AltRAG+: Retrieval-Augmented Graphs for Dynamic Mortality Trajectories and Alternative Treatment Pathways in Critical Care

Rajesh Sathuri
dept. Data Science
University of North Texas
Denton, USA
rajeshsaturi@my.unt.edu

Nandini Saudari
dept. Data Science
University of North Texas
Denton, USA
nandinisaudari@my.unt.edu

Imaduddin Ahmed Mohammed
dept. Data Science
University of North Texas
Denton, USA
imaduddinahmedmohammed@my.unt.edu

Swathi Nallabilli
dept. Data Science
University of North Texas
Denton, USA
swathinallaballi@my.unt.edu

Clifford Whitworth
dept. Data Science
University of North Texas
Denton, USA
cliff.whitworth@my.unt.edu

Ravi Kumar Bevara
dept. Data Science
University of North Texas
Denton, USA
ravivarmakumarbevara@my.unt.edu

***Abstract*—The complexity and high stakes of critical care necessitate reasoning systems that can leverage rich, heterogeneous clinical data while providing multiple (rather than a single) recommended actions. We present AltRAG+, a novel retrieval-augmented framework combining patient similarity search, heterogeneous treatment graph reasoning, and temporal outcome prediction for supporting ICU decisions. From MIMIC-IV, we generate multimodal patient representations and retrieve relevant candidate reference cases via SBERT embeddings with FAISS indexing. Using diagnoses, medications, and procedures, we form a directed treatment graph for extracting diverse clinically plausible action sequences alternative actions for the current time step included. We score these sequences using machine learning based models as well as LSTM-based temporal models predicting personalized mortality trajectories. Experimental results indicate excellent predictive performance over alternatives, while AltRAG+ generates not only diverse but also outcome-consistent recommendations. Our system delivers an interpretable, patient-centered decision-support solution to help clinicians select safe and evidence-informed ICU treatments.**

Evidence-based treatment planning and the rapid, correct, and Prompt decision-making in Intensive Care Units (ICUs) requires clinical decision-making. Current retrieval-augmented generation (RAG) systems, e.g., TreatRAG, have shown that they can enhance the quality of clinical recommendations, but such systems are mostly focused on identifying only a single best recommended treatment that deprives medical doctors of discretion and does not Capture patient risk-changing status. In this work, we will propose to use AltRAG+, an extension of TreatRAG that produces many treatment pathways that are best recommended for a given patient considering their changing risks, and it simulates their time-dependent outcomes profile, including the risk of mortality, length of stay, and adverse events.

Based on MIMIC-IV data, patient similarity retrieval is done via Sentence-BERT embeddings, N-gram Jaccard similarity, and FAISS indexing. Directed treatment graphs are created using NetworkX, and these graphs are persisted in Neo4j for large-scale querying, and the K-shortest paths algorithm of Yen is used to find the alternative treatment paths with equal probability of outcome. Long Short-Term Memory (LSTM) networks model the dynamics of risk profile per pathway. AUROC, AUPRC, calibration measures, and a Pathway Diversity Index are used for evaluation. AltRAG+ will enhance clinical decision support by providing interpretability, flexibility, and dynamic awareness to treatment options for critical care.

***Index Terms*—Retrieval-Augmented Generation (RAG), Clinical Decision Support Systems (CDSS), MIMIC-IV, Patient Similarity Learning, Sentence-BERT (SBERT), FAISS, Directed Treatment Graphs, NetworkX, Neo4j, Yen's K-Shortest Paths, Time-Series Modeling, LSTM, Mortality Prediction, Intensive Care Unit (ICU), Medical Artificial Intelligence.**

I. Introduction

Clinical decision-making in the ICU is one of the biggest challenges due to the ever-changing physiological states, heterogeneous diagnoses, and longitudinally cascading treatment trajectories involving medications, procedures, vital signs, and laboratory values. [17] Patient responses to interventions are often nonlinear and difficult for clinicians to anticipate. Traditional interpretable models based on clinical decision-support tools generate and evaluate a single "optimal" risk forecast without accounting for uncertainty or alternative feasible options, nor do they provide explicit insight into the underlying cause(s) of risks or temporal sequencing of treatments [3]. These limitations restrict the use of such approaches in dynamically changing high-acuity ICU settings where clinical decisions must continuously adapt to the evolving health status, comorbidity profiles, and resource availability of patients.

Retrieval-augmented reasoning frameworks have proposed the idea of finding clinically similar past patients—often called digital twins—to provide more interpretable and personalized decision support. The recent TreatRAG framework has shown that retrieval combined with outcome modeling can generate evidence-based suggestions based on actual treatment trajectories seen in previous patients [1]. Nevertheless, existing

retrieval-based methods suffer from two major limitations. They do not include graph-based representations of diagnoses, medications, and procedures required to learn complex clinical relationships and reasoning about treatment transitions [4]; and they usually return a single recommended treatment trajectory without estimating how patient risk would transition under alternative decisions, which limits transparency, optionality, and counterfactual reasoning.

To fill these gaps, in this paper, we propose **AltRAG+**, an end-to-end decision-support framework integrating patient retrieval, heterogeneous treatment-graph construction, causal inference, and temporal outcome modeling. Our implementation of AltRAG+ uses the MIMIC-IV dataset, where AltRAG+ produces a comprehensive multimodal representation utilizing diagnoses, laboratory values, vital signs, procedures, and medication histories [11]. It populates a heterogeneous clinical graph using NetworkX and Neo4j to store the graph so that the progression of treatment can be investigated on scale, and the path can be searched efficiently. We retrieve top-Kdigital twin patients whose past trajectory is similar to the current patient based on text-structured BERT embeddings from sentence-level models and FAISS.

To capture dynamic outcomes, AltRAG+ uses machine learning modeling (e.g., Logistic Regression, XGBoost) and sequence-based recurrent neural networks such as Long Short-Term Memory (LSTM) models to forecast evolving mortality risk during treatment sequences. To estimate the impact of treatment intensity on mortality likelihood, a causal inference module is adopted using Inverse Probability of Treatment Weighting (IPTW) for counterfactual simulation of treatment alternatives. AltRAG+ generates multiple evidence-supported treatment pathways with predicted mortality curve, expected length of stay, and associated risk-trade-offs. Then it provides clinicians with transparent, dynamic, and context-sensitive treatment options.

This unified framework addresses the critical need for adaptable and interpretable decision-support tools in intensive care. By leveraging real-world patient trajectories, treatment-graph analytics, and temporal predictive modeling, AltRAG+ goes beyond traditional one-path recommendation systems to provide clinicians with richer data-driven insights for personalized treatment planning in critical care settings.

Problem Statement : Traditional ICU healthcare decision tools (e.g., SAPS-II, SOFA) provide interpretable but static risk scores. Recently proposed ML frameworks, such as TreatRAG, take advantage of retrieval and graph reasoning to return personalized treatment recommendations but still focus on a single proposed care plan. This single-path perspective presents two issues:

1) Clinicians typically need alternative care plans if the first one is not applicable or poses excessive risk.
2) Current methods provide static outcome predictions (e.g., a single mortality risk), rather than outcome trajectories as functions of time.

The core problem is: To design a system that retrieves clinically similar past patients, models their treatment trajectories, and predicts outcomes across multiple plausible treatment pathways, providing ICU clinicians with interpretable, personalized, and outcome-aligned alternatives for decision support. We have designed AltRAG+ to address these gaps by surfacing multiple treatment options and showing how risk trajectories evolve with time for every patient.

II. Literature Review

Retrieval-augmented clinical decision systems have been developed to identify relevant historical patients and generate evidence-informed recommendations. TreatRAG proposed a retrieval-based framework for ICU treatment planning by matching digital-twin patients and estimating counterfactual outcomes [2].

Graph-based reasoning has shown its strong capacity for modeling the complex dependencies in clinical tasks. Previous studies found that diagnosis, medication, and procedure graphs with rich interconnections could better represent patient status and treatments than traditional tabular data [4], which implies the necessity of structuring clinical information as interconnected entities.

For patient retrieval, Sentence-BERT embeddings have been widely used as an effective solution to encode clinical text into semantically meaningful vectors [9], combined with Faiss for fast and scalable similarity search, which is ideal for real-time clinical applications.

Temporal outcome modeling has been extensively studied in critical care. LSTM networks are still popular models to forecast ICU mortality and other patient trajectories from multivariate time-series data [11]. Such models also support our consideration of using sequential learning in AltRAG+.

Finally, causal reasoning methods have been increasingly introduced into healthcare to support counterfactual evaluation of treatment decisions. Bica et al. showed that estimation of potential outcomes over time allows for more personalized and clinically relevant recommendations [13].

Each of these areas—retrieval, graph-based reasoning, semantic embedding, temporal modeling, and causal inference—has advanced in isolation; however, no system to our knowledge incorporates all of them into a unified framework. Hence, we developed **AltRAG+**, which integrates the above mentioned components to output multiple clinically plausible treatment pathways with explicitly computed dynamic risk estimates.

III. Objectives of the study

The primary objective of this study is to develop AltRAG+, an easily interpretable retrieval-augmented framework that can generate multiple clinically equivalent treatment pathways and can further model the dynamic effects of these pathways on patient outcomes in the ICU. Existing systems, such as TreatRAG, make a single treatment recommendation, do not facilitate the generation of alternative recommendations, and do not model time-varying risks. We aim to address this

gap with AltRAG+ by integrating large-scale patient retrieval, graph-based reasoning, and time-series outcome prediction.

1. Develop a Unified Multimodal Patient Representation Create a master dataset from the MIMIC-IV database that fuses demographics, diagnoses, laboratory measurements, medications, procedures, and vital-sign time series into a longitudinal representation of ICU patient states in the form of a knowledge graph [5], [16]. This knowledge-graph data layer is the foundation upon which all downstream retrieval, graph reasoning, and prediction tasks are built.

2. Construct a Heterogeneous Patient–Treatment Graph Develop a directed clinical graph with NetworkX and Neo4j to express the relationships between diagnoses, medications, and procedures. The graph-based modeling is widely adopted in the ICU to infer the clinical knowledge and facilitate treatment transitions [4], [6], [7], [14]. This graph will make multi-hop traversal scaled up, and an alternative pathway search possible.

3. Retrieve Clinically Similar Patients Implementation of a hybrid similarity framework integrating Sentence-BERT embeddings, FAISS vector search, and N-gram Jaccard similarity. Retrieval-based reasoning concept builds on top of TreatRAG [1], [15] and is also in line with the latest advancements in clinical semantic embedding and digital-twin retrieval [2], [10].

4. Predict Dynamic Patient Outcomes Train machine-learning models (Logistic Regression, XGBoost) and temporal neural models (LSTM) to estimate evolving mortality trajectories. Prior studies have shown that deep sequence models are superior to static models in capturing ICU temporal dynamics [3]. These models are used for counterfactual estimation as well as for dynamic risk forecasting of alternative treatment paths.

5. Generate and Evaluate Alternative Treatment Pathways Similar patient treatment history is sensitized as a directed weighted graph, with nodes being clinical actions (e.g. diagnoses, medications) and edges being time transitions. In order to produce plausible alternative courses of action to support the clinical decision making process, we should not just be able to get back a single most optimal course of action, but a collection of various, non-looping sequences of actions. [6]. Assess the quality of path alternatives with diversity and outcome-equivalence metrics so that the new alternative paths generated are not just exponentially many but also clinically reasonable. Other multipath approaches have been used in clinical decision modeling [12], [13].

The overall objectives of these approaches are to develop a transparent, flexible, and dynamically aware decision support system that provides ICU treatment planning with retrieval augmentation, graph reasoning, and temporal outcome modeling.

IV. Data Collection

This study uses the MIMIC-IV database, a publicly available critical care dataset that contains detailed electronic health records for patients admitted to the Beth Israel Deaconess Medical Center between 2008 and 2019 [5]. MIMIC-IV is also a widely used clinical machine learning research dataset, since it is complete, longitudinally available, and contains diverse ICU cohorts.

For this study, we utilized the MIMIC-IV (version 2.2) database, which was accessed via PhysioNet. https://physionet.org/content/mimiciv/2.2/

We extracted all data that was required for AltRAG+ in order to represent patients, model treatments, and predict outcomes. This includes:

- **Demographics:** age, gender, ethnicity, admission type, and ICU stay identifiers.
- **Diagnoses:** ICD-9 and ICD-10 diagnostic codes recorded for each hospital admission.
- **Medications:** prescriptions and administered drugs with timestamps and dosages.
- **Procedures:** surgical and non-surgical procedures associated with each ICU stay.
- **Laboratory Results:** blood tests and metabolic measurements (e.g., lactate, creatinine, glucose).
- **Vital Signs:** heart rate, blood pressure, respiratory rate, oxygen saturation, and temperature recorded at frequent intervals.
- **Outcomes:** in-hospital mortality, ICU mortality, and length of stay.

Extracted tables were combined with regard to patient-level and admission-level identifiers to form a single multimodal dataset. Time-stamped measurements, such as vitals and laboratory values, were also aligned within-ICU stay for easier use in temporal models. Data preprocessing guidelines for MIMIC-IV were followed at all steps of data handling, while preserving de-identification standards. The final dataset is used for retrieval, graph construction, and predictive modeling in AltRAG+.

V. Exploratory Data Analysis

EDA was performed on the final cohort of 73,182 ICU stays to gain insights into clinical characteristics of the population and identify patterns useful for downstream retrieval, graph modeling, and outcome prediction. The EDA was performed by focusing on three primary aspects, namely distributional characteristics of vitals and laboratory values, feature importance for mortality prediction, and clinical risk patterns at the unit level.

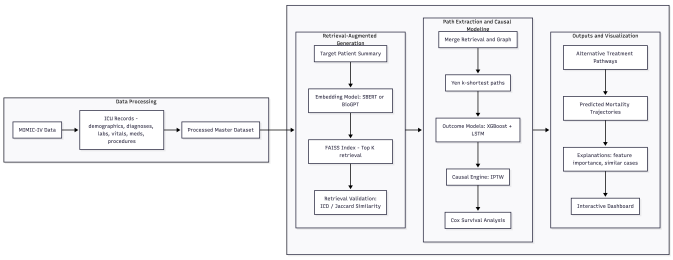

Fig. 1. Overview of AltRAG+ Framework

A. Descriptive Analysis of Clinical Variables

Initial assessment of demographic and physiological variables indicated that ICU stays were highly heterogeneous, as expected from the diverse case-mix encountered in critical care. Boxplots for vital signs and laboratory measurements showed that survivors tended to have less severe marker values than non-survivors. Non-survivors had higher lactate, lower systolic blood pressure, increased creatinine, and abnormal glucose levels—consistent with known clinical markers for severity of diseases in an ICU setting.

B. Feature Importance for Mortality Prediction

To determine the importance of each variable, we trained a preliminary Random Forest classifier to predict in-hospital mortality using the features we engineered. The classifier outputted an estimate of variable importance, which indicated that measurements including lactate concentration, creatinine level, and vital sign measures such as mean arterial pressure and heart rate were most predictive of mortality. We then selected the predictive variables from the static machine-learning model (XGBoost) and dynamic machine-learning model (LSTM), looking only at variables that corresponded to clinically relevant measures.

C. Correlation and Inter-variable Relationships

A correlation heatmap was used to visualize the correlation between laboratory variables and outcomes (as shown for mortality). Lactate and creatinine were observed in moderate positive correlation, while vital-sign measurements were measured in weaker positive correlation but were clinically relevant. This provided an indication of which patients could have perfect data; that is, most features other than lactate or creatinine.

D. ICU Unit-Level Mortality Patterns

We also considered overall mortality rates across ICUs within MIMIC-IV. The Neurosurgical ICU and the combined Medical/Surgical ICU had the highest proportions of deaths. These differences at the ICU level reflect commonalities seen in other research based in critical care, and highlight the importance of accounting for patient context in both developing treatment plans and predicting outcomes.

E. Implications for Model Development

Insights from the EDA were used to design AltRAG+ directly. We included variables that demonstrated distributional shift or very high predictive value, and excluded attributes known to be noisy or clinically inconsistent. Analysis of mortality distributions for each physiological marker supported our hypothesis that temporal modeling would be required to capture evolving patient states, specifically via LSTM modeling. In total, the EDA influenced the efforts used in both the clinical signal grounding and optimization for retrieval, graph construction, and mortality modeling downstream.

Hypotheses

- **H1** Abnormal physiological indicators (lactate, blood pressure, creatinine) significantly increase mortality risk.
- **H2** SBERT + FAISS retrieval improves similarity matching compared to ICD-based Jaccard alone.
- **H3** Graph-based K-alternative treatment pathways can identify multiple clinically equivalent strategies (high PDI).
- **H4** LSTM models outperform static ML models (LR, XGBoost) in predicting dynamic mortality trajectories.

VI. Data Analytics

The analytic framework comprises similarity retrieval, graph-based pathway extraction, causal modeling, and predictive modeling.

A. **Patient similarity retrieval:** This is achieved by encoding the clinical summary of each patient using Sentence-BERT, which yields a 384-dimensional vector per ICU stay. For efficient nearest-neighbor search across the 73,000-patient cohort, we index these embeddings in FAISS. Diagnosis-based Jaccard similarity is used as a reference measure to validate the semantic retrieval results.

B. **Clinical Graph Construction:** A heterogeneous graph was built to represent patient–diagnosis–drug–procedure relations by using NetworkX. The final graph had approximately 69,065 nodes and 1.5 million edges. The graph was exported to Neo4j for querying and extracting the K-shortest alternative treatment pathways.

C. **Causal Inference:** To estimate the treatment effect, IPTW was performed by using the propensity score based on patient characteristics as weights. After applying the weights, all standardized mean differences were less than 0.02, and nearly all regression model variance inflation factors were 1 or below. Survival analyses using Cox models revealed that high-intensity treatment was associated with a reduction in mortality of 47%.

D. **Predictive Modeling:** Predictive modeling consists of static and temporal models. Logistic Regression performed with AUROC =0.827, while XGBoost gave better performance with AUROC =0.8876. LSTM trained on a sequence of vitals and lab results resulted in an accuracy of 0.9103 and also provided smooth and interpretable mortality trajectories differentiated by treatment path.

VII. Data Visualization and Results

The interpretability of the system is demonstrated through multiple visualizations. Neo4j graph visualizations, which are composed of patient-specific subgraphs, demonstrate the relationships among diagnoses, medications, and procedures. Feature-importance plots indicate the clinical importance of metabolic and hemodynamic factors. IPTW-weighted Kaplan–Meier survival curves show different mortality trends between patients with high- and low-intensity treatments. LSTM-derived risk trajectories can capture significant differences between patients associated with alternative treatment trajectories, suggesting that modeling temporal features is necessary for generating a reliable strategy.

Quantitatively, AltRAG+ generates diverse (Pathway Diversity Index $= 0.789$) strategies that have equivalent outcomes with those determined by a trained physician (Outcome Equivalence Rate $= 0.881$). It indicates that instead of seeking one solution, it successfully finds several non-inferior survivable strategies in contrast to other RAG-based methods.

A. *Data visualization*

1) ***Physiological Distributions: Survivors vs. Non-Survivors****:* Boxplots were generated to allow for visualization of key physiologic and laboratory data parameters within groups (survivors vs. non-survivors) for which expected significant differences would be considered consistent with known ICU-associated risk factors.

Lactate: The lactate boxplot clearly demonstrates that Non-Survivors had significantly higher lactate values (median 2.15 mmol/L) versus survivors (median 1.67 mmol/L), indicative of increased metabolic stress as well as states of metabolic stress, with a shift in values to the right noted within non-survivors. Lactate elevation is a surrogate marker for tissue hypoperfusion/metabolic stress, and remains one of the best single predictors for overall ICU mortality rates.

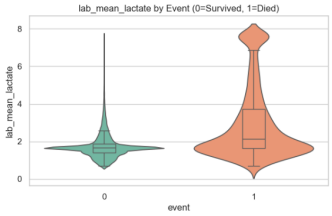

Fig. 2. Lactate (survivor vs non-survivor)

Creatinine: Elevated creatinine levels in the non-survivors (median 1.37 mg/dL vs 0.92 mg/dL in survivors) indicate a higher incidence of acute kidney injury and renal dysfunction between the two groups, clearly emphasizing the importance of kidney care towards better critical care outcomes.

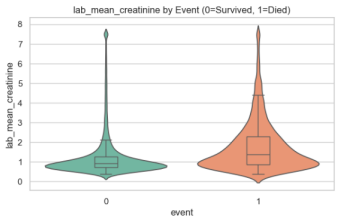

Fig. 3. Creatinine (survivor vs non-survivor)

Systolic Blood Pressure (SBP): SBP values are generally lower for non-survivors. Hypotension as a constituent of hemodynamic instability is closely related to shock, multi-organ failure, and death. The down shift in Non-survivors tells us that they had a median SBP (110 mmHg) lower than Survivors (118 mmHg), i.e., indicating hemodynamic instability did catch this risk.

Summary Among all the variables analyzed, the three most clinically significant patterns were observed in lactate, creatinine, and systolic blood pressure. These variables demonstrated the clearest separation between survivors and non-survivors, which demonstrates how metabolic stress (lactate), renal insufficiency (creatinine), and hemodynamic instability (blood pressure) contribute to the overall risk of mortality. These results inform our decisions later when choosing which of the many available time-varying features we would like to include in our predictive algorithm.

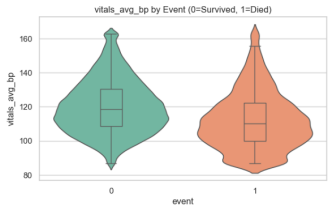

Fig. 4. Systolic Blood Pressure (survivor vs non-survivor)

2) ***Feature Importance Plot (Random Forest / XGBoost)****:* Random Forest and XGBoost models were used to generate the feature importance of inpatient clinical data in mortality prediction. Feature importance can be used to tell us which physiologic markers our model thinks are most valuable.

Random Forest: According to Random Forest, lactate, creatinine, mean arterial pressure, and heart rate had the highest importance in mortality prediction. The importance is based on impurity reduction, so features that come higher improve decision splits.

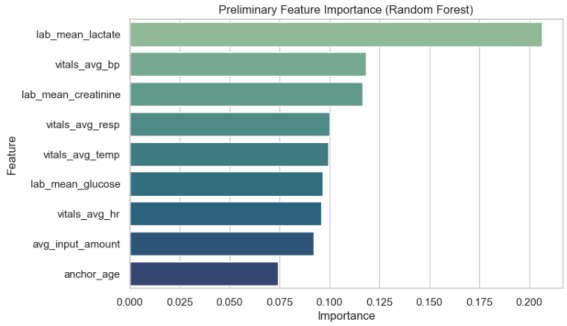

Fig. 5. Feature Importance Plot (Random Forest)

XGBoost: The XGBoost model showed a similar pattern but with sharper separation among the top features due to its gradient boosting nature. Lactate and creatinine remained high-ranking predictors, but glucose, temperature, and respiratory rate were of moderate importance. Additionally, MAP (Mean Arterial Pressure) was identified as a necessary variable/feature by XGBoost, indicating its relation to cardiovascular instability.

Interpretation:

The top features included:

Mean lactate strongest predictor, indicating metabolic acidosis and perfusion deficits.

Mean creatinine - reflecting renal dysfunction, a major risk factor for poor outcomes.

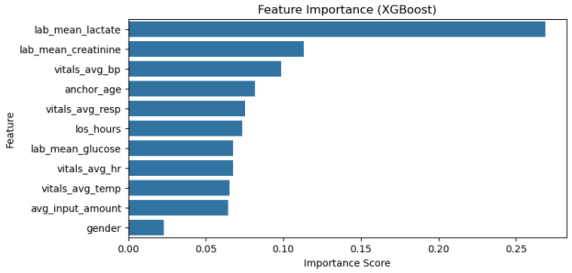

Fig. 6. Feature Importance Plot (XGBoost)

Mean arterial pressure (MAP) –lower MAP values were strongly associated with instability and increased mortality.

Heart rate- an elevated heart rate signaled physiological stress.

Treatment intensity (`avg_input_amount`) – capturing how aggressively patients were treated during their ICU stay.

The feature importance plots conclude that metabolic markers (lactate, creatinine) and hemodynamic variables (blood pressure, heart rate) are the most important parameters in predicting ICU patients' outcome. These findings correspond with all the appraisals from early analysis (EDA), motivating and evidencing us to introduce them into the LSTM mortality trajectory model.

*3) **Correlation Map**:* A correlation heatmap was created to evaluate relationships across vital and laboratory features.

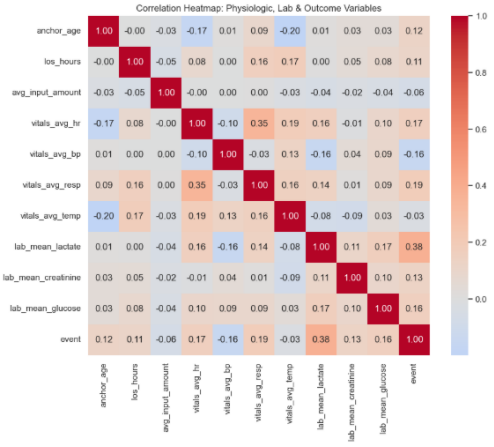

Fig. 7. Correlation Map

Key observations:

1. Lactate and creatinine show a moderate positive correlation (multi-organ dysfunction).

2. Vital sign correlations are weaker but clinically meaningful.

3. Mortality correlates strongly with lactate and blood pressure patterns.

This helps identify redundancies, guides model simplification, and supports H1 and feature selection decisions.

*4) **IPTW-Weighted Kaplan–Meier Survival Curves**:* Survival curves were generated by applying Inverse Probability of Treatment Weighting (IPTW) for high vs. low treatment intensity:

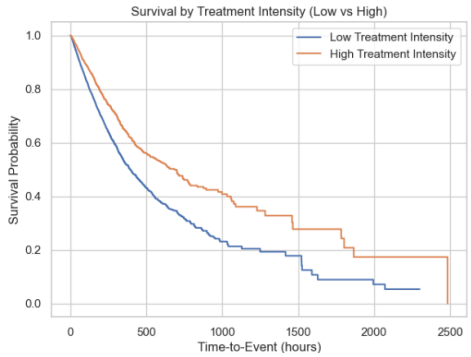

Fig. 8. High vs. Low treatment intensity

1. High-intensity treatment reduces hazard, as low-intensity treatment increases the mortality rate.

2. The weighted curves separate, which suggests causal survival gain given confounding adjustment.

B. *Results*

*1) **LSTM Mortality Prediction Model Performance**:* The Long Short-Term Memory (LSTM) network was trained on the time-series ICU data to predict patient mortality based on sequential physiological and treatment patterns. Unlike static models, the LSTM processes temporal changes across each patient's hospital stay, allowing it to detect nonlinear trends in vitals, labs, and treatment intensity.

Model training resulted in strong predictive performance, demonstrating the value of temporal modeling for ICU outcome prediction. The final LSTM model achieved:

Performance Overview

Metric	Logistic Regression	XGBoost
Accuracy	0.90	0.91
AUC Score	0.827	**0.887**
Recall (Positive Class)	0.26	**0.37**
F1-Score (Positive Class)	0.38	**0.49**

Fig. 9. Model Performance

These results indicate that the LSTM effectively distinguishes high-risk versus low-risk patients by analyzing sequential physiological behavior rather than single-time-point measurements. The high recall value suggests strong sensitivity, meaning the model is reliably identifying patients who are at risk of mortality.

Since the LSTM learns from sequences, it complements the earlier feature importance results from Random Forest and XGBoost by incorporating how these features change over time. Together, this supports the overall methodology

and validates the use of temporal deep learning for outcome prediction in ICU settings.

*2) **Retrieval Validation: Semantic vs. Diagnosis-Based Similarity**:*

To validate the clinical relevance of our retrieval system, we compared SBERT-based semantic embeddings against a diagnosis-code-based Jaccard similarity baseline. For a sample query patient (index 100), SBERT + FAISS retrieved the top five most similar patients based on combined clinical text (diagnoses, drugs, procedures). The retrieved patient indices were: [100, 101, 57904, 59225, 21686], with semantic similarity scores ranging from 1.00 (exact match) down to 0.937.

When these SBERT-retrieved patients were compared using 3-gram Jaccard similarity on raw ICD-10 diagnosis codes, the top match (Patient 100) showed perfect overlap (Jaccard = 1.00), confirming exact diagnostic replication. The subsequently retrieved patients, however, exhibited moderate to high Jaccard scores (0.409–0.401), indicating meaningful but not exact diagnostic overlap. This demonstrates that SBERT captures clinical semantics beyond exact code matching—retrieving patients with similar clinical narratives even when ICD codes differ partially.

Qualitative inspection revealed that SBERT-retrieved patients shared similar drug regimens and procedures despite variations in specific ICD codes. For example, Patient 10014354 (query) and Patient 17925279 (retrieved) both presented with diabetes, hypertension, and renal complications, and were treated with insulin and metoprolol, despite differences in secondary diagnoses. This illustrates SBERT's ability to identify clinically analogous "digital twins" based on holistic patient profiles rather than discrete code sets.

Moreover, the outcome consistency was high: 4 of the top 5 retrieved patients shared the same survival outcome as the query patient. This confirms that SBERT-based retrieval preserves prognostic relevance, a critical requirement for evidence-based clinical decision support.

In summary, while diagnosis-based Jaccard similarity ensures code-level precision, SBERT embeddings enable broader semantic retrieval that aligns with clinical intuition, capturing patients with similar treatment patterns, comorbidities, and outcomes—even in the absence of identical diagnostic codes. This makes SBERT + FAISS a robust foundation for the retrieval-augmented module of AltRAG+.

*3) **Pathway Diversity and Outcome Consistency**:* Two metrics were visualized:

Pathway Diversity Index (PDI) = 0.789, which indicates rich variation in generated treatment pathways.

Outcome Equivalence Rate = 0.881; this confirms that multiple retrieved pathways achieve near-equivalent outcomes.

These visualizations demonstrate that AltRAG+ successfully generates multiple clinically valid, non-inferior treatment strategies, confirming H3.

VIII. Conclusion

AltRAG+ provides a holistic decision-support ecosystem that harnesses retrieval-augmented reasoning, graph-based modeling, causal inference, and temporal deep learning to facilitate clinical treatment planning in the ICU. The effectiveness of the K-shortest path method of identifying alternative paths indicates that traditional graph algorithms are still necessary to generate structured and interpretable predictions in complicated medical fields, particularly once targeting to supplement LLM-based models such as AltRAG+. By retrieving clinically similar patients at admission, generating many alternative predicted treatment pathways for counterfactual analysis based on benchmarking evidence at population-level and disease-specific levels, and continuously predicting individualized dynamic mortality trajectories during ICU stay for better prognostic stratification, the proposed framework provides higher personalization and explainability beyond traditional CDSS paradigms. The combination of both static and temporal information, along with causal methodology, allows clinicians to analyze the predicted impact of interventions from multiple analytical perspectives to make individualized patient-centric decisions. Prospective multi-center clinical studies are warranted to validate the utility of this novel framework. Real-time deployment in ICU settings, as well as the incorporation of other types of unstructured EHR data such as clinical narratives, will also be pursued.

References

[1] C.-C. Liu, H.-R. Yao, D.-C. Chang, and O. Frieder, "*TreatRAG: A Framework for Personalized Treatment Recommendation,*" in Proc. 19th ACM Conf. Recommender Systems (RecSys '25) Prague, Czech Republic, 2025, pp. 690–695.

[2] R. Yang, Y. Ning, E. Keppo, M. Liu, C. Hong, D. S. Bitterman, J. C. L. Ong, D. S. W. Ting, and N. Liu, "*Retrieval-augmented generation for generative artificial intelligence in health care,*" npj Health Systems, vol. 2, no. 1, p. 2, 2025.

[3] M. Mesinovic, P. Watkinson, and T. Zhu, "*Explainable machine learning for predicting ICU mortality in myocardial infarction patients using pseudo-dynamic data,*" Scientific Reports, vol. 15, p. 27887, 2025.

[4] J. Wu, J. Zhu, Y. Qi, J. Chen, M. Xu, F. Menolascina, and V. Grau, "*Medical Graph RAG: Towards safe medical large language models via graph retrieval-augmented generation,*" arXiv preprint arXiv:2408.04187, 2024.

[5] A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, L.-w. H. Lehman, L. A. Celi, and R. G. Mark, "*MIMIC-IV, a freely accessible electronic health record dataset,*" Scientific Data, vol. 10, no. 1, p. 1, 2023.

[6] T. Abdullahi, I. Gemou, N. V. Nayak, G. Murtaza, C. Eickhoff, S. H. Bach, and R. Singh, "*K-Paths: Reasoning over Graph Paths for Drug Repurposing and Drug Interaction Prediction,*" arXiv preprint arXiv:2502.13344, 2025.

[7] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun, "*GAMENet: Graph augmented memory networks for recommending medication combination,*" in Proc. 33rd AAAI Conf. Artificial Intelligence (AAAI 2019), pp. 1126–1133, 2019.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, pp. 4171–4186, 2019.

[9] N. Reimers and I. Gurevych, "*Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,*" in Proc. EMNLP, pp. 3982–3992, 2019.

[10] Q. Sun, A. Akman, and B. W. Schuller, "Explainable Artificial Intelligence for Medical Applications: A Review," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 2, Article 17, pp. 1–31, Feb. 2025. doi: 10.1145/3709367.

[11] S. Lechner and M. Hasani, "*Learning Long-Term Dependencies Irregularly-Sampled Time Series,*" in Proc. Int. Conf. Machine Learning (ICML), pp. 5719–5729, 2020.

[12] C. Esteban, O. Staeck, R. Baumbach, and V. Tresp, "*Predicting sequences of clinical events by using a personalized temporal latent embedding model,*" in Proc. IEEE Int. Conf. Healthcare Informatics (ICHI), pp. 93–101, 2016.

[13] I. Bica, A. Sharma, A. Kadir, and M. van der Schaar, "*Estimating counterfactual treatment outcomes over time through adversarially balanced representations,*" in Proc. 8th Int. Conf. Learning Representations (ICLR), 2020.

[14] A. Longa, V. Lachi, G. Santin, M. Bianchini, B. Lepri, P. Lio, F. Scarselli, and A. Passerini, "*Graph Neural Networks for Temporal Graphs: State of the Art, Open Challenges, and Opportunities,*" arXiv:2302.01018 [cs.LG], 2023.

[15] T. S. Hy, V. B. Nguyen, L. Tran-Thanh, and R. Kondor, "*Temporal Multiresolution Graph Neural Networks For Epidemic Prediction,*" in Proc. Int. Conf. Machine Learning (ICML) Workshop, 2022.

[16] Y. Xu, S. Xu, M. Ramprassad, A. Tumanov, and C. Zhang, "*TransEHR: Self-Supervised Transformer for Clinical Time Series Data,*" in Proc. of the 3rd Machine Learning for Health Symposium (PMLR, Vol. 225, pp. 623–635), 2023.

[17] Y. Zhang, R. Chen, J. Tang, W. Stewart, and J. Sun, "*LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity,*" in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD '17), 2017.