RESEARCH

Open Access

Post-Anesthesia Care Unit (PACU) readiness predictions using machine learning: a comparative study of algorithms



Shahnam Sedigh Maroufi^{1†}, Maryam Soleimani Movahed^{2†}, Azar Ejmalian³, Maryam Sarkhosh^{1*} and Ali Behmanesh^{2,4*}

Abstract

Introduction Accurate and timely discharge from the Post-Anesthesia Care Unit (PACU) is essential to prevent postoperative complications and optimize hospital resource utilization. Premature discharge can lead to severe issues such as respiratory or cardiovascular complications, while delays can strain hospital capacity. Machine learning algorithms offer a promising solution by leveraging large amounts of patient data to predict optimal discharge times. Unlike prior studies relying on statistical models or single-algorithm methods, this research assesses multiple ML models to predict discharge readiness, comparing them against staff evaluations and the Aldrete checklist.

Methodology We conducted a cross-sectional study of 830 patients under general anesthesia from December 2023 to April 2024, collecting demographics, surgical details, and Aldrete scores. A power analysis ensured statistical robustness, targeting a 5% accuracy improvement (minimum clinically important difference, derived from Gabriel et al., 2017), with variance ($SD \approx 0.1$) from pilot data, using a two-sample t-test (power = 0.8, alpha = 0.05), confirming the sample size's adequacy. Two prediction approaches were tested: discharge timing in 15-minute intervals and binary classification (within 15 min or later). Models included Random Forest (RF), Support Vector Machines (SVM), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and XGBoost, assessed via accuracy, precision, recall, F1 score, and AUC. Predictions were benchmarked against staff and Aldrete scores, with 99.5% confidence intervals (Cls) adjusting for multiple comparisons.

Results he RF algorithm showed high performance in both prediction approaches. In the first approach, RF achieved an AUC of 0.75 (99.5% CI: 0.70–0.80) and accuracy of 0.87 (99.5% CI: 0.83–0.91) per staff evaluations, and an AUC of 0.87 (99.5% CI: 0.83–0.91) and accuracy of 0.71 (99.5% CI: 0.66–0.76) per Aldrete scores. In the second approach, RF recorded an AUC of 0.85 (99.5% CI: 0.81–0.89) and accuracy of 0.86 (99.5% CI: 0.82–0.90) per staff evaluations, with ANN also showing strong results (AUC = 0.88, 99.5% CI: 0.84–0.92; accuracy = 0.78, 99.5% CI: 0.74–0.82). Due to

⁺Shahnam Sedigh Maroufi and Maryam Soleimani Movahed shared as co-first authors.

*Correspondence: Maryam Sarkhosh Maryamsarkhosh1999@gmail.com Ali Behmanesh aa.behmanesh@gmail.com

Full list of author information is available at the end of the article



© 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 are shared 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/.

overlapping CIs, differences between models were not statistically significant (*P* > .005). According to the Aldrete checklist, RF, SVM, and ANN exhibited competitive predictive capability, with AUCs ranging from 0.80 to 0.86.

Conclusion The strong performance of Random Forest (RF) and Artificial Neural Network (ANN) models in predicting PACU discharge timing upon admission highlights their potential as effective tools for evaluating discharge readiness, as compared to staff assessments and the Aldrete checklist. This study focused on assessing these models, showing their ability to produce consistent predictions, though differences between top models were not statistically significant due to overlapping confidence intervals. Practical application of these findings to improve patient outcomes or hospital efficiency requires further investigation.

Keywords Post-anesthesia care unit, Machine learning, Discharge prediction, Length of stay, Recovery

Introduction

The Post-Anesthesia Care Unit (PACU) often referred to as the recovery room, is designed for the immediate postoperative care of patients as they emerge from anesthesia, closely monitoring their physiological state and addressing any complications that may arise. Inadequate or premature discharge from PACU can lead to severe postoperative complications, including respiratory, cardiovascular, and neurological issues.

The duration of patients' stay in the PACU holds significant importance within the realm of operating room management, exerting notable influence on operational efficiency, hospital expenditures, and staff workload [1]. This temporal aspect plays a pivotal role in facilitating the smooth transition of patients from the PACU to their designated post-operative care environments [2]. Prolonged PACU stays not only have the potential to engender patient dissatisfaction but also to escalate institutional costs [3]. Conversely, premature discharge from the PACU may precipitate residual complications arising from anesthesia and surgical interventions, thereby elevating mortality and morbidity rates [4]. Therefore, timely and appropriate discharge from PACU is crucial for patient safety, reducing morbidity, and enhancing the efficiency of hospital operations [5].

Traditionally, patient discharge from PACU has been based on standardized checklists, such as the Aldrete score, which assesses parameters like respiration, circulation, consciousness, activity, and oxygen saturation [6]. While these tools have been useful, they may not account for the nuanced variations in patient conditions, often leading to subjective decisions by healthcare providers. Additionally, with the increasing workload of healthcare professionals, especially in high-volume centers, calculating discharge readiness based solely on manual assessments can be prone to delays and errors [7].

Historically, predictive models have been devised to forecast PACU stay durations, furnishing invaluable clinical insights for streamlining discharge protocols [8, 9]. Estimating the time until patients are ready for discharge from the PACU involves understanding the recovery time variability among patients. This allows for better management of resources, timely communication with families, and improved patient flow, ensuring a smoother and more efficient process [10]. The emergence of machine learning (ML) algorithms presents a promising solution for enhancing the discharge decisionmaking process. Endeavors have been undertaken to curtail extended PACU stays through the deployment of machine learning frameworks, leveraging perioperative and patient data to bolster PACU efficiency and, consequently, mitigate mortality and morbidity rates [11]. These predictive modeling initiatives harbor the capacity to profoundly influence clinical decision-making and patient outcomes in the post-anesthesia milieu [8].

The primary decisions under study include determining when a patient is clinically ready for discharge based on standardized criteria, predicting discharge timing upon PACU admission, and comparing ML-based predictions with traditional staff assessments. These decisions impact hospital workflow, PACU resource allocation, and patient outcomes. Previous studies have analyzed PACU discharge timing and decision-support tools, highlighting challenges in achieving optimal predictions [12]. By addressing these decision points, our study aims to improve upon existing methodologies and provide datadriven insights into PACU discharge efficiency.

This study builds upon prior research in predictive analytics for PACU discharge while introducing novel contributions. Previous efforts, such as those by Gabriel et al. [8] and Tully et al. [11], focused primarily on statistical modeling or limited machine learning implementations for specific surgical contexts. In contrast, our approach integrates a comprehensive set of machine learning models to predict discharge readiness, addressing both precise time intervals and binary classification. Through the comparison of machine learning predictions with two reference points-staff evaluations embodying practical decision-making and Aldrete scores representing standardized clinical benchmarks-we establish a dual-validation system designed to improve reliability and adaptability across varied patient groups. These advancements aim to reduce subjective decision-making, optimize PACU resource allocation, and improve patient outcomes, offering a practical tool for integration into hospital management systems.

Methods

Study design and setting

This study was designed as a cross-sectional, descriptive, and analytical study conducted at Firouzgar Hospital, Tehran, Iran. The study aimed to evaluate the efficacy of machine learning algorithms in predicting patient discharge from the PACU, utilizing demographic, clinical, and procedural data and patient outcomes based on the Aldrete discharge criteria.

Patient selection

Inclusion criteria encompassed patients aged 18 years and older undergoing elective surgeries with general anesthesia, ensuring a focus on typical postoperative recovery scenarios. Exclusion criteria included patients with incomplete medical records, those receiving only local or regional anesthesia, and those experiencing intraoperative complications (e.g., excessive bleeding, cardiac events) that precluded normal postoperative recovery. This ensured the dataset reflected standard PACU discharge scenarios while minimizing confounding factors such as emergency procedures or atypical recovery trajectories.

Out of the original 872 patients, 42 (4.8%) were excluded because of unexpected complications such as severe bleeding or arrhythmias, resulting in a final dataset of 830. These data were gathered from patients who underwent general anesthesia at Firouzgar Hospital between December 2023 and April 2024. Variations in surgical procedures (e.g., ENT, orthopedics, gynecology) and patient conditions (e.g., ASA I-II classification, comorbidities like hypertension or diabetes) were retained to capture real-world diversity influencing discharge times.

Tak	ble	е	1	ΑI	d	rete	score	criteria	for	disc	harge	readi	ness
-----	-----	---	---	----	---	------	-------	----------	-----	------	-------	-------	------

Category	2	1	0
Respiration	Able to take deep breath and cough	Dyspnea/Shallow Breathing	Apnea
O ₂ Saturation	Main- tains > 92% on room air	Needs O ₂ inhalation to maintain O ₂ saturation > 90%	Satura- tion < 90% even with supple- mental O ₂
Consciousness	Fully awake	Arousable on calling	Not responding
Circulation	BP±20mmHg pre op	BP±20-50mmHg pre op	BP±50mmHg pre op
Activity	Able to move 4 extremities voluntarily or on command	Able to move 2 extremities voluntarily or on command	Able to move 0 extremities voluntarily or ² on command

Normal postoperative recovery was defined as the absence of significant intraoperative complications (e.g., severe hypotension, arrhythmias, or unplanned interventions) and a stable postoperative course in the PACU, as indicated by Aldrete scores trending toward 9 or higher within a typical timeframe (e.g., within 75 min, based on the study's observed PACU length of stay range of 5–75 min). Patients deviating from this trajectory due to unforeseen events were excluded to maintain focus on predictable recovery patterns.

The sample size of 830 patients was determined via power analysis to detect a clinically meaningful difference in predictive accuracy between ML models and traditional benchmarks. We targeted a minimum accuracy improvement of 5%, considered managerially significant based on prior PACU studies (Gabriel et al., 2017), with an estimated standard deviation of 0.1 derived from pilot data at Firouzgar Hospital. Using a two-sample t-test, we calculated a required sample size to achieve 80% power at an alpha of 0.05, confirming that 830 patients (split 80/20 for training/testing) provided sufficient statistical power for reliable, generalizable results.

Data collection

Data were gathered using convenience sampling based on the availability of eligible patient records during the study period. While this approach facilitated timely data collection within a single-center setting, it may introduce selection bias by overrepresenting certain patient profiles (e.g., elective surgeries at Firouzgar Hospital).

The data for each patient were collected based on three primary sources: (1) patient demographic information (e.g., age, gender, medical history) (2), surgical data (e.g., type of surgery, duration of surgery), and (3) PACU-specific clinical data including the Aldrete score and staff evaluations of readiness for discharge. Staff evaluations were recorded to compare with machine learning model predictions. Staff-based decisions for discharge generally aligned with the Aldrete score but exhibited subjectivity in borderline cases, particularly for patients with multiple comorbidities. Data were gathered using convenience sampling based on the availability of eligible patient records.

The Aldrete score is a postoperative assessment tool that evaluates several criteria including activity, respiration, circulation, level of consciousness, and arterial blood oxygen saturation [13]. Details of these criteria and their scoring are provided in Table 1. Patients achieving a score of 9 or higher were deemed fit for discharge [14, 15].

Machine learning algorithms

The choice of machine learning models—Random Forest (RF), Support Vector Machines (SVM), Logistic

Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and XGBoost-was guided by their proven efficacy in healthcare predictive analytics and their alignment with the dataset's characteristics. RF and XGBoost were selected for their robustness in handling high-dimensional, heterogeneous clinical data with both categorical (e.g., surgery type) and continuous (e.g., anesthesia duration) variables, as well as their ability to manage non-linear relationships common in postoperative recovery patterns. SVM and LR were included for their effectiveness in classification tasks and interpretability, respectively, offering a baseline for comparison. ANN was chosen to capture complex, non-linear interactions in the data, given its strength in modeling intricate patient recovery trajectories. DT and KNN were included to explore simpler, intuitive approaches, though their limitations (e.g., overfitting in DT, sensitivity to noise in KNN) were anticipated. This diverse ensemble ensures a comprehensive evaluation tailored to the clinical problem of predicting PACU discharge readiness.

Approach 1: Predicting exact time of discharge in 15-minute intervals

The first approach was designed as a multiclass classification task, where the goal was to predict the exact time window for a patient's discharge from the PACU. The discharge times were divided into 15-minute intervals, based on data collected about the patient's condition during their stay in PACU.

Approach 2: Binary classification of discharge within the first 15 min or after

This approach simplifies the discharge prediction by focusing on a binary classification task. Instead of predicting the exact discharge time, the goal was to determine whether a patient could be discharged from the PACU within the first 15 min of their post-surgical recovery or if they would require more than 15 min to stabilize. In this approach, a variety of features used to predict discharge timing, but the output was simplified to a binary decision:

- Discharge within 15 min (early discharge)
- Discharge after 15 min (late discharge)

The 15-minute threshold was determined according to the Guideline from the Association of Anaesthetists [16]. These guidelines specify that complete and accurate information about the patient's condition is recorded every 15 min, and this data is used to assess the patient's discharge status.

Model parameter tuning

For ANN, we used a predefined architecture (two hidden layers with 64 and 32 neurons, ReLU and softmax activations) and the Adam optimizer (learning rate 0.001), based on common practices and initial tests, though not systematically optimized. SVM hyperparameters ('C' and 'kernel') were tuned using GridSearchCV with 5-fold cross-validation over a specified range, offering a systematic yet not exhaustive optimization. XGBoost parameters were derived from a prior process yielding a best_params output, but we acknowledge the lack of detailed documentation on the method and ranges, which we have now noted as a limitation. Random Forest tuning utilized RandomizedSearchCV with 5-fold cross-validation across a defined parameter space, balancing efficiency and reliability, though not guaranteeing absolute optimality. KNN's n_neighbors (1-20) was optimized exhaustively via GridSearchCV with 5-fold crossvalidation, ensuring robust selection. Logistic Regression (LR) and Decision Tree (DT) were initially run with defaults due to their simpler parameter sets, but we have now applied GridSearchCV for consistency (e.g., 'C' for LR, max depth for DT).

We utilized the following Python packages with their respective versions: scikit-learn (v1.4.2) for implementing RF, SVM, LR, DT, KNN, and associated tuning methods (GridSearchCV, RandomizedSearchCV); XGBoost (v1.7.5) for the XGBoost model; and TensorFlow (v2.10.1) for the ANN. These versions reflect the software environment during the study period (December 2023 to April 2024), based on our records. Data preprocessing and evaluation were also conducted using NumPy (v1.23.4) and Pandas (v1.5.1). All analyses were performed in Python (v3.7). The file environment.txt, which includes full environment details, has been attached as a supplementary document, providing a comprehensive list of all packages and their versions for exact replication of the computational workflow.

Input data and features

To enhance clinical interpretability and manage continuous variables, we transformed BMI and blood pressure into categorical composite scores. BMI, derived from height and weight measurements, was categorized into five groups: underweight (<18.5 kg/m²), healthy weight (18.5–24.9 kg/m²), overweight (25–29.9 kg/m²), obese (30–34.4 kg/m²), and severely obese (\geq 35 kg/m²). Blood pressure, based on systolic and diastolic readings, was classified into four categories: normal (systolic <120 mmHg and diastolic < 80 mmHg), elevated (systolic 120– 129 mmHg and diastolic < 80 mmHg), Stage 1 hypertension (systolic 130–139 mmHg or diastolic 80–89 mmHg), and Stage 2 hypertension (systolic \geq 140 mmHg or diastolic \geq 90 mmHg). No features were excluded from the dataset; all other variables (e.g., age, anesthesia time) were retained as independent predictors. These categorical scores were one-hot encoded for compatibility with the machine learning algorithms.

To address potential multicollinearity and feature redundancy in the clinical dataset, feature engineering and selection were conducted prior to model training. Variance inflation factors (VIF) were calculated to assess multicollinearity among continuous variables (e.g., anesthesia time, PACU length of stay), with features exceeding a VIF threshold of 5 (e.g., highly correlated preoperative and postoperative blood pressure readings) combined into composite scores or excluded. Categorical variables (e.g., surgery type, ASA classification) were onehot encoded to ensure compatibility with the algorithms. Feature importance rankings from a preliminary RF model were used to select the top 15 predictive features (e.g., Aldrete score components, anesthesia duration, patient age), reducing dimensionality while retaining clinically relevant predictors. This process minimized redundancy and enhanced model performance.

Data were split into training (80%) and testing (20%) sets for model development and validation. ten-fold cross-validation was used to assess the performance and generalization ability of the machine learning models. This technique ensured that the models developed for predicting patient discharge times from PACU did not overfit the training data and could reliably predict discharge times on unseen data. The training phase included parameter tuning to optimize model performance based on accuracy (ACC), Area Under the Curve (AUC), and other performance metrics such as sensitivity and specificity.

Statistical adjustments

Model performance was evaluated using accuracy, precision, recall, F1 score, and AUC, with adjustments for multiple comparisons across the seven machine learning models. The Bonferroni correction was applied to control the family-wise error rate [17], adjusting the significance level to α corrected = 0.05/7 \approx 0.00714. For simplicity and added conservatism, we reported 99.5% confidence intervals, slightly stricter than the 99.29% implied by the Bonferroni adjustment, to ensure robust interpretation of model performance amidst multiple comparisons.

Results

Overview of the patient population

The study population consisted of 830 patients who underwent general anesthesia and were admitted to the PACU at Firouzgar Hospital. A wide range of patient data was gathered both before and after surgical procedures. The data is presented in Table 2.

	ruic guuicic		
Features	Frequency (%)	Range of values	Mean ± SD
Age		18–60	40.29±12.67
Gender			
Female	492 (59.3)		
Male	338 (40.7)		
BMI (Kq/M ²)		12.11-	28.23±12.29
		57.60	
ASA classification score			
ASA I	592 (71.32)		
ASA II	238 (28.67)		
ASA III	0		
ASA IV	0		
ASA V	0		
Past medical history			
Hypertension	124 (14.9)		
Diabetes Mellitus	77 (9.27)		
OSA	3 (0.36)		
Eavism	1 (0 1 2)		
Anemia	9 (1 08)		
Hyperlipidemia	16 (1 92)		
Rheumatoid arthritis	5 (0.60)		
	1 (0 1 2)		
Minor thalassemia	3 (0.36)		
Multiple sclerosis	1 (0 1 2)		
Cachovia	3 (0.36)		
Bro oclampsia	1 (0.12)		
Programsky	1 (0.12)		
Chronic Obstructive Pulme	1 (0.12)		
nary Disease	1 (0.12)		
hypothyroidism	43 (5.18)		
Asthma	22 (2.65)		
Surgery type	22 (2.05)		
ENIT	96 (116)		
Coporal surgery	102 (23 1)		
Orthopodics	192 (23.1)		
Neurosurgery	0(11)		
Pariatric	9(1.1)		
Ballatic	90 (11.0)		
Currenter	20(10.4)		
Gynecology	200 (24.0)		
Smoking	172 (20.0)		
Yes	173 (20.8)		
NO	657 (79.2)		
Alcohol consumption			
Yes	66 (7.9)		
No	/64 (92.2)		
Substance addiction	0.0 (0. ⁻)		
Yes	30 (3.6)		
No	800 (96.4)		
Anestnesia time (min) PACU length of stay (min)		15–120 5–75	//.28±33.26 21.17±12.09

Table 2 Variables of study in the gathered dataset

Patient demographics, including age, gender, and preexisting medical conditions, were recorded. The average patient age was 40.29 ± 12.67 years, with a slightly higher proportion of females (59.3%) compared to males (40.7%). A significant number of patients (24.17%) had underlying health issues, primarily hypertension (14.9%) and diabetes (9.27%). All patients included in this study were classified as American Society of Anesthesiologists (ASA) physical status classifications 1 and 2. Due to the high prevalence of hypertension, blood pressure was meticulously monitored and documented both before and after surgery, as well as upon discharge from the PACU.

Key patient data was used to improve the accuracy of discharge predictions, especially for patients at risk of heart problems. This data included the ASA, surgical duration, patient demographics (sex and age), substance use history (alcohol, cigarettes, and other addictions), body mass index (BMI), pre-operative blood pressure. The Aldrete score parameters included post-operative blood pressure, oxygen saturation (SpO2), movement, nausea, pain, respiratory rate, agitation, and level of consciousness prior to transfer to the recovery room.

Approach 1: Predicting exact time of discharge in 15-minute intervals

In the first approach to data analysis, several machine learning models—including XGBoost, LR, DT, RF, KNN, ANN, and SVM —were used to predict the specific 15-minute interval (first, second, third, and fourth) in which patients should be discharged from the PACU.

Machine learning vs. PACU staff

In predicting discharge times in 15-minute intervals based on PACU staff opinions, RF exhibited the highest point estimates across metrics, with an accuracy of 0.87 (99.5% CI: 0.83–0.91) and AUC of 0.75 (99.5% CI: 0.70– 0.80). XGBoost and SVM also showed strong performance (e.g., XGBoost AUC = 0.73, 99.5% CI: 0.68–0.78), while KNN had the lowest scores (AUC = 0.60, 99.5% CI: 0.55–0.65). Differences between top models like RF and SVM were not statistically significant (P>.005) due to overlapping CIs, suggesting competitive performance among leading algorithms (Fig. 1).

Machine learning vs. Aldret checklist

In predicting discharge times in 15-minute intervals based on Aldrete checklist scores, RF achieved the highest point estimates across all metrics, such as an AUC of 0.87 (99.5% CI: 0.83–0.91) and accuracy of 0.71 (99.5% CI: 0.66–0.76), positioning it as a strong candidate for predicting discharge timing. KNN consistently showed the lowest performance (e.g., AUC \approx 0.60, 99.5% CI: 0.55– 0.65), suggesting it is poorly suited for this task. Logistic Regression (LR), SVM, and XGBoost delivered competitive results (e.g., SVM AUC \approx 0.84, 99.5% CI: 0.80–0.88), though their CIs overlapped with RF's, indicating no statistically significant difference (P>.005). ANN also performed well (e.g., AUC \approx 0.85, 99.5% CI: 0.81–0.89), but



Fig. 1 Heatmap displaying the comparison of performance metrics (accuracy, precision, recall, F1 score, and AUC) across various models (LR, SVM, XG-Boost, DT, KNN, RF, ANN), with color intensity indicating performance levels, for predicting patient discharge timing based on PACU staff opinions

its metrics were not significantly superior to RF's due to overlapping CIs. These findings highlight RF's robust predictive potential, tempered by comparable performance among other leading models within statistical limits (Fig. 2).

Approach 2: Binary classification of discharge within or after 15 min

Approach 2 aimed to classify whether a patient could be discharged within the first 15 min of arriving in the PACU or if they would require a longer stay. In this analysis, the focus was on patients who were discharged within the initial 15 min compared to those who remained in the PACU longer. Patients discharged in the first 15 min were categorized into one group, while those discharged after this time were placed in another.

Machine learning vs. PACU staff

For binary classification (discharge within 15 min or later) per staff evaluations, RF achieved an accuracy of 0.86 (99.5% CI: 0.82–0.90) and AUC of 0.85 (99.5% CI: 0.81–0.89), with ANN showing a slightly higher AUC of 0.88 (99.5% CI: 0.84–0.92) but lower accuracy (0.78, 99.5% CI: 0.74–0.82). KNN consistently underperformed (AUC = 0.62, 99.5% CI: 0.57–0.67). Pairwise comparisons (e.g., RF vs. ANN) showed overlapping CIs, indicating no clear superiority at P<.005 (Fig. 3).

Machine learning vs. Aldret checklist

When predicting discharge within 15 min or later per the Aldrete checklist, RF, SVM, and LR showed comparable high performance (e.g., RF AUC = 0.86, 99.5% CI: 0.82–0.90; SVM AUC = 0.85, 99.5% CI: 0.81–0.89), with ANN slightly ahead in AUC (0.88, 99.5% CI: 0.84–0.92). KNN lagged (AUC = 0.61, 99.5% CI: 0.56–0.66). Overlapping CIs suggest these top models performed competitively, with no single model significantly outperforming others at P<.005 (Fig. 4).

Discussion

This research highlights the ability of machine learning algorithms, specifically Random Forest RF and ANN, to accurately predict PACU discharge timing at the point of admission. The benchmarks used include staff evaluations and the Aldrete checklist, with the Aldrete score recognized as the gold standard for assessing discharge readiness. However, staff opinions provide a valuable practical perspective, integrating situational clinical judgment. RF consistently showed the highest point estimates (e.g., AUC 0.85, 99.5% CI: 0.81-0.89 in Approach 2), though overlapping CIs with ANN and SVM suggest competitive performance rather than clear superiority (P>.005). Beyond demonstrating technical efficacy, these findings offer practical value by providing a data-driven tool to enhance PACU discharge decisions, potentially reducing premature discharges that risk complications and prolonged stays that strain resources [2]. Timely discharge predictions can optimize bed utilization, reduce



Fig. 2 Heatmap displaying the comparison of performance metrics (accuracy, precision, recall, F1 score, and AUC) across various models (LR, SVM, XG-Boost, DT, KNN, RF, ANN), with color intensity indicating performance levels, for predicting patient discharge timing according to the Aldrete checklist





Fig. 3 Heatmap displaying the comparison of performance metrics (accuracy, precision, recall, F1 score, and AUC) across various models (LR, SVM, XG-Boost, DT, KNN, RF, ANN), with color intensity indicating performance levels, for predicting patient discharge before and after first 15 min according to the opinions of PACU staff



Performance Metrics Heatmap for Machine Learning Models

Fig. 4 Heatmap displaying the comparison of performance metrics (accuracy, precision, recall, F1 score, and AUC) across various models (LR, SVM, XG-Boost, DT, KNN, RF, ANN), with color intensity indicating performance levels, for predicting patient discharge before and after first 15 min according to the Aldrete checklist

wait times, lower costs, and ultimately improve patient outcomes and hospital efficiency [1, 5].

In real-world settings, integrating these models into PACU workflows could involve a user-friendly interface where clinicians enter key patient metrics (e.g., Aldrete scores, surgical duration) and receive discharge probabilities within seconds. This approach would complement clinical expertise rather than replace it, helping to standardize decisions and reduce subjectivity, particularly in borderline cases such as patients with comorbidities [1, 5]. Such integration could be especially beneficial in high-volume centers, where time constraints and variability in subjective judgment pose challenges.

Implementing such ML tools, however, poses challenges. Healthcare providers may hesitate to trust blackbox predictions, necessitating explainable outputs to build confidence. Workflow integration requires seamless compatibility with electronic health records (EHRs) and minimal disruption—perhaps via a tablet-based app synced with existing systems—plus staff training to interpret ML recommendations alongside their assessments. Compared to staff evaluations alone, ML adds consistency and speed, processing nuanced patterns (e.g., subtle interactions between age and anesthesia time) that busy clinicians might overlook, as supported by Yang et al.'s [18] work on recovery predictions.

The success of RF can be attributed to its ability to handle complex, high-dimensional data, and its capacity for managing both continuous and categorical variables, which are prevalent in postoperative patient data. The robustness of RF in this setting mirrors findings from previous studies [18, 19], which identified RF and ANN as top-performing models in predicting post-surgical recovery and discharge times. Our results align with theirs, where RF excelled (AUC ≈ 0.85), but diverge from Tully et al.'s [11] XGBoost success in outpatient contexts, likely due to our diverse inpatient cohort favoring RF's complexity-handling. Unlike Kim et al.'s [19] ANN focus with the PARS checklist, our dual-benchmark approach broadens applicability. Still, feasibility hinges on addressing these hurdles-trust, training, and system integration-through pilot testing in varied PACU settings. By offering a standardized, rapid supplement to staff judgment, ML could optimize patient flow, cut costs, and enhance outcomes [5], making it a practical step forward if these challenges are met.

These findings contribute new insights by demonstrating the adaptability of machine learning across varied PACU scenarios and highlighting its potential to standardize discharge decisions beyond single-algorithm or context-specific applications.

Study limitations

Several limitations were identified in this study. First, premature discharges occurred for some patients based on subjective opinions of PACU staff, which may have influenced the accuracy of the model's predictions. Additionally, incomplete self-reporting by patients regarding alcohol consumption, smoking habits, and drug use may have introduced biases into the dataset. The severity of underlying medical conditions was not always thoroughly accounted for, which could affect recovery times and discharge readiness. Furthermore, pre-anesthesia stress and its impact on blood pressure were not consistently monitored, potentially skewing the assessment of patient stability before discharge.

The reliance on convenience sampling may limit the generalizability of our findings, as it may not fully represent the broader population of PACU patients across different hospitals or surgical contexts. Future studies could employ systematic or randomized sampling to enhance representativeness and validate these results in diverse settings.

While ten-fold cross-validation ensured internal robustness, the lack of external validation across different hospitals limits the generalizability of our findings. Variations in patient demographics, surgical practices, and PACU protocols at other institutions could affect model performance. Future research should validate these models using multi-center datasets to confirm their applicability in diverse clinical settings.

For future studies, including more detailed intraoperative data, such as blood pressure fluctuations during surgery, could provide deeper insights into patient recovery patterns and improve the accuracy of discharge predictions. Expanding the scope of data collection and considering these factors would enhance the robustness of machine learning models in predicting safe and timely discharges.

Conclusion

The high predictive accuracy of RF and ANN models suggests that machine learning could standardize PACU discharge decisions, reducing reliance on subjective staff assessments and minimizing risks of premature or delayed discharges. Accurate and timely predictions from such models may contribute to better bed management, shorter wait times, and reduced operational costs, ultimately supporting improved patient care and hospital efficiency. This study focused on assessing these models, showing their ability to produce consistent predictions, though differences between top models were not statistically significant due to overlapping confidence intervals. Practical application of these findings to improve patient outcomes or hospital efficiency requires further investigation. However, real-world implementation faces challenges, including the need for explainable models to gain clinician trust, staff training to integrate AI tools into workflows, and system integration complexities within existing hospital infrastructures. Addressing these barriers-through user-friendly interfaces, interdisciplinary collaboration, and validation in diverse settings-will be critical to translating these findings into practice. Future work should focus on overcoming these hurdles to fully realize the potential of AI-driven PACU management.

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12911-025-02982-0.

Supplementary Material 1 Supplementary Material 2

Acknowledgements

We gratefully acknowledge all participants and collaborators who contributed to this research.

Author contributions

S.S.M. was responsible for supervision and project administration. M.S.M. contributed to the methodology, formal analysis, and data curation. A.E. provided resources and participated in writing, review, and editing. M.S. handled data curation, validation, and writing the original draft. A.B. was involved in conceptualization, software development, visualization, and writing, review, and editing.

Funding

This study is a continuation of a Master's thesis completed at Iran University of Medical Sciences with Research registration number 1403-2-5-27914.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

This study, continuing from a Master's thesis at Iran University of Medical Sciences, was approved by the ethics committee (IR.IUMS.REC.1402.680). Participants provided informed consent, and the research adhered to the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Anesthesia, Faculty of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran

²Education Development Center, Iran University of Medical Sciences, Tehran, Iran

³Department of Anesthesiology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

⁴Bone and Joint Reconstruction Research Center, Department of

Orthopedics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

Received: 3 October 2024 / Accepted: 20 March 2025 Published online: 25 March 2025

References

- Lalani SB, Ali F, Kanji Z. Prolonged-stay patients in the PACU: a review of the literature. J Perianesthesia Nurs. 2013;28(3):151–5.
- 2. Seago JA, Weitz S, Walczak S. Factors influencing stay in the postanesthesia care unit: a prospective analysis. J Clin Anesth. 1998;10(7):579–87.
- Kiekkas P, Aretha D. PACU nurse staffing and patient outcomes: the evidence is still missing. J Perianesthesia Nurs. 2018;33(2):244–6.

- Samad K, Khan M, Khan FA, Hamid M, Khan FH. Unplanned prolonged postanaesthesia care unit length of stay and factors affecting it. J Pakistan Med Association. 2006;56(3):108.
- Carr DA, Saigal R, Zhang F, Bransford RJ, Bellabarba C, Dagal A. Enhanced perioperative care and decreased cost and length of stay after elective major spinal surgery. NeuroSurg Focus. 2019;46(4):E5.
- Bizuneh YB, Ashagrie HE, Lema GF, Fentie DY. Current Practice of Discharging Patients From Post Anesthesia Care Unit After Surgical Operations, 2018. 2020.
- Hazzard B, Johnson K, Dordunoo D, Klein T, Russell B, Walkowiak P. Work- and nonwork-related factors associated with PACU nurses' fatigue. J Perianesth Nurs. 2013;28(4):201–9.
- Gabriel RA, Waterman RS, Kim J, Ohno-Machado L. A predictive model for extended postanesthesia care unit length of stay in outpatient surgeries. Anesth Analgesia. 2017;124(5):1529–36.
- Fang F, Liu T, Li J, Yang Y, Hang W, Yan D, et al. A novel nomogram for predicting the prolonged length of stay in post-anesthesia care unit after elective operation. BMC Anesthesiol. 2023;23(1):404.
- Dexter F, Epstein RH. Implications of the log-normal distribution for updating estimates of the time remaining until ready for phase I post-anesthesia care unit discharge. Perioperative Care Operating Room Manage. 2021;23:100165.
- Tully JL, Zhong W, Simpson S, Curran BP, Macias AA, Waterman RS, et al. Machine learning prediction models to reduce length of stay at ambulatory surgery centers through case resequencing. J Med Syst. 2023;47(1):71.
- Ehrenfeld JM, Dexter F, Rothman BS, Minton BS, Johnson D, Sandberg WS, et al. Lack of utility of a decision support system to mitigate delays in admission from the operating room to the postanesthesia care unit. Anesth Analg. 2013;117(6):1444–52.
- Chekol B, Eshetie D, Temesgen N. Assessment of staffing and service provision in the post-anesthesia care unit of hospitals found in Amhara regional state, 2020. Drug, healthcare and patient safety. 2021:125–31.
- McMenamin L, Clarke J, Hopkins P. Basics of anesthesia. Br J Anaesth. 2018;120(5):1141.
- 15. Elisha S, Heiner J, Nagelhout JJ. Nurse anesthesia: Elsevier. 2023.
- Klein A, Meek T, Allcock E, Cook T, Mincher N, Morris C, et al. Recommendations for standards of monitoring during anaesthesia and recovery 2021: guideline from the association of anaesthetists. Anaesthesia. 2021;76(9):122–23.
- 17. Bland JM, Altman DG. Multiple significance tests: the bonferroni method. BMJ. 1995;310(6973):170.
- Yang S, Li H, Lin Z, Song Y, Lin C, Zhou T. Quantitative analysis of anesthesia recovery time by machine learning prediction models. Mathematics. 2022;10(15):2772.
- Kim WO, Kil HK, Kang JW, Park HR. Prediction on lengths of stay in the postanesthesia care unit following general anesthesia: preliminary study of the neural network and logistic regression modelling. 2000.

Publisher's note

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