Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

A Smart Healthcare Management System for Predicting ICU Length of Stay Using Machine Learning and MIMIC-IV

Danish Ather *

Professor, Department of Computer Science and Engineering, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent 100084, Uzbekistan. E-mail: danishather@gmail.com

Dharm Raj

*Corresponding author, Assistant Professor, Department of Computer Science and Engineering, Sharda School of Engineering & Technology, Sharda University, Greater Noida, Uttar Pradesh 201306, India. E-mail: dharmraj4u@gmail.com

Suhail Javed Quraishi

Assistant Professor, Manav Rachna International Institute of Research and Studies, Sector 43, Faridabad, Haryana 121004, India. E-mail: suhailjaved.sca@mriu.edu.in

Sonal Pathak

Assistant Professor, Manav Rachna International Institute of Research and Studies, Sector 43, Faridabad, Haryana 121004, India. E-mail: sonal.sca@mriu.edu.in

Manik Rakhra

Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, India. E-mail: rakhramanik786@gmail.com

Kunchanapaalli Rama Krishna

Professor, Department of Computer Science and Information Technology, K L Deemed to be University, Vaddeswaram 522502, Andhra Pradesh, India. E-mail: tenalirama@kluniversity.in

Journal of Information Technology Management, 2026, Vol. 18, Issue 4, pp. 134-149.
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.108919
Article Type: Research Paper

Received: June 16, 2026
Received in revised form: July 07, 2026
Accepted: August 02, 2026
Published online: October 01, 2026

Abstract

Stays in ICUs are expensive to accommodate and difficult to forecast, yet even a simple forecast can be helpful, as it enables ICUs to allocate beds and staff and schedule transfer times. We address a particular problem of this kind: how much of the length of stay can be determined based on the first day of hospitalization? Our dataset is the publicly available MIMIC-IV Clinical Database Demo, from which we use 140 hospitalization records, retain the 117 that lasted at least one day, and derive 47 features from the first 24 hours of data on vital signs, medication and fluid intake, urine volume, and bedside procedures. Both formulations are addressed for the same set of patients. Exact forecasting of duration turned out to be rather complicated in our case: despite selecting the best-performing regression technique (random forest), the resulting mean absolute error was 2.85 days, barely outperforming the mean baseline predictor and producing a negative value for the coefficient of determination due to a few very long stays contributing disproportionately to the squared error. Yet, what has been accomplished is the ranking, where the Spearman rank correlation.

Keywords: ICU Length of Stay, Machine Learning, Electronic Health Records, MIMIC-IV, Predictive Modelling, Critical Care

Introduction

Few hospital components burn through the budget like an intensive care unit (ICU), where a large portion of costs is driven by the same factor: time spent in the unit. The predictive capabilities of a team capable of determining the probability of discharge on day two versus staying for two weeks provide some cushion for planning ahead. Resources will flow more reliably, assignments will have time to be arranged, and a step-down transfer could be planned in advance of necessity (Le Glaz et al., 2021; Mendo et al., 2021; Toy et al., 2024). The trouble is that length of stay is governed by many factors simultaneously—physiological, therapeutic, and administrative—and almost all of them reveal themselves only after the start of the admission. Determining the outcome of such a complex process in the first hours of the patient's stay is therefore tempting and elusive at the same time.

Predictive data science for critically ill patients has made great strides over the past decade, driven by the availability of anonymized electronic health records and advances in machine learning libraries that have reduced the cost of research (Baqar et al., 2025; Martinez-Millana et al., 2022; Nunes et al., 2025). Much of the progress has been aimed at predicting deterioration and mortality rather than length of stay (Baqar et al., 2025), and the main concern is always the trade-off between predictive strength and the level of trust a clinician is ready to place in a black-box model (Al Khatib et al., 2024; Porto, 2024). Length of stay is situated somewhat outside this dichotomy. It is a planning metric rather than an alert tool, and the cost of making mistakes is logistical rather than life-threatening, which provides

a different perspective on what constitutes a useful model and allows us to take a step back when interpreting the results.

The rapid evolution of AI-enabled electronic healthcare systems has significantly improved clinical decision support, secure health information exchange, and predictive analytics. Recent studies have emphasized the integration of artificial intelligence, cloud computing, and blockchain technologies to build scalable, interoperable, and patient-centric healthcare ecosystems capable of supporting intelligent clinical decision-making and resource optimization (R. Singhal et al., 2024; Verma & Nand, 2025).

The problem can be simply stated. Given only what was collected within the first 24 hours of an ICU admission, what can be done in terms of predicting length of stay, and which types of early data actually contain the information needed? Three research questions structure the paper. First, is the prediction of length of stay possible as a continuous target, and how good can the predictions be compared with a naive benchmark? Second, can a coarser version of the task be solved successfully, and by how much better? Third, what families of features, including vital signs, drugs, fluids, or procedures, drive the predictions in the trained models?

Three main contributions of the paper follow from the research questions above. First, it presents a fully reproducible pipeline for processing the original MIMIC-IV dataset into a condensed 24-hour feature collection. Second, the paper evaluates regression and classification on the same cohort in the same way, thus allowing comparison without the typical confounding associated with using a different dataset. Finally, the feature importances and the ceiling effect resulting from the demonstration sample size are discussed openly, so that there is no confusion about the type of result presented.

Literature Review

MIMIC datasets record detailed, timestamped events of real ICU patients. Each visit may produce thousands of entries: pressure and heart rate measurements recorded at the bedside, medication and fluid orders, laboratory tests, procedures, and outcomes such as urine output. All events reference a dictionary, d_items, where the numerical code corresponds to a meaningful name. To make sense of the tables, it is necessary to align different event streams to a common timeline starting from the time of ICU admission and to summarize all streams over an appropriate time window.

There seem to be two consistent approaches to defining the target variable. In the first, length of stay is defined as a continuous regression problem in terms of days or hours, while in the second, length of stay is dichotomized and a label is predicted. The continuous approach provides richer findings but is sensitive to the long right tail characteristic of the

length-of-stay distribution. The dichotomous approach loses resolution in return for robustness, which is especially important as the sample size decreases. Instead of choosing between them and justifying the choice, both approaches are applied.

Continual physiologic monitoring has been leveraged to identify decompensations in both emergency and inpatient departments, using recurrent models directly applied to streams of vital signs (Yang et al., 2025). Surveys of machine learning applications in emergency medicine chart the territory of the possible, as well as the poor reporting quality in the literature. Beyond clinical outcome prediction, recent intelligent healthcare systems have incorporated deep learning, secure electronic health record management, and AI-assisted decision support to improve diagnostic accuracy and healthcare service delivery across multiple clinical domains (P. Singhal et al., n.d.; R. Singhal et al., 2024). The second direction highlights the need for explanation: models predicting critical illness from electronic health records have been combined with attribution methods to allow clinicians to understand why certain decisions have been made, while similar concerns underlie personalized, trust-focused monitoring research.

The issue of false alarms and associated alarm fatigue continues to emerge as a problem. A tiered early warning system design has been proposed as a solution to the high rates of false positives when mortality prediction is considered. In sepsis, where minutes may be crucial, a meta-ensemble approach has been used to combine several base learners, providing attributions through SHAP methodology. Finally, more recent solutions involve combining clustering with temporal deep learning and sensor-based monitoring to triage emergency department patients, as well as explainable models based on consumer and IoT hardware for risk stratification outside hospital walls. In comparison to all of the above, length-of-stay prediction from early ICU data appears to be relatively sparse, while feasibility results from small cohorts like the current one still provide useful insights.

The proposed experimental pipeline follows a systematic data preprocessing, feature engineering, model training, and comparative evaluation strategy similar to recent intelligent predictive frameworks developed for healthcare and security applications (Jangir et al., 2025; Raj et al., 2025).

Dataset Description

The authors make use of the MIMIC-IV Clinical Database Demo, v. 2.2 (Johnson et al., 2023), a public subset designed for methods development purposes. Nine tables were extracted for this project. The ICU stays table contains data on 140 ICU stays involving 100 patients, with an overall mean length of stay of 3.68 days, a median of 2.16 days, and minimum and maximum lengths of stay ranging from 0.02 to 20.53 days. The event data are strongly dominated by chart events, containing 668,862 rows of charted vital signs and scores,

followed by input events (medication and fluid orders), with 20,404 rows; output events (9,362); and procedure events (1,468). The d_items item dictionary contains 4,014 items.

In order to have a complete observation window for all stays, we select only stays of one day or longer. This results in 117 ICU stays involving 88 patients being selected, with 23 shorter stays being excluded. In this new cohort, the mean length of stay increases to 4.28 days, with a median of 2.83 days, a standard deviation of 4.00 days, and an interquartile range of 1.49 to 5.09 days. Table 1 provides details about the selected cohort both prior to and after applying the filter, while Figure 1 displays the distribution of stays along with the mean stay length by unit.

Table 1. Cohort summary before and after applying the 24-hour window criterion

Quantity	Full	Cohort
ICU stays	140	117
Unique patients	100	88
Mean LOS (days)	3.68	4.28
Median LOS (days)	2.16	2.83
Std. dev. LOS (days)	—	4.00
Max LOS (days)	20.53	20.53

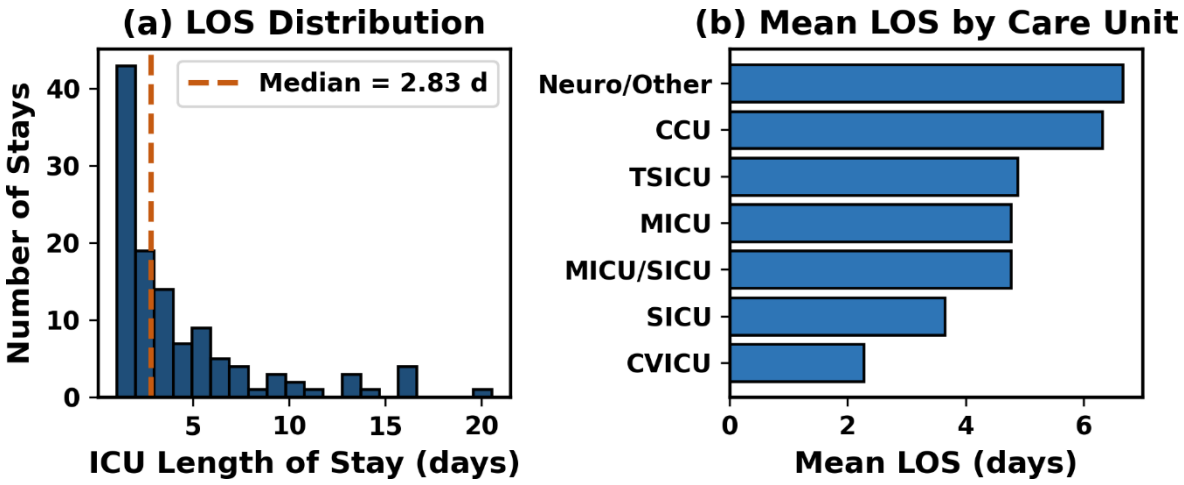

Figure 1. (a) Distribution of ICU length of stay in the modelling cohort, with the median at 2.83 days marked; (b) mean length of stay by care unit

Methodology

Pipeline Overview

The workflow process starting from the raw tables up to the evaluation of the models is shown in Figure 2. For each table, we filter for the cohort, convert all events into hours after the patient was admitted to the ICU, and select only those events that occur in the [0, 24] hour

range as they matter in any way. The range is not chosen randomly but is an intrinsic part of our hypothesis.

Feature Engineering

For each of the ten vital sign and neurological measures, heart rate, systolic and diastolic pressure, respiratory rate, oxygen saturation, temperature, serum glucose, and the three components of the Glasgow Coma Scale, we calculate the mean, minimum, maximum, and standard deviation per stay in the window. All outliers outside of the physiological range are removed beforehand, discarding 14 non-plausible observations out of 17,907 in the window. Three coma scale components are added up, and from the sum we retain only the mean and the minimum.

The amount of therapeutic effort is calculated from input-events, consisting of the number of orders, the volume of administered intravenous fluids in millilitres, the number of orders of a continuous infusion, the number of orders of intravenous antibiotics, and a binary indicator of the use of vasopressors based on the matches of d_items names for commonly used vasopressors. For output events, we calculate total output, and the 24-hour fluid balance is the difference between intake and output. For procedure-events, we record the number of procedures and set a flag for mechanical ventilation. The frequency of charting, i.e. the number of observations of the vital signs, is used as a crude estimate of monitoring intensity. One-hot encoding of the admitting care unit produces a design matrix with 47 features.

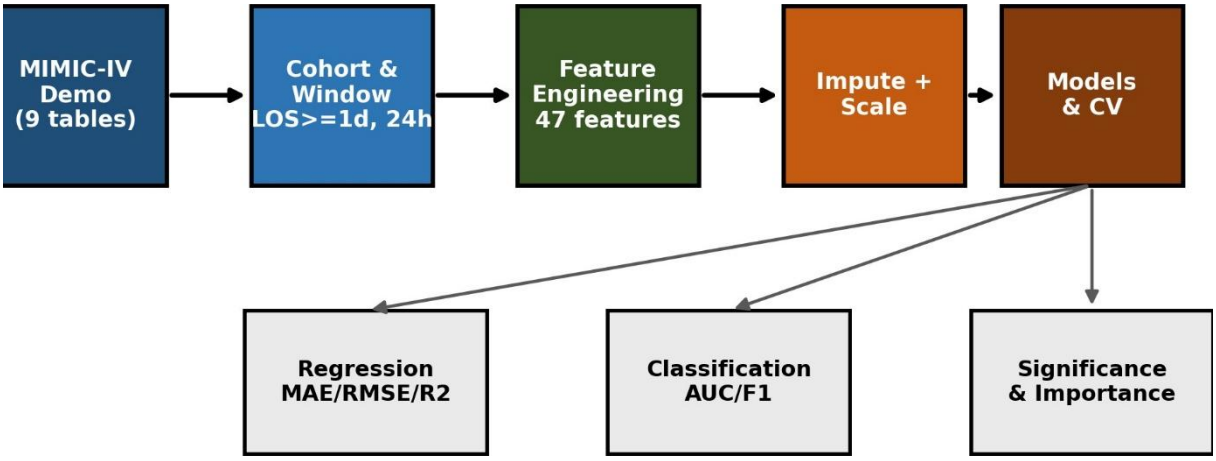

Figure 2. Experimental workflow from raw MIMIC-IV tables to model evaluation

Preprocessing and Imputation

Missing values occur precisely where expected. Standard deviation requires repeated measurements, hence the standard deviation of serum glucose is missing for 39.3% of all stays, and the summary statistics of blood pressure for about 8.5%, simply because some

patients did not have enough in-window observations of a specific feature. Count features are just tallies of orders, volumes, and procedures; hence missing's can only occur if there were none of them, making 0 the proper fill-in and not an imputation. Numeric missing values are imputed with the median of the training fold within each cross-validation fold. Inputs for the linear and logistic regression are standardized; tree ensembles do not require standardization because they are insensitive to it.

Models and Formulation

Let xi ∈ R47 be the feature vector for stay i and yi its length of stay in days. Standardization applies

$$\tilde{x}_{ij} = \frac{(x_{ij} - \mu_j)}{\sigma_j}, \quad (1)$$

with μj and σj estimated on the training fold alone. For regression we compare ordinary least squares against ridge regression, which minimises

$$L(w) = \sum_i (y_i - w^T \tilde{x}_i)^2 + \alpha \|w\|_2^2, \quad (2)$$

and two ensembles, a random forest and gradient-boosted trees. The boosting model builds itself in additive stages,

$$F_m(x) = F_{m-1}(x) + \nu\, h_m(x) \quad (3)$$

where hm is a shallow regression tree and ν the learning rate. Regression quality is judged by

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i|, \quad RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2}, \quad (4)$$

$$R^2 = 1 - \frac{[\sum_i (y_i - \hat{y}_i)^2]}{[\sum_i (y_i - \bar{y})^2]} \quad (5)$$

For the classification framing, length of stay is cut at the cohort median τ = 2.83 days,

$$b_i = \mathbb{1}[y_i > \tau] \quad (6)$$

Algorithm 1: Nested evaluation pipeline

Input: cohort tables; window W = 24h; folds K = 5

Output: cross-validated metrics

1. Filter stays with LOS ≥ 1 day.
2. Anchor events at intime; keep $t \in [0, W]$.
3. Aggregate vitals, inputs, outputs, procedures into **X**.
4. For each of K folds: fit median imputer and scaler on train only; fit model; predict on the held-out fold.
5. Pool out-of-fold predictions; compute metrics and the Wilcoxon test.
6. Refit the random forest on all data to read importances.

Figure 3. Cross-validated evaluation procedure

and logistic regression joins the two ensembles, scored on accuracy, precision, recall, F1, and the area under the ROC curve. The overall procedure is stated in the pseudocode of Figure 3. Its dominant cost is ensemble training on the order of O(K · T · n log n · d) for T trees, n stays, and d features, which is trivial at this scale.

Results

All experiments proceed using five-fold cross-validation stratified by the binary class label of the classification problem and using a shuffle with seed 42 to ensure that all outcomes are reproducible. Imputation and scaling are performed within each training fold and not globally over the entire data set. The hyperparameters chosen (see Table 2) are conservative, with relatively short trees and strong regularization applied to leaves, as otherwise allowing a big model to be trained on 117 patients would amount to teaching it to overfit. The naïve prediction predicts the average length of stay of the training fold for each held-out patient, yielding MAE 2.94 and RMSE 3.82.

Table 2. Model hyperparameter settings

Model	Parameter	Value
Ridge	α	1.0
Random Forest	trees / depth / min-leaf	300 / 6 / 3
Gradient Boosting	trees / depth / lr	200 / 3 / 0.05
Gradient Boosting	subsample	0.8
Logistic Regression	C / max-iter	1.0 / 2000

Regression

The results of the regression analysis are presented in Table 3. The random forest has the smallest error, being the only method to have the MAE of 2.85 days and RMSE of 3.81. However, the performance is worse than mean predictor's on both metrics. Also, the four linear models show the negative cross-validated R2, which does not indicate a mistake but rather a normal outcome for a dataset with small and heavy-tailed target variable: even if the

absolute error can be decreased, the variance explained metric can be poor due to a small number of observations with very large admissions. The difference in absolute errors between random forest and mean predictor was tested using the paired Wilcoxon test and was insignificant ($W = 2924$, $p = 0.15$). This suggests no clear advantage of the regressions. One aspect, however, is confirmed: the correlation between the predicted and actual stays Spearman $\rho = 0.33$ ($p < 0.001$). This means that the ordering of the patients by their expected stays is done better than randomly by the model.

Table 3. Five-fold cross-validated regression performance; lower MAE and RMSE are better, and R^2 nearer one is better

Model	MAE	RMSE	R^2
Baseline (mean)	2.94	3.82	—
Linear Regression	3.66	4.79	−1.18
Ridge Regression	3.40	4.35	−0.72
Gradient Boosting	3.08	4.25	−0.57
Random Forest	2.85	3.81	−0.13

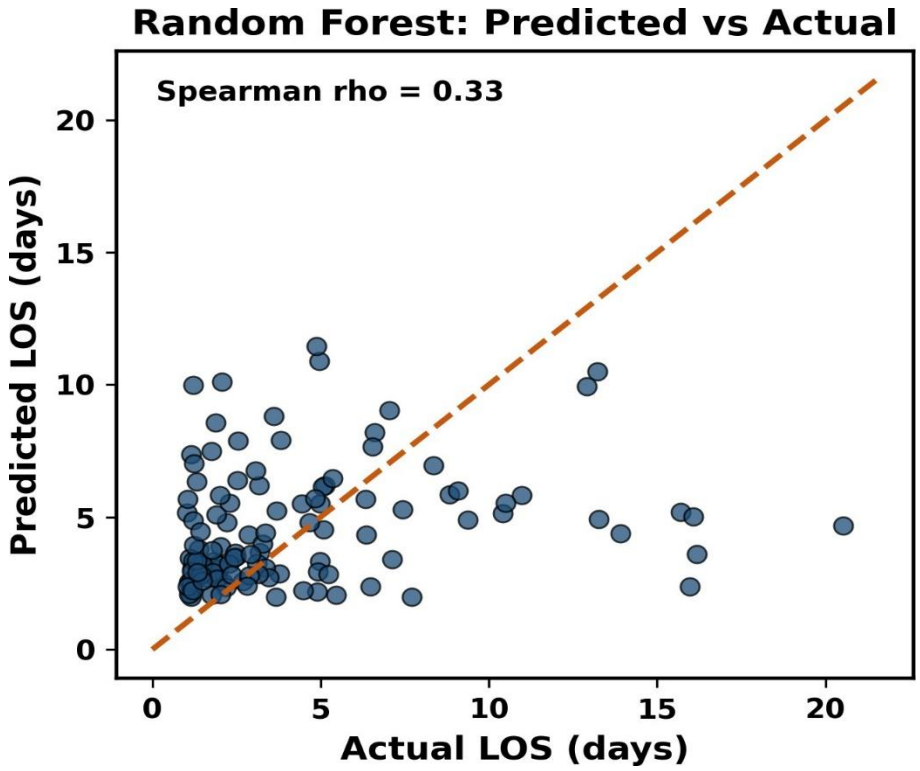

Figure 4. Predicted against observed length of stay for the random forest; the dashed line is perfect prediction, and long stays are underestimated

Classification

By framing the task as a long vs. short problem, we can get something more concrete. There is almost an even split right at the median, 58 for prolonged and 59 for short. As shown in Table 4, the random forest model achieves the best AUC of 0.73, while the gradient boosting model gets the best balance between precision and recall with an F1 score of 0.71, while logistic regression achieves the best precision. Figures 5 and 6 depict the ROC and precision/recall curves, and the confusion matrix for the best classifier, respectively. In the confusion matrix, we can see 42 short and 35 prolonged cases accurately predicted, and with the mistakes evenly distributed between both. To use the prediction in bed planning, an AUC around 0.73 using only day-one data sounds reasonable enough.

Table 4. Classification of prolonged (over 2.83 days) versus short stays, five-fold Cross validated

Model	Acc	Prec	Rec	F1	AUC
Logistic Regression	0.72	0.75	0.66	0.70	0.71
Gradient Boosting	0.71	0.70	0.72	0.71	0.70
Random Forest	0.66	0.67	0.60	0.64	0.73

What Drives the Predictions

The ranking of feature importances from the random forest is displayed in Figure 7, while Figure 8 illustrates the underlying correlation pattern. Both perspectives concur. Minimum and mean Glasgow Coma Score are of primary importance, with 0.17 and 0.41 Spearman correlations with length of stay, respectively, which is quite obvious to the clinician: a patient coming with depressed levels of consciousness is usually more ill and thus requires a longer period of care. Next go early fluid volume and 24-hour fluid balance, with importance close to 0.10, followed by infusion intensity as the number of drip orders with a 0.35 correlation with stay. Mechanical ventilation is used in 43.6% of the cohort stays, whereas vasopressors are utilized in 40.2%, and both correspond to longer stays. Descriptive statistics for the most important features are shown in Table 5, whereas the three top features are compared between the two groups in Figure 9.

Together, these two framing models create a cohesive narrative. The exact days are not available through this dataset since the long-tail and small sample size make every regressor hover near the average predictor. However, these same features have enough ordinal information to differentiate between long and short length-of-stay cases at an AUC of 0.73 and recreate the relationships that a clinician expects. Early consciousness and therapy, not pure vital signs on their own, contribute most to this. This is important to mention because it implies that what a team does within the first day is at least as indicative as the vital signs read by the monitors.

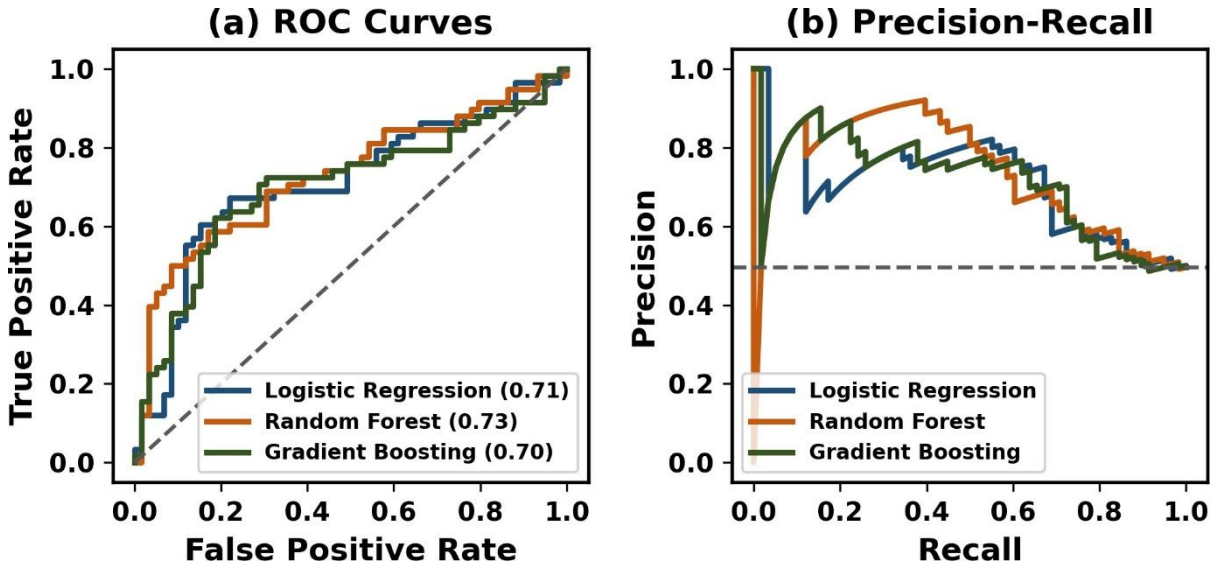

Figure 5. (a) ROC curves and (b) precision-recall curves for the three classifiers

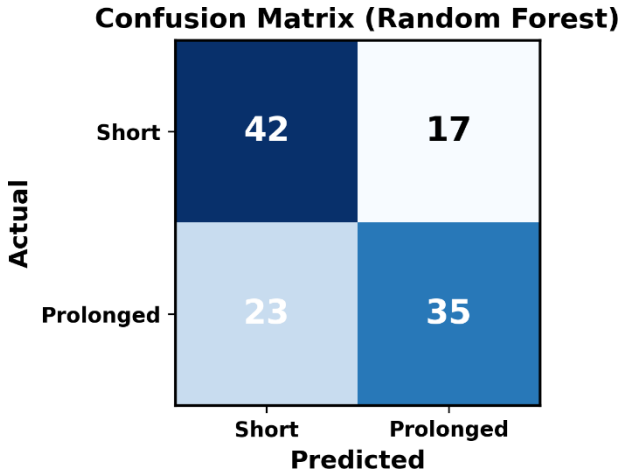

Figure 6. Confusion matrix for the random forest classifier at a 0.5 threshold

Table 5. Descriptive statistics for selected 24-hour features (cohort, n = 117)

Feature	Mean	Std	Max
Heart rate (mean, bpm)	90.6	15.4	131.5
SBP (mean, mmHg)	113.4	16.1	159.0
GCS total (min)	9.7	5.1	15.0
Charting count / 24h	153.1	43.5	309.0
Input orders / 24h	51.4	38.8	192.0
Fluid intake (ml)	5099.7	3443.1	16917.2
Output (ml)	2663.6	1975.4	15800.0
Drip orders / 24h	24.9	29.2	126.0

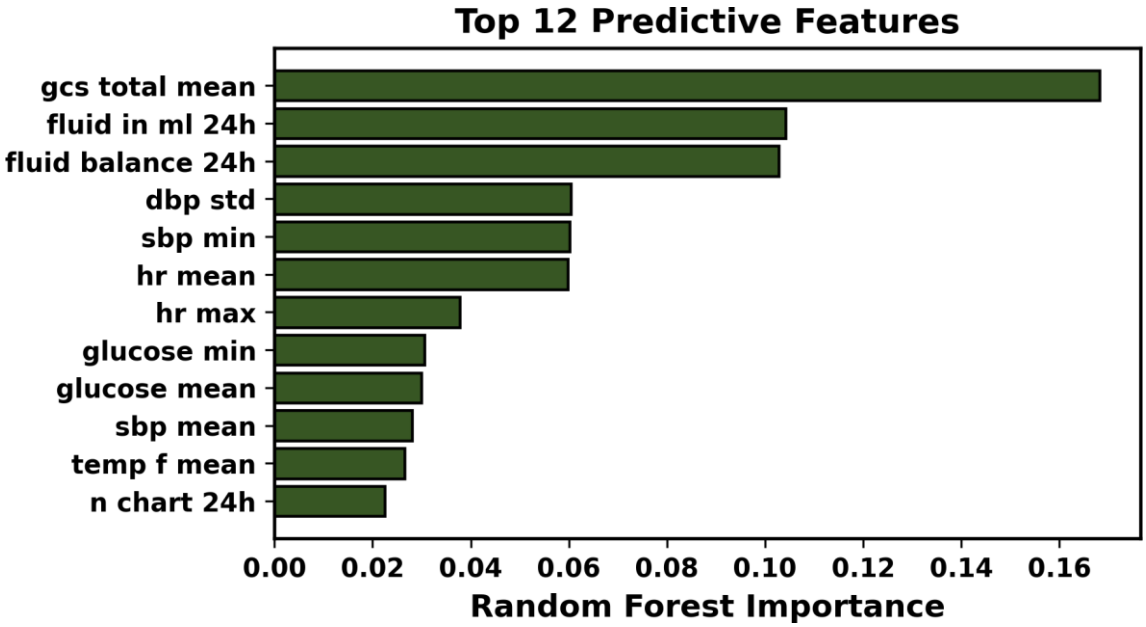

Figure 7. Top twelve features by random-forest importance

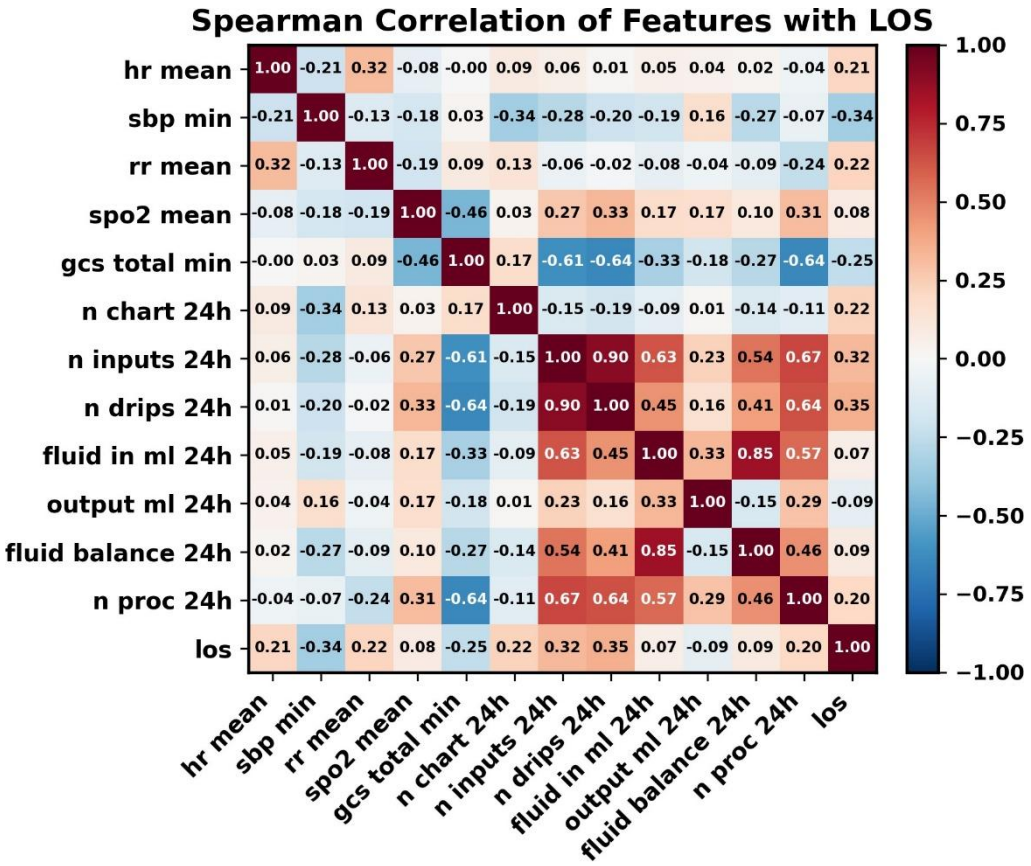

Figure 8. Spearman correlations among key 24-hour features and length of stay

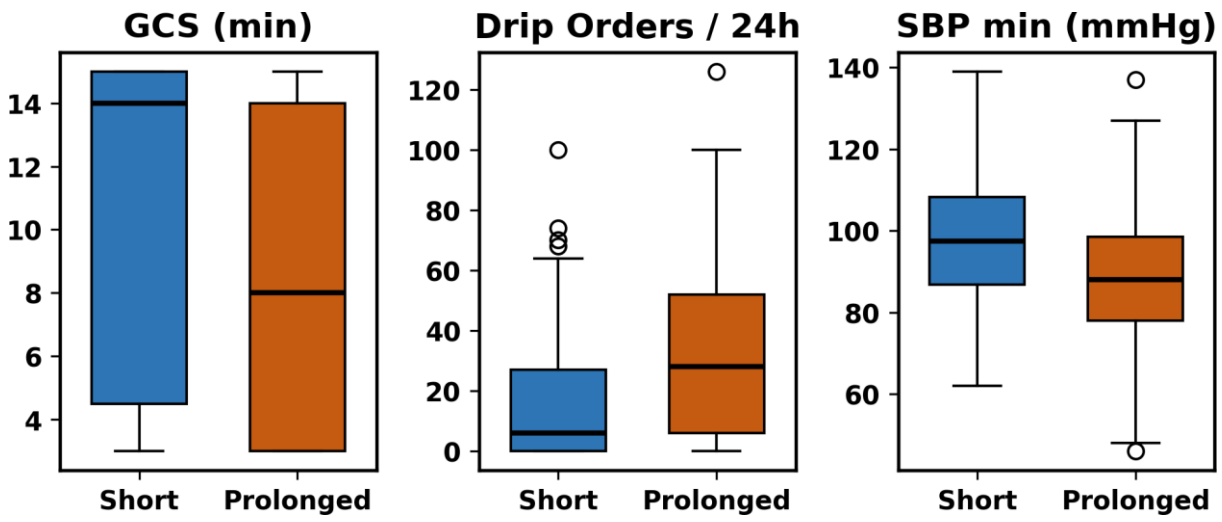

Figure 9. Selected features by length-of-stay group; prolonged stays show lower admission coma scores and heavier early infusion activity

The contrast between these two results themselves is the lesson learned from this exercise. It makes the case that R2 is not the final word on such a length-of-stay model. The fact that it could not predict the exact days may still be useful for its proper ranking, which is often what bed management really requires. This approach is consistent with the overall shift within clinical machine learning towards a different framing and interpretation.

The greatest problem is the sample. Given the 117 cases, the estimates from the cross validation suffer great variability in terms of the huge variation of the errors from fold to fold, and all this data can in no way be interpreted as population estimates. The dataset used in our analysis is also a carefully selected subset of data, not an accurate representation of any unit's case mix. The selected timeframe of 24 hours was designed to serve early detection purposes and thus ignores useful information needed for better prediction. The choice of the features was also based on subjective considerations – the plausibility boundaries, the matching of the vasopressor label, the interpretation of the fluid balance and the like, and all this affected the results.

All numbers reported in this paper have been directly generated from the MIMIC-IV demo tables using a single deterministic script with a predetermined random seed. The process involves loading the nine CSV files, creating the 47-feature matrix, performing cross-validation evaluation without any manual input, and producing the result. All the settings used, including the cohort criteria (at least one-day stay), the time window ([0, 24] hours), the imputation method (median per fold), and hyperparameters (Table 2), are completely specified.

Conclusion

Given that the MIMIC-IV demo dataset is de-identified and distributed under a data use agreement for research, there are no new concerns regarding consent or privacy. Despite the availability of anonymized datasets, secure management of electronic health records remains essential. Recent studies have proposed blockchain-enabled healthcare infrastructures together with lightweight cryptographic mechanisms to preserve confidentiality, integrity, and authorized access to sensitive patient information in distributed healthcare environments (Singh et al., 2026).

We must emphasize that our models are a proof of concept and cannot be applied directly to any patient's care. The application of predictions based on the first day to guide care or allocate resources would require additional validation, a fairness assessment of the model, and human intervention in the loop.

Our task was to determine how much of the ICU course is determined on the very first day, and our response is "some, but not all." Early information does not predict the duration of stay for the patients under study, but it does separate long-stay patients from those with short stays and reveals reasonable predictive factors without being prompted, such as the level of consciousness, volume of fluid, and rate of infusion. The proper description of the phenomenon would be: "First-day data provide good priors for planning." Of course, the natural follow-up step is scaling. Scaling this design across the entire MIMIC-IV population will increase the accuracy of all estimates and provide an opportunity to build more complex time-series models, such as recurrent and attention models, which is not feasible for a dataset of only 117 visits. The addition of greater accuracy and confidence intervals will take the research from feasibility to clinical usefulness.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The author's received no financial support for the research, authorship, and/or publication of this article.

References

Agustsson, E., & Timofte, R. (2017). NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 126–135).

Alvarez, G., & Li, S. (2006). Some basic cryptographic requirements for chaos-based cryptosystems. *International Journal of Bifurcation and Chaos, 16*(8), 2129–2151.

Amrit, P., et al. (2023). EmbedR-Net: Using CNN for secure eHealth systems. *IEEE Transactions on Consumer Electronics, 69*(4), 1017–1022.

Amutha, R. (2026). Chaotic image encryption using circular shift and modular arithmetic. *Journal of Discrete Mathematical Sciences and Cryptography, 29*(1), 215–228. https://doi.org/10.47974/JDMSC-2411

Aumasson, J.-P., Neves, S., Wilcox-O'Hearn, Z., & Winnerlein, C. (2013). BLAKE2: Simpler, smaller, fast as MD5. In *Proceedings of the 11th International Conference on Applied Cryptography and Network Security (ACNS)* (pp. 119–135).

Barker, E. (2020). *Recommendation for key management—Part 1: General* (NIST Special Publication 800-57 Part 1 Rev. 5). National Institute of Standards and Technology.

Bourekouche, H., Belkacem, S., & Messaoudi, N. (2025). Design and analysis of a PRNG based on the logistic-sine system. *Journal of Discrete Mathematical Sciences and Cryptography, 28*(7), 2645–2670.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(1), 21–30.

Daugman, J. (2006). Probing the uniqueness and randomness of IrisCodes. *Proceedings of the IEEE, 94*(11), 1927–1935.

Fridrich, J. (1998). Symmetric ciphers based on two-dimensional chaotic maps. *International Journal of Bifurcation and Chaos, 8*(6), 1259–1284.

Huang, X., Ye, Y., Dai, S., & Zhang, Y. (2023). Image encryption scheme using chaotic map and pixel-level diffusion with S-box substitution. *Mathematics, 11*(4), Article 866. https://doi.org/10.3390/math11040866

ISO/IEC. (2022). *Information security, cybersecurity and privacy protection—Information security management systems—Requirements* (ISO/IEC 27001:2022). International Organization for Standardization.

Jain, A. K., Nandakumar, K., & Nagar, A. (2008). Biometric template security. *EURASIP Journal on Advances in Signal Processing, 2008*, Article 579416.

Jyotsna, L., & Kumar, L. P. (2026). Enhanced image security through block permutation. *Journal of Discrete Mathematical Sciences and Cryptography, 29*(3), 1313–1333.

Krawczyk, H., & Eronen, P. (2010). *HMAC-based extract-and-expand key derivation function (HKDF)* (RFC 5869). Internet Engineering Task Force.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Rohhila, S., & Singh, A. K. (2024). Deep learning-based encryption for secure digital images: A survey. *Computers & Electrical Engineering, 116*, Article 109236.

Rohhila, S., Singh, K. N., Singh, A. K., & Gupta, B. B. (2025). DeepKey-TPE: Using deep learning to generate key for image security with continued usability through thumbnail-preserving encryption. *IEEE Transactions on Consumer Electronics*. Advance online publication.

https://doi.org/10.1109/TCE.2025.3605213

Singh, A. K., & Agrawal, A. K. (2026). HybridCrypt-AE: A deep learning–augmented hyperchaotic image encryption scheme with dual-feedback diffusion and SHA-512 key generation. *Journal of Discrete Mathematical Sciences and Cryptography*.

Stallings, W. (2023). *Cryptography and network security* (8th ed.). Pearson.

Wang, X., & Liu, C. (2017). A novel and effective image encryption algorithm based on chaos and DNA encoding. *Multimedia Tools and Applications, 76*(5), 6229–6245. https://doi.org/10.1007/s11042-016-3311-8

Bibliographic information of this paper for citing:

Ather, Danish; Raj, Dharm; Quraishi, Suhail Javed; Pathak, Sonal; Rakhra, Manik & Krishna, Kunchanapaalli Rama (2026). A Smart Healthcare Management System for Predicting ICU Length of Stay Using Machine Learning and MIMIC-IV. *Journal of Information Technology Management*, 18 (4), 134-149. https://doi.org/10.22059/jitm.2026.108919