



# Outcome Prediction in Burn Care: A Comparative Study of Machine Learning Algorithms on Structured Clinical Data

Abdolkhalegh Keshavarzi <sup>1,\*</sup>, Ali Sam <sup>2</sup>, Nahid Abolpour <sup>2,3,4</sup>, Zahra Taghados <sup>2</sup>, Iliya Keshavarzi <sup>5</sup>, Fardin Ghanbari <sup>1</sup>, Marzieh Mansouri <sup>6</sup>

<sup>1</sup> Department of Surgery, Faculty of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

<sup>2</sup> Artificial Intelligence Clinical Laboratory and Biological Data Bank, Shiraz University of Medical Sciences, Shiraz, Iran

<sup>3</sup> Department of Artificial Intelligence in Medical Sciences, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

<sup>4</sup> Maternal-Fetal Medicine Research Center, Shiraz, Iran

<sup>5</sup> Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran

<sup>6</sup> Burn and Wound Healing Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

\*Corresponding Author: Department of Surgery, Faculty of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran. Email: [iliakeshavarzi@yahoo.com](mailto:iliakeshavarzi@yahoo.com)

Received: 25 June, 2026; Revised: 9 July, 2026; Accepted: 20 July, 2026

## Abstract

**Background:** Early prediction of burn outcomes using structured clinical information available during prehospital care, triage, and early admission may support triage, counseling, and resource planning.

**Objectives:** To compare five prespecified classifiers for the early prediction of final outcome, number of surgeries, intensive care unit (ICU) length of stay, and hospital length of stay using routinely collected prehospital and admission registry data.

**Methods:** This retrospective single-center study used burn-registry records from Shiraz Burn Hospital for the period from 20 March 2020 to 20 March 2026. After records with final outcomes of transfer, discharge at the patient's own request, or unknown/not recorded were excluded, the analytic registry contained 4,127 records with 4,127 unique patient identifiers. Final outcome was modeled as recovery versus death. For the targets of number of surgeries and length of stay, deaths were excluded because death is a competing terminal event. Candidate predictors comprised demographic characteristics; referral and triage information; level of consciousness; vital signs and acute clinical findings; burn mechanism, intent, depth, total body surface area (TBSA), and anatomical regions; comorbidities; and prehospital treatment indicators. Target labels were not imputed. Logistic regression, decision tree, random forest, support vector machine (SVM), and k-nearest neighbors were compared using repeated stratified five-fold cross-validation, with macro-F1 as the primary model-selection metric, followed by evaluation on an independent 20% held-out test set.

**Results:** The final outcome was recovery in 3,175 patients (76.9%) and death in 952 (23.1%). Among the 3,175 recovered patients included in the resource-use analyses, surgery categories were 0 surgeries in 48.1%, 1 in 14.0%, 2 in 10.3%, and > 2 in 27.6%; ICU-stay categories were no ICU/0 days in 74.9%, 1 - 7 days in 12.6%, and > 7 days in 12.5%; and hospital-stay categories were ≤ 7 days in 26.4%, 8 - 14 days in 33.2%, and > 14 days in 40.4%. In the development-set comparison, SVM achieved the highest macro-F1 for final outcome (0.895), whereas random forest performed best in predicting the number of surgeries, ICU length of stay, and hospital length of stay, with macro-F1 values of 0.466, 0.618, and 0.575, respectively. On the held-out test set, the selected SVM achieved a macro-F1 of 0.905 and a balanced accuracy of 0.932 for final outcome. For death prediction, sensitivity was 93.7% and positive predictive value was 79.2%. Random forest achieved held-out macro-F1 values of 0.459, 0.699, and 0.575 for number of surgeries, ICU length of stay, and hospital length of stay, respectively.

**Conclusions:** Among the five evaluated classifiers, SVM provided the strongest performance for early prediction of final clinical outcome, whereas random forest performed best for the three resource-use outcomes. Early structured burn-registry variables enabled strong discrimination between recovery and death, including high sensitivity for mortality, whereas prediction of operative burden and hospital length of stay remained more challenging. These findings support further external and prospective validation before clinical implementation.

**Keywords:** Burn Injuries, Machine Learning, Patient Outcome Prediction, Mortality, Length Of Hospital Stay

## 1. Background

Burn injuries remain a substantial global public health and clinical care burden. The World Health Organization estimates that burns cause approximately 180,000 deaths each year, with the greatest burden in low- and middle-

income countries (1). Severe burn injury is associated with substantial mortality, prolonged hospitalization, intensive care requirements, and operative burden, placing heavy and sustained demands on hospital resources. Early estimation of clinical outcomes and resource needs is therefore important for patient management and service

planning (2). For burn clinicians, early estimation of death risk, the need for intensive care, expected hospitalization, and likely operative burden can support communication with families, triage discussions, operating-room planning, and registry-based quality improvement.

Traditionally, these challenges have been addressed through clinical judgment at triage and admission, supported by manual severity scores that use a few readily available patient and injury characteristics to estimate mortality risk. A range of burn mortality risk models has been developed and evaluated, although predictive performance and transportability may vary across populations and clinical settings (3). Evidence on the revised Baux score supports the relevance of age, total body surface area burned, and inhalation injury to mortality prediction (4), while the simpler Baux score remains a familiar clinical reference point when age and total body surface area burned are available (5).

However, these traditional tools have important limitations. Because they were designed to summarize mortality risk using a small number of variables, they focus mainly on a single endpoint and provide little information about other outcomes that determine hospital workload. Length of stay and operative burden depend on the evolving clinical course and local care pathways, making them more context dependent than mortality and difficult to capture with fixed, manually calculated scores. Manual scores also cannot readily exploit the many structured variables routinely recorded during prehospital care, triage, and early admission. These shortcomings have encouraged the use of artificial intelligence in burn care, where research has focused particularly on burn-depth assessment, wound segmentation, and mortality prediction, although substantial heterogeneity in study design and reporting remains (6). More broadly, large-scale clinical prediction studies have demonstrated the potential of routinely collected health data to support automated risk estimation across diverse clinical settings (7).

More recent single-center and registry-based studies have applied machine-learning methods to early mortality-risk estimation (8), evaluation of established burn-related risk features (9), postoperative mortality prediction (10), and prognostic modeling in high-risk burn populations (11). Collectively, these studies suggest that routinely available demographic, injury-severity, and clinical information can provide useful prognostic information, particularly for mortality. Machine-learning applications have also begun to address selected complications, including intensive care unit (ICU) delirium (12) and early sepsis (13). Prediction of resource-use outcomes is less mature. Previous studies have investigated prolonged

hospitalization, graft or operative requirements, and other adverse outcomes, showing that burn extent and depth, comorbidities, physiological status, and patient-level characteristics contribute to these endpoints (11, 14, 15). Nevertheless, important problems remain unresolved. Existing work has largely emphasized mortality or selected complications, whereas clinical outcome, operative burden, and length-of-stay measures are often investigated separately using different preprocessing and validation strategies. Relatively few investigations have compared multiple clinical and resource-use targets from the same early data source using a common preprocessing, model-comparison, and held-out evaluation framework. These shortcomings are compounded by broader methodological weaknesses in machine-learning prediction research, including inconsistent handling and reporting of missing data (16) and recurrent problems in study design, validation, and reporting that may lead to overly optimistic estimates of model performance (17).

Closing these gaps is important because early clinical prognosis and anticipated resource use are both needed: communication with families and triage depend on the likely outcome, whereas operating-room planning and quality improvement depend on expected operative burden and length of stay. Such comparisons are meaningful only when outcomes are modeled from the same data using the same validation strategy and when the prediction targets are clinically coherent. Against this background, the present study evaluated whether routinely collected prehospital, triage, and early-admission registry variables could support prediction of four distinct outcomes within a single analytical framework: final clinical outcome, number of surgeries, ICU length of stay, and hospital length of stay. The same five prespecified classifiers were compared for each target using a consistent development and held-out evaluation strategy. To improve clinical interpretability and reduce avoidable bias, administrative dispositions were separated from clinical outcome categories; deaths were excluded from the surgery and length-of-stay analyses because death terminates the opportunity to accumulate further procedures or inpatient days; and outcome-to-outcome information leakage was explicitly prevented. Final-outcome prediction was additionally examined against a Baux-score benchmark (5) and through probability calibration, and permutation-based feature importance was used to support interpretation of the selected models. These design choices reflect the guidance of reporting frameworks such as RECORD and TRIPOD+AI for observational health-data studies and clinical prediction models (18, 19), as well as the methodological emphasis on adequate sample size, calibration, and appropriate validation, because discrimination alone is insufficient for

judging model quality, particularly when events are limited or outcome classes are imbalanced (20, 21). The expanding application of machine learning to structured health data in regional research, including accident-related questionnaire data, telemedicine and care-quality applications, and gait-based health classification (22-24), further illustrates the feasibility of structured-data prediction and the need for careful methodological design. By integrating clinical and resource-use outcomes within one consistently defined framework, the study extends prior mortality-focused work toward a broader evaluation of early burn prognosis and resource planning.

## 2. Objectives

To address this gap, this study aimed to compare five prespecified classifiers for the early prediction of four burn outcomes: final outcome, number of surgeries, ICU length-of-stay category, and hospital length-of-stay category. The intended users of the models are burn clinicians and triage or registry-supported quality-improvement teams. The target setting is early assessment after emergency presentation or referral, where structured registry variables may assist with preliminary risk stratification and hospital resource planning. The models are not presented as deployable bedside decision tools.

## 3. Methods

### 3.1. Study Design, Setting, and Ethics

This was a retrospective, single-center prediction-model study using routinely collected burn-registry data from Shiraz Burn Hospital. The registry included patients recorded from 20 March 2020 through 20 March 2026. The study was approved by the Institutional Review Board of Shiraz University of Medical Sciences (approval code: IR.SUMS.REC.1405.091). Informed consent was waived because of the retrospective, anonymized nature of the study.

### 3.2. Eligibility and Analytic Datasets

The source population comprised all patients recorded in the burn registry during the study period. The original registry extract contained 4,807 unique records. Before model development, 680 records were excluded because the final outcome was transfer to another hospital ( $n = 62$ ), discharge at the patient's own request ( $n = 552$ ), or an unknown/not-recorded outcome ( $n = 66$ ). The final analytic registry therefore contained 4,127 records and 4,127 unique patient identifiers, with no repeated visits by the same patient identifier. For the number-of-surgeries, ICU

length-of-stay, and hospital length-of-stay models, patients whose final outcome was death were excluded because death competes with the opportunity to undergo additional operations or accumulate a full inpatient stay. No target labels were imputed. The final-outcome analysis included 4,127 patients, and the three resource-use analyses included 3,175 patients in the recovery category after the exclusion of 952 deaths.

### 3.3. Outcome Definitions

Clinical final outcome was modeled as recovery versus death. Full recovery and partial recovery ( $n = 39$ ) were combined into the recovery category for binary clinical final-outcome classification. Transfer to another hospital, discharge at the patient's own request, and unknown/not-recorded outcomes were excluded before modeling because transfer and self-discharge are administrative dispositions rather than homogeneous clinical states, and unknown outcomes cannot be interpreted as valid clinical labels. The number of surgeries was categorized as 0, 1, 2, or more than 2. ICU length of stay was categorized as no ICU stay/0 days, 1 - 7 days, or more than 7 days. Hospital length of stay was categorized as 7 days or less, 8 - 14 days, or more than 14 days. These thresholds were fixed before model training in consultation with clinicians to distinguish no resource use, short/intermediate use, and prolonged or higher-burden care. The same categorization also reflects common burn-service planning concepts, in which prolonged hospitalization beyond 14 days has been used as a clinically relevant target in burn prediction studies (8, 11, 12).

### 3.4. Predictors and Timing

Candidate predictors included age, sex, marital status, education, nationality, rural residence, referral pathway, triage level, level of consciousness, vital signs, pregnancy status, airway compromise, respiratory distress, signs of shock, lethargy or drowsiness, severe pain or distress, burn intent, place of injury, burn mechanism, burn depth, total body surface area burned, grouped anatomical burn regions, prehospital topical treatments, substance use, and individual comorbidities. Clinically important features were retained based on input from clinical experts. Vital signs and triage variables were recorded at triage, and additional burn characteristics and clinical variables were obtained early after admission. Practical implementation would require local evaluation of documentation flow to confirm that all predictors are available before the intended time of model use.

### 3.5. Preprocessing and Leakage Prevention

Explicit unknown or not-recorded markers were converted to missing values and were not treated as a separate category. Predictor missingness was handled within the training data using a mixed random forest imputation procedure for continuous and categorical variables. Continuous and ordinal variables were standardized after imputation. Nominal categorical predictors were one-hot encoded, with unseen categories ignored during validation or testing. Triage level and level of consciousness were treated as ordered clinical variables and ordinally encoded. Detailed anatomical site indicators were collapsed into grouped body-region variables, and initial topical treatment indicators were collapsed into clinically interpretable treatment-type groups. For each target, the other outcome variables were excluded from the predictor set to prevent outcome-to-outcome leakage. Patient identifiers and variables unavailable for practical model use were excluded. All preprocessing steps were fitted within training folds or within the development set before held-out test evaluation.

### 3.6. Model Development and Evaluation

The same five candidate classifiers were evaluated for every target: logistic regression, decision tree, random forest, support vector machine (SVM), and k-nearest neighbors (KNN). Logistic regression is reported as a conventional supervised baseline classifier; this does not imply that regression and flexible machine-learning algorithms are identical. Prespecified hyperparameters were used consistently across targets, and no data-driven hyperparameter tuning was performed. Logistic regression used the lbfgs solver, a maximum of 3,000 iterations, and balanced class weights. The decision tree used balanced class weights and a minimum leaf size of 10. Random forest used 150 trees, balanced-subsample class weighting, a minimum leaf size of 5, unrestricted maximum depth, and all available processor cores. SVM used a radial-basis-function kernel, balanced class weights, the scikit-learn defaults  $C = 1.0$  and  $\gamma = 'scale'$ , and probability estimation disabled for classifier comparison. KNN used  $k = 15$  and distance weighting. The dataset was split into a development set (80%) and an independent held-out test set (20%) using stratified sampling. Algorithm comparison and model selection were performed only within the development set using repeated stratified cross-validation with 5 folds and 2 repeats. The selected model for each target was refit on the full development set and evaluated once on the held-out test set.

### 3.7. Performance Measures and Calibration

Macro-F1 was the primary selection metric because it gives equal weight to each outcome class in the presence of

class imbalance. Accuracy, balanced accuracy, macro precision, macro recall, weighted F1-score, and class-specific precision, recall, and F1-score were also reported. Balanced accuracy was calculated as the average of class-specific recall and was used to supplement macro-F1. Bootstrap 95% confidence intervals were calculated for held-out macro-F1 and balanced accuracy using 2,000 resamples of the held-out predictions. Confusion matrices were generated for all four selected models. For final-outcome probability estimation, SVM probabilities were calibrated using sigmoid Platt scaling with three-fold cross-validation within the development set; Brier score, log loss, bootstrap 95% confidence intervals, and calibration plots were reported (21). A Baux-score benchmark, defined as age plus total body surface area burned, was evaluated for final-outcome prediction because the required variables were available (5). To convert the Baux score to a binary recovery/death prediction, a logistic model using the Baux score was fitted in the development set. The probability threshold maximizing development-set macro-F1 was selected by grid search from 0.05 to 0.95, and the resulting threshold of 0.695 was fixed before held-out testing. The held-out test set was not used to select the Baux threshold. Class-specific precision, recall (class-wise accuracy), F1-score, specificity, and one-vs-rest accuracy were calculated for all five classifiers in every validation fold of the development-set repeated cross-validation and summarized as mean (SD) across the 10 splits. Recall for a class is the proportion of patients in that class who were classified correctly. The held-out test set was used only for the selected model for each target.

### 3.8. Model Interpretation and Software

Permutation feature importance was calculated for the selected model for each target as the decrease in macro-F1 after predictor permutation. Rank stability was assessed across cross-validation folds. Main-text feature-importance figures were restricted to clinically interpretable predictors. Importance estimates were interpreted as predictive associations, not causal effects. Analyses were performed in Python using pandas, NumPy, scikit-learn, and matplotlib. Random seeds were fixed for data splitting, cross-validation, imputation, bootstrapping, and stochastic model components.

## 4. Results

The original registry extract included 4,807 unique records. Preanalysis exclusions removed transfers to another hospital ( $n = 62$ ), discharges at the patients' own request ( $n = 552$ ), and unknown or unrecorded final outcomes ( $n = 66$ ), leaving 4,127 unique patients for clinical final-outcome prediction. The clinical final outcome was

**Table 2.** Target-Specific Inclusion and Exclusion Flow

| Prediction Target              | Original Registry Records | Pre-Analysis Final-Outcome Exclusions                                                           | Deaths Excluded | Analytic Records | Unique Patient IDs | Outcome Distribution                                                          |
|--------------------------------|---------------------------|-------------------------------------------------------------------------------------------------|-----------------|------------------|--------------------|-------------------------------------------------------------------------------|
| <b>Final outcome</b>           | 4,807                     | 680 (transfer to another hospital: 62; discharge at own request: 552; unknown/not recorded: 66) | 0               | 4,127            | 4,127              | Recovered: 3,175; Death: 952                                                  |
| <b>Number of surgeries</b>     | 4,807                     | 680 (transfer to another hospital: 62; discharge at own request: 552; unknown/not recorded: 66) | 952             | 3,175            | 3,175              | 0: 1,526; > 2: 876; 1: 446; 2: 327                                            |
| <b>ICU length of stay</b>      | 4,807                     | 680 (transfer to another hospital: 62; discharge at own request: 552; unknown/not recorded: 66) | 952             | 3,175            | 3,175              | No ICU stay / 0 days: 2,377; ICU stay 1 - 7 days: 400; ICU stay > 7 days: 398 |
| <b>Hospital length of stay</b> | 4,807                     | 680 (transfer to another hospital: 62; discharge at own request: 552; unknown/not recorded: 66) | 952             | 3,175            | 3,175              | > 14 days: 1,282; 8 - 14 days: 1,055; ≤ 7 days: 838                           |

**Table 3.** Development-Set Repeated Cross-Validation Performance for the Five Prespecified Classifiers

| Target and model               | Macro precision | Macro recall | Macro-F1 | Weighted F1 | Accuracy |
|--------------------------------|-----------------|--------------|----------|-------------|----------|
| <b>Final outcome</b>           |                 |              |          |             |          |
| SVM <sup>a</sup>               | 0.878           | 0.918        | 0.895    | 0.924       | 0.922    |
| Random Forest                  | 0.879           | 0.904        | 0.890    | 0.921       | 0.919    |
| Logistic Regression            | 0.873           | 0.910        | 0.889    | 0.919       | 0.917    |
| Decision Tree                  | 0.841           | 0.892        | 0.861    | 0.897       | 0.893    |
| KNN                            | 0.911           | 0.817        | 0.851    | 0.900       | 0.906    |
| <b>Hospital length of stay</b> |                 |              |          |             |          |
| Random Forest <sup>a</sup>     | 0.573           | 0.582        | 0.575    | 0.583       | 0.587    |
| SVM                            | 0.565           | 0.570        | 0.562    | 0.568       | 0.567    |
| KNN                            | 0.566           | 0.558        | 0.561    | 0.570       | 0.570    |
| Logistic Regression            | 0.540           | 0.551        | 0.540    | 0.546       | 0.549    |
| Decision Tree                  | 0.538           | 0.539        | 0.535    | 0.542       | 0.539    |
| <b>ICU length of stay</b>      |                 |              |          |             |          |
| Random Forest <sup>a</sup>     | 0.599           | 0.646        | 0.618    | 0.773       | 0.762    |
| SVM                            | 0.566           | 0.649        | 0.582    | 0.725       | 0.692    |
| KNN                            | 0.748           | 0.533        | 0.576    | 0.775       | 0.812    |
| Logistic Regression            | 0.552           | 0.635        | 0.570    | 0.716       | 0.685    |
| Decision Tree                  | 0.512           | 0.597        | 0.521    | 0.669       | 0.629    |
| <b>Number of surgeries</b>     |                 |              |          |             |          |
| Random Forest <sup>a</sup>     | 0.493           | 0.467        | 0.466    | 0.576       | 0.590    |
| SVM                            | 0.438           | 0.445        | 0.436    | 0.532       | 0.515    |
| KNN                            | 0.513           | 0.409        | 0.414    | 0.538       | 0.580    |
| Logistic Regression            | 0.406           | 0.413        | 0.399    | 0.483       | 0.461    |
| Decision Tree                  | 0.371           | 0.367        | 0.348    | 0.419       | 0.389    |

<sup>a</sup> Identifies the model with the highest mean macro-F1 within each target in the development-set cross-validation comparison.

recovery in 3,175 patients (76.9%; including 39 records of partial recovery) and death in 952 patients (23.1%). After deaths were excluded from resource-use prediction, 3,175 patients in the recovery category remained. In this subset, surgery counts were 0 for 1,526 patients, 1 for 446, 2 for 327, and more than 2 for 876. ICU length of stay was no ICU stay (0 days) for 2,377 patients, 1 - 7 days for 400, and more than 7 days for 398. Hospital length of stay was 7 days or less for 838 patients, 8 - 14 days for 1,055, and more than 14 days for

1,282. Consolidated descriptive characteristics and outcome distributions are provided in [Table 1](#), and target-specific inclusion and exclusion flow is summarized in [Table 2](#).

Across repeated stratified cross-validation in the development set, SVM achieved the highest mean macro-F1 for final outcome, whereas random forest achieved the highest mean macro-F1 for number of surgeries, ICU length of stay, and hospital length of stay ([Table 3](#)). Accuracy was

**Table 4.** Held-Out Overall and Class-Specific Performance of Selected Models <sup>a</sup>

| Target and Class or Summary       | Selected Model | Precision | Recall | F1-Score              | Accuracy       | Balanced Accuracy     | Support |
|-----------------------------------|----------------|-----------|--------|-----------------------|----------------|-----------------------|---------|
| <b>Final outcome</b>              |                |           |        |                       |                |                       |         |
| Overall (macro / held-out)        | SVM            | 0.886     | 0.932  | 0.905 (0.881 - 0.927) | 0.929          | 0.932 (0.910 - 0.951) | 826     |
| Death                             | SVM            | 0.792     | 0.937  | 0.859                 | Not applicable | Not applicable        | 191     |
| Recovered                         | SVM            | 0.980     | 0.926  | 0.952                 | Not applicable | Not applicable        | 635     |
| <b>Number of surgeries</b>        |                |           |        |                       |                |                       |         |
| Overall (macro / held-out)        | Random Forest  | 0.479     | 0.457  | 0.459 (0.414 - 0.502) | 0.573          | 0.457 (0.416 - 0.497) | 635     |
| 0                                 | Random Forest  | 0.677     | 0.702  | 0.689                 | Not applicable | Not applicable        | 305     |
| 1                                 | Random Forest  | 0.361     | 0.337  | 0.349                 | Not applicable | Not applicable        | 89      |
| 2                                 | Random Forest  | 0.344     | 0.167  | 0.224                 | Not applicable | Not applicable        | 66      |
| > 2                               | Random Forest  | 0.534     | 0.623  | 0.575                 | Not applicable | Not applicable        | 175     |
| <b>ICU length of stay, d</b>      |                |           |        |                       |                |                       |         |
| Overall (macro / held-out)        | Random Forest  | 0.687     | 0.714  | 0.699 (0.649 - 0.746) | 0.817          | 0.714 (0.663 - 0.766) | 635     |
| ICU stay 1 - 7                    | Random Forest  | 0.494     | 0.550  | 0.521                 | Not applicable | Not applicable        | 80      |
| ICU stay > 7                      | Random Forest  | 0.655     | 0.713  | 0.683                 | Not applicable | Not applicable        | 80      |
| No ICU stay / 0                   | Random Forest  | 0.911     | 0.880  | 0.895                 | Not applicable | Not applicable        | 475     |
| <b>Hospital length of stay, d</b> |                |           |        |                       |                |                       |         |
| Overall (macro / held-out)        | Random Forest  | 0.574     | 0.578  | 0.575 (0.537 - 0.613) | 0.587          | 0.578 (0.540 - 0.616) | 635     |
| 8 - 14                            | Random Forest  | 0.490     | 0.445  | 0.467                 | Not applicable | Not applicable        | 211     |
| ≤ 7                               | Random Forest  | 0.542     | 0.577  | 0.559                 | Not applicable | Not applicable        | 168     |
| >14                               | Random Forest  | 0.689     | 0.711  | 0.700                 | Not applicable | Not applicable        | 256     |

<sup>a</sup> For overall held-out rows, values in parentheses for F1-score and balanced accuracy are bootstrap 95% confidence intervals from 2,000 resamples of held-out predictions.

consistently higher than macro-F1 for imbalanced tasks, confirming that accuracy alone was not an adequate summary metric.

On the held-out test set, the selected SVM for final outcome achieved an accuracy of 0.929, a balanced accuracy of 0.932, and a macro-F1 of 0.905. Random forest was selected for number of surgeries, ICU length of stay, and hospital length of stay, with held-out macro-F1 values of 0.459, 0.699, and 0.575, respectively (Table 4). Class-specific results in Table 4 showed strong recognition of the recovery and death classes for final outcome, moderate recognition of ICU classes, and weaker performance for the intermediate surgery and hospital-stay classes. Confusion matrices for all four targets are shown in Figure 1.

Class-specific cross-validation results for all five classifiers are given in Tables 5 and 6. For all four targets, the larger classes were recognized more readily than the smaller or intermediate classes, irrespective of the algorithm. For final outcome, class-wise accuracy (recall) for recovery ranged from 0.894 to 0.932 and for death from 0.875 to 0.912 across logistic regression, decision tree, random forest, and SVM, whereas KNN had the highest recovery recall of 0.983 but a death recall of only 0.651. SVM achieved the highest F1-score for both recovery and death, at 0.948 and 0.843, respectively, although its margins over

the next-best classifier (0.001 and 0.009) were smaller than the cross-validation SD. For the number of surgeries, no classifier reliably recognized patients with two operations (recall 0.082 - 0.295; F1-score 0.138 - 0.207), whereas random forest performed best for more than two operations (recall 0.670; F1-score 0.591). For ICU length of stay, random forest yielded the highest F1-scores for the 1 - 7-day and > 7-day classes, at 0.406 and 0.580, respectively, and correctly classified 83.3% of patients without an ICU stay; KNN had a higher recall of 0.982 for no ICU stay, but its specificity for that class was only 0.376, and its recall for the 1 - 7-day class was 0.188. For hospital length of stay, differences among classifiers were smaller; random forest performed best for > 14 days (recall 0.697; F1-score 0.699), and the 8 - 14-day class was the hardest for all classifiers (F1-score 0.406 - 0.483).

For calibrated final-outcome prediction, the calibrated SVM achieved an accuracy of 0.929, a macro-F1 of 0.899, a Brier score for death of 0.055 (95% CI, 0.044 - 0.066), and a log loss of 0.195 (95% CI, 0.156 - 0.234). The Baux-score benchmark used a development-set-selected probability cutoff of 0.695 and achieved a held-out macro-F1 of 0.838 and an area under the receiver operating characteristic curve (ROC AUC) of 0.920 for death. The calibration plot for the final-outcome model is shown in Figure 2.

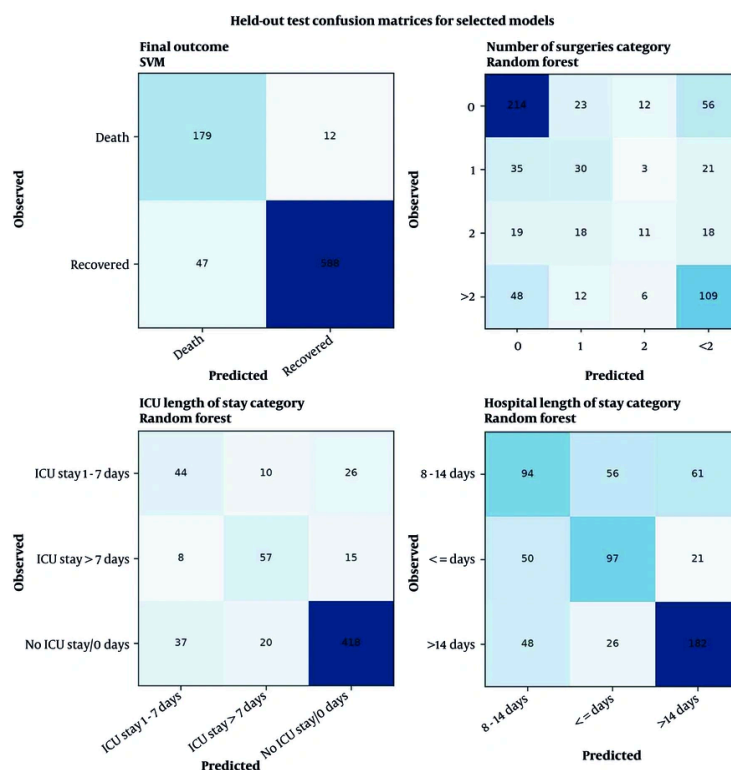


**Figure 1.** Held-out test confusion matrices for selected models for final outcome, number of surgeries, ICU length of stay, and hospital length of stay.

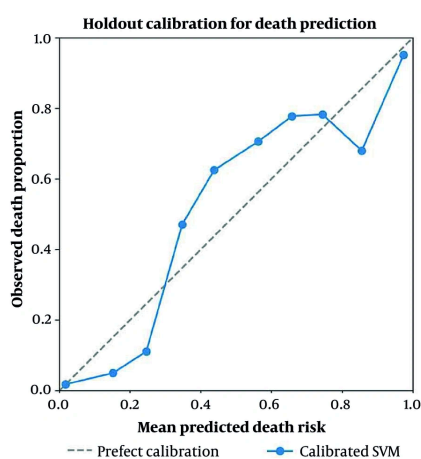


**Figure 2.** Calibration of predicted death risk for the calibrated SVM final-outcome model.

Permutation importance identified total body surface area burned as the dominant predictor for final outcome and the length-of-stay targets. Age, triage level, level of consciousness, systolic blood pressure, referral pathway, burn mechanism, burn depth, anatomical burn region, and comorbidities also contributed across targets. Figure 3 presents compact feature-importance plots for the four selected models.

## 5. Discussion

This study found that structured prehospital and admission burn-registry data can support early prediction of several clinically relevant burn outcomes, although performance differed substantially by target. Final outcome was predicted with strong class-balanced performance, whereas performance for ICU length of stay was moderate. Hospital length of stay and number of surgeries were more difficult to predict. This pattern is clinically plausible: mortality and ICU need are strongly linked to burn extent, burn depth, age, physiological compromise, and level of consciousness, whereas

**Table 5.** Class-Wise Accuracy (Recall) of the Five Classifiers, Mean (SD) Over 5 × 2 Cross-Validation Splits in the Development Set <sup>a</sup>

| Target and Class                  | n/Fold | Logistic Regression        | Decision Tree              | Random Forest              | SVM                        | KNN                        |
|-----------------------------------|--------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| <b>Final outcome</b>              |        |                            |                            |                            |                            |                            |
| Recovered                         | 508    | 0.923 (0.012)              | 0.894 (0.020)              | 0.932 (0.012)              | 0.924 (0.011)              | 0.983 (0.003) <sup>b</sup> |
| Death                             | 152    | 0.898 (0.021)              | 0.890 (0.023)              | 0.875 (0.021)              | 0.912 (0.020) <sup>b</sup> | 0.651 (0.052)              |
| <b>Number of surgeries</b>        |        |                            |                            |                            |                            |                            |
| 0                                 | 244    | 0.498 (0.032)              | 0.401 (0.048)              | 0.709 (0.025)              | 0.589 (0.044)              | 0.854 (0.024) <sup>b</sup> |
| 1                                 | 71     | 0.382 (0.039) <sup>b</sup> | 0.348 (0.062)              | 0.364 (0.039)              | 0.381 (0.030)              | 0.232 (0.029)              |
| 2                                 | 52     | 0.259 (0.043)              | 0.295 (0.035) <sup>b</sup> | 0.126 (0.035)              | 0.265 (0.064)              | 0.082 (0.031)              |
| > 2                               | 140    | 0.510 (0.065)              | 0.424 (0.027)              | 0.670 (0.041) <sup>b</sup> | 0.547 (0.049)              | 0.466 (0.043)              |
| <b>ICU length of stay, d</b>      |        |                            |                            |                            |                            |                            |
| No ICU stay / 0                   | 380    | 0.714 (0.020)              | 0.649 (0.028)              | 0.833 (0.009)              | 0.717 (0.016)              | 0.982 (0.010) <sup>b</sup> |
| ICU stay 1 - 7                    | 64     | 0.520 (0.053)              | 0.520 (0.055)              | 0.445 (0.038)              | 0.573 (0.037) <sup>b</sup> | 0.188 (0.043)              |
| ICU stay > 7                      | 64     | 0.670 (0.077) <sup>b</sup> | 0.621 (0.058)              | 0.659 (0.085)              | 0.657 (0.088)              | 0.429 (0.058)              |
| <b>Hospital length of stay, d</b> |        |                            |                            |                            |                            |                            |
| ≤ 7                               | 134    | 0.655 (0.035)              | 0.576 (0.059)              | 0.636 (0.032)              | 0.663 (0.032) <sup>b</sup> | 0.508 (0.038)              |
| 8 - 14                            | 169    | 0.376 (0.032)              | 0.470 (0.036)              | 0.414 (0.050)              | 0.444 (0.037)              | 0.496 (0.023) <sup>b</sup> |
| > 14                              | 205    | 0.622 (0.048)              | 0.572 (0.059)              | 0.697 (0.043) <sup>b</sup> | 0.605 (0.037)              | 0.671 (0.032)              |

<sup>a</sup> n/fold = mean number of patients of that class per validation fold. Recall is the proportion of that class correctly classified.

<sup>b</sup> Identifies the highest mean recall for that class.

operations and total hospitalization are also influenced by wound evolution, timing of excision or grafting, complications, rehabilitation needs, operating-room access, and institutional discharge practices.

The final-outcome findings are consistent with published burn prediction literature showing that total body surface area burned and age are central mortality predictors (4, 13, 14). Recent machine-learning studies of burn mortality have also reported strong performance when models include burn extent, age, depth, physiological markers, or laboratory variables (5-7). In this registry, our final-outcome model exceeded the Baux-score benchmark in macro-F1, suggesting incremental classification value from the broader structured predictor set. This comparison supports the premise that routinely collected early registry variables can provide useful predictive information beyond a simple severity summary.

The resource-use findings should be interpreted more cautiously. Previous studies of length-of-stay prediction in burn patients have shown that total body surface area burned, burn depth, inhalation trauma, comorbidity, and social factors influence hospitalization, and that prediction errors increase with greater burn severity (11, 12). Yeh et al. also modeled hospitalization longer than 14 days in high-risk burn patients and emphasized the role of total body surface area burned and clinical variables (11). Our moderate macro-F1 values for hospital length of stay

and ICU length of stay are therefore credible but not definitive. These targets are not purely biological outcomes; they also reflect care pathways and local capacity.

Class imbalance was a central methodological issue. Accuracy alone would have overstated performance because large classes dominated several targets. Macro-F1 was therefore emphasized because it assigns equal weight to each class and better reflects whether minority classes are recognized. The number-of-surgeries model remained the weakest, especially for the two-surgery category, indicating that the early registry variables did not fully capture the later operative trajectory. This finding is clinically important: a model that performs well for common classes but poorly for smaller classes may still be useful for population-level planning, but it should not be used to make individual operative predictions. The class-wise comparison shows that the classifiers differed mainly in how they balanced recognition of the largest class against recognition of the smaller classes. KNN achieved the highest recall for the largest class in several targets (recovery 0.983, no ICU stay 0.982, zero surgeries 0.854), but had low specificity for these classes and poor recall for the smaller ones (death 0.651, ICU stay of 1 - 7 days 0.188, two surgeries 0.082). Its overall accuracy of 0.812 for ICU length of stay was the highest of the five models, whereas its balanced accuracy of 0.533 was the lowest, again demonstrating why accuracy alone was not an adequate

**Table 6.** Class-Wise F1-Score of the Five Classifiers, Mean (SD) Over 5 × 2 Cross-Validation Splits in the Development Set

| Target and Class                  | n/Fold | Logistic Regression | Decision Tree | Random Forest              | SVM                        | KNN                        |
|-----------------------------------|--------|---------------------|---------------|----------------------------|----------------------------|----------------------------|
| <b>Final outcome</b>              |        |                     |               |                            |                            |                            |
| Recovered                         | 508    | 0.945 (0.006)       | 0.928 (0.009) | 0.947 (0.005)              | 0.948 (0.007) <sup>a</sup> | 0.942 (0.008)              |
| Death                             | 152    | 0.833 (0.016)       | 0.793 (0.017) | 0.833 (0.013)              | 0.843 (0.018) <sup>a</sup> | 0.761 (0.039)              |
| <b>Number of surgeries</b>        |        |                     |               |                            |                            |                            |
| 0                                 | 244    | 0.581 (0.031)       | 0.503 (0.049) | 0.707 (0.019)              | 0.654 (0.033)              | 0.709 (0.020) <sup>a</sup> |
| 1                                 | 71     | 0.318 (0.031)       | 0.269 (0.040) | 0.383 (0.041) <sup>a</sup> | 0.350 (0.022)              | 0.298 (0.027)              |
| 2                                 | 52     | 0.191 (0.034)       | 0.179 (0.016) | 0.181 (0.049)              | 0.207 (0.049) <sup>a</sup> | 0.138 (0.047)              |
| > 2                               | 140    | 0.503 (0.040)       | 0.440 (0.033) | 0.591 (0.031) <sup>a</sup> | 0.533 (0.035)              | 0.514 (0.040)              |
| <b>ICU length of stay, d</b>      |        |                     |               |                            |                            |                            |
| No ICU stay / 0                   | 380    | 0.805 (0.011)       | 0.758 (0.021) | 0.866 (0.008)              | 0.811 (0.012)              | 0.896 (0.011) <sup>a</sup> |
| ICU stay 1 - 7                    | 64     | 0.365 (0.037)       | 0.344 (0.041) | 0.406 (0.041) <sup>a</sup> | 0.382 (0.025)              | 0.291 (0.056)              |
| ICU stay > 7                      | 64     | 0.538 (0.053)       | 0.462 (0.036) | 0.580 (0.060) <sup>a</sup> | 0.551 (0.050)              | 0.541 (0.056)              |
| <b>Hospital length of stay, d</b> |        |                     |               |                            |                            |                            |
| ≤ 7                               | 134    | 0.565 (0.020)       | 0.521 (0.037) | 0.577 (0.031) <sup>a</sup> | 0.575 (0.022)              | 0.537 (0.035)              |
| 8 - 14                            | 169    | 0.406 (0.029)       | 0.466 (0.032) | 0.448 (0.040)              | 0.454 (0.032)              | 0.483 (0.014) <sup>a</sup> |
| > 14                              | 205    | 0.648 (0.028)       | 0.617 (0.042) | 0.699 (0.032) <sup>a</sup> | 0.657 (0.025)              | 0.662 (0.019)              |

<sup>a</sup> Identifies the highest mean F1-score for that class.

summary. Random forest and SVM, selected on macro-F1, provided the best balance across classes: the selected model had the highest F1-score in 8 of the 12 classes and was never ranked last, although it was not best in every class (for example, SVM recalled more patients with an ICU stay of 1 - 7 days than random forest, 0.573 versus 0.445, at the cost of lower precision, 0.287 versus 0.374). Performance was weakest for the intermediate categories (two surgeries, 1 - 7 ICU days, 8 - 14 hospital days) for every algorithm, which suggests that early registry variables carry limited information for separating adjacent ordinal categories rather than indicating a weakness of one specific algorithm.

The feature-importance results were clinically coherent. Total body surface area burned was the leading predictor for final outcome and length-of-stay targets, reflecting the central role of burn extent in systemic injury burden (2, 4). Age contributed to final outcome, consistent with Baux-based mortality prediction (13, 14). Burn depth, anatomical region, referral pathway, vital signs, level of consciousness, and comorbidities likely provide complementary information about injury severity, physiological compromise, access route, and baseline vulnerability. These findings should be interpreted as model-specific predictive patterns, not causal effects. Correlated predictors can share or mask importance, and importance values should not be compared directly across different target models.

The intended clinical value is early support for clinicians and hospital teams, not automatic treatment assignment. In the target setting, a prediction could help identify patients who may require intensive care unit (ICU) care, prolonged hospitalization, or higher operative resource use, supporting early planning, audit, and resource discussions. Practical deployment would require prospective testing in the actual triage and admission workflow, confirmation that variables are available before the intended prediction time, calibration monitoring, threshold selection aligned with clinical consequences, and governance to prevent overreliance on imperfect predictions.

### 5.1. Study Limitations

This study has important limitations. Its retrospective, single-center design means that the reported performance may not generalize to other hospitals, regions, registry systems, or practice patterns. External validation in multicenter and geographically distinct settings is needed. Several predictors were documented during triage or early admission, but practical deployment would require confirming local information flow at the intended prediction time. Length of stay and number of surgeries reflect clinical severity as well as institutional capacity and practice. Although preprocessing reduced leakage and coding artifacts, residual bias, unmeasured confounding, and documentation heterogeneity may remain. The

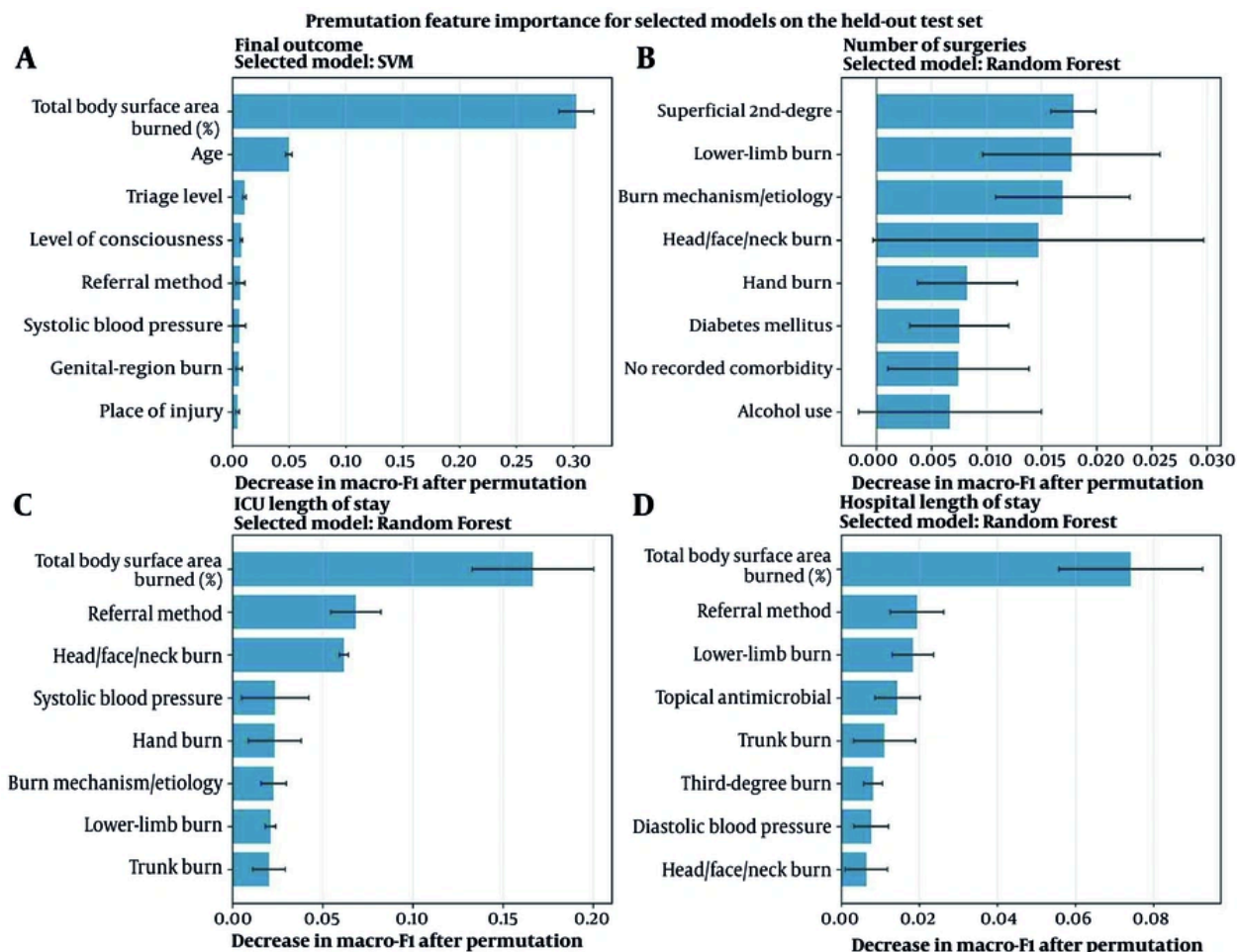


**Figure 3.** Permutation feature importance for selected models. Error bars show standard deviation across permutation repeats. Feature labels are presented in clinically interpretable wording.

models are exploratory and should not be used for individual bedside decisions without prospective validation.

## 5.2. Conclusions

This study demonstrates the feasibility of using routinely collected prehospital, triage, and early-admission burn registry data to predict clinically relevant burn outcomes. Model performance was strongest for clinical final outcome and moderate for ICU length-of-stay classification, whereas hospital length of stay and the number of surgeries were more difficult to predict, reflecting the influence of the evolving clinical course and institutional care pathways. The findings support the

potential value of structured early registry data for risk stratification and resource-planning research, but they should not be interpreted as evidence for immediate bedside deployment. External multicenter validation, prospective workflow assessment, calibration monitoring, and evaluation of clinical utility are required before these models can be considered for routine clinical use.

## Footnotes

**AI Use Disclosure:** ChatGPT was used for text editing in the etc section at a Moderate level.

**Authors' Contribution:** Study Concept and Design: A. K., A. S., and N. A.; Data Collection Protocol Design and

Oversight: A. K.; Machine Learning Study Design and Target-Specific Modeling Framework: A. S.; Methodological Supervision: N. A.; Data Acquisition, Preparation, and Clinical Registry Support: F. G. and M. M.; Data Preprocessing, Model Development and Evaluation, Machine Learning Results Interpretation, and Statistical Analysis: A. S.; Model Development Support: Z. T.; Data Analysis Support: I. K.; Clinical Interpretation of Data and Findings: F. G. and M. M.; Clinical and Methodological Interpretation of Results: A. K. and N. A.; Manuscript Drafting: A. S.; Critical Revision for Important Intellectual Content: A. K., N. A., F. G., M. M., Z. T., I. K., and A. S.; Administrative, Technical, and Material Support: A. K., F. G., M. M., and N. A.; Study Supervision: A. K. and N. A.; All authors read and approved the final manuscript.

**Conflict of Interests Statement:** The authors do not declare any conflicts of interests for this study.

**Data Availability:** The dataset presented in the study is available on request from the corresponding author during submission or after publication.

**Ethical Approval:** Ethical approval was obtained under the code IR.SUMS.REC.1405.091 on 2026/05/04, and the approval is accessible at ris.sums.ac.ir.

**Funding/Support:** No funding was received for this study.

## References

- World Health Organization. *World Health Organization*. World Health Organization; 2023, [cited 2026 Aug 23]. Burns. Available from: <https://www.who.int/news-room/fact-sheets/detail/burns/>.
- Jeschke MG, van Baar ME, Choudhry MA, Chung KK, Gibran NS, Logsetty S. Burn injury. *Nat Rev Dis Primers*. 2020;**6**(1). 11. [PubMed ID: 32054846]. [PubMed Central ID: PMC7224101]. <https://doi.org/10.1038/s41572-020-0145-5>.
- Mrad MA, Al Qurashi AA, Shah Mardan QNM, Al Jabr FA, Almenhali AA, Bamakhrama B, et al. Risk models to predict mortality in burn patients: a systematic review and meta-analysis. *Plast Reconstr Surg Glob Open*. 2022;**10**(12). e4694. [PubMed ID: 36569241]. [PubMed Central ID: PMC9760622]. <https://doi.org/10.1097/GOX.0000000000004694>.
- Edgar MC, Bond SM, Jiang SH, Scharf IM, Bejarano G, Vrouwe SQ. The Revised Baux score as a predictor of burn mortality: a systematic review and meta-analysis. *J Burn Care Res*. 2023;**44**(6):1278-1288. [PubMed ID: 37220881]. <https://doi.org/10.1093/jbcr/irad075>.
- Osler T, Glance LG, Hosmer DW. Simplified estimates of the probability of death after burn injuries: extending and updating the Baux score. *J Trauma*. 2010;**68**(3):690-697. [PubMed ID: 20038856]. <https://doi.org/10.1097/TA.0b013e3181c453b3>.
- Taib BG, Karwath A, Wensley K, Minku L, Gkoutos GV, Moiemien N. Artificial intelligence in the management and treatment of burns: a systematic review and meta-analyses. *J Plast Reconstr Aesthet Surg*. 2023;**77**:133-161. [PubMed ID: 36571960]. <https://doi.org/10.1016/j.bjps.2022.11.049>.
- Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;**1**(1). 18. [PubMed ID: 31304302]. [PubMed Central ID: PMC6550175]. <https://doi.org/10.1038/s41746-018-0029-1>.
- Çinar MA, Ölmez E, Erkiş A, Bayramlar K, Er O. Machine learning models for early prediction of mortality risk in patients with burns: a single center experience. *J Plast Reconstr Aesthet Surg*. 2024;**89**:14-20. [PubMed ID: 38118361]. <https://doi.org/10.1016/j.bjps.2023.11.048>.
- Yazici H, Ugurlu O, Aygül Y, Yildirim M, Ucar AD. The effect of well-known burn-related features on machine learning algorithms in burn patients' mortality prediction. *Ulus Travma Acil Cerrahi Derg*. 2023;**29**(10):1130-1137. [PubMed ID: 37791433]. [PubMed Central ID: PMC10644077]. <https://doi.org/10.14744/tjtes.2023.79968>.
- Park JH, Cho Y, Shin D, Choi SS. Prediction of mortality after burn surgery in critically ill burn patients using machine learning models. *J Pers Med*. 2022;**12**(8):1293. [PubMed ID: 36013242]. [PubMed Central ID: PMC9410169]. <https://doi.org/10.3390/jpm12081293>.
- Yeh CC, Lin YS, Chen CC, Liu CF. Implementing AI models for prognostic predictions in high-risk burn patients. *Diagnostics (Basel)*. 2023;**13**(18):2984. [PubMed ID: 37761351]. [PubMed Central ID: PMC10528558]. <https://doi.org/10.3390/diagnostics13182984>.
- Esumi R, Funao H, Kawamoto E, Sakamoto R, Ito-Masui A, Okuno F, et al. Machine learning-based prediction of delirium and risk factor identification in intensive care unit patients with burns: retrospective observational study. *JMIR Form Res*. 2025;**9**. e65190. [PubMed ID: 39895101]. [PubMed Central ID: PMC11923481]. <https://doi.org/10.2196/65190>.
- Drysch M, Reinkemeier F, Pusch F, Hinzmann J, Paul Christian F, Lehnhardt M, et al. Streamlined machine learning model for early sepsis risk prediction in burn patients. *NPJ Digit Med*. 2025;**8**(1). 621. [PubMed ID: 41120704]. [PubMed Central ID: PMC12540978]. <https://doi.org/10.1038/s41746-025-02078-z>.
- Elrod J, Mohr C, Wolff R, Boettcher M, Reinshagen K, Bartels P, et al. Using artificial intelligence to obtain more evidence? Prediction of length of hospitalization in pediatric burn patients. *Front Pediatr*. 2020;**8**. 613736. [PubMed ID: 33537267]. [PubMed Central ID: PMC7849450]. <https://doi.org/10.3389/fped.2020.613736>.
- Onah CN, Allmendinger R, Handl J, Dunn KW. Surviving burn injury: drivers of length of hospital stay. *Int J Environ Res Public Health*. 2021;**18**(2):761. [PubMed ID: 33477442]. [PubMed Central ID: PMC7829802]. <https://doi.org/10.3390/ijerph18020761>.
- Nijman SWJ, Leeuwenberg AM, Beekers I, Verkouter I, Jacobs JLL, Bots ML, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *J Clin Epidemiol*. 2022;**142**:218-229. [PubMed ID: 34798287]. [PubMed Central ID: PMC11743349]. <https://doi.org/10.1016/j.jclinepi.2021.11.023>.
- Andaur Navarro CL, Damen JAA, van Smeden M, Takada T, Nijman SWJ, Dhiman P, et al. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. *J Clin Epidemiol*. 2023;**154**:8-22. [PubMed ID: 36436815]. <https://doi.org/10.1016/j.jclinepi.2022.11.015>.
- Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of Studies Conducted using Observational Routinely-collected Health Data (RECORD) statement. *PLoS Med*. 2015;**12**(10). e1001885. [PubMed ID: 26440803]. [PubMed Central ID: PMC4595218]. <https://doi.org/10.1371/journal.pmed.1001885>.
- Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;**385**. e078378. [PubMed ID: 38626948]. [PubMed Central ID: PMC11019967]. <https://doi.org/10.1136/bmj-2023-078378>.
- Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical

- prediction model. *BMJ*. 2020;**368**:m441. [PubMed ID: [32188600](#)]. <https://doi.org/10.1136/bmj.m441>.
21. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;**17**(1). 230. [PubMed ID: [31842878](#)]. [PubMed Central ID: [PMC6912996](#)]. <https://doi.org/10.1186/s12916-019-1466-7>.
  22. Manouchehri T, Fereidooni R, Heydari ST, Bagheri Lankarani K. Application of machine learning in accident data analysis: a case study using self-report questionnaire. *Shiraz E-Med J*. 2025;**26**(6). e158678. <https://doi.org/10.5812/semj-158678>.
  23. Rabeifar F. Big data and machine learning in telemedicine: implications for nursing practice, patient safety, and care quality. *Mod Care J*. 2026;**23**(1). e166124. <https://doi.org/10.69107/mcj-166124>.
  24. Haj Lotfalian M, Samadi H, Abootalebi V. Main features for detecting mental fatigue in elderly gait: a machine learning approach. *J Motor Control Learn*. 2025;**7**(2). e161864. <https://doi.org/10.5812/jmcl-161864>.

**Table 1.** Consolidated Descriptive Characteristics and Outcome Distributions of the Analytic Registry <sup>a</sup>

| Characteristic                     | Overall Registry     | Missing |
|------------------------------------|----------------------|---------|
| <b>Demographic and burn extent</b> |                      |         |
| Age, y                             | 34.0 [17.0, 46.0]    | 25      |
| Total body surface area burned, %  | 20.0 [8.0, 41.0]     | 0       |
| <b>Demographics and referral</b>   |                      |         |
| Sex                                |                      | 35      |
| Male                               | 3,044 (73.8)         |         |
| Female                             | 1,048 (25.4)         |         |
| Missing                            | 35 (0.8)             |         |
| Nationality                        |                      | 0       |
| Iranian                            | 3,848 (93.2)         |         |
| Non-Iranian                        | 279 (6.8)            |         |
| <b>Marital status</b>              |                      |         |
| Married                            | 2,088 (50.6)         | 3       |
| Single                             | 2,036 (49.3)         |         |
| Missing                            | 3 (0.1)              |         |
| <b>Education level</b>             |                      |         |
| Missing                            | 1,287 (31.2)         | 1287    |
| Less than high school diploma      | 1,248 (30.2)         |         |
| High school diploma                | 766 (18.6)           |         |
| University education               | 358 (8.7)            |         |
| No formal education                | 358 (8.7)            |         |
| Student                            | 110 (2.7)            |         |
| <b>Rural residence</b>             |                      |         |
| False                              | 2,696 (65.3)         | 13      |
| True                               | 1,418 (34.4)         |         |
| Missing                            | 13 (0.3)             |         |
| <b>Referral pathway</b>            |                      |         |
| Personal                           | 2,007 (48.6)         | 70      |
| Transfer from other hospital       | 1,081 (26.2)         |         |
| EMS                                | 969 (23.5)           |         |
| Missing                            | 70 (1.7)             |         |
| <b>Vital signs and triage</b>      |                      |         |
| Systolic blood pressure, mmHg      | 116.0 [110.0, 129.0] | 544     |
| Diastolic blood pressure, mmHg     | 70.0 [65.0, 80.0]    | 544     |
| Pulse rate, beats/min              | 94.0 [86.0, 110.0]   | 526     |
| Respiratory rate, breaths/min      | 20.0 [18.0, 21.0]    | 535     |
| Temperature, °C                    | 37.0 [36.5, 37.0]    | 559     |
| Peripheral oxygen saturation, %    | 96.0 [95.0, 97.0]    | 791     |
| <b>Triage level</b>                |                      |         |
| Level 2                            | 2,805 (68.0)         | 280     |
| Level 1                            | 543 (13.2)           |         |
| Level 3                            | 350 (8.5)            |         |
| Missing                            | 280 (6.8)            |         |
| Level 4                            | 122 (3.0)            |         |
| Level 5                            | 27 (0.7)             |         |
| <b>Level of consciousness</b>      |                      |         |
| A                                  | 2,743 (66.5)         | 915     |
| Missing                            | 915 (22.2)           |         |
| U                                  | 300 (7.3)            |         |
| P                                  | 132 (3.2)            |         |
| V                                  | 37 (0.9)             |         |
| <b>Acute clinical signs</b>        |                      |         |
| Pregnancy status                   | 16 (0.4)             | 0       |
| Airway compromise                  | 28 (0.7)             | 0       |
| Respiratory distress               | 11 (0.3)             | 0       |
| Signs of shock                     | 13 (0.3)             | 0       |
| Lethargy / drowsiness              | 12 (0.3)             | 0       |
| Severe pain or distress            | 6 (0.1)              | 0       |
| <b>Burn depth and context</b>      |                      |         |
| First-degree burn                  | 36 (0.9)             | 0       |
| Superficial second-degree burn     | 737 (17.9)           | 0       |
| Deep second-degree burn            | 2,540 (61.5)         | 0       |

| Characteristic                                             | Overall Registry                                   | Missing |
|------------------------------------------------------------|----------------------------------------------------|---------|
| Third-degree burn                                          | 3,364 (81.5)                                       | 0       |
| Fourth-degree burn                                         | 453 (11.0)                                         | 0       |
| <b>Place of injury</b>                                     |                                                    | 98      |
| Home                                                       | 2,646 (64.1)                                       |         |
| Public place                                               | 729 (17.7)                                         |         |
| Workplace                                                  | 654 (15.8)                                         |         |
| Missing                                                    | 98 (2.4)                                           |         |
| <b>Burn intent</b>                                         |                                                    | 57      |
| Unintentional                                              | 3,788 (91.8)                                       |         |
| Self-immolation                                            | 234 (5.7)                                          |         |
| Missing                                                    | 57 (1.4)                                           |         |
| Assault-related burn                                       | 48 (1.2)                                           |         |
| <b>Burn mechanism/etiology</b>                             |                                                    | 93      |
| Flame / fire                                               | 1,415 (34.3)                                       |         |
| Explosion                                                  | 1,226 (29.7)                                       |         |
| Hot liquids / steam                                        | 735 (17.8)                                         |         |
| Contact with hot object                                    | 274 (6.6)                                          |         |
| Electrical                                                 | 262 (6.3)                                          |         |
| Chemical                                                   | 122 (3.0)                                          |         |
| Missing                                                    | 93 (2.3)                                           |         |
| <b>Outcomes</b>                                            |                                                    |         |
| Clinical final outcome                                     |                                                    | 0       |
| Recovery                                                   | 3,175 (76.9; includes 39 partial recovery records) |         |
| Death                                                      | 952 (23.1)                                         |         |
| <b>Number of surgeries among recovered patients</b>        |                                                    | 0       |
| 0                                                          | 1,526                                              |         |
| 1                                                          | 446                                                |         |
| 2                                                          | 327                                                |         |
| > 2                                                        | 876                                                |         |
| <b>ICU length of stay among recovered patients, d</b>      |                                                    | 0       |
| No ICU stay/0                                              | 2,377                                              |         |
| 1-7                                                        | 400                                                |         |
| > 7                                                        | 398                                                |         |
| <b>Hospital length of stay among recovered patients, d</b> |                                                    | 0       |
| ≤ 7                                                        | 838                                                |         |
| 8 - 14                                                     | 1,055                                              |         |
| > 14                                                       | 1,282                                              |         |

<sup>a</sup> Values are expressed as median [IQR], No. (%) or distribution summaries. Missing values are reported before imputation. Abbreviation: SpO<sub>2</sub>, peripheral oxygen saturation.