

Methodological development study: Dynamic mask attention graph neural network for mechanical ventilation in elderly intensive care unit patients

Yi Xie^{1,2}, Ni Xie³ and Jiao Guo¹ 

Abstract

Objective: With the intensifying global population aging, the demand for mechanical ventilation in geriatric patients is rising. Given their complex physiological traits and sparse intensive care unit (ICU) data, accurate intubation prediction is difficult. Premature intubation may raise the risk of hypoxic organ damage, whereas delayed intubation can lead to increased ventilator-associated mortality. Therefore, developing precise intubation prediction models is vital for elderly ICU patients.

Methods: This study retrospectively analyzed data from ICU patients aged over 65 years in the MIMIC-IV and eICU databases. The intubation prediction task was formulated using a sliding window with a strict temporal data split to avoid data leakage. We propose a dynamic mask attention graph neural network (DymaGNN) to capture the time-varying relationship of key physiological variables by constructing a dynamic heterogeneous graph structure and an adaptive edge-weighting mechanism. The mask attention layer is designed to identify the key timesteps in the irregular sampling data.

Results: The experiments showed that DymaGNN achieved an area under the curve (AUC) value of 0.8363 and 0.8557 on the intubation prediction task on MIMIC-IV and eICU datasets, respectively, and maintained an AUC of 0.8115 under a 15% data missing rate. Visualization of the feature interaction graph revealed the relationship between important features such as respiratory rate and oxygen saturation. These interaction patterns matched much clinical knowledge, significantly improving doctors' trust in the model prediction.

Conclusion: Our proposed DymaGNN establishes a useful method for mechanical ventilation prediction in elderly ICU patients, achieving high predictive accuracy and remaining robust under a 10% data missing rate. Its interpretable feature interaction graphs provide transparent insights, aligning with established medical knowledge to build trustworthy tools for real-world ICU intubation decisions.

Keywords

Intensive care units, mechanical ventilation, graph neural network, machine learning algorithms, interpretability

Received: 18 March 2025; accepted: 9 July 2025

Introduction

The growing number of geriatric patients in the world is causing a rise in intensive care unit (ICU) admissions.¹ Their pathophysiological characteristics, including multiple comorbidities,^{2,3} such as reduced hypoxemia tolerance and respiratory muscle fatigue, often enable earlier intubation to alleviate respiratory burden. However, mechanical ventilation in geriatric patients carries significant risks, with mortality rates 1.8 times higher than in younger patients and 3.07 times greater compared to nonventilated elderly patients.⁴ This creates a critical clinical dilemma: how to handle low

¹Department of Anesthesiology, Shaanxi Provincial People's Hospital, Xi'an, China

²Faculty of Business, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

³School of Medicine, Shanghai Jiao tong University, Shanghai, China

Corresponding author:

Jiao Guo, Department of Anesthesiology, Shaanxi Provincial People's Hospital, No.256, Youyi West Road, Beilin District, Xi'an, Shaanxi 710000, China.

Email: guojane1989@gmail.com



oxygen risks without increasing the chances of death from ventilation.⁵

Current studies about intubation prediction in elderly ICU patients mainly use traditional machine learning and deep learning models.⁶ Traditional machine learning methods (such as support vector machine (SVM), and XGBoost) usually rely on statistical feature construction (such as the 1-h average respiratory rate (RR))^{7,8} and heuristic feature selection.⁹ While computationally efficient and interpretable, such methods rely on domain-specific prior knowledge in feature engineering, assuming its broad applicability. This overlooks the significant variability in key physiological features across elderly ICU patients. Consequently, features derived from potentially incomplete prior knowledge may fail to fully capture the complex physiology of this population.

Deep learning models, particularly sequence models like long short-term memory (LSTM)^{10,11} and transformer-based architectures,¹² have superior performance in some ICU-related predictive tasks, including intubation and mortality prediction. For instance, Nora et al. developed an LSTM-based model capable of predicting intubation requirement, mortality risk, and ventilation duration with an area under the curve (AUC) exceeding 0.95, demonstrating the correlation between intubation necessity and acute respiratory distress syndrome mortality.¹³ However, these sequence models intrinsically presuppose uniformly sampled time-series data.¹⁴ Elderly ICU records exhibit marked sampling frequency variability, resulting in inherently irregular and sparse patterns.

Modeling how different features connect has become key for better predicting intubation in ICUs.¹⁵ The cross-attention mechanism, leveraging multi-head attention to model inter-variable dependencies, enhanced AUC by 0.0379 and reduced the false alarm rate to 17.8% in intubation prediction.¹⁶ Li et al. employed a temporal convolutional network to model temporal dependencies and identify key risk factors such as mean blood pressure and oxygen saturation (SpO₂).¹⁷ A representative study by Kim et al. employed pulmonary expert-curated feature relationships to construct static graph structures for predicting spontaneous breathing trial success. Although achieving 0.85 AUC, their FT-GAT model's fixed topological configuration fails to capture the temporal evolution of feature interactions during acute respiratory failure episodes.¹⁸ These approaches focus either on temporal correlations or feature-level dependencies, but they struggle to effectively model the time-evolving interactions between features simultaneously. This limitation is particularly acute in the context of ICU intubation prediction, where the influence of feature relationships varies significantly across different clinical states and time points.

Graph neural networks (GNNs) have recently demonstrated remarkable performance in heterogeneous data analysis, particularly for multivariate time-series forecasting. Some studies have utilized GNNs to extract directed

dependencies among variables, enhancing predictive accuracy.^{19–21} Furthermore, integrating neural ordinary differential equations strengthened GNNs' ability to model temporal dynamics.²² Zhang et al. employed GNNs to model inter-variable associations and predict misaligned data points based on adjacent time steps.²³ Current GNNs are usually applied to observation windows spanning days to weeks to learn stable patterns or slowly evolving trends. This is inconsistent with the needs of ICU intubation prediction, which relies on short and sparse observation windows. As a result, existing GNNs often suffer from over-smoothing and performance degradation.²⁴

In summary, while progress has been made in developing ICU intubation prediction models, solutions tailored for elderly patients remain underdeveloped. Existing approaches often fail to adequately capture the dynamic and interrelated nature of clinical variables, instead treating vital signs and lab values as isolated rather than interdependent. This shortcoming is further complicated by data sparsity due to irregular sampling, which can degrade model performance. Moreover, while current explainability methods can identify key predictive features, they fall short in elucidating the complex interactions between clinical variables.

To address these limitations, we propose the dynamic mask attention graph neural network (DymaGNN) to enhance intubation prediction in geriatric ICU patients. The main contributions of this study are as follows:

1. We construct a time-evolving graph structure where dynamic physiological variables are represented as nodes. Adaptive edge weights capture cooperative and antagonistic effects between variables, allowing a more comprehensive representation of complex interactions.
2. We employ masked attention temporal aggregation to identify critical time windows, improving the utilization of high-quality data segments while mitigating the impact of missing data.
3. DymaGNN shows good performance in multiple mechanical ventilation prediction tasks. The generated feature interaction graph is consistent with clinical knowledge, which enhances doctors' trust in the model's judgments.

This study provides a new method with both high prediction accuracy and interpretability²⁵ for the intubation prediction of elderly ICU patients.

Method

We first establish the ventilation prediction task through expert-guided ventilation status reclassification and sliding window segmentation. Subsequently, we conduct rigorous data splits to avoid temporal data leakage. Second, we detail the dynamic graph construction that encodes variable relationships through adaptive edge weights and mask attention

Table 1. Reclassification of ventilation status in the MIMIC-IV dataset.

Ventilation status	Type	Category
Tracheostomy	4	intubation
InvasiveVent	4	intubation
NonInvasiveVent	3	assisted ventilation
SupplementalOxygen	1	assisted ventilation
HFNC	2	assisted ventilation
None	0	None

HFNC: high-flow nasal cannula.

mechanisms to handle irregular sampling. The integrated framework enables simultaneous learning of cross-feature dependencies and temporal dynamics for intubation prediction.

Data source

This study is based on the MIMIC-IV 2.2 database and focuses on elderly ICU patients aged 65²⁶ years and older to analyze their ventilation requirements and develop predictive models. In the MIMIC dataset, the `mimiciv_derived.ventilation` table classifies ventilation status into five categories based on the type of oxygen therapy and ventilatory support: (0) no oxygen therapy, (1) supplemental oxygen, (2) high-flow nasal cannula (HFNC), (3) non-invasive ventilation (NIV), and (4) invasive mechanical ventilation (IMV), which includes both endotracheal intubation and tracheostomy.

To facilitate analysis, we reclassified the ventilation status into two groups: intubation and assisted ventilation. Specifically, category 4 (IMV) is designated as intubation, encompassing both endotracheal intubation and tracheostomy. Meanwhile, categories 1 (supplemental oxygen), 2 (HFNC), and 3 (NIV) are grouped under assisted ventilation.

For the intubation prediction task, patient outcomes are categorized into two groups: those who underwent intubation (including endotracheal intubation and tracheostomy) and those who did not receive intubation (including patients with no oxygen therapy, supplemental oxygen, HFNC, or NIV). The mapping of ventilation status to categories is summarized in Table 1.

The variables utilized in our study, extracted from the MIMIC-IV database, are summarized in Table 2. Static variables comprise demographic information like gender and admission age. We incorporate diagnostic information with specific constraints: Given the absence of temporal

annotations for diagnoses in MIMIC-IV, our analysis was rigorously limited to chronic diseases and pre-existing conditions to prevent potential data leakage. This focus is particularly clinically appropriate for geriatric ICU patients, as healthcare services in this population pay attention to chronic diseases.²⁷ Additionally, only procedures performed before the prediction window were included as inputs.

Dynamic variables encompass time-series measurements of vital signs and laboratory test results. For vital sign features, we retained all seven high-frequency recorded in the MIMIC-IV database due to their general availability and foundational role in intensive care monitoring. Laboratory variables present wide variation in measurement frequency. To select the most representative and clinically relevant laboratory variables while addressing challenges related to high missingness rates, we implemented the following strategy: We calculated the measurement frequency (i.e. the count of times each variable was performed) for the entire study cohort. These variables were then ranked from the highest to the lowest measurement frequency. The top 25 most frequently measured items were selected as the final input features. This strategy is grounded in clinical practice as high measurement frequency indicates the high level of clinical attention and prioritization in ICU.

Data processing

To formulate the prediction task for intubation requirements in ICU patients, we employ a sliding window mechanism to segment the time-series data. Each sample consists of three consecutive windows: an observation window (6 h), a gap window (2 h), and a prediction window (1 h). The physiological data within the observation window serve as input to predict whether the patient will require mechanical ventilation in the subsequent 1-h period following the 3-h gap, denoted as y . Specifically, $y=1$ indicates that the patient undergoes intubation during the prediction window, while $y=0$ signifies no need for mechanical ventilation. In the MIMIC-IV dataset, we constructed 18,835 samples with a positive-negative ratio of approximately 0.721. The eICU dataset comprised 42,842 samples, with a positive-to-negative ratio of approximately 0.947.

To prevent data leakage, which may arise from direct random splitting and compromise the model's generalization ability, we strictly adhere to a chronological fivefold cross-validation strategy. This ensures that training and test sets in different folds do not share temporal information, thereby maintaining the validity and robustness of the experimental results.

For a given sample i , the input features comprise multiple physiological variables, structured as illustrated in Figure 1. In our DymaGNN model, besides the observed values at each time step, we incorporate a mask matrix to explicitly represent missing data. Specifically, assuming

Table 2. Features in the processed dataset.

Type	Variables
Demographic information	Gender, admission age, hospitalization sequence, hours from ICU admission, ICU stay sequence
Diagnosis	Hypertension, Paralysis, other neurological, Chronic pulmonary, Diabetes uncomplicated, Diabetes complicated, Hypothyroidism, Liver disease, Peptic ulcer, Aids, Lymphoma, Metastatic cancer, Solid tumor, Rheumatoid arthritis, Coagulopathy, Obesity, Fluid electrolyte, Deficiency anemias, Alcohol abuse, Drug abuse, Psychoses, Depression
Procedure	Arterial catheterization, Central venous catheter placement with guidance, Closed endoscopic biopsy of bronchus, Enteral infusion of concentrated nutritional substances, Hemodialysis, Insertion of feeding device into the stomach, Percutaneous approach, Insertion of infusion device into superior vena cava, Percutaneous approach, Inspection of the tracheobronchial tree, Via natural or artificial opening endoscopic, Introduction of nutritional substance into upper GI, Via natural or artificial opening, Percutaneous endoscopic gastrostomy [PEG], Percutaneous abdominal drainage, Venous catheterization for renal dialysis, Venous catheterization (not elsewhere classified)
Vital signs	Heart Rate, Arterial Blood Pressure systolic, Arterial Blood Pressure diastolic, Arterial Blood Pressure mean, O2 saturation pulseoxymetry, Respiratory rate, Blood Temperature CCO (C)
Laboratory tests	Sodium, Potassium, Chloride, Calcium, Magnesium, Phosphate, pH, O2 Pressure (PO2), Bicarbonate, Ionized Calcium, Lactate, Hemoglobin (Hb), Platelet, White Blood Cell (WBC), Red Blood Cells, Hemoglobin Concentration (MCHC), Mean Corpuscular Hemoglobin (MCH), Red Cell Distribution Width (RDW), Blood Urea Nitrogen (BUN), Creatinine, Glucose, Anion Gap, Temperature, Glasgow Coma Score (GCS)

ICU: intensive care unit.

the ICU dataset contains D dynamic features and the maximum sequence length is set to T , the mask matrix for sample i has a dimension of $D \times T$. The mask matrix element $M_u^{i,t}$ indicates whether feature u is observed at time step t in sample i , where $M_u^{i,t} = 1$ denotes the presence of an observation, while missing values are represented otherwise.

Dynamic mask attention graph neural network model

Patient physiological data in the ICU exhibit dynamic and heterogeneous characteristics, encompassing static and time-series features. The irregularity of data, and complex variable interactions in disease progression, makes intubation prediction highly challenging. To tackle these issues, we propose DymaGNN, a dynamic GNN that adaptively models heterogeneous ICU data, captures inter-variable dependencies and highlights critical time steps to improve predictive performance.

DymaGNN represents ICU data as a feature interaction graph in Figure 2, where nodes correspond to dynamic variables and edges capture their interactions, encoded via edge embeddings. While the graph structure remains static, node representations and edge weights evolve over time. To handle temporal irregularity, a masked attention mechanism

identifies key time steps, enhancing feature aggregation and prediction accuracy.

For a given sample i , the observation value of variable u at time step t is denoted as $x_u^{i,t}$. The corresponding time embedding is denoted as $p^{i,t}$. The initial node embedding is obtained through a multi-layer perceptron (MLP):

$$h_u^{i,t} = \text{MLP}(x_u^{i,t}) \quad (1)$$

For an edge (u, v) between variables u and v , we introduce an edge embedding $e_{u,v}^i$ and the time embedding $p^{i,t}$, computing a d -dimension interaction vector whose mean value serves as the edge weight. The edge embedding $e_{u,v}^i$ enables adaptive learning of inter-feature relationships:

$$\beta_{u,v}^{i,t} = \frac{1}{d} \sum_{j=1}^d ([e_{u,v}^i \odot h_u^{i,t}] \oplus [p^{i,t} \odot h_v^{i,t}])_j \quad (2)$$

A softmax normalization is then applied to derive the final edge weight $\gamma_{u,v}^{i,t}$:

$$\gamma_{u,v}^{i,t} = \frac{\exp(\beta_{u,v}^{i,t})}{\sum_{w \in N(u)} \exp(\beta_{u,w}^{i,t})} \quad (3)$$

The threshold top k retains top 50% strongest edge weights ($k = 50\%$) to prevent over-smoothing. The

Gap time (multi-time scale)							Label (if intube)
Feature name	2024/01/02 10:00:00	2024/01/02 10:15:00	2024/01/02 11:07:00	2024/01/02 11:10:00	...	max length	
feature u (heart rate)	90	95	☒	☒		92	1
feature v (respiratory)	15	☒	15	☒		☒	0
...	☒	☒	400	300		☒	...
feature N	☒	☒	☒	☒		120	
observe window							prediction window

Figure 1. Illustration of input data for ICU ventilation prediction.
ICU: intensive care unit.

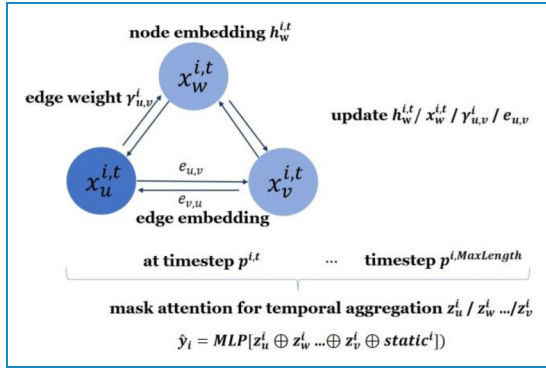


Figure 2. Hierarchical architecture of GNN layer in DymaGNN.
DymaGNN: dynamic mask attention graph neural network;
GNN: graph neural network.

indicator function $(\cdot)$ returns 1 when the condition holds. The variable representation is updated via neighborhood aggregation:

$$z_u^{i,t} = \sum_{w \in \mathcal{N}(u)} MLP(x_w^{i,t}) \cdot \gamma_{u,w}^{i,t} \cdot (\gamma_{u,w}^{i,t} > \text{top}_k). \quad (4)$$

To select key time steps, we employ a masked attention mechanism. We first compute the query, key, and value matrices:

$$Q_u^{i,t} = \mathbf{W}_Q z_u^{i,t}, \quad K_u^{i,t} = \mathbf{W}_K z_u^{i,t}, \quad V_u^{i,t} = \mathbf{W}_V z_u^{i,t} \quad (5)$$

where $\langle \cdot, \cdot \rangle$ denotes dot product, and $\sqrt{d}$ scales the gradient stability with $d = 64$ being the embedding dimension.

The attention weight is then computed as follows:

$$\alpha_u^{i,t} = \frac{\exp(\langle Q_u^{i,t}, K_u^{i,t} \rangle / \sqrt{d}) M_u^{i,t}}{\sum_{t'} \exp(\langle Q_u^{i,t}, K_u^{i,t'} \rangle / \sqrt{d}) M_u^{i,t'}} \quad (6)$$

The missing mask $M_u^{i,t}$ equals 1 for observed values to highlight the importance of real observed data, where the

mask $M_u^{i,t}$ is defined as follows:

$$M_u^{i,t} = \begin{cases} 1, & \text{if feature } u \text{ is observed at time step } t \\ 0.1, & \text{if feature } u \text{ is missing at time step } t \end{cases} \quad (7)$$

The final temporal aggregation representation is computed as follows:

$$\bar{z}_u^i = \sum_{t=1}^T \alpha_u^{i,t} \cdot V_u^{i,t} \quad (8)$$

To construct the final feature representation, we concatenate the time-aggregated representations of all dynamic variables $\bar{z}_u^i, \bar{z}_v^i, \dots$ with static features:

$$z^i = [\bar{z}_u^i \oplus \bar{z}_v^i \oplus \dots \oplus \bar{z}_w^i \oplus \text{static}^i] \quad (9)$$

The prediction of intubation need is then obtained via an MLP:

$$\hat{y}_i = MLP(z^i). \quad (10)$$

During training, we minimize the binary cross-entropy loss:

$$L = - \sum_i [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \quad (11)$$

DymaGNN effectively integrates feature interaction information with temporal dynamics in ICU data. By leveraging a dynamic GNN to model variable interactions and employing a masked attention mechanism to aggregate key time steps, our approach enhances the predictive accuracy of intubation need assessment in ICU patients.

Experiment results

We conducted several experiments to demonstrate the effectiveness of DymaGNN. We first compare the

performance of intubation predictions with various models. Then we conduct our model on other ventilation prediction tasks, including assistant ventilation prediction based on MIMIC-IV and intubation prediction based on eICU dataset. We also visualized DymaGNN's edge weight and found that many of the interrelationships between the features corresponded with medical knowledge, which greatly increased doctors' trust in DymaGNN.

We conducted ablation experiments and experimental effects of the model under different data loss percentages to demonstrate the effectiveness and robustness of our model. Experiments were conducted on the Google Tesla T4 GPU with 16GB memory.

Results in intubation prediction

We evaluate the performance of our proposed method against several baseline models, including SVM,⁶ XGBoost,²⁸ LSTM,²⁹ Transformer,³⁰ Multivariate Time Series Graph Neural Network (MTGNN),²² and Raindrop.²³ SVM, a traditional machine learning approach, is known for its stability in high-dimensional classification tasks. XGBoost, a widely used ensemble learning method for tabular data, exhibits robustness in handling missing values and outliers. LSTM and Transformer, both designed for sequential data modeling, effectively capture long-term dependencies. MTGNN, a GNN-based time-series forecasting model, constructs a graph structure to represent complex multivariate dependencies and demonstrates advantages in imputing missing values.

To ensure the reliability of the experimental results, we employ fivefold cross-validation, with five repeated experiments, reporting the average performance across all folds. The evaluation metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC), providing a comprehensive assessment of classification performance and model robustness. The results are presented in Table 3.

Our method achieves the best performance in most metrics, particularly showing a significant advantage in AUC (AUC = 0.8388, $p < 0.05$ vs. all baselines) over other methods. However, in accuracy, our method is outperformed by SVM (0.7711 vs. our 0.7725, $p > 0.05$), though not significantly. Similarly, XGBoost shows a slight edge in precision (0.7634 vs. our 0.7533, $p > 0.05$), and the difference is not significant. This implies that traditional models may still have some competitiveness in specific metrics. Overall, traditional machine learning models generally underperform compared to time-series and GNN models.

Among time-series models, LSTM performs worse than Transformer. Within GNN models, while our method has the highest F1 score on average, the difference is not significant compared to other methods.

Analysis of feature interaction in dynamic mask attention graph neural network

The edge weights $\gamma_{(u,v)}^{i,t} \in [0, 1]$ quantifies the dynamic relationship between features u and v at timestep t for sample i , derived from DymaGNN. To characterize global feature interactions, we compute the mean coupling weights $\bar{\gamma}_{(u,v)} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \gamma_{(u,v)}^{i,t}$, where N denotes the size of positive samples and T the maximum length of observation window.

As Figure 3 shows, the interaction pattern of physiological indicators revealed in this study is highly consistent with existing clinical knowledge, and this explainability helps to enhance clinicians' trust in the predicted results of our model. Specific findings are as follows:

RR exhibits correlations with blood glucose, calcium ions (Ca^{2+}), anion gap, hemoglobin (Hb), mean corpuscular hemoglobin concentration (MCHC), and red cell distribution width (RDW). Specifically, a patient in the MIMIC-IV dataset exhibited significantly low blood calcium levels 6.3 mg/dL at 05-04 17:30, the RR was observed to sharply increase to 26 breaths/min. This spike likely represents compensatory hyperventilation triggered by the hypocalcemic state. Subsequently, blood calcium levels gradually rose, increasing to 7.9 mg/dL by 05-04 15:00. In parallel, the RR steadily decreased and ultimately stabilized at a normal range of 14 breaths/min.

Elevated blood glucose levels are associated with increased RR,¹² as observed in diabetic ketoacidosis (DKA) patients who develop tachypnea to compensate for metabolic acidosis. This relationship is illustrated in the following clinical observation from a DKA patient: At 06-05 22:08:00, the patient's blood glucose was measured 145.0 mg/dL, with a corresponding 30-min average RR of 21.0 breaths/min. At the subsequent measurement 06-07 18:24:00, blood glucose increased to 335.0 mg/dL, and the average RR increased to 24.6 breaths/min during the same monitoring interval.

Hypocalcemia may impair neuromuscular function and induce respiratory distress,³¹ while anion gap expansion reflecting metabolic acidosis typically triggers compensatory hyperventilation (Kussmaul respiration).³² Reduced Hb, abnormal MCHC, and elevated RDW may further modulate RR through impaired oxygen transport efficiency.³³

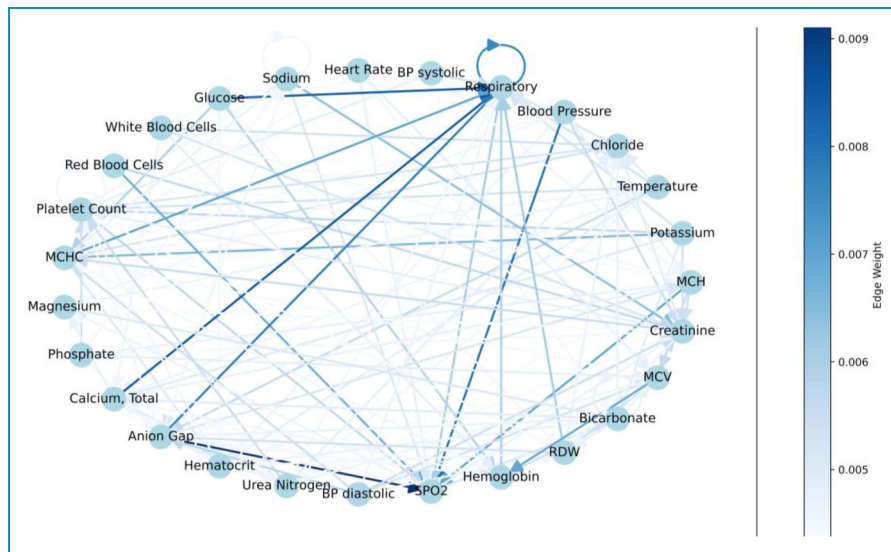
Oxygen Saturation (SpO_2) demonstrates multifactorial regulation involving RR, erythrocyte count, mean corpuscular hemoglobin (MCH), Hb, blood pressure, platelet count, and anion gap. Tachypnea enhances gas exchange efficiency,³⁴ whereas hypotension-induced hemodynamic alterations correlate with SpO_2 decline.³⁵ Hb concentration directly determines oxygen-carrying capacity, while erythrocyte count and MCH variations influence SpO_2 .

Table 3. Performance comparison of different models in intubation prediction.

Model	Accuracy	Precision	Recall	FI	AUC
XGBoost	0.6511 ^a	0.7634	0.4417 ^a	0.5587 ^a	0.7230 ^a
SVM	0.7711	0.5981 ^a	0.3442 ^a	0.4317 ^a	0.7388 ^a
LSTM	0.6653 ^a	0.6336 ^a	0.5973 ^a	0.6082 ^a	0.6974 ^a
Transformer	0.7165 ^a	0.6887 ^a	0.6144 ^a	0.6484 ^a	0.7776 ^a
MTGNN	0.7170 ^a	0.7510	0.7442 ^a	0.7465	0.7783 ^a
Raindrop	0.7354 ^a	0.7196 ^a	0.7698	0.7567	0.8082 ^a
Our Method	0.7725	0.7533	0.7704	0.7608	0.8388

Note. ^aindicates that the value is significantly different from that of our method using the Mann–Whitney U test ($p < 0.05$).

LSTM: long short-term memory; AUC: area under the curve; SVM: support vector machine; MTGNN: Multivariate Time Series Graph Neural Network.

**Figure 3.** Feature interaction.

through functional erythrocyte modifications.³⁶ Platelet dysregulation may indirectly impair oxygenation via micro-circulatory disturbances in pathological states.³⁷

Creatinine levels show associations with sodium ions (Na^+), leukocyte count, erythrocyte parameters, platelet count, MCHC, and MCH. Hyponatremia frequently accompanies renal dysfunction and altered creatinine metabolism.³⁸ Leukocytosis, indicative of systemic inflammation, may elevate creatinine through renal impairment.³⁹ Thrombocytopenia correlates with postoperative renal dysfunction, potentially via altered renal perfusion.⁴⁰ For instance, following surgery on 13 November, the patient remained mechanically ventilated. During this period, creatinine levels progressively decreased from 0.45 mg/dL on the operative day to 0.33 mg/dL by 26 November.

Concurrently, platelet counts increased from $136.5 \times 10^9/\text{L}$ to $608 \times 10^9/\text{L}$ over the same time.

Erythrocyte parameter abnormalities may affect renal hemodynamics through blood viscosity changes and oxygen delivery alterations.⁴¹ These findings underscore the necessity for multimodal renal function assessment in clinical practice.

Analysis of feature importance in XGBoost

Figure 4 shows the dynamic feature importance ranking given by the XGBoost model. Respiratory function-related features are prioritized, with SpO_2 as the most important, followed by mean arterial pressure and diastolic blood pressure. Subsequently, hematological features such as

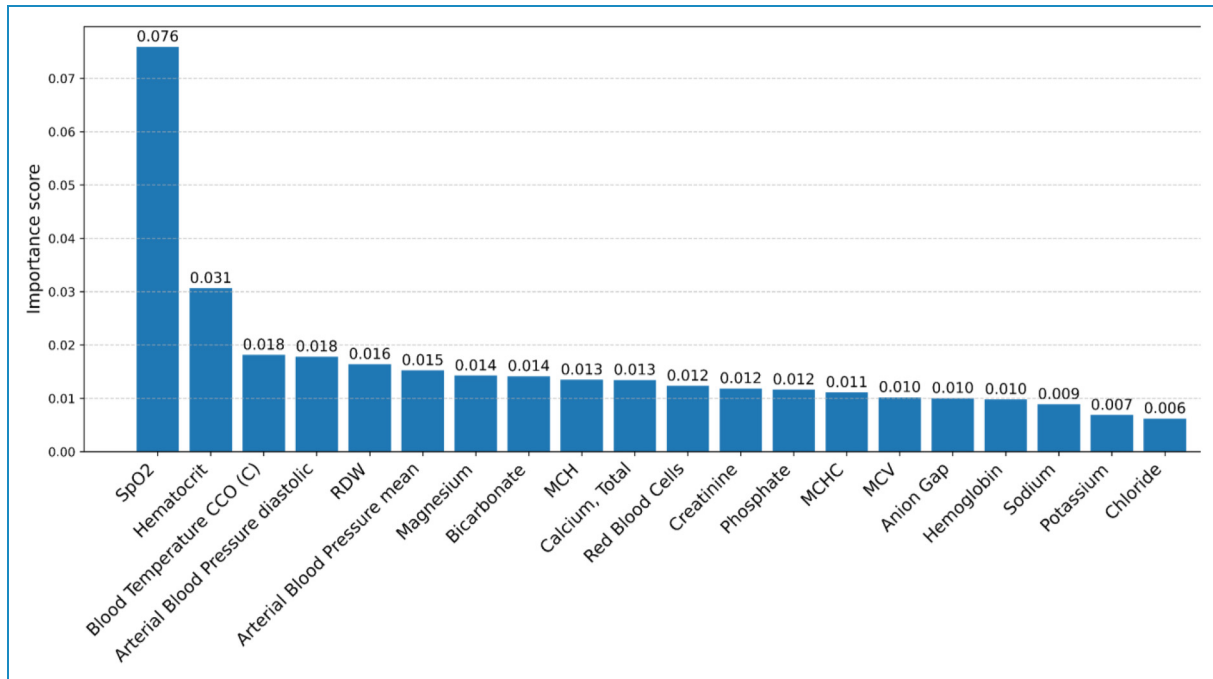


Figure 4. Importance of dynamic features in XGBoost.

Table 4. Performance comparison on different ventilation tasks.

Task	Accuracy	Precision	Recall	FI	AUC
NIV	0.6821	0.4661	0.6736	0.5495	0.7246
eICU	0.7579	0.8790	0.6795	0.7665	0.8557

ICU: intensive care unit; AUC: area under the curve; NIV: non-invasive ventilation.

hematocrit, Hb, mean corpuscular volume, MCH, and MCHC are also of significant importance. Additionally, features related to electrolyte balance, such as calcium and sodium ion concentrations, exhibit high significance. The prominence of these features indicates their crucial role in predicting the need for intubation in elderly ICU patients.

Performance in other ventilation tasks

To further evaluate the model's generalization capability across different ventilation prediction tasks for elderly ICU patients, we conducted experiments on NIV prediction using the MIMIC-IV dataset and intubation prediction using the eICU dataset. As shown in Table 4, the model demonstrates good performance in the eICU intubation prediction task (AUROC = 0.8557), indicating its efficacy in identifying critically ill patients requiring urgent mechanical ventilation. However, its performance significantly

declines in the NIV prediction task (AUROC = 0.7246, precision = 0.4661).

Influence of the missing rate

We evaluate three representative models—traditional machine learning SVM, sequence model Transformer, and our proposed GNN DymaGNN—to analyze their performance under varying data missing rates in Figure 5.

The SVM shows marked performance degradation, with a 12% AUC decrease at a 5% missing rate, confirming its strong feature dependency. Its performance progressively declines as the missing rates increase. The Transformer maintains robust performance (AUC > 0.78) below 10% missing rate, but exhibits accelerated deterioration when missing rates surpass 15%. Our DymaGNN, while following a similar trend to Transformer, demonstrated superior robustness. It sustains AUC above 0.81 at a 15% missing rate through dynamic graph structure adoption.

Figure 6 illustrates other metrics under varying data missing rates. SVM demonstrates remarkable stability in accuracy despite performance degradation in other metrics—precision drops 25% (from 0.72 to 0.53) and recall decreases 68% (from 0.54 to 0.17) at a 20% missing rate.

Our proposed DymaGNN shows trends with Transformer but significantly outperforms in recall metrics. Notably, the recall gap reverses from -0.04 (DymaGNN 0.77 vs Transformer 0.81) at 0% missing rate to $+0.14$ (DymaGNN 0.70 vs Transformer 0.56) at 20% missing rate.

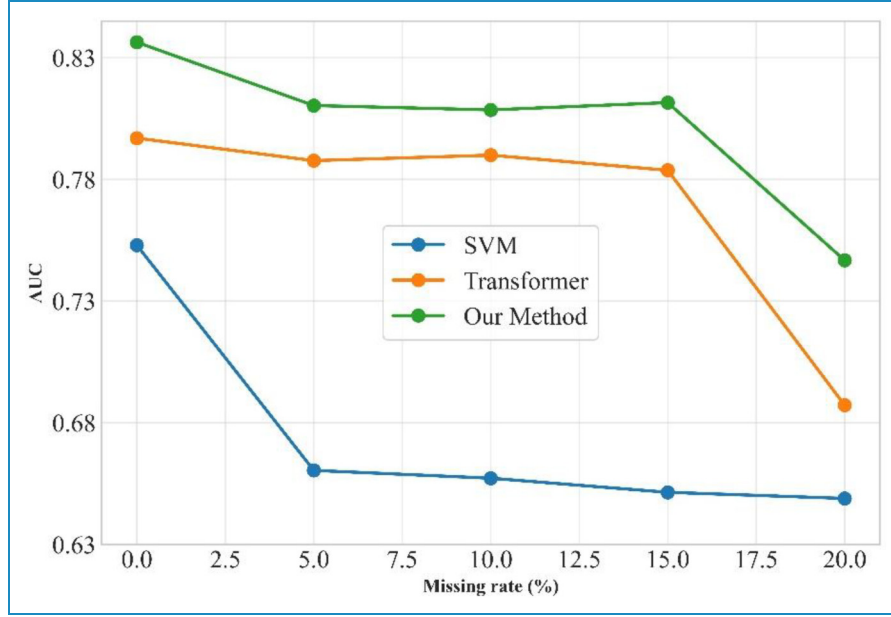


Figure 5. AUC performance under varying missing rates.
AUC: area under the curve.

Ablation experiment

To further validate the efficacy of DymaGNN, we conducted two ablation studies: 1) “weights” means fixing the time-varying edge weights to 1, and 2) “attention” means replacing the masked attention-based temporal aggregation with mean pooling.

As shown in Table 5, fixed edge weights caused significant performance degradation (AUC decreased by 2.85%; F1 decreased by 5.78%) compared to the original DymaGNN, confirming the critical role of dynamic edge weights in capturing time-evolving physiological interactions. Removing the masked attention module resulted in a marginal AUC decline to 0.8262, suggesting that this component enhances predictive stability through prioritized integration of critical time windows, while future work could explore better temporal aggregation strategies.

Discussion

Key findings

This study proposes DymaGNN to address intubation prediction in elderly ICU patients. Through comprehensive evaluations, our model demonstrates superior performance, particularly in the AUC metric, reflecting robust discriminative capacity in distinguishing between positive (intubation-required) and negative (non-intubation) patient groups. Although DymaGNN has the highest F1 score, it lacks statistical significance when compared to other GNN models. This indicates potential for refinement in balancing precision

and recall—a clinically critical consideration given that false negatives (indicating delayed intubation) may increase patient mortality, while false positives (leading to unnecessary intubation) escalate resource burdens.

Traditional machine learning approaches (SVM and XGBoost) exhibit limited capability in processing ICU data’s temporal and multivariate characteristics. While sequence models show improvement, the Transformer’s advantage over LSTM confirms the value of attention mechanisms in capturing time-series patterns, as established in Ayad et al.’s work.⁴² Notably, graph-based models excel at handling irregular data, managing different sampling frequencies effectively. Raindrop concatenates observation values with timestamp information, while DymaGNN employs a decoupled processing of observation values and timestamps. This design may preserve model expressiveness and reduce parameters, enhancing adaptability to sparse medical data.

Under varying missing data rates, the deterioration across multiple metrics in traditional machine learning models reveals their high dependency on complete data sets, where even a 5% data loss triggers a substantial performance decline.⁴³ Graph models exhibit significantly stronger robustness compared to traditional methods. This advantage may stem from DymaGNN’s ability to infer missing values through cross-feature dependencies—a capability aligned with Yalavarthi et al.’s findings on GNN superiority in missing data imputation.⁴⁴ This capability proves clinically critical for real-time monitoring scenarios where partial sensor failures occur. While most studies report strictly declining performance with increasing

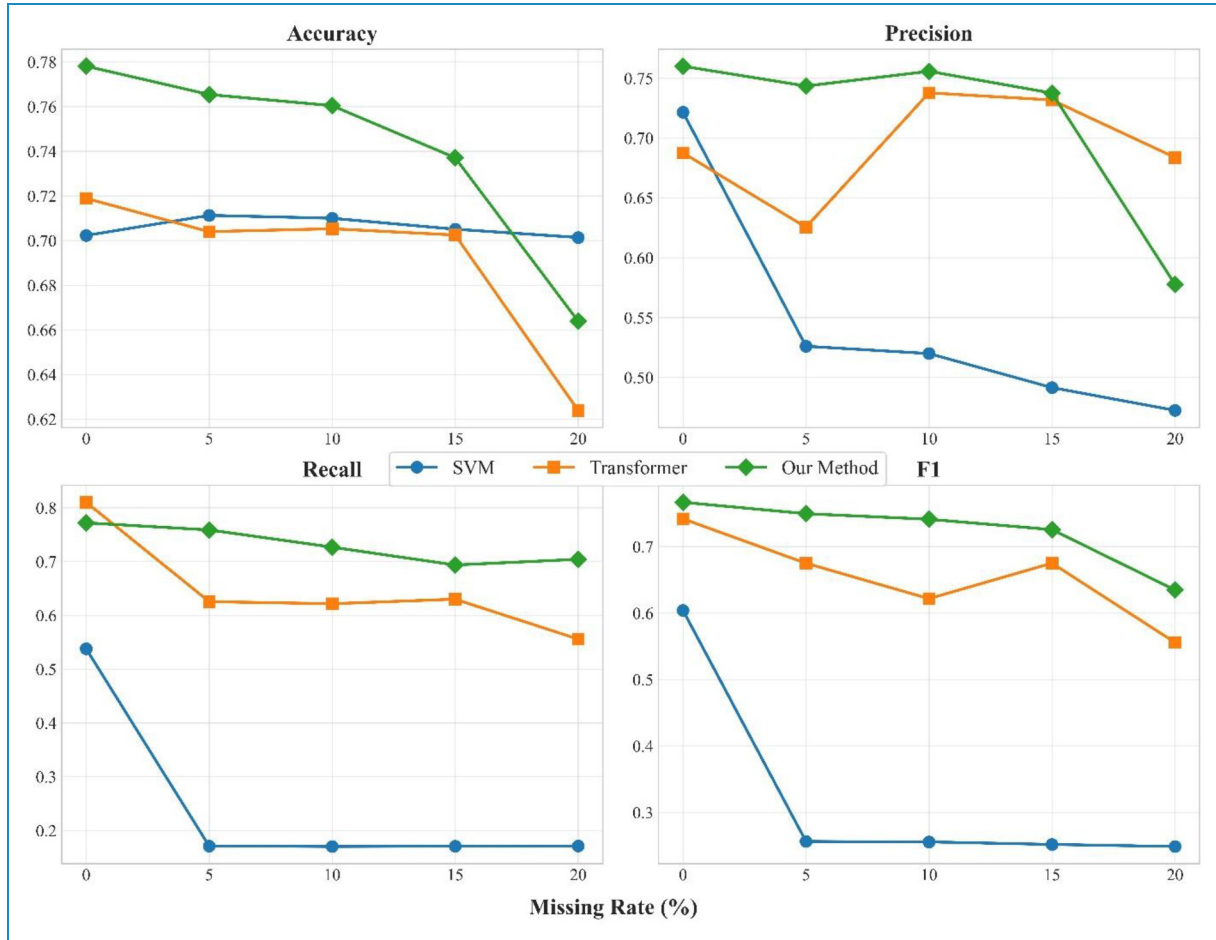


Figure 6. Performance under varying missing rates.

Table 5. Performance comparison of different model variations.

Task	Accuracy	Precision	Recall	F1	AUC
weight	0.7781	0.7600	0.7719	0.7666	0.8262
attention	0.7247	0.7092	0.7427	0.7223	0.8125

AUC: area under the curve.

missing rates,⁴⁵ our experiments reveal non-monotonic degradation patterns. This divergence may be attributed to contextual factors and experimental design, where random deletion occasionally removes critical features (e.g. SpO₂) despite lower overall missing rates.0.7608.

The XGBoost model reveals that respiratory function indicators (e.g. SpO₂), hematological parameters (e.g. Hb), and electrolyte balance indicators (e.g. Ca²⁺) are highly important, aligning with the key nodes (high degree or high edge weights) identified by DymaGNN. Notably, despite RR being a core physiological indicator of respiratory distress and compensation, and a proven predictor of intubation needs in prior studies,⁴⁶ it ranks relatively low

in feature importance in our XGBoost analysis. This may stem from the model's handling of strongly correlated features. SpO₂, a key output of respiratory function, is strongly correlated with RR and may partially “capture” or “replace” RR's predictive information in XGBoost, thereby underestimating RR's independent contribution. Kazemitabar⁴⁷ also supports that correlated variables can impact feature importance in XGBoost (the model remains robust in the presence of multicollinearity). In contrast, DymaGNN's dynamic interaction graph explicitly positions RR as a core node. DymaGNN models potential feature interaction pathways through its graph structure, allowing it to identify RR's importance even when its changes are often accompanied by SpO₂ changes. This mechanism-based explanation⁴⁸ better resonates with clinical practitioners' understanding.

Clinical implications

Our proposed model enhances ICU intubation prediction accuracy, which can potentially improve patient outcomes. Experiments show GNNs excel in handling ICU data with

irregular sampling, offering a robust benchmark for future ICU research.

Data quality, especially missing data rates, critically impacts model performance. For practical deployment during data collection, it is advisable to adhere to MIMIC-IV's data acquisition protocols⁴⁹ regarding sampling frequency and quality control standards. The minimum compliance threshold should not fall below 85%⁵⁰ of these benchmarks, as model efficacy substantially degrades when missing data rates exceed MIMIC-IV's baseline by approximately 15%.⁵¹

Moreover, the edge-weight-based feature importance analysis method introduced in the study improves model interpretability. It can be cross-verified with other feature importance methods, enhancing conclusion reliability while revealing feature interactions. This enables clinicians to assess whether the model's key features match clinical knowledge, boosting trust in model decisions and facilitating practical application.

Limitations and future work

Future research directions mainly include the following three aspects. DymaGNN performed well in the intubation prediction task, but decreased in the NIV prediction task. We may attribute this to two main factors. First, this difference may stem from the inherent uncertainty of assisted ventilation decision-making in clinical practice.⁵² Due to the lack of clear clinical indicators and significant individual differences in patients, there are significant differences in physician judgment in NIV decision-making. This uncertainty may lead more conservative physicians to adopt assisted ventilation prematurely.⁵³ Second, in the NIV prediction task with a positive-to-negative sample ratio of 2.94, the model exhibited inferior performance on imbalanced data. This imbalance likely induced a prediction bias toward the positive class, as evidenced by a relatively low precision—indicating a high false positive rate. Consequently, clinical deployment risks perpetuating unfair outcomes for minority patient subgroups, raising ethical concerns regarding algorithmic fairness.⁵⁴ To mitigate this limitation in future work, we suggest exploring the use of a weighted loss function to improve the model's performance.⁵⁵

Secondly, the dynamic weight graph in the current DymaGNN is entirely learned adaptively from data. In the next step of research, medical prior knowledge such as respiratory physiological mechanisms and the laws of blood gas compensation can be introduced to guide the construction of GNNs. This prior knowledge can not only reduce the scale of training parameters but also enhance clinical interpretability. Meanwhile, the aggregation method for time-level and feature-level⁵⁶ needs to be optimized to capture the dynamic evolution patterns of physiological variables more accurately.

Thirdly, this study mainly focuses on short-term intubation prediction. It can be further expanded, such as multiple ventilations, repeated ventilations, and prognosis assessment. For example, studying indicators like survival, rehabilitation status, and quality of life after receiving intubation can more comprehensively evaluate the effectiveness of mechanical ventilation decisions,⁵⁷ which can provide more long-term guidance for clinical treatment.

Conclusion

Our proposed DymaGNN presents a clinically valuable solution for predicting mechanical ventilation in elderly ICU patients. By dynamically modeling the critical, time-evolving interactions between physiological variables, DymaGNN achieves high predictive accuracy and maintains reliability even with 10% missing data. Crucially, its interpretable feature interaction graphs, which align with established clinical knowledge, provide clinicians with transparent insights into the model's reasoning. This integration of precision and explainability establishes a critical foundation for deploying trustworthy AI tools in real-world ICU intubation prediction.⁵⁸

Acknowledgements

This work was also supported by funding from the Shaanxi Province Health and Wellness Discipline Leader Visiting Scholar Program.

Ethical considerations

The MIMIC-IV and eICU databases are publicly available, and the creation of the research resource was reviewed by the Institutional Review Board at the Beth Israel Deaconess Medical Center. Permission to use the data was available in the Supplemental Material. The patient consent was waived as the data are wholly deidentified and retrospective from public databases.

Author contributions

Conceptualization done by XY and GJ; methodology by XY; software by XY; data curation by GJ; writing—original draft preparation by XY; writing—review and editing by GJ and XN; visualization done by XN. All authors have read and agreed to the published version of the manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the Scientific and Technological Talent Support Program at Shaanxi Provincial People's Hospital: (2023JY-37); Research Incubation Fund of Shaanxi Provincial People's Hospital (2023YJY-32).

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work; there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Data availability statement

Data and code are available from the corresponding author upon request.

ORCID iD

Jiao Guo  <https://orcid.org/0009-0001-0850-3943>

Supplemental material

Supplemental material for this article is available online.

References

1. Brunker LB, Bonczyk CS, Rengel KF, et al. Elderly patients and management in intensive care units (ICU): clinical challenges. *Clin Interv Aging* 2023; 18: 93–112.
2. Cheng A, Zhang Y, Qian Z, et al. Integrating multi-task and cost-sensitive learning for predicting mortality risk of chronic diseases in the elderly using real-world data. *Int J Med Inf* 2024; 191: 105567.
3. Huang J, Liu X and Jin W. Clinical decision support systems for 3-month mortality in elderly patients admitted to ICU with ischemic stroke using interpretable machine learning. *Digit Health* 2024; 10: 20552076241280126.
4. Lee KG, Roca O, Casey JD, et al. When to intubate in acute hypoxaemic respiratory failure? Options and opportunities for evidence-informed decision making in the intensive care unit. *Lancet Respir Med* 2024; 12: 642–654.
5. Laghi F, Shaikh H and Caccani N. Basing intubation of acutely hypoxemic patients on physiologic principles. *Ann Intensive Care* 2024; 14: 86.
6. Ossai CI and Wickramasinghe N. Intelligent decision support with machine learning for efficient management of mechanical ventilation in the intensive care unit—a critical overview. *Int J Med Inf* 2021; 150: 104469.
7. Castineira D, Schlosser KR, Geva A, et al. Adding continuous vital sign information to static clinical data improves the prediction of length of stay after intubation: a data-driven machine learning approach. *Respir Care* 2020; 65: 1367–1377.
8. Siu BMK, Kwak GH, Ling L, et al. Predicting the need for intubation in the first 24h after critical care admission using machine learning approaches. *Sci Rep* 2020; 10: 20931.
9. Fu Y, Liu Y, Zhong C, et al. An enhanced machine learning-based prognostic prediction model for patients with AECOPD on invasive mechanical ventilation. *iScience* 2024; 27: 111230.
10. Li H and Xie X. Deep learning-based prediction of mechanical ventilation reintubation in intensive care units. In: *INFORMS international conference on service science*, 2022, pp.15–22. Springer.
11. Kim Y, Kim H, Choi J, et al. Early prediction of need for invasive mechanical ventilation in the neonatal intensive care unit using artificial intelligence and electronic health records: a clinical study. *BMC Pediatr* 2023; 23: 525.
12. Im JE, Yoon SA, Shin YM, et al. Realtime prediction for neonatal endotracheal intubation using multimodal transformer network. *IEEE J Biomed Health Inform* 2023; 27: 2625–2634.
13. El-Rashidy N, Tarek Z, Elshewey AM, et al. Multitask multilayer-prediction model for predicting mechanical ventilation and the associated mortality rate. *Neural Comput Appl* 2025; 37: 1321–1343.
14. Sun C, Li H, Song M, et al. Time pattern reconstruction for classification of irregularly sampled time series. *Pattern Recognit* 2024; 147: 110075.
15. Zhang Y, Sheng M, Liu X, et al. A heterogeneous multimodal medical data fusion framework supporting hybrid data exploration. *Health Inf Sci Syst* 2022; 10: 22.
16. Mohanty A, Shashikumar SP, Lam JY, et al. Improving prediction of need for mechanical ventilation using cross-attention. In: 2024 46th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2024, pp.1–4. IEEE.
17. Li R, Xu Z, Xu J, et al. Predicting intubation for intensive care units patients: a deep learning approach to improve patient management. *Int J Med Inf* 2024; 186: 105425.
18. Kim GH, Kim JW, Kim KH, et al. FT-GAT: graph neural network for predicting spontaneous breathing trial success in patients with mechanical ventilation. *Comput Methods Programs Biomed* 2023; 240: 107673.
19. Fan S, Zhang S, Wang X, et al. Directed acyclic graph structure learning from dynamic graphs. *Proc AAAI Conf Artif Intell* 2023; 37: 7512–7521.
20. Guo H, Zhou Z, Zhao D, et al. EGNN: energy efficient anomaly detection for IoT multivariate time series data using graph neural network. *Future Gener Comput Syst* 2024; 151: 45–56.
21. Qiu X, Qian J, Wang H, et al. An attentive Copula-based spatio-temporal graph model for multivariate time-series forecasting. *Appl Soft Comput* 2024; 154: 111324.
22. Wu Z, Pan S, Long G, et al. Connecting the dots: multivariate time series forecasting with graph neural networks. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp.753–763.
23. Zhang X, Zeman M, Tsiligkaridis T, et al. Graphguided network for irregularly sampled multivariate time series. *arXiv preprint arXiv:2110.05357*.
24. Qureshi S. Limits of depth: over-smoothing and over-squashing in GNNs. *Big Data Min Anal* 2023; 7: 205–216.
25. Musbahi O, Syed L, Le Feuvre P, et al. Public patient views of artificial intelligence in healthcare: a nominal group technique study. *Digit Health* 2021; 7: 20552076211063682.

26. Navaneetham K and Arunachalam D. Global population aging, 1950–2050. In: *Handbook of aging, health and public policy: perspectives from Asia*. Singapore: Springer Nature Singapore, 2023, pp.1–18.
27. Lu C and Wenjuan J. Construction and evaluation of acquired weakness nomogram model in patients with mechanical ventilation in intensive care unit. *Digit Health* 2024; 10: 20552076241261604.
28. Wang Z, Zhang L, Huang T, et al. Developing an explainable machine learning model to predict the mechanical ventilation duration of patients with ARDS in intensive care units. *Heart Lung* 2023; 58: 74–81.
29. Li H and Xie X. Deep learning-based prediction of mechanical ventilation reintubation in intensive care units. In: *INFORMS international conference on service science*, 2022, pp.15–22.
30. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017; 30: 5998–6008.
31. Vilas-Boas EA, Cabral-Costa JV, Ramos VM, et al. Goldilocks calcium concentrations and the regulation of oxidative phosphorylation: too much, too little, or just right. *J Biol Chem* 2023; 299: 102904.
32. Maniscalco J, Hoffmeyer F, Monse C, et al. Physiological responses, self-reported health effects, and cognitive performance during exposure to carbon dioxide at 20 000ppm. *Indoor Air* 2022; 32: e12939.
33. Hough R, Cox SC, Chimelski E, et al. Prehospital critical care blood product administration: quantifying clinical benefit. *Dimens Crit Care Nurs* 2023; 42: 333–338.
34. Singhal A, Prafull K, Daulatabad V, et al. Arterial oxygen saturation: a vital sign? *Niger J Clin Pract* 2023; 26: 1591–1594.
35. Yan Y, Mao Z, Jia Q, et al. Changes in blood pressure, oxygen saturation, hemoglobin concentration, and heart rate among low-altitude migrants living at high altitude (5380m) for 360 days. *Am J Hum Biol* 2023; 35: e23913.
36. Yoshida T, McMahon E, Croxon H, et al. The oxygen saturation of red blood cell concentrates: the basis for a novel index of red cell oxidative stress. *Transfusion* 2022; 62: 183–193.
37. Van Aardt WJ, Bronner G and Buffenstein R. Hemoglobin–oxygen-affinity and acid-base properties of blood from the fossorial mole-rat, cryptomys hottentotus pretoriae. *Comp Biochem Physiol A Mol Integr Physiol* 2007; 147: 50–56.
38. Panizza J, Swee YJS, Edmundson A, et al. Renal dysfunction occurs following ileostomy formation and is independent of readmission. *ANZ J Surg* 2023; 93: 622–628.
39. Zhu W, Lu J, Lu J, et al. Clinical analysis of newly diagnosed multiple myeloma patients with renal dysfunction. *Zhonghua yi xue za zhi* 2015; 95: 741–744.
40. Hadipourzadeh F, Rastravan R, Totonchi Z, et al. Evaluating the relationship between lactate levels during coronary artery bypass graft surgery and postoperative renal dysfunction. *J Cardiovasc Thorac Res* 2024; 16: 129–134.
41. Sloop GD, Moore C, Pop G, et al. New onset anemia, Worsened plasma creatinine concentration, and hyperviscosity in a patient with a monoclonal IgM paraprotein. *Cureus* 2023; 15: e41657.
42. Ayad A, Hallawa A, Peine A, et al. Predicting abnormalities in laboratory values of patients in the intensive care unit using different deep learning models: comparative study. *JMIR Med Inform* 2022; 10: e37658.
43. Zeng Z, Liu Y, Yao S, et al. Neural networks based on attention architecture are robust to data missingness for early predicting hospital mortality in intensive care unit patients. *Digit Health* 2023; 9: 20552076231171482.
44. Yalavarthi VK, Madhusudhanan K, Scholz R, et al. Graffiti: graphs for forecasting irregularly sampled time series. *Proc AAAI Conf Artif Intell* 2024; 38: 16255–16263.
45. Hayakawa K, Uchino S, Endo H, et al. Impact of missing values on the ability of the acute physiology and chronic health evaluation III and Japan risk of death models to predict mortality. *J Crit Care* 2024; 79: 154432.
46. Heo JH, Kim T, Shin TG, et al. Using machine learning techniques for early prediction of tracheal intubation in patients with septic shock: a multi-center study in Korea. *Acute Crit Care* 2025; 40: 221–234.
47. Kazemitabar J, Amini A, Bloniarz A, et al. Variable importance using decision trees. *Adv Neural Inf Process Syst* 2017; 30: 425–434.
48. Delaunay J. *Explainability for machine learning models: from data adaptability to user perception*. Doctoral dissertation, Université de Rennes.
49. Piasecki J and Cheah PY. Ownership of individual-level health data, data sharing, and data governance. *BMC Med Ethics* 2022; 23: 104.
50. Rodemund N, Wernly B, Stundner O, et al. Beyond perfection: why imperfect routinely collected intensive care data still hold value. *Intensive Care Med* 2025; 51: 829–830.
51. Sittig DF and Singh H. Recommendations to ensure safety of AI in real-world clinical care. *JAMA* 2025; 333: 457–458.
52. Coppola S, Radovanovic D, Pozzi T, et al. Non-invasive respiratory support in elderly hospitalized patients. *Expert Rev Respir Med* 2024; 18: 789–804.
53. Lin MY, Chi HY and Chao WC. Multitask learning to predict successful weaning in critically ill ventilated patients: a retrospective analysis of the MIMIC-IV database. *Digit Health* 2024; 10: 20552076241289732.
54. Hanna MG, Pantanowitz L, Jackson B, et al. Ethical and bias considerations in artificial intelligence/machine learning. *Mod Pathol* 2025; 38: 100686.
55. Wang L, Guo X, Shi H, et al. CRISP: a causal relationships-guided deep learning framework for advanced ICU mortality prediction. *BMC Med Inform Decis Mak* 2025; 25: 165.
56. Zhao C, Tang H, Zhao H, et al. Beyond sequential patterns: rethinking healthcare predictions with contextual insights. *ACM Trans Inf Syst* 2025; 43: 1–32.

57. Luo J, Xie W, Hong S, et al. Comparison of outcomes between early and late tracheostomy. *Respir Care* 2024; 69: 76–81.

58. Pinsky MR, Bedoya A, Bihorac A, et al. Use of artificial intelligence in critical care: opportunities and obstacles. *Crit Care* 2024; 28: 113.

Appendix

Parameter selection for Top *k*

We adopted the Top *k* parameter to retain the strongest edge weights and prevent model over-smoothing. Experiments showed that *k* = 50% (retaining the top 50% strongest edge weights) delivered optimal results. As indicated in Table A1, AUC improvements were minimal when *k* exceeded 50%. Thus, we chose *k* = 50% for model simplicity. AUC: area under the curve.

Parameter selection for embedding dimension *d*

We evaluated the model’s performance with varying embedding dimensions *d*. The optimal results were achieved when *d* was set to 64, as shown in Table A2.

Table A1. AUC values for different *k* values.
AUC: area under the curve.

<i>k</i>	AUC
10%	0.6126
30%	0.7959
50%	0.8388
70%	0.8399
100%	0.8225

Table A2. AUC values for different *d* values.
AUC: area under the curve.

<i>d</i>	AUC
32	0.8122
64	0.8388
128	0.8307