

An Entropy-Weighted Biological Energy Index for Decomposable Mortality Prediction and Patient-Specific Tipping Dimension Identification in Sepsis: A Hypothesis-Generating Methodological Study

Gaurav Chandra^a, Kirsten Bischof^a, Suraj Saggar^b, Shubham Chandra^a

^aAdnexus Biotechnologies Inc., Dallas, Texas, USA

^bHoly Name Medical Center, Teaneck, New Jersey, USA

DOI:10.37819/Jaike.26.0015

Article history:

Received : 07.03.2026

Revised : 29.09.2026

Accept : 05.10.2026

Published : 21.10.2026

Corresponding author:

E-mail: gaurav.chandra@adnexusbiotech.com

Abstract: Sepsis remains a leading cause of global mortality, yet current severity scores provide only aggregate risk estimates without identifying the specific biological systems driving individual deterioration. This limits precision triage and mechanism-guided intervention. To develop and evaluate the entropy-weighted Biological Energy Index (BEI), an explainable AI framework that quantifies a composite biomarker reserve and identifies the dominant “tipping dimension” responsible for mortality risk in individual patients. Here “energy” is an information-theoretic construct derived from biomarker values; it is not a metabolic (ATP) or physical measurement. We analyzed 17 biomarkers from the Davoudian et al. (2022) sepsis cohort (N=135 from 178 enrolled; 34.3% 90-day mortality among 105 patients with confirmed outcome). Entropy-based weights were derived from the geometric mean of Shannon information gain, mutual information, and Spearman $|\rho|$, each computed on quintile-binned energy scores, with normalization to unit sum applied once after the geometric mean. BEI was computed as a weighted energy score with per-patient tipping-contribution decomposition. SHAP analysis examined independent prognostic signals. Discrimination was assessed out-of-fold under 5-fold stratified cross-validation with the entropy weights recomputed inside each training fold; bootstrap confidence intervals, leave-one-out cross-validation, calibration, decision-curve and survival analyses were also performed. A complete-case sensitivity analysis was restricted to the 45 patients with all 17 markers observed, of whom 33 had a confirmed 90-day outcome (13 deaths). BEI achieved an out-of-fold AUC=0.793, comparable to the equal-weighted BEI (AUC=0.799) — entropy weighting conferred no discriminative advantage under leakage-free cross-validation — and to standard classifiers trained on the same 17 markers: logistic regression (AUC 0.787), random forest (0.819) and XGBoost (0.797), while retaining full interpretability. In the complete-case sensitivity analysis the out-of-fold AUC was 0.885 (95% CI 0.752–0.981), an imprecise estimate that overlaps the primary result. Mortality exhibited a sharp nonlinear gradient: 86% in the lowest BEI quintile versus 19% in the highest (n=21 per quintile). A four-tier tipping taxonomy (KILL, MODERATE, NOISE, BENIGN) identified PTX3 as the most frequent tipping dimension (17 of 105 patients), suggesting endothelial collapse as a candidate early failure mode warranting prospective investigation. Any consciousness impairment (GCS ≤ 14) was associated with 67% mortality

(16/24). sIL-1R2 outperformed IL-18 for inflammasome-related mortality prediction (AUC 0.726 vs 0.610). BEI provides a transparent, decomposable framework for identifying per-patient biological failure modes in sepsis. Its contribution is interpretive rather than discriminative. All findings are derived from a single-center retrospective cohort and require external validation before clinical application. The framework is presented as a hypothesis-generating methodological contribution.

Keywords: Biological Energy Index (BEI), sepsis mortality, entropy weighting, tipping point analysis, explainable AI, precision ICU medicine, clinical decision support systems

1. INTRODUCTION

Sepsis — the life-threatening organ dysfunction caused by dysregulated host response to infection — affects an estimated 48.9 million individuals and causes 11 million deaths annually (Rudd et al., 2020). Despite decades of research, the 90-day mortality of hospitalized sepsis patients remains between 20% and 35% in academic centers (Singer et al., 2016). The core clinical challenge is stratifying patients with sufficient biological precision to match interventions to mechanisms.

Current severity scores — SOFA, APACHE-II, SAPS-III — provide aggregate estimates of organ dysfunction severity but do not answer the clinically decisive question: which biological system is failing fastest, and what should be done about it first? A patient with equivalent SOFA scores can be deteriorating from endothelial collapse (PTX3-driven), coagulation cascade failure (D-dimer-driven), or metabolic energy crisis (Lactate-driven) — conditions requiring mechanistically distinct interventions that no aggregate score can differentiate (Figure 1).

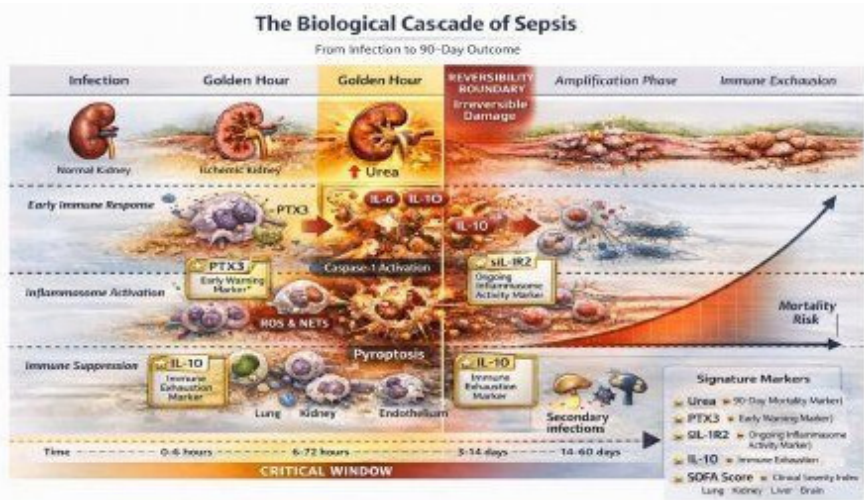


Figure 1. The biological cascade of sepsis from infection onset to 90-day outcome. Signature markers are positioned temporally: PTX3 (early warning, 0–72 h), sIL-1R2 (ongoing inflammasome activity, 3–60 days), IL-10 (immune exhaustion marker, persistent), Urea (90-day mortality marker), and SOFA Score (clinical severity index spanning all phases)

The sepsis trial failure rate over three decades is partly attributable to this heterogeneity problem. Trials targeting a single mediator consistently fail because they treat biologically distinct populations as one (Abraham et al., 2003; Bernard et al., 2001; Warren et al., 2001).

Entropy, as formalized by Shannon (1948), provides a natural framework for measuring information

content and systemic disorder. Applied to biomarker distributions, entropy-based weighting assigns higher discriminatory weight to markers whose variation carries more prognostic information. The combination of information gain, mutual information, and rank correlation as complementary signals for biomarker weighting has not been

previously applied to a composite clinical severity index.

The Biological Energy Index (BEI) was developed from this foundation: a composite measure of the reserve available to maintain homeostatic function, computed as the entropy-weighted sum of normalized biomarker energy scores. Terminology note: throughout this manuscript, “energy” is an information-theoretic construct — a normalized, direction-aligned score derived from measured biomarker values — and does not refer to ATP, metabolic equivalents or any other physical or metabolic quantity; for brevity the index is referred to as BEI hereafter. BEI is designed as a decomposable score that can be algebraically partitioned into per-marker contributions, enabling per-patient identification of the “tipping dimension”: the biological dimension most responsible for pulling an individual patient across the survival threshold.

This manuscript presents the BEI framework as a hypothesis-generating methodological study, incorporating sensitivity analyses, tipping taxonomy evaluation, inflammasome mechanism testing, GCS threshold characterization, and ablation studies across 17 clinical biomarkers measured in 135 sepsis patients (Davoudian et al., 2022). The work addresses the JAIKE theme of explainable AI frameworks for healthcare, demonstrating predictive ANALYTICS FOR CLINICAL DECISION SUPPORT WITH FULL MODEL TRANSPARENCY.

2. LITERATURE REVIEW

2.1. Severity Scoring and Its Limitations

SOFA score was introduced by Vincent et al. (1996) as a sequential organ failure assessment across six organ systems. While SOFA achieves AUC of 0.74–0.79 for ICU mortality prediction in external validation (Raith et al., 2017), its fundamental limitation is additivity without biological specificity. An equivalent SOFA score of 8 can arise from isolated renal failure, combined hepatic-respiratory failure, or pan-organ involvement — conditions with profoundly different prognoses. Serial rather than single-timepoint SOFA assessment improves outcome prediction (Ferreira et al., 2001), indicating that trajectory carries information beyond any admission value — a point returned to in the Conclusion. APACHE-II achieves AUC 0.80–0.85 in validation studies (Knaus et al., 1985) but does

not incorporate any cytokine or immunological dimension (Hotchkiss et al., 2016).

2.2. Cytokine Biomarkers in Sepsis

Conceptions of sepsis pathophysiology have been repeatedly revised over the past century, and biomarker interpretation has shifted with them (Ward & Bosmann, 2012). PTX3 is synthesized by endothelial cells and macrophages in response to inflammatory stimuli (Porte et al., 2019). Elevated PTX3 reflects endothelial barrier dysfunction and has been associated with sepsis severity and mortality in multiple cohorts (Huttunen et al., 2011; Mauri et al., 2010). sIL-1R2 is a decoy receptor reflecting NLRP3 inflammasome activity (Garlanda et al., 2013). IL-10 paradoxically associates with poor outcomes when persistently elevated, reflecting immunoparalysis (van der Poll et al., 2017). IL-6 has been shown to be prognostically redundant when sIL-1R2, IL-10, and PTX3 are available (Riedemann et al., 2003).

2.3. Information-Theoretic Approaches

Shannon entropy (Shannon, 1948) quantifies the average information content of a probability distribution. Applied to biomarker distributions, entropy-based methods have been used in genomics for gene selection (Vergara et al., 2014), multi-omics integration (Hasin et al., 2017), and physiological monitoring for sepsis (Seely & Macklem, 2004). Mutual information captures non-linear dependencies missed by linear correlation (Ross, 2014). The geometric mean of information gain, MI, and Spearman $|\rho|$ (Spearman, 1904) enforces broad-spectrum discriminatory power across all three criteria simultaneously.

2.4. Tipping Point Frameworks in Complex Systems

Tipping points — critical thresholds beyond which a system transitions to a qualitatively different state — have been studied extensively in ecology (Scheffer et al., 2009), climate science (Lenton, 2011), and neuroscience (Haken, 2002). Seely and Macklem (2004) proposed that critical illness represents a failure of the complex adaptive system of physiology. The present framework formalizes this concept: BEI is the system energy state, the threshold is the empirically identified survival boundary, and the tipping contribution decomposition identifies which subsystem is driving the transition.

2.5. Recent Advances in Explainable AI for Healthcare

Recent work in explainable AI for healthcare has emphasized the need for models that provide both strong performance and clinician-interpretable outputs. Lauritsen et al. (2020) demonstrated that an explainable early-warning model could provide clinically meaningful per-feature attributions for acute critical illness. Shickel et al. (2018) reviewed deep learning approaches to electronic health record analysis, identifying interpretability as a central barrier to clinical adoption. The TRIPOD+AI extension (Collins et al., 2024) formalized reporting standards for AI-based prediction models, requiring transparent documentation of model architecture, feature selection, and explainability methods. Knowledge-engineering work published in this journal has likewise emphasised transparent, interpretable AI-based decision support (Sewak, 2026). BEI addresses this gap by providing a model-free, fully decomposable scoring framework where every component of the final score has a direct biological interpretation.

3. METHODS

3.1. Study Population

The dataset was collected and published by Davoudian et al. (2022) under a Creative Commons Attribution

(CC BY) license. The original study enrolled 178 patients with suspected sepsis at the Emergency Department of IRCCS Humanitas Research Hospital, Milan, Italy. Of these, 6 were excluded pre-assessment and 172 were stratified by Sepsis-3 criteria (Singer et al., 2016): 37 patients with SOFA < 2 were classified as non-sepsis, yielding N=135 patients with sepsis (n=99) or septic shock (n=36) in the published dataset (Figure 2). The primary outcome was 90-day all-cause mortality. Vital status was confirmed in 105 patients (77.8%); 30 (22.2%) had missing outcome data and were excluded from all outcome analyses. Among confirmed patients, 36 deaths were recorded (34.3% mortality). All results are based on this 105-patient analytical cohort unless otherwise stated; every analysis set is enumerated in §3.6.

Biomarkers were measured at ICU admission (Day 0) and Day 5 for paired temporal analyses. Day 0 cytokine panels were 100% complete across all 135 patients (9 cytokines measured by ELLA automated immunoassay). Day 5 samples were available for 79 patients (58.5%); 56 patients lacked Day 5 data, including 15 who died before collection was possible, 40 who were alive but had no sample obtained, and 1 whose status was unknown (Figure 2). This survivorship bias in temporal data is explicitly acknowledged: the sickest patients who died earliest are structurally excluded from Day 0 to Day 5 trajectory comparisons.

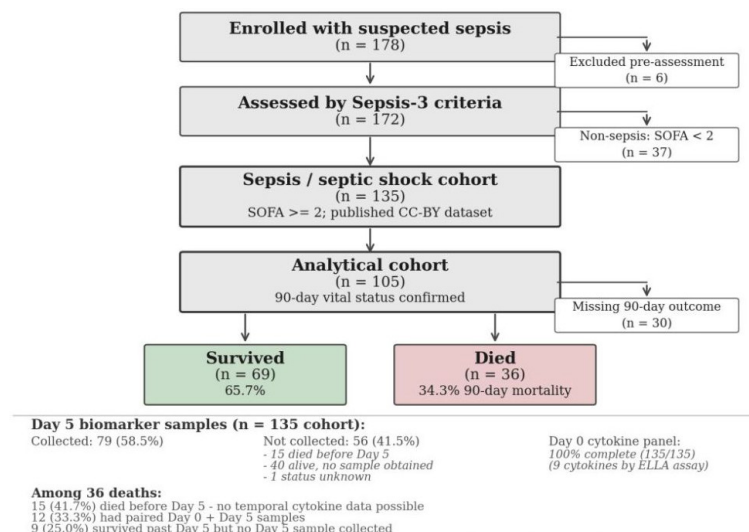


Figure 2. Participant flow diagram (TRIPOD+AI Item 3a). Of 178 patients enrolled with suspected sepsis, 37 were classified as non-sepsis (SOFA < 2), yielding the 135-patient published cohort. Thirty patients lacked confirmed 90-day vital status, leaving an analytical cohort of 105 (36 deaths, 34.3% mortality)

Seventeen biomarkers were analyzed: SOFA, Lactate, PTX3, Urea, IL-10, IL-8, GCS, sIL-1R2, MAP, PaO₂/FiO₂, D-dimer, IL-6, CRP, IL-18, IL-1Ra, TNF-α, and Platelets. Missing data rates varied from 0% (SOFA, GCS) to 47% (D-dimer). The original study was approved by the Institutional Review Board of IRCCS Humanitas Research Hospital (approval no. 820/18; Davoudian et al., 2022); the present work is a secondary analysis of the de-identified public dataset (see Ethics Approval and Consent to Participate).

$$v_m(i) = \log_{10}(x_m(i) + 1); s_m(i) = [v_m(i) - \min v_m] / [\max v_m - \min v_m] \quad (1)$$

$$E_m(i) = 1 - s_m(i) \text{ for failure markers}; E_m(i) = s_m(i) \text{ for protective markers} \quad (2)$$

Choice of scaling and sensitivity to outliers. Min–max scaling was chosen because it places every marker on a common, bounded [0,1] energy scale, so that per-marker energy scores, the weighted BEI and the tipping contributions of Equations (5)–(6) are directly comparable across markers and interpretable as fractions of the observed cohort range. The log₁₀(x+1) transform is applied first to compress the right-skewed tails typical of cytokine and acute-phase marker distributions; no winsorisation, clipping or outlier removal was applied beyond this transform. Residual outliers affect the pipeline asymmetrically. The entropy weights are unaffected: min–max scaling is monotone within each marker, and all three weighting signals (§3.3) — information gain and mutual information on quantile bins, and Spearman |ρ| — depend only on the rank order of the energy scores, so the weights are identical under min–max, robust (e.g., 5th–95th percentile) or any other monotone scaling. The BEI score and the tipping contributions are not invariant: a single extreme value compresses the remaining patients' energy scores for that marker into a narrower band, reducing that marker's spread within the composite and the magnitude of its tipping contribution. Robust

3.2. Biomarker Energy Normalization

Each biomarker was log-transformed (log₁₀(x+1)) then min-max normalized to [0,1]. Directional alignment ensured an energy score of 1.0 represents intact function: failure markers (SOFA, Lactate, Urea, PTX3, D-dimer, IL-10, IL-8, IL-6, IL-1Ra, TNF-α, IL-18, CRP, sIL-1R2) were inverted; protective markers (GCS, MAP, PaO₂/FiO₂, Platelets) were retained as-is (Equations 1–2; Algorithm 1, Step 1).

percentile scaling with clipping would limit this effect at the cost of saturating the most extreme patients; this sensitivity is acknowledged in §5.6.

3.3. Entropy Weight Computation

Three information-theoretic signals were computed for each biomarker *m* against the binary 90-day mortality outcome *Y*: (1) Information Gain: IG(*m*) = H(*Y*) – H(*Y*|*m*), with continuous markers discretised into five quantile bins (quintiles); (2) Mutual Information: estimated with the discrete mutual-information estimator (scikit-learn `mutual_info_score`) applied to the same quantile-binned energy scores; (3) Spearman |ρ|: absolute Spearman rank correlation between energy score and mortality (Spearman, 1904). The composite weight was computed as the geometric mean of the three raw signals, $W(m) = ((IG(m)+\epsilon) \cdot (MI(m)+\epsilon) \cdot (|\rho(m)|+\epsilon))^{1/3}$ with $\epsilon = 1 \times 10^{-6}$, and the resulting weights *W*(*m*) were then normalised to sum to 1.0 across the 17 markers. Geometric mean requires a marker to score well across all three criteria simultaneously, preventing single-signal artifacts from dominating.

$$g(m) = [(IG(m)+\epsilon) \cdot (MI(m)+\epsilon) \cdot (|\rho(m)|+\epsilon)]^{1/3}, \epsilon = 1 \times 10^{-6} \quad (3)$$

$$W(m) = g(m) / \sum_k g(k), \text{ with } \sum_m W(m) = 1 \quad (4)$$

Three properties of the implemented procedure are stated explicitly. First, the three signals enter the geometric mean in their raw form: no per-signal rescaling is applied beforehand, and normalization to unit sum is applied once, to *g*(*m*), in Equation (4). Second, both IG and MI are computed on the same five quantile bins of the energy score

(KBinsDiscretizer, strategy='quantile', n_bins=5); no continuous mutual-information estimator is used. Third, all quantities in Equations (3)–(4) are recomputed on the training partition alone within each cross-validation fold (§3.5), so no held-out outcome information enters weight estimation. The in-sample weights (reported in §4.1 and §4.4) are

used only for interpretive decomposition, never for the reported discrimination estimates.

3.4. BEI Formula and Tipping Contribution

$BEI(i) = \Sigma[W(m) \times Energy_m(i)] / \Sigma[W(m)]$ for non-missing markers in patient i . Only available markers contribute to both numerator and denominator. A patient's BEI is computed only when at least 60% of the 17 markers are observed (minimum-coverage

threshold $\tau = 0.60$); patients below this coverage are left unscored. All 105 analytical-cohort patients met this threshold and were scored, which is why the quintile ($n=21$ each; §4.1) and energy-tier ($n=35$ each; §4.8) strata sum to 105. The tipping contribution of marker m for patient i is: $TC(m, i) = W(m) \times [Energy_m(i) - Cohort_mean(Energy_m)]$. The marker with the largest negative TC is designated the tipping marker. The complete pipeline is specified in Algorithm 1 (§3.9).

$$BEI(i) = \Sigma_{\{m \in O(i)\}} W(m) \cdot E_m(i) / \Sigma_{\{m \in O(i)\}} W(m), \text{ computed when } |O(i)|/17 \geq \tau = 0.60 \quad (5)$$

$$TC(m, i) = W(m) \cdot [E_m(i) - \text{mean}_i E_m]; \text{ tipping}(i) = \text{argmin}_m TC(m, i) \quad (6)$$

3.5. Statistical Assessment and Explainability

Discriminative performance was evaluated using AUC with 95% bootstrap confidence intervals (1000 iterations). Generalization was assessed via 5-fold stratified and leave-one-out cross-validation. All reported BEI AUCs are out-of-fold: within each of the five stratified folds, the entropy weights of Equations (3)–(4) are recomputed on the training partition only and applied to the held-out partition, so the reported discrimination carries no target leakage. Survival analysis used Kaplan-Meier estimators with log-rank testing. Cox proportional hazards regression used lifelines with penalizer=0.1. Calibration was assessed on the out-of-fold score using Hosmer–Lemeshow goodness-of-fit, the calibration slope, calibration-in-the-large, the Brier score and calibration plots; predicted probabilities were obtained from a logistic recalibration map fitted within training folds only, so the calibration assessment is leakage-free (§4.1). Decision-curve analysis evaluated clinical utility across threshold probabilities from 10% to 80%.

SHAP feature attribution used a Random Forest classifier (500 trees, class_weight="balanced", random_state=42). SHAP values estimated marginal contributions of each biomarker, complementing Cox regression, in which collinearity can suppress individual signals. Systematic ablation across nine feature combinations quantified predictive lift from cytokines versus clinical variables; the ablation models are Random Forest classifiers fitted to the stated feature subsets and are independent of the BEI weighting scheme.

3.6. Missing Data Strategies and Sample-Size Accounting

Three parallel strategies assessed robustness. Strategy 1 (primary — dynamic weight normalisation): BEI is computed from the markers available for each patient, and the entropy weights of those markers are rescaled to sum to 1.0 (the denominator of Equation 5). Strategy 2 (sensitivity — complete-case analysis): BEI restricted to the 45 patients with all 17 markers observed, of whom 33 have a confirmed 90-day outcome (13 deaths). Strategy 3 (sensitivity — multiple imputation): MICE with 20 imputed datasets.

A single sample-size accounting is used throughout and is set out in Table 1. Two distinctions matter. First, the complete-case set has two sizes: 45 patients in the published cohort have all 17 markers observed, of whom 33 also have a confirmed 90-day outcome; every outcome-dependent complete-case result (AUC, confidence interval, event count) is based on $n=33$. Second, the figure 57 denotes the number of patients with a missing Lactate measurement (42% of 135); it is not a complete-case sample size.

Lactate missingness (42%, $n=57$ of 135) was associated with lower illness severity: lactate-missing patients had lower SOFA scores than lactate-present patients (median 4 vs 5, $p=0.009$; logistic odds ratio 0.81 per SOFA point, 95% CI 0.70–0.95, $p=0.008$) and lower 90-day mortality (31% vs 37%), consistent with clinician-driven selective ordering in which Lactate is preferentially measured in the most acutely ill patients (MNAR). D-dimer missingness (47%) followed the same direction (31% vs 38% mortality in missing vs present patients). A composite score derived from different marker subsets across patients

is not strictly the same instrument; the complete-case analysis therefore serves as a sensitivity analysis in which all patients are scored on identical marker sets.

Set	n	Deaths	Used for
Enrolled with suspected sepsis	178	—	Screening (Figure 2)
Assessed by Sepsis-3 criteria	172	—	6 excluded pre-assessment
Non-sepsis (SOFA < 2), excluded	37	—	Excluded from all analyses
Published sepsis / septic shock cohort	135	—	Normalization ranges (Eq. 1); missingness analyses; sepsis n=99, septic shock n=36
Missing 90-day vital status	30	—	Excluded from outcome analyses
Analytical cohort (outcome confirmed)	105	36 (34.3%)	All primary results: entropy weights (Eq. 3–4, recomputed within training folds), AUC, quintiles, tipping taxonomy, Kaplan–Meier, Cox, GCS strata
Marker-complete patients (all 17 markers)	45	—	Definition of the complete-case instrument
Complete-case sensitivity cohort (marker-complete and outcome-confirmed)	33	13	Complete-case sensitivity AUC 0.885 (95% CI 0.752–0.981) only
Lactate missing	57 (42% of 135)	—	Missingness / MNAR analysis only — not a complete-case set
D-dimer missing	63 (47% of 135)	—	Missingness analysis only
Day 5 samples collected	79 (58.5% of 135)	—	Day 0 → Day 5 temporal comparisons only

3.7. Sample Size and Events-Per-Variable

The cohort comprises 36 events. For the primary BEI composite (single score vs. binary outcome), $EPV=36/1=36$, above the threshold of 10. For 17-marker models, $EPV=36/17=2.1$, below the recommended minimum. This limitation is addressed by: (1) cross-validation providing honest performance estimates; (2) SHAP importance from Random Forest, less sensitive to EPV than linear regression; (3) individual marker AUCs reported with appropriately wide bootstrap CIs. In the complete-case sensitivity analysis the event count falls to 13 ($EPV=13$ for the composite); the correspondingly wide confidence interval is reported in §4.1 and the analysis is not treated as validation. The BEI composite is the primary instrument; individual 17-marker models are exploratory, and all clinical observations from this study are hypotheses requiring prospective validation.

3.8. Baseline AI Model Comparison

To contextualize BEI performance, logistic regression, random forest and XGBoost classifiers were trained on the same 17 biomarkers under identical 5-fold stratified cross-validation, with median imputation of missing values inside each training fold. BEI achieved an out-of-fold $AUC=0.793$, comparable to all three: logistic regression 0.787 (SD 0.102), random forest 0.819 (SD 0.096) and XGBoost 0.797 (SD 0.086). These three models are fitted to the 17 BEI markers and are distinct from the ablation models of §4.9, which are fitted to other feature subsets and are not directly comparable with them. Unlike these models, BEI provides full per-patient decomposability and tipping dimension identification without requiring model retraining.

3.9. Formal Algorithm Specification

The complete BEI pipeline — energy normalization, entropy-weight computation, scoring, tipping decomposition, taxonomy and risk stratification — is specified formally in Algorithm 1.

Formal specification of the implemented pipeline (Python 3.10.12; scikit-learn 1.3.0, lifelines 0.27.8; random seed = 42), consistent with §3.2–§3.4 and §4.4, including Equations (1)–(6). Comments (#) are placed on their own lines.

Algorithm 1. Entropy-Weighted Biological Energy Index (BEI) and Tipping-Contribution Decomposition

```

INPUTS
N patients; M = 17 Day-0 biomarkers; binary 90-day mortality Y in {0,1}
direction d_m in {failure, protective}; coverage threshold tau = 0.60
STEP 1 Biomarker energy normalization (per marker m, observed values)
v_m(i) = log10( x_m(i) + 1 )
s_m(i) = ( v_m(i) - min_i v_m ) / ( max_i v_m - min_i v_m )
E_m(i) = 1 - s_m(i) if d_m = failure
E_m(i) = s_m(i) if d_m = protective
STEP 2 Entropy weights (per marker m, patients with observed E_m and Y)
# quantile bins: KBinsDiscretizer, strategy = "quantile"
b_m = quantile_bins( E_m, n_bins = 5 )
# information gain on binned E_m
IG_m = H(Y) - H(Y | b_m)
# discrete mutual_info_score on the same bins
MI_m = I( b_m ; Y )
rho_m = | Spearman( E_m , Y ) |
# the three signals enter g_m raw (no per-signal rescaling)
g_m = ( (IG_m + eps)(MI_m + eps)(rho_m + eps) )^(1/3), eps = 1e-6
# normalisation applied once, to g_m
W_m = g_m / sum_k g_k
STEP 3 BEI score (per patient i)
O_i = { m : E_m(i) observed }
# require >= 60% marker coverage
if |O_i| / M < tau : BEI(i) = NA
else : BEI(i) = sum_{m in O_i} W_m E_m(i) / sum_{m in O_i} W_m
STEP 4 Tipping contribution and tipping marker (per patient i)
Ebar_m = mean_i E_m(i)
TC_m(i) = W_m * ( E_m(i) - Ebar_m )
tipping_marker(i) = argmin_m TC_m(i)
STEP 5 Tipping-dimension taxonomy (per marker m, patients tipped by m)
MortRate_m = mean( Y | tipping_marker = m ); N_m = count
Tier_m = KILL if MortRate_m >= 0.40
MODERATE if 0.25 <= MortRate_m < 0.40
NOISE if 0.20 < MortRate_m < 0.25
BENIGN if MortRate_m <= 0.20
# markers with N_m = 0 are left unclassified
# (IL-1Ra, TNF-alpha, IL-6, IL-18 in this cohort)
STEP 6 Risk stratification (reporting)
Energy tiers : BEI tertiles (33rd / 67th cohort percentiles)
-> Very Low / At-Risk / Sufficient
Quintiles : Q1..Q5 of BEI; 90-day mortality per quintile
Discrimination: risk score = (1 - BEI) vs Y
# out-of-fold: in each of 5 stratified CV folds, W_m (Step 2) is
# recomputed on the training partition only and applied to the
# held-out partition, so the AUC carries no target leakage

```

Algorithm 1. Continued

```

Uncertainty      : 1000x bootstrap 95% CIs; leave-one-out CV also reported
OUTPUTS
W_m; BEI(i); TC_m(i); tipping_marker(i); tier assignments;
quintile and energy-tier mortality

```

3.10. Reproducibility

All analyses were implemented in Python 3.10.12 (scikit-learn 1.3.0, shap 0.42.1, lifelines 0.27.8). Random seeds fixed at 42 throughout. Algorithm 1 (§3.9) provides the complete formal specification. Three scripts accompany the submission and together regenerate every reported value. BEI_Manuscript_Figures_Standalone.py produces the entropy weights, BEI scores, out-of-fold ROC, quintile and energy-tier stratification, tipping taxonomy, individual-marker AUCs, Cox model, ablation and Kaplan–Meier curves. reproduce_reported_values.py produces the baseline classifiers, the complete-case sensitivity analysis, the missingness statistics and the Figure 4 right-shift test. BEI_R5_phenotype_and_calibration.py produces the Golden Hour signature AUCs of §4.3, and the calibration, decision-curve and multiple-imputation robustness results of §4.1 and §4.4. The scripts are supplied with the submission as separate code files. Patient-level source data are not reproduced in this article; they are publicly available from the original authors' Zenodo repository (record 7612571; Davoudian et al., 2022). Code is archived at [DOI to be assigned on acceptance].

4. RESULTS

4.1. BEI Framework and Discriminative Performance

Under strict 5-fold cross-validation with entropy weights recomputed inside each training fold and applied out-of-fold, the entropy-weighted BEI achieved an out-of-fold AUC=0.793 for 90-day mortality prediction, comparable to the equal-weighted BEI (AUC=0.799); under leakage-free cross-validation, entropy weighting did not confer a discriminative advantage over equal weighting (Figure 3B). A BEI recomputed with Lactate excluded for all patients is reported in the missing-data analysis (AUC=0.805; Section 4.2). In a complete-case sensitivity analysis restricted to the 33 patients with all 17 markers observed and a confirmed outcome (of the 45 marker-complete patients; 13 deaths; Table 1), the out-of-fold AUC was 0.885 (95% CI 0.752–0.981); the wide confidence interval, expected at this sample size, overlaps the primary estimate and indicates that the fixed-instrument score remains discriminative without implying superior performance.

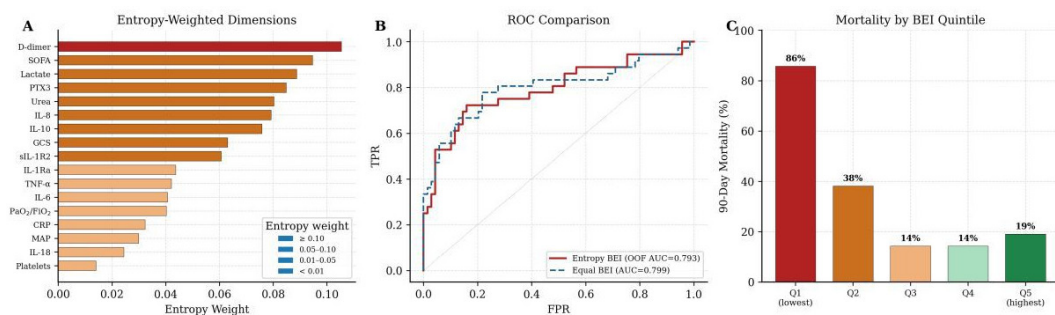


Figure 3. BEI framework and discriminative performance (analytical cohort, $n=105$). (A) Entropy-weighted dimensions for all 17 biomarkers, coloured by entropy-weight magnitude; the eight highest-weighted markers carry approximately two-thirds of the total weight. (B) ROC curve comparison: entropy-weighted BEI (out-of-fold) vs equal-weighted BEI. (C) 90-day mortality by BEI quintile ($n=21$ per quintile), stratified on the out-of-fold score

The quintile dose-response (Figure 3C) demonstrated a pronounced non-linear pattern: Q1 patients (lowest energy) carried 86% mortality versus 38% in Q2, declining to 14% in Q3 and Q4. The Q5 anomaly (19% mortality) is addressed in Section 4.2. Each quintile contains 21 patients (18, 8, 3, 3 and 4 deaths from Q1 to Q5, summing to the 36 cohort deaths). The eight highest-weighted dimensions (D-dimer, SOFA, Lactate, PTX3, Urea, IL-8, IL-10, GCS) accounted for approximately two-thirds of the total entropy weight (Figure 3A; per-marker values are tabulated with the taxonomy in §4.4).

Calibration was assessed on the out-of-fold score. Predicted probabilities were obtained from a logistic recalibration map whose coefficients were fitted on the training partition of each fold and applied to the held-out partition, so neither the entropy weights nor the recalibration used held-out outcome information. On this leakage-free basis the calibration slope was 0.850 (95% CI 0.46–1.56, 2000 bootstrap resamples), calibration-in-the-large was -0.085 and the Brier score was 0.168, while the Hosmer–Lemeshow test rejected ($\chi^2=20.20$, $df=8$, $p=0.010$). Calibration is therefore not adequate at this sample size: the score ranks patients well but its absolute risk estimates are not reliable, and BEI must not be read as a calibrated probability of death without refitting in the deployment population (§5.6). Decision-curve analysis on the same out-of-fold

probabilities showed positive net benefit over treatment for threshold probabilities of 20% and above, with the maximum advantage at 40%.

4.2. Lactate Sensitivity Analysis

Lactate-missing patients had higher BEI scores than Lactate-present patients (Mann-Whitney $p=0.001$; median 0.70 vs 0.62, Figure 4A). Two factors contribute. First, the lactate-missing group was genuinely somewhat less severely ill (Section 3.6), with lower 90-day mortality (31% vs 37%). Second, the dynamic denominator introduces a mechanical right-shift: omitting Lactate — a failure marker whose elevated values lower the energy score — removes a downward contribution and raises BEI for patients in whom it was not measured. Within the highest BEI quintile, 90-day mortality was similar whether Lactate was measured (20%, $n=5$) or missing (19%, $n=16$; together the 21 patients of Q5; Figure 4B), so the elevated-quintile pattern is not attributable to Lactate missingness alone. Recomputing BEI with Lactate excluded for every patient, so that all patients are scored on an identical 16-marker panel, gives an out-of-fold AUC of 0.805. Because the score depends on which markers are available, Lactate should be measured routinely in any prospective deployment so that patients are compared on a consistent panel.

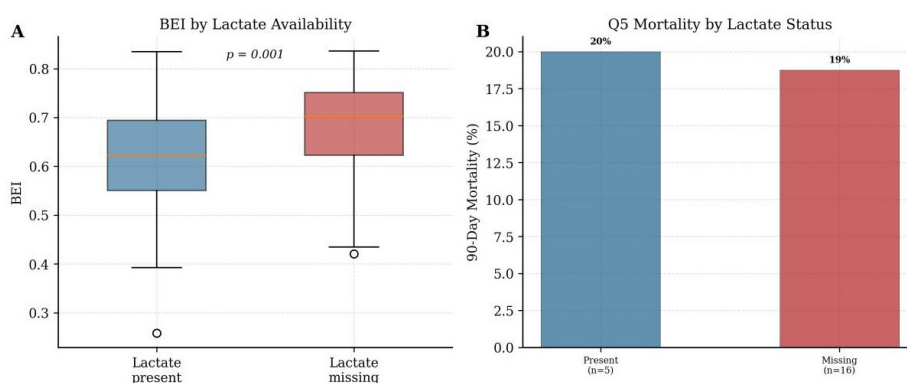


Figure 4. Sensitivity analysis for Lactate missingness bias (analytical cohort, $n=105$). (A) BEI distribution by Lactate availability (Mann-Whitney $p=0.001$). (B) Q5 mortality by Lactate status (present $n=5$, missing $n=16$)

4.3. Golden Hour Signature

The five-marker Golden Hour Signature (GH5: Urea + IL-10 + sIL-1R2 + PTX3 + SOFA) achieved a 5-fold cross-validated AUC of 0.801 (SD 0.091)

and a leave-one-out AUC of 0.786; removing SOFA lowered these to 0.763 and 0.768 respectively, so SOFA is retained wherever computable. Adding IL-8 to the signature made no material difference (5-fold

0.803). Individual-marker AUCs for all 17 markers are shown in Figure 5. PTX3 temporal kinetics showed persistent elevation in non-survivors from Day 0 to Day 5, consistent with sustained endothelial damage — a state-dependent signal. IL-8 in non-survivors decreased from Day 0 to Day 5, consistent with neutrophil compartment exhaustion rather than resolution, making it potentially misleading as a Day 0 triage marker for late mortality.

4.4. Tipping Dimension Taxonomy

The four-tier tipping taxonomy classifies markers by associated mortality when they are the primary tipping dimension (Figure 6). Using the cutoffs KILL $\geq 40\%$, MODERATE 25–40%, NOISE 20–25%, and BENIGN $\leq 20\%$, markers are assigned as follows. KILL dimensions: GCS (~67%, N=3), D-dimer (~50%, N=8), SOFA (~46%, N=13), PaO₂/FiO₂ (~45%, N=11), Lactate (~44%, N=9), and PTX3 (~41%, N=17); the highest-mortality assignments (e.g., GCS) rest on small numbers of tipped patients and are reported as hypothesis-generating. PTX3 was

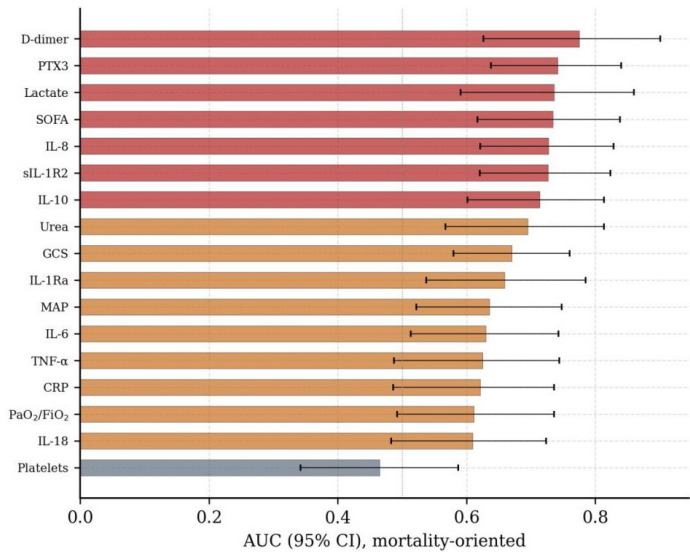


Figure 5. Individual marker AUC ranking with 95% bootstrap confidence intervals for 90-day mortality prediction (mortality-oriented; analytical cohort, n=105)

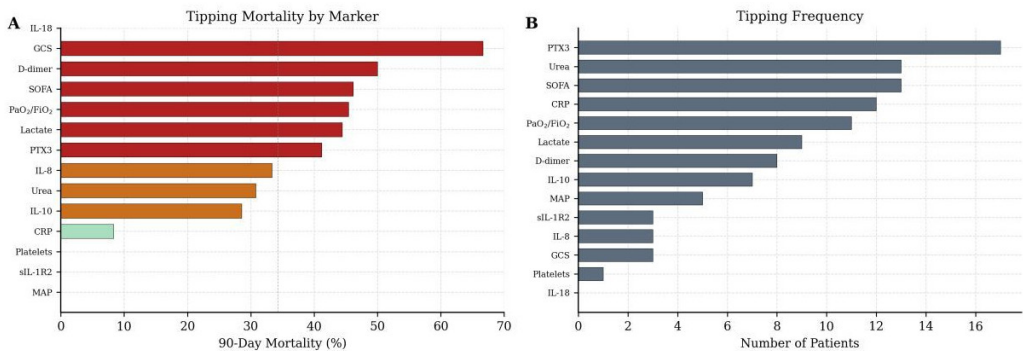


Figure 6. Tipping dimension taxonomy (analytical cohort, n=105). (A) 90-day mortality rate among patients for whom each marker is the primary tipping dimension, sorted by mortality rate and coloured by operational tier (tier cutoffs: KILL $\geq 40\%$, MODERATE 25–40%, NOISE 20–25%, BENIGN $\leq 20\%$). (B) Tipping dimension frequency; per-marker N is listed in Table 2. Markers with N=0 (IL-18 shown; IL-1Ra, TNF-α and IL-6 omitted) are unclassified

the most frequent tipping marker ($n=17$), suggesting it warrants investigation as a monitoring target in prospective studies. MODERATE dimensions: IL-8 (~33%, $N=3$), Urea (~31%, $N=13$), and IL-10 (~29%, $N=7$). No marker fell in the NOISE band (20–25%) in this cohort. BENIGN dimensions: CRP (~8%, $N=12$), sIL-1R2, MAP, and Platelets (all 0%; $N=3$, 5 and 1 respectively).

Four markers — IL-1Ra, TNF- α , IL-6 and IL-18 — were never the primary tipping dimension in this cohort ($N=0$) and are therefore left unclassified rather than assigned to a tier; IL-18 appears on the axis of Figure 6 with $N=0$ and no mortality bar. The thirteen classified markers together account for all 105 analytical-cohort patients and all 36 deaths (Table 2), which provides an internal check that the decomposition partitions the cohort exactly once.

Marker	Entropy weight	Patients tipped (N)	Tipping-associated 90-day mortality
KILL (tipping mortality • 40%)			
GCS	0.063	3	67%
D-dimer	0.106	8	50%
SOFA	0.095	13	46%
PaO ₂ /FiO ₂	0.040	11	45%
Lactate	0.090	9	44%
PTX3	0.085	17	41%
MODERATE (25–40%)			
IL-8	0.079	3	33%
Urea	0.080	13	31%
IL-10	0.077	7	29%
NOISE (20–25%) — no marker in this cohort			
BENIGN (• 20%)			
CRP	0.032	12	8%
MAP	0.031	5	0%
sIL-1R2	0.061	3	0%
Platelets	0.014	1	0%
Not classified (never the primary tipping dimension; N = 0)			
IL-1Ra	0.044	0	—
TNF- α	0.042	0	—
IL-6	0.041	0	—
IL-18	0.024	0	—
Total	1.000	105	36 deaths

Patients tipped (N) is the number of patients in the analytical cohort ($n = 105$) for whom the marker is the argmin tipping dimension; mortality is computed within that subset. Tier assignments resting on few tipped patients (e.g., GCS and IL-8, $N = 3$) are hypothesis-generating and should not be read as stable. PTX3 was the most frequent tipping dimension ($N = 17$); no marker fell in the NOISE

band; IL-1Ra, TNF- α , IL-6 and IL-18 were never the primary tipping dimension and are left unclassified.

Tier classification was assessed across all three missing-data strategies. Under multiple imputation (20 imputed datasets; §3.6, Strategy 3) the KILL tier was stable, losing 0.1% of its assignments on average (range 0.0–1.6%). The identity of the individual tipping marker was markedly less stable, changing for 18.9% of the 105 patients on average

(range 13.3–23.8%). This is consistent with the small per-marker subgroups and reinforces that per-patient marker assignments are hypothesis-generating, even where the tier is not.

4.5. Cox Model and SOFA vs Age Analysis

Cox regression identified Urea (HR $\approx$ 1.66, $p=0.044$) and age (HR $\approx$ 1.05, $p=0.002$) as the independently significant predictors, with the remaining covariates not reaching significance (Figure 7), consistent with collinearity. SHAP analysis complemented this: sIL-1R2, Urea, IL-10, and PTX3 all demonstrated substantial independent contributions within the Random Forest model. The value of SOFA within the Golden Hour Signature is quantified in §4.3.

4.6. NLRP3 Inflammasome Analysis

sIL-1R2 alone achieved AUC=0.726 for 90-day mortality, while IL-18 alone achieved AUC=0.610 (Figure 8A). Combining both (AUC=0.725) provided no additive value, indicating that IL-18 is informationally redundant conditional on sIL-1R2 in this cohort. The Spearman $\rho=0.30$ ($p=0.002$) indicates moderate correlation. Temporal kinetics revealed distinct trajectories: sIL-1R2 in non-survivors maintained elevated levels while IL-18 increased from Day 0 to Day 5, suggesting these markers capture mechanistically distinct aspects of inflammasome biology.

4.7. GCS as a Threshold Alarm

GCS 13–14 ($n=15$) carried 67% 90-day mortality — identical to GCS ≤ 12 (67%, $n=9$) — whereas GCS=15

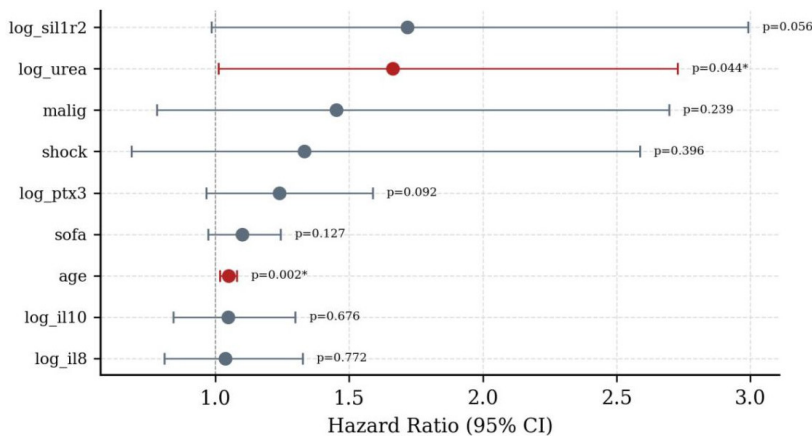


Figure 7. Cox proportional hazards forest plot ($n=105$; penalizer=0.1). Urea (HR $\approx$ 1.66, $p=0.044$) and age (HR $\approx$ 1.05, $p=0.002$) reach significance, marked with an asterisk; all other covariates are non-significant, reflecting collinearity

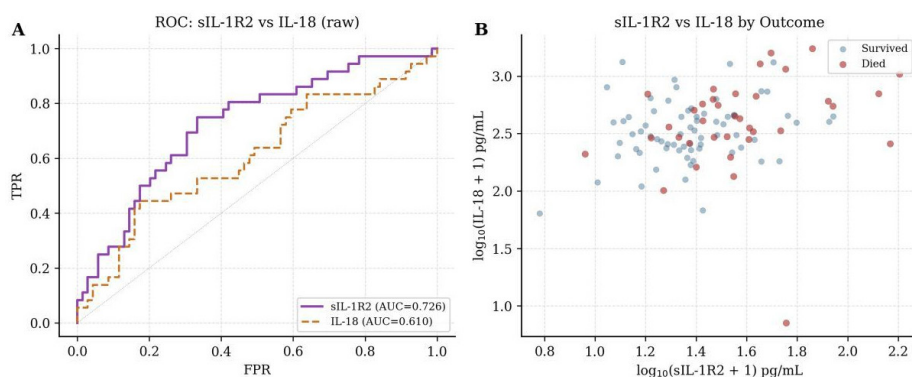


Figure 8. NLRP3 inflammasome analysis (analytical cohort, $n=105$). (A) ROC comparison: sIL-1R2 (AUC=0.726) vs IL-18 (AUC=0.610). (B) Scatter plot of sIL-1R2 vs IL-18 by outcome

(n=81) carried 25% mortality (9 + 15 + 81 = 105 patients; 6 + 10 + 20 = 36 deaths; any impairment, GCS ≤ 14 , therefore corresponds to 16 deaths among 24 patients, 67%). This step-change between GCS=15 and GCS ≤ 14 challenges the conventional ≤ 12 threshold (Figure 9). Continuous GCS (AUC=0.670) outperformed all binary thresholds, with GCS ≤ 12 performing worst (AUC=0.562). These observations suggest that any reduction from full consciousness warrants heightened monitoring, though this finding requires validation in larger, independent cohorts before informing clinical thresholds.

4.8. Kaplan-Meier Survival Analysis

BEI energy tier stratification produced significant survival separation (log-rank $p < 0.001$); the three tiers contain 35 patients each and carry 71%, 17% and 14% 90-day mortality from Very Low to Sufficient (Figure 10). PTX3 tertile stratification (log-rank $p = 0.0001$) demonstrated steep early mortality in the high tertile (Days 0–15) followed by plateau, while sIL-1R2 stratification showed gradual attrition through Day 90 — consistent with distinct temporal mortality signatures.

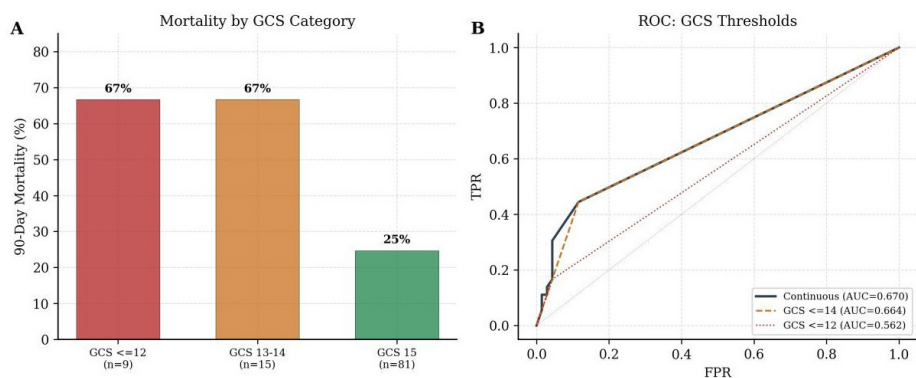


Figure 9. GCS as a sepsis mortality threshold (analytical cohort, n=105). (A) 90-day mortality by GCS category (n=9, 15 and 81). (B) ROC curves for continuous vs binary GCS thresholds

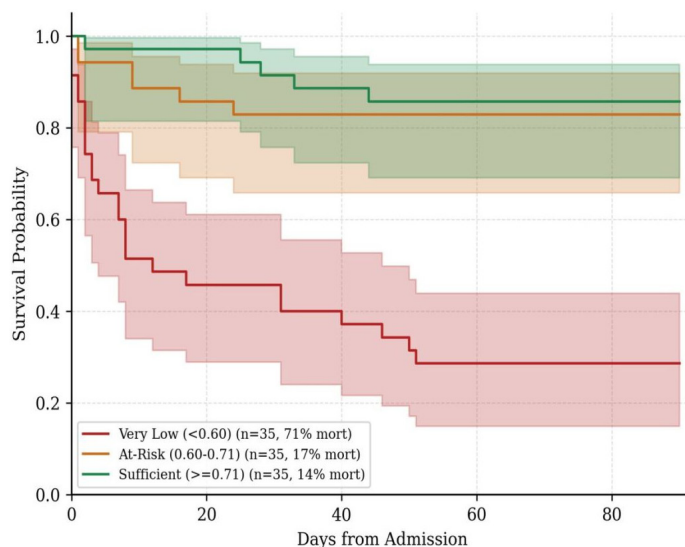


Figure 10. Kaplan-Meier survival by BEI energy tier (three BEI energy tiers; n=105; n=35 per tier; tier boundaries are the 33rd and 67th cohort percentiles of the out-of-fold score). BEI stratification produces significant survival separation (log-rank $p < 0.001$); shaded bands are 95% confidence intervals

4.9. Ablation Study

Systematic ablation across nine feature combinations (Figure 11), each fitted as a random forest under identical 5-fold stratified cross-validation, showed that the combined Labs + Cytokines set (Urea, SOFA, GCS, PTX3, sIL-1R2, IL-10, IL-8, IL-6) achieved the highest mean AUC at 0.832 (SD 0.082). It exceeded the broadest set evaluated — 15 of the 17 BEI markers together with age, septic shock and malignancy, at 0.802 (SD 0.077) — and the cytokine-only set at 0.756 (SD 0.096). Clinical variables alone (SOFA, GCS, MAP, PaO₂/FiO₂)

reached 0.777 and comorbidities alone 0.596. The broadest set is neither the 17-marker BEI panel nor the 17-marker random forest of §3.8; the exact composition of every ablation set is given in the released code. The competitive performance of compact laboratory–cytokine combinations is consistent with the entropy weighting concentrating discriminatory information in a subset of markers. Error bars overlap substantially across feature sets (Figure 11), so the ordering of the intermediate sets should not be over-interpreted.

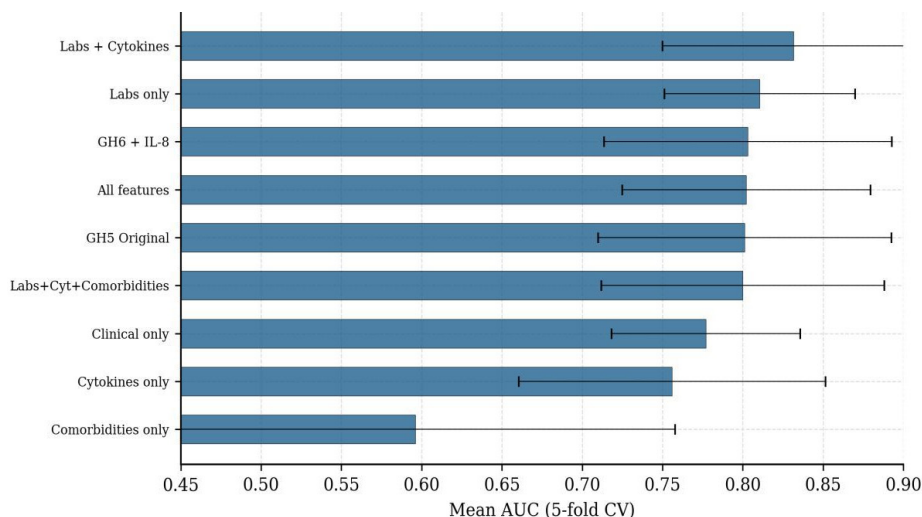


Figure 11. Ablation study — feature set comparison (mean AUC, 5-fold stratified cross-validation; error bars = SD; random forest, 300 trees, max depth 5, minimum leaf 5, seed 42; n=105). The combined Labs + Cytokines set achieves the highest mean AUC (0.832); the broadest set evaluated — 15 of the 17 BEI markers plus age, septic shock and malignancy — reaches 0.802 and the cytokine-only set 0.756. This broadest set is not the 17-marker BEI panel; set compositions are given in the released code

5. DISCUSSION

This study introduces the Biological Energy Index, an entropy-weighted framework designed to quantify composite biological energy reserve and identify per-patient biological failure modes in sepsis. The central finding is that BEI achieves moderate out-of-fold discriminative performance (AUC=0.793) while providing mechanistic interpretability through tipping contribution decomposition. This dual capability addresses a gap in sepsis prognostication: existing scores estimate aggregate risk but do not reveal which biological system is failing fastest.

5.1. BEI as a Precision Triage Framework

The four-tier tipping taxonomy operationalizes the BEI decomposition into categories that could, if validated prospectively, inform intervention prioritization. KILL-tier markers identify patients who may require immediate mechanism-targeted intervention; MODERATE-tier markers indicate specific subsystem failures that may retain sufficient reserve elsewhere. Under the leakage-free stratification, 90-day mortality decreased monotonically across BEI energy tiers — 71% in the Very Low tier, 17% in the At-Risk tier, and 14% in the Sufficient tier (n=35 per tier; Figure 10).

5.2. Positioning Within the AI Literature

BEI is architecturally a filter-method ensemble: it assigns importance weights using information-theoretic criteria prior to model fitting. This distinguishes it from wrapper methods (recursive feature elimination) and embedded methods (LASSO, Random Forest importance). Related work includes JMIM (Bennasar et al., 2015), mRMR (Peng et al., 2005), and Boruta (Kursa & Rudnicki, 2010). BEI differs in combining selected features into a weighted composite score with direct interpretive meaning, and in providing per-instance tipping decomposition absent from standard filter methods. The algorithm is domain-agnostic: any setting with continuous features measured against a binary outcome can be mapped to the framework. In this cohort the information-theoretic weighting did not improve discrimination over uniform weighting, so the contribution of the filter step here is interpretive — it orders the dimensions of the decomposition — rather than predictive.

5.3. Clinical Deployment Considerations

Prospective deployment of BEI would require integration with EHR systems through a dashboard displaying the current BEI score, threshold position, and top tipping contributions. Mandatory Lactate measurement at admission is essential to prevent the systematic score shift documented in Section 4.2. PTX3 and sIL-1R2, though not yet universally available, can be incorporated into rapid immunoassay panels. The tipping taxonomy also identifies biologically coherent failure modes that could, in principle, guide mechanism-aligned patient enrichment in clinical trials — addressing the heterogeneity problem that has contributed to negative sepsis trials. BEI outputs must be advisory: clinicians retain full autonomy, with BEI functioning as an augmentative tool.

5.4. Ethical and Data Governance Considerations

MNAR patterns in this cohort demonstrate that clinician-driven ordering behavior can encode bias into datasets. Prospective deployment must enforce mandatory measurement of high-weight biomarkers to prevent structural bias. Model drift from evolving ICU populations requires periodic recalibration. BEI weights may require adjustment in populations

with different biomarker distributions or clinical workflows.

5.5. TRIPOD+AI Compliance

This study was designed following the TRIPOD+AI statement (Collins et al., 2024) for transparent reporting of multivariable prediction models. The study type is development-only (Type 1a), with external validation committed to as a mandatory next step. All 17 candidate predictors were pre-specified based on biological plausibility. Calibration and decision-curve analyses are presented in Section 4.1. The TRIPOD+AI items are addressed in the corresponding sections of this article: participant flow (§3.1, Figure 2), predictors (§3.2), missing data (§3.6), sample size (§3.7), model evaluation (§3.5, §4.1) and limitations (§5.6).

5.6. STUDY LIMITATIONS

Several limitations constrain these findings. First, the analysis is based on a single retrospective cohort of N=135 patients (105 with a confirmed outcome) from one center. All thresholds, entropy weights, tipping frequencies, and tier classifications are cohort-specific and at high risk of overfitting. External validation in at minimum two independent sepsis cohorts is mandatory before any clinical application. Second, Lactate and D-dimer — two of the highest-mortality tipping markers — have 42–47% missingness, and the dynamic weight normalization strategy means different patients' BEI scores are computed from different marker subsets, making cross-patient comparisons of BEI fundamentally instrument-variable. The complete-case sensitivity analysis (n=33 outcome-confirmed patients, 13 deaths) addresses this but at the cost of substantial sample reduction and correspondingly wide confidence intervals. Third, the EPV ratio of 2.1 for 17-marker models is well below the recommended minimum of 10; all conclusions from multi-marker models are exploratory. Fourth, the tipping contribution is computed from Day 0 only and may shift as sepsis evolves. Fifth, the intervention mapping from tipping dimension to treatment is mechanistically grounded but not prospectively validated.

Sixth, and central to the interpretation of this work, the entropy-weighting step that motivates the construct conferred no discriminative advantage over equal weighting under leakage-free cross-

validation (0.793 vs 0.799). The tier boundaries of the taxonomy also rest on per-marker subgroups as small as one to three patients (Table 2), so individual tier assignments — in particular GCS and IL-8 — are unstable and are reported only as candidate structure.

Seventh, the score is not calibrated. Under leakage-free assessment the Hosmer–Lemeshow test rejects ($p=0.010$) and the calibration slope is 0.850 with a confidence interval spanning 0.46 to 1.56 (§4.1). BEI supports ranking and stratification in this cohort but does not yield trustworthy absolute risks, and any prospective use would require recalibration in the target population; the reported calibration values are regenerated by the analysis script named in §3.10.

Eighth, min–max normalization (Equation 1) is sensitive to extreme values. Although the log10 transform compresses skewed tails and the entropy weights are invariant to any monotone rescaling (§3.2), a single outlying measurement narrows the energy-score range of the remaining patients for that marker and can therefore alter BEI values and tipping-marker assignments. Robust percentile-based scaling (e.g., 5th–95th percentile with clipping) should be evaluated alongside min–max scaling in external validation cohorts.

6. CONCLUSION

We present the Biological Energy Index as an interpretable, information-theoretically weighted composite of routinely available sepsis biomarkers, together with an algebraic per-patient decomposition that attributes an individual's aggregate risk to a single dominant “tipping” dimension. On discrimination alone, BEI performed comparably to standard baselines and to organ-dysfunction scoring, and entropy weighting conferred no advantage over equal weighting under leakage-free cross-validation; we therefore do not advance BEI as a superior mortality predictor. Its intended contribution is decomposability: rather than assigning a patient to one of a fixed number of population-level clusters, the framework yields an individual-level, marker-specific characterization of which physiological axis dominates a given patient's risk, computed analytically from the patient's own data without a trained model.

We position this contribution deliberately within, not apart from, the emerging framework of sepsis subphenotypes, endotypes, and treatable traits

(Davenport et al., 2016), and the dynamical-systems view of critical physiological transitions. Where population-level subphenotyping partitions cohorts into discrete states, the present approach is closer in spirit to a per-patient treatable-trait attribution, and is offered as hypothesis-generating rather than confirmatory: several results — notably the tipping taxonomy, whose tier assignments rest on small per-marker subgroups — should be read as candidate structure for prospective testing, not as established phenotypes.

Two extensions define the path from this single-timepoint analysis to clinical relevance, and neither is demonstrated here. First, the decomposition can be computed on serial rather than admission-only measurements, so that it reflects the direction in which a patient is moving; the temporal trajectories observed in this cohort suggest that this directional formulation is where the signal is most likely to reside. Second, and most consequentially, the value of a per-patient attribution rests on whether the identified dimension predicts differential response to a mechanistically matched intervention — a biomarker-by-treatment interaction — rather than prognosis alone; this is testable in existing intervention-annotated critical-care cohorts before any prospective commitment.

Subject to external validation in independent, multi-center cohorts, an interpretable, patient-specific decomposition of this kind could inform biomarker-guided, mechanism-targeted decision support in sepsis. The present analysis establishes the feasibility and interpretability of the decomposition, not its clinical utility or readiness for deployment; until the trajectory and treatment-interaction steps are taken, the Biological Energy Index is best understood as an interpretable analytic framework and a source of testable, mechanism-specific hypotheses regarding the patient-specific drivers of sepsis outcome.

Ethics Approval and Consent to Participate: This article reports a secondary analysis of a de-identified dataset that was collected, published and deposited in a public repository by Davoudian et al. (2022) under a Creative Commons Attribution (CC BY) licence (Zenodo record 7612571). The original prospective study was approved by the Institutional Review Board of IRCCS Humanitas Research Hospital, Milan, Italy (approval no. 820/18), and participant enrolment and consent were handled by the original investigators under that approval, as described in their publication. The present authors did not

recruit, contact or interact with any participant, had no access to identifying information or to any key linking records to individuals, and did not attempt re-identification. Because the analysis used only fully de-identified, publicly released data, no additional ethics committee approval or participant consent was required for this secondary analysis.

Human and Animal Rights: No animals were used in this research. No experiments, interventions or procedures were performed on human participants by the authors of this study; all human data were obtained from the publicly available, de-identified dataset described above, which was originally collected under institutional ethics approval (IRCCS Humanitas Research Hospital, no. 820/18). The secondary analysis was conducted in accordance with the principles of the Declaration of Helsinki as they apply to research using de-identified human data.

Consent for Publication: Not applicable. The article contains no individual person's data in any form — no images, case details, or other information that could identify a participant. All results are reported as aggregate statistics computed on the de-identified cohort, and no patient-level data are reproduced in the article.

Availability of Data and Materials: The source dataset analysed in this study is publicly available from the original authors' Zenodo repository (record 7612571; <https://zenodo.org/records/7612571>) under a CC BY licence (Davoudian et al., 2022). Patient-level data are not reproduced in this article. The three analysis scripts listed in §3.10 (BEI_Manuscript_Figures_Standalone.py, reproduce_reported_values.py and BEI_R5_phenotype_and_calibration.py) regenerate every value reported in the article from the public dataset; they were supplied with the submission and will be archived in a public repository with a persistent DOI. Further information is available from the corresponding author on reasonable request.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' Contributions

Conceptualization: Gaurav Chandra and Shubham Chandra jointly conceived the Biological Energy Index framework, the entropy-weighting approach

and the tipping-dimension decomposition, and designed the study.

Writing: Gaurav Chandra and Shubham Chandra wrote the original draft of the manuscript.

Reviewing: Kirsten Bischof and Suraj Saggarr critically reviewed the manuscript for intellectual content; Suraj Saggarr reviewed the clinical interpretation of the findings.

Supervision: Gaurav Chandra supervised the study.

Editing: Shubham Chandra and Kirsten Bischof edited the manuscript through the revision rounds.

All authors read and approved the final version of the manuscript.

Conflict of interest: Gaurav Chandra is the Founder, Chairman and Chief Executive Officer of Adnexus Biotechnologies Inc.; Kirsten Bischof and Shubham Chandra are affiliated with Adnexus Biotechnologies Inc. The study received no funding from Adnexus Biotechnologies Inc. or from any other source. The authors declare no other competing interests.

Acknowledgements: The authors thank Davoudian et al. and the investigators of IRCCS Humanitas Research Hospital for making their sepsis cohort dataset publicly available under a CC BY licence, and the patients and families whose participation made the original study possible. The authors also thank the editors and anonymous reviewers of the Journal of Artificial Intelligence and Knowledge Engineering for their constructive comments, which improved the manuscript.

REFERENCES

- Abraham, E., et al. (2003). Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: a randomized controlled trial. *JAMA*, 290(2), 238–247.
- Bennasar, M., Hicks, Y., & Setchi, R. (2015). Feature selection using Joint Mutual Information Maximisation. *Expert Systems with Applications*, 42(22), 8520–8532.
- Bernard, G. R., et al. (2001). Efficacy and safety of recombinant human activated protein C for severe sepsis. *New England Journal of Medicine*, 344(10), 699–709.
- Collins, G. S., et al. (2024). TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378.

- Davenport, E. E., et al. (2016). Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *The Lancet Respiratory Medicine*, 4(4), 259–271.
- Davoudian, S., et al. (2022). A cytokine/PTX3 prognostic index as a predictor of mortality in sepsis. *Frontiers in Immunology*, 13, 979232.
- Ferreira, F. L., et al. (2001). Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*, 286(14), 1754–1758.
- Garlanda, C., Dinarello, C. A., & Mantovani, A. (2013). The interleukin-1 family: back to the future. *Immunity*, 39(6), 1003–1018.
- Haken, H. (2002). *Brain dynamics: synchronization and activity patterns in pulse-coupled neural nets with delays and noise*. Berlin: Springer.
- Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18(1), 83.
- Hotchkiss, R. S., et al. (2016). Sepsis and septic shock. *Nature Reviews Disease Primers*, 2, 16045.
- Huttunen, R., et al. (2011). High plasma level of long pentraxin 3 (PTX3) is associated with fatal disease in bacteremic patients: a prospective cohort study. *PLoS ONE*, 6(3), e17653.
- Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical Care Medicine*, 13(10), 818–829.
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11), 1–13.
- Lauritsen, S. M., et al. (2020). Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nature Communications*, 11, 3852.
- Lenton, T. M. (2011). Early warning of climate tipping points. *Nature Climate Change*, 1(4), 201–209.
- Mauri, T., et al. (2010). Persisting high levels of plasma pentraxin 3 over the first days after severe sepsis and septic shock onset are associated with mortality. *Intensive Care Medicine*, 36(4), 621–629.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238.
- Porte, R., et al. (2019). The long pentraxin PTX3 as a humoral innate immunity functional player and biomarker of infections and sepsis. *Frontiers in Immunology*, 10, 794.
- Raith, E. P., et al. (2017). Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA*, 317(3), 290–300.
- Riedemann, N. C., Guo, R.-F., & Ward, P. A. (2003). The enigma of sepsis. *Journal of Clinical Investigation*, 112(4), 460–467.
- Ross, B. C. (2014). Mutual information between discrete and continuous data sets. *PLoS ONE*, 9(2), e87357.
- Rudd, K. E., et al. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*, 395(10219), 200–211.
- Scheffer, M., et al. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59.
- Seely, A. J., & Macklem, P. T. (2004). Complex systems and the technology of variability analysis. *Critical Care*, 8(6), R367–R384.
- Sewak, M. (2026). The ascent of machine cognition: a systematic review of large reasoning models. *Journal of Artificial Intelligence and Knowledge Engineering*, 1(1), 1–15. <https://doi.org/10.37819/Jaike.25.0010>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604.
- Singer, M., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801–810.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72–101.

- van der Poll, T., van de Veerdonk, F. L., Scicluna, B. P., & Netea, M. G. (2017). The immunopathology of sepsis and potential therapeutic targets. *Nature Reviews Immunology*, 17(7), 407–420.
- Vergara, I. A., et al. (2014). Gene expression profile discriminates sepsis patients from healthy controls. *Critical Care*, 18(5), 584.
- Vincent, J.-L., et al. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7), 707–710.
- Ward, P. A., & Bosmann, M. (2012). A historical perspective on sepsis. *American Journal of Pathology*, 181(1), 2–7.
- Warren, B. L., et al. (2001). High-dose antithrombin III in severe sepsis: a randomized controlled trial. *JAMA*, 286(15), 1869–1878.



Publisher's note: Eurasia Academic Publishing Group (EAPG) and AAKS remain neutral about jurisdictional claims in published maps and institutional affiliations.

Open Access. This article is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) licence, which permits copy and redistribute the material in any medium or format for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the licence terms. Under the following terms you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorsed you or your use. If you remix, transform, or build upon the material, you may not distribute the modified material. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nd/4.0/>.