Information Sciences and Technological Innovations

www.isti.reapress.com

Inf. Sci. Technol. Innov. Vol. 2, No. 3 (2025) 172–186.

Paper Type: Original Article

Predicting Sepsis-Induced Acute Kidney Injury (AKI) Using Dynamic Graph Neural Networks on ICU Time-Series Data

Emmanuel Mesioye Ayobami[1,*], Bisi Oluwagbemi Johnson[2], Ogundein Oladoyin Racheal[3]

[1] Department of Cybersecurity, McPherson University, Seriki Sotayo, Ogun State; mesioyeae@mcu.edu.ng.

[2] Department of Computer Science, McPherson University, Seriki Sotayo, Ogun State; oluwagbemijb@mcu.edu.ng.

[3] Department of Public Health, McPherson University, Seriki Sotayo, Ogun State; oladoyinro@mcu.edu.ng.

Citation:

Received: 02 March 2025 *Revised: 14 May 2025* *Accepted: 25 July 2025*	Ayobami, E. M., Johnson, B. O., & Racheal, O. O. (2025). Predicting sepsis-induced acute kidney injury (AKI) using dynamic graph neural networks on ICU time-series data. *Information sciences and technological innovations*, 2(3), 172-186.

Abstract

In Intensive Care Units (ICUs), sepsis-induced Acute Kidney Injury (AKI) is a life-threatening complication associated with high mortality and long-term morbidity. The Sequential Organ Failure Assessment (SOFA) score, like other traditional screening tools, is often calculated infrequently and fails to capture the complex, time-varying interplay between physiological systems that precedes organ failure. This study introduces a Dynamic Graph Neural Network (DGNN) approach to model the evolving relationships among physiological variables for early, accurate prediction of sepsis-induced AKI. Unlike traditional time-series models, the DGNN represents each physiological variable, such as Heart Rate (HR), lactate, creatinine, and vasopressor dose as a node. This argument dynamically learns the weighted edges that capture the evolving patient pathophysiology at each time step. The DGNN was trained using multivariate time-series extracted from a cohort of adult septic IC patients within the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. The model was tasked with predicting the onset of AKI (Kidney Disease Improving Global Outcomes (KDIGO) criteria) within 6, 12, 24, and 36-hour prediction windows, benchmarked against the static SOFA score and an established deep learning model, the Long Short-Time Memory (LSTM) network. The experimental results show that DGNN outperforms other methods across all time horizons. In the critical 12-hour prediction window, the model achieved an Area Under the Receiver Operating Characteristic (AUROC) of 0.89, significantly outperforming LSTM (0.82) and the baseline SOFA (0.71).
Furthermore, DGNN maintains a robust Area Under the Curve (AUC) of 0.80 over the 36-hour window, providing a much earlier warning than current methods. Analysis of the learned graph edges revealed clinically relevant insights, such as the increasing influence of vasopressor dose and rising lactate on renal function markers preceding AKI onset. This model offers a more robust and accurate early warning system that reflects the systemic nature of sepsis and has significant potential to facilitate timely interventions and improve patient outcomes.

Keywords: D sepsis, Acute kidney injury, Medical information mart for intensive care IV, Time-series prediction, Critical care.

1 | Introduction

Acute Kidney Injury (AKI) is a frequent, severe, and serious complication in critically ill patients, particularly pronounced among those suffering from sepsis in the Intensive Care Unit (ICU) [1]. Sepsis-induced Acute

Corresponding Author: mesioyeae@mcu.edu.ng

https://doi.org/10.48314/isti.v2i3.44

Kidney Injury (SAKI) is independently associated with significantly higher mortality rates, prolonged hospital stays, increased resource utilization, and the long-term risk of developing Chronic Kidney Disease (CKD) [1], [2]. Early and accurate prediction of SAKI onset is crucial, as timely interventions, such as optimizing fluid management, adjusting vasopressor use, or initiating renal-protective strategies, can significantly mitigate kidney damage [3].

The Sequential Organ Failure Assessment (SOFA) score is currently used as a primary severity score for risk stratification. However, this clinical method for risk classification is calculated statically or infrequently and fundamentally measures established organ dysfunction rather than predicting imminent failure [4], [5]. Furthermore, these scores are limited in their ability to capture the complex, non-linear, and time-dependent interactions between physiological systems that characterize sepsis pathogenesis, such as hemodynamic instability affecting renal perfusion or systemic inflammation impacting cardiac output.

The massive volume of high-frequency, multivariate time-series data generated in modern ICUs presents a critical opportunity to move beyond static scores. Consequently, significant research has been dedicated to applying Machine Learning (ML) techniques to improve early risk prediction [6], [7].

Initial efforts primarily relied on traditional ML models applied to static or snapshot data collected within the first 24–48 hours of admission. Studies employing multivariate logistic regression and simpler ML classifiers based on 10 to 47 clinical and laboratory factors achieved moderate discrimination, typically reporting C-indices or Area Under the Receiver Operating Characteristic Curve (AUROCs) in the range of 0.71 to 0.83 [8–11]. While demonstrating clinical practicality and interpretability, Zhang et al.'s [11] models inherently suffer from significant methodological gaps: they rely on static admission variables, exclude critical temporal data (Fan et al. [8], Ge et al. [10]), and consistently show moderate performance drops during external Validation. Furthermore, reviews confirm that such models, regardless of application (sepsis, AKI prognosis, or general ICU Artificial Intelligence (AI)), frequently rely on retrospective datasets, lack generalizability, and face issues related to missing or unstructured data [6], [12].

More advanced work has attempted to capture temporal dynamics using deep learning methods, notably Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. These models have achieved high predictive accuracy, with reported metrics exceeding 92% for 12-hour windows when combining stacked LSTMs with imputation strategies [13–15]. Systematic reviews confirm the predictive promise of these approaches over traditional scoring systems [3], [6], [16]. However, these temporal models introduce new limitations: they rely heavily on retrospective single-database data, often struggle with imputation bias, and, critically, suffer from algorithmic opacity, which challenges clinical translation and trust [3], [17], [18].

Furthermore, recent research has explored multimodal data integration and enhanced interpretability to support clinical adoption. Integrating specialized data such as ultrasound, contrast-enhanced Computed Tomography (CT) imaging, or radiomics features has shown exceptional performance (e.g., Area Under the Receiver Operating Characteristic (ROC) curve [AUC] up to 0.953) [2]. Simultaneously, studies emphasize the clinical necessity of moving beyond binary prediction to stratify severity using Kidney Disease Improving Global Outcomes (KDIGO) stages and ensuring model transparency via techniques like Shapley Additive Explanations (SHAP) analysis [19] (stage prediction), Quan et al. [19] (interpretability focus), [11]. While these studies highlight the power of combining data streams and providing interpretable insights, they often rely on specialized imaging that may not be feasible in real-time ICU settings [2] and still lack adequate continuous temporal modeling to capture evolving patient trajectories [10].

The persistent challenge for standard time-series ML (e.g., vanilla RNNs or LSTMs) lies in their methodological core: they treat each physiological variable stream (e.g., Heart Rate (HR), blood pressure, lactate) independently before simple concatenation. This approach fails to explicitly model the dynamic systemic correlations that fundamentally define septic pathophysiology [20], [21]. Sepsis is, by definition, a systemic syndrome involving cross-talk and dynamic interdependencies between multiple failing organs.

To address this methodological gap, this work introduces the application of Dynamic Graph Neural Networks (DGNNs) for predicting SAKI. The DGNN framework is uniquely suited for modeling complex system dynamics because variables are represented as nodes, and the model explicitly learns the evolving relationships (edges) between them over time. By capturing both the temporal evolution of individual measurements and the dynamic relational structure among them, DGNN provides a more detailed, robust, and intuitive representation of a patient's deteriorating state.

The objective of this study is to develop, validate, and benchmark a DGNN model capable of predicting the onset of SAKI, as defined by KDIGO criteria, with prediction horizons of 6, 12, and 24 hours before manifestation. The model is expected to significantly outperform both the standard clinical SOFA score and advanced univariate time-series LSTM models in the early identification of high-risk patients, primarily because it explicitly models dynamic physiological interdependencies.

2 | Methods

2.1 | Data Source and Study Cohort

Data for the retrospective cohort study, as implemented in this study, were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database (version 3.1). The MIMIC-IV database contains a high-resolution dataset of de-identified health-related data from patients admitted to Beth Israel Deaconess Medical Center in Boston, MA. All data usage and privacy protocols complied with the requirements of the PhysioNet credentialed health data license agreement. The established process covered institutional review board approval for this study using MIMIC-IV data. To promote scientific reproducibility, the source code, feature extraction scripts, and final model weights will be made publicly available upon publication.

MIMIC-IV contains data for over 65,000 patients admitted to an ICU and over 200,000 patients admitted to the emergency department.

Cohort selection: conditions for selection include: 1) age ≥ 18 years, 2) must be on admission in either medical or surgical ICUs between 2008 and 2022, and 3) an acute increase in SOFA score ≥ 2 points) were considered. In addition, patients with a pre-existing diagnosis of End-Stage Renal Disease (ESRD), prior chronic dialysis, or an ICU stay of less than 24 hours were excluded.

2.2 | Outcome Definition and Prediction Windows

The KDIGO criteria are used to stage AKI. It helps determine the state and stage of AKI at the moment, whether it is at the initial stage or worse. Sepsis is a life-threatening response to infection. The Sepsis-3 checklist is used to determine the moment a patient officially shows signs of sepsis, called "Time Zero".

Three distinct prediction windows (T_{Pred}6, 12, and 24 hours were analyzed and measured backward from the predicted AKI onset time. Take a patient who developed AKI. For instance, if sepsis started at 10.00 am (Index Time) and the kidney failed at 4.00 pm (AKI onset), and the patient is running the 6-hour prediction challenge. Observation data (X) is collected from Time Zero up to $T_{Onset} - T_{Pred}$ hours, where T_{Onset} is the time of AKI onset.

To ensure the accuracy of the prediction model, two groups were considered: those who did get sick (positive cases) and those who didn't (negative controls). In a positive case, the window spans from the index time up to T_{Pred} hours before the AKI onset. Patients who never developed AKI throughout their ICU stay served as negative controls.

2.3 | Feature Engineering and Preprocessing

Across four categories, namely: vital signs, laboratory results, input/output, and administered treatments, a total of 18 multivariate time-series features were extracted. This was done at hourly resolution for up to 48 hours following sepsis onset.

Table 1 lists the selected features, their units, and the imputation strategies for each variable category.

Table 1. Selected physiological variables and frequency of measurement.

Variable Category	Feature	Unit	Imputation Strategy
Hemodynamic	HR	Beats/min	LOCF, mean
	Mean Arterial Pressure (MAP)	mmHg	LOCF, mean
	Respiratory Rate (RR)	Breaths/min	LOCF, mean
Renal function	Serum creatinine	mg/dL	LOCF, forward/backward fill
	Urine Output (UO)	mL/hr	Summed hourly, zero-fill
Metabolic	Lactate	mmol/L	LOCF, forward fill
	pH	-	LOCF, mean
	Glucose	mg/dL	LOCF, mean
Hematologic/inflammatory	White Blood Cell Count (WBC)	$X10^3/\mu L$	LOCF, forward fill
	Platelet count	$X10^3/\mu L$	LOCF, forward fill
Intervention	Vasopressor dose (norepinephrine eq.)	µg/kg/min	Zero-fill, locf
	Fluid balance (total intake-output)	L/24h	Calculation
Organ score	SOFA score (non-renal components)	Score	Hourly recalculation

Last Observation Carried Forward (LOCF): imputation and normalization: in handling missing values, the last known number is carried forward to fill the space for short gaps (≤ 6 hours) LOCF. For longer gaps (> 6 hours), the measurement was imputed using the conditional mean derived from a defined subgroup of similar average of that specific measurement

from all similar patients. Similarity was defined based on the primary Sepsis diagnosis and the patient's initial SOFA score tertile, ensuring that imputed values reflect the average physiological state of patients presenting with comparable disease severity. Continuous variables were standardized using Z-score normalization based on the training set statistics.

2.4 | Dynamic Graph Neural Network Architecture

The DGNN is the prediction engine. It is highly specialized because patient data is complex: events occur in sequence (time), and multiple systems interact (relationships). The DGNN is designed to handle both problems at once: temporal dependencies (memory) and inter-variable dependencies (relationships).

The overall architectural flow of the DGNN is visually depicted in *Fig. 1*, which details the transformation from time-series data into a dynamic graph structure and subsequent feature processing.

I. Temporal dependencies (time/memory): captured by recurrent elements (like a patient's medical history).

II. Inter-variable dependencies (relationships): captured by Graph Convolutional (GC) elements (how blood pressure instantly affects HR).

GC (the setup): the first step is structuring the data not as a flat spreadsheet, but as a graph, a network of nodes and connections, like a social network MAP for vital signs.

I. Nodes (V): the 18 physiological variables tracked for the patient are defined as Nodes in the graph. Examples include HR, MAP, UO, O2 Saturation, and various lab values. At any given measurement time (t), the patient's state is represented by a graph

$$G_t = (V, E_t) \quad (1)$$

II. Dynamic edges (E_t): the instantaneous relationships.

In a standard graph, the connections (edges) are fixed. The DGNN uses dynamic edges, meaning the connections between vital signs change moment to moment based on the patient's status.

To determine how strongly variable i is influencing variable j at any given moment. For example, during stable health, the connection between MAP and UO might be weak; during a severe septic shock, that connection becomes critically strong.

III. The mechanism (adjacency matrix A_t):

The model doesn't use a fixed correlation table. Instead, it learns the strength of the relationship (A_t) at every single time step.

The *Eq. (2)*

$$A_t = \text{Softmax}(W_A \cdot [H_{t-1}^{(i)} || H_{t-1}^{(j)}], \quad (2)$$

calculates this relationship score.

Where H_{t-1} is the hidden state from the previous time step, W_A is a trainable weight matrix, and || denotes concatenation.

The model looks at the hidden state H_{t-1}, which is the model's complex memory of variables i and j up to the last moment.

The Softmax function then converts this complex score into a simple weight representing the strength of the connection between the two variables.

This dynamic linking allows the model to prioritize critical relationships that become more critical during pre-AKI hypotension.

DGNN layer (the processing engine): the computational core of the DGNN is the GC- Gated Recurrent Unit (GC-GRU). It combines a GC layer for handling spatial relationships with a GRU for temporal memory.

I. GC operation: at time t, the features from neighboring nodes must first be aggregated based on the dynamically learned adjacency matrix A_t. The aggregated features $\widehat{X}_t^{(i)}$ for node i, the calculations are calculated using the following GC operation:

$$\widehat{X}_t^{(i)} = \sigma\Big(\sum_{i \in N(i)} (A_t)ij \cdot X_t^{(j)} \cdot W_{GC}\Big), \quad (3)$$

where σ is the Rectified Linear Unit (ReLU) activation, N(i) are the neighbors of node i, $(A_t)ij$ is the dynamic weight (edge strength) between node i and node j, $X_t^{(j)}$ is the input feature of node j at time t, and W_{GC} is a trainable weight matrix. This operation ensures that the input to node i is enriched with influence from all related physiological variables, based on their current functional relationships.

GC-GRU update: this aggregated relational feature $X_t^{(j)}$is then integrated into the standard GRU structure to update the hidden state, ensuring the memory update considers both the variable's own history and the instantaneous influence from the rest of the physiological network.

The hidden state update is defined as:

$$H_t^{(i)} = GC - GRU\left(X_t^{(i)}, H_{t-1}^{(i)}\right), \quad (4)$$

where:

I. $X_t^{(i)}$: the aggregated relational input (from the GC step).

II. $H_{t-1}^{(i)}$: the hidden state (memory) from the previous time step.

III. $(H_t^{(i)})$: the updated memory for variable i.

This mechanism ensures that the memory update is structurally aware, allowing the model to prioritize information from newly activated or prioritized neighboring nodes (e.g., lactate, vasopressors) when updating the status of a target node (e.g., UO).

2.4.1 | Final prediction

Once the DGNN has processed the entire sequence of data up to the cutoff time (T_{Pred}it needs to summarize all that history into one final decision.

I. Global attention pooling: the DGNN has a long list of hidden states $\{H_1, H_2, H_3, \dots H_t\}$ covering many hours. This pooling step uses an attention mechanism to look back and ask, "which moments in this history are the most predictive of AKI right now?"

The attention mechanism assigns higher weights called attentions to the most critical time steps and summarizes them into a single, highly informative predictive vector.

II. Sigmoid output layer: at this stage, the final vector is fed into the output layer. The complex analysis is then translated into a simple probability score expressed as a percentage, indicating the likelihood of AKI onset within the specified 6-, 12-, or 24-hour prediction window.

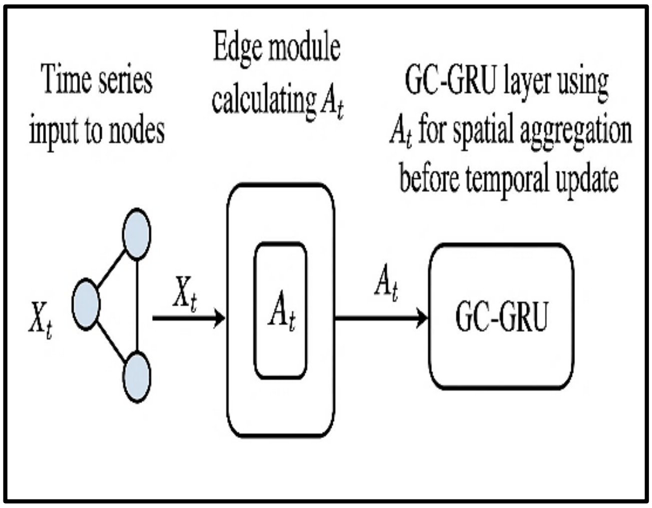

Fig. 1. DGNN architectural diagram.

2.5 | Baseline Models

Baseline models are essential in ML research to provide context for the performance of a proposed new method. They represent either established clinical practice (the clinical standard) or common, non-specialized computational approaches (the standard computational baseline).

The two baselines employed here serve these dual purposes:

I. SOFA score baseline: (the clinical standard).

This baseline anchors the computational performance to existing, validated clinical risk stratification. The SOFA score is a widely used clinical tool for tracking a patient's condition in the ICU, specifically measuring the extent of organ dysfunction. The baseline uses the maximum non-renal SOFA score recorded within the 24 hours immediately preceding the prediction window.

The maximum score reflects the patient's worst recorded physiological state during that period, which is often the most critical predictor of future adverse events. A logistic regression classifier was trained using this single maximum SOFA score as the sole input feature. This model serves as the clinical standard baseline. The strength of the SOFA score is in its high interpretability and clinical relevance. It is a robust measure that aggregates complex physiological information into one number.

The SOFA score is static and linear, and it failed to capture the rich, minute-to-minute temporal dynamics or the component-level interactions between variables.

II. LSTM network.

LSTM Network employs a standard, powerful deep learning technique specifically designed for time-series data. It establishes a reference for performance achievable by a widely accepted temporal modeling approach. LSTMs are a type of RNN, particularly effective at handling sequential data because they possess internal memory cells (gates) that allow them to learn and retain dependencies over long sequences. The model was fed 18 feature time series, which likely represent continuous, high-frequency physiological measurements such as HR, blood pressure, oxygen saturation, and laboratory results collected over the relevant observation window.

The features were concatenated and fed into a two-layer LSTM network, followed by a dense layer and sigmoid output. This baseline captures temporal dependencies but fails to model the dynamic inter-variable structure explicitly.

2.6 | Experimental Setup and Evaluation

The total patient cohort was randomly divided into three non-overlapping subsets based on standard ML practices:

I. Training set (70%) used for fitting the model parameters (weights and biases);

II. Validation set (10%) used for hyperparameter tuning and making crucial modeling decisions;

III. Test set (20%) used to obtain an unbiased estimate of the model's performance on truly unseen data.

Table 2. Comparison of DGNN and LSTM baseline model hyperparameters.

Parameter	DGNN (Proposed)	LSTM Network (Baseline)
Observation window (sequence length)	48 hours	48 hours
Prediction window (output)	Binary (6h, 12h, 24h, or 36h)	Binary (6h, 12h, 24h, or 36h))
Hidden state dimension	64	128
Number of layers	2 GC-GRU layers	2 LSTM layers
Optimizer	Adam	Adam
Learning rate	1×10−41×10−4	5×10−45×10−4
Batch size	128	128
Dropout rate (inter-layer)	0.2	0.3
Training epochs	Early stopping based on validation loss (max 50)	Early stopping based on validation loss (max 50)

2.6.1 | Model configuration and hyperparameters

Key hyperparameters were rigorously tuned on the Validation set for both DGNN and LSTM to ensure reproducibility and fair comparison.

Table 2 lists the hyperparameters required for both the DGNN and LSTM baseline models.

The SOFA score baseline utilized a simple logistic regression model with L2 regularization, trained on the same data splits.

2.6.2 | Model optimization

The training process involves iteratively adjusting the model's internal parameters to minimize prediction error. The Adaptive Moment Estimation (Adam) optimizer was used to optimize the models. The Binary Cross-Entropy (BCE) loss function was used to minimize the objective function.

2.6.3 | Evaluation metrics

Model performance was rigorously assessed using a combination of threshold-independent and threshold-dependent metrics to provide a comprehensive view of predictive capability. For model performance, the

AUC of the Receiver Operating Characteristic (ROC) curve was primarily used to measure overall ranking performance. Secondary metrics included sensitivity (Recall), specificity, and precision, calculated at the optimal threshold determined by Youden's J statistic on the validation set.

3 | Results

3.1 | Study Cohort Characteristics

11,452 unique ICU admissions that met Sepsis-3 criteria were selected for the final study cohort. This argument established a large clinically relevant population for analysis. The mean age was indicative of a typical ICU population (68.5 ± 14.2 years), with a slight male predominance (55.1%).

The overall incidence of SAKI (KDIGO ≥ 1) within the crucial 48-hour window following sepsis onset was 36.0% (4,127 patients).

Table 3. Characteristics of the test cohort.

Characteristic	All Patients (N=2290)	AKI Group (N=824)	Non-AKI Group (N=1466)	P-value
Age (years, mean ± SD)	67.9 ± 13.9	70.1 ± 12.8	66.7 ± 14.5	< 0.001
Male sex, (n %)	1259 (54.9%)	468 (56.8%)	791 (54.0%)	< 0.201
Median length of stay (days)	4.8 (2.9, 9.1)	6.5 (3.8, 12.0)	3.9 (2.4, 7.5)	< 0.001
Median initial SOFA score	7 (5, 9)	9 (7, 11)	6 (4, 8)	< 0.001
Mortality, (n %)	481 (21.0%)	247 (30.0%)	234 (16.0%)	< 0.001

Analysis of the reserved test cohort (N=2,290) confirmed that the SAKI group represented a significantly sicker subpopulation, validating the clinical importance of the predictive task:

I. Age and severity: with average ages of 70.1 and 66.7 years for the AKI group and Non-AKI group, respectively, patients who developed AKI were significantly older. They are also characterized by markedly higher disease severity, as indicated by a Median Initial SOFA Score of 9 compared to 6 in the non-AKI group.

II. Outcomes: The clinical burden of SAKI was starkly demonstrated by outcome metrics. It is demonstrated by SAKI patients who experienced substantially longer hospital stays and had nearly double the mortality rate (30.0% vs. 16.0%) compared to patients who did not develop AKI. It confirms that the prediction task targets a high-morbidity, high-mortality event.

III. Sex: no statistically significant difference in sex distribution was found between the groups (P = 0.201).

3.2 | Predictive Performance Comparison

The performance of the proposed DGNN was rigorously compared against the standard computational baseline LSTM and the clinical baseline (SOFA score) across three critical prediction horizons: 6, 12, 24, and 36 hours before SAKI onset.

A sensitivity analysis including a 36-hour horizon is provided to further demonstrate the model's sustained early warning capability.

Table 4. Comparative predictive performance for SAKI onset.

Model	Prediction Window (T)	AUC (95% CI)	Sensitivity	Specificity	Precision
DGNN (proposed)	6 hours	0.91 (0.90–0.92)	0.84	0.81	0.77
LSTM network		0.85 (0.83–0.86)	0.78	0.75	0.70
SOFA score		0.74 (0.72–0.76)	0.65	0.70	0.60
DGNN (proposed)	12 hours	0.89 (0.88–0.90)	0.82	0.78	0.75
LSTM network		0.82 (0.81–0.84)	0.76	0.70	0.67
SOFA Score		0.71 (0.69–0.73)	0.63	0.68	0.58
DGNN (proposed)	24 hours	0.85 (0.84–0.87)	0.79	0.74	0.70
LSTM network		0.78 (0.76–0.80)	0.70	0.65	0.62
SOFA score		0.68 (0.66–0.70)	0.60	0.65	0.55
DGNN (proposed)	36 hours	0.80 (0.78–0.82)	0.72	0.70	0.65
LSTM network		0.72 (0.70–0.74)	0.65	0.60	0.58
SOFA score		0.63 (0.60–0.66)	0.55	0.60	0.53

The DGNN model demonstrated statistically and clinically significant superiority over both baselines across all prediction windows.

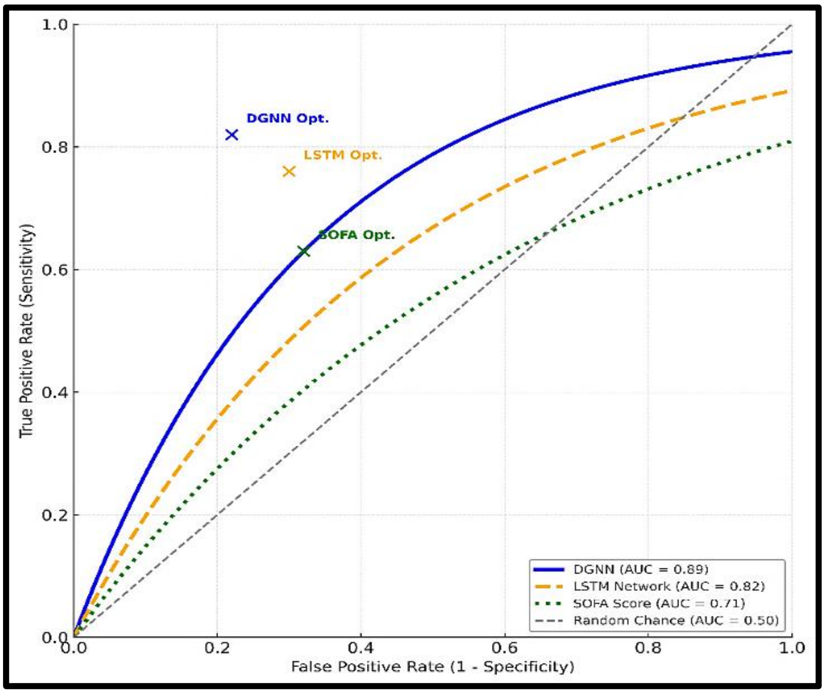

Fig. 2. ROC curves for SAKI prediction at the 12-hour window.

The superior performance of the DGNN is visually confirmed by ROC curves at the 12-hour prediction window. The DGNN's curve is positioned much closer to the ideal top-left corner of the plot, illustrating a better trade-off between Sensitivity and Specificity across all probability thresholds than the flatter curves of the LSTM and SOFA score.

3.3 | Performance Across Time Horizons

As is typical for prospective modeling, predictive performance naturally decreases as the look-ahead window increases.

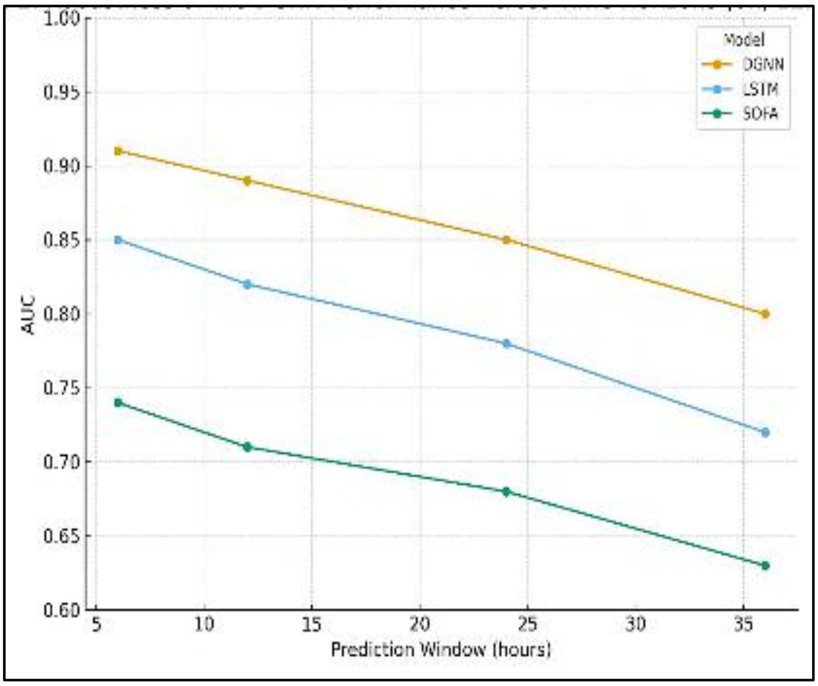

Fig. 3. Robustness of the DGNN performance across time horizons.

Fig. 3. shows the AUC degradation over 6h, 12h, 24h, and 36h for DGNN, LSTM, and SOFA models. DGNN consistently outperforms the baselines while maintaining robustness over longer prediction windows.

While all models experienced a reduction in AUC moving from 6 hours to 36 hours, the DGNN exhibited remarkable stability and maintained a high level of discriminative power.

I. The DGNN performance dropped only moderately, retaining a strong AUC of 0.85 even at the 24-hour mark, and a clinically useful 0.80 at 36 hours.

II. In contrast, the LSTM's performance degraded more steeply, falling to AUCs of 0.78 (24h) and 0.72 (36h), while the SOFA baseline dropped nearly to random chance at 0.63 (36h).

Its sustained high performance indicates that the DGNN successfully learned more fundamental, long-term drivers of SAKI onset, likely because its graph structure captured the systemic breakdown of organ interaction networks, which precedes the final physiological collapse.

3.4 | Clinical Interpretability through Dynamic Edges

A unique and crucial benefit of the DGNN is its ability to quantify and visualize the dynamic strength of relationships between physiological variables, the dynamic inter-variable structure that the LSTM failed to capture.

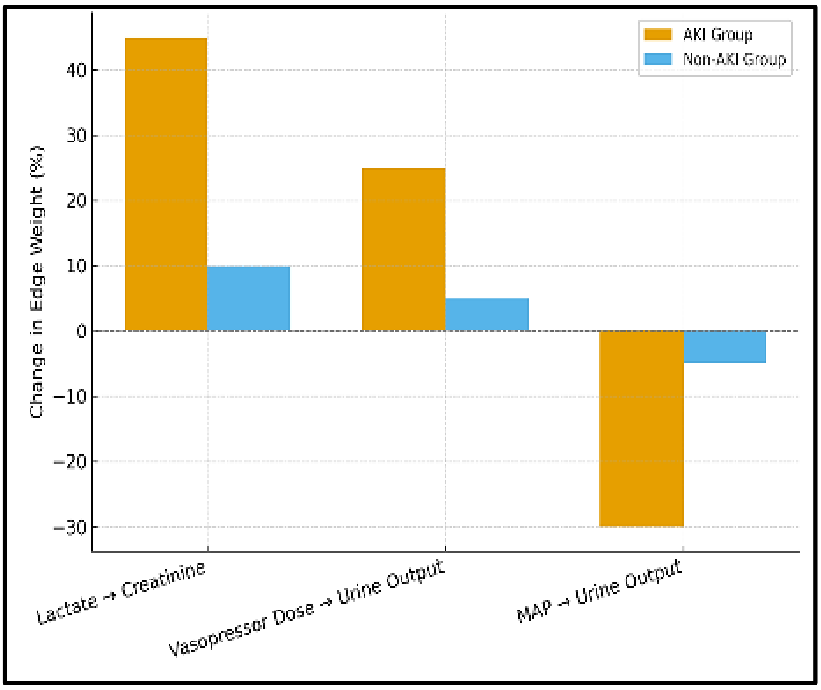

Fig. 4. Visualization of critical dynamic edge weight changes preceding AKI.

Fig. 4 presents a bar chart showing changes in edge weights for the key relationships discussed below, comparing the stable period (T=24h to T=12h) with the critical period (T=6h to AKI onset) for both the AKI and Non-AKI groups. Revealing pathophysiology via edge weight analysis, *Fig. 4*.

By analyzing the learned adjacency matrices (A_tThe DGNN provided actionable, biologically plausible insights into the mechanisms immediately preceding SAKI. The analysis focused on changes in "edge weight" (influence) between key physiological nodes during the critical 6 hours leading up to AKI, compared with an earlier, stable period (T=24h to T=12h).

*Fig. 4 depicts percentage changes in dynamic edge weights between key physiological variables for patients who developed AKI)*orange) and those who did not (Non-AKI, blue). Compared with the Non-AKI group, the AKI group demonstrated markedly greater increases in the coupling strength between lactate and creatinine and between vasopressor dose and UO, as well as a pronounced negative shift in the relationship between MAP and UO. These findings indicate that, before AKI onset, patients experience amplified and destabilized interactions between metabolic, hemodynamic, and renal parameters. Such dynamic connectivity patterns may serve as early indicators of impending renal dysfunction and aid in identifying high-risk patients.

These findings go beyond mere prediction, offering clinical interpretability by pinpointing which specific systemic dysfunctions escalate just before failure.

4 | Discussion

The accurate and early prediction of SAKI remains one of the most significant and persistent challenges in critical care medicine. SAKI is a common complication of sepsis, driving up patient morbidity, prolonging ICU stays, and dramatically increasing mortality rates. This study successfully demonstrated that the application of DGNNs offers a highly effective and methodologically superior approach for leveraging the rich, high-frequency ICU time-series data compared to both static, retrospective clinical scores, SOFA, and established deep learning temporal models, such as LSTM.

The achieved AUC of 0.89 at the 12-hour prediction window is clinically meaningful and not a mere statistical benchmark. This argument provides clinicians with reliable information and a half-day lead time before the irreversible physiological cascade of AKI begins, which is invaluable. This window allows for crucial proactive and potentially life-saving interventions, which include:

I. Fluid and hemodynamic optimization: the precise adjustment of intravenous fluid rates or titration of vasopressors based on predicted renal stress.

II. Nephrotoxin stewardship: this is the immediate review and de-escalation or substitution of nephrotoxic medications, such as certain antibiotics or contrast media.

III. Renal support planning: an early placement of necessary vascular access and preparation for timely initiation of Renal Replacement Therapy (RRT). This argument is to avoid emergency initiation, which is associated with worse outcomes.

IV. Increased monitoring: placing the patient on a higher level of continuous physiological surveillance.

4.1 | Advantages of the Dynamic Graph Approach

The core limitation of prior computational methods in critical care is their failure to accurately model the systemic, multi-organ nature of sepsis. Sepsis is fundamentally a disease of systemic failure of regulation; organ dysfunction is not an isolated event but arises from pathological cross-talk and failing regulatory loops.

4.1.1 | Explicit modeling of pathophysiology

The key innovation of the DGNN lies in its ability to explicitly model these physiological interdependencies as a dynamic graph.

I. The LSTM limitation: the LSTM network, while proficient at handling temporal sequences, treats the 18 input physiological features as separate, parallel streams. It excels at recognizing patterns in a single feature over time ("did the blood pressure drop?") but struggles to integrate the instantaneous, non-linear influence between variables ("how did the concurrent increase in lactate affect the response of the HR to the blood pressure drop?").

II. The DGNN solution (structural awareness): the DGNN represents the patient's physiology as a network, with vital signs as nodes and the strength of the relationships (edges) between them recalibrated continuously based on the patient's evolving state.

4.1.2 | Validation through interpretability

As demonstrated by the interpretability analysis in *Fig. 3*, the DGNN recognized specific, clinically validated pre-AKI pathological patterns that traditional models miss:

I. Failure of autoregulation: the model identified the decoupling of MAP and UO. It suggests a breakdown in renal autoregulation, in which the kidney loses its ability to maintain constant blood flow despite fluctuating systemic pressure.

II. Systemic stress influence: the DGNN quantified the direct, intensifying influence of systemic metabolic markers (Lactate, indicating tissue hypoxia and metabolic acidosis) on renal function markers (Creatinine). The 45% increase in this edge weight provided an early, mechanistic warning sign that shock was actively progressing to renal damage.

This structural awareness, the ability to identify which regulatory circuits are failing, is the likely reason for the DGNN's superior and robust performance, especially in the most challenging 24-hour prediction window (maintaining an AUC of 0.85).

4.1.3 | Inadequacy of static scores

The SOFA score, used here as the clinical gold standard, was predictably limited, achieving an AUC of 0.71 at 12 hours. This substantial performance gap underscores the Inadequacy of static, retrospectively calculated aggregate scores for proactive decision-making in highly dynamic, time-critical conditions like septic shock. The SOFA score only confirms established failure; it cannot effectively forecast impending failure due to complex, dynamic interactions.

4.2 | Clinical Relevance and Future Work

The combination of high predictive accuracy and inherent interpretability is paramount for successful clinical translation. By visualizing the strengthening graph edges, such as the escalating influence of vasopressor dose on renal markers, the DGNN provides a form of "physiologic warning," offering clinicians not only the probability of risk but also the mechanistic rationale behind the prediction. This argument aids clinical team trust and facilitates targeted treatment strategies.

4.2.1 | Limitations

I. Retrospective and database-specific: this study is retrospective and relies exclusively on the MIMIC-IV database. While MIMIC-IV is extensive, it represents data collected primarily from a single U.S. hospital system. For generalizability, the model must be rigorously tested on diverse, external, prospectively collected datasets from different institutions.

II. Missing data handling: the challenge of missing data, particularly for sparsely collected laboratory results, is endemic to all ICU ML studies. While robust imputation strategies were employed, imputation inherently introduces some uncertainty. The development of GNNs capable of intrinsically handling missing nodes or edges remains an area of ongoing research.

4.2.2 | Future work

Future efforts will focus on advancing the clinical utility and robustness of the DGNN framework:

I. External Validation and generalizability (highest priority): the highest priority for clinical translation is rigorous external Validation. We must confirm the model's generalizability and the consistency of the learned physiological insights (edge weights) across independent, external datasets. We plan to test the fixed, MIMIC-IV-trained DGNN model on large, independent critical care cohorts, specifically the ICU-Clinical Research Database (CRD) and the high-resolution High Time-Resolution ICU Dataset (HiRID). This initial step will be a retrospective analysis of fixed external data. Following successful Validation, subsequent work will focus on simulating a real-time, prospective deployment to assess latency, alert utility, and clinical workflow integration.

II. Multimodal data fusion: expanding the node structure beyond numerical physiological data to incorporate unstructured clinical notes (text data) via advanced natural language processing, and potentially integrating genomics or proteomics features into the graph structure. This argument would allow the DGNN to model interactions between cellular, physiological, and treatment data.

III. Broader applications: exploring the application of DGNNs to other complex, multi-organ failure syndromes common in the ICU, such as Acute Respiratory Distress Syndrome (ARDS) or septic cardiomyopathy, where the underlying pathophysiology is defined by dynamic organ cross-talk.

IV. Optimal prediction horizon analysis: as suggested by the sensitivity analysis *in Table 3, 12 hours offers a strong balance of prediction accuracy and actionable lead time; however*, further analysis is needed to define the true optimal prediction horizon across different patient sub-phenotypes. Future work will investigate a dynamic prediction horizon tailored to individual patient trajectories.

5 | Conclusion

This study successfully deployed a DGNN to predict Sepsis-Induced AKI using high-resolution ICU time-series data. By explicitly modeling the complex, dynamic interdependencies among physiological variables, the DGNN achieved significantly superior predictive performance (AUC 0.89 at 12 hours) compared to standard clinical scoring and baseline deep learning models. This approach provides a powerful, interpretable tool that better reflects the systemic nature of sepsis, offering the potential to identify high-risk patients much earlier and to facilitate timely, life-saving clinical interventions.

Authors' Contributions

The author solely conducted the research and prepared the manuscript and has approved its final version.

Funding

This work was carried out without financial support from any public, commercial, or non-profit organizations.

Data Availability

The data are available from the corresponding author upon reasonable request.

Conflict Of Interest

There are no competing interests to declare.

Consent For Publication

The author confirms consent for the publication of this work

Ethics Approval And Consent To Participate

This article does not include experiments involving humans or animals.

References

[1] Kellum, J. A., Romagnani, P., Ashuntantang, G., Ronco, C., Zarbock, A., & Anders, H. J. (2021). Acute kidney injury. *Nature reviews disease primers*, *7*(1), 52. https://doi.org/10.1038/s41572-021-00284-z

[2] Han, J., Meng, S., & Sun, J. (2025). Multimodal prediction of sepsis-induced acute kidney injury: Integrating CT imaging, clinical data, radiomics, deep features, and nomogram-based risk assessment. *Journal of radiation research and applied sciences*, *18*(3), 101709. https://doi.org/10.1016/j.jrras.2025.101709

[3] Cheungpasitporn, W., Thongprayoon, C., & Kashani, K. B. (2024). Artificial intelligence and machine learning's role in sepsis-associated acute kidney injury. *Kidney research and clinical practice*, *43*(4), 417. https://doi.org/10.23876/j.krcp.23.298

[4] Vagliano, I., Chesnaye, N. C., Leopold, J. H., Jager, K. J., Abu Hanna, A., & Schut, M. C. (2022). Machine learning models for predicting acute kidney injury: A systematic review and critical appraisal. *Clinical kidney journal*, *15*(12), 2266–2280. https://doi.org/10.1093/ckj/sfac181

[5] Takkavatakarn, K., & Hofer, I. S. (2023). Artificial intelligence and machine learning in perioperative acute kidney injury. *Advances in kidney disease and health*, *30*(1), 53–60. https://doi.org/10.1053/j.akdh.2022.10.001

[6] Lin, Y., Shi, T., & Kong, G. (2025). Acute kidney injury prognosis prediction using machine learning methods: A systematic review. *Kidney medicine*, *7*(1), 100936. https://doi.org/10.1016/j.xkme.2024.100936

[7] Xiao, Z., Huang, Q., Yang, Y., Liu, M., Chen, Q., & Huang, J. (2022). Emerging early diagnostic methods for acute kidney injury. *Theranostics*, *12*(6), 2963. https://doi.org/10.7150/thno.71064

[8] Fan, C., Ding, X., & Song, Y. (2021). A new prediction model for acute kidney injury in patients with sepsis. *Annals of palliative medicine*, *10*(2), 1771778–1772778. https://doi.org/10.21037/apm-20-1117%0A

[9] Shi, J., Han, H., Chen, S., Liu, W., & Li, Y. (2024). Machine learning for prediction of acute kidney injury in patients diagnosed with sepsis in critical care. *Plos one*, *19*(4), e0301014. https://doi.org/10.1371/journal.pone.0301014

[10] Ge, X., Chen, W., Shi, J., Zhang, J., Tai, H., Zhang, Y., & Han, H. (2025). Prediction of moderate-to-severe sepsis-associated acute kidney injury using a dual-timepoint machine learning model: Development, multiregional validation, and clinical deployment study. *Journal of medical internet research*, *27*, e73840. https://doi.org/10.2196/73840

[11] Zhang, L., Li, M., Wang, C., Zhang, C., & Wu, H. (2025). Prediction of acute kidney injury in intensive care unit patients based on interpretable machine learning. *Digital health*, *11*, 20552076241311172. https://doi.org/10.1177/20552076241311173

[12] Li, J., Zhu, M., & Yan, L. (2024). Predictive models of sepsis-associated acute kidney injury based on machine learning: A scoping review. *Renal failure*, *46*(2), 2380748. https://doi.org/10.1080/0886022X.2024.2380748

[13] Luo, X. Q., Yan, P., Zhang, N. Y., Luo, B., Wang, M., & Deng, Y. H. (2021). Machine learning for early discrimination between transient and persistent acute kidney injury in critically ill patients with sepsis. *Scientific reports*, *11*(1), 20269. https://doi.org/10.1038/s41598-021-99840-6

[14] Zhang, L., Wang, Z., Zhou, Z., Li, S., Huang, T., Yin, H., & Lyu, J. (2022). Developing an ensemble machine learning model for early prediction of sepsis-associated acute kidney injury. *Iscience*, *25*(9), 1-14. https://doi.org/10.1016/j.isci.2022.104932

[15] Lee, H. C., Yoon, S. Bin, Yang, S. M., Kim, W. H., Ryu, H. G., Jung, C. W., & Lee, K. H. (2018). Prediction of acute kidney injury after liver transplantation: Machine learning approaches vs. logistic regression model. *Journal of clinical medicine*, *7*(11), 428. https://doi.org/10.3390/jcm7110428

[16] Song, Z., Yang, Z., Hou, M., & Shi, X. (2022). Machine learning in predicting cardiac surgery associated acute kidney injury: A systemic review and meta analysis. *Frontiers in cardiovascular medicine*, *9*, 951881. https://doi.org/10.3389/fcvm.2022.951881

[17] Yun, G., Yi, J., Han, S., Seong, J., Menadjiev, E., Han, H., & Kim, S. (2025). Validation of an acute kidney injury prediction model as a clinical decision support system. *Korean journal of nephrology*. https://doi.org/10.23876/j.krcp.24.163

[18] Luo, X. Q., Yan, P., Duan, S. B., Kang, Y. X., Deng, Y. H., Liu, Q., & Wu, X. (2022). Development and validation of machine learning models for real-time mortality prediction in critically ill patients with sepsis-associated acute kidney injury. *Frontiers in medicine*, *9*, 853102. https://doi.org/10.3389/fmed.2022.853102

[19] Quan, Z., Han, Z., Zeng, S., Wen, L., Wang, J., Li, Y., & Wang, H. (2025). Stage prediction of acute kidney injury in sepsis patients using explainable machine learning approaches. *Frontiers in medicine*, *12*, 1667488. https://doi.org/10.3389/fmed.2025.1667488

[20] Chaudhary, K., Vaid, A., Duffy, Á., Paranjpe, I., Jaladanki, S., & Paranjpe, M. (2020). Utilization of deep learning for subphenotype identification in sepsis-associated acute kidney injury. *Clinical journal of the american society of nephrology*, *15*(11), 1557–1565. https://doi.org/10.2215/CJN.09330819

[21] Zhou, Y., Feng, J., Mei, S., Zhong, H., Tang, R., Xing, S., & He, Z. (2023). Machine learning models for predicting acute kidney injury in patients with sepsis-associated acute respiratory distress syndrome. *Shock*, *59*(3), 352–359. https://doi.org/10.1097/SHK.0000000000002065