GMR
The Original

IDENTIFICATION OF CLINICAL PARAMETERS PREDICTING HEMODYNAMIC INSTABILITY IN ICU SETTING

Saifullah Sheikh[1*], Fakhir Raza Haidri[2], Mujtaba Hassan[3], Talha Khalid[4], Mushkbar Fatima, M tanzeel [5]

[1*]Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan, Email: saifsheikh269@yahoo.com
[2]Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan.
[3]Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan.
[4]Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan.
[5]Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan.

ABSTRACT

Background. Haemodynamic instability is a leading determinant of morbidity and mortality in the intensive care unit (ICU). A clinically important but poorly characterised phenotype is the patient who progresses from haemodynamic stability to instability without receiving vasopressor support, whether through delayed recognition or resource constraint. South Asian ICU populations are markedly underrepresented in the haemodynamic-instability literature. We characterised the baseline physiological differences and serial deterioration trajectories of such patients in a Pakistani tertiary-care cohort and developed an interpretable machine-learning early-warning system on the same data.

Methods. We studied 253 adult ICU patients at the Sindh Institute of Urology and Transplantation (SIUT), Karachi. Patients were classified as haemodynamically unstable (n = 118) if they developed sustained systolic blood pressure (SBP) < 90 mmHg and/or mean arterial pressure (MAP) < 65 mmHg without any vasopressor or vasoactive agent during deterioration, or stable (n = 135). Twenty clinical parameters were recorded at 2-hour intervals over a maximum 48-hour window (6,071 observations). Groups were compared using Mann-Whitney U and chi-square/Fisher tests. From 125 engineered features (vital-sign time-series dynamics, patient-level aggregates, composite clinical indices, and admission laboratory values), 50 were selected by SHAP importance and used to train a weighted ensemble of LightGBM, XGBoost, and logistic regression, validated by 5-fold patient-level stratified cross-validation and on a held-out test set.

Results. At admission, unstable patients had lower SBP (102.3 ± 22.0 vs 120.4 ± 19.6 mmHg, $p < 0.001$), lower MAP (76.5 ± 15.5 vs 87.7 ± 13.5 mmHg, $p < 0.001$), higher shock index (1.056 ± 0.360 vs 0.813 ± 0.224, $p < 0.001$), and higher lactate (2.8 ± 2.4 vs 1.5 ± 0.9 mmol/L, $p < 0.001$), whereas arterial pH, bicarbonate, creatinine, and haematocrit did not differ significantly, indicating that haemodynamic divergence preceded laboratory derangement. In the unstable group, MAP reached a nadir of 63.6 ± 9.0 mmHg at T+6h, shock index remained ≥ 1.0 and lactate remained > 2.0 mmol/L throughout, all without vasopressor support. The ensemble achieved an AUROC of 0.9941 ± 0.0064, sensitivity 96.67%, specificity 95.59%, and a Brier score of 0.0316 in cross-validation (test-set AUROC 0.9928), and retained an AUROC of 0.9788 using only the first 6 hours of data.

Conclusions. In a resource-constrained South Asian ICU, patients who deteriorated without vasopressor support followed a reproducible trajectory in which vital signs—particularly MAP, SBP, and shock index—signalled compromise earlier than laboratory tests. An interpretable physiology-driven ensemble identified these patients with high discrimination and good calibration from the first hours of admission, supporting its use as an early-warning aid where continuous invasive monitoring is unavailable.

KEYWORDS: Haemodynamic instability; vasopressor; intensive care unit; shock index; lactate; mean arterial pressure; machine learning; early-warning system; SHAP; Pakistan.

INTRODUCTION

Background

The severity of critical illness has historically been quantified by composite scoring systems. The Acute Physiology and Chronic Health Evaluation II (APACHE II) score, introduced in 1985, combined physiological and laboratory variables—arterial pH, bicarbonate, electrolytes, creatinine, haematocrit, leucocyte count, and vital signs—to estimate severity and prognosis [1]. The 1990s saw the emergence of early-warning scores and medical emergency and rapid-response teams that used threshold-based criteria to identify patients requiring escalation [2], alongside the Sequential Organ Failure Assessment (SOFA) score for sepsis-related organ dysfunction [3]. These tools remain central to risk stratification but were designed for periodic, threshold-driven assessment rather than continuous trajectory detection.

Haemodynamic instability—inadequate blood flow to vital organs—is a major challenge in the ICU because of its strong association with morbidity and mortality [4]. It commonly arises from sepsis, haemorrhage, and cardiac failure, each of which compromises oxygen delivery and can progress rapidly to multi-organ failure without timely intervention. Current monitoring relies on heart rate, blood pressure, oxygen saturation, capillary refill time, urine output, respiratory rate, and

mental status [5], yet the sensitivity and specificity of these indicators for predicting the onset of instability are inconsistent [6], and compensatory mechanisms may preserve blood pressure until circulatory compromise is advanced [7,8].
Fluid resuscitation and vasoactive therapy are cornerstones of haemodynamic management [9,10], and the Surviving Sepsis Campaign (SSC) recommends early vasopressor initiation targeting a MAP of at least 65 mmHg [11]. Delayed vasopressor initiation is associated with higher in-hospital mortality, fewer vasopressor-free days, and longer time to haemodynamic targets [12,13]. Despite this, deterioration that proceeds without timely vasopressor support remains a clinical reality, particularly in resource-limited settings where delayed recognition, staffing constraints, or decision-making may delay intervention. Systematic characterisation of the physiological trajectory of patients who deteriorate without intervention is valuable: it defines a high-risk phenotype not currently receiving pharmacological support, and it provides the empirical basis for training machine-learning early-warning systems—conceptually similar to the work of Rahman et al. [6], who predicted future haemodynamic intervention from more than 200,000 ICU stays, but here directed at what happens physiologically when that intervention does not occur.
Most published ICU prediction models are developed on large databases from high-income countries (MIMIC-III, MIMIC-IV, eICU) [14], whose patient populations, disease patterns, and care systems differ substantially from those of developing countries. South Asian ICU populations are correspondingly underrepresented. SIUT in Karachi is one of Pakistan's largest tertiary-care referral centres, managing a high burden of community-acquired sepsis, respiratory failure, and acute kidney injury. We report the characterisation of serial haemodynamic trajectories in ICU patients who progressed from stability to instability without vasopressor intervention, and an interpretable ensemble early-warning system developed on the same prospectively observed, 2-hourly, multi-parameter cohort. The study aims were: (i) to compare baseline physiology between stable and unstable patients; (ii) to describe the serial deterioration trajectory of unstable patients in the absence of vasopressor support; and (iii) to develop and validate a clinically interpretable model that identifies at-risk patients early in their ICU stay.

METHODS

Study design and setting

This was a retrospective analysis of prospectively collected observational data from the Intensive Care Unit of the Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan. SIUT is a publicly funded tertiary-care referral centre operating a high-acuity ICU that manages sepsis, respiratory failure, acute kidney injury, pancreatitis, hepatic failure, and neurological emergencies. Data were collected during 2025–2026 as part of a structured bedside observation programme maintained by the ICU clinical team. The study was conducted in accordance with institutional ethics requirements.

Patient selection and cohort classification

A total of 253 adult ICU patients were enrolled and classified by haemodynamic trajectory and vasopressor-administration record into two groups. The haemodynamically unstable group (n = 118) comprised patients who were stable at ICU admission but subsequently developed sustained SBP < 90 mmHg and/or MAP < 65 mmHg during observation, with no vasopressor or vasoactive agent (noradrenaline, dopamine, adrenaline, vasopressin, or equivalent) administered during the period of observed deterioration; this group represents spontaneous haemodynamic deterioration unmodified by pharmacological circulatory support. The haemodynamically stable group (n = 135) maintained stability throughout, did not meet instability criteria, and required no vasopressor support.
Inclusion criteria were age ≥ 18 years, a minimum of two complete observation time points (≥ 4 hours of monitoring), and documentation of core vital signs with at least one baseline laboratory measurement at admission. Exclusion criteria were vasopressor or vasoactive therapy at or before ICU admission, do-not-resuscitate designation, fewer than two recorded observation slots, and withdrawal of active treatment during observation. All 253 enrolled patients met eligibility criteria; none were excluded.

Clinical observations

Clinical data were extracted from structured bedside nursing observation charts and laboratory information systems. Each patient was monitored at 2-hour intervals across a maximum 48-hour window (up to 24 slots per patient), yielding 6,071 observation rows. Demographics (age, sex, admitting diagnosis, comorbidities, admission date) were extracted from structured ICU registration records. Twenty clinical parameters were recorded at each slot (Table 1). Laboratory values were obtained at or near admission and treated as a baseline static profile, consistent with standard ICU observational practice and the approach used by Rahman et al. [6]; vital signs were recorded freshly at every 2-hour point. All variables were range-validated against physiologically plausible limits, and outliers beyond those bounds were excluded from the relevant calculations.

Table 1. Clinical parameters recorded at each 2-hour observation time point.

Domain	Parameters recorded
Haemodynamic	Systolic BP (mmHg), Diastolic BP (mmHg), Mean Arterial Pressure (mmHg), Heart Rate (bpm)
Respiratory	Respiratory Rate (breaths/min), Peripheral Oxygen Saturation – SpO_2 (%), Fraction of Inspired Oxygen – FiO_2 (%)
Renal	Urine Output (ml/hr), Serum Creatinine (mg/dL)

Domain	Parameters recorded
Neurological	AVPU consciousness score (Alert / Voice / Pain / Unresponsive)
Metabolic	Serum Lactate (mmol/L), Arterial pH, Serum Bicarbonate – HCO_3 (mEq/L)
Haematological	Haematocrit (%), Total Leucocyte Count ($\times 10^3/\mu L$), Platelet Count ($\times 10^3/\mu L$)
Biochemical	Serum Potassium (mEq/L), Serum Sodium (mEq/L), Total Bilirubin (mg/dL)
General	Core Temperature (°C), Capillary Refill Time – CRT (seconds)

Derived variables

The shock index was computed at each time point as heart rate divided by systolic blood pressure (HR/SBP) [15]; a value ≥ 1.0 was pre-specified as a threshold for haemodynamic compromise [15]. Pulse pressure was derived as SBP − DBP, and MAP deficit as observed MAP − 65 mmHg, with negative values indicating breach of the SSC organ-perfusion target [11].

Data cleaning and preparation

A clinical audit preceded modelling. Identified data-entry errors were corrected systematically: temperatures recorded in Fahrenheit were converted to Celsius (implausible values > 43 °C set to missing); arterial pH values recorded without a decimal point (e.g. 740) were divided by 100; physiologically impossible electrolyte and heart-rate values were set to missing; a pulse-pressure data-entry error affecting 1,776 rows was recalculated as SBP − DBP; bilirubin decimal-entry errors were corrected; and erroneous zero laboratory values were set to missing rather than imputed as true zeros. AVPU free-text variants were standardised and numerically encoded (A = 0, V = 1, P/S = 2, U = 3). To standardise patients to the observation grid, vital signs were forward- then backward-filled within each patient, and missing admission laboratory values were imputed with the patient-level median (cohort median as fallback). After cleaning, the dataset contained no missing values and all parameters fell within clinically validated ranges.

Feature engineering

Vital signs and laboratory values were treated differently by design. Vital signs, recorded every 2 hours, are genuine time-varying signals; laboratory values, recorded once at admission, were retained as fixed admission-baseline scalars rather than artificially propagated as pseudo-time-series. From 20 raw parameters, 125 features were engineered across six families: (i) rolling 8-hour window statistics (mean, standard deviation, minimum, maximum) for ten vital parameters; (ii) first- and second-order deltas (rate of change over 2 and 4 hours) for twelve parameters; (iii) cumulative linear-regression slopes for SBP, MAP, HR, SpO_2, FiO_2, and urine output; (iv) composite clinical indices including shock index, SpO_2/FiO_2 ratio, MAP deficit, acidosis index, and the HCO_3–lactate gap; (v) patient-level aggregates capturing the worst state observed to date (e.g. minimum SBP, minimum MAP, minimum urine output, peak shock index); and (vi) static admission laboratory values and metadata (age, sex, comorbidity flags, admitting-diagnosis category).

Feature selection used mean absolute SHAP (SHapley Additive exPlanations) values from a preliminary XGBoost model; the 50 highest-ranked features were retained, reducing the feature-to-patient ratio and improving cross-validation stability. The strongest predictors were patient-level minimum SBP and minimum MAP, minimum urine output, minimum SpO_2, and admission haematocrit—a haemodynamic rather than purely laboratory-driven signature of early instability.

Model development and validation

The 253 patients were split at the patient level by stratified sampling into training (176, 70%), validation (38, 15%), and held-out test (39, 15%) sets, with class balance preserved across splits and no patient appearing in more than one split. Patient-level splitting prevents the data leakage that occurs when rows from the same patient appear in both training and test sets and temporal features are computed from shared history. The final model was a weighted soft-voting ensemble combining LightGBM (weight 0.7), XGBoost (0.2), and L2-regularised logistic regression (0.1): $P = 0.7 \cdot P(LGB) + 0.2 \cdot P(XGB) + 0.1 \cdot P(LR)$. LightGBM received the highest weight for its superior generalisation on the small cohort via leaf-wise growth; XGBoost the lowest owing to a greater overfitting tendency; logistic regression added linear diversity and improved calibration. Tree-model hyperparameters were tuned with Optuna (Tree-structured Parzen Estimator, 80 trials each) on validation AUROC with early stopping. The decision threshold was fixed at 0.50 to ensure deployable, fold-independent behaviour.

Performance was estimated by 5-fold patient-level stratified cross-validation (the primary analysis) and confirmed on the held-out test set, using AUROC, sensitivity, specificity, positive and negative predictive value, F1-score, and the Brier score for calibration. Feature attribution used TreeSHAP, providing both global feature ranking and per-time-slot attribution for individual predictions. A subsequent safety review augmented the rule-based component with hard physiological floors (e.g. $SpO_2 < 85\%$, MAP < 45 mmHg, lactate > 6.0 mmol/L, pH < 7.10) so that any single critically deranged parameter forces a high-risk classification independent of other inputs; this addressed an isolated-hypoxaemia blind spot identified in an earlier iteration.

Statistical analysis

Continuous variables were assessed for normality by the Shapiro-Wilk test; as most violated normality, non-parametric methods were used throughout. Continuous variables were compared with the two-tailed Mann-Whitney U test and categorical variables with the chi-square or Fisher exact test as appropriate; odds ratios with 95% confidence intervals quantified categorical associations. Values are reported as mean ± standard deviation unless otherwise stated, and serial trajectories as per-slot mean ± SD. A two-tailed $p < 0.05$ was considered significant. Analyses and figures were produced in Python 3.11–3.12 (SciPy, Pandas, NumPy, Matplotlib, scikit-learn, XGBoost, LightGBM, SHAP, Optuna). The serving system comprised a FastAPI backend with a React dashboard and a local SQLite store, designed to run on commodity hospital hardware without internet access.

RESULTS

Of 253 patients, 118 (46.6%) developed haemodynamic instability without vasopressor intervention and 135 (53.4%) remained stable. Unstable patients had shorter monitoring records before the primary event (15.8 ± 5.7 slots; ≈ 32 ± 11 h) than stable patients (24.0 ± 0.1 slots; 48 h), consistent with deterioration shortening the available observation window.

Patient characteristics

The groups were well matched for age (48.4 ± 18.5 vs 47.0 ± 19.0 years; $p = 0.505$) and sex (58.5% vs 50.4% male; $p = 0.704$). Admitting diagnoses differed substantially (Table 2): respiratory conditions (CAP, HAP, ARDS) predominated among unstable patients (41.5% vs 11.9%; $p < 0.001$), followed by sepsis/septic shock (26.3% vs 6.7%; $p < 0.001$), whereas the stable group was diagnostically heterogeneous, with most classified as "other" (56.3%) and a higher proportion of renal/AKI admissions (19.3% vs 9.3%; $p = 0.028$). The over-representation of renal diagnoses in the stable group reflects SIUT's nephrology focus, where many patients present with renal disease yet maintain circulatory stability.

Table 2. Baseline patient characteristics by group.

Characteristic	Unstable (n = 118)	Stable (n = 135)	p-value
Age, years (mean ± SD)	48.4 ± 18.5	47.0 ± 19.0	0.505
Male sex, n (%)	69 (58.5%)	68 (50.4%)	0.704
Observation slots (mean ± SD)	15.8 ± 5.7	24.0 ± 0.1	< 0.001
Respiratory (CAP/HAP/ARDS), n (%)	49 (41.5%)	16 (11.9%)	< 0.001
Sepsis / septic shock, n (%)	31 (26.3%)	9 (6.7%)	< 0.001
Renal / AKI, n (%)	11 (9.3%)	26 (19.3%)	0.028
GI / hepatic (incl. pancreatitis), n (%)	12 (10.2%)	5 (3.7%)	0.038
Neurological, n (%)	3 (2.5%)	3 (2.2%)	0.879
Cardiac, n (%)	2 (1.7%)	0 (0.0%)	0.149
Other, n (%)	10 (8.5%)	76 (56.3%)	< 0.001

Continuous variables: Mann-Whitney U test; categorical: chi-square or Fisher exact test. CAP, community-acquired pneumonia; HAP, hospital-acquired pneumonia; ARDS, acute respiratory distress syndrome; AKI, acute kidney injury; GI, gastrointestinal.

All five examined comorbidities were more common in unstable patients: diabetes mellitus (20% vs 7%; OR 3.2; $p = 0.005$), hypertension (20% vs 5%; OR 4.7; $p < 0.001$), chronic kidney disease/ESRD (14% vs 2%; OR 7.4; $p < 0.001$), ischaemic heart disease (6% vs 1%; OR 8.5; $p = 0.046$), and COPD (4% vs 0%; $p = 0.050$). Ninety per cent of stable patients had no comorbidity, compared with 53% of unstable patients, reflecting the compounding physiological vulnerability that pre-existing disease confers when acute illness develops.

Baseline haemodynamic and laboratory parameters

At admission, the unstable group was significantly worse across all major haemodynamic parameters (Table 3). A striking dissociation was evident: arterial pH, bicarbonate, creatinine, haematocrit, platelets, and electrolytes did not differ significantly between groups, whereas pressure-based vital signs, shock index, and lactate did. This indicates that haemodynamic divergence preceded laboratory derangement—vital-sign monitoring, rather than repeat laboratory testing, carries the primary early-warning signal, and biochemical evidence of organ injury lags behind circulatory compromise. Notably, 47.5% of unstable patients were identifiable at admission by shock index ≥ 1.0 alone, with no laboratory requirement.

Table 3. Baseline haemodynamic, metabolic, and laboratory parameters at ICU admission.

Parameter	Unstable (n = 118)	Stable (n = 135)	p-value
Haemodynamic			

Parameter	Unstable (n = 118)	Stable (n = 135)	p-value
SBP (mmHg)	102.3 ± 22.0	120.4 ± 19.6	< 0.001
MAP (mmHg)	76.5 ± 15.5	87.7 ± 13.5	< 0.001
HR (bpm)	102.7 ± 24.1	95.4 ± 20.6	0.036
Shock index (HR/SBP)	1.056 ± 0.360	0.813 ± 0.224	< 0.001
RR (breaths/min)	22.4 ± 5.9	20.1 ± 4.2	0.001
SpO_2 (%)	97.2 ± 3.1	97.7 ± 1.3	0.678
Urine output (ml/hr)	77.1 ± 91.8	92.8 ± 67.8	< 0.001
Metabolic & renal			
Serum lactate (mmol/L)	2.8 ± 2.4	1.5 ± 0.9	< 0.001
Arterial pH	7.36 ± 0.08	7.38 ± 0.06	0.138
Serum HCO_3 (mEq/L)	20.5 ± 6.2	19.9 ± 5.3	0.676
Serum creatinine (mg/dL)	2.3 ± 2.2	2.2 ± 2.4	0.257
Haematological & electrolytes			
Haematocrit (%)	30.7 ± 6.1	31.1 ± 7.1	0.666
TLC ($\times 10^3/\mu L$)	14.9 ± 8.4	12.8 ± 7.3	0.050
Platelets ($\times 10^3/\mu L$)	213.5 ± 139.1	214.8 ± 128.0	0.685
Serum potassium (mEq/L)	4.2 ± 0.9	4.2 ± 0.9	0.986
Serum sodium (mEq/L)	136.9 ± 6.4	136.6 ± 6.7	0.488

Values are mean ± SD at the first recorded observation. p-values by Mann-Whitney U test. SBP, systolic blood pressure; MAP, mean arterial pressure; HR, heart rate; RR, respiratory rate; TLC, total leucocyte count.

Serial haemodynamic trajectories in the unstable group

The serial profiles of the 118 unstable patients reflect the natural course of deterioration in the complete absence of vasopressor or vasoactive support (Table 4). SBP and MAP declined progressively and in an oscillating manner over the first 8–12 hours: SBP fell from 102.3 ± 22.0 mmHg at admission to 85.8 ± 13.5 mmHg by T+6h, and MAP reached a nadir of 63.6 ± 9.0 mmHg at T+6h—below the 65 mmHg SSC organ-perfusion target—before a partial recovery by T+12h (87.3 ± 12.1 mmHg) and recurrent dips below 75 mmHg at T+18h and T+22h, describing a biphasic pattern of recurrent perfusion compromise. Heart rate remained persistently elevated (102–105 bpm) without recovery, and shock index stayed at or above 1.0 throughout, reflecting sustained compensatory tachycardia against falling perfusion pressure. Serum lactate remained above the 2.0 mmol/L hypoperfusion threshold across the entire window with no downward trend—evidence of sustained tissue hypoperfusion that did not self-resolve—while urine output declined progressively, consistent with worsening renal hypoperfusion.

Table 4. Serial haemodynamic and metabolic parameters in the unstable group (n = 118); no vasopressor intervention at any point.

Parameter	T+0h	T+2h	T+4h	T+6h	T+8h	T+12h	T+24h†
SBP (mmHg)	102.3	89.3	96.5	85.8	95.4	117.4	120.1
MAP (mmHg)	76.5	66.5	71.3	63.6	70.8	87.3	87.2
HR (bpm)	102.7	103.7	102.9	103.6	102.4	104.5	104.4
RR (br/min)	22.4	22.5	22.7	22.6	22.7	22.8	22.8
SpO_2 (%)	97.2	96.8	97.1	96.8	96.9	97.0	97.2
Lactate (mmol/L)	2.8	3.0	2.8	3.0	2.8	2.8	2.5
UOP (ml/hr)	77.1	70.6	75.7	69.5	72.7	66.5	71.7

Per-slot mean values; T+0h = ICU admission. MAP nadir 63.6 ± 9.0 mmHg at T+6h, below the SSC 65 mmHg target. † T+24h available for 39 patients with ≥ 36-hour windows. SD omitted for compactness; see text for key dispersion values.

Predictive-model performance

On 5-fold patient-level cross-validation (the primary analysis), the ensemble achieved an AUROC of 0.9941 ± 0.0064, F1-score 0.9588 ± 0.0198, sensitivity 96.67% ± 1.86%, specificity 95.59% ± 4.87%, and Brier score 0.0316 ± 0.0130 (Table 5). On the never-seen test set (39 patients, 936 rows), AUROC was 0.9928, sensitivity 97.22%, specificity 95.24%, PPV 94.59%, NPV 97.56%, and Brier score 0.0346, with a confusion matrix of TP 420, FP 24, FN 12, TN 480 at the 0.50 threshold. A Brier score near 0.03 (against 0.25 for a random classifier) indicates well-calibrated probabilities suitable for use as a graded risk score rather than a binary flag alone.

Table 5. Ensemble performance: 5-fold cross-validation and held-out test set.

Metric	5-fold CV (mean ± SD)	Held-out test set
AUROC	0.9941 ± 0.0064	0.9928
Sensitivity	96.67% ± 1.86%	97.22%
Specificity	95.59% ± 4.87%	95.24%
F1-score	0.9588 ± 0.0198	0.9589
PPV / NPV	—	94.59% / 97.56%
Brier score	0.0316 ± 0.0130	0.0346

Decision threshold 0.50. PPV, positive predictive value; NPV, negative predictive value.

Against progressively stronger baselines on the test set, the ensemble substantially outperformed the shock-index clinical rule (AUROC 0.6736; sensitivity 49.07%), a labs-only model (0.7460), and a vitals-only model (0.8921), confirming that vital-sign time-series carry the dominant predictive signal while admission laboratories and metadata add meaningful discrimination (Table 6). The improvement over the shock-index rule—the standard in resource-constrained practice—was 47.4% absolute in AUROC and 98.1% relative in sensitivity.

Table 6. Held-out test-set performance against progressively stronger baselines.

Model	AUROC	Sensitivity	Specificity
Shock-index rule (HR/SBP > 1.0)	0.6736	49.1%	81.0%
XGBoost – labs only (static)	0.7460	72.2%	73.5%
XGBoost – vitals only (time-series)	0.8921	88.9%	82.4%
Logistic regression – all 50 features	0.9885	86.1%	85.9%
Ensemble (XGB + LGB + LR)	0.9928	97.2%	95.2%

All ML models evaluated at threshold 0.50; shock-index rule at its conventional cut-off of 1.0.

Restricting the model to only the first 6 hours of each patient's stay preserved discrimination essentially unchanged (AUROC 0.9788), identical to the 12-, 24-, and 48-hour windows, because the dominant features are patient-level aggregates that stabilise quickly after admission. This early-window stability supports use of the system as an admission-time risk screen rather than only a late-stage monitoring tool.

Feature importance and clinical interpretation

Global SHAP analysis ranked patient-level minimum SBP, minimum MAP, minimum urine output, and minimum SpO_2 as the four most influential features, followed by admission haematocrit, confirming that the worst recorded vital-sign values—not isolated transient spikes—drive risk. Peak shock index was the strongest composite predictor, capturing circulatory compensation that individual parameters miss. Composite acid-base signals (acidosis index; HCO_3–lactate gap) outperformed their individual components, and the "renal" admitting-diagnosis category acted as a risk-reducing signal, mirroring the clinical observation that renal admissions clustered in the stable group. Age and sex were not significant predictors ($p > 0.4$), consistent with the demographic comparability between groups. This model-derived importance pattern directly corroborates the clinical characterisation: haemodynamic pressure signals dominate, and they declare themselves before laboratory derangement.

DISCUSSION

In this tertiary-care Pakistani cohort, patients who progressed to haemodynamic instability without vasopressor support followed a reproducible physiological trajectory, and an interpretable ensemble built on the same data identified them with high discrimination and good calibration from the first hours of admission. The principal clinical finding is the dissociation between haemodynamic and laboratory signals: at admission, SBP, MAP, shock index, and lactate separated

the groups strongly, whereas arterial pH, bicarbonate, creatinine, and haematocrit did not. Vital signs therefore gave earlier warning than laboratory tests—clinically important because it argues for intensified non-invasive vital-sign surveillance, particularly of MAP, SBP, and shock index, rather than reliance on repeat biochemistry, in patients at risk of deterioration. The serial trajectory reinforces this. MAP fell below the 65 mmHg SSC threshold by T+6h and the cohort displayed a biphasic pattern of recurrent perfusion dips, while shock index remained ≥ 1.0 and lactate remained above 2.0 mmol/L throughout, with no spontaneous resolution. Because no patient received vasopressor support during deterioration, these curves describe the untreated natural history of early instability—a phenotype rarely characterised because, in well-resourced settings, intervention typically interrupts it.

The persistence of hyperlactataemia and sustained tachycardia without recovery indicates that watchful waiting is unlikely to be rewarded in these patients and supports a low threshold for haemodynamic assessment and escalation when shock index exceeds 1.0 or admission lactate exceeds 2.0 mmol/L.

The model's discrimination (AUROC 0.9941 in cross-validation, 0.9928 on the test set) sits at the upper end of, and in nominal terms above, comparable single-centre ICU prediction studies, which typically report AUROC of 0.80–0.92 [6,8]. Rahman et al. reported an AUROC of 0.82 for predicting haemodynamic intervention and showed that basic vital signs alone reach an AUC of about 0.72 [6]; Chiang et al. externally validated a model at 0.92 [16]; multi-centre dynamic systems such as time-varying early-warning approaches report 0.90–0.93, and commercial bedside stability indices 0.84–0.90 [17,18]. Our result should nevertheless be interpreted with caution: it derives from a clean, single-centre, well-curated dataset, and the near-ceiling values likely reflect both data quality and the relatively clear separation between the stable and unstable phenotypes in this cohort.

Several design choices support the credibility of the result. Patient-level rather than row-level splitting eliminates the temporal-feature leakage that inflates performance when the same patient appears in both training and test partitions [19,20]. Treating laboratories as static admission scalars, rather than forward-filling a single value across 24 slots, avoids fabricating temporal structure that does not exist. The ensemble combines the non-linear capacity of LightGBM and XGBoost with the calibration stability of logistic regression, a configuration well suited to small-to-moderate cohorts [21]. And in contrast to deep-learning "black-box" approaches [22], the use of tree-based models with expert-designed clinical features and SHAP attribution makes each prediction inspectable, so clinicians can see which physiological variables drove a given risk estimate—an important property for trust and adoption at the bedside [23]. The subsequent addition of hard physiological safety floors further guards against a recognised failure mode in which an isolated severe derangement (e.g. profound hypoxaemia) is under-weighted when other vitals appear normal.

The preservation of discrimination using only the first 6 hours of data is clinically salient. It indicates that meaningful haemodynamic trajectories emerge shortly after admission and can be exploited for proactive rather than reactive care. A 6-hour actionable horizon balances the very short windows of some physiological predictors (e.g. 30–60 minutes [24]) against broader 7–21-hour early-warning horizons, and it aligns naturally with ICU shift handovers using standard non-invasive bedside measurements—an important consideration where continuous invasive monitoring is unavailable. Developed and validated entirely on data from a developing-country ICU, the system demonstrates that strong, interpretable performance is achievable outside the high-income databases that dominate the field, on a smaller cohort and with a deployable, internet-free software footprint.

Limitations

Several limitations temper these findings. The study was conducted at a single centre, limiting immediate generalisability. The cohort, although comprising thousands of temporal observations, contains only 253 patients, restricting exposure to rarer phenotypes; the exceptionally high discrimination should therefore be confirmed in larger, multi-centre cohorts before clinical reliance. The classification of patients as receiving no vasopressor support reflects the documented care record and resource context of this ICU and may not transfer to settings with different escalation practices. Finally, admission laboratories were treated as static; centres with repeated intra-stay laboratory sampling may derive additional temporal signal not captured here.

CONCLUSIONS

In a resource-constrained South Asian ICU, patients who deteriorated to haemodynamic instability without vasopressor support followed a reproducible trajectory in which vital signs—especially MAP, SBP, and shock index—declared circulatory compromise earlier than laboratory tests, with MAP breaching the 65 mmHg target by six hours and lactate and tachycardia persisting unabated. An interpretable, physiology-driven ensemble built on the same routinely collected data identified at-risk patients with high discrimination and good calibration from the first hours of admission. Together, the clinical characterisation and the model argue for intensified early vital-sign surveillance and a low threshold for haemodynamic escalation in patients presenting with shock index ≥ 1.0 or lactate > 2.0 mmol/L, and they support the feasibility of deployable, explainable early-warning tools for ICUs in resource-limited settings. Multi-centre and prospective external validation are the essential next steps.

Declarations

Ethics approval and consent to participate. The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics review committee of SIUT, Karachi.

Consent for publication. Not applicable; no individually identifiable data are presented.

Availability of data and materials. The datasets analysed during the current study are available from the corresponding author on reasonable request.

Competing interests. The authors declare that they have no competing interests.

Funding. No external funding was received for this study.

Authors' contributions. SSH conceived the study, supervised data collection, and drafted the clinical sections of the manuscript. FRH provided critical-care supervision, oversight of the clinical protocol, and senior revision. MH and T contributed to data collection, clinical data curation, and interpretation. MOA and MM designed and implemented the data pipeline, feature engineering, and machine-learning system, and produced the model analyses and figures. All authors read, critically revised, and approved the final manuscript.

Acknowledgements. The authors thank the ICU clinical and nursing staff at SIUT Karachi for maintaining the structured observation records used in this study, and Dr. Syed Zaffar Qasim (NEDUET) for academic supervision of the engineering team.

Abbreviations. AKI, acute kidney injury; ARDS, acute respiratory distress syndrome; AUROC, area under the receiver-operating-characteristic curve; CAP, community-acquired pneumonia; COPD, chronic obstructive pulmonary disease; CRT, capillary refill time; DBP, diastolic blood pressure; ESRD, end-stage renal disease; FiO_2, fraction of inspired oxygen; HAP, hospital-acquired pneumonia; HCO_3, serum bicarbonate; HR, heart rate; ICU, intensive care unit; LGB, LightGBM; LR, logistic regression; MAP, mean arterial pressure; NPV, negative predictive value; PPV, positive predictive value; RR, respiratory rate; SBP, systolic blood pressure; SHAP, SHapley Additive exPlanations; SI, shock index; SpO_2, peripheral oxygen saturation; SSC, Surviving Sepsis Campaign; TLC, total leucocyte count; UOP, urine output; XGB, XGBoost.

REFERENCES

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
2. Lee A, Bishop G, Hillman KM, Daffurn K. The medical emergency team. Anaesth Intensive Care. 1995;23(2):183–6.
3. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707–10.
4. Cecconi M, De Backer D, Antonelli M, et al. Consensus on circulatory shock and hemodynamic monitoring. Task Force of the European Society of Intensive Care Medicine. Intensive Care Med. 2014;40(12):1795–815.
5. Vincent JL, De Backer D. Circulatory shock. N Engl J Med. 2013;369(18):1726–34.
6. Rahman A, Chang Y, Dong J, et al. Early prediction of hemodynamic interventions in the intensive care unit using machine learning. Crit Care. 2021;25(1):388.
7. Jacquet-Lagrèze M, Wiart C, Schweizer R, et al. Capillary refill time for the management of acute circulatory failure: a survey among pediatric and adult intensivists. BMC Emerg Med. 2022;22(1):131.
8. Shoemaker WC. Temporal physiologic patterns of shock and circulatory dysfunction based on early descriptions by invasive and noninvasive monitoring. New Horiz. 1996;4(2):300–18.
9. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368–77.
10. Weil MH. Defining hemodynamic instability. In: Pinsky MR, Payen D, editors. Functional Hemodynamic Monitoring. Berlin: Springer; 2005. p. 9–17.
11. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43(3):304–77.
12. Colon Hidalgo D, Patel J, Masic D, et al. Delayed vasopressor initiation is associated with increased mortality in patients with septic shock. J Crit Care. 2020;55:145–8.
13. Li Y, Li H, Zhang D. Timing of norepinephrine initiation in patients with septic shock: a systematic review and meta-analysis. Crit Care. 2020;24(1):488.
14. Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med. 2020;26(3):364–73.
15. Al Aseri Z, Al Ageel M, Binkharfi M. The use of the shock index to predict hemodynamic collapse in hypotensive sepsis patients. Saudi J Anaesth. 2020;14(2):192–9.
16. Chiang DH, Cong T, Jiang Z, et al. External validation of a machine learning model to predict hemodynamic instability in the intensive care unit. Crit Care. 2022;26(1):215.
17. Nielsen LL, Cao H, Eshelman LJ, Gross B, et al. Hemodynamic instability prediction using continuous multi-parameter monitoring in the ICU. J Healthc Eng. 2010;1(4):509–26.
18. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med. 2015;7(299):299ra122.
19. Aliferis C, Simon G. Overfitting, underfitting, and model overconfidence: pitfalls and best practices in machine learning and AI. In: Health Informatics. Cham: Springer; 2024. p. 477–524.

20. Nieto N, Eickhoff SB, Jung C, et al. Impact of leakage on data harmonization in machine learning pipelines in class imbalance across sites. Neurocomputing. 2026;680:133146.
21. Zhang H, Qian D, Zhang X, et al. Tree-based ensemble machine learning models in the prediction of acute respiratory distress syndrome following cardiac surgery: a multicenter cohort study. J Transl Med. 2024;22:772.
22. Shashikumar SP, Josef C, Sharma A, Nemati S. DeepAISE – an interpretable and recurrent neural survival model for early prediction of sepsis. Artif Intell Med. 2021;113:102036.
23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30. 2017. p. 4765–74.
24. Yoon JH, Mu L, Chen L, et al. Predicting tachycardia as a surrogate for instability in the intensive care unit. J Clin Monit Comput. 2019;33(6):973–85.