probabilities and actual outcomes. As shown in **Figure 4 (Calibration Plot)**, the calibration curve closely follows the ideal diagonal line, suggesting that the model provides reliable probability estimates and is not suffering from significant overfitting.

3.5 Contribution explanation of each factor to the occurrence of PICC-UE

To further clarify the contribution explanation of each risk factor, this study employed the SHAP tool for feature interpretation. Figure 5A shows the SHAP summary plot of the feature importance ranking of various significant factors for PICC-UE risk. Figure 5B shows the three highest-ranked feature importance factors. WQOL played a crucial role in PICC-UE risk, followed by self-management and upper arm circumference on the PICC side, with SHAP values of 1.02, 0.22, and 0.22, respectively.

Figures 5C and 5D illustrated the SHAP waterfall plots to interpret the feature effects of all factors on PICC-UE risk. The arrow size of the waterfall represented the degree of impact of each factor on PICC-UE risk in an individual, and the color identified the increasing or decreasing risk (red or blue). Figure 5C depicted a non-PICC-UE individual, with a total score of -1.301, to predict a low risk of PICC-UE occurrence, while Figure 5D depicted a PICC-UE individual, with a total score of 1.45, to predict a high risk of PICC-UE occurrence. Notably, only WQOL demonstrated significantly discriminatory power between the two completely different individuals (SHAP-value=-0.98 and 1.67) despite having identical self-management scores (both

SHAP-value=0.36) and similar interpretability of upper arm circumference on the PICC side (SHAP-value=-0.17 and -0.07).

4. DISCUSSION

The reported PICC-UE incidence varies considerably, ranging from 7.5% to 22.0% in China [16] and 2.5% to 40.7% in other countries [7]. Our results showed a 10.8% PICC-UE incidence, which is lower than that reported in most previous studies [7,16]. Variability in PICC-UE incidence across studies is likely attributable to differences in study populations and sample sizes. PICC maintenance often extends beyond the hospital setting into long-term home care for the majority of patients. In the home environment, the individual patient's ability plays a crucial role in PICC-UE risk. Among patient-specific factors, quality of life reflects a patient's functional capacity to perform certain activities, whereas self-management pertains to the actual execution of those activities. While prior research has highlighted the role of self-management in influencing PICC-UE outcomes, the association between quality of life and PICC-UE remains insufficiently examined. Our study found a significant correlation between quality of life and self-management, which

accounts for the historical neglect of quality of life in previous studies. Furthermore, patient ability and willingness to engage in self-care tasks vary, encompassing four distinct categories: capable and willing, capable but unwilling, incapable but willing, and incapable and unwilling. PICC-UE risk likely varies across these scenarios. Therefore, we integrated patients' quality of life and self-management, introducing WQOL to evaluate self-ability comprehensively. Our machine learning-based optimization analysis demonstrated that WQOL exhibited significantly greater feature importance compared to other indicators, serving as a more robust predictor of PICC-UE. In addition, the machine learning-based optimization algorithm also identified two other risk factors with relatively lower contributions, namely self-management and upper arm circumference. In the two typical non-PICC-UE and PICC-UE cases presented, the SHAP tool accurately interpreted the positive relationship between WQOL and PICC-UE risk. Notably, in both typical cases, good self-management was associated with a risk reduction, whereas upper arm circumference was linked to an increased risk. In contrast, WQOL, which integrates

assessments of quality of life and self-management, demonstrates greater practical utility in predicting PICC-UE. Conventional wisdom has long held that there is no direct association between quality of life and PICC-UE. However, our SHAP analysis demonstrates that the WQOL significantly influences PICC-UE occurrence. Specifically, higher WQOL values correspond to higher SHAP values and are associated with a lower PICC-UE probability. Our finding shows a negative correlation between WQOL and PICC-UE, and demonstrate a notable influence on PICC-UE occurrence. In this study, five cases with low WQOL (manifested as high self-management ability and low HRQOL) experienced significant declines in their activity and daily living abilities, hindering timely hospital visits for maintenance. This led to tube blockages and eventual PICC-UE. Therefore, in clinical decision-making, it is advisable to provide on-site services by intravenous therapy specialists for patients with low WQOL who cannot visit the hospital. This ensures timely medical assistance.

Better self-management is essential to prevent complications in patients taking catheters home during chemotherapy intervals. These patients return home during the interval between

chemotherapy sessions. They are required to engage in PICC self-management, such as monitoring the puncture site for bleeding, curling of the transparent dressing, and catheter dislodgement. They also need to perform reasonable functional exercises for the upper limb on the PICC side [17]. Similar to our research, [12] Jiang L et al. also observed that patients with PICC exhibit low levels of catheter self-management at home, leading to a high incidence of PICC-UE. In our study, self-management's SHAP value of 0.22 ranked second. Of 12 patients with home catheters, some experienced catheter removal due to inadequate protection during bathing and dressing, resulting in PICC-UE. This suggests patients frequently neglect catheter maintenance or lack the skills for proper self-care. Thus, it is crucial for nursing staff to provide comprehensive health education on PICC self-maintenance and train patients in necessary skills.

This study innovatively identified upper arm circumference at the PICC insertion site, which was underemphasized in previous research, as the third most significant factor associated with an increased risk of PICC-UE. Among the 23 patients who developed PICC-UE in our study, 14 were women aged 55 to 92

years with arm circumferences ranging from 27 and 32 cm. This indicates that PICC-UE occurred in more than 50% of older women with larger arm circumferences. Loose skin and greater upper arm circumference in older women may lead to greater local tissue mobility during activities owing to gravity, increasing the risk of dressing misalignment or curling, which can cause catheter dislodgement. Therefore, nursing staff should be alert to dislodgement risks in older women with larger arm circumferences undergoing PICC placement. We recommend using catheter immobilizers and advising such patients to avoid strenuous upper-limb activities to minimize the occurrence of PICC-UE.

Our study had several strengths. We used machine learning algorithms to optimize PICC-UE risk assessment, highlighting for the first time WQOL's significant impact on PICC-UE as a measure of patients' comprehensive abilities. To address data imbalance, we improved the focal loss function, ensuring greater attention to rare PICC-UE cases. Additionally, the SHAP tool interpreted the prediction model from both global and local perspectives.

Limitations: This study had some limitations. As a single-center study, the conclusions require further external validation to assess their clinical utility. While the newly constructed metric in this study serves as a predictor of PICC-UE risk, PICC self-management ability, and patient HRQOL are dynamic in nature. Thus, these factors must be assessed continuously, and interventions should be individualized. Given the considerable variability in HRQOL during long-term PICC use among cancer patients, this study conducted HRQOL assessments one week before catheter removal. Further research is warranted to evaluate alternative assessment time points. Of the two factors in relation to WQOL, only self-management demonstrated a negative association with PICC-UE, suggesting a potential reverse causality between HRQOL measured one week before removal and the occurrence of PICC-UE. However, it remains unclear whether this association is influenced by self-management behaviors. Future research should focus on multicenter studies with broader populations and incorporate more clinical risk factors to better guide clinical decision-making.

5. Conclusion

PICC-UE occurrence in patients with cancer and PICC is linked to poor clinical outcomes and increased healthcare resource use. Therefore, this study developed a newly constructed key metric (i.e., WQOL) to facilitate early PICC-UE risk prediction in such patients. The clinical applicability of the XGB algorithm-based PICC-UE risk prediction requires testing in actual clinical practice. Furthermore, the SHAP tool's interpretability suggests WQOL has significant potential for future clinical application and could aid nursing staff in the early and dynamic identification of PICC-UE risk.

Declarations

Authors' contributions

Yanying Yang: data curation, formal analysis, Funding acquisition, methodology, writing - original draft;

Yuwei Yang: data curation, formal analysis, writing - review;

Qiong Liu: Conceptualization, writing - review;

Chenxi Li: investigation;

Yin Long: methodology;

Xuanxuan Li: methodology;

Hongmei Liu: methodology, Investigation;

Run He: Investigation;

Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this article.

Ethical Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Mianyang Central Hospital approved this study (protocol code: S2018085, date of approval: 2018). Written informed consent was obtained from all participating individuals.

Consent to Publish declaration

The Author confirms that the work described has not been published before; that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors, if any; that its publication has been approved tacitly or explicitly by the responsible authorities at the institution where the work is carried out.

Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] Johansson E, Hammarskjöld F, Lundberg D, Arnlind MH. Advantages and disadvantages of peripherally inserted central venous catheters (PICC) compared to other central venous lines: a systematic review of the literature. Acta Oncol. 2013 Jun;52(5):886-92.

https://doi.org/10.3109/0284186X.2013.773072.

[2] Glauser F, Kivrak S, Righini M. Cathéters centraux insérés par voie périphérique,Indications, contre-indications et complications [Peripherally inserted central catheters : indications, contraindications, complications]. Rev Med Suisse. 2018 Dec 5;14(630):2211-2213. French.

[3] Sakai T, Kohda K, Konuma Y, Hiraoka Y, Ichikawa Y, Ono K, et al. Horiguchi H, Tatekoshi A, Takada K, Iyama S, Kato J. A

role for peripherally inserted central venous catheters in the prevention of catheter-related bloodstream infections in patients with hematological malignancies. Int J Hematol. 2014 Dec;100(6):592-8. https://doi.org/10.1007/s12185-014-1677-9.

[4] Duwadi S, Zhao Q, Budal BS. Peripherally inserted central catheters in critically ill patients - complications and its prevention: A review. Int J Nurs Sci. 2018 Dec 21;6(1):99-105. https://doi.org/10.1016/j.ijnss.2018.12.007.

[5] Al-Asadi O, Almusarhed M, Eldeeb H. Predictive risk factors of venous thromboembolism (VTE) associated with peripherally inserted central catheters (PICC) in ambulant solid cancer patients: retrospective single Centre cohort study. Thromb J. 2019 Jan 25;17:2. https://doi.org/10.1186/s12959-019-0191-y.

[6] Gao W, Luan XR, Sun YY, Zhang M, Li K, Li QH, et al. Experiences of patients with abnormal extubation of PICC tubes: a qualitative study. Int J Clin Exp Med. 2015 Oct 15;8(10):19297-303.

[7] Silva PS, Reis ME, Aguiar VE, Fonseca MC. Unplanned extubation in the neonatal ICU: a systematic review, critical

appraisal, and evidence-based recommendations. Respir Care. 2013 Jul;58(7):1237-45. https://doi.org/10.4187/respcare.02164.

[8] Li P, Sun Z, Xu J. Unplanned extubation among critically ill adults: A systematic review and meta-analysis. Intensive Crit Care Nurs. 2022;70:103219. https://doi.org/10.1016/j.iccn.2022.103219.

[9] Mielke D, Wittig A, Teichgräber U. Peripherally inserted central venous catheter (PICC) in outpatient and inpatient oncological treatment. Support Care Cancer. 2020;28(10):4753-4760. https://doi.org/10.1007/s00520-019-05276-0.

[10] Chan RJ, Northfield S, Larsen E, Mihala G, Ullman A, Hancock P, et al. Central venous Access device SeCurement And Dressing Effectiveness for peripherally inserted central catheters in adult acute hospital patients (CASCADE): a pilot randomised controlled trial. Trials. 2017;18(1):458. https://doi.org/10.1186/s13063-017-2207-x.

[11] Kang JR, Long LH, Yan SW, Wei WW, Jun HZ, Chen W. Peripherally Inserted Central Catheter-Related Vein Thrombosis in Patients With Lung Cancer. Clin Appl Thromb Hemost. 2017;23(2):181-186. https://doi.org/10.1177/1076029615595880

[12] Liu X, Tao S, Ji H, Chen S, Gu Y, Jin X. Risk factors for peripherally inserted central catheter (PICC)-associated infections in patients receiving chemotherapy and the preventive effect of a self-efficacy intervention program: a randomized controlled trial. Ann Palliat Med. 2021 Sep;10(9):9398-9405. doi: 10.21037/apm-21-1848.

[13] Zhang J, Ma G, Peng S, Hou J, Xu R, Luo L, et al. Risk Factors and Predictive Models for Peripherally Inserted Central Catheter Unplanned Extubation in Patients With Cancer: Prospective, Machine Learning Study. J Med Internet Res. 2023;25:e49016. Published 2023 Nov 16. https://doi.org/10.2196/49016 .

[14] Li L, Gao Z, Guo L. [Comfort Status and Its Influencing Factors in Patients With Gastrointestinal Cancer After PICC Catheterization]. Hu Li Za Zhi. 2020;67(6):32-39. https://doi.org/10.6224/JN.202012_67(6).06.

[15] Oliveira JS, Hayes A. Clinimetrics: The EuroQol-5 Dimension (EQ-5D). J Physiother. 2020;66(2):133. https://doi.org/10.1016/j.jphys.2020.02.012.

[16] Kang J, Chen W, Sun W, Ge R, Li H, Ma E, et al. Peripherally inserted central catheter-related complications in

cancer patients: a prospective study of over 50,000 catheter days. J Vasc Access. 2017;18(2):153-157. https://doi.org/10.5301/jva.5000670.

[17] Wang Y, Li J, Wang Y, Wang L, Xiang Y, Huang M, et al. The influential factors and intervention strategies that engage malignant cancer patients in health-promoting behaviors during PICC line maintenance. Am J Transl Res. 2021;13(5):5208-5215. Published 2021 May 15.

Figure Legends

Figure 1. Flow Chart of Patient Enrollment, Study Participation and Data Analysis

Note: PICC, peripherally inserted central catheter; PICC-UE, unplanned extubation of peripherally inserted central catheters.

Figure 2. Feature Construction

Note: A) The distribution of the sample in a two-d imensional planar plot of self-management and HRQOL total scores; B) The effect of self-management and HRQOL indicators on subjects; C) The effect of self-management, HRQOL and Self-management* indicators on subjects. HRQOL, health-related quality of life; PICC-UE, unplanned extubation.

Figure 3. Optimal Multiple Analysis Based on Machine-learning Algorithm

Note: A) Comparison of three mainstream variable encoding methods; B) Comparison of four methods for data imbalance processing; C) Comparison of four ML models; D) Validation of the advantages and disadvantages of the four XGB models; E) Decision curve analysis of three mainstream variable encoding

methods; F) Decision curve analysis of four methods for data imbalance processing; G) Decision curve analysis of four ML models; H) Decision curve analysis of the four XGB models. WoE, weight of encoding; CE, counting encoder; ENN, edited nearest neighbor; SMOTE, synthetic minority over-sampling technique; XGB, extreme gradient boosting; SVM, support vector machine; FC, feature construction; FL, focal loss.

Figure 4. Calibration Plot of the Predictive Model.

Figure 5. SHAP Interpretability analysis

Note: A) SHAP summary plot of the feature importance ranking; B) The three highest ranked feature importance metrics; C) The waterfall plots from the SHAP analysis for one non-PICC-UE patients; D) The waterfall plots from the SHAP analysis for one PICC-UE patients. WQOL, weighted quality of life; HRQOL, health-related quality of life; BMI, body mass index; SHAP, shapley additive explanations.

```
┌─────────────────────────────────┐
│  patients with PICC catheterization  │
│            (n=278)              │
└─────────────────────────────────┘
```

```
┌──────────────────────────────┐        ┌──────────────────────────────┐
│ Patients with non-cancer PICC │        │ Patients with cancer PICC    │
│      catheterization          │        │      catheterization         │
│         (n=24)                │        │         (n=254)              │
└──────────────────────────────┘        └──────────────────────────────┘
```

┌──────────────────────────────────┐
│ Eleven patients with age ≤18 years │
│ were excluded   │
└──────────────────────────────────┘

┌──────────────────────────────────┐
│ Two patients with psychiatric   │
│ abnormalities were excluded   │
└──────────────────────────────────┘

┌──────────────────────────────────┐
│ Three patients with communication │
│ disorders were excluded   │
└──────────────────────────────────┘

┌──────────────────────────────────┐
│ Informed consent was signed   │
│ in 234 cases   │
└──────────────────────────────────┘

```
┌──────────────────┐                    ┌──────────────────┐
│   Non-PICC-UE    │                    │     PICC-UE      │
└──────────────────┘                    └──────────────────┘
```

┌──────────────────────────────────┐   ┌──────────────────────────────────┐
│ Seven cases of catheter maintenance │   │ Nine cases of catheter maintenance │
│ not performed according to the   │   │ not performed according to the prescribed │
│ prescribed time were excluded   │   │ time were excluded   │
└──────────────────────────────────┘   └──────────────────────────────────┘

┌──────────────────────────────────┐   ┌──────────────────────────────────┐
│ One cases with death were excluded │   │ One cases with death were excluded │
└──────────────────────────────────┘   └──────────────────────────────────┘

┌──────────────────────────────────┐   ┌──────────────────────────────────┐
│ Two cases with incomplete medical │   │ Two cases with incomplete medical │
│ records were excluded   │   │ records were excluded   │
└──────────────────────────────────┘   └──────────────────────────────────┘

```
┌──────────────────┐                    ┌──────────────────┐
│    189 cases     │                    │     23  cases    │
└──────────────────┘                    └──────────────────┘
```

```
┌──────────────────────────┐
│     Included(n=212)       │
│   Non-PICC-UE(n=189)      │
│      PICC-UE(n=23)        │
└──────────────────────────┘
```

Tests of Between-Subjects Effects

Dependent Variable: PICC-UE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	18.303[a]	55	.333	23.577	.000
Intercept	3.431	1	3.431	243.120	.000
Self_management	11.911	39	.305	21.639	.000
HRQOL	2.739	15	.183	12.936	.000
Error	2.202	156	.014		
Total	23.000	212			
Corrected Total	20.505	211			

a. R Squared = .893 (Adjusted R Squared = .855)

B

Tests of Between-Subjects Effects

Dependent Variable: PICC-UE

| Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	19.005[a]	86	.221	18.415	.000
Intercept	4.508	1	4.508	375.687	.000
Self_management	11.568	39	.297	24.718	.000
HRQOL	2.130	15	.142	11.832	.000
Self_management * HRQOL	.702	31	.023	1.887	.008
Error	1.500	125	.012		
Total	23.000	212			
Corrected Total	20.505	211			

a. R Squared = .927 (Adjusted R Squared = .877)

C

Calibration Curve
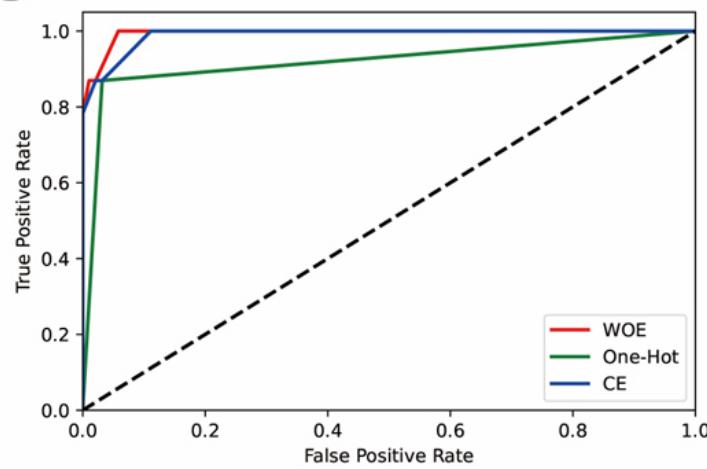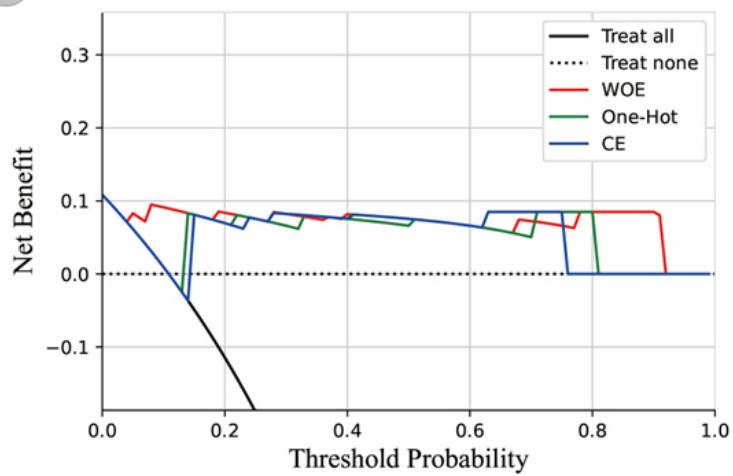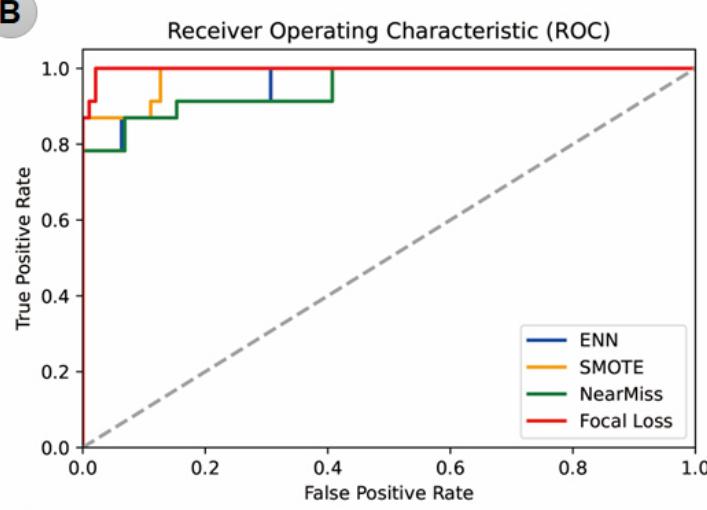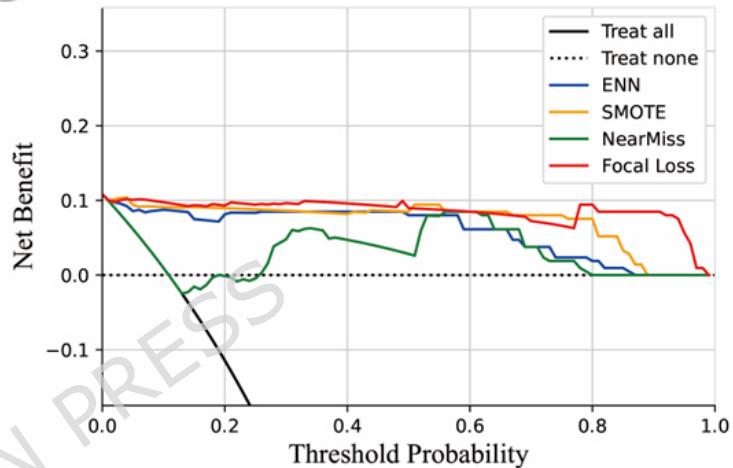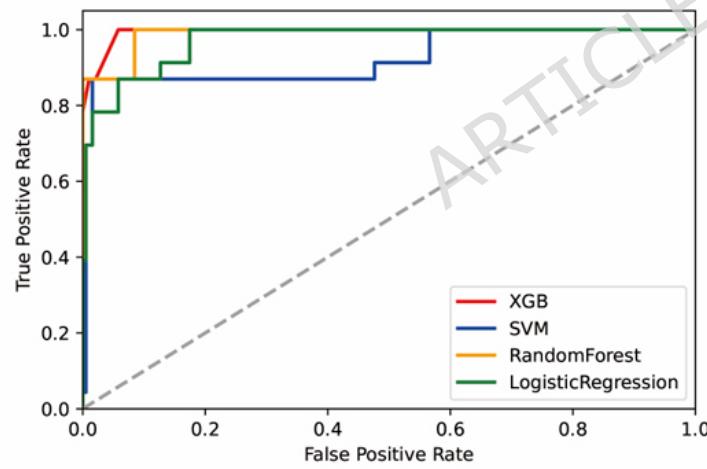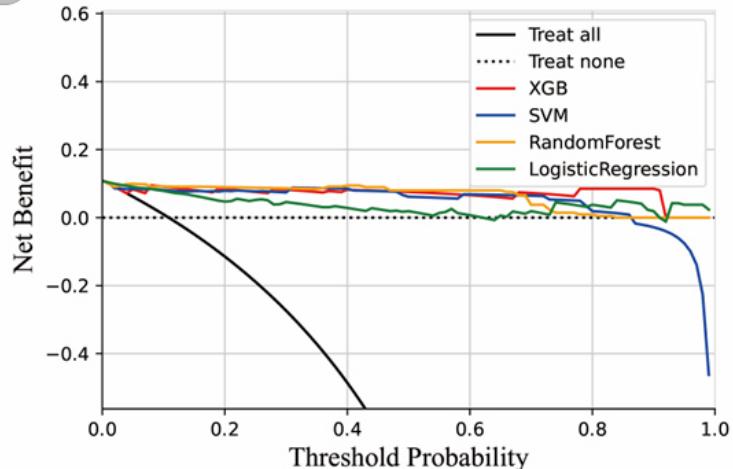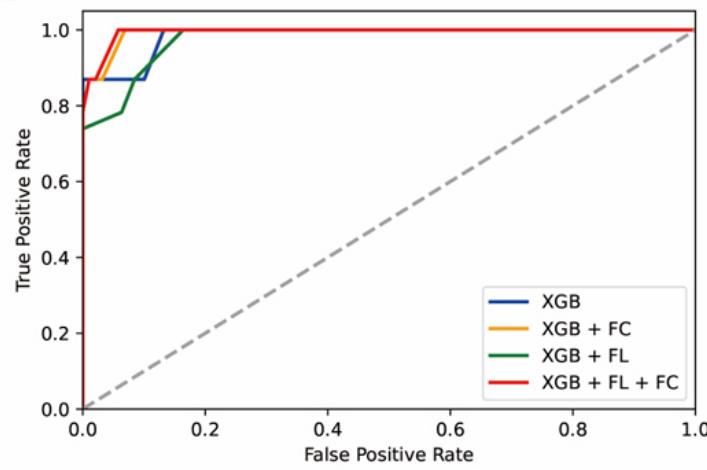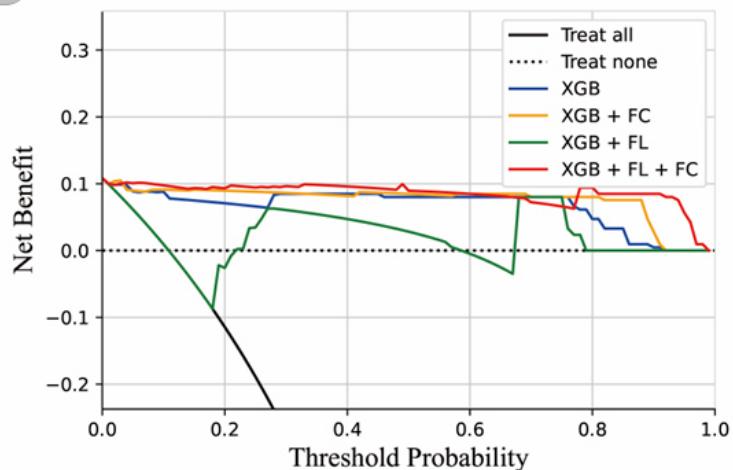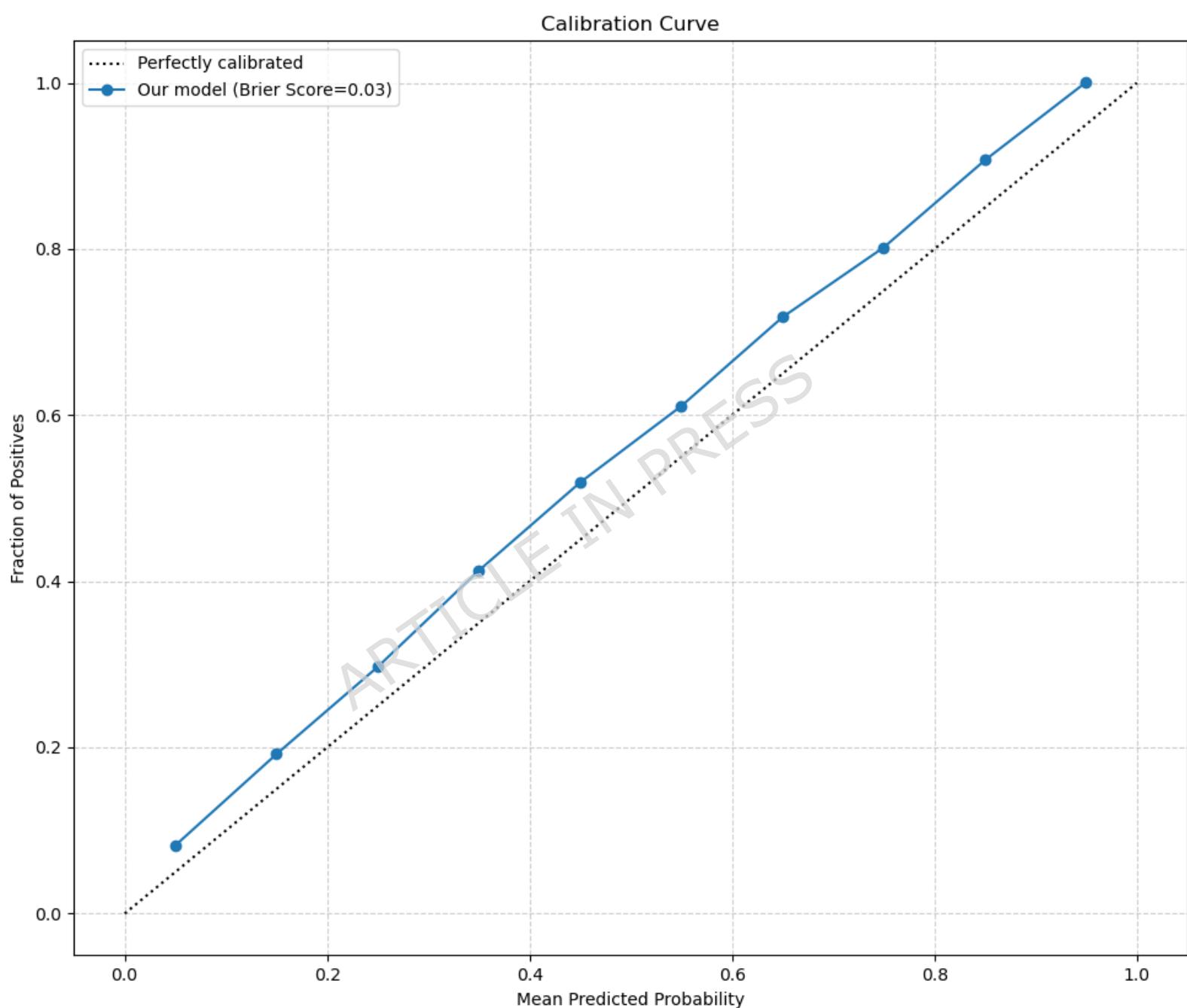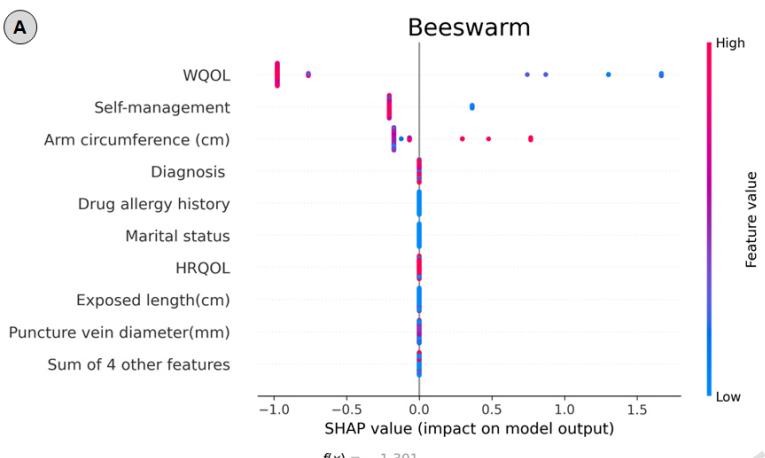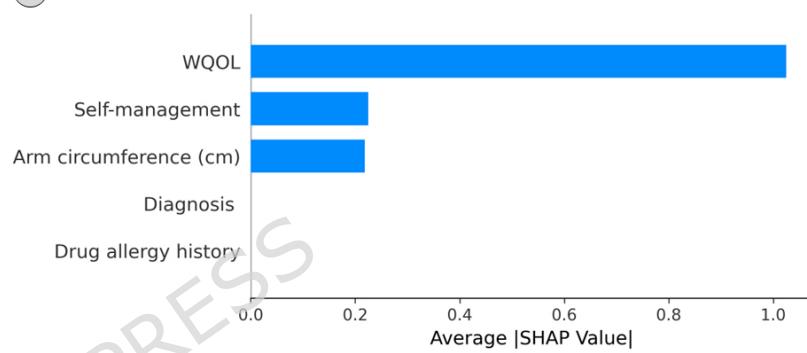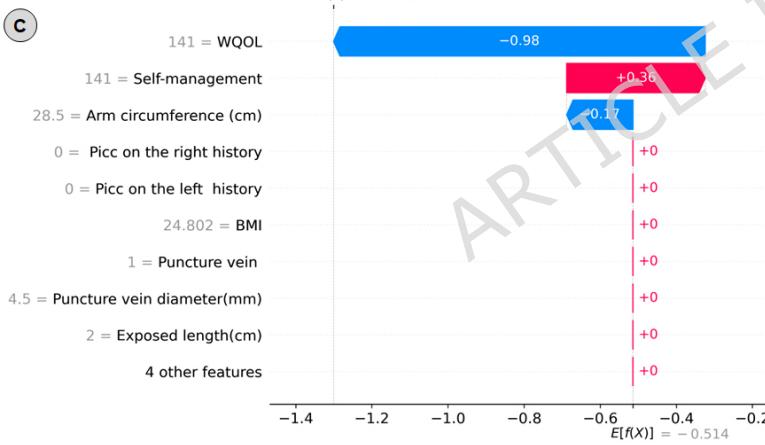
Table 1. Collinearity assessment of risk factors (Variance Inflation Factor) before and after mean-centering

Variable	VIF (Before Centering)	VIF (After Centering)
Self-management	29.62	1.11
HRQOL	72.04	1.15
WQOL (Interaction)	118.41	1.17
Arm Circumference	2.37	2.37
BMI	2.34	2.34

Table-2 PICC-UE single factor analysis (n = 212)

Observed factors	Non-UE (n=189)	UE (n=23)	χ^2/Z	P
Age (years)	55(50~63)	56(48~68)	-0.71	0.478
Gender			0.12	0.724
Men [n(%)]	56(29.6)	6(26.1)		
Women [n(%)]	133(70.4)	17(73.9)		
Height (cm)	158(155~164)	158(150~163)	-0.66	0.507
Weight (kg)	59(51~65)	60(55~64)	-1.21	0.305
BMI (kg/m2)	22.49(20.73~25.10)	24.44(21.83~28.44)	-1.84	0.066
Drug allergy history	3(1.6)	2(8.7)	4.50	0.034
Diagnosis			14.18	0.015
Breast cancer [n(%)]	99(52.4)	19(82.6)		
Lung cancer [n(%)]	5(2.5)	0(0.0%)		
Digestive System Cancer [n(%)]	58(30.7)	0(0.0%)		
Gynecologic Cancer [n(%)]	9(4.8)	0(0.0%)		
Hematological malignancies [n(%)]	2(1.1)	1(4.3)		
Others [n(%)]	16(8.5)	3(13.1)		
Education leve			1.88	0.171
Middle school or below [n(%)]	130(68.8)	19(82.6)		
High school or above [n(%)]	59(31.2)	4(7.4)		
Puncture vein diameter(mm)	5(4.5~6)	5(4~5.6)	-2.02	0.043

Arm circumference(cm)	26(24~28)	28(25~30)	-2.39	0.017
Insertion length(cm)	38(36~40)	38(37~39)	-0.61	0.541
Exposed length(cm)	2(2~2)	2(2~3)	-1.82	0.068
ALB(g/L)	43.65(40.02~46.15)	42.55(37.90~46.15)	-0.34	0.731
WBC(*109/L)	5.71(4.64~7.08)	5.97(5~8.35)	-0.97	0.334
PLT(*109/L)	204(163~242)	209(135~362)	-0.60	0.550
INR(s)	0.96(0.93~1.01)	0.97(0.89~1.00)	-0.79	0.432
PT(s)	10.70(10.20~11.20)	10.80(10.00~11.40)	-0.45	0.656
APTT(s)	26.40(24.90~27.40)	25.90(24.90~29.20)	-0.26	0.797
Self-management	164(156~167)	133(124~141)	-6.70	<0.001
HRQOL	1.00(0.95~1.00)	0.95(0.73~1.00)	-3.57	<0.001
WQOL	160(150~167)	121(92~124)	-7.01	<0.001
Marital status			13.59	0.004
Unmarried [n(%)]	2(1.1)	0(0.0)		
Married [n(%)]	179(94.6)	20(87.0)		
Divorced [n(%)]	6(3.2)	0(0.0)		
Widowed [n(%)]	2(1.1)	3(13)		
PICC on the left history			9.31	0.009
No [n(%)]	180(95.2)	22(95.7)		
One time [n(%)]	9(4.8)	0(0.0)		
Two or more times [n(%)]	0(0.0)	1(4.3)		
PICC on the right history			5.45	0.065
No [n(%)]	164(86.8)	19(82.6)		

One time [n(%)]	19(10.0)	1(4.3)		
Two or more times [n(%)]	6(3.2)	3(13.1)		
Hypertension	23(12.2)	2(8.7)	0.24	0.626
Diabetes	7(3.7)	0(0.0)	0.88	0.348
Picc on the left or right arm			0.33	0.566
Left [n(%)]	69(36.5)	7(30.4)		
Right [n(%)]	120(63.5)	16(69.6)		
puncture times			0.14	0.706
One time [n(%)]	177(93.7)	22(95.6)		
Two or more times [n(%)]	12(6.3)	1(4.4)		
Handedness	119(63.0)	13(56.5)	0.36	0.547
Number of catheter lumen			0.75	0.386
Single lumen [n(%)]	183(96.8)	23(100.0)		
double lumen [n(%)]	6(3.2)	0(0.0)		
Puncture vein			7.49	0.024
Basilic vein [n(%)]	161(85.2)	15(65.2)		
Brachial vein [n(%)]	27(14.3)	7(30.4)		
Cephalic vein [n(%)]	1(0.5)	1(4.4)		

Note: UE, unplanned extubation; PICC-UE, unplanned extubation of peripherally inserted central catheters; BMI, Body Mass Index; HRQOL, health-related quality of life; ALB, albumin; WBC, White Blood Cell; PLT,platelet; INR, International Normalized Ration; PT, Prothrombin Time; APTT,Activated Partial Thromboplastin Time.The items (age, height, weight, BMI, puncture vein diameter, arm circumference,insertion length, exposed length, ALB, WBC, PLT, INR, PT, APTT, self-management, HRQOL, weighted quality of life) with the Mann-Whitney U test, the other items with χ^2 test.

Table-3 Comparison of models (n = 212)

Methods	Three mainstream variable encoding methods			Four for data imbalance processing methods				Four ML models				Four XGB models			
	WOE	One-Hot	CE	NearMiss	ENN	SMOTE	FL	XGB	SVM	Random Forest	Logistic Regression	XGB	XGB + FC	XGB + FL	XGB + FL + FC
Accuracy	0.967	0.958	0.967	0.967	0.962	0.965	0.967	0.967	0.972	0.962	0.953	0.972	0.976	0.920	0.967
F1	0.851	0.816	0.851	0.796	0.811	0.824	0.851	0.851	0.856	0.810	0.770	0.850	0.878	0.679	0.851
Recall	0.870	0.870	0.870	0.786	0.812	0.835	0.870	0.870	0.820	0.780	0.740	0.739	0.783	0.783	0.870
AUC	0.994	0.990	0.990	0.907	0.932	0.945	0.994	0.994	0.952	0.993	0.972	0.985	0.993	0.976	0.994

Note: FC,.feature construction;AUC, area under the ROC curve; ENN, edited nearest neighbor; FL, focal loss; LR, logistic regression; ML, machine-learning; ROC, receiver operating characteristic; RFC, random forest classifier; SVM, support vector machine; WOE, weight of evidence.