Submitted: 22 March, 2025 Accepted: 08 May, 2025 Published: 08 January, 2026 DOI:10.22514/sv.2025.161

Open Access
Signa Vitae

ORIGINAL RESEARCH

Early sepsis prediction in elderly patients with urinary tract infections: a machine learning

İzzet Ustaalioğlu[1,*], Ferhat Yıldız[2]

[1]Department of Emergency Medicine, Gönen State Hospital, 10900 Balıkesir, Türkiye
[2]Department of Emergency Medicine, Mardin Training and Research Hospital, 47100 Mardin, Türkiye

***Correspondence**
izzetustaalioglu@gmail.com
(İzzet Ustaalioğlu)

Abstract

Background: Sepsis remains a leading cause of morbidity and mortality, particularly among elderly patients, for whom urinary tract infections (UTIs) are a significant trigger. This study applies machine learning methods to identify early predictors of sepsis in elderly patients presenting with UTIs in the emergency department (ED). **Methods**: This retrospective study analyzed elderly patients with UTIs over five years, excluding those with sepsis at presentation, secondary infections or incomplete records. Logistic regression, Generalized Additive Modeling (GAM), Least Absolute Shrinkage and Selection Operator (LASSO) regression, and a Decision Tree model were evaluated for sepsis prediction within 72 hours. Model performance was assessed using area under the curve (AUC), Brier scores, and Net Reclassification Improvement. **Results**: Of 1176 patients, 139 (11.8%) developed sepsis within 72 hours. Independent predictors included age (adjusted odds ratio (aOR) 1.10, 95% confidence interval (CI) 1.05–1.14), blood urea nitrogen (BUN) (aOR 1.03, 95% CI 1.00–1.05), C-reactive protein (CRP) (aOR 1.04, 95% CI 1.03–1.05), creatinine (aOR 1.79, 95% CI 1.11–3.14), respiratory rate (aOR 1.10, 95% CI 1.05–1.16), temperature (aOR 2.52, 95% CI 1.65–4.38), and lower systolic blood pressure (aOR 0.95, 95% CI 0.92–0.97). GAM (AUC 0.954) and LASSO (AUC 0.942) outperformed logistic regression (AUC 0.792, $p < 0.001$). GAM showed superior discrimination over the Decision Tree (AUC 0.915, $p = 0.046$). **Conclusions**: This study highlights that clinical parameters including age, BUN, CRP, creatinine, respiratory rate, body temperature and systolic blood pressure are independent risk factors for early sepsis in elderly patients with UTIs. These factors should be carefully considered when assessing elderly patients presenting to the ED for sepsis risk.

Keywords

Sepsis prediction; Machine learning; Artificial intelligence; Urinary tract infections; Elderly patients; Biomarkers; Logistic regression

1. Introduction

Despite advances in early warning systems, changing diagnostic criteria and increased awareness, sepsis continues to be a leading cause of poor clinical outcomes [1, 2]. In 2016, the definition of sepsis was updated to include the concepts of Sepsis-Related Organ Failure Assessment (SOFA) and quick-SOFA (qSOFA), thereby clarifying the criteria for organ dysfunction in sepsis [3]. Elderly patients, due to factors like polypharmacy, chronic diseases, and altered physiology and metabolism, are particularly vulnerable to developing organ dysfunction [4, 5]. Tools such as SOFA and qSOFA, are effective at identifying patients with existing organ dysfunction [6]. However, these tools are primarily designed to assess the presence and extent of existing organ dysfunction at the time of evaluation.

Urinary tract infections (UTIs) are common in the elderly and often manifest atypically, including altered mental status or functional decline, posing diagnostic challenges in the emergency department (ED) [7]. The lack of specific and reliable risk scores tailored to the ED setting, combined with the insufficient identification of factors influencing poor outcomes, complicates the management of these patients. Although several risk models exist for sepsis prediction, most have not been validated in elderly patients with UTIs—a population with distinct clinical trajectories. Furthermore, conventional tools are largely based on linear associations and may overlook complex, interaction-driven risk patterns common in geriatric presentations. Machine learning (ML) methods such as generalized additive models (GAM), least absolute shrinkage and selection operator (LASSO), and decision tree algorithms offer the potential to uncover complex, non-linear relationships among clinical features, which may enhance early risk stratification beyond traditional models [8, 9]. These approaches have shown promise in improving prediction accuracy while maintaining clinical interpretability [10]. However, their ap-

plication specifically in elderly patients with UTIs remains limited, with few studies directly comparing ML models to conventional regression in this context.

This study, therefore, aims to identify predictors of early sepsis within 72 hours in elderly patients presenting to the ED with UTIs. The predictive performance of ML models including GAM, LASSO regression, and decision trees is evaluated against multivariable logistic regression.

2. Methods

2.1 Study design and setting

This retrospective observational study was conducted in the ED of a tertiary care hospital. Ethical approval for the study was obtained from Mardin Artuklu University Non-Interventional Clinical Research Ethics Committee (date: 05 November 2024; approval number: 2024/11-12). Due to the retrospective nature of the study, the ethics committee waived the requirement for informed consent. The research was conducted in accordance with the Declaration of Helsinki.

2.2 Study population

Patients aged over 65 years who presented to the ED with a diagnosis of urinary tract infection (UTI) between January 2019 and December 2023 were eligible for inclusion. The diagnosis was based on the International Classification of Diseases (ICD)-10 codes assigned by emergency physicians at the time of admission, in line with institutional protocols, as detailed in **Supplementary Table 1**. These codes were retrospectively reviewed to identify eligible cases. Inclusion criteria encompassed patients with typical UTI symptoms, such as dysuria, urgency or frequency. These symptoms were supported by urinalysis findings. These included urine microscopy showing $\geq$10 white blood cells/μL, and positive dipstick tests for leukocyte esterase and nitrites. We also included patients with atypical symptoms often seen in older adults, such as altered mental status, general malaise or functional decline. These cases were eligible only if they did not meet Sepsis-3 criteria at presentation [7, 11]. Exclusion criteria included a diagnosis of sepsis upon ED arrival, defined by Sepsis-3 as an acute increase of two or more points in the SOFA score; the presence of concurrent secondary infections such as pneumonia or cellulitis; and incomplete clinical or laboratory data. All diagnoses were made by the treating physician and verified retrospectively through ICD-10 coding, without modification, by the study team. Urine blood and protein were defined as positive if dipstick testing or microscopy confirmed hematuria or proteinuria, respectively. The presence of urine microorganisms referred to bacteria observed via microscopic urinalysis.

2.3 Data quality and management

Data quality was ensured through independent extraction by two reviewers. Inter-observer reliability was assessed using a random sample of 120 patient records, representing approximately 10 percent of the cohort, which were independently reviewed by both assessors. Every variable included in the study—demographics, comorbidities, vital signs, laboratory values and clinical assessments—was cross-checked between reviewers. Agreement was excellent, with Cohen's Kappa for categorical variables reaching 0.94 and intraclass correlation coefficients (ICCs) for continuous variables exceeding 0.98. Discrepancies were resolved by a third reviewer blinded to the initial assessments. A prespecified multiple imputation protocol was established to address missing data if any variable exceeded 5% missingness. However, as no variable exceeded the 5% missing data threshold, complete case analysis was used without the need for imputation.

2.4 Data collection and outcome measures

Demographic data, chronic comorbidities, vital signs at presentation, Charlson Comorbidity Index scores, laboratory parameters at presentation, urine analysis parameters, and the development of sepsis within 72 hours were recorded. Data were retrieved from the hospital information management system. All data were anonymized before analysis to maintain patient confidentiality. Data collection was performed by the researchers to ensure consistent and accurate information gathering. Body temperature values were obtained retrospectively from the triage vital signs recorded at ED presentation, but the method of measurement was not consistently specified in the electronic medical records. Sepsis was defined, according to the Sepsis-3 criteria, as life-threatening organ dysfunction caused by infection, indicated by an acute increase of $\geq$2 points in the SOFA score [3]. The SOFA score was calculated by assigning a score from 0 to 4 for dysfunction in six organ systems—respiratory, coagulation, liver, cardiovascular, kidney and nervous—resulting in a total score ranging from 0 to 24 [12]. The primary outcome of the study was the development of sepsis within 72 hours of the ED visit. SOFA scores were retrospectively calculated using clinical and laboratory parameters recorded in the electronic health record (EHR) at the time of admission. All values used in the SOFA calculation were derived from physician and nurse documented data. Parameters not measured at admission were not included in the SOFA score calculation. If a parameter required for SOFA calculation was not measured, a default score of 0 was assigned for that component, as per standard retrospective sepsis study methodology. This reflects clinical practice, where unmeasured values are often considered normal unless clinically indicated. Sepsis development within 72 hours was determined by reviewing subsequent clinical and laboratory data recorded in the EHR during hospitalization. If a patient was discharged before 72 hours, their last available documented SOFA score was used to assess progression. The secondary outcomes assessed included mortality, intensive care unit (ICU) admission, vasopressor use, antibiotic changes and clinical deterioration during follow-up. Mortality was defined as death occurring at any point during hospitalization or follow-up. ICU admission was recorded for patients who required transfer to intensive care. Vasopressor use was defined as the administration of any vasoactive medication (*e.g.*, norepinephrine, dopamine or vasopressin) due to hypotension. Antibiotic change was recorded when a patient's antimicrobial regimen was modified during hospitalization due to treatment

failure, antimicrobial resistance or clinical deterioration. Clinical deterioration was defined as a worsening of hemodynamic parameters, respiratory distress, or the need for escalation of care despite appropriate antibiotic therapy. These outcomes were extracted from electronic medical records and verified by independent reviewers. All predictor variables, including vital signs, laboratory data and clinical features, were recorded at the time of ED presentation.

2.5 Sample selection

For observational studies involving logistic regression analysis, it is generally recommended that the sample size includes at least 100 patients, plus an additional 50 patients for each independent variable in the final model. Alternatively, a minimum of 500 patients is advised to ensure statistical robustness [13]. Given the study's aim and available data, the sample size of 1176 patients was deemed appropriate to meet these requirements without compromising the reliability of the results.

2.6 Analysis

All statistical analyses were performed using R (version 4.4.2, R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics were used to summarize baseline characteristics. Categorical variables were reported as counts (n) and percentages (%), while continuous variables were presented as medians with interquartile ranges (IQR) or means with standard deviations (SD), as appropriate. The Kolmogorov-Smirnov test, accompanied by visual inspections using histograms, was applied to assess the normality of continuous variables. Univariable logistic regression analysis was performed to assess the association between individual variables and sepsis development within 72 hours. Continuous variables were dichotomized at their median values and results were expressed as odds ratios (OR) with 95% confidence intervals (CI). Variables in the univariable analysis with an OR whose 95% CI did not include 1 were selected for inclusion in the multivariable logistic regression model to identify independent predictors while controlling for confounding factors. A stepwise backward elimination approach was applied to retain variables with $p < 0.05$ in the final model, and results were reported as adjusted ORs (aOR) with 95% CI. All ML models used the same predictors identified as significant in multivariable logistic regression to ensure comparability. This approach ensured that any differences in performance were due to the modeling technique rather than discrepancies in variable selection. GAM, LASSO regression, and decision trees were implemented using age, BUN, CRP, creatinine, systolic blood pressure, respiratory rate, and temperature as predictors. These variables were chosen based on their statistical significance in the logistic regression model, allowing for direct comparison across methods. For logistic regression, continuous variables were dichotomized at their median values to facilitate interpretability and address non-linearity. In contrast, ML models (GAM, LASSO and decision tree) were constructed using continuous predictors without transformation, allowing for non-linear relationships assessment and threshold-based effects to emerge naturally.

GAM was employed to capture potential non-linear relationships among these predictors. LASSO regression was applied to assess whether penalization techniques could improve predictive accuracy while maintaining interpretability. A decision tree model was developed to explore threshold-based risk stratification for bedside utility. Model performance was evaluated using area under the receiver operating characteristic curve (AUC), Brier scores for calibration, and DeLong's test for statistical comparisons between models. To assess the robustness of the logistic regression model, 5-fold cross-validation was performed. The dataset was randomly divided into five equal parts, and the model was trained on four parts and tested on the remaining part in an iterative manner. The AUC was calculated for both the training and testing datasets in each fold to evaluate the discriminatory ability of the model. The overall AUC values were also calculated by pooling the results across folds, with 95% CIs. Comparative model performance for GAM, LASSO, and decision tree models was evaluated using the test dataset from the 80/20 train-test split. This allowed for direct comparison using identical data inputs. Logistic regression was also evaluated via 5-fold cross-validation for internal validation, with its final performance also reported on the test set for consistency across models. Multicollinearity among predictors was checked using the variance inflation factor (VIF), with a threshold of VIF <5 considered acceptable. Missing data were minimized by careful data extraction. The handling of missing data followed a predefined strategy approved by the ethics committee. If missing values were present in fewer than 5% of cases for a given variable, multiple imputation would have been applied to maintain statistical robustness [14]. For variables with more than 5% missing data, affected cases would have been excluded from the analysis to minimize potential bias. However, in the final dataset, no variable exceeded the 5% threshold for missingness, and therefore, all available data were included in the analysis without the need for imputation or case exclusion. Summary of missingness across predictors and outcome is given in **Supplementary Table 2**.

3. Results

3.1 Incidence of sepsis and univariable analysis

In this study, 1176 elderly patients presenting with UTI to the ED were analyzed (Fig. 1), of which 139 patients (11.8%) developed sepsis within the first 72 hours, whereas 1037 patients (88.2%) did not. Among patients who developed sepsis, the median SOFA score at the time of diagnosis was 4 (IQR 3–6), while the median SOFA score in the non-sepsis group remained 1 (IQR 0–1). Univariable analysis of patient characteristics and vital signs (Table 1) showed that older age (OR 1.15, 95% CI 1.12–1.19), diabetes mellitus (OR 3.18, 95% CI 2.18–4.61), and dementia (OR 2.86, 95% CI 1.76–4.54) were associated with a higher risk of sepsis. Patients with higher respiratory rates (OR 1.32, 95% CI 1.27–1.38), altered mental status (OR 2.55, 95% CI 1.78–3.66), and elevated temperature (OR 2.09, 95% CI 1.72–2.55) had increased odds of sepsis. In contrast, higher systolic (OR 0.86, 95% CI 0.84–0.88) and diastolic

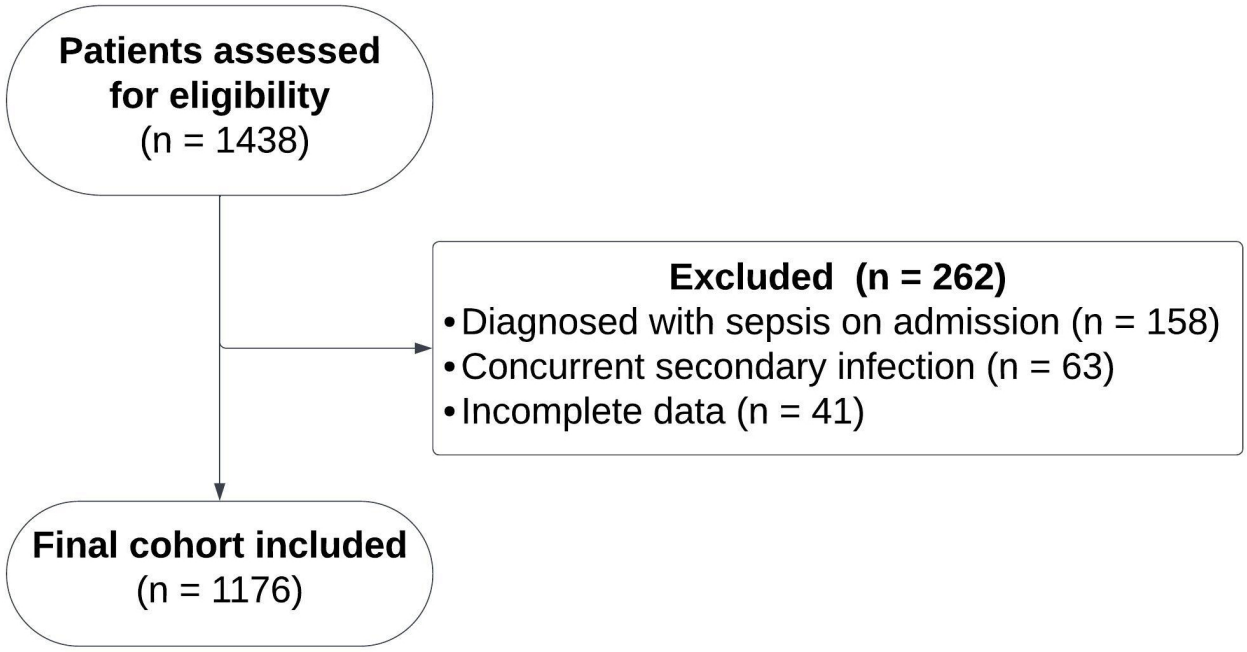

FIGURE 1. Patient flowchart.

TABLE 1. Patient characteristics and vital signs.

Variable	Non-Sepsis Group (n = 1037) (Median (IQR) or Count (%))	Sepsis Group (n = 139) (Median (IQR) or Count (%))	OR (95% CI)
Age (yr)	72 (68–77)	80 (74–87)	1.15 (1.12–1.19)
Sex (Female)	646 (62.3%)	86 (61.9%)	0.98 (0.68–1.42)
Hypertension	330 (31.8%)	37 (26.6%)	0.78 (0.52–1.15)
Diabetes Mellitus	186 (17.9%)	57 (41.0%)	3.18 (2.18–4.61)
Coronary Artery Disease	125 (12.1%)	14 (10.1%)	0.82 (0.44–1.42)
Chronic Kidney Disease	75 (7.2%)	13 (9.4%)	1.32 (0.68–2.38)
Dementia	84 (8.1%)	28 (20.1%)	2.86 (1.76–4.54)
Congestive Heart Failure	25 (2.4%)	2 (1.4%)	0.59 (0.09–2.01)
Stroke	58 (5.6%)	10 (7.2%)	1.31 (0.63–2.51)
Malignancy	25 (2.4%)	7 (5.0%)	2.15 (0.84–4.81)
Systolic BP (mmHg)	123 (116–140)	107 (98–123)	0.86 (0.84–0.88)
Diastolic BP (mmHg)	81 (71–85)	66 (59–79)	0.54 (0.47–0.60)
Heart Rate (bpm)	79 (67–92)	77 (61–98)	1.00 (0.99–1.01)
Respiratory Rate (breaths/min)	17 (13–20)	25 (15–32)	1.32 (1.27–1.38)
Altered Mental Status	320 (30.9%)	74 (53.2%)	2.55 (1.78–3.66)
Temperature (°C)	36.7 (36.1–37.7)	37.3 (36.3–38.4)	2.09 (1.72–2.55)
Charlson Comorbidity Index	3 (3–4)	4 (3–5)	1.82 (1.54–2.16)

BP: Blood pressure; IQR: Interquartile range; OR: odds ratios; CI: confidence intervals.

blood pressure (OR 0.54, 95% CI 0.47–0.60) were negatively associated with sepsis.

Univariable analysis of laboratory parameters (Table 2) identified elevated CRP (OR 4.44, 95% CI 2.93–6.91), higher glucose (OR 3.00, 95% CI 2.04–4.50), BUN (OR 2.08, 95% CI 1.44–3.05), creatinine (OR 1.92, 95% CI 1.32–2.82), and lactate (OR 1.90, 95% CI 1.32–2.76) as significant risk factors. Higher platelet count (OR 0.52, 95% CI 0.36–0.74) was inversely associated with sepsis. The presence of urinary microorganisms (OR 2.10, 95% CI 1.47–3.01) was a significant predictor, while urinary blood and proteinuria were not significantly associated with sepsis.

3.2 Multivariable modeling

Logistic regression analysis identified age (aOR 1.10, 95% CI 1.05–1.14), BUN (aOR 1.03, 95% CI 1.00–1.05), CRP (aOR 1.04, 95% CI 1.03–1.05), creatinine (aOR 1.79, 95%

TABLE 2. Laboratory parameters associated with sepsis.

Variable	Non-Sepsis Group (n = 1037) (Median (IQR) or Mean ± SD)	Sepsis Group (n = 139) (Median (IQR) or Mean ± SD)	OR (95% CI)
WBC (10^3/μL)	8447 ± 1934	9217 ± 2441	1.46 (1.02–2.10)
Glucose (mg/dL)	111.4 ± 42.6	144.1 ± 56.7	3.00 (2.04–4.50)
Hgb (g/dL)	10.4 ± 2.3	10.0 ± 3.1	0.87 (0.61–1.24)
PLT (10^3/μL)	233 (177–287)	191 (126–259)	0.52 (0.36–0.74)
Lactate (mmoL/L)	1.9 ± 1.0	2.4 ± 1.2	1.90 (1.32–2.76)
BUN (mg/dL)	25 (19–30)	33 (20–45)	2.08 (1.44–3.05)
CRP (mg/L)	35.8 (26.0–45.2)	77.8 (45.2–106.4)	4.44 (2.93–6.91)
Creatinine (mg/dL)	1.1 (0.7–1.7)	1.5 (0.9–2.2)	1.92 (1.32–2.82)
Urine Blood	635 (61.2%)	84 (60.4%)	0.97 (0.68–1.40)
Urine Protein	136 (13.1%)	25 (18.0%)	1.45 (0.89–2.29)
Urine Microorganism	378 (36.5%)	76 (54.7%)	2.10 (1.47–3.01)

IQR: Interquartile range; CI: Confidence interval; WBC: White blood cell count; BUN: Blood urea nitrogen; CRP: C-Reactive protein; PLT: Platelet count; Hgb: Hemoglobin; SD: standard deviations; OR: odds ratios.

CI 1.11–3.14), systolic blood pressure (BP) (aOR 0.95, 95% CI 0.92–0.97), respiratory rate (aOR 1.10, 95% CI 1.05–1.16), and temperature (aOR 2.52, 95% CI 1.65–4.38) as statistically significant predictors of sepsis development (Table 3). Age, elevated BUN, CRP, creatinine, increased respiratory rate, and elevated temperature were associated with higher odds of developing sepsis, while lower systolic BP was also independently associated with sepsis risk. Factors such as Diabetes Mellitus, Dementia, white blood cell count, glucose, platelets, lactate, urine microorganism, altered mental status, and Charlson comorbidity index were not statistically significant in the multivariable analysis for predicting sepsis development.

TABLE 3. Adjusted odds ratios from multivariable logistic regression for predictors of sepsis development.

Variable	Adjusted ORs (95% CI)
Age	1.10 (1.05–1.14)
Blood Urea Nitrogen	1.03 (1.00–1.05)
C-Reactive Protein	1.04 (1.03–1.05)
Creatinine	1.79 (1.11–3.14)
Systolic Blood Pressure	0.95 (0.92–0.97)
Respiratory Rate	1.10 (1.05–1.16)
Temperature	2.52 (1.65–4.38)

OR: Odds ratio; CI: Confidence interval.

The area under the curve (AUC) for the multivariable logistic regression model was evaluated for both training and testing datasets. The overall training AUC was 0.792 (95% CI 0.742–0.841, $p < 0.001$), while the overall testing AUC was 0.824 (95% CI 0.733–0.915, $p < 0.001$) (Table 4). Fold-specific training AUCs ranged from 0.779 to 0.824, with testing AUCs ranging from 0.730 to 0.840 across the five folds (Table 4). To facilitate clinical interpretation, we also analyzed age as a continuous variable in 10-year increments. In this additional analysis, a 10-year increase in age was associated with 4.20 times higher odds of sepsis (95% CI 3.22–5.56, $p < 0.001$). The goodness-of-fit of the model was evaluated using the Cox & Snell R Square, Nagelkerke R Square, and the Hosmer and Lemeshow test. The Cox & Snell R Square was 0.320, and the Nagelkerke R Square was 0.433. The Hosmer and Lemeshow test indicated an acceptable fit (Chi-square = 6.654, $df = 8$, $p = 0.284$).

TABLE 4. Area under the receiver operating characteristic curve for predicted sepsis probabilities.

Variable(s)	Training AUC (95% CI)	Testing AUC (95% CI)
All	0.792 (0.742–0.841)	0.824 (0.733–0.915)
Fold 1	0.824 (0.785–0.863)	0.730 (0.682–0.778)
Fold 2	0.779 (0.738–0.820)	0.807 (0.759–0.855)
Fold 3	0.792 (0.752–0.832)	0.787 (0.739–0.835)
Fold 4	0.810 (0.771–0.849)	0.747 (0.699–0.795)
Fold 5	0.789 (0.749–0.829)	0.840 (0.792–0.888)

AUC: Area under curve; CI: Confidence interval.

To assess predictive performance, we compared logistic regression with GAM, LASSO regression and Decision Tree models (Table 5). The area under the curve (AUC) for the multivariable logistic regression model was 0.792 (95% CI 0.742–0.841) in the training set and 0.824 (95% CI 0.733–0.915) in the test set. GAM analysis demonstrated an AUC of 0.9545 (95% CI 0.9317–0.9773, $p < 0.001$). Unlike traditional regression models, GAM allows for smooth, non-linear relationships between predictors and sepsis risk. As shown in Fig. 2, the risk of sepsis increases non-linearly with age, particularly accelerating beyond 78 years. Similarly, blood urea nitrogen and C-reactive protein exhibit threshold effects, where risk significantly rises beyond 35 mg/dL and 70 mg/L, respectively. The effect of respiratory rate is also non-linear, with a notable increase in risk occurring beyond 22 breaths/min. The deviance explained was 63.5%, with

TABLE 5. Diagnostic performance of predictive models.

Model	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 Score	AUC (95% CI)
Logistic Regression	0.725 (0.650–0.790)	0.937 (0.921–0.953)	0.640 (0.565–0.715)	0.964 (0.951–0.975)	0.679	0.792 (0.742–0.841)
GAM	0.871 (0.803–0.921)	0.959 (0.946–0.971)	0.742 (0.668–0.808)	0.982 (0.972–0.989)	0.801	0.954 (0.932–0.977)
LASSO	0.906 (0.845–0.949)	0.986 (0.976–0.992)	0.894 (0.831–0.939)	0.987 (0.979–0.993)	0.900	0.942 (0.910–0.974)
Decision Tree	0.856 (0.787–0.910)	0.933 (0.917–0.948)	0.633 (0.560–0.702)	0.980 (0.969–0.988)	0.728	0.915 (0.884–0.946)

GAM: Generalized Additive Model; LASSO: Least Absolute Shrinkage and Selection Operator; PPV: Positive Predictive Value; NPV: Negative Predictive Value; AUC: Area Under the Curve; CI: Confidence Interval.

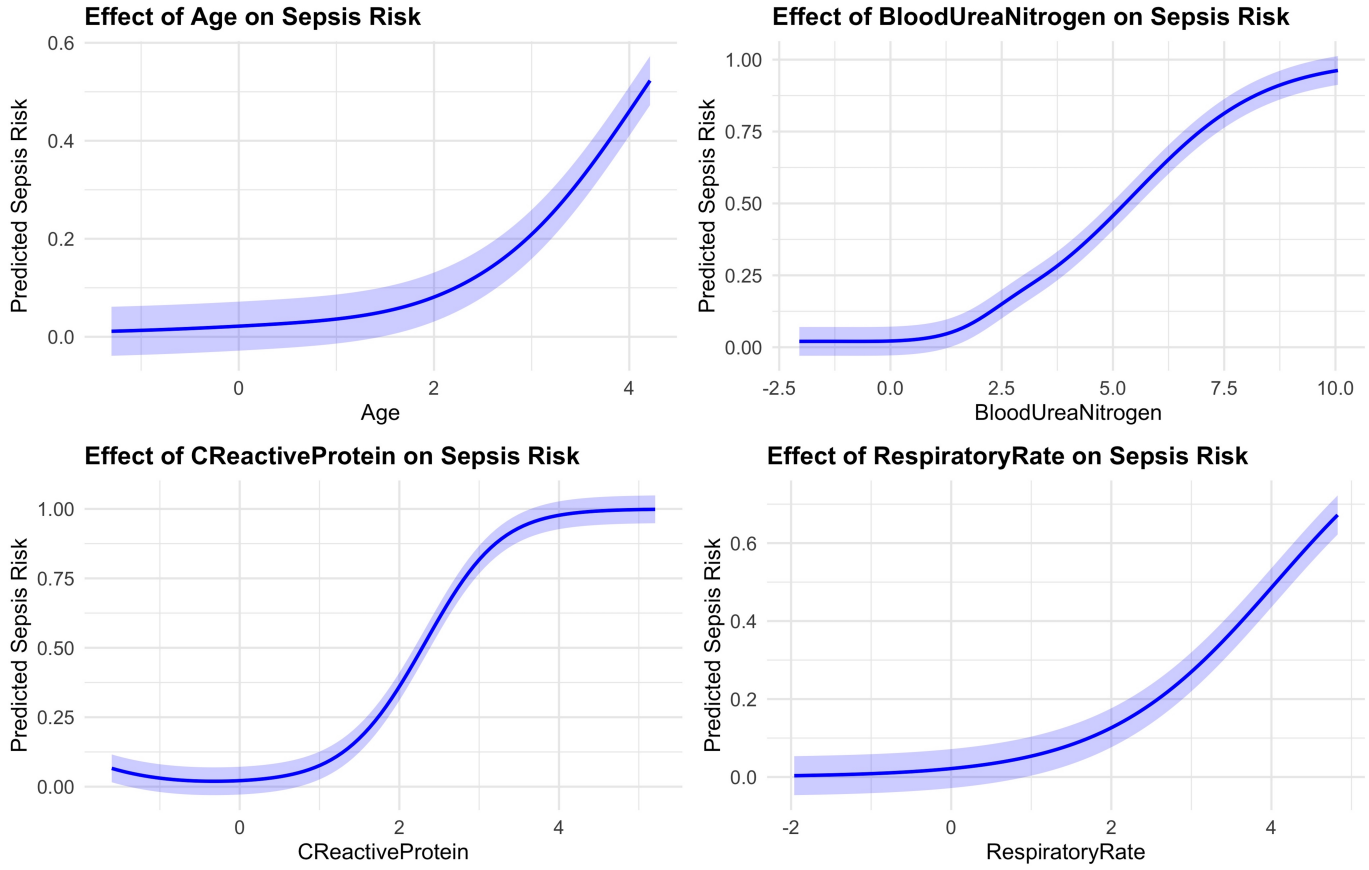

FIGURE 2. Predicted effects of age, blood urea nitrogen, C-reactive protein, and respiratory rate on sepsis risk. Partial effect plots from the Generalized Additive Model showing non-linear associations between predictors and sepsis probability. Threshold effects for age (>78 years), CRP (>70 mg/L) and BUN (>35 mg/dL) are apparent.

an adjusted R^2 of 67.8%, supporting the improved predictive performance of GAM over logistic regression.

LASSO regression achieved an AUC of 0.9421 (95% CI 0.9103–0.9738). The model retained the key predictors identified in logistic regression: age, BUN, CRP, systolic blood pressure, respiratory rate and temperature. The coefficient paths across different regularization strengths (Fig. 3) illustrate how predictor importance shifts as the penalty increases, reinforcing the robustness of the selected features. The decision tree model identified clear stratification thresholds, with CRP ≥70 mg/L, respiratory rate ≥22 breaths/min, and BUN ≥35 mg/dL emerging as the strongest predictors (Fig. 4). The structure of the decision tree provides an intuitive, rule-based approach to sepsis risk prediction. Each branching point represents a clinical cutoff, guiding risk assessment in a stepwise manner. For instance, patients with CRP levels above 70 mg/L were categorized as high risk, while those with lower CRP but elevated respiratory rates and BUN levels were also identified as at-risk subgroups. This visualization supports clinical interpretation without requiring complex calculations. The model had an AUC of 0.9152 (95% CI 0.8842–0.9462).

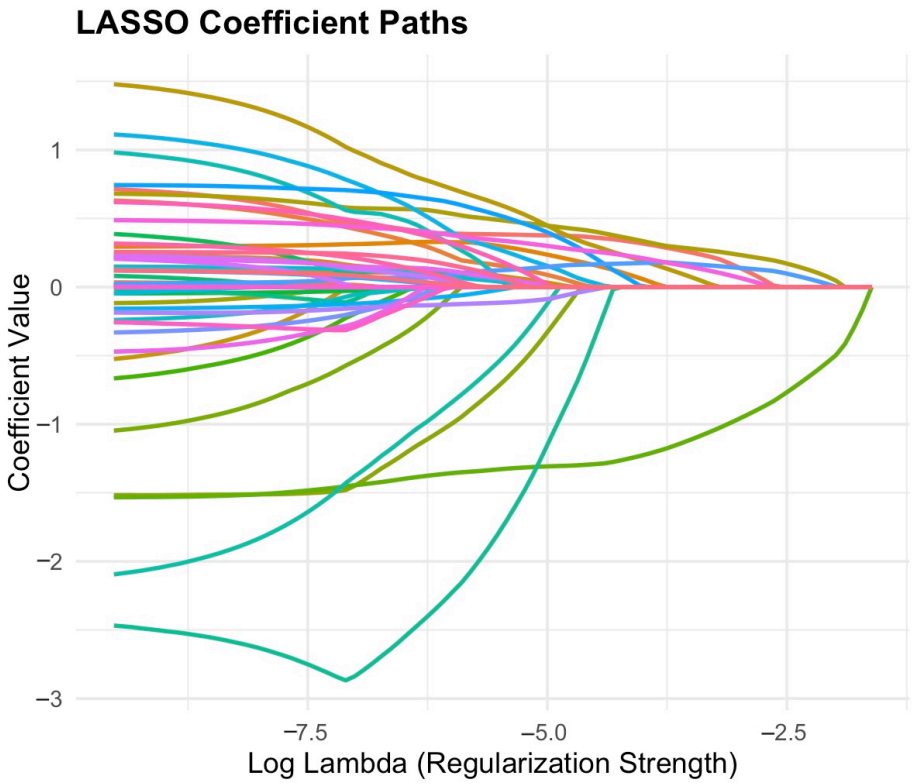

FIGURE 3. Coefficient paths in least absolute shrinkage and selection operator (LASSO) regression. Each colored line represents a predictor variable. As the regularization strength increases (log lambda moves right), less informative coefficients are progressively shrunk toward zero and excluded from the model. This visualization illustrates the variable selection process and helps identify which predictors remain robust across a range of penalization levels. LASSO: Least Absolute Shrinkage and Selection Operator.

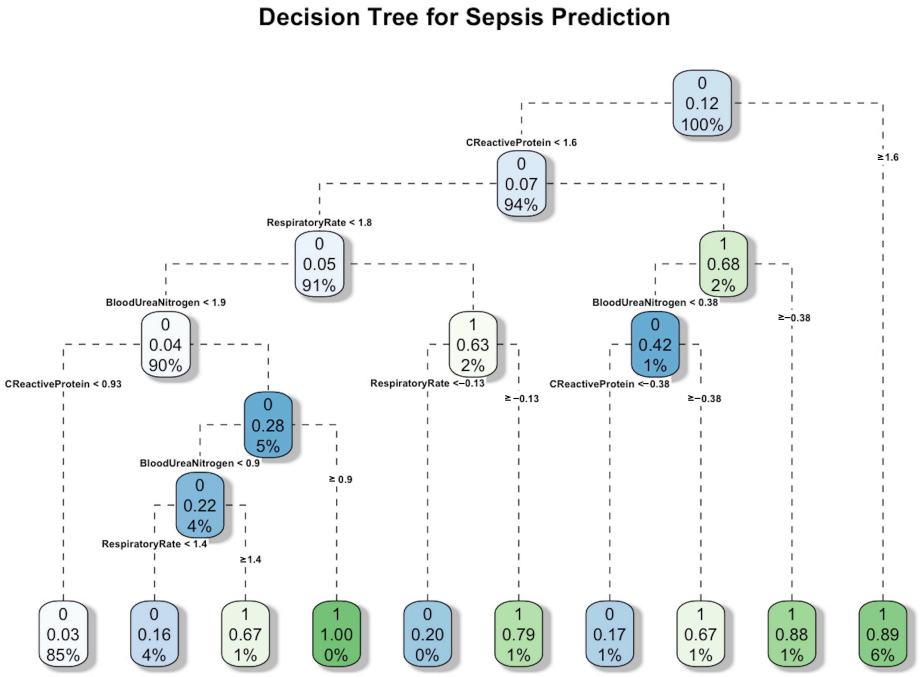

FIGURE 4. Decision tree model for sepsis prediction. Each node represents a decision point based on a clinical variable and threshold. The top value in each box is the predicted class (0 = no sepsis, 1 = sepsis). The middle value represents the predicted probability of sepsis for patients in that node, and the bottom percentage indicates the proportion of the total population falling into that category. Green boxes represent terminal nodes predicting sepsis (class 1), and blue boxes represent terminal nodes predicting no sepsis (class 0).

3.3 Machine learning performance

Comparative analysis using DeLong's test revealed statistically significant differences in AUC performance across the models. GAM significantly outperformed logistic regression ($p < 0.001$), as did LASSO ($p < 0.001$). Additionally, GAM demonstrated a statistically significant improvement over the decision tree model ($p = 0.046$). However, no significant difference was observed between LASSO and the decision tree ($p = 0.235$) or between GAM and LASSO ($p = 0.548$). These results suggest that GAM and LASSO are both superior to logistic regression, while GAM also holds a slight edge over the decision tree. For calibration assessment, Brier scores were computed, with LASSO exhibiting the best calibration (Brier Score = 0.0240), followed by GAM (0.0331), logistic regression (0.0367) and the decision tree (0.0379). While GAM had the highest AUC, LASSO demonstrated superior specificity (0.986 *vs.* 0.959), positive predictive value (PPV) (0.894 *vs.* 0.742), and F1 score (0.900 *vs.* 0.801), suggesting improved performance in identifying true positive cases with fewer false positives.

Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI) analyses further demonstrated the superior predictive value of GAM compared to logistic regression. The NRI for GAM versus logistic regression was 0.1397 (13.97%), indicating an improved classification of sepsis cases. The decision tree model, while useful, showed a modest improvement with an NRI of 0.0441 (4.41%). These findings suggest that GAM offers the most substantial enhancement in predictive accuracy for early sepsis detection in elderly patients with UTI.

4. Discussion

In this study, we identified several independent risk factors for early sepsis development in elderly patients who presented to the ED with UTIs. Notably, increased age, elevated levels of BUN, CRP, creatinine, heightened respiratory rate and body temperature, and reduced systolic blood pressure were found to be significant predictors of sepsis. However, traditional clinical risk stratification methods may be insufficient for timely identification, as ML-based approaches demonstrated superior predictive performance. The non-linearity captured by GAM allows for a more nuanced representation of risk factors, particularly for age-related acceleration in sepsis risk, while LASSO provides a streamlined yet effective selection of key predictors. These models hold potential for real-time application in ED settings, where rapid risk assessment is critical for optimizing early intervention.

Age demonstrated a dose-response relationship with sepsis risk. The GAM model identified an inflection point beyond 78 years, where the risk rose more sharply. This highlights that even within an elderly population ($\geq$65 years), sepsis risk continues to rise with each additional year, reinforcing the need for closer monitoring of older individuals. This threshold effect is consistent with prior research indicating that very elderly patients (>80 years) exhibit disproportionate vulnerability to sepsis-related mortality [15]. These results suggest that in clinical practice, simply labeling a patient as "elderly" is insufficient—older individuals within this group may have vastly different risk profiles.

Renal function markers were among the most consistently predictive variables across the models. Declining renal function compromises immune defenses, increasing susceptibility to severe infections and systemic inflammation [16]. Several studies have established a direct link between impaired renal function and an increased risk of infection-related morbidity and mortality [17, 18]. GAM revealed a threshold effect for BUN, with sepsis risk increasing disproportionately beyond 35 mg/dL. BUN and creatinine were also independently associated with early sepsis and showed consistent thresholds across ML models. Even under penalization, LASSO retained BUN and creatinine, reinforcing their predictive robustness. These thresholds may aid bedside decision-making and triage. While these findings may reflect known clinical risk factors, the added value of machine learning lies in its ability to refine prediction through nuanced modeling of their interactions and thresholds.

Inflammatory markers such as CRP and fever play a crucial role in sepsis risk assessment and severity stratification [18]. CRP, in particular, demonstrated strong discriminatory ability [19]. Liu *et al.* [15] identified 60 mg/L as a critical threshold for bacterial infections in elderly patients, while Zincircioğlu *et al.* [20] found higher CRP levels in septic versus non-septic elderly patients (88.3 *vs.* 68.8 mg/L). Notably, GAM analysis in our study revealed a threshold effect, with sepsis risk sharply increasing beyond 70 mg/L CRP. Fever's role in sepsis is less consistent, as 30–50% of elderly sepsis patients lack a febrile response [21, 22]. Shimazui *et al.* [23] found that while hypothermia was common, it was not associated with increased mortality. A systematic review of 42 studies reported higher mortality (47.3%) in hypothermic sepsis patients compared to 22.2% in febrile patients [24]. Our findings align with prior literature suggesting a non-linear relationship between fever and sepsis risk [25]. This variability highlights the need for comprehensive risk models incorporating multiple inflammatory markers rather than relying solely on fever for early sepsis detection.

One of the key challenges in applying artificial intelligence (AI) to clinical settings is the trade-off between interpretability and predictive accuracy. While GAM and LASSO demonstrated higher discrimination in sepsis prediction, Decision Tree models offer a more intuitive, threshold-based approach that may be more accessible to frontline clinicians. Although GAM had the highest AUC, LASSO excelled in specificity, positive predictive value, and F1 score. This characteristic may make LASSO more suitable in resource-limited environments, where false positives carry clinical or operational consequences. We acknowledge that stepwise regression, used in our logistic model, may risk overfitting. This concern was mitigated by consistent predictor retention across GAM and LASSO. These differences highlight that model selection should be guided not only by AUC but also by context-specific performance metrics. To facilitate comparison, our study included direct evaluation of all models across multiple statistical metrics, including AUC, Brier score and DeLong's test. Future research should focus on the integration of explainable AI techniques to enhance model transparency, ensuring that

high-performing models are both interpretable and clinically actionable. ML-based sepsis risk models could be integrated into ED workflows to trigger automated alerts and support early identification of high-risk patients. Integration with EHR and real-time monitoring systems could enhance early intervention strategies, reducing time to sepsis recognition and treatment initiation. However, external validation in multi-center cohorts is required before widespread implementation.

5. Limitations

This study has several limitations. As a retrospective single-center study, it was dependent on the accuracy and completeness of existing clinical records, which may introduce data inconsistencies or missingness and limit generalizability. While we excluded patients with incomplete data and no variable exceeded 5% missingness, some parameters (*e.g.*, SOFA components) were not universally measured at admission, which may have affected the capture of dynamic changes over 72 hours. For patients discharged early, the final documented SOFA score was used to assess sepsis progression, potentially underestimating post-discharge deterioration. However, most high-risk patients remained hospitalized. The study did not include infection etiology or presence of indwelling urinary catheters and obstructive uropathy as independent variables, although these are not always available in the ED setting. Finally, external validation and decision curve analysis were not performed. Although our sample met the $\geq$10 events-per-predictor rule of thumb, we recognize that more advanced approaches such as the pmsampsize methodology could enhance model development in future prospective studies [26].

6. Conclusions

Advanced age, elevated BUN, CRP, creatinine, increased respiratory rate and temperature, and low systolic BP were identified as independent predictors of early sepsis in elderly patients with UTI. While logistic regression provided useful risk stratification, ML models such as GAM and LASSO outperformed it by capturing threshold effects and complex patterns, highlighting their added value in early detection and clinical decision support. Future studies should focus on real-time integration of predictive models to optimize ED workflows and early intervention.

AVAILABILITY OF DATA AND MATERIALS

The dataset and statistical code used for model development and validation are available upon reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

İU—conceptualization, methodology, data collection, data curation, formal analysis, visualization, writing–original draft, and writing–review & editing. FY—conceptualization, visualization, writing–review & editing, supervision.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval for the study was obtained from Mardin Artuklu University Non-Interventional Clinical Research Ethics Committee (date: 05 November 2024; approval number: 2024/11-12). Due to the retrospective nature of the study, the ethics committee waived the requirement for informed consent.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found, in the online version, at https://oss.signavitae.com/mre-signavitae/article/1981287684935434240/attachment/Supplementary%20material.docx.

REFERENCES

[1] Chiscano-Camón L, Ruiz-Sanmartin A, Bajaña I, Bastidas J, Lopez-Martinez R, Franco-Jarava C, *et al.* Current perspectives in the management of sepsis and septic shock. Frontiers in Medicine. 2024; 11: 1431791.

[2] Kule A, Stassen W, Flores GE, Djarv T, Singletary E, Kule A, *et al.* Recognition and awareness of sepsis by first-aid providers in adults with suspected infection: a scoping review. Cureus. 2024; 16: e61612.

[3] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016; 315: 801–810.

[4] Ibarz M, Haas LEM, Ceccato A, Artigas A. The critically ill older patient with sepsis: a narrative review. Annals of Intensive Care. 2024; 14: 6.

[5] Wang X, Yang C, Jiang J, Hu Y, Hao Y, Dong JY. Polypharmacy, chronic kidney disease, and mortality among older adults: a prospective study of National Health and nutrition examination survey, 1999–2018. Frontiers in Public Health. 2023; 11: 1116583.

[6] Andaluz D, Ferrer R. SIRS, qSOFA, and organ failure for assessing sepsis at the emergency department. Journal of Thoracic Disease. 2017; 9: 1459–1462.

[7] Dutta C, Pasha K, Paul S, Abbas MS, Nassar ST, Tasha T, *et al.* Urinary tract infection induced delirium in elderly patients: a systematic review. Cureus. 2022; 14: e32321.

[8] Zhen J, Li J, Liao F, Zhang J, Liu C, Xie H, *et al.* Development and validation of machine learning models for young-onset colorectal cancer risk stratification. Npj Precision Oncology. 2024; 8: 1–14.

[9] Mittas N, Chatzopoulou F, Kyritsis KA, Papagiannopoulos CI, Theodoroula NF, Papazoglou AS, *et al.* A risk-stratification machine learning framework for the prediction of coronary artery disease severity: insights from the GESS trial. Frontiers in Cardiovascular Medicine. 2022; 8: 812182.

[10] van Doorn WPTM, Helmich F, van Dam PMEL, Jacobs LHJ, Stassen

PM, Bekers O, *et al.* Explainable machine learning models for rapid risk stratification in the emergency department: a multicenter study. The Journal of Applied Laboratory Medicine. 2024; 9: 212–222.

[11] Lawati HA, Blair BM, Larnard J. Urinary tract infections: core curriculum 2024. American Journal of Kidney Diseases. 2024; 83: 90–100.

[12] Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Medicine. 1996; 22: 707–710.

[13] Bujang MA, Sa'at N, Sidik TMITAB, Joo LC. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. The Malaysian Journal of Medical Sciences. 2018; 25: 122.

[14] He Y. Missing data analysis using multiple imputation: getting to the heart of the matter. Circulation Cardiovascular Quality and Outcomes. 2010; 3: 98–105.

[15] Liu A, Bui T, Van Nguyen H, Ong B, Shen Q, Kamalasena D. Serum C-reactive protein as a biomarker for early detection of bacterial infection in the older patient. Age and Ageing. 2010; 39: 559–565.

[16] Tang Y, Jiang J, Zhao Y, Du D. Aging and chronic kidney disease: epidemiology, therapy, management and the role of immunity. Clinical Kidney Journal. 2024; 17: sfae235.

[17] Francis A, Harhay MN, Ong ACM, Tummalapalli SL, Ortiz A, Fogo AB, *et al.* Chronic kidney disease and the global public health agenda: an international consensus. Nature Reviews Nephrology. 2024; 20: 473–485.

[18] Maringhini S, Zoccali C. Chronic kidney disease progression—a challenge. Biomedicines. 2024; 12: 2203.

[19] Jiang X, Zhang C, Pan Y, Cheng X, Zhang W. Effects of C-reactive protein trajectories of critically ill patients with sepsis on in-hospital mortality rate. Scientific Reports. 2023; 13: 15223.

[20] Zincircioğlu Ç, Rollas K, Güldoğan IK, Sarıtaş A, Özkarakaş H, Ersan G, *et al.* Diagnostic value of procalcitonin and C reactive protein for infection and sepsis inelderly patients. Turkish Journal of Medical Sciences. 2021; 51: 2649–2656.

[21] Nasa P, Juneja D, Singh O. Severe sepsis and septic shock in the elderly: an overview. World Journal of Critical Care Medicine. 2012; 1: 23.

[22] Rowe TA, McKoy JM. Sepsis in older adults. Infectious Disease Clinics. 2017; 31: 731–742.

[23] Shimazui T, Nakada T, Walley KR, Oshima T, Abe T, Ogura H, *et al.* Significance of body temperature in elderly patients with sepsis. Critical Care. 2020; 24: 387.

[24] Rumbus Z, Matics R, Hegyi P, Zsiboras C, Szabo I, Illes A, *et al.* Fever is associated with reduced, hypothermia with increased mortality in septic patients: a meta-analysis of clinical trials. PLOS ONE. 2017; 12: e0170152.

[25] Doman M, Thy M, Dessajan J, Dlela M, Do Rego H, Cariou E, *et al.* Temperature control in sepsis. Frontiers in Medicine. 2023; 10: 1292468.

[26] 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. The BMJ. 2020; 368: m441.

How to cite this article: İzzet Ustaalioğlu, Ferhat Yıldız. Early sepsis prediction in elderly patients with urinary tract infections: a machine learning. Signa Vitae. 2026; 22(1): 75-84. doi: 10.22514/sv.2025.161.