European Heart Journal - Digital Health (2026) **7**, ztaf150
https://doi.org/10.1093/ehjdh/ztaf150

ORIGINAL ARTICLE

From haemodynamics to kidney risk: AI-based early prediction validated in general and burn ICU populations

Louis Boutin [1,2,3,*], Fedi Kadri[4], Arij Chaftar[4], Benjamin Deniau [2,5], Sakura Minani[2,5], Stefanny M. Figueroa[3], Christos E. Chadjichristos[3], Anis Ghorbel[4], Alexandre Mebazaa [2,5], and François Dépret[2,5]

[1]Department of Anaesthesiology and Intensive Care, Université Paris Cité, Hôpital Européen Georges Pompidou, AP-HP, 20 rue Leblanc, Paris 75015, France; [2]INSERM, UMR 942, MASCOT: Cardiovascular Marker in Stress Condition, Université Paris Cité, Lariboisière Hospital, 43 bld de la chapelle, 75010 Paris, France; [3]INSERM, UMRS 1155, CORAKID, Sorbonne Université, Tenon Hospital, 4 rue de la chine, 75020 Paris, France; [4]Precisia Care SA, 8 route de la corniche, 1066, Lausanne, Switzerland; and [5]Department of Anaesthesiology, Critical Care Medicine and Burn Unit, AP-HP, Saint-Louis Hospital, DMU Parabol, FHU PROMICE, Université de Paris, Paris 75010, France

Received 27 August 2025; revised 12 November 2025; accepted 29 November 2025; online publish-ahead-of-print 21 January 2026

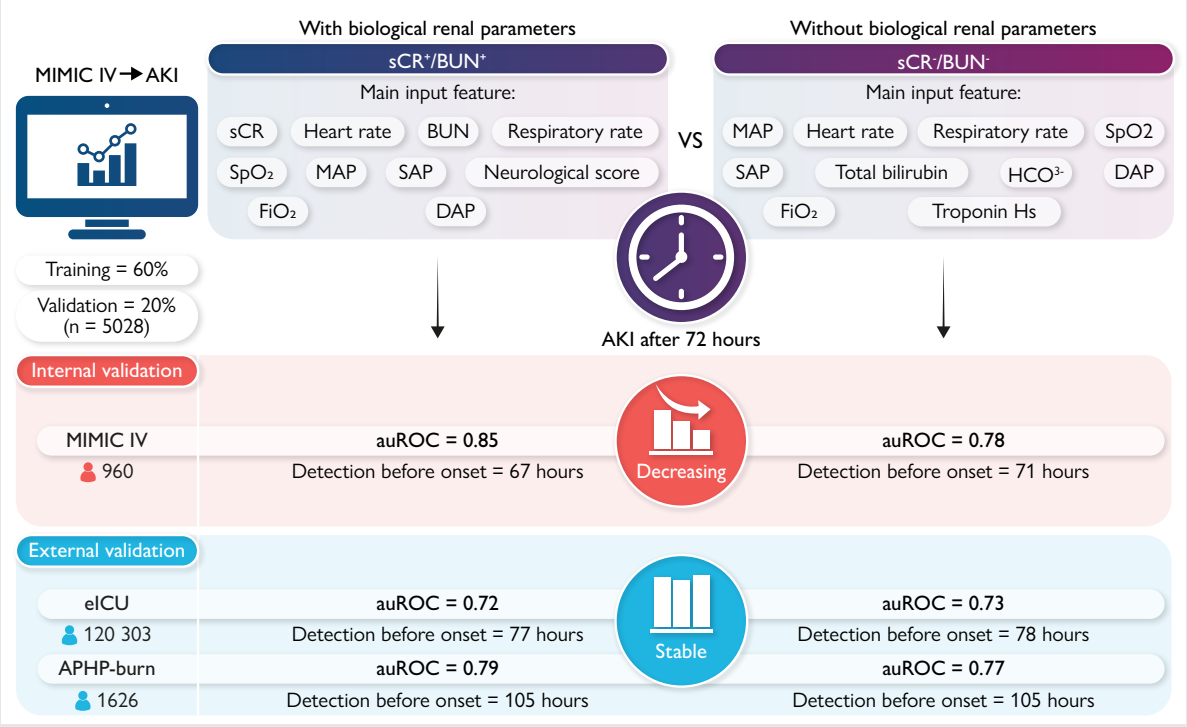
Aims Acute kidney injury (AKI) is a frequent and severe complication in critically ill patients with cardiovascular instability. Current risk scores rely on delayed renal biomarkers such as serum creatinine (sCr) and blood urea nitrogen (BUN). We aimed to develop and validate machine learning (ML) models predicting AKI and major adverse kidney events (MAKE) exclusively from systemic physiological and haemodynamic data.

Methods and results Two ML models were trained on the MIMIC-IV database: one including (sCr+/BUN+) and one excluding (sCr−/BUN−) renal parameters. External validation was performed in the eICU database and in a cohort of burn ICU patients from AP-HP. Model performance was assessed for early AKI and MAKE prediction up to 100 h before diagnosis. Systemic haemodynamic and physiological variables were the strongest predictors of AKI. In MIMIC-IV, the sCr−/BUN− model achieved auROC 0.78 at 72 h, approaching the sCr+/BUN+ model. In eICU, it outperformed the biomarker-based model at later time points (auROC 0.73). In the burn ICU cohort—representing a high-stress systemic environment—it maintained robust accuracy (auROC 0.75 at 24 h, 0.77 at 72 h). For MAKE prediction, the sCr−/BUN− model achieved auROC 0.87 (burn cohort), 0.67 (eICU), and 0.77 (MIMIC-IV). Median lead time for AKI prediction exceeded 70 h.

Conclusion AI models based solely on non-renal parameters can accurately predict AKI and MAKE, even under extreme systemic stress such as severe burns. Haemodynamic signatures carry sufficient information to anticipate kidney dysfunction well in advance, opening the way to real-time, proactive cardio-renal risk stratification in ICU patients with acute heart failure, cardiogenic shock, and after cardiac surgery.

* Corresponding author. Tel: +33156093515, Email: louis.boutin@aphp.fr

Graphical Abstract



'Decreasing' indicates that the performance advantage of the biomarker-based (sCr+/BUN+) model diminished over time, whereas 'Stable' refers to comparable or superior performance of the haemodynamics-only (sCr−/BUN−) model in the external validation cohorts

Keywords Acute kidney injury • Major adverse kidney event • Prediction • Serum creatinine

Take Home Message

AI models based solely on systemic haemodynamic and physiological variables can anticipate AKI more than 70 h before diagnosis by renal biomarkers. This performance is comparable to renal biomarker–based models, even under extreme stress conditions such as severe burns. Haemodynamic signatures thus offer a complementary and earlier window for proactive cardio-renal risk management.

Clinical Perspective

What is new?

Machine learning models relying on systemic haemodynamic and physiological parameters (without renal biomarkers) accurately predict acute kidney injury (AKI) and major adverse kidney events (MAKE). These models anticipate AKI up to 70–100 h before clinical diagnosis, even in high-stress ICU settings such as severe burns. Haemodynamic signatures act as early, reliable indicators of impending renal dysfunction.

What are the clinical implications?

Traditional renal biomarkers (serum creatinine, BUN) remain important but are delayed indicators of injury. Haemodynamic monitoring, already standard in critically ill and cardiac patients, can be leveraged by AI to provide earlier and actionable kidney risk prediction. This complementary paradigm may enable earlier intervention in cardio-renal syndromes, with relevance for acute heart failure, cardiogenic shock, and post–cardiac surgery patients.

Introduction

Acute Kidney Injury (AKI) is a frequent and serious condition associated with morbidity and mortality.[1,2] According to the Kidney Disease: Improving Global Outcomes (KDIGO) criteria, AKI is defined by an abrupt decline in renal function, detected by elevations in serum creatinine (sCr) and/or reductions in urine output (UO).[3] Although this definition is well-established, AKI is often detected too late in the clinical course, limiting opportunities for effective intervention and contributing to poor patient outcomes.[4]

Importantly, AKI usually arises as a secondary manifestation of systemic insults—such as haemodynamic instability, inflammation, sepsis, or hypoxia—that precede measurable renal dysfunction.[5] These upstream mechanisms alter cardiovascular and metabolic homeostasis and can generate detectable changes in systemic physiology and haemodynamics hours to days before renal parameters deteriorate.[6,7] This observation highlights the potential of non-renal indicators to serve as early signatures of kidney risk, enabling clinicians to anticipate injury rather than react to its biochemical confirmation.

Efforts to improve early detection have included the development of novel urinary and plasma biomarkers capable of identifying subclinical

AKI.[5,8] While promising, their use is constrained by assay cost, limited availability, and the need for intermittent sampling.[9] In contrast, modern intensive care units already generate a continuous stream of haemodynamic and physiological data—from heart rate and blood pressure to ventilation and oxygenation parameters—that reflect the systemic stresses leading to renal injury. These routinely collected data offer a unique, low-cost substrate for predictive modelling.

Artificial intelligence (AI) and machine learning (ML) methods are particularly well-suited to capture the complex, non-linear interactions between systemic physiology and subsequent kidney injury.[10] Yet most existing prediction tools remain anchored in traditional renal markers such as sCr and blood urea nitrogen (BUN), which inherently constrain their predictive horizon.[11] By excluding delayed renal biomarkers and focusing solely on systemic haemodynamic and physiological parameters, AI-based models may provide a longer lead time for AKI prediction, creating an actionable window for preventive interventions.[12]

Here, we demonstrate that AI models trained exclusively on systemic haemodynamic and physiological data can accurately predict AKI and major adverse kidney events (MAKE) up to 70–100 h before diagnosis. Their performance was validated in general and burn ICU populations, showing that haemodynamic signatures alone carry sufficient information to anticipate kidney dysfunction and support earlier, actionable cardio-renal risk stratification.

Method

Study design

Data were collected from three distinct cohorts to ensure population heterogeneity: the MIMIC-IV database,[13] the eICU Collaborative Research Database,[14] and a French Burn ICU cohort (APHP-burn). The MIMIC-IV database is a publicly available single-centre ICU dataset from an academic medical centre in the USA. The eICU Collaborative Research Database comprises multicentre ICU data from a diverse network of US hospitals, enhancing generalizability. The APHP-burn cohort originates from the POOF (Perio-Operative Organ Failure) study conducted at Saint-Louis Hospital (AP-HP, Paris), representing a high-acuity, domain-specific population of critically ill burn patients.

Inclusion

Patients were eligible for inclusion if they met the following criteria: age ≥18 years and an ICU stay of more than 24 h. Patients who were re-admitted to the ICU during the study period were also included. Nevertheless, patients with chronic kidney disease (CKD) or with AKI before the first 24 h following admission were not included.

Ethical approval

This study adhered to the principles of the Declaration of Helsinki. All datasets were fully de-identified prior to analysis and classified as non-human subject research; therefore, IRB approval and informed consent were not required. The MIMIC-IV database was approved by the Massachusetts Institute of Technology (IRB No. 0403000206) and the Beth Israel Deaconess Medical Center (Protocol No. 2001-P-001699/14).[13] For the eICU Collaborative Research Database, IRB approval was waived due to its retrospective nature and HIPAA-compliant de-identification certified by Privacert (Certification No. 1031219-2).[14] The APHP-burn cohort was approved by the Ethics Committee of the Société Française d'Anesthésie et de Réanimation (CERAR, SFAR; IRB 00010254-2025-091).

Definition of acute kidney injury

AKI was defined according to the KDIGO criteria.[3] For patients with an estimated glomerular filtration rate (eGFR) < 75 mL/min/1.73 m^2 at ICU admission (calculated using the CKD-EPI equation), baseline sCr was back-estimated using the MDRD formula, assuming a reference eGFR of 75 mL/min/1.73 m^2. For outcome analysis, AKI was classified based on the component of the KDIGO definition: AKI detected by sCr variation (AKI-sCr) and AKI detected by urine output (AKI-UO). Unless stated otherwise, the term AKI refers to the occurrence of either or both components.

Definition of major adverse kidney events

The MAKE at ICU discharge was defined as a composite outcome including: (1) ICU mortality, (2) need for renal replacement therapy (RRT), and (3) non-recovery of kidney function, defined as a sCr at discharge ≥150% of baseline sCr. MAKE was assessed only in patients with an ICU length of stay >7 days, to ensure sufficient time for renal outcome evaluation.

Prediction task and input features

The occurrence of AKI was modelled as a binary classification task, with patients labelled positive if they developed AKI during their ICU stay and negative otherwise. The model produced hourly updated risk scores representing the probability of AKI onset throughout the ICU stay.

Input features included routinely collected clinical, physiological, biochemical, and treatment-related variables. Feature selection was based on impurity-based importance, where variables were iteratively ranked according to their contribution to reducing classification uncertainty. A minimal, clinically relevant set of features was selected by optimizing the area under the receiver operating characteristic curve (auROC).

For characteristic tables, in the eICU cohort, comorbidities were extracted from ICD-9 diagnostic codes available prior to ICU admission. The detailed ICD-9 lists used for comorbidity identification in both eICU and MIMIC-IV datasets are now provided in the Supplementary material online, *Table S1*. Notably, comorbidity variables were not used as model inputs, which relied solely on physiological, haemodynamic, and biochemical data.

Data preprocessing

Comprehensive preprocessing was performed to ensure data quality and model reliability. Input variables were standardized using *Z*-score normalization.

Missing values were not imputed manually, as the XGBoost algorithm natively handles missing data by learning optimal split directions during training. This approach avoids introducing bias and preserves the structure of clinical variability. Outlier handling was performed using predefined physiological ranges established in collaboration with medical experts; values outside these clinically plausible limits were excluded to ensure data quality and consistency.

Model development and validation

Before model training, selected physiological and biochemical variables were transformed using proprietary mathematical functions developed by Precisia Care SA. These transformations were designed to enhance the model's ability to capture non-linear interactions and temporal dynamics between systemic parameters. Although the exact functions cannot be disclosed for intellectual-property reasons, all transformations were applied uniformly across datasets to ensure methodological consistency and preserve clinical interpretability of the resulting features. In this line, detailed disclosure of internal transformations and SHAP analyses cannot be fully detailed. To ensure transparency, we report impurity-based feature importance, highlighting the key haemodynamic and physiological variables driving AKI prediction.

Hyperparameter tuning was conducted using a Bayesian optimization approach coupled with 4-fold cross-validation. During each iteration, three folds were used for training and one for validation. The

search space included learning rate, maximum tree depth, number of estimators, and subsampling parameters. The combination achieving the highest mean area under the ROC curve across folds was selected for the final model to maximize robustness and prevent overfitting.

The model was developed using the MIMIC-IV dataset. A total of 5028 patients were randomly selected, with data split into training (60%) and validation (20%) sets. The model generated hourly AKI risk scores during the ICU stay. Performance metrics included auROC, sensitivity (Se), specificity (Sp), and positive and negative likelihood ratios. Two versions of the model were evaluated: one including sCr and BUN (sCr+/BUN+), and one excluding them (sCr−/BUN−). A prediction was classified as positive if it exceeded the threshold maximizing the Youden index (sensitivity + specificity − 1). Model architecture and workflow are illustrated in Supplementary material online, *Figure S1*.

Outcomes

The primary outcome was the prediction of AKI. To ensure the availability of a minimal set of patient data following admission, prediction performance was specifically assessed in subpopulations of patients who developed AKI within 24, 48, or 72 h after admission.

Secondary outcomes included the prediction of MAKE at ICU discharge, AKI-sCr, AKI-UO, and the time interval between model prediction and AKI onset. These outcomes are detailed in Supplementary material online, *Figure S2*.

Statistical analysis

Continuous variables were summarized as medians with interquartile ranges (IQR), and categorical variables were reported as absolute counts and percentages. The model's predictive performance was assessed by calculating the auROC and determining Se, Sp, and likelihood ratios. These metrics were used to evaluate the ability of the ML algorithm to predict AKI accurately and reliably in ICU patients. Statistical analyses and model development were performed using Python (version 3.9), leveraging libraries such as scikit-learn, pandas, and numpy for data processing, ML, and evaluation metrics.

Results

Datasets, models, and features

Two models were developed: one including renal biomarkers (sCr+/BUN+) and one excluding them (sCr−/BUN−). In both settings, the most influential predictors were systemic physiological and haemodynamic parameters rather than renal biomarkers. In the sCr+/BUN+ model, although sCr and BUN were among the top-ranked variables, heart rate, respiratory rate, oxygen saturation, mean arterial pressure (MAP), systolic (SAP), and diastolic arterial pressure (DAP), Glasgow Coma Scale, and inspired oxygen fraction also emerged as key drivers. Strikingly, in the sCr−/BUN− model, the leading predictors remained exclusively haemodynamic and physiological variables—MAP, SAP, DAP, heart rate, respiratory rate, oxygen saturation—complemented by systemic markers such as total bilirubin, bicarbonate, inspired oxygen fraction, and troponin. Feature importance rankings (see Supplementary material online, *Figure S3*) highlight that haemodynamic signatures consistently dominated model performance, underscoring their central role as early indicators of AKI risk.

Within each dataset, three cohorts were defined based on the timing of AKI onset (no AKI within the first 24, 48, or 72 h after ICU admission). The study flow and cohort definition are illustrated in Supplementary material online, *Figure S4*. Baseline characteristics, including demographics, admission diagnoses, severity scores, and ICU outcomes, are summarized in *Table 1* for each dataset.

Acute kidney injury prediction performance with model sCr+/BUN+

In all cohorts, prediction accuracy improved with longer horizons. Importantly, while sCr and BUN contributed when included, systemic haemodynamic and physiological features already enabled robust prediction in the absence of renal biomarkers. In MIMIC-IV, the auROC increased from 0.84 (95%CI, 0.80–0.89) at 24 h to 0.85 (5%CI, 0.80–0.90) at 72 h. The correlation between predicted risk and sCr was: 0.45 at 24 h ($P < 0.001$), 0.44 at 48 h ($P < 0.001$), and 0.43 at 72 h ($P < 0.001$).

In eICU, auROC rose from 0.64 (95%CI, 0.64–0.65) at 24 h to 0.72 (95%CI, 0.71–0.73) at 72 h, with stable correlations to sCr: 0.26 at 24 h ($P < 0.001$) and 48 h ($P < 0.001$), and 0.25 at 72 h ($P < 0.001$).

In the APHP-burn cohort, auROC increased from 0.75 (95%CI, 0.72–0.78) to 0.79 (95%CI, 0.75–0.82) between 24 and 72 h. Correlation with sCr was 0.27 at 24 h ($P < 0.001$) to 0.16 at 72 h ($P < 0.001$).

Additional metrics and visualizations are presented in *Table 2* and Supplementary material online, *Figures S5* and *S6*.

Acute kidney injury prediction performance of model sCr−/BUN−

In the MIMIC-IV cohort, the sCr−/BUN− model yielded auROCs of 0.74 (95%CI, 0.68–0.79), 0.75 (95%CI, 0.69–0.80), and 0.78 (95%CI, 0.71–0.85) at 24, 48, and 72 h, respectively. These were significantly lower than those of the sCr+/BUN+ model at all time points ($P < 0.001$, $P < 0.001$, and $P = 0.001$), with NRIs favouring the sCr+/BUN+ model: 0.22 (95%CI, 0.12–0.32; $P < 0.001$), 0.19 (95%CI, 0.09–0.29; $P < 0.001$), and 0.14 (95%CI, 0.01–0.27; $P = 0.03$). In the eICU cohort, auROCs for the sCr−/BUN− model were 0.63 (95%CI, 0.62–0.63), 0.70 (95%CI, 0.70–0.71), and 0.73 (95%CI, 0.72–0.74) at 24, 48, and 72 h, respectively. At 24 h, the sCr+/BUN+ model performed significantly better [$P < 0.001$; NRI = 0.02 (95%CI, 0.01–0.03); $P < 0.001$]. However, at 48 and 72 h, performance significantly favoured the sCr−/BUN− model ($P < 0.001$ and $P = 0.001$), with negative NRIs of −0.02 (95%CI, −0.03 to −0.01; $P = 0.003$) and −0.02 (95%CI, −0.03 to −0.001; $P = 0.03$), respectively. In the APHP-burn cohort, the sCr−/BUN− model achieved auROCs of 0.75 (95%CI, 0.72–0.78), 0.78 (95%CI, 0.75–0.81), and 0.77 (95%CI, 0.71–0.73) at 24, 48, and 72 h. Notably, in this high-stress environment, adding sCr and BUN did not yield any performance benefit, underscoring the sufficiency of systemic signatures under extreme physiological stress (*Figure 1* and Supplementary material online, *Table S2*).

All additional performance metrics are provided in *Table 2* and Supplementary material online, *Figures S5* and *S6*.

Renal prognosis prediction performance

MAKE prediction was limited to patients with an ICU stay >7 days and a prediction window ≥4 days before event onset to ensure prognostic relevance. In the MIMIC-IV cohort ($n = 104$; 17.3% with MAKE), prediction performance was comparable between models: auROC 0.73 (95% CI, 0.61–0.84) for the sCr+/BUN+ model vs. 0.77 (95%CI, 0.65–0.89) for the sCr−/BUN− model (P = ns). No significant reclassification improvement was observed [NRI = −0.05 (95%CI, −0.18 to 0.07); $P = 0.41$]. In the eICU cohort ($n = 13\,536$; 23.6% with MAKE), the sCr+/BUN+ model significantly outperformed the sCr−/BUN− model [auROC 0.76 (95%CI, 0.75–0.77) vs. 0.67 (95%CI, 0.65–0.68); $P < 0.001$], with improved reclassification [NRI = 0.15 (95%CI, 0.13–0.17); $P < 0.001$]. In the APHP-burn cohort, 72 patients (5.8%) experienced MAKE. Both models had similar discrimination [auROC 0.88 (95%CI, 0.84–0.92) vs. 0.87 (95%CI, 0.83–0.90); $P = 0.176$], but reclassification significantly favoured the sCr+/BUN+ model [NRI = 0.07

Table 1 Patients characteristics in cohorts

Variable	APHP-burn (*n* = 1626)	eICU (*n* = 120 303)	MIMIC IV (*n* = 960)
General			
Age, years, median, [IQR]	50.6 [34.4, 64.9]	65.0 [53.0–76.0]	60.0 [48.0–72.0]
Sex, male, *n* (%)	965 (59.3%)	65 066 (54.1%)	488 (50.8%)
IMC, median, [IQR]	24.8 [22.2, 28.7]	27.5 [23.5–32.9]	27.3 [26.5–35.5]
Comorbidities			
Charlson score, median, [IQR]	1.0 [0.0, 3.0]	3.0 [2.0–5.0]	4.0 [2.0–6.0]
Cardiovascular comorbidities [*n* (%)]	400 (24.6%)	45 961 (38.2%)	336 (35.0%)
Neurologic comorbidities [*n* (%)]	39 (2.4%)	11 913 (9.9%)	214 (22.3%)
Admission			
Sepsis [*n* (%)]	NC	16 709 (13.9%)	NC
Trauma [*n* (%)]	NC	4751 (3.9%)	NC
Other			
Heart rate (beats/minute)	90.5 [78.0, 104.0]	87.0 [74.0–103.0]	87.0 [75.0–103.0]
MAP (mmHg)	84.0 [74.0, 97.0]	83.0 [71.0–96.0]	84.0 [73.0–97.0]
Respiratory rate (beats/minute)	19.0 [15.0–23.0]	19.0 [16.0–23.0]	18.0 [15.0–22.0]
Body temperature (C)	36.9 [36.4–37.4]	36.7 [36.4–37.1]	36.8 [36.5–37.2]
SpO2 (%)	98.0 [96.0–100.0]	98.0 [96.0–100.0]	97.0 [94.5–98.0]
Haemoglobin (g/dL)	13.1 [11.2–14.9]	11.0 [9.3–12.6]	10.2 [8.8–11.6]
Platelets (K/µL)	255.0 [192.5–334.5]	192.0 [142.0–251.0]	205.0 [151.0–269.2]
WBC (K/µL)	11.2 [8.3–16.0]	10.7 [7.7–14.8]	10.4 [7.5–14.5]
BUN (mg/dL)	4.8 [3.6–6.4]	19.0 [13.0–31.0]	14.0 [10.0–20.0]
Serum creatinine (mg/dL)	0.8 [0.6–1.0]	1.0 [0.7–1.4]	0.7 [0.6–0.9]
Serum glucose (mg/dL)	115.3 [99.1–147.7]	128.0 [104.0–167.0]	123.0 [103.0–151.0]
Serum chloride (mEq/L)	102.0 [99.0–105.0]	105.0 [101.0–108.0]	105.0 [101.0–108.0]
Bicarbonate (mmol/L)	22.0 [20.0–24.0]	24.0 [21.0–27.0]	23.0 [21.0–26.0]
Serum potassium (mEq/L)	4.0 [3.7–4.3]	4.0 [3.7–4.4]	4.0 [3.7–4.4]
Serum sodium (mEq/L)	138.0 [136.0–140.6]	139.0 [136.0–141.0]	139.0 [136.0–141.0]
eGFR (mL/min/1.73 m^2)	104.6 [85.6–119.3]	76.9 [45.6–99.5]	97.3 [79.4–110.5]
Severity score			
IGSII	18.0 [11.0, 29.2]	52.0 [43.0–61.0]	29.0 [22.0–37.0]
SOFA	4.0 [1.0, 5.0]	5.0 [0.0–9.0]	3.0 [2.0–5.0]
Outcomes			
Length of stay (hours)	351.5 [171.3, 731.1]	57.0 [39.0–98.0]	177.6 [140.0–257.1]
Mortality [*n* (%)]	247 (15.2%)	10 738 (8.9%)	34 (3.5%)
Renal replacement therapy [*n* (%)]	89 (5.5%)	5231 (4.3%)	7 (0.7%)
Vasopressors therapy [*n* (%)]	292 (18.0%)	17 069 (14.2%)	241 (25.1%)
AKI after 24 h	642 (39.5%)	18 610 (15.4%)	93 (9.7%)
AKI after 48 h	387 (23.8%)	6290 (5.2%)	84 (8.8%)
AKI after 72 h	257 (15.8%)	3210 (2.7%)	44 (4.7%)
MAKE	72 (4.4%)	3190 (2.6%)	19 (2.0%)

(95%CI, 0.004–0.14); $P = 0.037$], suggesting that systemic physiology carries prognostic value beyond short-term AKI prediction (*Figure 2*; *Table 2*; Supplementary material online, *Table S2*).

Models prediction according to KDIGO physiological parameter

In the MIMIC-IV cohort, both models achieved their highest performance for AKI-UO, with the sCr+/BUN+ model significantly outperforming the sCr−/BUN− model [auROC 0.91 (95%CI, 0.88–0.94) vs. 0.83 (95%CI, 0.79–0.87) at 24 h; $P < 0.001$, and 0.91 (95%CI, 0.88–0.94) vs. 0.83 (95%CI, 0.78–0.88) at 72 h; $P = 0.001$]. In the eICU cohort, the best discrimination was for AKI-sCr using the sCr+/BUN+ model, with auROC increasing from 0.64 [95%CI, 0.63–0.64] at 24 h to 0.77 [95%CI, 0.76–0.77] at 72 h. In the APHP-burn cohort, both models performed well for AKI-sCr: AUROCs for sCr+/BUN+ rose from 0.85 [95%CI, 0.82–0.88] to 0.88 [95%CI, 0.85–0.91], and for sCr−/BUN− from 0.86 [95%CI, 0.83–0.88] to 0.87 [95%CI, 0.84–0.90]. For AKI-UO, predictive performance was lower and similar between models across all time points (see Supplementary material online, *Figures S7–S9*; Supplementary material online, *Tables S3* and *S4*). These findings suggest that haemodynamic signatures are

Table 2 Predictive performance value and performance parameters value according to cohorts and models

	auROC	95%CI	Sensitivity	Specificity	PPV	NPV	Accuracy
Models with sCr and BUN							
MIMIC IV							
24 h	0.842	[0.8, 0.885]	0.806	0.775	0.288	0.973	0.778
48 h	0.838	[0.79, 0.881]	0.81	0.775	0.269	0.975	0.778
72 h	0.85	[0.795, 0.9]	0.818	0.775	0.163	0.988	0.777
MAKE	0.728	[0.607, 0.841]	0.579	0.8	0.393	0.895	0.76
eICU cohort							
24 h	0.641	[0.636, 0.645]	0.576	0.626	0.369	0.795	0.612
48 h	0.687	[0.68, 0.693]	0.713	0.554	0.17	0.938	0.572
72 h	0.719	[0.71, 0.727]	0.699	0.623	0.108	0.969	0.627
MAKE	0.758	[0.749, 0.767]	0.734	0.655	0.396	0.889	0.674
APH-burn cohort							
24 h	0.753	[0.721, 0.781]	0.611	0.787	0.794	0.601	0.686
48 h	0.79	[0.76, 0.818]	0.661	0.781	0.715	0.742	0.731
72 h	0.79	[0.753, 0.823]	0.642	0.808	0.642	0.808	0.75
MAKE	0.879	[0.838, 0.918]	0.792	0.834	0.228	0.985	0.832
Models without sCr and BUN							
MIMIC IV							
24 h	0.737	[0.68, 0.79]	0,57	0.787	0.232	0.942	0.765
48 h	0.752	[0.691, 0.803]	0.607	0.787	0.226	0.951	0.77
72 h	0.781	[0.706, 0.846]	0.705	0.784	0.13	0.979	0.746
MAKE	0.767	[0.648, 0.885]	0.632	0.8	0.414	0.907	0.769
eICU cohort							
24 h	0.627	[0.622, 0.632]	0.649	0.53	0.344	0.799	0.563
48 h	0.699	[0.692, 0.706]	0.643	0.642	0.187	0.933	0.642
72 h	0.731	[0.722, 0.739]	0.69	0.648	0.114	0.97	0.651
MAKE	0.665	[0.654, 0.676]	0.632	0.609	0.333	0.843	0.615
APH-burn cohort							
24 h	0.751	[0.72, 0.779]	0.556	0.835	0.819	0.584	0.675
48 h	0.778	[0.746, 0.807]	0.607	0.835	0.748	0.725	0.733
72 h	0.769	[0.73, 0.805]	0.704	0.72	0.575	0.819	0.715
MAKE	0.866	[0.825, 0.904]	0.792	0.762	0.171	0.983	0.764

particularly informative for creatinine-defined AKI in high-stress contexts, but less sensitive for urine-output–driven definitions.

Timing of models for prediction

The time interval between model prediction and actual onset of AKI was assessed across the different time-point cohorts (24, 48, 72 h). Across all datasets, both models anticipated AKI well in advance, with median lead times exceeding 70 h. Importantly, the model using haemodynamic and physiological parameters alone (without renal biomarkers) maintained similar or even longer lead times compared with the biomarker-based model, reinforcing its clinical utility as an early warning tool.

In the MIMIC-IV cohort, the lead time from prediction to AKI onset increased with earlier prediction windows. Notably, the sCr−/BUN− model demonstrated longer lead time with a median time in hours of 45 [IQR: 25–74] (24 h) to 71 [IQR: 53–111] (72 h) compared to the sCr+/BUN+ model with a median time of 36 [IQR: 26–64] (24 h) to 67 [IQR: 52–117] (72 h) (*Figure 3*, Supplementary material online, *Table S5*).

In the eICU cohort, median time increased from 15 h [IQR: 7–35] (24 h) to 78 h [IQR: 55–126] (72 h) for the sCr+/BUN+ model, and from 15 h [IQR: 7–37] (24 h) to 78 h [IQR: 56–127] (72 h) for the sCr−/BUN− model. No significant differences in lead times were observed between the two models at any prediction timepoint (*Figure 3*, Supplementary material online, *Table S5*).

In the APHP-burn study at the 24-h prediction timepoint, the median lead time was 31 h [IQR: 12–87] for the sCr+/BUN+ model and 31.81 h [IQR: 13–79] for the sCr−/BUN− model. At the 72 h prediction timepoint, the lead time increased to 105 h [IQR: 57–181] and 105 h [IQR: 59–171], respectively (*Figure 3*, Supplementary material online, *Table S5*).

To evaluate prediction performance over time before AKI onset, a sensitivity analysis was conducted at 12, 24, and 36 h prior to diagnosis. In the MIMIC-IV cohort, auROCs decreased slightly with increasing prediction horizon: from 0.84 (95%CI, 0.78–0.89) to 0.82 (95%CI, 0.75–0.88) for the sCr+/BUN+ model, and from 0.77 (95%CI, 0.69–0.83) to 0.74 (95%CI, 0.67–0.81) for the sCr−/BUN− model. In the eICU cohort, performance also declined: sCr+/BUN+ auROC dropped from 0.71 (95%CI, 0.70–0.72) to 0.68 (95%CI, 0.67–0.69), and sCr−/BUN− from 0.72 (95%CI, 0.71–0.73) to 0.70 (95%CI, 0.69–0.71). In the APHP-burn cohort, the sCr+/BUN+ model decreased from 0.78 (95%CI, 0.74–0.81) to 0.75 (95%CI, 0.71–0.79), and the sCr−/BUN−

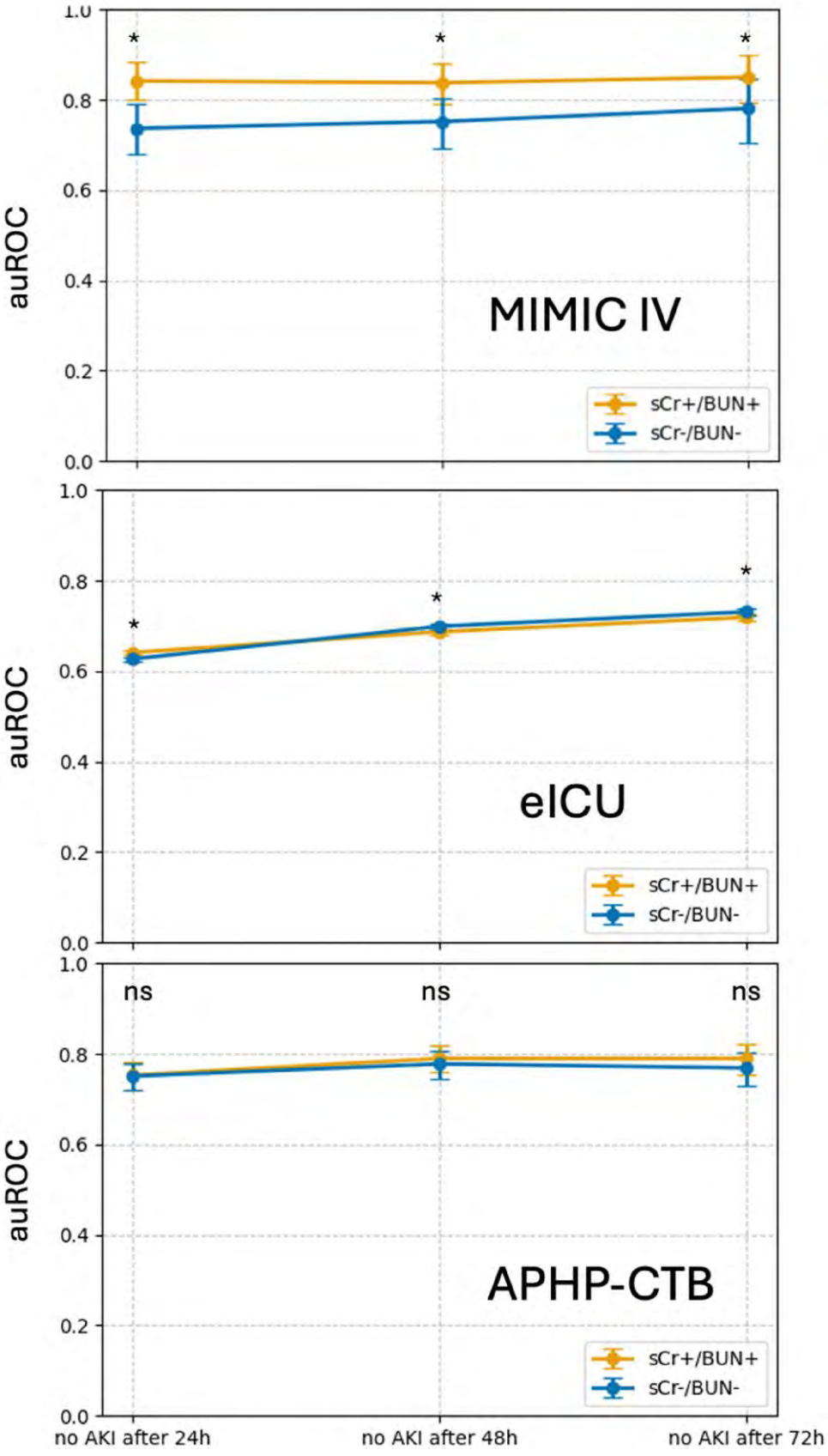


Figure 1 Prediction performance according to cohorts and dataset using sCr+/BUN+ and sCr−/BUN− model. auROC, area under the curve of the receiver operating curve; AKI, acute kidney injury. Tests were assessed for sCr+/BUN+ vs. sCr−/BUN−: ns, non-significant. * = *P*-value <0.001.

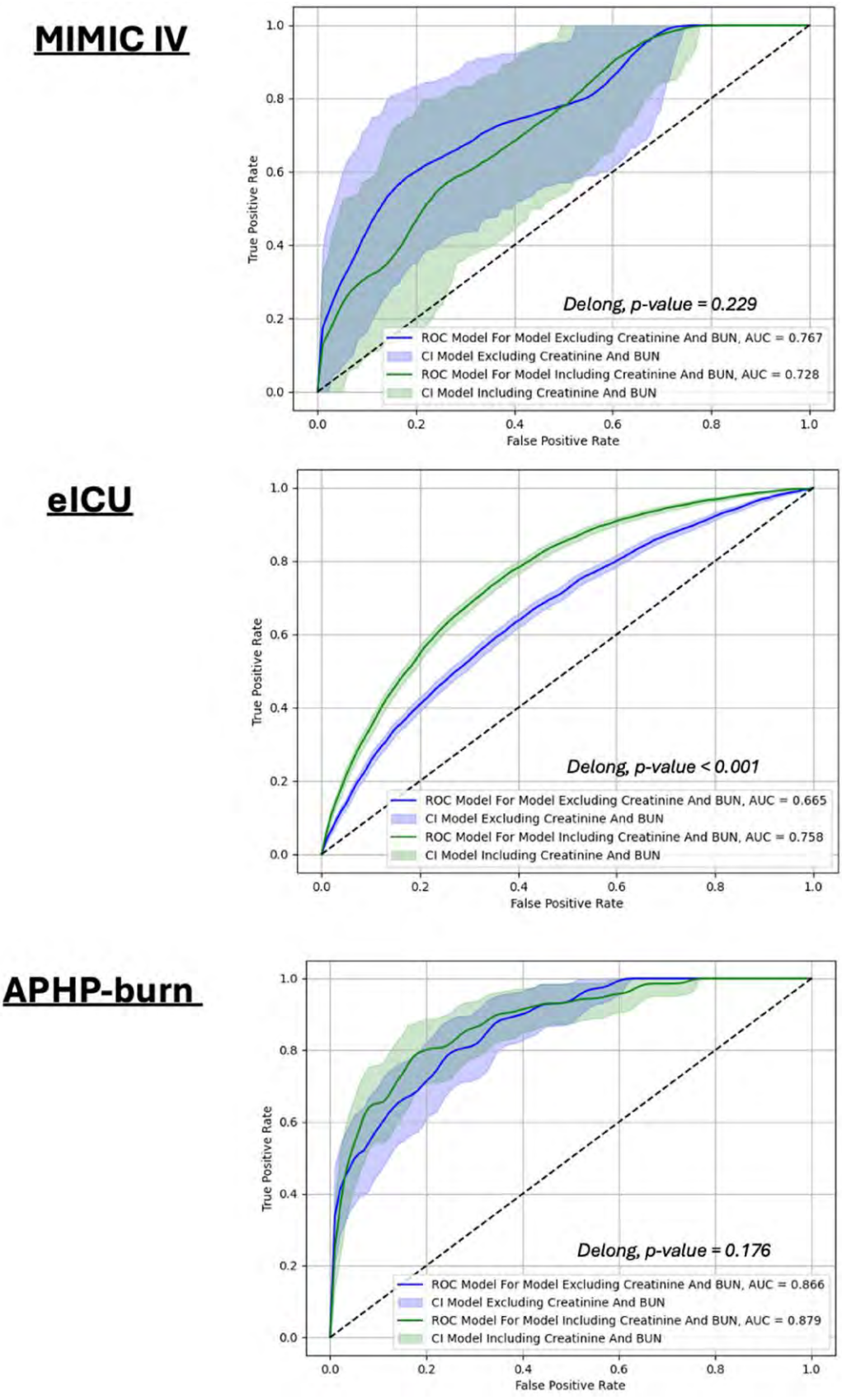


Figure 2 Prediction performance of models for MAKE. MAKE, major adverse kidney event.

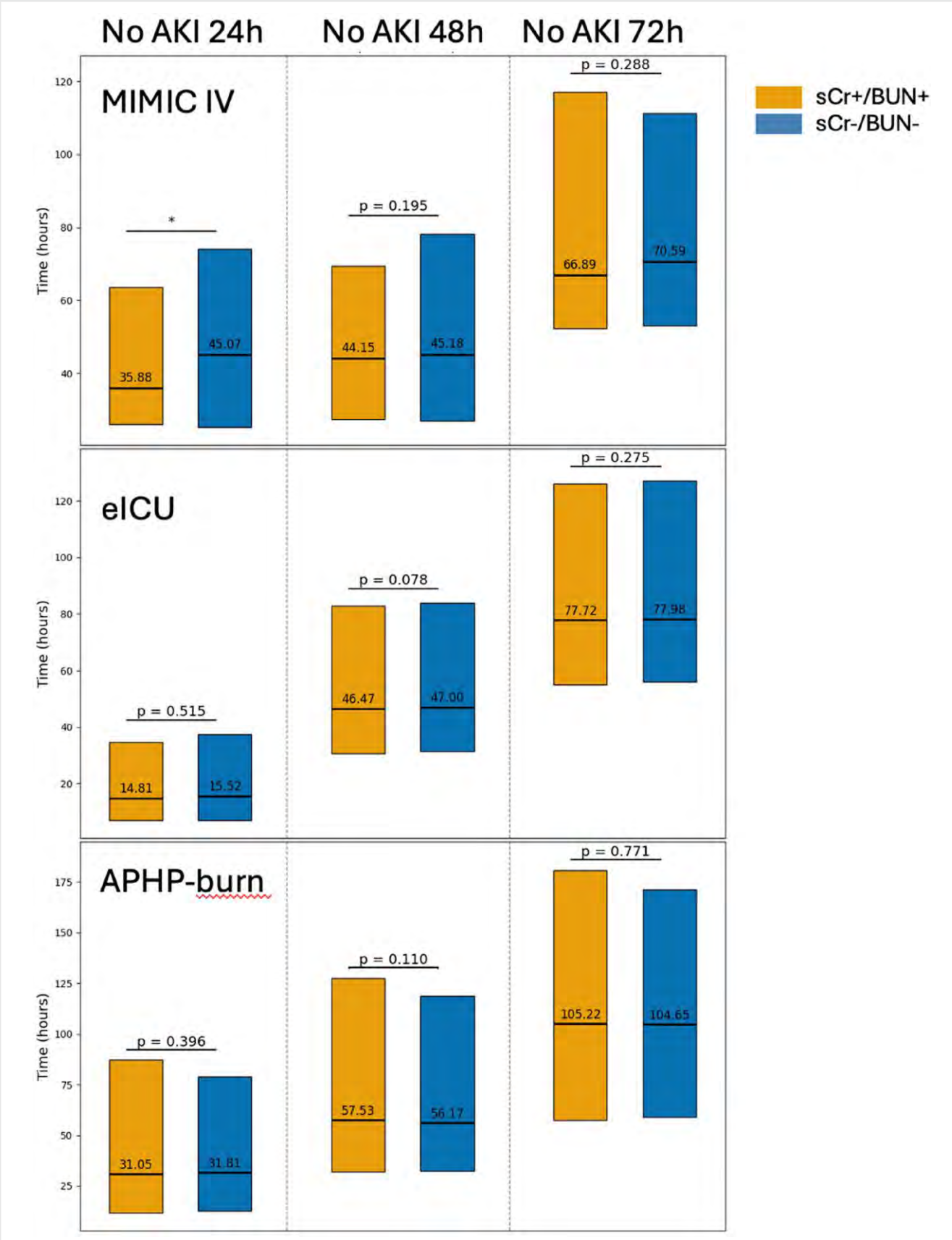


Figure 3 Duration between high prediction probability of AKI and onset of AKI.

model from 0.76 (95%CI, 0.72–0.80) to 0.75 (95%CI, 0.70–0.79) (see Supplementary material online, *Figure S10*; Supplementary material online, *Table S6*).

Discussion

This multicentre external validation study demonstrates that the ML models developed in this work could accurately predict AKI and MAKE across diverse ICU populations—even in the absence of sCr and BUN as input features. Our findings show that systemic haemodynamic and physiological signatures alone carry sufficient information to anticipate kidney injury, achieving predictive performance comparable to models including sCr and BUN. Remarkably, the sCr−/BUN− model maintained good performance even when predicting AKI defined by sCr criteria. Moreover, median lead times exceeded 70 h, providing a clinically actionable window to intervene well before conventional renal biomarkers signal injury.

These results support a paradigm shift: reliable AKI prediction can be achieved with renal-independent, continuously available bedside data. Haemodynamic-driven models thus offer a practical and scalable solution for early kidney risk stratification, particularly relevant in critical care settings where continuous monitoring is standard. This promotes a new non-invasive method to manage AKI and its prognosis.

The term 'haemodynamic signature' refers to the multivariate and time-dependent combinations of cardiovascular and respiratory parameters that precede AKI onset. Rather than detecting isolated abnormalities such as hypotension or tachycardia, the model captures complex patterns—sustained elevations in heart rate relative to mean arterial pressure, progressive loss of blood pressure variability, concurrent increases in respiratory rate and oxygen extraction, or subtle acid–base shifts—that collectively signal early systemic stress and microcirculatory compromise. These non-linear interactions, continuously integrated by the model, likely represent the transition from compensated to decompensated physiology preceding overt renal dysfunction. The dominance of MAP, heart rate, and respiratory rate among top importance features thus reflects physiologically interpretable, clinically plausible dynamics rather than opaque model behaviour.

Despite substantial differences in patient characteristics and AKI incidence across the APHP-burn, eICU, and MIMIC-IV cohorts, the predictive performance of the models remained robust.[15,16] This highlights the generalizability and transferability of our approach across diverse ICU populations.

Several AI-based models have been developed for AKI prediction, but most of these approaches rely heavily on renal biomarkers such as sCr, urine output, or laboratory-specific data, and are generally designed for static risk assessment at admission rather than real-time physiological monitoring. In contrast, our model focuses exclusively on systemic haemodynamic and physiological variables, enabling biomarker-independent early prediction and continuous risk updating. Despite this difference in design, its discrimination (auROC ≈ 0.75–0.85 across cohorts) compares favourably with previously reported externally validated models, which typically range between 0.80 and 0.82.[11,17] This performance, combined with a median predictive lead time exceeding 70 h, highlights that systemic signatures alone provide sufficient information for actionable, early kidney-risk stratification.

Interestingly, model performance was highest in the MIMIC-IV cohort, likely reflecting that the algorithm was primarily trained on this dataset and was fitted to MIMIC IV characteristic population.[18] In contrast, the APHP-burn and eICU cohorts provided critical insights into performance in more heterogeneous and real-world ICU settings, including burn patients and a multicentre ICU population, respectively. The burn ICU cohort is especially important: in this extreme stress setting, inclusion of sCr and BUN conferred no benefit, underscoring the robustness of systemic physiological signatures as early warning signals. In burn patients, the superior performance of non-renal parameters likely reflects the distinct pathophysiology of burn-associated AKI. Severe burns trigger a systemic inflammatory and hypermetabolic response characterized by massive capillary leak, hypovolemia, and catecholamine-driven vasoplegia, which profoundly alter cardiovascular and microcirculatory homeostasis before renal injury becomes biochemically apparent. These perturbations manifest as measurable changes in heart rate, arterial pressure, oxygenation, and acid–base balance—features continuously captured by haemodynamic monitoring. In contrast, sCr and BUN are unreliable early indicators in this setting, as they are affected by fluid resuscitation, muscle catabolism, sepsis (resulting in a decrease in sCr production)[19] and non-steady-state kinetics. Hence, haemodynamic and physiological data more directly represent the early systemic drivers of renal dysfunction in burns, explaining the comparable or even superior predictive accuracy of biomarker-independent models in this cohort.

One of the main strengths of this study is the demonstration that reliable AKI prediction is feasible without relying on sCr or BUN biomarkers that define AKI but may not always accurately capture early kidney injury. The superior late-stage prediction observed in eICU with the sCr−/BUN− model, and the equivalent performance in burns, both highlight that systemic haemodynamics, respiratory signals, and basic biochemical variables are early markers of cardio-renal stress, preceding overt dysfunction captured by delayed renal biomarkers.

Prediction of MAKE was also effective, reinforcing the potential role of non-renal features in prognostic modelling. Notably, lead times from prediction to AKI onset were substantial—often exceeding 72 h—indicating a clinically meaningful window for preventive intervention.[20] Although model performance declined modestly when extending the prediction horizon, early warnings remained feasible, suggesting practical utility for real-time monitoring. The definition of MAKE at ICU discharge and limited to patients with a length of stay exceeding 7 days was chosen to ensure sufficient follow-up for assessing long-term renal recovery and RRT use. However, this approach may introduce survivorship bias, excluding patients with early mortality or short ICU stays, and variable follow-up bias, since the observation period differs across patients. This limitation reflects the constraints of retrospective datasets without standardized post-discharge follow-up. To mitigate this, a sensitivity analysis using a fixed 28-day mortality including all admitted patient endpoint confirmed consistent model performance, suggesting that these potential biases did not materially affect our conclusions (see Supplementary material online, *Figure S11*).

Model performance in the external validation cohort was highest for AKI-sCr, which is expected given the reliance on sCr in both AKI definitions and model training. However, the ability of the sCr−/BUN− model to achieve strong predictive accuracy for creatinine-defined endpoints—without using creatinine itself—confirms the robustness of systemic haemodynamic signatures as anticipatory markers of kidney dysfunction.

Despite the major above-mentioned advances our study has some limitations. First, its retrospective design may introduce biases related to data quality, missingness, and unmeasured confounders. Second, variation in data frequency, measurement practices, and clinical documentation across datasets may have influenced model performance and comparability. XGBoost was chosen for its strong performance on structured clinical data, robustness across heterogeneous ICU populations, and ability to provide interpretable feature importance. Compared with other algorithms previously tested (not published) (logistic regression, SVM, random forest), it consistently achieved higher accuracy and stability. Its balance between predictive performance, computational efficiency, and explainability justified its use in this study.

While full SHAP cannot be fully shared for intellectual-property reasons, representative feature importance plots are provided in the supplementary material. These confirm that the model's predictions are primarily driven by physiologically meaningful variables such as mean arterial pressure, heart rate, oxygen saturation, and respiratory rate.

Finally, while the sCr−/BUN− model demonstrates promising results, clinical applicability remains to be established and will require prospective validation in real-time settings, ideally through integration into clinical decision-support systems for proactive cardio-renal management.

From a clinical perspective, a real-time alert generated by this model could serve as an early trigger to reassess kidney-protective strategies. When a patient is identified as high-risk, clinicians may verify and optimize haemodynamic stability—ensuring adequate perfusion pressure, reassessing fluid balance, and adjusting vasopressor support as needed. It should also prompt a careful review of ongoing or planned exposure to nephrotoxic agents and a reinforcement of renal monitoring through close follow-up of urine output and sCr trends. Importantly, such an alert may also guide the search for underlying complications, such as sepsis or bleeding, that could further compromise renal and overall prognosis. In selected cases, early multidisciplinary consultation with nephrology or critical care specialists may be warranted. These actions align with current KDIGO and ESICM recommendations and illustrate how physiology-based predictive tools can translate into actionable, preventive bedside interventions.[4]

The clinical translation of this work will follow a progressive pathway, starting with prospective validation of the model in real-time ICU monitoring systems to confirm predictive accuracy and lead time in daily practice. Once validated, the algorithm could be integrated into bedside monitoring or electronic health records as an automated kidney risk alert, displaying dynamic risk trends to support early preventive actions. A subsequent interventional trial will be required to evaluate whether early alerts effectively reduce AKI incidence and improve renal outcomes. From a regulatory perspective, the model would qualify as medical device software and would require certification, with specific attention to explainability, auditability, and data protection.

Although several studies have already evaluated similar AI-based alert systems for AKI prediction,[21,22] the overall level of evidence supporting a tangible clinical benefit from such implementations remains limited. The key differentiating feature of the present approach lies in its early detection capability (72 h) compared to existing study, which may provide a broader actionable window for intervention and renal protection before overt injury occurs.

Conclusion

This study demonstrates the feasibility of predicting AKI and renal prognosis without relying on traditional renal biomarkers, using primarily systemic physiological and haemodynamic parameters. Haemodynamic-driven AI models achieved early and robust performance across heterogeneous ICU populations, including burn patients, underscoring that systemic signatures alone carry sufficient information to anticipate kidney dysfunction. These findings represent a meaningful step towards real-time, accessible, and proactive cardio-renal risk monitoring tools in critical care settings.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

None.

Authors' contributions

Louis Boutin, François Dépret, Fedi Kadri, Arij Chaftar, Anis Ghorbel, and Alexandre Mebazaa have contributed to the study conception, design, writing—original draft preparation, and conceptualization. All authors commented and contributed to revise the manuscript. All authors read and approved the final manuscript.

Louis Boutin (Conceptualization [equal]; Data curation [equal]; Validation [equal]; Writing—original draft [equal]; Writing—review & editing [equal]), Fedi Kadri (Data curation [equal]; Formal analysis [equal]; Visualization [equal]), Arij Chaftar (Data curation [equal]; Formal analysis [equal]; Visualization [equal]), Benjamin Deniau (Writing—review & editing [equal]), Sakura Minani (Writing—review & editing [equal]), Stefanny M Figueroa (Writing—review & editing [equal]), Christos E Chadjichristos (Writing—review & editing [equal]), Anis Ghorbel (Conceptualization [equal]; Supervision [equal]; Validation [equal]; Writing—original draft [equal]; Writing—review & editing [equal]), Alexandre Mebazaa (Conceptualization [equal]; Supervision [equal]; Validation [equal]; Writing—original draft [equal]; Writing—review & editing [equal]), and François Dépret (Conceptualization [equal]; Supervision [equal]; Validation [equal]; Writing—original draft [equal]; Writing—review & editing [equal])

Funding

None declared.

Conflict of interest: Louis Boutin, Benjamin Deniau, Sakura Minani, Stefanny M. Figueroa, Christos E. Chadjichristos, and François Dépret have no conflict of interest to declare. Fedi Kadri, Anis Ghorbel, Arij Chaftar, and Alexandre Mebazaa are collaborating with Precisia Care an AI working company.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Pickkers P, Darmon M, Hoste E, Joannidis M, Legrand M, Ostermann M, *et al.* Acute kidney injury in the critically ill: an updated review on pathophysiology and management. *Intensive Care Med* 2021;**47**:835–850.
2. Ostermann M, Zarbock A, Goldstein S, Kashani K, Macedo E, Murugan R, *et al.* Recommendations on acute kidney injury biomarkers from the acute disease quality initiative consensus conference: a consensus statement. *JAMA Netw Open* 2020;**3**:e2019209.
3. Kellum JA, Lameire N, Aspelin P, Barsoum RS, Burdmann EA, Goldstein SL, *et al.* Kidney disease: improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl* 2012;**2**:1–138.
4. Joannidis M, Druml W, Forni LG, Groeneveld ABJ, Honore PM, Hoste E, *et al.* Prevention of acute kidney injury and protection of renal function in the intensive care unit: update 2017: expert opinion of the Working Group on Prevention, AKI section, European Society of Intensive Care Medicine. *Intensive Care Med* 2017;**43**:730–749.
5. Ostermann M, Lumlertgul N, Jeong R, See E, Joannidis M, James M. Acute kidney injury. *Lancet* 2025;**405**:241–256.
6. Kellum JA, Ronco C, Bellomo R. Conceptual advances and evolving terminology in acute kidney disease. *Nat Rev Nephrol* 2021;**17**:493–502.
7. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders HJ. Acute kidney injury. *Nat Rev Dis Primers* 2021;**7**:52.
8. Yang H, Chen Y, He J, Li Y, Feng Y. Advances in the diagnosis of early biomarkers for acute kidney injury: a literature review. *BMC Nephrol* 2025;**26**:115.
9. Ostermann M, Legrand M, Meersch M, Srisawat N, Zarbock A, Kellum JA. Biomarkers in acute kidney injury. *Ann Intensive Care* 2024;**14**:145.
10. Tran TT, Yun G, Kim S. Artificial intelligence and predictive models for early detection of acute kidney injury: transforming clinical practice. *BMC Nephrol* 2024;**25**:353.
11. Cama-Olivares A, Braun C, Takeuchi T, O'Hagan EC, Kaiser KA, Ghazi L, *et al.* Systematic review and meta-analysis of machine learning models for acute kidney injury risk classification. *J Am Soc Nephrol* 2025;**36**:1969–1983.
12. Cama-Olivares A, Braun CG, Takeuchi T, Kaiser KA, Ghazi L, Chen J, *et al.* Artificial intelligence/machine learning externally validated models for AKI risk-classification: a systematic review and meta-analysis. *J Am Soc Nephrol* 2024;**36**:1869–1983.

13. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, *et al.* MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10:1–9.
14. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data* 2018;**5**:1–13.
15. Chen S, Fan J, Pishgar E, Alaei K, Placencia G, Pishgar M. Machine learning-based prediction of ICU mortality in sepsis-associated acute kidney injury patients using MIMIC-IV database with validation from eICU database. arXiv 2025; https://doi.org/10.48550/arXiv.2502.17978
16. Lin Y, Shi T, Kong G. Acute kidney injury prognosis prediction using machine learning methods: a systematic review. *Kidney Med* 2025;**7**:100936.
17. Al-Absi DT, Simsekler MCE, Omar MA, Anwar S. Exploring the role of artificial intelligence in acute kidney injury management: a comprehensive review and future research agenda. *BMC Med Inform Decis Mak* 2024;**24**:1–20.
18. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, Van Dlepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;**14**:49–58.
19. Doi K, Yuen PST, Eisner C, Hu X, Leelahavanichkul A, Schnermann J, *et al.* Reduced production of creatinine limits its use as marker of kidney injury in sepsis. *J Am Soc Nephrol* 2009;**20**:1217–1221.
20. Swan JT, Moore LW, Sparrow HG, Frost AE, Gaber AO, Suki WN. Optimization of acute kidney injury (AKI) time definitions using the electronic health record: a first step in automating in-hospital AKI detection. *J Clin Med* 2021;**10**:3304.
21. Chen JJ, Lee TH, Chan MJ, Tsai TY, Fan PC, Lee CC, *et al.* Electronic alert systems for patients with acute kidney injury: a systematic review and meta-analysis. *JAMA Netw Open* 2024;**7**:e2430401–e2430401.
22. Wang H, Deng L, Li T, Liu K, Mao H, Wu B. The influence of electronic AKI alert on prognosis of adult hospitalized patients: a systematic review and meta-analysis. *Crit Care* 2025;**29**:1–18.