Multicenter validation of a machine learning model to predict intensive care unit readmission within 48 hours after discharge

Leerang Lim,[a,f] Mincheol Kim,[b,c,f] Kyungjae Cho,[b] Dongjoon Yoo,[b,d] Dayeon Sim,[a] Ho Geol Ryu,[a,e] and Hyung-Chul Lee[a,*]

[a]Department of Anesthesiology and Pain Medicine, Seoul National University College of Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea
[b]VUNO, 479 Gangnam-daero, Seocho-gu, Seoul, 06541, Republic of Korea
[c]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, 115 Irwon-ro, Gangnam-gu, Seoul, 06351, Republic of Korea
[d]Department of Critical Care Medicine and Emergency Medicine, Inha University College of Medicine, 100 Inha-ro, Michuhol-gu, Incheon, 22212, Republic of Korea
[e]Department of Critical Care Medicine, Seoul National University College of Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea

eClinicalMedicine
2025;81: 103112
Published Online xxx
https://doi.org/10.1016/j.eclinm.2025.103112

Summary

Background Intensive care unit (ICU) readmission is a crucial indicator of patient safety. However, discharge decisions often rely on subjective assessment due to a lack of standardized guidelines. We aimed to develop a machine-learning model to predict ICU readmission within 48 h and compare its performance to traditional scoring systems.

Methods We developed an ensemble model, iREAD, that generates a probability score at ICU discharge, representing the likelihood of the patient being readmitted to the ICU within 48 h, using data from Seoul National University Hospital (SNUH) and validated it using the MIMIC-III and eICU-CRD datasets. From September 2007 to August 2021, a total of 70,842 patients were included from SNUH. The MIMIC-III datasets comprised 43,237 patients admitted to ICUs between 2001 and 2012 at Beth Israel Deaconess Medical Center, and the eICU-CRD datasets included 90,271 ICU admissions across 208 hospitals between 2014 and 2015. Patients younger than 18, those who died in ICUs, or who refused life-sustaining treatment were excluded from the final analysis. The model's performance was evaluated using the area under the receiver operating characteristic curve (AUROC) and compared to the traditional scores and conventional machine learning models. Kaplan–Meier analysis was performed to compare the outcome between the high-risk and low-risk groups.

Findings We developed the iREAD, that utilized 30 input features, encompassing demographics, length of stay, vital signs, GCS, and laboratory values. iREAD demonstrated superior performance compared with other models across all cohorts (all $P < 0.001$). In the internal validation, iREAD achieved AUROCs of 0.771 (95% CI 0.743–0.798), 0.834 (0.821–0.846), and 0.820 (0.808–0.832) for early (≤48 h), late (>48 h), and overall ICU readmissions, respectively. External validations with MIMIC-III and eICU-CRD also showed modest performance with AUROCs of 0.768 (0.748–0.787) and 0.725 (0.712–0.739) for overall readmission in MIMIC-III and eICU-CRD respectively, demonstrating superior performance compared to other models (All $P < 0.001$; higher than other models). Kaplan–Meier analysis revealed that over 40% of high-risk patients predicted by iREAD were readmitted within 48 h, representing a more than four-fold increase in predictive performance compared to the traditional scores.

Interpretation iREAD demonstrates superior performance in predicting ICU readmission within 48 h after discharge compared to traditional scoring systems or conventional machine learning models in both internal and external validations. While the performance degradation observed in the external validations suggests the need for further prospective validation on diverse patient populations, the robust performance and ability to identify high-risk patients have the potential to guide clinical decision-making.

*Corresponding author.
E-mail address: vital@snu.ac.kr (H.-C. Lee).
[f]Leerang Lim and Mincheol Kim contributed equally to this work.

Funding This work was supported by the Korea Health Technology Research & Development Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant number RS-2021-KH114109).

Keywords: Intensive care unit; Readmission; Prediction; Machine learning

Research in context

Evidence before this study

We searched PubMed and IEEE Xplore to explore relevant studies that attempted to develop prediction model for intensive care unit (ICU) readmission within 48 h, using the term ("intensive care unit" OR "ICU") AND "readmission" AND "prediction" published in English until 15 December 2023. Most studies have demonstrated modest prediction performance; however, they predominantly focused on developing models for long-term or overall ICU readmission rather than short-term readmission. Additionally, external validation was scarce, which limited the assessment of performance on external datasets. Moreover, machine learning models often exhibited black-box characteristics, meaning that their internal decision-making processes were not transparent or easily interpretable to clinicians utilized specialized tests, utilized specialized tests or treatment outcomes as some input variables and incorporated an excessive number of input variables, thereby making it difficult to be integrated into clinical workflows. Therefore, this study aimed to develop an explainable machine learning model that predict ICU readmission within 48 h, allowing for easy integration of the model into existing clinical workflows. Furthermore, the study intended to demonstrate the model's generalizability through extensive external validation using external datasets, thereby contributing to the prediction of ICU readmissions.

Added value of this study

This study suggests a machine learning model, iREAD, designed to predict short-term (≤48 h), long-term (>48 h), and overall readmission risks for both surgical and medical ICU patients at the time of discharge from ICU. To facilitate integration into real-world clinical workflows, the model utilizes variables that are routinely and easily measured in clinical settings. Additionally, by incorporating serial measurements obtained after ICU admission rather than relying on single time-point observations, the model captures trend changes during the ICU stay. Furthermore, the model's generalizability was evaluated through external validations using patient cohorts differing in race and nationality.

Implications of all the available evidence

iREAD demonstrates superior predictive performance and robust calibration compared to various existing prediction models in both internal and external validations. Furthermore, by identifying variables that contribute to increased readmission risk, it allows the reassessment of the appropriateness of the ICU discharge and adequate post-discharge monitoring, thereby mitigating of the readmission risk. iREAD highlights its potential to be integrated into clinical workflows across diverse ICUs, contributing not only to the prediction but also to the prevention of ICU readmissions.

Introduction

Deciding on a patient's discharge from the intensive care unit (ICU) to the general ward or step-down unit is a daily challenge for intensivists. Prolonged ICU stays can increase the risk of infection and medical costs, while premature ICU discharge can result in higher mortality and readmission to the ICU.[1–4] Therefore, the ICU readmission rate has been a key indicator of ICU safety. Despite its importance, most discharge decisions rely on subjective assessment due to the lack of standardized guidelines.[1–6] Developing an objective tool for patient discharge is complex because it must account for both the patient's medical condition and the non-medical factors, such as the capacity of the receiving unit in the hospital (e.g., ventilators, beds, or specialists).[4,6,7]

Previous studies have identified several risk factors for ICU readmission, including age, comorbidities, severity of illness, route of admission, and diagnosis at admission.[8–12] Some researchers have proposed specialized scoring systems, such as the Stability and Workload Index for Transfer (SWIFT) score, to predict ICU readmission.[12] Severity scores, such as the Modified Early Warning Score (MEWS) or Simplified Acute Physiology Score (SAPS), have also been used as alternatives, albeit with inconsistent predictive performance.[13–15]

A vast amount of clinical data collected from daily care has facilitated the development of clinical decision support systems using machine learning techniques in the field of intensive care medicine.[16] These techniques have helped to create prediction models for mortality, cardiac arrest, and sepsis in ICU patients.[17–19] Similarly, prediction models using machine learning algorithms have been proposed to assess the risk of ICU readmission before discharging patients from the ICU to

general wards.[20,21] Furthermore, if a model incorporates modifiable factors and suggests each variable's impact on the readmission risk, it could provide intensivists with objective scores and risk factors to support their clinical decisions. However, previous ICU readmission models have not yet been thoroughly validated in the external cohorts.

This study aims to develop an explainable machine learning-based prediction model to identify patients at high risk for readmission. We focused on the patients who were discharged from the ICU and attempted to create a model predicting readmission to the ICU within 48 h. We validated our model's performance using the temporally independent dataset and the cohorts from different countries and ethnicities. We hypothesized that the machine learning-based prediction model for ICU readmission would demonstrate excellent performance in both internal and external validations.

Methods

Ethical approval

This study was approved by the Institutional Review Board (IRB) and the Data Review Board of Seoul National University Hospital (SNUH) in the Republic of Korea (IRB No. 2111-140-1275). Due to the retrospective study design and the use of deidentified patient information, the requirement for informed patient consent was waived by the IRB.

Two publicly available datasets were used for external validation in this study: the Medical Information Mart for Intensive Care (MIMIC)-III and the eICU Collaborative Research Database (eICU-CRD). The MIMIC-III dataset was approved by the IRBs of Beth Israel Deaconess Medical Center (IRB No. 2001-P-001699/14) and the Massachusetts Institute of Technology (IRB No. 0403000206) in the USA.[22] The eICU-CRD dataset was exempted from the IRB approval in the USA as it has been certified for re-identification risk from the Health Insurance Portability and Accountability Act (HIPAA) (Certification No. 1031219-2).[23] Both datasets are publicly available under credentialed access on PhysioNet.[22,23]

Study design

This multicenter retrospective cohort study aimed to develop and validate a machine learning model for ICU readmission using the Korean cohort data from SNUH. The model's generalizability was further evaluated through external validation using the MIMIC-III and eICU-CRD datasets.[17,24–26] An overview of the study design is shown in Fig. 1.

All patients admitted to the ICU in SNUH from September 2007 to August 2021 were included in this study. The MIMIC-III datasets included patients admitted to the ICUs between 2001 and 2012 in the Beth Israel Deaconess Medical Center in Boston, MA, United States. The eICU-CRD datasets included patients who were admitted to the ICUs between 2014 and 2015 across 208 hospitals in the United States.[22,23] Patients who were younger than 18 years, those who died in the ICUs or refused life-sustaining treatment (e.g., do-not-resuscitate (DNR)) were excluded. Supplementary Figure S1 shows the detailed exclusion criteria. In the MIMIC-III dataset, adjacent ICU admissions within 24 h were merged because the transfers between the ICUs and operating rooms could not be verified.[22] For the eICU-CRD dataset, step-down units were considered general wards, and admissions to those units were excluded from the study.[23]

Data collection and preprocessing

Fig. 1a illustrates the data collection process. Data for the development and internal validation were collected from the electronic health records and clinical data warehouse of the SNUH. The external validation data was collected from the MIMIC-III and eICU-CRD after signing the data use agreement on PhysioNet. Fig. 1b shows the annotation process, where the 'normal' label was defined as hospital discharge after ICU discharge. The 'event' label was defined as either ICU readmission from the general ward, a DNR order in the general ward, or death in the general ward. The criteria for ICU readmission refer to the patient being readmitted to the same ICU during the same hospitalization after being originally discharged based on a clinical assessment.[1] Events were categorized as early (≤48 h after discharge) and late (>48 h) ICU readmission. The SNUH cohort proceeded with a strict review by clinical experts (LL, DS), whereas the MIMIC-III and eICU-CRD datasets relied on technical annotations based on publicly available data.

Prediction is performed at the time of ICU discharge, and the model predicts the risk of subsequent outcome labels using input variables collected during the entire ICU stay, from ICU admission to ICU discharge. Various candidate input variables were initially selected by the clinical experts (LL, DY) through literature review and subsequently refined via greedy backward elimination, enhancing model simplicity and performance.[13,27] The complete list of input variables is presented in Supplementary Table S1. Medical codes corresponding to the identical input variable were integrated, and measurement timestamps were collected in minute-level units. Multiple measurements per timestamp were averaged. Outliers beyond the valid data range were excluded (Supplementary Table S2).[24] Continuous variables were normalized to a standard normal distribution, while the categorical variables were one-hot encoded; both were then arranged in a time-by-feature matrix. Missing values were handled by forward-fill imputation, and any remaining were replaced with variable means.

The difference in the distribution of the input variables was assessed across the cohorts. Differences in continuous

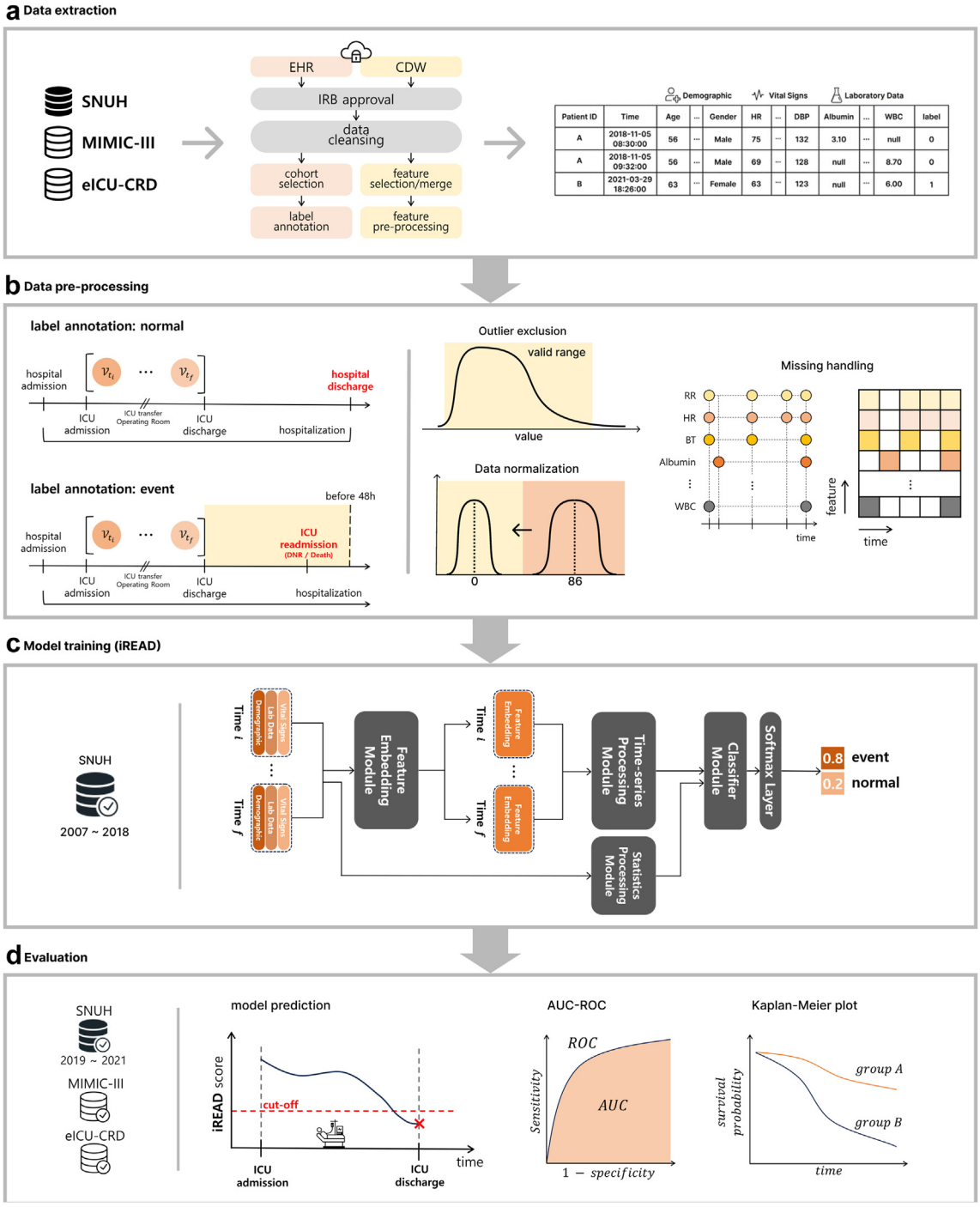

***Fig. 1:* An overview of model development and validation process**. (a) A process of extracting data from various hospital databases, including electronic health records (EHR) and clinical data warehouse (CDW). (b) Preprocessing steps, including label annotation for ICU readmissions within 48 h, outlier exclusion, data normalization, and handling of missing data. (c) An architecture of the iREAD model. (d) Evaluation of the model's performance, including quantitative metrics with AUROC, Kaplan–Meier plot, and qualitative metrics with prediction score trend.

variables were examined using a one-way analysis of variance or the Kruskal–Wallis H test, based on normality and variance homogeneity, and categorical variables were evaluated using the Chi-square test.[28] The box plot method was used to visualize differences in distributions across cohorts and label groups.

Model development

An ensemble machine learning model was developed to predict ICU readmission using a deep learning model for processing time-series data and a LightGBM model for analyzing statistical data. Both utilized patient information from the ICU stay only, as depicted in Fig. 1c.[29] The model generates a probability score at ICU discharge, representing the likelihood of the patient being readmitted to the ICU within 48 h. The deep learning component comprises modules for feature embedding, time series fitting, and classification. The feature embedding module consists of a fully connected network (FCN), layer normalization, and dropout, designed to effectively represent patient information at each time step.[30,31] The time-series processing module consists of stacked Long Short-Term Memory (LSTM) layers.[32] The classifier module consists of layer normalization and FCN. Finally, the outputs from the deep learning model and the LightGBM model are ensembled with equal weights of 0.5. Detailed architecture is illustrated in Supplementary Figure S2. The architecture was enhanced based on our previous research, demonstrating superior predictive performance across various studies, including prospective multicenter designs.[33–37] To optimize the model and prevent overfitting under the constraint of a low incidence rate in the dataset, we trained it by constructing selective balanced mini-batches corresponding to the incident data and scheduling a learning rate with early stopping.

The SNUH cohort dataset was split based on the period into the development dataset (2007–2018), which was split again into train and holdout datasets, for model training and hyperparameter optimization and the validation dataset (2019–2021) for internal validation. The MIMIC-III and eICU-CRD datasets were utilized only for external validation. All possible combinations of internal and external validation performance were assessed in supplementary experiments to examine cohort characteristics and internal performance for external datasets.

Performance evaluation

To evaluate predictive performance, we employed the area under the receiver operating characteristic curve (AUROC) to compare the developed model against clinical criteria, including the MEWS, national early warning score (NEWS), SWIFT, acute physiology and chronic health evaluation II (APACHE-II), and the single-parameter track-and-trigger system (SPTTS), as well as conventional machine learning models, including logistic regression (LR), random forest (RF), and Cox proportional hazard regression.[12,38–41] The AUROC scores were calculated from the model score at ICU discharge along with the outcome labels (Fig. 1d). We also calculated additional performance metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Youden's index. For a fair comparison, we set the threshold at the same specificity of 0.9 or close to 0.9 across models, as high specificity for reducing false positive alarms regarding the low event ratio. We conducted a comprehensive cross-dataset validation by training and testing the model on all possible combinations of the datasets to identify the potential margin of generalization.

Bootstrapping of 1000 random subsamples from each dataset, with all subsample sizes equal to the dataset size, and DeLong's method were performed to calculate 95% confidence intervals. Differences between our model and each baseline model were tested using a two-tailed paired t-test.[42]

To identify performance reliability, we assessed calibration curves to the prediction models with expected calibration errors. The curves compared the mean predicted probability of model outcomes against the observed fraction of positives, with normalized corrections to mitigate the extreme imbalance rate. The expected calibration error was the average absolute difference between the mean predicted probability and the fraction of positives. All models, including traditional scores, were recalibrated on the development set for a fair comparison.

Feature importance

The Shapley additive explanation (SHAP) method was used for model interpretability, whereby the SHAP value of each feature explains the impact of the variable on the outcome.[43] Therefore, a positive SHAP value denoted an increased likelihood of ICU readmission, while a negative SHAP value suggested a decreased likelihood.

Sensitivity analysis

We assessed the robustness of the model to missing input variables. We incrementally introduced missing rates to the input variables, ranging from 0% to 100% in 10% intervals, by randomly removing data. The process was repeated 10 times to calculate the mean and standard deviation of the degradation in model performance. Additionally, since each input variable may have a different missing rate due to differences in measurement frequencies, we examined the model's performance when each input variable was completely missing, one at a time.

Subgroup analysis

We assessed the model's performance in the following subgroups: age, APACHE-II score, duration from hospital admission to ICU admission, ethnicity, ICU length of stay, ICU Type, language, and sex. The model performance was evaluated in each subgroup using AUROC and compared to the clinical criteria as a baseline.

Trend analysis

The trajectory of prediction scores leading up to ICU discharge was analyzed to understand the comparative performance of various clinical models, including iREAD, NEWS, SWIFT, and APACHE-II. Patients were divided into two groups, and their risk prediction scores were averaged in hourly intervals over 120 h before ICU discharge. The non-event group included patients discharged from the ICU without subsequent readmission, while the event group comprised patients readmitted to the ICU, reflecting the outcome label.

Survival analysis

We conducted the Kaplan–Meier analysis on both internal and external datasets to estimate survival functions for two patient groups, categorized into high-risk and low-risk groups.[44] The high-risk groups had risk prediction scores above the threshold, while the low-risk groups had scores below it. The threshold was set at the same specificity between predictive models for a fair comparison. The specificity was set to 0.99, as high specificity is required in practical usage to reduce false positive alarms, which is common when event ratios are extremely low. The Kaplan–Meier analysis provides a visual representation of the time-to-event outcome across different patient groups. The differences between these curves were assessed using the log-rank test.[44] This non-parametric statistic enabled the comparison of prognoses between target groups and statistical information on the time points at which events occurred most frequently. We used the Kaplan–Meier cumulative event probability approach to emphasize event occurrences in contrast to the survival probability.[44]

Role of the funding source

The funder of this study had no role in the study design, data collection, analysis, and interpretation, model development and validation, reporting of the results, approval of the manuscript, and decision to submit the manuscript for publication.

Results

Study cohorts

Data were extracted from the cohorts, as detailed in Fig. 1a and b. After exclusion, the final cohorts consisted of 70,842, 43,237, and 90,271 ICU admissions for SNUH, MIMIC-III, and eICU-CRD, respectively (Supplementary Figure S1). Overall ICU readmission rates were 9.34%, 8.73%, and 9.83%, and early readmission rates were 2.23%, 2.23%, and 4.80% for SNUH, MIMIC-III, and eICU-CRD, respectively. Supplementary Figure S3 shows the distribution of ICU readmission over time. Supplementary Table S1 presents the characteristics of input variables according to the cohorts and Supplementary Table S3 shows the general characteristics and additional information, indicating significant differences across the datasets. Additionally, Supplementary Figure S4 offers a visual comparison of input feature distributions across the cohorts.

Model performance

A deep learning-based prediction model for ICU readmission, named iREAD, was developed (Fig. 1c and d). Fig. 2 demonstrates its performance against traditional scores, such as MEWS, NEWS, SWIFT, APACHE-II, and SPTTS, as well as conventional machine learning models, such as LR, RF, and Cox proportional hazard regression. In the internal validation using the SNUH dataset, the iREAD showed superior performance in early (≤48 h), late (>48 h), and overall ICU readmissions, with AUROCs of 0.771 (95% confidence interval 0.743–0.798), 0.834 (0.821–0.846), and 0.820 (0.808–0.832) for early (≤48 h), late (>48 h), and overall ICU readmissions, respectively (All $P < 0.001$) (Table 1). External validations on the MIMIC-III and eICU-CRD datasets suggested their higher performance and generalizability, with AUROCs of 0.726 (0.687–0.764), 0.782 (0.760–0.804), and 0.768 (0.748–0.787) in MIMIC-III, and 0.686 (0.665–0.707), 0.759 (0.742–0.776), and 0.725 (0.712–0.739) in eICU-CRD, for early (≤48 h), late (>48 h), and overall ICU readmissions, respectively (All $P < 0.001$) (Table 1). The confidence interval calculated with DeLong's method also showed similar results (Supplementary Table S4). Additional performance metrics are presented in Supplementary Table S5, demonstrating that iREAD achieved the highest F1 score and Youden's index among all models in both internal and external validation.

Supplementary Table S6 details iREAD's performance in AUROCs across all combinations of the training and testing datasets, identifying cohort characteristics and internal performance for external datasets. Similar performance was observed between MIMIC-III and eICU-CRD, in contrast to SNUH, indicating distinct cohort characteristics. Notably, training on the multicenter eICU-CRD dataset resulted in high performance on the single-center MIMIC-III dataset, but the reverse was less effective.

Calibration curves for various prediction models are presented in Fig. 3 iREAD followed the dashed diagonal line, with the expected calibration error of 0.047, 0.045, and 0.029, in SNUH, MIMIC-III, and eICU-CRD, respectively, while the others did not. Therefore, iREAD was almost perfectly calibrated; the predicted probability precisely matched the observed frequencies.

Feature importance

The interpretability of our model was provided using the SHAP method. Fig. 4 illustrates the individual feature importance. Fig. 4a enumerates the input features ranked by their mean absolute SHAP values, providing a hierarchy of their impact on the model's predictions. Fig. 4b further explores the joint distribution of selected features

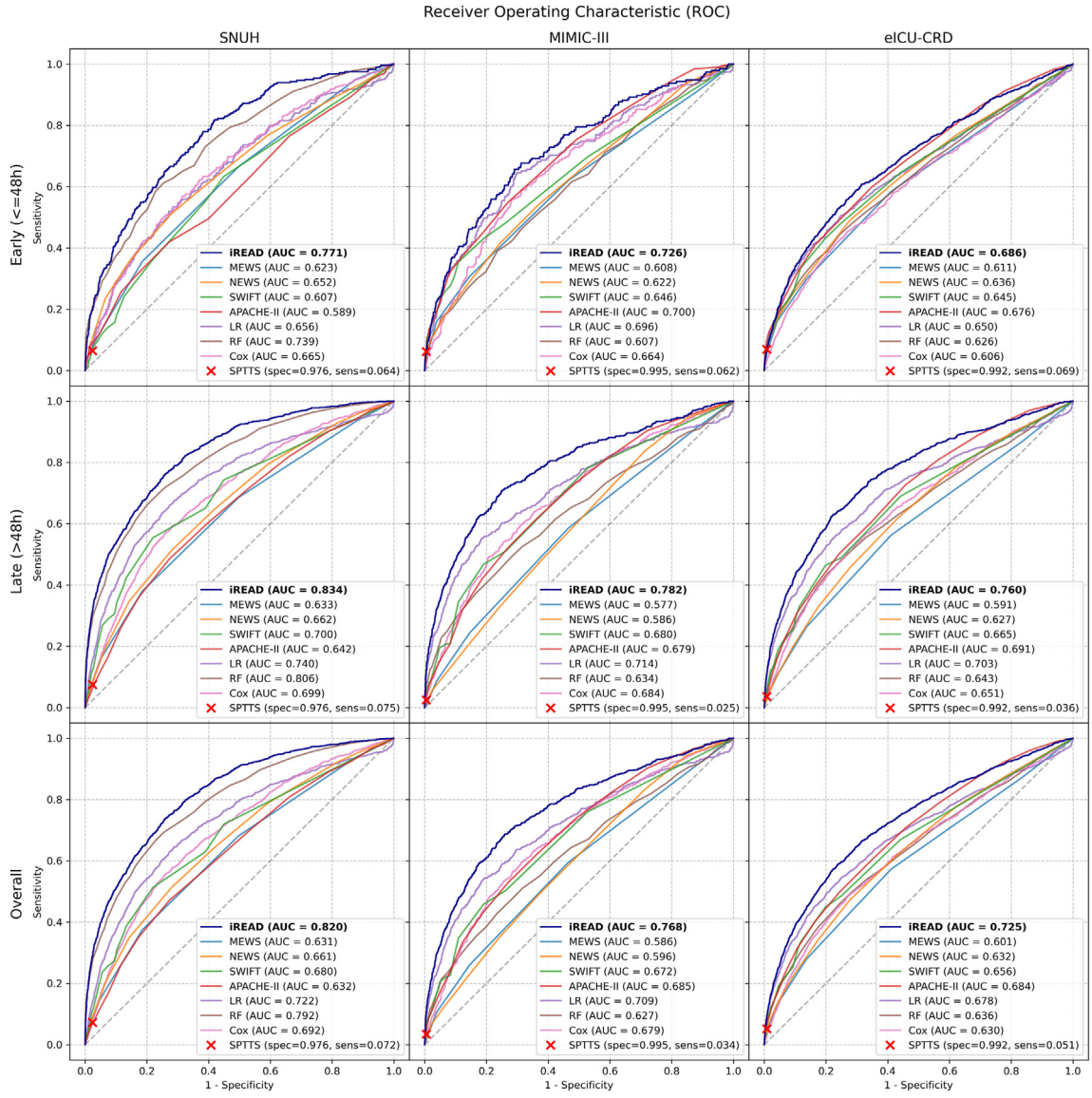

Fig. 2: **Performance of the model predicting ICU readmission based on a 48-h threshold for internal validation and MIMIC-III and eICU-CRD for external validation**. '**' denotes a P-value <0.001. SNUH, Seoul National University Hospital; MIMIC-III, Medical Information Mart for Intensive Care III; eICU-CRD, eICU Collaborative Research Database; MEWS, modified early warning score; NEWS, national early warning score; SWIFT, stability and workload index for transfer; APACHE-II, acute physiology and chronic health evaluation II; SPTTS, single-parameter track-and-trigger system; AUC, area under the curve.

and SHAP values, detailing how variations in input features interact with the risk of ICU readmission. The top three time-dependent variables were peripheral oxygen saturation (SpO_2), respiratory rate, and heart rate, excluding the static variables of demographics. Longer ICU stays, and hospital admission before the ICU stay also emerged as significant risk factors for ICU readmission.

Sensitivity analysis

The missing rates are described in Supplementary Table S7. Supplementary Figure S5 presents the results of the sensitivity analysis on early ICU readmission using the iREAD model according to the rate of the missing input data. The iREAD model shows robustness to missing data in input variables with high missing rates, such as lab results, while it is more sensitive to missing data in variables with low missing rates, such as vital signs (Supplementary Figure S5b). When a random missing rate was artificially introduced, even with a missing rate as high as 90%, the performance drop remained within −5% (Supplementary Figure S5a).

Model	Internal validation	External validation	
	SNUH	MIMIC-III	eICU
	AUROC (95% CI)	AUROC (95% CI)	AUROC (95% CI)
Early readmission (≤48 hrs)			
iREAD	0.771 (0.743–0.798)	0.726 (0.687–0.764)	0.686 (0.665–0.707)
MEWS	0.623 (0.590–0.655)	0.608 (0.565–0.650)	0.611 (0.591–0.632)
NEWS	0.652 (0.619–0.685)	0.622 (0.581–0.663)	0.636 (0.616–0.656)
SWIFT	0.607 (0.576–0.638)	0.646 (0.603–0.689)	0.646 (0.625–0.666)
APACHE-II	0.589 (0.554–0.625)	0.700 (0.662–0.738)	0.676 (0.657–0.695)
LR	0.656 (0.622–0.689)	0.696 (0.654–0.739)	0.650 (0.629–0.672)
RF	0.739 (0.710–0.769)	0.607 (0.563–0.650)	0.626 (0.605–0.648)
Cox	0.665 (0.633–0.697)	0.664 (0.625–0.703)	0.606 (0.584–0.628)
Late readmission (>48 hrs)			
iREAD	0.834 (0.821–0.846)	0.782 (0.760–0.804)	0.759 (0.742–0.776)
MEWS	0.633 (0.616–0.651)	0.577 (0.553–0.602)	0.591 (0.572–0.610)
NEWS	0.662 (0.645–0.680)	0.586 (0.563–0.610)	0.627 (0.609–0.646)
SWIFT	0.700 (0.682–0.717)	0.680 (0.657–0.703)	0.666 (0.648–0.684)
APACHE-II	0.642 (0.624–0.660)	0.679 (0.656–0.702)	0.691 (0.674–0.708)
LR	0.740 (0.721–0.758)	0.714 (0.689–0.740)	0.703 (0.683–0.723)
RF	0.806 (0.792–0.821)	0.634 (0.608–0.660)	0.643 (0.624–0.663)
Cox	0.699 (0.682–0.716)	0.684 (0.660–0.707)	0.651 (0.631–0.670)
Overall readmission			
iREAD	0.820 (0.808–0.832)	0.768 (0.748–0.787)	0.725 (0.712–0.739)
MEWS	0.631 (0.615–0.647)	0.586 (0.563–0.608)	0.601 (0.586–0.615)
NEWS	0.661 (0.644–0.677)	0.596 (0.575–0.617)	0.632 (0.618–0.646)
SWIFT	0.680 (0.664–0.695)	0.672 (0.651–0.694)	0.656 (0.642–0.671)
APACHE-II	0.632 (0.615–0.648)	0.685 (0.665–0.706)	0.684 (0.670–0.697)
LR	0.722 (0.706–0.738)	0.709 (0.687–0.732)	0.678 (0.664–0.692)
RF	0.792 (0.779–0.806)	0.627 (0.604–0.651)	0.636 (0.621–0.650)
Cox	0.692 (0.677–0.708)	0.679 (0.658–0.700)	0.630 (0.616–0.644)

P-value <0.001 for all comparisons; iREAD as a reference. SNUH, Seoul National University Hospital; MIMIC-III, Medical Information Mart for Intensive Care III; eICU-CRD, eICU Collaborative Research Database; MEWS, modified early warning score; NEWS, national early warning score; SWIFT, stability and workload index for transfer; APACHE-II, acute physiology and chronic health evaluation II; LR, Logistic Regression; RF, Random Forest; Cox, Cox Proportional Hazards Regression.

Table 1: **Performance of the model predicting ICU readmission based on a 48-h threshold.**

Subgroup analysis

Supplementary Table S8 presents the results of the subgroup analysis on early ICU readmission. The iREAD model consistently outperformed the clinical criteria baseline in AUROC. Particularly, for the length of ICU stay and duration from hospital admission to ICU admission, the iREAD model's performance was outstanding across subgroups and cohorts. For ICU type and APACHE-II score, SNUH showed consistent performance across subgroups, whereas MIMIC-III and eICU-CRD presented slight variations. Regarding ethnicity, SNUH was entirely composed of Asians (specifically Koreans), and in MIMIC-III, Asians showed the highest performance, and in eICU-CRD, Caucasians showed the highest performance.

Trend analysis

In the trend analysis, the risk from the iREAD model clearly distinguished the two groups (Supplementary Figure S6). For the non-event group, the prediction trends exhibited a stable decreasing pattern as discharge approached, aligning with the expected reduction in clinical risk. Conversely, for the event group, the scores fluctuated more noticeably, often increasing risk levels as the discharge time got closer, suggesting possible patient deterioration. This separates the two trend lines clearly in the figure. In contrast, the NEWS, SWIFT, and APACHE-II scores showed overlaps between the two groups, with fluctuating scores for both. The lack of a clear trend in these models complicates the ability to distinguish between outcomes, highlighting the iREAD model's superior predictability.

Survival analysis

The Kaplan–Meier analysis following ICU discharge, as depicted in Fig. 5, illustrated the predictive capabilities of various clinical models, including iREAD, NEWS,

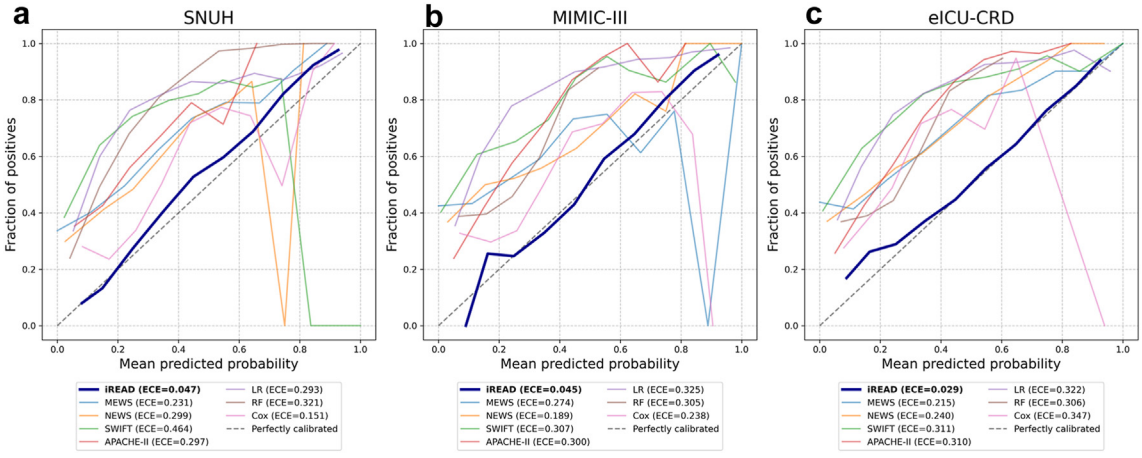

Fig. 3: **Calibration curves of the model predicting ICU readmission across.** (a) SNUH, (b) MIMIC-III, and (c) eICU-CRD. A perfectly calibrated model would follow the dashed diagonal line, where the predicted probability precisely matches the observed frequencies. ECE, expected calibration error.

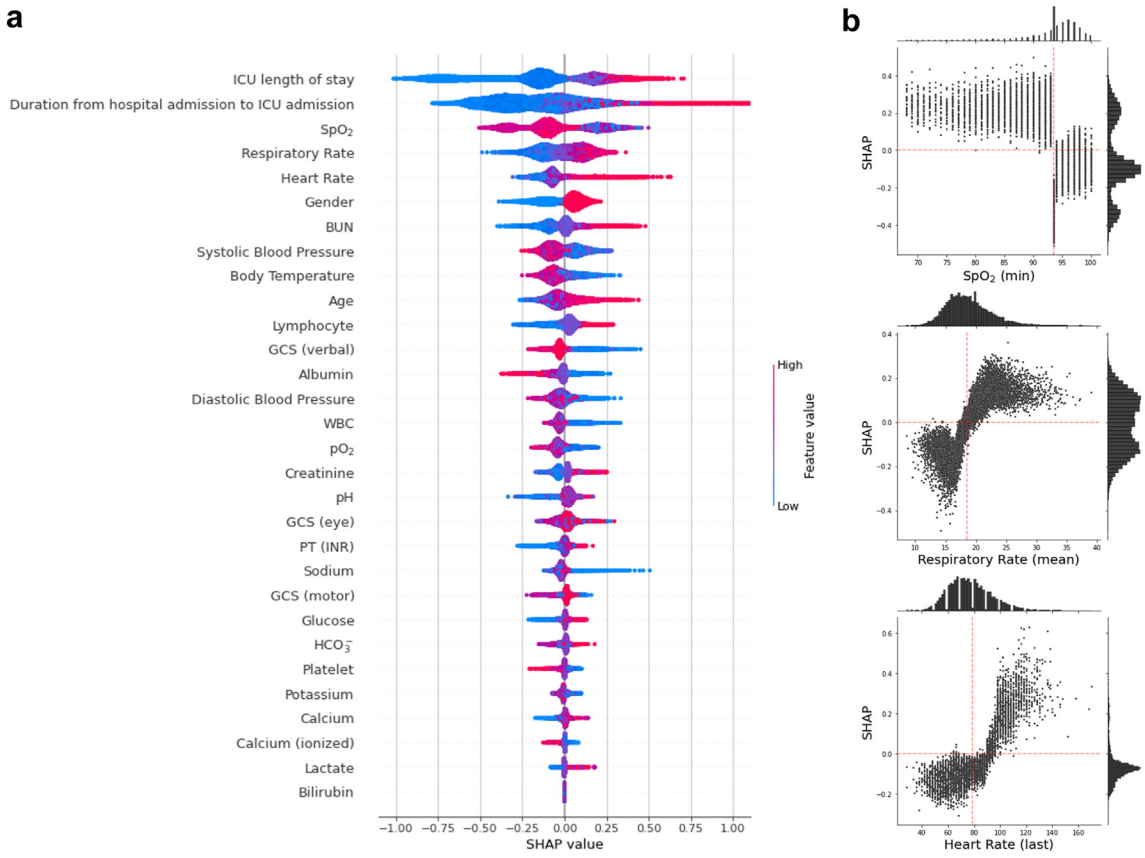

***Fig. 4:* Visualization of individual feature importance for predicting ICU readmission**. (a) Ranks of the mean absolute Shapley additive explanations (SHAP) values of all input features. (b) Detailed joint distribution of selected features; the correlation between feature values and their corresponding SHAP values. Red dashed lines are drawn at the x-axis mean feature value, indicating normal level, and at the y-axis SHAP value of 0, indicating no impact on the outcome.

SWIFT, and APACHE-II for internal validation. The iREAD model demonstrated a substantial differentiation in the occurrence of ICU readmission between patients in the high-risk and low-risk groups ($P < 0.001$ in the log-rank test). In traditional clinical models such as NEWS, SWIFT, and APACHE-II, ICU readmissions occurred in only about 10% of the high-risk group. Moreover, in the case of SWIFT, no significant difference ($P = 0.655$) was found between the two groups. However, with the iREAD model, more than 40 percent of patients in the high-risk group were readmitted. In the external validation, as presented in Supplementary Figure S7, the iREAD model outperforms traditional clinical models.

Discussion

This study developed and validated iREAD, a machine learning-based model for predicting ICU readmission within 48 h after discharge. The model demonstrated superior performance compared to traditional scores in both internal and external validation cohorts. While prediction for ICU readmission prediction has been widely investigated by researchers, iREAD introduces several unique contributions that distinguish it from prior models, emphasizing its potential to advance clinical practice.

Previous studies on ICU readmission have mainly utilized static variables such as measurements at admission or discharge, diagnoses that led to ICU admission, or routes of ICU admission.[1,8,12,21,27,45] However, models based on static variables have inherent limitations in capturing trends that reflect changes in patient conditions during ICU stays. The iREAD utilized time-series data using a deep learning model, incorporating all input variables recorded from ICU admission to the point of discharge decision as input variables, enabling the model to capture trends and reflect changes in patient condition during the ICU stay. Furthermore, our feature importance results showed that the time series data, such as SpO_2, respiratory rate, and heart rate, are important for readmission prediction (Fig. 4). In contrast to the SWIFT score, one of the most widely used scoring systems for the prediction of

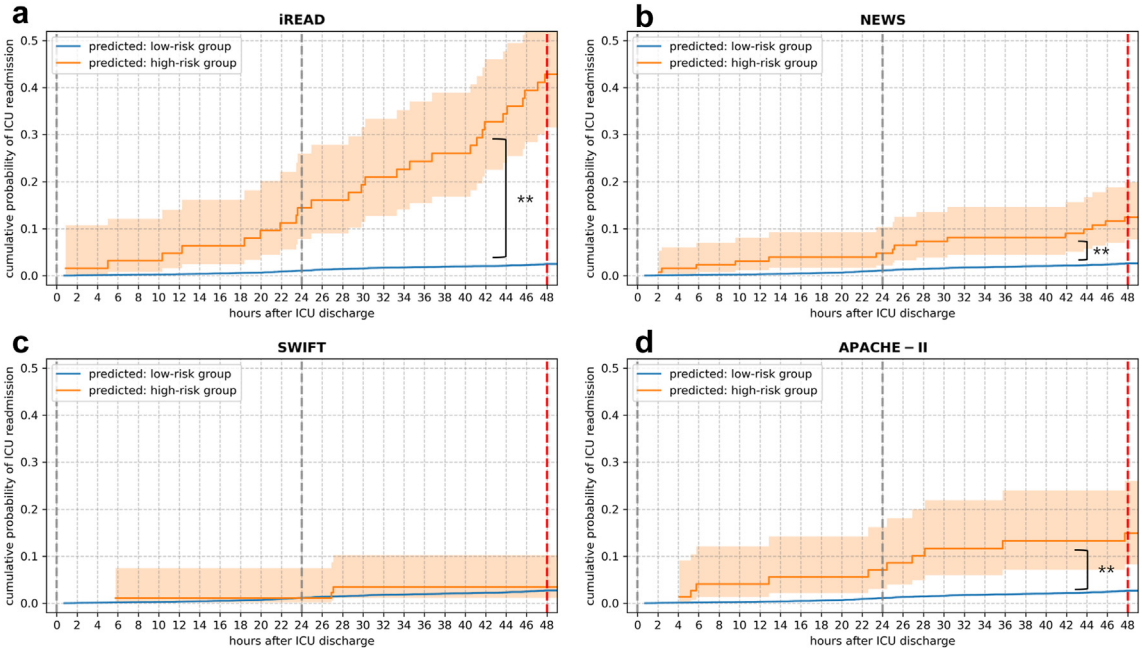

Fig. 5: **Kaplan–Meier cumulative probability of ICU readmission of each prediction score within 48 h after ICU discharge**. (a) iREAD, (b) NEWS, (c) SWIFT, and (d) APACHE-II. Grey dashed lines are drawn at 24-h intervals, and the time 48 h after ICU discharge is marked with a red dashed line. '**' denotes P-value <0.001.

unplanned ICU readmission based on only five static variables,[12] exhibits limited utility in clinical practice due to its low performance,[13,46] iREAD demonstrated superior performance on both internal and external datasets, highlighting its potential for the use as a clinical decision-supporting tool.

Although several ICU readmission prediction models based on machine learning techniques have been suggested with the advancement of machine learning techniques and the accumulation of patient data during ICU stays,[27,45,47–49] they showed modest performances, used static rather than dynamic data with time series data, and focused on the prediction of long-term readmission[27,47–49] or were not externally validated.[27] The iREAD was designed to address these limitations by assessing the predictive performance for ICU readmission separately within or beyond 48 h, demonstrating robust performance for both short- and long-term readmission risk. The iREAD's performance was also assessed through various approaches, including external validations on different populations with different nationality and races (MIMIC-III and eICU-CRD) and comparing the performances not only to traditional prediction models like SWIFT but also to conventional machine learning models such as RF and LR, consistently demonstrating superior outcomes (Table 1 and Fig. 2). The potential generalizability of the iREAD across varied healthcare settings and patient populations also mitigates the risk of models based on homogeneous patient groups, such as neurologic or cardiac surgery patients, or patients with treatment limitations such as DNR.[21]

The SHAP method applied to the iREAD model addresses the inherent black-box nature of deep learning algorithms by providing interpretability and explaining how and why specific predictions are made.[50] The method offers insights into the factors influencing the model's decision-making process by quantifying the SHAP values of variables contributing to the predictions. Furthermore, it holds the potential to identify modifiable risk factors that contribute to increased readmission risk, thus enabling clinicians to take corrective actions. For instance, Fig. 4b illustrates how specific dynamic variables influenced the risk of ICU readmission by analyzing the trends of the top three ranked modifiable variables. Minimum SpO_2, mean respiratory rate, and heart rate measured at the time of ICU discharge were identified as being associated with readmission risk. Clinicians can evaluate the trends and values of these variables throughout the ICU stay, identify modifiable or correctable factors, and reassess the appropriateness of ICU discharge to mitigate the risk of readmission. This can also apply to other variables, as iREAD primarily utilizes modifiable variables, with the exception of non-modifiable factors such as length of ICU stay, duration from hospital admission to ICU admission, age, and gender. Furthermore, after ICU discharge, implementing more intensive

monitoring of these variables and utilizing supportive devices could facilitate early identification of patients at risk, thereby reducing the likelihood of ICU readmission.

With regard to the real-world application of predictive models, Thoral et al. recently proposed an explainable machine learning model for predicting mortality or ICU readmission within 7 days after discharge.[20] This model demonstrated good performance in both internal and external validations, achieving a significant relative risk reduction (14%) using 180 variables, including those requiring specific observations such as bronchial suctioning and cough reflex.[20,51] In contrast, the iREAD utilizes only 30 routinely measured or easily collectible variables from the EHR (e.g., GCS), without the need for specialized observations or additional monitoring. This simplicity allows the iREAD to be seamlessly implemented into clinical workflows, providing automated risk calculations at the time of discharge decision-making without requiring extra devices or assessments. The robustness of the iREAD, even with a high rate of missing data, also supports the clinical utility of the model (Supplementary Figure S5).

Furthermore, the iREAD's unique features enhance its utility in clinical practice. The risk score, combined with SHAP values, enables clinicians to reassess the appropriateness of discharge decisions and either proceed with or delay discharge based on individualized risk assessments. While SHAP values for modifiable factors provide actionable insights—helping clinicians understand why a patient is at high risk of readmission and how to mitigate that risk—non-modifiable factors complement these insights by contributing to risk stratification when combined with modifiable variables. Based on this comprehensive assessment, clinicians can determine appropriate post-discharge interventions, such as increased frequency of laboratory tests or vital sign monitoring, the application of advanced therapeutic devices, or transfer to a step-down unit. These measures, coupled with high-level surveillance and rapid response systems, may reduce ICU readmission risk. The examples of the clinical application of the iREAD are illustrated in Supplementary Figure S8. However, future studies are needed to validate the effectiveness of the iREAD in diverse clinical settings and patient populations.

Even for patients whose ICU discharge is delayed, iREAD offers guidance on mitigating risk for a safer eventual discharge. Moreover, in critical care settings where ICU bed shortages necessitate difficult discharge decisions,[4] iREAD can be a tool for optimizing resource utilization. By enabling rapid assessment of individual patient readmission risks, iREAD supports clinicians in screening and selecting patients for safe discharge while prioritizing resources for those in greater need. This capability ensures that discharge decisions are both clinically sound and resource-efficient, addressing the challenges posed by limited ICU capacity.

The excellent calibration of the iREAD with low expected calibration errors in both internal and external datasets indicates that the predicted probabilities closely match the observed frequencies of ICU readmission, enhancing the model's reliability in clinical decision-making. While several readmission prediction models or risk scoring systems have demonstrated good calibration on internal datasets, these studies did not provide calibration results on external datasets.[8,52–54] In previous studies, the SWIFT exhibited poor calibration on external datasets,[46] and another predictive model required recalibration due to poor calibration.[51] In contrast, the iREAD exhibited excellent calibration on both internal and external datasets (MIMIC-III and eICU-CRD), supporting its generalizability across diverse clinical settings. The well-calibrated prediction model can be useful for planning and optimizing clinical care and allocating resources in ICU settings.

The trend and Kaplan–Meier analysis are distinct features of this study that were rarely evaluated in other studies. In the trend analysis, iREAD consistently distinguished the two groups (non-event and event) from 120 h before discharge. Although the SWIFT also exhibited a similar trend over time, it showed a tendency to fluctuate with a reduced discriminative power between the two groups, except during the 18 h prior to discharge. This dynamic risk assessment capability and early discrimination of the risk groups could provide clinicians with the possibility of planning for discharge and post-ICU care strategies. Notably, over 40% of patients classified as high-risk by iREAD were readmitted within 48 h, compared to approximately 10% for traditional scores at the same specificity, demonstrating its potential to improve clinical outcomes. In external cohorts, while traditional scores showed increased cumulative probabilities of readmission—exceeding 30% in the eICU-CRD dataset—iREAD consistently maintained cumulative probabilities above 40%, outperforming traditional models. Despite these promising results, further studies are needed to prospectively validate iREAD's performance, integrate it into clinical decision-making processes, and evaluate its impact on patient outcomes.

The results of the subgroup analysis highlighted the limitations of iREAD and emphasized the need for adjustments before its application to different populations. Although iREAD demonstrated modest performance in the external validations, it exhibited performance degradation in several subgroups, particularly among patients with an APACHE-II score of less than 10 or 15, certain age groups, and specific ethnicities. We suggested that the relatively small number of events in these subgroups might have contributed to the lower performance of iREAD. However, performance degradation in AI models for specific subgroups raises concerns about fairness in healthcare and may indicate potential data biases, such as minority, informativeness,

or training-serving skew bias, originating from the development phase. To address these issues, adjustments to iREAD through data augmentation or fine-tuning should be implemented to ensure consistent predictive performance across all subgroups. Additionally, future studies using data from diverse patient populations could further mitigate performance degradation in specific subgroups.

The results of the subgroup analysis showed the weakness of the iREAD and suggested the requirement of the adjustment before its applications to the different populations. Though the iREAD showed modest performances in the external validations, the iREAD showed performance degradations in several subgroups, especially in patients with APACHE-II score less than 10 or 15, several age groups, or ethinicities. We suggested that the relatively small number of events that occurred in such groups might affect the low performances of the iREAD. However, as the performance degradation of the AI model in specific subgroups can compromise the fairness of healthcare and can reflect the potential of the model's data bias such as minority, informativeness, or training-serving skew bias a datda bias arrised from the development phase,[55] the adjustment of the iREAD with data augmentation or fine tunning should be proceed to maintain adequate predictive performance for all subgroups. Future prospective studies on different patient populations can also help mitigate the performance degradation on specific subgroups.

This study has several limitations. First, our model was developed using data from a single tertiary center with several types of sub-ICU that were specialized for close monitoring or capable of applying devices such as low-level mechanical ventilators in the general ward, suggesting higher severity of illness of patients in general wards. Second, all validations were retrospective, resulting in the possibility of inaccuracy in outcome labels or the measurement of some variables and inconsistency in the criteria for initial ICU discharge and readmission across patients. Therefore, prospective validation of the model across diverse patient populations is essential before its clinical application. Third, the model's performance varied across subgroups, necessitating further refinement may be needed to ensure consistent accuracy across diverse patient populations. And fourth, there is a potential risk of overfitting. Since our model was developed using data from a single tertiary center within one country, achieving international generalization may be demanding, not only in terms of model performance but also in model calibration for a consistent, practical application. To mitigate this risk, deep learning techniques, such as selective balanced mini-batches and scheduling a learning rate with early stop, were introduced into the model training process, with isolation from the validation dataset. Although the iREAD showed statistically superior performance (Table 1 and Supplementary Table S5) and excellent calibration (Fig. 3) on both internal and external datasets, and efforts were made to mitigate the risk of overfitting, the potential margin of generalization remains (Supplementary Table S6). Comprehensive prospective research or fine-tuning should be conducted before its application in real practice.

In conclusion, iREAD represents a significant advancement in predicting ICU readmission risk, outperforming traditional scoring systems across diverse patient cohorts. Its robust performance, excellent calibration, and ability to provide dynamic risk assessments make it a promising tool for improving ICU discharge decision-making and potentially reducing unnecessary readmissions. If future studies validate the utility of the model in diverse clinical settings, models like iREAD could play a crucial role in enhancing patient safety in the ICU.

Contributors

LL made substantial contributions to the study's conception and design, data acquisition, access and verification of the data, analysis, and the initial drafting of the manuscript. MK was instrumental in accessing and verifying the data, processing and analyzing data, conducting experiments, collecting and visualizing results, and writing the sections of the manuscript related to methodology and results. KC contributed significantly by enhancing the performance of the model, supervising the study, and proposing critical experiments. DY provided essential suggestions for experiments, offered clinical insights for selecting input features, and confirmed the validity of the experiments. DS was responsible for institutional data extraction, collection, cleaning, verifying, and structuring. HGR was involved in the selection and interpretation of variables and data analysis. HL participated in drafting the article and critically revising it for important intellectual content. All authors have reviewed and approved the final version of the manuscript for publication.

Data sharing statement

The institutional dataset used for model development and internal validation, along with the de-identified results from this study, can be made available upon reasonable request to the corresponding author.

Declaration of interests

All authors declared that they have no competing interests that could influence the work.

Acknowledgements

This work was supported by the Korea Health Technology Research & Development Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number RS-2021-KH114109).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2025.103112.

References

1 Woldhek AL, Rijkenberg S, Bosman RJ, van der Voort PH. Readmission of ICU patients: a quality indicator? *J Crit Care*. 2017;38:328–334.
2 Kramer AA, Higgins TL, Zimmerman JE. The association between ICU readmission rate and patient outcomes. *Crit Care Med*. 2013;41:24–33.

3 Berenholtz SM, Dorman T, Ngo K, Pronovost PJ. Qualitative review of intensive care unit quality indicators. *J Crit Care*. 2002;17:1–12.
4 Kramer AA, Higgins TL, Zimmerman JE. Can this patient be safely discharged from the ICU? *Intensive Care Med*. 2016;42:580–582.
5 Stelfox HT, Lane D, Boyd JM, et al. A scoping review of patient discharge from intensive care: opportunities and tools to improve care. *Chest*. 2015;147:317–327.
6 Plotnikoff KM, Krewulak KD, Hernandez L, et al. Patient discharge from intensive care: an updated scoping review to identify tools and practices to inform high-quality care. *Crit Care*. 2021;25:438.
7 Bai J, Fugener A, Gonsch J, Brunner JO, Blobner M. Managing admission and discharge processes in intensive care units. *Health Care Manag Sci*. 2021;24:666–685.
8 Badawi O, Breslow MJ. Readmissions and death after ICU discharge: development and validation of two predictive models. *PLoS One*. 2012;7:e48758.
9 Frost SA, Alexandrou E, Bogdanovski T, et al. Severity of illness and risk of readmission to intensive care: a meta-analysis. *Resuscitation*. 2009;80:505–510.
10 Timmers TK, Verhofstad MH, Moons KG, Leenen LP. Patients' characteristics associated with readmission to a surgical intensive care unit. *Am J Crit Care*. 2012;21:e120–e128.
11 Campbell AJ, Cook JA, Adey G, Cuthbertson BH. Predicting death and readmission after intensive care discharge. *Br J Anaesth*. 2008;100:656–662.
12 Gajic O, Malinchoc M, Comfere TB, et al. The Stability and Workload Index for Transfer score predicts unplanned intensive care unit patient readmission: initial development and validation. *Crit Care Med*. 2008;36:676–682.
13 Loreto M, Lisboa T, Moreira VP. Early prediction of ICU readmissions using classification algorithms. *Comput Biol Med*. 2020;118:103636.
14 Reini K, Fredrikson M, Oscarsson A. The prognostic value of the Modified Early Warning Score in critically ill patients: a prospective, observational study. *Eur J Anaesthesiol*. 2012;29:152–157.
15 Jahn M, Rekowski J, Gerken G, Kribben A, Canbay A, Katsounas A. The predictive performance of SAPS 2 and SAPS 3 in an intermediate care unit for internal medicine at a German university transplant center; a retrospective analysis. *PLoS One*. 2019;14:e0222164.
16 Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care*. 2019;23:284.
17 Lim L, Gim U, Cho K, Yoo D, Ryu HG, Lee HC. Real-time machine learning model to predict short-term mortality in critically ill patients: development and international validation. *Crit Care*. 2024;28:76.
18 Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020;46:383–400.
19 Hong N, Liu C, Gao J, et al. State of the art of machine learning-enabled clinical decision support in intensive care units: literature review. *JMIR Med Inform*. 2022;10:e28781.
20 Thoral PJ, Fornasa M, de Bruin DP, et al. Explainable machine learning on AmsterdamUMCdb for ICU discharge decision support: uniting intensivists and data scientists. *Crit Care Explor*. 2021;3:e0529.
21 Ruppert MM, Loftus TJ, Small C, et al. Predictive modeling for readmission to intensive care: a systematic review. *Crit Care Explor*. 2023;5:e0848.
22 Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035.
23 Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178.
24 Hyland SL, Faltys M, Huser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med*. 2020;26:364–373.
25 Moor M, Bennett N, Plecko D, et al. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. *eClinicalMedicine*. 2023;62:102124.
26 Liu X, Hu P, Yeung W, et al. Illness severity assessment of older adults in critical illness using machine learning (ELDER-ICU): an international multicentre study with subgroup bias evaluation. *Lancet Digit Health*. 2023;5:e657–e667.
27 Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc*. 2018;15:846–853.
28 Lee SW, Lee HC, Suh J, et al. Multi-center validation of machine learning model for preoperative prediction of postoperative mortality. *NPJ Digit Med*. 2022;5:91.
29 Ke GL, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3149–3157. Nips 2017.
30 Ba JL. Layer normalization. *arXiv*. 2016.
31 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
32 Hochreiter S. *Long short-term memory. Neural Computation MIT-Press*. 1997.
33 Kim Y, Kim H, Choi J, et al. Early prediction of need for invasive mechanical ventilation in the neonatal intensive care unit using artificial intelligence and electronic health records: a clinical study. *BMC Pediatr*. 2023;23:525.
34 Cho KJ, Kim JS, Lee DH, et al. Prospective, multicenter validation of the deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest or unplanned intensive care unit transfer in patients admitted to general wards. *Crit Care*. 2023;27:346.
35 Cho KJ, Kim KH, Choi J, Yoo D, Kim J. External validation of deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest in patients admitted to general wards based on rapid response system operating and nonoperating periods: a single-center study. *Crit Care Med*. 2024;52:e110–e120.
36 Kim T, Tae Y, Yeo HJ, et al. Development and validation of deep-learning-based sepsis and septic shock early prediction system (DeepSEPS) using real-world ICU data. *J Clin Med*. 2023;12:7156.
37 Lee YJ, Cho KJ, Kwon O, et al. A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards. *Resuscitation*. 2021;163:78–85.
38 Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. *QJM*. 2001;94:521–526.
39 Jones M. NEWSDIG: the national early warning score development and implementation group. *Clin Med*. 2012;12:501–503.
40 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–829.
41 Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Higgins B. A review, and performance evaluation, of single-parameter "track and trigger" systems. *Resuscitation*. 2008;79:11–21.
42 James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. Berlin: Springer; 2013.
43 Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4768–4777. Nips 2017.
44 Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. *ACM Comput Surv*. 2019;51:1–36.
45 Pakbin A, Rafi P, Hurley N, Schulz W, Krumholz MH, Mortazavi JB. *Prediction of ICU readmissions using data at patient discharge. 2018 40th annual international conference of the IEEE engineering in medicine and biology society*. EMBC; 2018:4932–4935.
46 Kastrup M, Powollik R, Balzer F, et al. Predictive ability of the stability and workload index for transfer score to predict unplanned readmissions after ICU discharge. *Crit Care Med*. 2013;41:1608–1615.
47 Venugopalan J, Chanani N, Maher K, Wang MD. *Combination of static and temporal data analysis to predict mortality and readmission in the intensive care. 2017 39th annual international conference of the IEEE engineering in medicine and biology society*. EMBC; 2017:2570–2573.
48 Barbieri S, Kemp J, Perez-Concha O, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. *Sci Rep*. 2020;10:1111.
49 Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS One*. 2019;14:e0218942.
50 Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. *Inf Fusion*. 2022;77:29–52.
51 de Hond AAH, Kant IMJ, Fornasa M, et al. Predicting readmission or death after discharge from the ICU: external validation and retraining of a machine learning model. *Crit Care Med*. 2023;51:291–300.

52 Frost SA, Tam V, Alexandrou E, et al. Readmission to intensive care: development of a nomogram for individualising risk. *Crit Care Resusc*. 2010;12:83–89.
53 Jo YS, Lee YJ, Park JS, et al. Readmission to medical intensive care units: risk factors and prediction. *Yonsei Med J*. 2015;56: 543–549.
54 Hammer M, Grabitz SD, Teja B, et al. A tool to predict readmission to the intensive care unit in surgical critical care patients-the RISC score. *J Intensive Care Med*. 2021;36:1296–1304.
55 Ueda D, Kakinuma T, Fujita S, et al. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol*. 2024;42:3–15.