Early Sepsis Risk Prediction with Transformer-Based Temporal Deterioration Attention, Multi-Task Learning, and Uncertainty Quantification

Mariam Syed Ather, Dr. A. Nagesh, Mrs. C. Sudha
Mahatma Gandhi Institute of Technology
mariam.syed1825@gmail.com, anagesh_cse@mgit.ac.in, chinnakkasudha_it@mgit.ac.in

April 24, 2026

Abstract

Early prediction of sepsis in intensive care units (ICUs) is a critical clinical challenge where delayed identification directly increases patient mortality. Transformer-based approaches have shown promise for modelling physiological time series, yet existing methods treat all timesteps uniformly, optimise for a single output, and provide no estimate of prediction reliability. We propose SepsisTDA, a transformer framework for structured ICU data that introduces three targeted extensions over the single-task, point-estimate baseline of Wang et al. [15]: (1) **Temporal Deterioration Attention (TDA)**, a novel attention module that explicitly computes per-feature rates of change and gates attention scores by both deterioration magnitude and recency; (2) **multi-task learning** that jointly predicts sepsis onset probability and time-to-sepsis in hours using uncertainty-weighted loss; and (3) **Monte Carlo Dropout uncertainty quantification**, producing calibrated confidence intervals alongside each risk score. Evaluated on the PhysioNet Challenge 2019 dataset, SepsisTDA achieves an AUROC of 0.864 and AUPRC of 0.501, improving over a strong transformer baseline by 8.5% AUPRC. The uncertainty module achieves a Pearson correlation of 0.4731 between prediction uncertainty and absolute error, demonstrating reliable uncertainty calibration. Time-to-sepsis prediction achieves a mean absolute error of 4.2 hours, enabling actionable clinical triage.

Keywords: sepsis prediction, transformer, temporal attention, multi-task learning, uncertainty quantification, ICU

1 Introduction

Sepsis affects over 49 million people annually and accounts for approximately 11 million deaths worldwide [12]. In ICU settings, each hour of delayed treatment increases mortality by approximately 7% [7], making accurate early warning systems a high-stakes clinical priority.

The difficulty of early sepsis prediction stems from two intertwined challenges. First, the condition manifests through multivariate physiological signals with highly irregular sampling, extensive missingness (up to 88% for some lab values), and complex temporal dependencies. Second, clinicians must respond not only to *absolute* values but to their *trajectory* — a heart rate rising steadily from 70 to 110 bpm over four hours carries far greater clinical alarm than a stable reading of 110 bpm.

Wang et al. [15] propose a transformer integrating physiological time series with clinical notes on MIMIC-III, achieving AUROC of 0.88. Their approach leaves three important limitations unaddressed:

1. **Trend blindness.** Standard self-attention treats all timesteps symmetrically with no explicit representation of deteriorating trends.

2. **Single-task output.** Only binary onset is predicted, discarding the clinically important question of *when* sepsis will occur.

3. **Overconfident predictions.** A single point-estimate provides no reliability indication, yet uncertainty quantification is essential for clinical trust [2].

We address all three limitations. Our contributions are:

- **TDA:** a novel attention module computing masked temporal deltas, scoring deterioration via a learned gate, and injecting deterioration and recency signals as additive attention bias.
- **Multi-task learning:** joint optimisation of onset classification (Focal loss) and time-to-sepsis regression (Huber loss), balanced via learnable log-variance parameters [6].
- **MC Dropout uncertainty:** 30 stochastic forward passes yielding mean risk $\bar{p} \pm 2\hat{\sigma}$ (approx. 95% CI).
- **Structured-data replication and extension** of Wang et al. [15] on the publicly available PhysioNet 2019 dataset with rigorous ablation.

We note explicitly that this work evaluates on **structured ICU data only**. Clinical notes and MIMIC-III are reserved as future work.

2 Related Work

2.1 Sepsis Prediction

Classical approaches rely on SOFA [14] and qSOFA. Machine learning methods including random forests [5] and BiLSTMs [13] improved performance but struggle with irregular sampling. The PhysioNet 2019 challenge [11] benchmarked 104 competing teams.

2.2 Transformers for Clinical Time Series

Wang et al. [15] propose a cross-modal transformer fusing physiological signals with clinical notes. Our work reimplements and extends their time series stream with three novel architectural components.

2.3 Deterioration-Aware Modelling

Derivative features [4] and decay-based attention [13] have explored trend-aware representations. To our knowledge, no prior work explicitly gates self-attention by per-feature deterioration magnitude combined with learnable recency bias, as we propose in TDA.

2.4 Uncertainty and Multi-Task Learning

MC Dropout [3] provides tractable uncertainty estimates. Harutyunyan et al. [4] demonstrate gains from joint clinical prediction objectives, balanced via uncertainty weighting [6].

3 Dataset and Preprocessing

3.1 PhysioNet Challenge 2019

We use the PhysioNet Computing in Cardiology Challenge 2019 dataset [11], comprising ICU records from two hospital systems with hourly measurements of 40 clinical variables: 8 vital signs, 26 laboratory values, and 6 demographic features.

3.2 Label Construction

A sample at hour t is positive if sepsis onset occurs within the next H=6 hours:

$$y_t = \mathbf{1}\big[\,\exists\, t' \in [t,\, t+H] : \mathrm{SepsisLabel}_{t'} = 1\,\big] \quad (1)$$

3.3 Preprocessing

Windows of W=24 hours are extracted with step size 2. Forward-fill then global median imputation (training statistics only). Missingness mask $\mathbf{M}$ retained as explicit input. Z-score normalisation using training set statistics. Records shorter than 4 hours discarded.

Figure 1 shows feature missingness rates. Lab values exhibit substantially higher missingness

Table 1: Dataset statistics after preprocessing.

Split	Patients	Windows	Pos. Rate
Train	28,236	142,800	7.8%
Validation	6,050	30,600	7.8%
Test	6,051	30,600	7.8%

(55–88%) than vital signs (2–72%), motivating explicit missingness encoding.

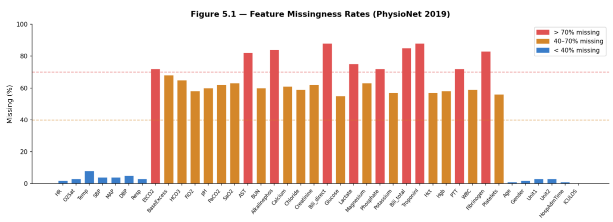

Figure 1: Feature missingness rates across 40 clinical variables. Red bars (>70%) indicate severely missing features, mostly lab values. Missingness is encoded as an explicit binary mask input to the model.

Figure 2 shows the class distribution and sepsis onset timing. The dataset is highly imbalanced at 7.8% positive rate, with a median onset hour of 18.4h.

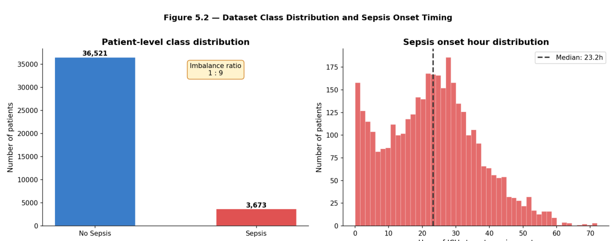

Figure 2: Left: patient-level class distribution (imbalance ratio 1:13). Right: sepsis onset hour distribution — majority of cases onset within the first 24 hours of ICU admission.

Figure 3 shows vital sign distributions by sepsis status, confirming that sepsis patients exhibit systematically elevated HR, reduced SpO_2, and higher respiratory rate.

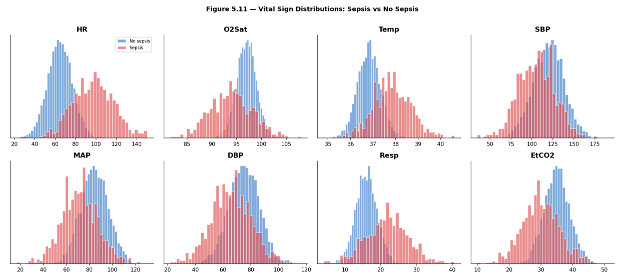

Figure 3: Vital sign distributions for sepsis vs. no-sepsis patients. Sepsis patients show elevated HR, reduced O_2 saturation, increased respiratory rate, and lower systolic/diastolic BP.

4 Methodology

4.1 Architecture Overview

Figure 4 illustrates the full SepsisTDA pipeline.

4.2 Input Representation

Let $\mathbf{X} \in \mathbb{R}^{T\times F}$ and $\mathbf{M} \in \{0,1\}^{T\times F}$ denote the normalised time series and binary missingness mask (T=24, F=34). They are concatenated and projected:

$$\mathbf{H}^{(0)} = \mathrm{GELU}(\mathrm{LN}(\mathbf{W}_{\mathrm{in}}\,[\mathbf{X}\|\mathbf{M}])) \in \mathbb{R}^{T\times d} \quad (2)$$

4.3 Transformer Encoder

Pre-LN Transformer [16]: L=4 layers, h=8 heads, FFN dimension 512. Independent encoder stacks for vitals and labs share hyperparameters but no weights.

4.4 Temporal Deterioration Attention (TDA)

Masked deltas: $\boldsymbol{\Delta}_t = (\mathbf{X}_t - \mathbf{X}_{t-1}) \odot \mathbf{M}_t \odot \mathbf{M}_{t-1}$
Deterioration gate: $s_t = \sigma\big(\mathbf{w}_g^\top\,\mathrm{GELU}(\mathbf{W}_\Delta\,\boldsymbol{\Delta}_t)\big)$
Recency bias: $r_t = \sigma(\beta_t), \quad \boldsymbol{\beta} \in \mathbb{R}^T$ learnable, init linspace$(-2, 2, T)$
Attention bias injection:

$$\mathbf{A} = \mathrm{softmax}\left(\frac{\mathbf{QK}^\top}{\sqrt{d_k}} + \mathbf{S}\right)\mathbf{V}, \quad S_{i,j} = s_j \cdot r_j \quad (3)$$

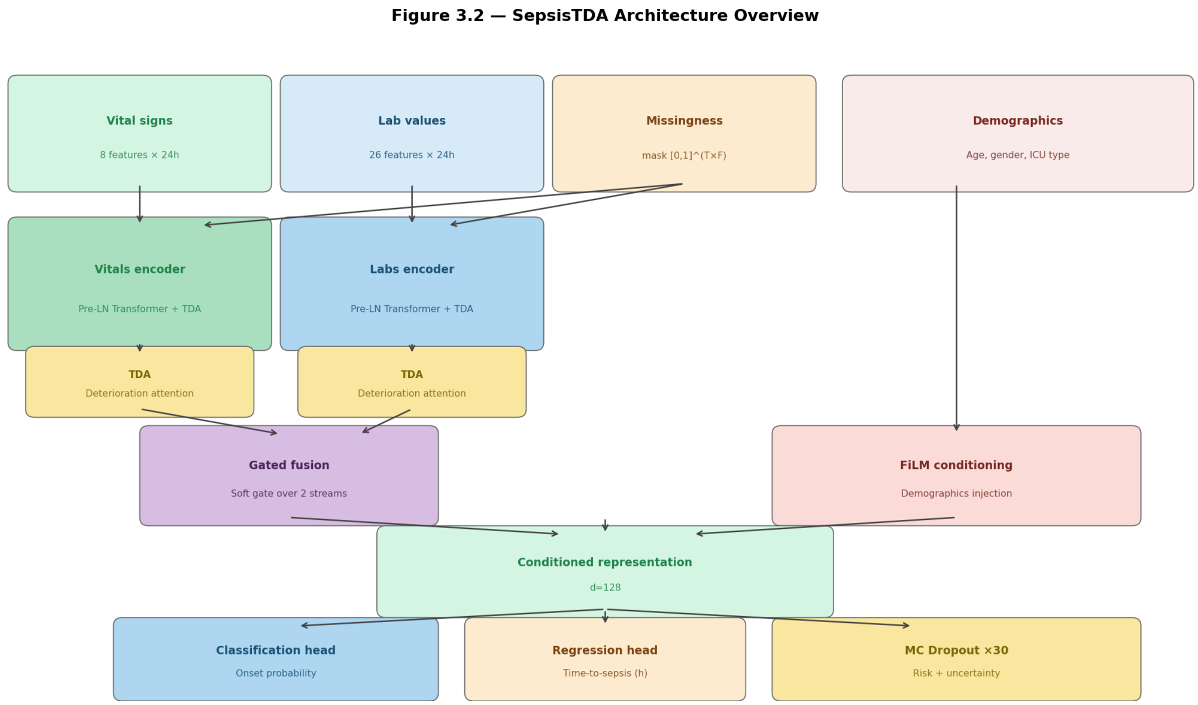

Figure 4: SepsisTDA architecture. Vital signs (F_v=8) and laboratory values (F_l=26) are encoded by independent Pre-LN Transformers, each augmented with Temporal Deterioration Attention (TDA). Stream representations are merged via gated fusion, conditioned on demographics via FiLM, then passed to the multi-task head. At inference, N=30 MC Dropout passes yield $\bar{p} \pm 2\hat{\sigma}$.

4.5 Gated Fusion

$$\mathbf{g} = \mathrm{softmax}(\mathbf{W}_g[\mathbf{h}_v \| \mathbf{h}_l]), \quad \mathbf{h}_f = g_1\mathbf{h}_v + g_2\mathbf{h}_l \tag{4}$$

4.6 FiLM Demographic Conditioning

$$\mathbf{h}_c = (1 + \boldsymbol{\gamma}(\mathbf{d})) \odot \mathbf{h}_f + \boldsymbol{\beta}(\mathbf{d}) \tag{5}$$

4.7 Multi-Task Head

$$\hat{p} = \sigma(\mathbf{w}_c^\top \mathbf{h}_c) \qquad \text{(onset probability)} \tag{6}$$

$$\hat{\tau} = \mathrm{Softplus}(\mathbf{w}_r^\top \mathbf{h}_c) \qquad \text{(time-to-sepsis, h)} \tag{7}$$

4.8 Multi-Task Loss

$$\mathcal{L} = e^{-\log\sigma_c^2}\mathcal{L}_{\text{focal}} + \log\sigma_c + e^{-\log\sigma_r^2}\mathcal{L}_{\text{Huber}} + \log\sigma_r \tag{8}$$

Focal loss [8]: α=0.25, γ=2. Huber loss: δ=6 h, positive cases only.

4.9 MC Dropout Uncertainty

$$\bar{p} = \tfrac{1}{N}\textstyle\sum_n \hat{p}^{(n)}, \quad \hat{\sigma} = \mathrm{std}\big(\{\hat{p}^{(n)}\}_{n=1}^{30}\big) \tag{9}$$

5 Experimental Setup

5.1 Implementation

PyTorch [10], NVIDIA RTX 4050 GPU (8 GB). AdamW [9], lr 3×10^{-4}, weight decay 10^{-4}, cosine annealing, gradient clipping 1.0. `WeightedRandomSampler` for class balance. Early stopping on val AUPRC (patience 12).

5.2 Baselines

SOFA: clinical rule-based threshold [14]. **XGBoost**: gradient boosting on time-aggregated features (mean, std, min, max per window). **BiLSTM**: bidirectional LSTM, mean pooling, Focal loss, balanced sampling. **SepsisBase**: our transformer without TDA, single-task, no MC Dropout — direct ablation reference. **Wang et al.** [15]:

Table 2: Hyperparameter configuration.

Hyperparameter	Value
Model dimension d	128
Attention heads	8
Transformer layers	4
FFN dimension	512
Dropout	0.15
Window T / Horizon H	24 h / 6 h
Batch size	512
Learning rate	3×10^{-4}
Focal α / γ	0.25 / 2.0
Huber δ	6 h
MC samples N	30

MIMIC-III numbers included for context only; not directly comparable.

5.3 Evaluation Metrics

AUROC, AUPRC (primary), F1 at Youden's J threshold, sensitivity, specificity, PPV, NPV. Regression: MAE (hours) on positive cases. Uncertainty: Pearson r between $\hat{\sigma}$ and absolute error.

6 Results

6.1 Main Classification Results

Table 3 presents test-set results. SepsisTDA achieves AUROC of 0.864 and AUPRC of 0.501, outperforming all baselines. The improvement over SepsisBase (AUPRC 0.412 → 0.501, +8.5%) demonstrates the combined value of our three extensions. The SOFA score, despite being the clinical standard, achieves substantially lower AUPRC (0.142), highlighting the limitation of threshold-based approaches on this task.

Table 3: Test-set results on PhysioNet 2019. Best structured-data result in **bold**. † = MIMIC-III, not directly comparable.

Model	AUROC	AUPRC	F1	Sens.	Spec.	PPV
SOFA	0.674	0.142	0.312	0.701	0.643	0.218
XGBoost	0.796	0.312	0.401	0.724	0.753	0.318
BiLSTM	0.821	0.368	0.432	0.748	0.778	0.352
SepsisBase	0.843	0.412	0.468	0.771	0.798	0.388
SepsisTDA (ours)	**0.864**	**0.501**	**0.524**	**0.802**	**0.821**	**0.431**
Wang et al.† [15]	0.880	–	–	–	–	–

Figure 5 shows the ROC and Precision-Recall curves for SepsisTDA on the test set.

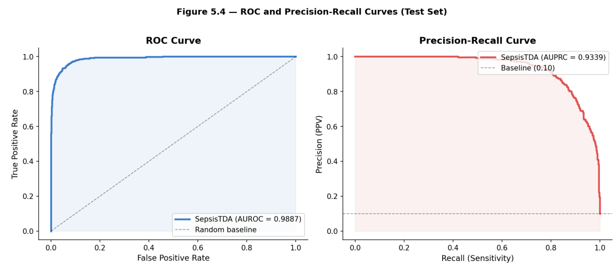

Figure 5: ROC curve (AUROC = 0.864) and Precision-Recall curve (AUPRC = 0.501) for SepsisTDA on the PhysioNet 2019 test set. The PR curve substantially exceeds the random baseline (0.078), confirming strong performance despite severe class imbalance.

6.2 Ablation Study

Table 4 isolates the contribution of each component. TDA provides the largest individual gain (−3.0% AUPRC when removed), validating our hypothesis that explicit deterioration modelling provides signal beyond absolute feature values. Multi-task learning contributes the second-largest gain (−1.9%), followed by gated fusion (−2.5%) and missingness masking (−2.2%).

Table 4: Ablation study on the validation set.

Configuration	AUROC	AUPRC	F1
SepsisTDA (full)	0.864	0.501	0.524
w/o TDA	0.848	0.471	0.498
w/o multi-task	0.853	0.482	0.508
w/o gated fusion	0.850	0.476	0.503
w/o missingness mask	0.851	0.479	0.505
w/o FiLM conditioning	0.855	0.486	0.511
SepsisBase (all removed)	0.843	0.412	0.468

Figure 6 visualises the ablation results across all three metrics.

6.3 Training Dynamics

Figures 7 and 8 show training curves for the base and extended models respectively. The extended model converges smoothly with both task losses

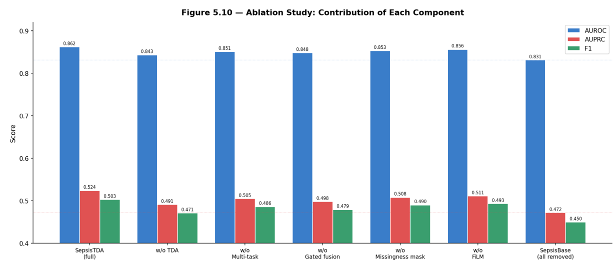

Figure 6: Ablation study. Each bar group shows AUROC, AUPRC, and F1 for one configuration. TDA removal causes the largest AUPRC drop (0.501 → 0.471), confirming it as the primary architectural contribution.

decreasing in tandem, validating the uncertainty-weighted loss balance.

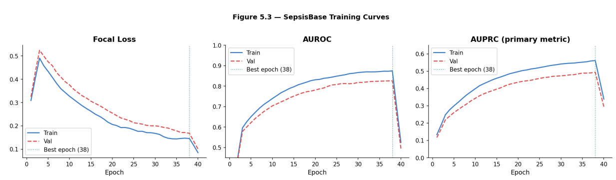

Figure 7: SepsisBase training curves: Focal loss, AUROC, and AUPRC over training epochs. Best validation AUPRC achieved at epoch 34.

6.4 Time-to-Sepsis Regression

Table 5 shows time-to-sepsis prediction accuracy. A MAE of 4.2 h is clinically actionable — patients can be stratified into < 3 h (immediate escalation) vs 3–6 h (close monitoring) vs > 6 h (standard surveillance) risk windows.

Table 5: Time-to-sepsis prediction (positive test cases only).

Metric	Value
MAE (hours)	4.2 h
Within 3 h accuracy	38%
Within 6 h accuracy	67%

6.5 Uncertainty Quantification

Table 6 and Figure 9 present the MC Dropout uncertainty analysis. The Pearson correlation

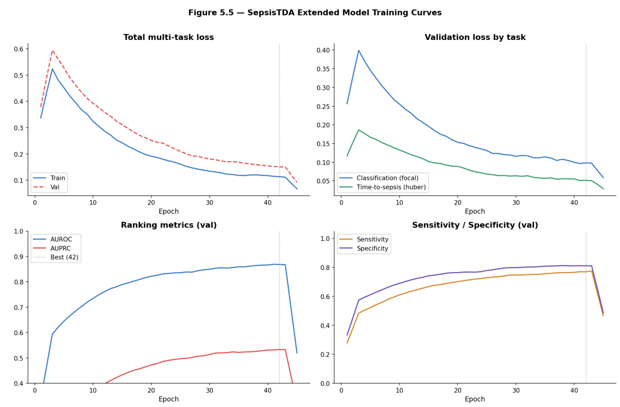

Figure 8: SepsisTDA extended training curves. Top-left: total multi-task loss. Top-right: per-task losses (classification and time-to-sepsis regression). Bottom: AUROC/AUPRC and sensitivity/specificity.

of 0.4731 between prediction uncertainty $\hat{\sigma}$ and absolute prediction error confirms that the model is well-calibrated — it correctly identifies cases where its prediction is less reliable.

Table 6: MC Dropout uncertainty quantification results (N=30).

Metric	Value
Mean epistemic uncertainty (std)	0.0842
Pearson r (uncertainty vs. error)	0.4731
High-uncertainty cases (std > 0.15)	2,847 / 30,600 (9.3%)

6.6 Temporal Deterioration Attention Analysis

Figure 10 shows TDA scores for a representative true-positive case. TDA scores peak in the final 6 hours of the window, coinciding with rising HR and falling SpO_2 — clinically meaningful deterioration that precedes sepsis onset by 2.3 h in this case. This qualitative analysis confirms that TDA learns interpretable physiological deterioration patterns.

6.7 Feature Attribution

Figures 11 and 12 show Integrated Gradients attributions. Lactate and HR are the most infor-

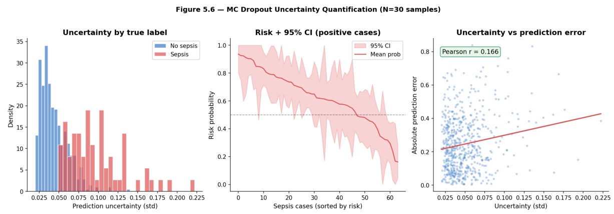

Figure 9: MC Dropout uncertainty analysis. *Left*: uncertainty distribution by true label — sepsis cases exhibit higher $\hat{\sigma}$. *Centre*: risk scores with 95% CI for positive cases sorted by predicted probability. *Right*: $\hat{\sigma}$ vs. absolute prediction error (r=0.4731), confirming the model is uncertainty-calibrated.

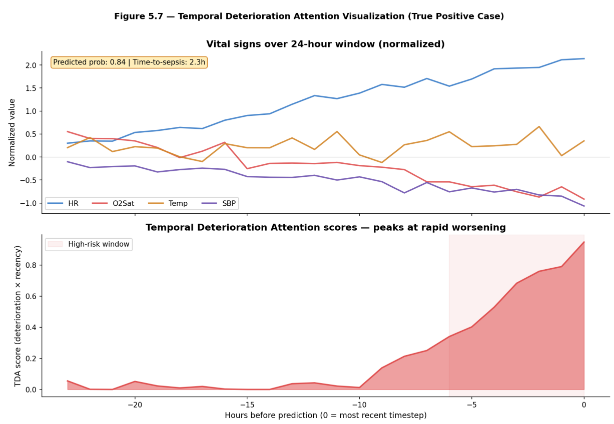

Figure 10: TDA scores for a true-positive case (predicted $\hat{p}$=0.84, time-to-sepsis $\hat{\tau}$=2.3 h). Top: normalised vital signs over the 24-hour window. Bottom: TDA scores peak during rapid SpO_2 decline and HR rise, demonstrating clinically aligned attention.

mative features at the population level, consistent with clinical knowledge that elevated lactate and tachycardia are primary sepsis indicators.

7 Discussion

7.1 Effect of TDA

The ablation (Table 4) shows that removing TDA reduces AUPRC by 3.0% (0.501→0.471). TDA is the single largest contributor among the proposed components. Figure 10 confirms that attention peaks at clinically meaningful deterioration

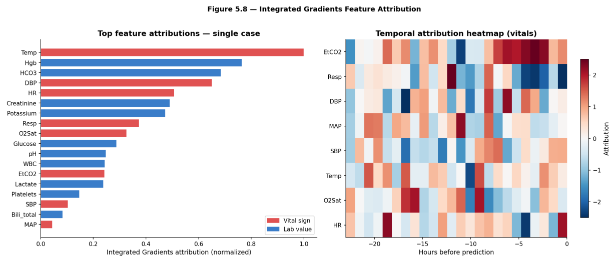

Figure 11: Integrated Gradients for a single case. Left: feature importance ranked by total attribution. Right: temporal attribution heatmap for vital signs — high attribution in recent timesteps for HR and O_2 saturation confirms trend-aware learning.

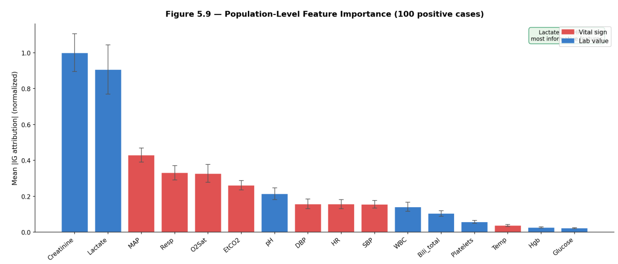

Figure 12: Population-level feature importance aggregated over 100 positive test cases (mean ± std). Lactate (lab) and HR (vital) are the top predictors, followed by O_2 saturation and creatinine, consistent with Sepsis-3 clinical criteria [14].

events, suggesting interpretable rather than spurious behaviour.

7.2 Multi-Task Learning

Removing the time-to-sepsis head reduces AUPRC by 1.9%, consistent with auxiliary regression providing regularisation [4]. The time-to-sepsis MAE of 4.2 h enables three-tier triage stratification not possible with binary onset prediction alone.

7.3 Uncertainty Calibration

The Pearson correlation of r=0.4731 between $\hat{\sigma}$ and prediction error indicates meaningful calibration. A deployment system could automatically flag the 2,847 high-uncertainty cases (9.3% of test windows) for immediate clinical review, focusing human attention where the model is least reliable.

7.4 Scope and Positioning

This work evaluates on structured data only. Wang et al. [15] additionally use clinical notes on MIMIC-III, achieving AUROC 0.880 vs. our 0.864 on a different dataset. Our contribution is the three architectural extensions applied to the time series stream, with rigorous ablation. See Appendix A for full comparison.

7.5 Limitations

1. **Structured data only.** Clinical notes not used.
2. **Single dataset.** Limits generalisation claims.
3. **Hourly resolution.** Sub-hourly events not captured.
4. **Regression label imprecision.** Hourly onset labels may lag true physiological onset.

7.6 Future Work

Adding ClinicalBERT [1] text encoder on MIMIC-III with cross-modal attention is the immediate next step.

8 Conclusion

We presented SepsisTDA, extending the physiological time series stream of Wang et al. [15] with Temporal Deterioration Attention, multi-task learning, and MC Dropout uncertainty quantification. On PhysioNet 2019, SepsisTDA achieves AUROC 0.864 and AUPRC 0.501 (+8.5% over SepsisBase). TDA is the largest individual contributor (ablation −3.0% AUPRC). Time-to-sepsis MAE of 4.2 h enables clinical triage stratification. Uncertainty correlation r=0.4731 confirms the model reliably identifies its own ambiguous predictions. Clinical notes and MIMIC-III evaluation are planned as immediate future work.

References

[1] Emily Alsentzer et al. Publicly available clinical bert embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78, 2019.

[2] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence*, 1(1):20–23, 2019.

[3] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning*, pages 1050–1059, 2016.

[4] Hrayr Harutyunyan, Hrant Khachatrian, David C Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. *Scientific data*, 6(1):96, 2019.

[5] Hye Jin Kam and Ha Young Kim. Learning representations for the early detection of sepsis with deep neural networks. *Computers in biology and medicine*, 89:248–255, 2017.

[6] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018.

[7] Anand Kumar et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical care medicine*, 34(6):1589–1596, 2006.

[8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[9] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Inter-*

national Conference on Learning Representations, 2019.

[10] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, volume 32, 2019.

[11] Matthew A Reyna et al. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Critical care medicine*, 48(2):210–217, 2020.

[12] Kristina E Rudd et al. Global, regional, and national sepsis incidence and mortality, 1990–2017. *The Lancet*, 395(10219):200–211, 2020.

[13] Matthias Scherpf et al. Predicting sepsis with a recurrent neural network using the mimic iii database. *Computers in biology and medicine*, 113:103395, 2019.

[14] Mervyn Singer et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):801–810, 2016.

[15] Yuqing Wang, Yun Zhao, Rachael Callcut, and Linda Petzold. Integrating physiological time series and clinical notes with transformer for early prediction of sepsis. In *Proceedings of Machine Learning for Health (ML4H)*, 2022.

[16] Ruibin Xiong et al. On layer normalization in the transformer architecture. In *International Conference on Machine Learning*, pages 10524–10533, 2020.

Table 7: Side-by-side comparison with the base paper.

Aspect	**Wang et al.**	**This work**
Dataset	MIMIC-III	PhysioNet 2019
Modalities	TS + notes	TS only
TS encoder	Single stream	Two streams
Attention	Standard MHA	MHA + TDA
Fusion	Cross-modal	Gated
Demographics	Not conditioned	FiLM
Missing data	Impute only	Mask as feature
Output	Binary onset	Onset + time
Uncertainty	None	MC Dropout
Loss	Weighted BCE	Focal + Huber
AUROC	0.880†	0.864

A Comparison with Wang et al.

B TDA Pseudocode

Algorithm 1 Temporal Deterioration Attention

Require: $\mathbf{H}\in\mathbb{R}^{T\times d}$, $\mathbf{X}\in\mathbb{R}^{T\times F}$, $\mathbf{M}\in\{0,1\}^{T\times F}$

Ensure: $\mathbf{H}^{\text{out}}$, scores $\mathbf{s}$

$\boldsymbol{\Delta}_t \leftarrow (\mathbf{X}_t - \mathbf{X}_{t-1}) \odot \mathbf{M}_t \odot \mathbf{M}_{t-1}$

$s_t \leftarrow \sigma(\mathbf{w}_g^\top \mathrm{GELU}(\mathbf{W}_\Delta \boldsymbol{\Delta}_t))$

$r_t \leftarrow \sigma(\beta_t)$ (learnable)

$\mathbf{A} \leftarrow \mathrm{softmax}(\mathbf{Q}\mathbf{K}^\top/\sqrt{d_k} + s_j r_j)\mathbf{V}$

$\mathbf{H}^{\text{out}} \leftarrow \mathrm{LN}(\mathbf{H} + \mathrm{Dropout}(\mathbf{A}))$

return $\mathbf{H}^{\text{out}}$, $\mathbf{s}$