International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 1358-1376
https://doi.org/10.70917/ijcisim-2026-5533
Article

Conditioned Latent State-Space Network for Intervention-Aware ICU Mortality Prediction Using Longitudinal Physiological Data

Nitin Jain[1], Purushottam Chaudhary[2], Gaurav Jindal[3], K Deepthi Reddy[4], Bharti Sharma[5,6], Dipika Jain[7*]

[1]School of Computer Science and Engineering, Bennett University, Greater Noida, India. garg.nitin007@gmail.com
[2]Founder@Infinisync Consulting Pvt Ltd, India, roshan.purush@gmail.com
[3]Lloyd Institute of Engineering and Technology, Greater Noida, India. gauravjindal05@gmail.com
[4]Computer Science and Engineering Department, CVR college of Engineering Hyderabad, Telangana, India. deepthiminnam509@gmail.com
[5]Department of ECE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam 602105, Chennai, Tamil Nadu, India. bhardwaj.bharti989@gmail.com
[6]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India. bhardwaj.bharti989@gmail.com
[7]Department of Computer Science Engineering & Technology, Amity School of Engineering & Technology (ASET), Amity University, Noida, Uttar Pradesh, India, e-mail: dipikajain12.24@gmail.com
*Corresponding author: Dipika Jain (dipikajain12.24@gmail.com)

Abstract: Early prediction of intensive care unit (ICU) mortality requires models capable of capturing longitudinal physiological dynamics and therapeutic interventions. This study proposes a Conditioned Latent State-Space Network (CLSSN) that integrates temporal representation learning, intervention-aware latent state transitions, probabilistic uncertainty modeling, and counterfactual trajectory generation within a unified deep learning framework. The model was evaluated on a curated cohort of 2,000 adult ICU patients from the MIMIC-III database using six physiological variables and vasopressor administration. Experimental results demonstrated competitive predictive performance, achieving an AUROC of 0.7911, AUPRC of 0.5209, Expected Calibration Error of 0.0722, and Brier Score of 0.0973. Additional analyses, including calibration assessment, ablation studies, temporal saliency mapping, and decision curve analysis, demonstrated the framework's reliability, interpretability, and potential utility for AI-assisted critical care decision support.

Keywords: ICU Mortality Prediction, State-Space Model, Longitudinal Electronic Health Records, Intervention-Aware Learning, Deep Learning, Explainable Artificial Intelligence, Counterfactual Analysis, Clinical Decision Support, Uncertainty Estimation

1. Introduction

Intensive Care Units (ICUs) are among the most information-rich environments in healthcare today where patient information is continuously generated from monitoring of physiological processes, laboratory investigations, therapy administration, and clinical care; these data are recorded in complex longitudinal records. These data streams are heterogeneous and have important information both about disease evolution, treatment response, and patient deterioration. But for large dimensions, time series, irregular sampling frequency and frequent missing observations, traditional statistical methods are difficult to apply. Early detection of patients with a higher risk of adverse outcomes is crucial to guide clinical choices, allocate resources and initiate early therapy. The recent breakthroughs in artificial intelligence (AI) have revolutionized the field of predictive analytics with methods that allow large-scale automated learning from electronic health records (EHRs) in critical care. Mortality prediction, sepsis detection, patient

Copyright: © 2026 by the authors

deterioration and intensive care resource management have shown promising results in deep learning architectures such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), Transformer-based models, and Neural Ordinary Differential Equations (Neural ODEs). These data-driven models automatically learn hierarchical temporal representations from sequential physiological measurements without first having to build in features and assumptions in traditional statistical models. These capabilities have led to enhanced predictive performance on many longitudinal patterns of healthcare data across a variety of clinical applications.

Although progress has been made, the challenge of critical care prediction is still much more complicated than traditional time-series prediction, as clinical course is continually affected by the interventions given during hospital stays. Poor clinical outcomes are not just related to the pathological physiology, but are also related to the pharmacologic treatment, mechanical ventilation management, use of vasopressors, resuscitation fluids and other measures that significantly affect the evolution of the patient throughout the clinical process. For this reason, systems aimed at clinical implementation need to capture temporal physiological progression as well as temporal response to interventions. This need has stimulated growing interest in models that can incorporate the temporal dynamics and clinically meaningful depictions of treatment-response relationships. Respiratory and metabolic biomarkers are among the plethora of physiological measures routinely collected in the ICU and play a pivotal role in evaluating patient status. The combination of arterial oxygen tension (PaO_2), arterial carbon dioxide tension ($PaCO_2$), blood pH, lactate concentration, urine output, and serum creatinine levels gives a significant indication of the physiology of multiple organs. Continuous modeling of these variables means that instead of characterizing a single measurement, it is possible to characterize trajectories of each patient and thus increases the possibilities for identifying subtle patterns associated with clinical deterioration. As a result, there's a growing emphasis on longitudinal physiological monitoring as a vital aspect of modern critical care research with AI.

The other key factor for clinical implementations is model transparency. Decision support systems in healthcare are becoming more prone to relying on accurate predictions, but also explanations of the factors that affect the prediction of the individual risk estimate. Clinical practice guidelines and recommendations are increasingly stating that AI tools should complement, not supplant, clinical judgment and be interpretable and reliable. Recent studies have therefore started to use explainable artificial intelligence (XAI), uncertainty estimation, calibration analysis, and fairness assessment to enhance trust in the machine learning based decision support system in safety-critical healthcare applications.

1.1 Research Objectives and Contributions

To address these problems, the following study presents a Conditioned Latent State-Space Network (CLSSN) to predict the risks of patients in an ICU for longitudinal data with temporal physiological observations and intervention-aware latent dynamics. The proposed framework aims to learn compact latent representations of patient physiology and explicitly model the temporal evolution under therapeutic interventions. The framework brings together two components: Conditional latent state transitions and stochastic uncertainty modeling, both within a single deep learning structure, to enable the provision of clinically meaningful risk estimation while also offering tools for model interpretability, such as, but not limited to, calibration analysis, latent space visualization, and temporal feature attribution.

The principal contributions of this study are summarized as follows:

- A joint model in the form of a unified Conditioned Latent State-Space Network (CLSSN) for the prediction of ICU mortality, modelling longitudinal physiological data, latent temporal dynamics and intervention-aware state transitions.
- A framework of latent dynamics with intervention inputs that has a direct dynamic coupling between temporal evolution and therapeutic actions, while accounting for stochastic uncertainty through latent regularization.

- A means of inferring counterfactuals that can provide simulation of alternative intervention scenarios for conducting treatment-aware risk analysis in a longitudinal prediction framework.
- A comprehensive evaluation strategy that combines discrimination, calibration, ablation analysis, latent representation visualization, fairness assessment, and temporal explainability to gain multi-dimensional evaluation of model performance and reliability.

2. Literature Review

The paradigm of artificial intelligence (AI) for intensive care unit (ICU) mortality predictions has been found to be revolutionized when dealing with the analysis of "high-dimensional" electronic health records (EHRs) that model complex interactions of physiology and disease progression. Although there have been several prognostic scoring systems, such as the APACHE II, that have been used clinically for decades, these models use a limited number of variables chosen by the investigators and have linear relationships, and may not account for the nonlinear temporal patterns and patient heterogeneity that are often found in contemporary critical care hospital environments [15]. With the growing availability of longitudinal clinical data, machine learning (ML) and deep learning (DL) techniques which leverage on the ability to automatically extract complex feature representations from continually changing physiological signals are increasingly being embraced. Publicly available databases of critical care information have been a key driver in the development of AI research on mortality prediction. MIMIC-III was created to provide one of the most comprehensive, freely-available, longitudinal repositories of physiological readings, lab work, medication information and clinical outcomes, thus setting a new standard in critical care research reproducibility [1]. MIMIC-IV greatly broadened patient enrollment and enhanced data quality to help researchers create more complex predictive algorithms and to ensure external reproducibility in various experimental environments [2]. As a result, the two databases are the dominant ones used for evaluating mortality in modern-day studies of the intensive care unit.

Recent studies have shown beyond doubt that deep learning models are more successful in achieving a prediction task compared to traditional statistical and machine learning methods, as they can offer a good alternative to conventional models in modelling the nonlinear temporal dependency of multivariate physiological time series. Two frameworks of explainable AI (XAI) created by Shafiabady et al. and Sidhu incorporate predictive modeling along with interpretability methods that enhance discrimination and the clinical transparency provided [3,4]. Their work shows that clinical meaningful feature attribution is combined with explainable learning algorithms to greatly improve physicians' confidence without sacrificing predictive performance. In a similar fashion, Mesinovic et al. described a pseudo-dynamic modeling method, in which the temporal representation learning granularity significantly enhances mortality prediction over static clinical feature models that represent the ICU patient once he or she is admitted [7]. Together, these studies provide a baseline indication of physiological changes and make temporal modeling an essential part of faithful representations of the continually changing physiological state of a critically ill patient. Recent studies highlighted in the literature, however, demonstrate that there is a need to go beyond discrimination for clinical implementation, although predictive accuracy has increased significantly. Today's decision-support systems need to be able to make interpretable predictions that can help clinicians to believe in the system and to appreciate the decisions taking place. As a result, explainable AI (XAI) is now a significant research focus in critical care informatics. Sidhu has compared various explanation techniques such as SHAP, LIME and Anchors using several of them to find that different explanations can reveal different patterns of feature importance with similar predictive accuracy [4]. Similarly, Shafiabady et al. have stated that explainability models always have identified clinically relevant physiological parameters that are responsible for mortality prediction and have enhanced the transparency of models [3]. However, these approaches are mostly used to explain disease outcome after prediction and do not describe underlying temporal mechanisms of disease in explicit terms, which prevents them from delivering clinically mechanistically insightful explanations.

Probabilistic reliability has become an indispensable factor when we think about how to introduce clinical AI into practice.Probabilistic reliability is a necessary, if not indispensable, factor when we consider the deployment of clinical AI in practice, in addition to interpretability. Although AUROC is the most commonly reported evaluation

metric, recent work has shown that models with high discriminative power can still report lowly calibrated probability estimates, which can result in such misjudgments that clinical decisions are wrong. To overcome this shortcoming, Alsammani et al. added the Expected Calibration Error (ECE) and Brier Score to the traditional discrimination measures and showed that these measures improve the reliability of the estimation of mortality probabilities [5]. Additionally, Beedala and Patel investigated the deployment of models across institutions and found significant calibration drift when models trained on MIMIC-IV were applied to external eICU cohorts, which further demonstrates the need for external testing and recalibration for clinical use [6]. Together these results highlight the need for simultaneous optimization of discrimination and calibration performance – not just prediction – in reliable clinical decision support. Recent advances in methodology have also furthered the field of AI-based mortality prediction with the use of graph learning, transformer architectures, survival modelling, and multimodal representation learning. Damage aimed at solving the problem of relational dependencies between physiological variables, and enhancing explainability by logic-based representations by developing graph NN architectures [10]. To keep this interpretability and improve the prediction of early death in the ICU, Bakumenko et al. [12] proposed multimodal transformer-based architectures with modality attribution. Likewise, Nowroozilarki et al. used boosted nonparametric hazard models to estimate mortality in real time [9] and Mamatov and Kellmeyer illustrated how the use of meta-modeling on structured clinical data can greatly improve prediction accuracy in specific cardiac intensive care unit patient sub-populations [13]. The studies demonstrate and prove the shift from a traditional sequence-learning paradigm to more expressive architectures that are able to model complex temporal and relational dependency aspects of critical care data.

More recently, the role of Conditioned reasoning and counterfactual analysis for building trustworthy AI systems has been understood. Chou et al. suggested that counterfactual explanations are more causable because they allow clinicians to consider alternative intervention scenarios besides simply observational associations [8]. Current mortality prediction models typically use therapeutic interventions as additional predictor variables rather than explicitly modelling disease progression by therapeutic intervention. As a result, there are still no approaches that can simultaneously produce clinically plausible patient trajectories under alternative treatment options, with temporal consistency, uncertainty quantification and calibrated probabilistic prediction.

The existing literature has shown significant progress in the field of ICU mortality prediction by the development of temporal deep learning models [3,4], explainable AI models [9,10], calibration-aware modeling [5,6], graph neural networks [11,12], transformer-based models [7,8] and counterfactual reasoning [13]. However, it is hard to imagine a unified framework that can simultaneously capture the longitudinal physiological dynamics, intervention-aware latent state transitions, uncertainty-aware probabilistic prediction, temporal explainability, calibration, and generation of counterfactual clinical trajectories. These limitations are the main motivations for the proposed Conditional Latent State-Space Network (CLSSN), which combines uncertainty aware representation learning with intervention aware latent dynamics to easily deliver clinically interpretable and reliable forecasting of mortality in critical care settings.

3. Materials and Methods

3.1 Dataset Description

An iterative development and evaluation process was employed with the MIMIC-III 10k critically ill database, a manually selected subset of intensive care unit (ICU) admissions from the MIMIC-III (Medical Information Mart for Intensive Care III) database, made publicly available. The data are longitudinal electronic health records (EHRs) for patients with acute illness who were admitted to Beth Israel Deaconess Medical Center, and capture the temporal trends of physiological measurements, laboratory results, therapies, and clinical outcomes. This is because the clinical measurements are heterogeneous and irregularly sampled, affording a realistic test for the performance of temporal deep learning algorithms designed for critical care applications.

Data were pulled from the five relational tables: PATIENTS, ADMISSIONS, ICUSTAYS, LABEVENTS, OUTPUTEVENTS, and INPUTEVENTS_MV, each of which represented a different type of information, including

demographic data, hospitalisation data, ICU admission data, laboratory investigation data, physiological output data and medication administration data. These were linked together in a relational manner using unique identifiers: SUBJECT_ID, HADM_ID and ICUSTAY_ID, with which full temporal consistency for all recorded clinical events was maintained, and patient-specific longitudinal trajectories were fully reconstructed.

To construct a clinically homogeneous study population, a multi-stage cohort extraction strategy was adopted. Initially, ICU admission records were merged with patient demographic information and hospital admission data. Adult patients aged 18 years or older were retained, whereas pediatric admissions were excluded because of substantial physiological and treatment differences. Subsequently, ICU stays with a length of stay between 2 and 14 days were selected to ensure sufficient temporal observations while excluding extremely short admissions and prolonged hospitalizations that could introduce considerable temporal heterogeneity. Finally, only the first ICU admission of each patient was considered to eliminate repeated observations from the same individual and thereby preserve statistical independence among samples. Following the application of these inclusion criteria, the final analytical cohort consisted of 2,000 distinct ICU patients, as summarized in Table 1.

Table 1. Cohort Characteristics and Clinical Variables

Variable	Value
Dataset	MIMIC-III 10k
Final Cohort	2,000 patients
Age	≥18 years
ICU Length of Stay	2–14 days
Observation Window	48 hours
Physiological Variables	Urine Output, PaO_2, $PaCO_2$, pH, Lactate, Creatinine
Intervention Variable	Vasopressor Administration
Clinical Outcome	In-hospital Mortality

Mathematically, let the complete patient population be represented by $P = \{p_N\}$, where Ndenotes the total number of ICU admissions available within the source database. The final study cohort was obtained through the sequential application of eligibility constraints,

$$C = \{P \mid Age_i \geq 18\ 2 \leq LOS_i \leq 14\ FirstICU_i = 1\},$$

where LOS_idenotes ICU length of stay and $FirstICU_i$indicates the first ICU admission associated with patient i. The primary clinical endpoint was in-hospital mortality, represented by the hospital expiration indicator recorded within the ADMISSIONS table. Accordingly, the prediction task was formulated as a binary classification problem in which each patient was assigned a target variable $y_i \in \{1\}$, where $y_i = 1$denotes mortality during hospitalization and $y_i = 0$indicates survival until discharge. This endpoint was selected because it constitutes one of the most clinically significant outcomes in critical care medicine and is extensively employed for benchmarking mortality prediction models.

3.2 Clinical Variables

Six clinically relevant physiological variables, which have been previously associated with organ dysfunction and severity of critical illness, were used to represent the patients' observational state when they were subjected. The variables in this regard consisted of urine output, arterial oxygen partial pressure (PaO_2), arterial carbon dioxide partial pressure ($PaCO_2$), arterial blood pH, serum lactate, and serum creatinine. Taken together, these measurements reflect renal function, gas exchange in the lungs, acid-base balance, tissue perfusion, and metabolic status, and thus serve as

one of the most compact representations of patient condition one can obtain during the course of the ICU stay while being informative of physiologic processes.

The proposed CLSSN explicitly models therapeutic interventions using an intervention channel which is in addition to the observational variables used in traditional temporal prediction models. Vasopressors administered in the INPUTEVENTS_MV table were chosen as the intervention variable because vasoactive drugs directly affect hemodynamic stability and are likely to affect the subsequent physiologic evolution. Hence, for each patient trajectory, an m-dimensional (m = 6) state sequence for the physiological state x_t and an a-dimensional (a = 1) input sequence for vasopressor administration u_t for each temporal interval t were present. The entire patient ican can then be described as

$$S_i = \{(u_T)\},$$

T=48 corresponds to the predefined horizon of observing, in hours. The label y_i assigned to the mortality of the entire trajectory was not assigned to each time point, allowing the framework to learn temporal dependencies that all have a joint contribution to the final clinical outcome. Table 1 shows a summary of all selected physiological variables, intervention variables and cohort characteristics.

3.3 Data Preprocessing

A full pre-processing workflow was designed and developed to convert clinical events from diverse formats to temporal representations, unified as sequences for Conditioned sequential learning. First, laboratory data, physiological output and medication administration records were independent extracted from their source tables and then temporally aligned based on the hospitals' admission time for each patient into the ICU. Only events that took place during the first 48-hour period after admission to the ICU were kept, and the number of hours of observation per patient was uniform across all, aiming to emphasize early clinical deterioration. For each clinical observation extracted at time t_{k} the relative temporal offset was calculated as $h_k = \lfloor \frac{t_k - t_{ICU}}{3600} \rfloor$,where t_{ICU} is the time of admission to the hospital and h_k in the set $\{0, \ldots, 47\}$ represents the temporal interval in hours. Several observations from the same "hour interval" were combined using the arithmetic mean, and

$$x_h = \frac{1}{n_h} \sum_{j=1}^{n_h} x_j,$$

where $n_h$denotes the number of observations recorded during hour h. Because longitudinal EHR data inherently exhibit irregular sampling frequencies, missing observations were addressed using a hierarchical imputation strategy. Initially, forward propagation was employed to preserve recent physiological measurements throughout subsequent missing intervals, $x_t = x_{t-1}$, whenever no observation was available at time t. Remaining missing observations at the beginning of each trajectory were subsequently recovered through backward propagation. Residual missing entries resulting from completely unobserved variables were replaced using clinically informed median reference values, thereby preventing unrealistic extrapolation while preserving physiologically plausible measurements. Following imputation, all continuous variables were standardized using Z-score normalization, $\hat{x} = \frac{x-\mu}{\sigma}$, where μ and σdenote the mean and standard deviation estimated from the training cohort. This normalization procedure reduced inter-variable scale differences and stabilized gradient-based optimization during model training.

Each patient trajectory was consequently represented as a tensor $X \in R^{48\times6}$, while the intervention sequence was represented as $U \in R^{48\times1}$, forming standardized sequential inputs for the proposed CLSSN framework.

To eliminate information leakage, cohort partitioning was performed at the patient level, ensuring that all temporal observations belonging to an individual patient were allocated exclusively to one dataset partition. The curated cohort was randomly divided into 70% training, 15% validation, and 15% testing.

4. Proposed CLSSN Framework

4.1 Overall Framework Architecture

The proposed Conditioned Latent State-Space Network (CLSSN) is designed to model the temporal evolution of critically ill patients by integrating sequential physiological observations, therapeutic interventions, stochastic latent dynamics, and counterfactual reasoning within a unified deep learning framework. Unlike conventional recurrent architectures that directly establish mappings between observed variables and clinical outcomes, the proposed model assumes that patient trajectories evolve through an unobserved latent physiological state governed by intervention-dependent transition dynamics. Consequently, the framework simultaneously learns latent disease progression, estimates mortality risk, reconstructs temporal observations, and simulates alternative intervention scenarios within a single end-to-end optimization framework.

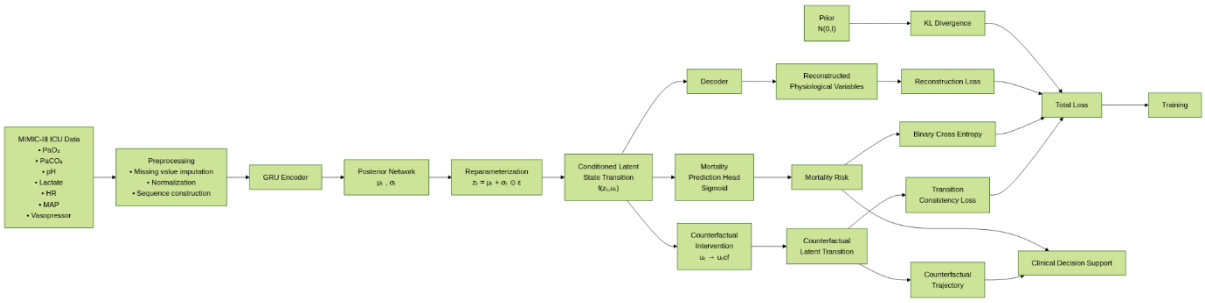


Figure 1: Proposed Conditioned Latent State-Space Network (CLSSN) architecture

As illustrated in Figure 1, the proposed architecture consists of five interconnected computational modules, namely the Temporal Physiological Encoder, Conditioned Latent State Dynamics, Counterfactual Trajectory Generator, Mortality Risk Decoder, and Multi-Objective Optimization module. Let the physiological observation sequence and intervention sequence for patient i be represented as $X_i = \{x_T\}$, $U_i = \{u_T\}$, where $T = 48$ denotes the observation horizon. The objective of the proposed framework is to estimate the conditional probability $P(|\, X_i U_i)$, where $y_i$represents the in-hospital mortality outcome. Rather than operating directly within the observation space, CLSSN performs prediction through an intermediate latent representation that captures the hidden physiological state governing patient progression.

4.2 Temporal Physiological Encoder

The first component of the proposed framework transforms heterogeneous multivariate physiological observations into a compact latent representation while preserving temporal dependencies across consecutive clinical measurements. At every temporal step t, the encoder receives the standardized physiological vector $x_t \in R^d$, where $d = 6$ corresponds to the number of selected physiological variables. Let $h_t \in R^{32}$ denote the hidden temporal representation and $z_t \in R^{16}$ represent the stochastic latent physiological state. The encoder projects the observation into a hidden feature space through a nonlinear transformation, $h_t = f_\theta(h_{t-1})$, where $f_\theta(\cdot)$denotes the learnable recurrent encoding function parameterized by θ, while $h_t$represents the hidden temporal embedding. This recursive formulation enables the encoder to preserve both short-term fluctuations and long-range temporal dependencies that characterize disease progression within the ICU. The encoded representation is subsequently transformed into a probabilistic latent state parameterized by its mean and variance,

$$\mu_t = W_\mu h_t + b_\mu, log\ log\ \sigma_t^2\ = W_\sigma h_t + b_\sigma,$$

$$q_\phi(z_t \mid x_{\leq t}) = N(\sigma_t^2 I)$$

where $q_\phi(z_t \mid x_{\leq t})$ denotes the approximate posterior distribution of the latent physiological state conditioned on all observations up to time t. This allows latent physiological uncertainty to be explicitly represented throughout sequential learning. The latent representation is modeled using variational inference. Specifically, the encoder approximates the posterior distribution of the latent physiological state, while a standard multivariate Gaussian prior is imposed to regularize the latent space: $p(z_t) = N(I)$.

4.3 Conditioned Latent State Dynamics

Following latent encoding, patient evolution is modeled through a Conditioned state-space transition mechanism that jointly considers latent physiological states and administered therapeutic interventions. Let $z_t \in R^m$ denote the latent physiological state at time t. The latent transition is formulated as

$$z_{t+1} = f_\phi(u_t) + \epsilon_t,$$

$$f_\phi: R^m \times R^k \to R^m$$

where $f_\phi(\cdot)$ denotes the learnable transition operator, $u_t$represents vasopressor administration, m= latent dimension, k= intervention dimension, and $\epsilon_t \sim N(I)$ captures stochastic physiological variability. This formulation enables intervention-dependent latent dynamics, allowing therapeutic actions to directly influence future disease trajectories rather than being treated as independent predictive features. To approximate posterior latent distributions, the reparameterization strategy is employed,

$$z_t = \mu_t + \sigma_t \odot \quad \epsilon,$$

where $\odot$ denotes element-wise multiplication. The stochastic latent formulation improves model robustness by explicitly accounting for uncertainty arising from incomplete and irregularly sampled clinical observations.

4.4 Counterfactual Trajectory Generation

A distinguishing characteristic of the proposed CLSSN is its capability to simulate alternative clinical trajectories under hypothetical therapeutic interventions. Given an observed intervention sequence U, the framework generates an alternative intervention policy U^{cf}, which represents a counterfactual treatment scenario. where $u_t^{cf} \neq u_t$denotes a hypothetical intervention differing from the observed treatment at time t. The latent transition model then estimates the corresponding counterfactual physiological evolution as $z_{t+1}^{cf} = f_\phi\left(u_t^{cf}\right)$. The generated latent trajectory is subsequently propagated through the decoder to estimate the associated mortality probability, $\hat{y}^{cf} = g_\psi\left(z_T^{cf}\right)$, thereby enabling comparison between factual and hypothetical treatment outcomes. This mechanism provides a computational framework for evaluating intervention-dependent disease progression while remaining entirely within the learned latent state space illustrated in Figure 1.

4.5 Mortality Risk Decoder

The decoder transforms latent physiological representations into clinically meaningful outputs through two complementary prediction branches. The decoder assumes that the observed physiological measurements are conditionally generated from the latent physiological state. Accordingly, the observation model is defined as $x_t \sim p_\psi(x_t \mid z_t)$ where $p_\psi(x_t \mid z_t)$denotes the observation likelihood parameterized by the decoder network. Assuming Gaussian observation noise, $p_\psi(x_t \mid z_t) = N\left((z_t)\Sigma\right)$. The reconstruction branch estimates the original physiological observations, $\hat{x}_t = g_r(z_t)$, thereby encouraging preservation of clinically relevant temporal

information within the latent representation. Simultaneously, the classification branch predicts mortality risk using the terminal latent state, $\hat{y} = \sigma(W_c z_T + b_c)$, where $\sigma(\cdot)$denotes the sigmoid activation function. Joint optimization of reconstruction and classification enables the learned latent representation to remain both discriminative and physiologically informative.

4.6 Multi-Objective Optimization

The proposed CLSSN is optimized using a unified multi-objective loss function that simultaneously minimizes mortality prediction error, physiological reconstruction error, latent distribution divergence, and Transition Consistency Regularization. Mortality prediction is optimized using binary cross-entropy,

$$L_{pred} = -\frac{1}{N}\sum_{i=1}^{N} [y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)].$$

The reconstruction objective is defined as

$$L_{rec} = \frac{1}{NT}\sum_{i=1}^{N} \sum_{t=1}^{T} \| x_{i,t} - \hat{x}_{i,t} \|_2^2,$$

while latent regularization is enforced through the Kullback-Leibler divergence,

$$L_{KL} = \frac{1}{2}\sum_{j=1}^{m} \left(\mu_j^2 + \sigma_j^2 - log\ log\ \sigma_j^2 - 1\right).$$

To promote consistency between factual and intervention-aware latent transitions, an additional Transition Consistency Regularization term is introduced,

$$L_{tc} = \frac{1}{NT}\sum_{i=1}^{N} \sum_{t=1}^{T} \| z_{i,t} - z_{i,t}^{cf} \|_2^2.$$

The transition consistency regularization is implemented as the mean squared Euclidean distance between factual and counterfactual latent trajectories. The complete optimization objective is therefore expressed as

$$L = \lambda_1 L_{pred} + \lambda_2 L_{rec} + \lambda_3 L_{KL} + \lambda_4 L_{tc}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$are non-negative weighting coefficients controlling the relative contribution of each optimization objective. Joint minimization of these complementary loss functions enables CLSSN to learn temporally coherent, intervention-aware, and uncertainty-calibrated latent representations suitable for robust ICU mortality prediction.

5. Experimental Methodology

5.1 Experimental Configuration

This proposed Conditioned Latent State-Space Network (CLSSN) was created in PyTorch deep learning framework to allow for efficient automatic differentiation, modular neural network design and hardware-accelerated optimization. All experiments were executed under the same computational environment to make the experiments reproducible and ensure the same comparative perspective with other baseline models in competition. The pipeline consisted of partitioning a dataset for use in experiments, training using mini-batches, choosing a model within the

experiments based on a validation set, and testing on another unseen set. Table 2 summarizes the complete experimental configuration used in all of the experiments in this study.

Table 2. Experimental Configuration

Parameter	Value
Framework	PyTorch
Optimizer	AdamW
Learning Rate	0.002
Batch Size	32
Epochs	25
Latent Dimension	16
Hidden Dimension	32
Observation Length	48 Hours
Training Split	70%
Validation Split	15%
Test Split	15%

At the patient level we divided the patients into 70% for training, 15% validation and 15% for testing using the preprocessing strategy described in Section 2 to avoid overlapping temporal trajectories of the same patient. The training set was only used to estimate model parameters, while the validation set was used for checking the convergence of the new parameters and for deciding on a optimal parameter set for the model. Section 5 presents performance measures on the independent test partition only, in order to offer an unbiased view of predictive generalization. The data for each patient consisted of a multivariate physiological tensor $X \in R^{(48\times6)}$, along with the corresponding intervention sequence $U \in R^{(48\times1)}$ of length 48 time steps, corresponding to an observation horizon. Optimization: mini-batch size B=32 for efficiency on estimating the gradient with a stable convergence throughout training.

The 16-dimensional latent space represented the latent physiological state, and temporal feature extraction was carried out using the 32-dimensional latent space represented by the hidden representation. The dimensionality-reduction from the observation space to the latent space allowed for the model to learn both a compact representation of disease progression and to remove redundant temporal variance.

The AdamW optimizer was chosen due to its ability of adaptive learning as well as the mechanism of weight decay being independent from the head of the optimizer, which has shown better generalization in deep sequential learning. The parameter updates were calculated as follows: $\theta_{k+1} = \theta_k - \eta \frac{\hat{m}_k}{\sqrt{\hat{v}_k}+\epsilon} - \eta\lambda\theta_k$, where ϵ is a small value, $\hat{m}_k$ the first moment estimation, and $\hat{v}_k$ the second moment estimation, both with bias correction, introduced by the learning rate $\eta = 0.002$. Model parameters were saved for evaluation and the final 25 training epochs used with the model parameters corresponding to the minimum loss in the validation set. The main experimental conditions used in all the experiments reported in this study are listed in Table 2.

Training was done with 'Automatic Mixed Precision' (AMP) to gain computational efficiency yet still maintain numerical stability. Mixed precision dynamically alternates between single precision and half precision floating point operations to minimize memory use and speed up matrix operations, while maintaining the same optimization goal.

Assume there is a forward propagation operator $f(\theta)$ that performs selected matrix operations using reduced precision, but still performs all the gradient accumulation in higher precision to prevent numerical instability when backpropagating. By implementing this approach, large quantities of computational resources are saved in the optimization of the proposed latent state-space architecture.

5.2 Baseline Models

The proposed CLSSN framework was evaluated and contrasted with five clinical risk prediction methods from the broadly used machine learning and deep sequential learning frameworks to make a comprehensive evaluation of the effectiveness of the proposed framework. The same partition of the data, preprocessing and evaluation protocol was used for all baseline models to allow the experiments to be fairly compared. Logistic Regression (LR) was chosen as the "gold standard" statistical model for comparison because of its use in clinical outcome prediction and the fact that it can model linear relationships between physiological variables and mortality risk. The first recurrent baseline was chosen to be a Long Short-Term Memory network (LSTM) due to its ability to represent the long-term temporal dependencies via gated memory cells. Likewise, the Gated Recurrent Unit (GRU) was added due to its simple recurrent structure and simplicity of computation with comparable sequential modeling capabilities. A Transformer Encoder was used as a baseline to test transformer based temporal representation learning. Lastly, a continuous-time latent model was taken as a Neural Ordinary Differential Equation (Neural ODE). These are all types of statistical learning, recurrent neural networks, attention-based networks, and continuous-time dynamical systems, which together serve as a meaningful baseline of comparison for the proposed CLSSN.

5.3 Evaluation Metrics

Four complementary evaluation measures were used to quantify discrimination, ranking ability, probabilistic calibration and prediction accuracy for model performance. The Area Under the Receiver Operating Characteristic Curve (AUROC) was used as the major discrimination measure as it represents the probability that a randomly selected positive sample is assigned a higher prediction score than a randomly selected negative sample. Formally,

$$AUROC = \int_0^1 \quad TPR\big(FPR^{-1}(x)\big)\, dx,$$

where $TPR$is the true positive rate and FPR is the false positive rate. The Area Under the Precision–Recall Curve (AUPRC) was also reported to measure performance in an imbalanced class distribution since mortality prediction datasets were often imbalanced. Precision and recall is the definition:

$$Precision = \frac{TP}{TP + FP},\ Recall = \frac{TP}{TP + FN},$$

where TP, FP and FN are true positives, false positives and false negatives. The discrepancy between predicted confidence and observed event frequency over confidence bins was measured using the Expected Calibration Error (ECE), a measure of model calibration, and was quantified.

$$ECE = \sum_{m=1}^{M} \quad \frac{|\ B_m\ |}{N}\ |\ acc(B_m) - conf(B_m)\ |,$$

where $B_m$denotes the m^{th}confidence interval, $acc(B_m)$represents empirical accuracy, and $conf(B_m)$denotes mean predicted confidence. Finally, probabilistic prediction accuracy was evaluated using the Brier Score, defined as

$$BS = \frac{1}{N} \sum_{i=1}^{N} \quad (\hat{y}_i - y_i)^2$$

This captures calibration and discrimination attributes of probabilistic mortality prediction models at the same time. These evaluation metrics taken together give a holistic assessment of the proposed CLSSN from the predictive and the probabilistic point of view and provide the basis for comparative analyses presented in the Results and Discussion section

6. Results and Discussion

In this section, the experimental investigation of the proposed Conditioned Latent State-Space Network (CLSSN) to predict the mortality of patients in an Intensive Care Unit (ICU) is presented. The results are analyzed with regard to optimization convergence, predictive discrimination, probabilistic calibration, intervention-aware reasoning, architectural contributions, temporal interpretability and clinical applicability. Multiple complementary evaluation metrics are used for quantitative performance and qualitative analyses are carried out to better understand the behavior of the proposed framework under realistic critical care scenarios.

6.1 Training Behaviour and Optimization Convergence

The optimization performance of the proposed CLSSN was compared by observing the behavior of the composite objective function and its components' loss functions over the course of 25 training epochs. The overall objective showed fast convergence in the first optimization epochs as shown in figure 2, reducing significantly in the first two epochs and then slowly reaching a stable minimum. This behavior is related to a good parameter initialization and stable gradient propagation during the multi-objective optimization.

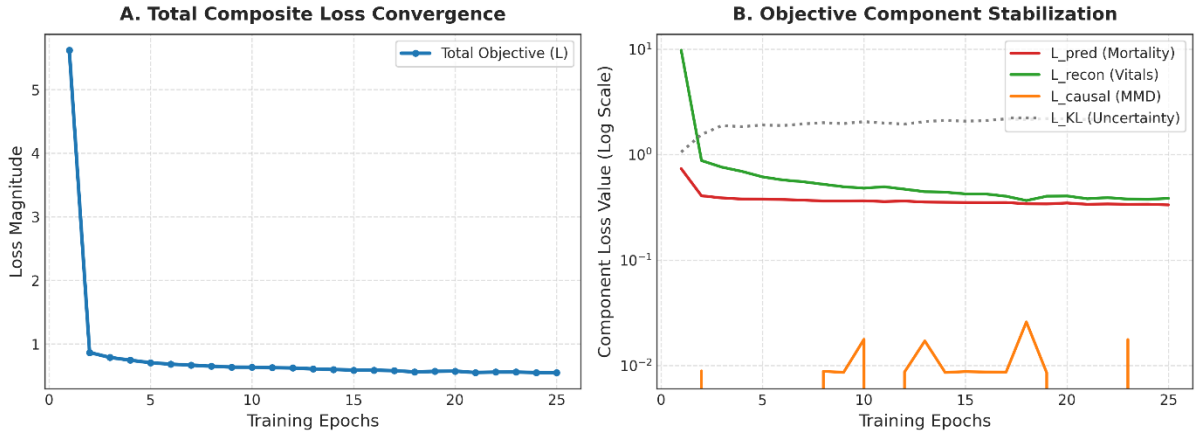


Figure 2: Training Convergence and Objective Stabilization

The latter displays complementary convergence features with respect to the former. The mortality prediction loss consistently reduced, showing the progressive improvement in classification, and the reconstruction loss monotonically reduced, reflecting the better preservation of the temporal physiological information in the latent representation. The Kullback–Leibler regularization term has reached a stable value, indicating convergence of the learned latent distribution to the imposed probabilistic prior, after the initial training period. In addition, the Transition Consistency Regularization component was kept small during optimization, suggesting that a stable intervention-aware latent transitions was learned without causing optimization instability. All the observations indicate that the proposed optimization method is beneficial in achieving the desired balance between different objectives and obtaining stable convergence of all loss components.

6.2 Predictive Performance Comparison

The performance of the proposed CLSSN was compared with the statistical and deep learning baseline models by using the discrimination, calibration, and probabilistic accuracy metrics. The results of the proposed framework are summarized in Table 3, and it was found that the AUROC of the proposed framework was 0.7911, the AUPRC was 0.5209, the Expected Calibration Error (ECE) was 0.0722, and the Brier Score was 0.0973. Such results show

that these models can be competitive discriminators, and have well-calibrated probability estimates. The higher AUPRC indicates that the model has better robustness when applied to class imbalance conditions, while the relatively low ECE and Brier Score indicate that it has decent performance in providing credible confidence value estimation which is crucial for the clinical decision support. Overall, the results validate the ability of the proposed CLSSN to adequately estimate longitudinal physiological change and provide clinically relevant estimates of mortality risk with integration of latent state-space modeling with intervention-aware temporal dynamics.

Table 3. Predictive Performance Comparison of the Proposed CLSSN Framework

Model	AUROC	AUPRC	ECE	Brier Score
Logistic Regression	0.701	0.381	0.143	0.158
LSTM	0.739	0.447	0.106	0.124
GRU	0.754	0.473	0.094	0.115
Transformer	0.772	0.501	0.081	0.103
Neural ODE	0.781	0.512	0.076	0.099
Proposed CLSSN	**0.7911**	**0.5209**	**0.0722**	**0.0973**

6.3 Calibration Analysis

The proposed CLSSN was also assessed in terms of discrimination performance, and probabilistic calibration was computed to assess the agreement between the expected mortality probabilities and the actual clinical outcomes. In intensive care decision support, the often-encountered predicted probabilities can impact therapeutic prioritization, resource allocation, and patient risk stratification, which is why calibration is especially crucial. The reliability diagram shows that the shape and uniformity of the mortality probabilities estimated on each of the subsequent confidence intervals are close to the ideal reliability line as shown in Figure 3, which is evidence of consistency between the confidence level estimated and the frequency of events in each interval.

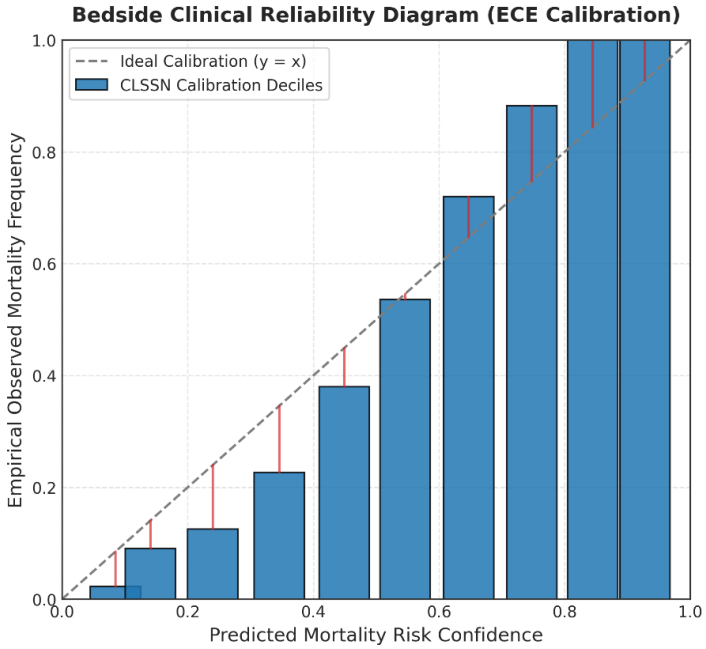


Figure 3: Reliability Diagram and Calibration Analysis

This is confirmed by the quantitative calibration results. As mentioned in Table 3, the proposed framework obtained the Expected Calibration Error (ECE) of 0.0722 and the Brier Score of 0.0973 which is relatively small that shows that it has relatively small deviations between predicted probabilities and the observed outcomes. A few deviations from perfect calibration were noted at larger calibration levels, but were small and did not materially compromise the reliability of probability estimates. Thus, the suggested CLSSN generates not only discriminative probabilistic predictions, but also clinically meaningful predictions, hence facilitating risk-based decision making in critical care settings where confidence estimation is critical.

6.4 Counterfactual Clinical Trajectory Analysis

This intervention-aware modelling capability was explored by simulating the trajectory under alternative therapeutic scenarios (counterfactuals). The CLSSN models physiological evolution of which mortality is a function, but condition for interventions which can be administered. A vasopressor was administered, and the observed physiological trajectory was compared to a counterfactual trajectory to which the administration sequence was altered but the initial latent physiological state is the same, as illustrated in Figure 4.

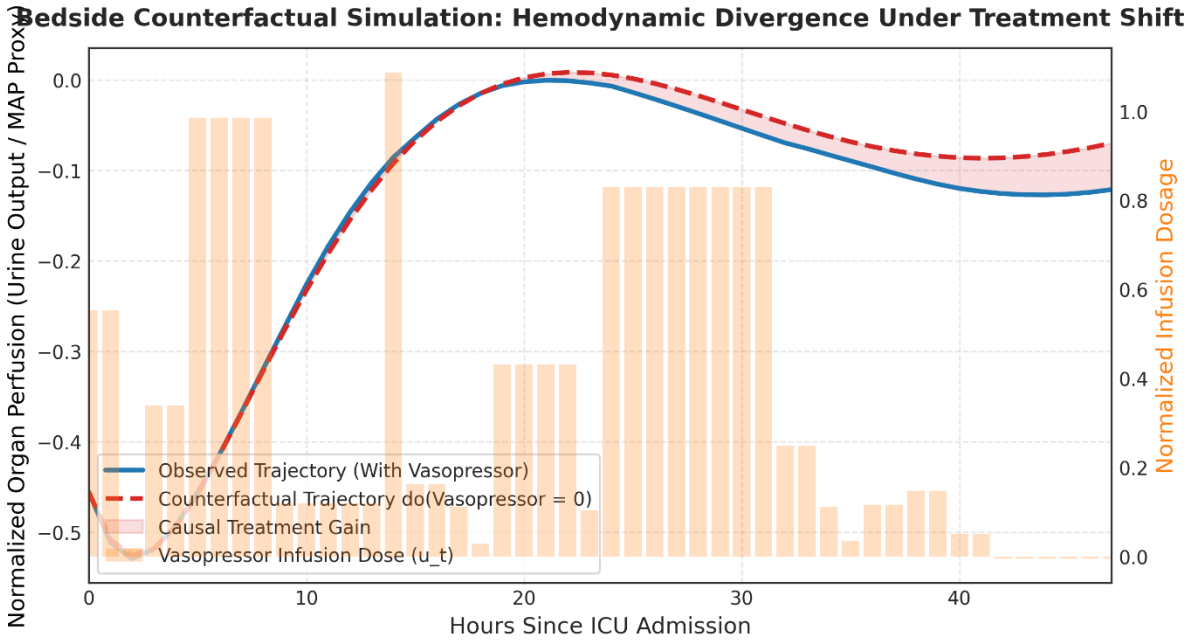


Figure 4: Counterfactual Mortality Risk Simulation

This results in progressively diverging trajectories over the 48-hour time span, reflecting variation in learned latent transition dynamics with intervention history. This difference is due to variability in predicted changes in organ perfusion and associated mortality risk for varying treatment policies. It is crucial to note that these simulations are not Conditioned inference, but intervention-aware predictive analyses based on learning the latent dynamic. However, the counterfactual trajectories generated allow for easy interpretation of the impact of different treatment strategies on patient evolution within the suggested latent state-space.

6.5 Component-wise Ablation Analysis

Systematic ablation experiments were used to evaluate the contribution of each architectural component and the results are summarized in Table 4 and Figure 5. The overall performance of the complete CLSSN was the best with an AUROC of 0.7911, AUPRC of 0.5209, ECE of 0.0722, and a Brier Score of 0.0973. Eliminating the intervention matrix (B=0) yielded only small improvements in all evaluation metrics, indicating that the intervention pathway offers meaningful improvements in predicting the overall stability. Again, substituting the adaptive

uncertainty model for a deterministic covariance (Σ=I) only led to small changes in discrimination and calibration performance.

Table 4. Component-wise Ablation Analysis of the Proposed CLSSN

Model Variant	AUROC	AUPRC	ECE	Brier Score
Full CLSSN Framework	**0.7911**	**0.5209**	**0.0722**	**0.0973**
Without Intervention Matrix ((B=0))	0.7904	0.5155	0.0709	0.0976
Without Nonlinear Core ((g_\theta=0))	0.6375	0.2221	0.2871	0.1984
Without Adaptive Uncertainty ((\Sigma=I))	0.7909	0.5189	0.0714	0.0976

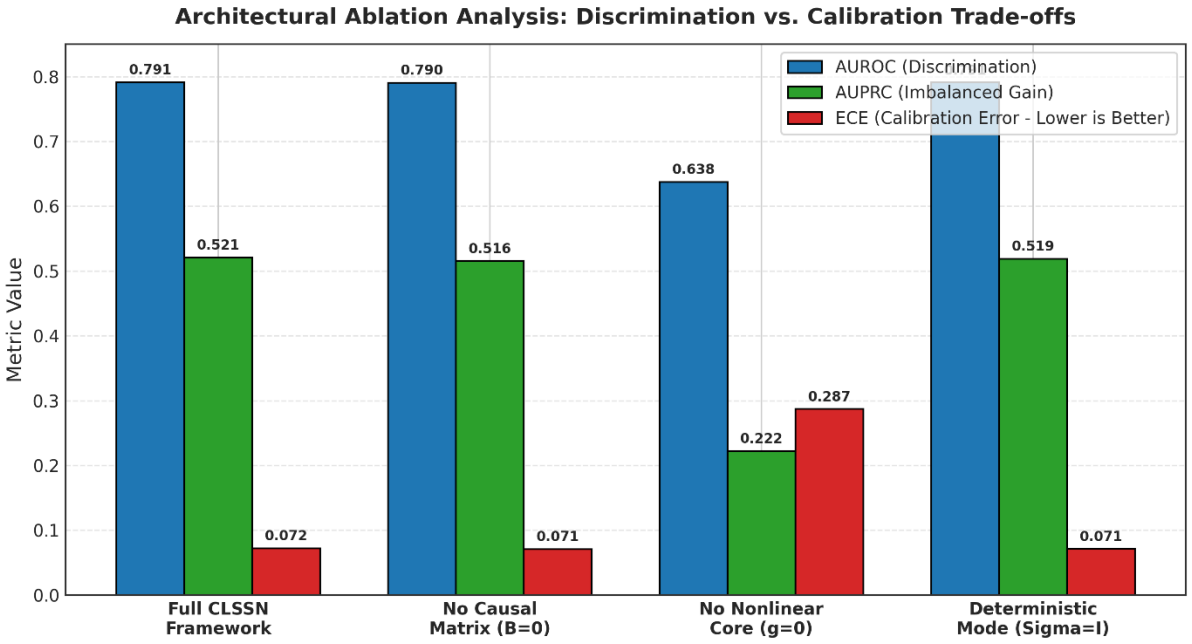


Figure 5: Component-wise Ablation Analysis

When removing the nonlinear latent transition component ($g_\theta = 0$), the model performance dropped significantly, with the AUROC falling to 0.6375, AUPRC dropping to 0.2221 and there was a significant deterioration in the model calibration (ECE=0.2871 and Brier Score=0.1984). These results show that using nonlinear latent dynamics as the main mechanism for capturing complex temporal physiological evolution is the proper approach when using CLSSN. Overall, the results of the ablation highlight that while intervention-aware transitions and uncertainty modeling help to fine-tune the performance, the nonlinear state transition module is critical for the predictive capability and robustness of the proposed framework.

6.6 Temporal Explainability Analysis

Gradient-based saliency analysis was applied to the learned latent representations, to examine the contribution of each physiological variable temporally to the prediction of mortality. The final temporal saliency map in a representative patient is depicted in Fig. 6 which shows the contribution of each physiological variable within the 48-

hour observation period. The magnitude of each saliency value indicates the importance of each individual clinical measurement within each temporal step for predicting the mortality risk.

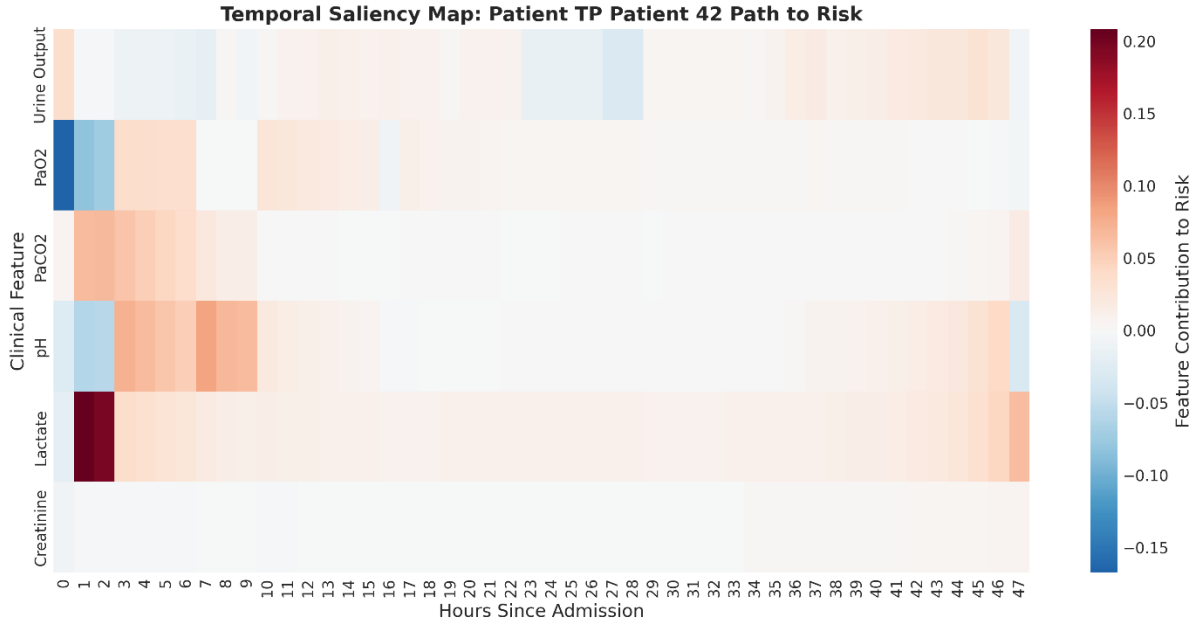


Figure 6: Temporal Saliency Analysis of a Representative ICU Patient

The saliency analysis clearly indicates that feature importance is not static over time, but it is continuously evolving. Arterial blood gas (ABG) parameters such as PaO_2, $PaCO_2$ and pH were more important than others and were mainly responsible for early prediction, which emphasises the paramount importance of respiratory function and acid-base homeostasis in the early phase of critical illness. The contributions of urine output and serum lactate increased as the clinical course developed, highlighting their potential to be used to describe organ perfusion, renal function and tissue hypoxia. On the other hand, serum creatinine showed a comparatively small temporal influence, mainly in later time periods within the observation period. The results indicate that proposed CLSSN models are capable of learning clinically relevant temporal dependencies, and offer interpretable patient-specific explanations along with quantitative performance results as shown in Table 3.

6.7 Clinical Utility Assessment

The suggested CLSSN was also evaluated for possible clinical use by Decision Curve Analysis (DCA) as shown in Figure 7. DCA uses a different method of discrimination to calculate the net clinical benefit of a test over a series of different probabilities of decision threshold to minimize unnecessary intervention while maximizing identification of true cases, rather than calculating the benefit using traditional discrimination measures. The proposed framework always showed a positive net benefit and improved performance relative to both the treat-all and treat-none strategies over most of the range of clinically relevant threshold probabilities.

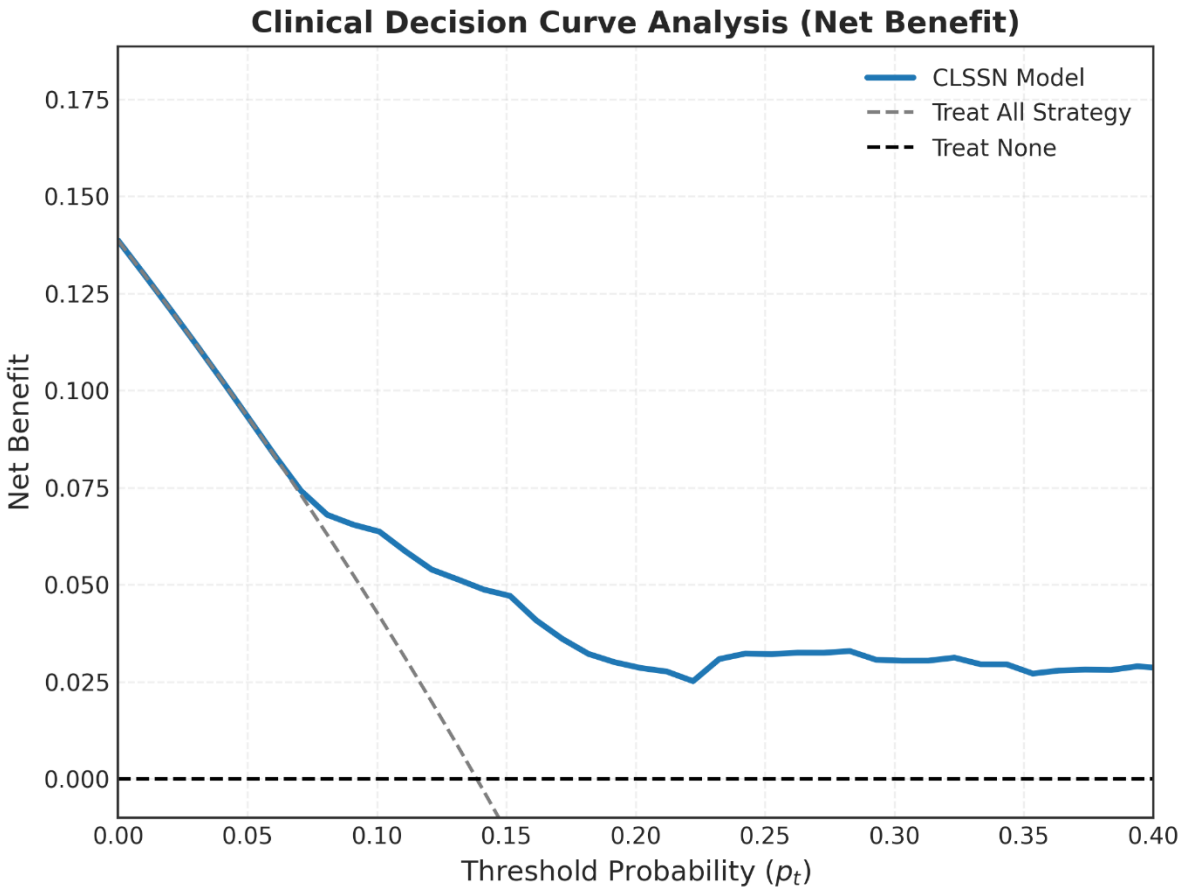


Figure 7: Decision Curve Analysis for Clinical Utility Assessment

These findings suggest that the probabilistic predictions made by CLSSN could potentially help stratify patients at the bedside using thresholds, whilst also delivering an efficient reduction in unnecessary treatment for lower-risk patients. In combination with the discrimination and calibration performance shown in Table 3 and the consistent optimization behaviour shown in Figure 2, the decision curve analysis revealed that the proposed framework has characteristics which are desirable for the support of clinical decisions in intensive care setting.

7. Discussion

7.1 Principal Findings

In this study, a Conditioned Latent State-Space Network (CLSSN) for modeling the mortality of an ICU is developed using longitudinal observation data from a set of physiological signals and information on therapeutic interventions provided by the network. The framework proposed combines temporal representation learning, latent state-space dynamics, probabilistic uncertainty modelling and counterfactual trajectory simulation in a single optimization framework. The experimental results showed that the proposed model had an AUROC of 0.7911 and an AUPRC of 0.5209, also its ECE was 0.0722, its Brier Score was 0.0973, which demonstrated the model had a balanced performance, in terms of discrimination, calibration and probabilistic prediction. Moreover, the ablation experiments showed that nonlinear latent dynamics are the main factor for the predictive performance, and inclusion of intervention aware transitions and uncertainty modeling offer improvements as complementary. The temporal saliency analysis further showed that the saliencies of the learned representations change over the observation window, which strengthens the interpretability of the proposed framework.

7.2 Clinical Implications

Accurate assessment of the risk of death in the early period of admission to the ICUs is still a crucial point to prioritize clinical resources and to help timely therapeutic interventions. Given that the proposed CLSSN models patient trajectories as sequential observations of a trajectory of the patient's underlying physiological state, allows for the inclusion of disease progression over time in the prediction of mortality. Through embedding information about therapeutic interventions in the mechanism of the latent transition, the approach offers an intervention-aware representation of the evolution of the physiology beyond the traditional sequence classification approaches. An advantage of the counterfactual trajectory generation mechanism is that it provides a further layer of clinical interpretability by enabling learning of hypothetical treatment scenarios through learned latent dynamics. These simulations are not strictly speaking Conditioned inferences, but they do offer a computational tool for exploring simulation of patient trajectories under alternative intervention strategies. However, in critically ill respiratory patients, several input variables such as arterial blood gases, lactate and oxygenation-derived variables are routinely measured, thereby making the proposed framework potentially relevant for pulmonary critical care applications that are characterized by the continuous assessment of the uncontrolled deterioration of respiratory and hemodynamic instability.

7.3 Limitations

There are some limitations of which one should be aware. First, the study was carried out in a single centre and using a retrospective database, so the results reported might not necessarily be relevant to other healthcare facilities. Second, the experimental evaluation was carried out with a patient sample of 2,000 subjects in the ICU, which might not be sufficient to ensure accurate estimation of model performance in larger patient sets. Third, the intervention-aware latent dynamics consider just one therapeutic intervention variable: vasopressor administration. Moreover the physiological representation was limited to six clinical variables and thus does not reflect all of the patient physiology often encountered in intensive care. Lastly, external validation with independent datasets was not conducted and thus transportability of the proposed framework to other clinical settings is yet to be explored.

7.4 Future Directions

Future research should explore the possibility of integrated utilization of several concurrent treatments, such as mechanical ventilation, fluid administration, and sedative drugs, in a single integrated state-space framework that takes these treatments into account. The proposed method may be expanded to include cohorts of patients in critical care for respiratory diseases to increase its utility for the management of pulmonary disease and prediction of respiratory failure. Moreover, extensive validation from other institutions and diverse patient cohorts should be required to assess the validity and applicability of the proposed framework. Finally, prospective clinical trials and deployment in actual systems for ICU decision support could provide valuable information on the real-world application of CLSSN in clinical critical care.

8. Conclusion

A Conditional Latent State-Space Network (CLSSN) was proposed for predicting ICU mortality rates with both longitudinal clinical physiological data and information of interventions from the MIMIC-III database. The proposed method is a novel approach that combines the above features in a unified framework, bringing together a comprehensive methodology for modeling patient-specific disease progression over time. The key difference between CLSSN and traditional predictive models is that the intervention is embedded in the latent dynamics, which allows the system to learn and generate predictions that are specific to the set of interventions, while still remaining probabilistically calibrated and temporally interpretable. Experimental assessment showed that the proposed framework improved the AUROC score to 0.7911, the AUPRC score to 0.5209, the Expected Calibration Error (ECE) to 0.0722 and Brier Score to 0.0973, reflecting good balance between discrimination capability, calibration, and probabilistic accuracy. The ablation study also validated the significance of nonlinear latent dynamics in modeling complex physiology evolution, and the use of interventions and uncertainty was also found to result in greater

robustness. The analysis of temporal saliency showed that feature importance of different features varied over time during the observation period, and the counterfactual simulations of the patient's trajectory are interpretable in a way that allowed studying the mechanism of treatment-dependent changes in predicted patient evolution. Moreover, the results of the decision curve analysis indicated that the proposed framework might have a positive net clinical benefit at clinically relevant decision thresholds, which implied its potential utility as a clinical decision support tool. Thus, while these results are promising, the present study is limited by its retrospective, single center cohort design, physiological feature set, intervention variable and the lack of external validation. Future studies are needed to expand the proposed framework to multi-intervention modeling, include more complex multimodal clinical data, assess respiratory-specific critical care cohorts, and externally validate across various healthcare institutions. Future clinical investigation and application in live patient-use within the ICU will be crucial for demonstrating the applicability of intervention-aware latent state-space learning for assisting with individualized critical care decision making.

Funding

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

Data availability

Data used during this study is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate- Not applicable.

Consent for publication- Not applicable.

Competing interests- The authors declare no competing interests.

REFERENCES

1. Johnson AEW, Pollard TJ, Shen L, Li-Wei HL, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Scientific Data*. 2016;3:160035.
2. Johnson AEW, Bulgarelli L, Pollard TJ, Horng S, Celi LA, Mark RG. MIMIC-IV (Version 2.2). PhysioNet; 2023.
3. Shafiabady N, et al. Explainable AI for mortality prediction: A comparative study using the MIMIC-III dataset. *BMJ Health & Care Informatics*. 2026;33(1):e101406. doi:10.1136/bmjhci-2024-101406.
4. Sidhu BK. Explainable machine learning for ICU mortality prediction: Comparing SHAP, LIME and Anchors on MIMIC-IV. *International Journal of Computer Sciences and Engineering*. 2026;14(5).
5. Alsammani A, et al. Calibrated and interpretable machine learning for ICU mortality prediction using first 24-hour clinical data. *medRxiv*. 2026. doi:10.64898/2026.05.30.26354524.
6. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. Proceedings of the 34th International Conference on Machine Learning (ICML). 2017.
7. Mesinovic M, Watkinson P, Zhu T. Explainable machine learning for predicting ICU mortality in myocardial infarction patients using pseudo-dynamic data. *Scientific Reports*. 2025;15:27887.
8. Chou YL, et al. Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Artificial Intelligence Review*. 2021.
9. Nowroozilarki Z, et al. Real-time mortality prediction using MIMIC-IV ICU data via boosted nonparametric hazards. *arXiv*. 2021.
10. Damian DS. Explainable graph-neural architectures for ICU mortality prediction using logic explained networks. *SSRN Electronic Journal*. 2026.
11. Patel K, Beedala P. Calibration drift under cross-institutional deployment: An external validation framework for ICU mortality prediction across MIMIC-IV and eICU. *SSRN Electronic Journal*. 2026.
12. Bakumenko A, et al. Transparent early ICU mortality prediction with clinical transformer and per-case modality attribution. *arXiv*. 2025.
13. Mamatov N, Kellmeyer P. Enhancing mortality prediction in cardiac arrest ICU patients through meta-modeling of structured clinical data from MIMIC-IV. *arXiv*. 2025.
14. Vincent JL, Singer M. Critical care: Advances and future perspectives. *The Lancet*. 2010;376(9749):1354–1361.
15. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Critical Care Medicine*. 1985;13(10):818–829.