Relational Learning of Temporal and Semantic Clinical Structured EHRs for Enhanced Patient Outcome Forecasting

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Accurate prediction of clinical outcome requires models that can jointly reason over longitudinal diagnostic narratives and structured electronic health records (EHRs). Current multimodel approaches often treat patient data as static or rely on early/late feature concatenation, overlooking the temporal progression of clinical events and the semantic relationships between diagnostic findings. We propose TS-GNN, a lightweight Temporal-semantic Graph Neural Network that represents each hospital admission as a multimodal graph composed of sequential chest X-ray reports enriched with structured clinical features. Each node fuses hierarchical ClinicalBioBERT embeddings with a six-phase feature schema capturing demographics, acuity, surgical status, care intensity, interventions, and biological state. Temporal edges encode the chart-time progression of diagnostic events, while semantic edges connect clinically similar findings, enabling long-range evidence propagation. A multitask GNN architecture jointly predicts in-hospital survival, length of stay, and 30-day readmission. We evaluate on MIMIC-IV; TS-GNN demonstrates consistent improvements across hierarchical feature phases and temporal windows, outperforming strong tabular and sequential baseline models. Despite integrating rich multimodal context, the framework remains computationally lightweight with sparse graph construction and moderate parameterization. These results highlight the value of temporal-semantic graph reasoning for robust and scalable clinical decision support.

Keywords: Graph Neural Networks · Multi-modal Fusion · Clinical Outcome Prediction · MIMIC-IV · Radiology Report Analysis.

1 Introduction

Accurate prediction of clinical outcomes, such as in-hospital mortality, length of stay (LOS), and readmission risk, is fundamental to proactive clinical decision support and hospital resource optimization [6,9]. In practice, clinicians do not rely on isolated measurements; rather, they synthesize heterogeneous and temporally evolving information, including structured EHR data, laboratory trends, treatment intensity, and unstructured diagnostic narratives such as CXR reports and discharge summaries.

Despite the richness of large-scale EHR resources such as MIMIC-IV, current predictive models exhibit fundamental limitations. First, clinical data are irregular, sparsely paired, and temporally misaligned [2]. For example, MedFuse [5] combines numeric EHR data with images to predict in-hospital mortality. Multimodal components of datasets such as MIMIC-IV [9,6,7] and MIMIC-CXR [8] remain fragmented across modules, requiring extensive preprocessing and limited reproducibility [1,3]. Existing hierarchical models treat patient records as static bags of features even when graph-based depend on dense graph construction, limiting their ability to model long patient history as it evolves through serial examinations [4]. While TAMME [18] frequently relies on simple concatenation of features, which ignores the complex semantic relationship between radiologist reports and the underlying physiological state. For instance, a finding of pulmonary edema in a chest X-ray report carries different prognostic weight depending on the patient's acuity markers, such as ICU status or intravenous medication intensity.

To address these challenges, we propose **TS-GNN**, a **Multi-modal Temporal-Semantic Graph Neural Network** that represent each admission as a structured graph integrating longitudinal diagnostic text with hierarchical EHR features. In our formulation, serial Chest X-ray (CXR) reports form the nodes of the graph, each enriched with a multi-phase clinical feature representation capturing demographics, acuity, surgical complexity, care intensity, interventions, and biological state. Two complementary edge types define the relational structure: (i) temporal edges, which encode the chart-time progression of diagnostic events, and (ii) semantic similarity edges, which link clinically related findings via ClinicalBioBERT [11] embeddings. This dual formulation enables the model to follow both the temporal evolution of the patient and long-range diagnostic continuity, substantially improving the aggregation of pathophysiological evidence. Our design is further motivated by practical constraints in large-scale clinical modeling. Transformer-based multimodal architectures, while expressive, tend to be computationally heavy, requiring large parameter counts and significant GPU memory. By contrast, TS-GNN uses sparse graph construction, moderate hidden dimensionality, and a lightweight GNN backbone, making it substantially more efficient while retaining strong representational capacity.

The primary contributions of this work are as follows:

1. A novel Temporal-Semantic graph representation of hospital admissions that integrates longitudinal radiology reports with structured EHR context using a six-phase hierarchical clinical feature schema.
2. A lightweight multi-task GNN architecture that jointly predicts LOS, in-hospital survival, and readmission, enabling shared representation learning across clinically related outcomes.
3. We conduct a rigorous ablation study to evaluate the incremental value of clinical data phases, demonstrating that the fusion of longitudinal diagnostic text with biological state indicators significantly outperforms tabular-only baseline models.

2 Related Work

Clinical outcome prediction on the MIMIC-IV dataset has been widely explored across diverse modeling paradigms. Traditional approaches rely heavily on structured EHR features, using models such as decision trees and XGBoost for ICU survival prediction [13] and random forests and LSTMs for predicting length of stay and readmission [3]. More recent transformer-based work, such as medical transformers for ICU survival and hierarchical attention networks for readmission, demonstrates the values of sequence modeling but typically focuses on limited time windows or handcrafted clinical inputs [10]. Several methods incorporate the longitudinal nature of EHRs – BEHRT [12] and MOTOR [17] represent patient history as a sequence of timestamped categorical events. Some of encodes numerics using Fourier features and leverages pretrained encoders such as RAD-DINO [14] for images and BioLORD [15] for text. Multimodal fusion methods have attempted to bridge this gap. MedFuse [5] combines numerical EHR data with CXR images, while graph-based [4,16] hierarchical models treat patient data as static bags of features or rely on dense graph construction, limiting their ability to capture evolving clinical trajectories. Similarly, time-aware multimodal approaches such as TAMME incorporate multiple modalities but primarily fuse features at the token level without modeling semantic relationships across serial diagnostic events [18].

3 Methodology

Our framework utilizes a multi-modal approach, synchronizing unstructured clinical narratives with structured electronic health record (EHR) data from MIMIC-IV [9], MIMIC-IV-Note [7]. Fig. 1 We integrate three primary data streams from the MIMIC-IV database:

Radiology Reports (Diagnostic Evidence) Serial Chest X-ray (CXR) reports prvide temporally dense, physiologically grounded observation (e.g., edema, consolidation). These serve as the primary temporal markers of withing-stay progression.

Discharge Summaries (Post-Clinical Synthesis) High-level narrative summaries capturing admission context, clinical reasoning, and definitive diagnoses. We extract clinically relevant sections (chief complaint, history of present illness, past medical history, physical examination, family history, and social history).

Structured Clinical Data (Tabular Features) Fourteen tabular variables are grouped into six hierarchical phases-demographics, acuity, surgical complexity, care intensity, interventions, and biological state. Each feature is normalized and fused into node attributes. The feature hierarchy is detailed in Table 1.

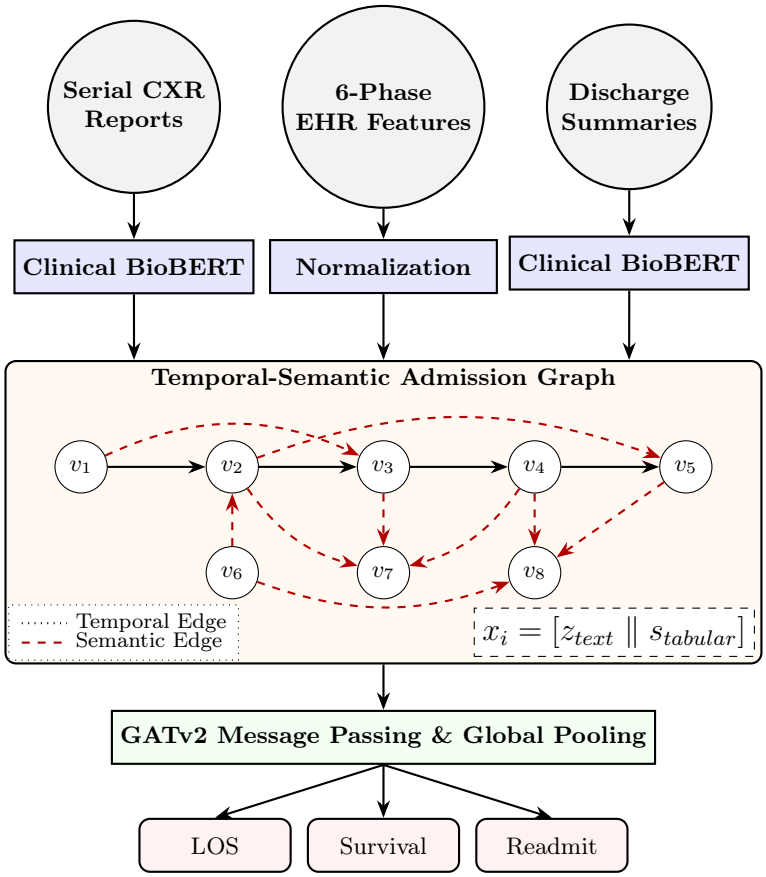

Fig. 1: TS-GNN Architecture: Longitudinal radiology and EHR fusion via a dense Temporal-Semantic graph.

3.1 Node Representation

Each node corresponds to a single CXR report or discharge summary. Given a raw text s, we segment it into tokenized chunks $c_1, c_2, \ldots, c_m$. Each chunk is encoded using ClinicalBioBERT: $h_j = \text{BioBERT}(c_j)$. Chunk embeddings are aggregated with lightweight multi-head attention: $z_i = \text{MHA}([h_1, h_2, \ldots, h_m])$. To contextualize diagnostic findings, we concatenate textual embeddings with the structured feature vector s $\mathbf{x_i} = [z_i \,|\, \mathbf{s}]$ This produces a unified representation capturing both semantic meaning and patient-specific physiological context.

3.2 Temporal-Semantic Graph Construction

The admission graph $G = (V, E)$ uses two complementary edge types: Temporal Edges and Semantic Similarity Edges.

$$(v_i, v_{i+1}) \in E \quad \text{iff} \quad t_i < t_{i+1}. \tag{1}$$

Directed edges connect reports in chart-time order as shown in eqn. 1. These edges preserve the evolution of the stay and encode patterns of deterioration, stabilization, or recovery. To link clinically related findings across distant time points, we compute cosine similarity between node embeddings and connect each node to its top-k nearest neighbors (here $k = 3$) given by eqn. 2

$$v_i \leftrightarrow \mathcal{N}_K(v_i), \tag{2}$$

These edges enable long-range aggregation of pathophysiologically similar evidence (e.g., repeated edema findings). Together, these edges form a Temporal-Semantic Admission Graph that captures both short-term chronology and long-range semantic continuity.

3.3 GNN Backbone and Multi-Task Prediction

We employ a lightweight GATv2-based GCN backbone with two message-passing layers (hidden dimension 256, four attention head). Node features are updated via attention-weighted neighborhood aggregation: $h_i^{(l+1)}$:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{c_{ij}} \mathbf{W}^{(l)} h_j^{(l)}\right) \tag{3}$$

A global mean pooling layer summarizes the admission graph:

$$\mathbf{z_{adm}} = \frac{1}{|V|} \sum_{v_i \in V} h_i^{(L)} \tag{4}$$

This embedding feeds three independent prediction heads: in-hospital survival, Length of stay, and 30-day readmission.

Optimization We use a weighted multi-task objective:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{Surv} + \beta \mathcal{L}_{LOS} + \gamma \mathcal{L}_{Readmit} \tag{5}$$

Class imbalance is handled using inverse-frequency weights:

$$w_c = \frac{N}{\text{count}(c)} \tag{6}$$

4 Experiments

This research utilizes the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, a large-scale de-identified EHR corpus. The study cohort is restricted to adult patients (Age $\geq$ 18) who underwent Chest X-ray (CXR) imaging during their hospitalization. After merging the tabular EHR data with

Table 1: Hierarchical Clinical Feature Phases for Ablation Study

Phase	Category	Features Included
Phase 1	Baseline Demographics	Age, Gender, Insurance Type, Admission Type
Phase 2	Acuity & Urgency	Phase 1 + Emergency Admission Flag
Phase 3	Surgical Complexity	Phase 2 + Surgical Admission Flag
Phase 4	Care Intensity	Phase 3 + ICU Admission History (*ever_icu*)
Phase 5	Clinical Intervention	Phase 4 + Total Med Admins, IV Count, STAT Meds, IV Ratio
Phase 6	Biological State	Phase 5 + Abnormal Lab Flags, Positive Blood Cultures

the *MIMIC-IV-Note* module, we filtered for admissions containing at least one radiology report and complete records for the 14 clinical features defined in our ablation phases. A key component of our experimental design is a six-phase incrementally incorporates higher-level clinical complexity. Table 1 summarizes this hierarchy. Phase 1 begins with foundational demographic and administrative attributes. Phase 2 to 4 progressively introduce indicators of acuity, surgical status, and care intensity, representing increasingly discriminative hospital-level descriptors. Phase 5 and 6 incorporate treatment-level and physiological variables, including medication burden, intravenous therapy patterns, abnormal laboratory flags, and microbiology findings-which provide the stongest signals of evolving severity.

The model is developed using the `PyTorch` and `PyTorch Geometric (PyG)` libraries. Textual embeddings are generated using the `ClinicalBioBERT` pretrained model. The GCN backbone consists of two GATv2 convolutional layers with a hidden dimensionality of 256 distributed across four attention heads. To ensure graph sparsity and focus on the most relevant semantic links, the K value for similarity edges is set to 3. We use the Adam optimizer with a learning rate of 10^{-4} and a dropout rate of 0.3. Training is performed with a batch size of 64 graphs. Given the memory requirements for processing BERT embeddings within a graph structure, all experiments were performed on an NVIDIA RTX A6000 GPU.

5 Results and Discussion

Table 2 summarizes the performance of our TS-GNN against the transformer-based TAMME model, MedFuse multi-model [5], and [3] model across varying temporal windows. Overall, our model demonstrates consistently strong behaviour across all tasks, with substantial improvements in Length of stay and In-hospital survival prediction.

For LOS classification, TS-GNN achieves markedly higher AUC across all temporal windows, showing that temporal-semantic graph modeling captures clinically important progression patterns that the baseline models fail to represent. Notably, even without historical context, LOS AUC improves from 0.83

to 0.95, indicating that serial radiology evidence fused with hierarchical tabular descriptors provides a richer representation than token-level fusion approaches. In-hospital survival exhibits the strongest gains in AUC across all historical windows. Although ACC remains close to the baseline models, the increased AUC demonstrates better discriminative ability, especially for high-risk patient subgrounds that are difficult for sequential models to identify. Short-range (7-day) context proves particularly beneficial, suggesting that recent diagnostic evolution carries the most prognostic value. Readmission prediction remains a challenging task for all models. While the baseline models marginally surpasses TS-GNN in AUC, our method demonstrates competitive accuracy with smaller variance across history windows. This suggests that semantic connectivity between radiology reports provides useful auxiliary information, even though readmission is influenced by additional non-clinical factors not fully captured in MIMIC-IV.

Table 2: Classification performance comparison of the proposed TS-GNN with existing state-of-the-art methods across different clinical prediction tasks and patient history settings.

Target	**Model**	**No History**		**7 Days**		**6 Months**		**1 Year**	
		ACC	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**
Hospital LOS	susetzky et al.[18]	0.79	0.83	0.80	0.85	0.82	0.88	0.80	0.87
	Gupta et al. [3]	—	0.85	—	0.85	—	0.85	—	0.85
	Ours	**0.82**	**0.95**	**0.84**	**0.95**	**0.83**	**0.95**	**0.84**	**0.95**
In-Hosp Survival	susetzky et al. [18]	**0.85**	0.87	0.86	0.86	0.89	0.89	0.90	0.91
	Hayat et al. [5]	—	0.87	—	0.87	—	0.87	—	0.87
	Ours	0.83*	**0.94***	**0.88***	**0.95***	0.88	**0.95***	—	—
Hospital Readmit	susetzky et al. [18]	0.65	**0.73**	0.68	**0.77**	0.81	0.83	0.82	0.81
	Gupta et al. [3]	—	0.79	—	0.79	—	0.79	—	**0.79**
	Ours	**0.66**	0.72	**0.69**	0.72	0.64	0.71	0.66	0.72

*Note: * Indicates results without using discharge summary for the training.*

5.1 Ablation Study

We assess the contribution of hierarchical tabular features and multimodal inputs across six feature phases (Table 1) and multiple temporal windows. As shown in Figs. 2 and 3, performance generally improves as richer clinical features are added.

In the multimodal setting (radiology + discharge summary), LOS prediction improves progressively, reaching an AUC of **0.95** in Phases 5–6 after adding intervention and biological-state features. Survival prediction reaches an AUC of **0.99** with discharge summaries, while radiology-only performance improves from 0.87 to **0.94**, highlighting the value of hierarchical clinical features when textual

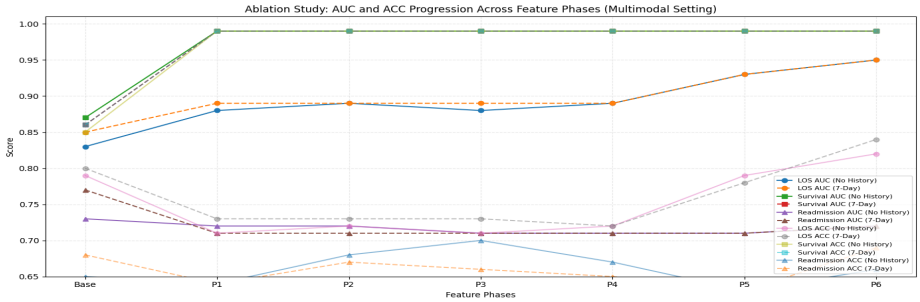

Fig. 2: ACC & AUC progression across hierarchical clinical feature phases for Length of Stay, In-Hospital Survival, and Readmission under No-History and 7-Day multimodal settings.

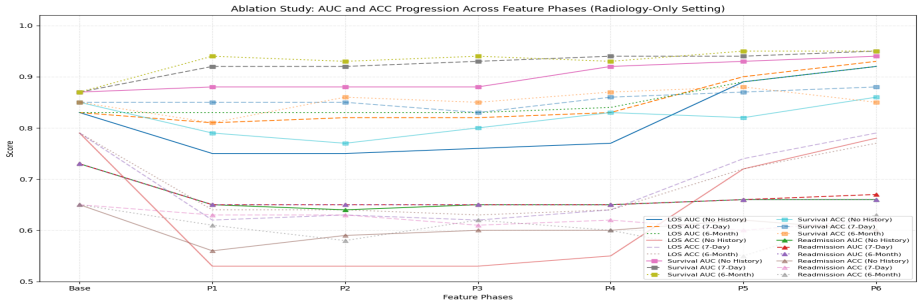

Fig. 3: ACC & AUC progression across hierarchical clinical feature phases for Length of Stay, In-Hospital Survival, and Readmission under No-History, 7-Day , and Six-month multimodal settings.

information is limited. Across temporal windows, the 7-day history provides the best overall performance, whereas longer windows benefit LOS and readmission but slightly reduce survival performance, likely due to less relevant older events.

6 Conclusion

We present TS-GNN, a scalable temporal-semantic graph framework for multimodal clinical outcome prediction in critical care. By integrating clinical text, hierarchical tabular features, and temporal information, TS-GNN captures longitudinal and semantic relationships in EHR data. Experiments on the MIMIC-IV dataset show consistent gains from multimodal fusion and structured clinical context across multiple prediction tasks. Unlike parameter-heavy transformer-based models, TS-GNN uses sparse graph construction and a lightweight architecture, enabling efficient training on a single GPU. Future work will explore additional modalities, improved text encoders, and validation across diverse clinical cohorts.

References

1. Adiba, F.I., Danduri, V., Piya, F.L., Abbasi, A., Gupta, M., Beheshti, R.: A multimodal data processing pipeline for mimic-iv dataset. arXiv preprint arXiv:2601.11606 (2026)
2. Bannur, S., Hyland, S., Liu, Q., Perez-Garcia, F., Ilse, M., Castro, D.C., Boecking, B., Sharma, H., Bouzid, K., Thieme, A., et al.: Learning to exploit temporal structure for biomedical vision-language processing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15016–15027 (2023)
3. Gupta, M., Gallamoza, B., Cutrona, N., Dhakal, P., Poulain, R., Beheshti, R.: An extensive data processing pipeline for mimic-iv. In: Machine learning for health. pp. 311–325. PMLR (2022)
4. Gupta, S., Sharma, S., Sharma, R., Chandra, J.: Healing with hierarchy: Hierarchical attention empowered graph neural networks for predictive analysis in medical data. Artificial Intelligence in Medicine **165**, 103134 (2025)
5. Hayat, N., Geras, K.J., Shamout, F.E.: Medfuse: Multi-modal fusion with clinical time-series data and chest x-ray images. In: Machine Learning for Healthcare Conference. pp. 479–503. PMLR (2022)
6. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L.A., Mark, R.: MIMIC-IV. PhysioNet (Oct 2024). `https://doi.org/10.13026/kpb9-mt58`, `https://doi.org/10.13026/kpb9-mt58`, version 3.1
7. Johnson, A., Pollard, T., Horng, S., Celi, L.A., Mark, R.: MIMIC-IV-Note: Deidentified free-text clinical notes. PhysioNet (Jan 2023). `https://doi.org/10.13026/1n74-ne17`, `https://doi.org/10.13026/1n74-ne17`, version 2.2
8. Johnson, A., Pollard, T., Mark, R., Berkowitz, S., Horng, S.: MIMIC-CXR Database. PhysioNet (Jul 2024). `https://doi.org/10.13026/4jqj-jw95`, `https://doi.org/10.13026/4jqj-jw95`, version 2.1.0
9. Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., et al.: Mimic-iv, a freely accessible electronic health record dataset. Scientific data **10**(1), 1 (2023)
10. Khader, F., Kather, J.N., Müller-Franzes, G., Wang, T., Han, T., Tayebi Arasteh, S., Hamesch, K., Bressem, K., Haarburger, C., Stegmaier, J., et al.: Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data. Scientific Reports **13**(1), 10666 (2023)
11. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.: Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**(4), 1234–1240 (2020)
12. Li, Y., Rao, S., Solares, J.R.A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., Salimi-Khorshidi, G.: Behrt: transformer for electronic health records. Scientific reports **10**(1), 7155 (2020)
13. Pang, K., Li, L., Ouyang, W., Liu, X., Tang, Y.: Establishment of icu mortality risk prediction models with machine learning algorithm using mimic-iv database. Diagnostics **12**(5), 1068 (2022)
14. Pérez-García, F., Sharma, H., Bond-Taylor, S., Bouzid, K., Salvatelli, V., Ilse, M., Bannur, S., Castro, D.C., Schwaighofer, A., Lungren, M.P., et al.: Exploring scalable medical image encoders beyond text supervision. Nature Machine Intelligence **7**(1), 119–130 (2025)
15. Remy, F., Demuynck, K., Demeester, T.: Biolord-2023: semantic textual representations fusing large language models and clinical knowledge graph insights. Journal of the American Medical Informatics Association **31**(9), 1844–1855 (2024)

16. Rocheteau, E., Tong, C., Veličković, P., Lane, N., Liò, P.: Predicting patient outcomes with graph representation learning. arXiv preprint arXiv:2101.03940 (2021)
17. Steinberg, E., Fries, J., Xu, Y., Shah, N.: Motor: A time-to-event foundation model for structured medical records. arXiv preprint arXiv:2301.03150 (2023)
18. Susetzky, T., Qiu, H., Braren, R., Rueckert, D.: A holistic time-aware classification model for multimodal longitudinal patient data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–33. Springer (2025)