

TSKG-ICU: temporal–semantic knowledge graph embeddings for ICU readmission prediction

Ricardo M. S. Carvalho · Andreia Reia Sofia Teixeira · Catia Pesquita

Received: 11 March 2026 / Accepted: 15 September 2026

© The Author(s) 2026

Abstract

Background Unplanned Intensive Care Unit (ICU) readmissions emerge from complex, time-dependent clinical processes that are poorly captured by existing electronic health records-based prediction models. Existing approaches capture either temporal patterns without understanding clinical meaning, or clinical relationships without preserving temporal context. This separation limits their ability to simultaneously model how clinical events relate to one another with how they evolve across an ICU stay. Bridging this gap requires models that integrate both temporal dynamics and semantic clinical knowledge within a unified framework.

Methods We introduce TSKG-ICU, a novel framework that jointly models temporal evolution and semantic relationships in ICU trajectories using temporal knowledge graph embeddings. Leveraging the MIMIC-III cohort, we constructed a temporal–semantic knowledge graph with over 11 million triples representing 60,000+ admissions and 200,000+ biomedical concepts. Clinical events are temporally encoded relative to ICU admission, preserving the full longitudinal structure of clinical records, and semantically annotated through ontology mapping. Patient trajectories are encoded as temporal knowledge graph embeddings and used as feature representations for supervised readmission prediction.

Results TSKG-ICU achieves state-of-the-art performance (ROC-AUC: 0.786 and PR-AUC: 0.641). Ablation studies reveal that temporal modelling provides the primary predictive signal, whilst semantic enrichment offers complementary improvements, an ordering consistent across all three classifiers evaluated. Removing temporal information displaces roughly an order of magnitude more predictions than removing semantic structure, which instead refines a representation temporal encoding has already organised, confirming that unified temporal–semantic representations shape the clinical decisions a deployed model would support. Architectural evaluation demonstrates that asymmetric, complex-valued embeddings with explicit temporal decomposition outperform both simpler and more recent alternatives.

Conclusion TSKG-ICU establishes temporal–semantic knowledge graph embeddings as a new paradigm for ICU readmission prediction. By leveraging complete ICU timelines and grounding clinical events in a shared semantic space, this approach provides a clinically meaningful foundation for identifying high-risk patients and supporting critical care decision-making.

Keywords Temporal semantic knowledge graphs · Temporal knowledge graph embeddings · ICU readmission prediction

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



Abbreviations

AI	Artificial Intelligence
DOD	Date of Death
DRON	The Drug Ontology
EHR	Electronic Health Record(s)
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
ICD-9	International Classification of Diseases, 9th Modification
ICU	Intensive Care Unit
KG	Knowledge Graph
KGE	Knowledge Graph Embedding
LOINC	Logical Observation Identifiers Names and Codes
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MRR	Mean Reciprocal Rank
NCIt	National Cancer Institute Thesaurus
NDC	National Drug Code
ODE	Ordinary Differential Equation
OWL	Web Ontology Language
PR-AUC/AUPRC	Area Under the Precision{Recall Curve
RF	Random Forest
ROC-AUC/AUROC	Area Under the Receiver Operating Characteristic Curve
RxNorm	Normalized naming system for clinical d
TKG	Temporal Knowledge Graph
TKGE	Temporal Knowledge Graph Embedding
TSKG	Temporal-Semantic Knowledge Graph
TSKGE	Temporal-Semantic Knowledge Graph Embedding
XGB/XGBoost	Extreme Gradient Boosting

Introduction

Artificial intelligence (AI) plays an increasingly important role in intensive care units (ICU), where timely and precise decision-making determines patient outcomes. However, most AI-based applications in critical care miss the opportunity to explicitly leverage the semantic relationships among complex physiological states, prescriptions, interventions, and care processes. Conversely, emerging approaches that incorporate rich clinical semantics into electronic health records (EHRs), such as those based on knowledge graphs (KGs), fail to capture the time-dependent nature of critical care. This disconnect reveals a fundamental gap: AI methods must be both semantic and time-aware, capable of reasoning about what clinical events mean and how they evolve. Unplanned ICU readmissions represent a particularly challenging prediction task within this context, arising from complex interactions amongst physiological, therapeutic, and procedural factors during a patient's stay, and closely associated with patient deterioration, increased mortality, prolonged hospitalisation, and greater financial costs to healthcare institutions [1, 2].

Risk however does not reside in any single measurement; it emerges from how events interact and unfold across a stay. Accurately predicting ICU readmissions therefore requires models that capture not only the timing of laboratory abnormalities, medication administrations, and procedural interventions, but also the interactions

and relations amongst these factors throughout a patient's stay. EHRs capture both, encoding clinical variables using standardised vocabularies such as ICD, LOINC, and RxNorm [3, 4]. Nonetheless, their structure prioritises documentation over reasoning, leaving important semantic relationships between clinical entities implicit. A diagnosis of acute myocardial infarction may be followed by aspirin prescription or coronary angiography, yet the EHR does not explicitly encode that aspirin is an antiplatelet therapy typically used to treat myocardial infarction, nor that angiography is a diagnostic procedure commonly performed in this context. Without these explicit links, models must infer relationships from data alone, limiting their ability to interpret patient trajectories.

Capturing temporal dynamics presents a parallel challenge. EHRs record time as absolute timestamps, either as single events or start–end intervals, and collect different data types at vastly different granularities: laboratory tests such as serum lactate every few hours, vital signs such as heart rate every minute, and medication administrations such as intravenous norepinephrine with precise start and end times. These heterogeneous temporal patterns make it difficult for conventional models to reconstruct the ordering, duration, and proximity of clinical events, and the difficulty compounds when trajectories of different lengths must be compared. Existing work has largely addressed these two problems separately: temporal representations capture how events unfold but treat clinical concepts as opaque entities, while semantic representations capture the relationships between concepts but describe a static snapshot.

We introduce TSKG-ICU, a framework that jointly models the timing and clinical meaning of events in ICU patient trajectories through temporal–semantic knowledge graph embeddings, enabling improved prediction of unplanned readmission. By encoding time relationally rather than sequentially, this design places all patient trajectories within a common representational space, allowing direct comparison across ICU stays with different lengths, event densities, and clinical profiles. This does not conflict with other temporal models, such as time-series approaches, but rather addresses a representational requirement they are not designed to meet: the simultaneous embedding of heterogeneous clinical events, ontological knowledge, and temporal relations within a unified relational space. Our contributions are: (i) a methodology for constructing temporal–semantic clinical KGs from ICU-specific EHR data; (ii) an evaluation of TKGE methods for ICU readmission prediction; and (iii) an in-depth analysis of the distinct and combined effects of temporal and semantic structure on classification performance. With this approach, TSKG-ICU matches or surpasses reported state-of-the-art ICU readmission prediction performance whilst avoiding temporal bias and leveraging clinical semantic knowledge, making it more suitable for prospective clinical deployment.

Related work

The application of machine learning to electronic health records (EHR) spans risk stratification, diagnosis support, and outcome prediction across critical care settings [3, 4]. Within this domain, predicting unplanned ICU readmission is widely recognized as vital for patient care and resource management [5, 6]. Unplanned readmission is typically framed as binary classification over a post-discharge window most commonly 30 days as in our work [7–9], though shorter or alternative horizons like 72-hour are also explored [10].

Existing readmission models differ particularly in how they handle time. A first line of work restricts model inputs to fixed pre-discharge windows, such as the final 24 to 48 hours of a stay, using conventional classifiers or LSTMs [7, 10]. While computationally convenient, this approach discards early clinical trajectory and fails across variable lengths of stay. A second line models full patient trajectories using sequence-aware architectures such as GRU encoders, multi-task models, attention mechanisms, or neural ODEs over multimodal data [9, 11–13]. While these models demonstrate that sequence structure improves prediction, they share a key limitation: clinical events are treated as isolated sequences rather than semantically related entities. Relations between diagnoses, procedures, and prescriptions are not explicitly represented and must instead be inferred purely from co-occurrence.

Knowledge Graphs (KGs) and KG embeddings address this representational gap by organizing entities into structured networks governed by biomedical ontologies [14–17]. In healthcare, KGs have been developed and applied for tasks such as cohort discovery, patient trajectory modelling, or readmission analysis [18–20]. However, clinical KGs are almost exclusively static; they provide a semantic scaffold of what occurred without representing when events happened, omitting the temporal signals crucial for clinical risk prediction.

Temporal Knowledge Graphs (TKGs) offer a principled route to unifying both paradigms by attaching validity times or intervals to each edge [21–23]. This structure maps naturally onto EHR data, where clinical events occur either at discrete time points (a laboratory test, a procedure) or across extended intervals (a course of antibiotics, ventilator use), enabling the integration of biomedical knowledge with dynamic patterns of care [24]. Consequently, Temporal Knowledge Graph Embedding (TKGE) methods learn representations that jointly encode relational structure and event timing, rather than treating time or semantics as independent annotations [25].

Several TKGE architectures descend directly from static formulations, extending an existing scoring function with a temporal component. Translational models extend TransE by treating time either as a relational shift vector (T-TransE) [26] or as a sequence processed by a recurrent encoder (TA-TransE) [27]; bilinear models apply similar recurrent extensions to DistMult (TA-DistMult) [27]; and tensor factorization models treat time as a fourth decomposition mode over quadruples, with T-ComplEx modulating relations by a temporal factor and TNT-ComplEx separating static from time-varying components [28]. Recent refinements further expand these lineages: TeAST [29] maps relations onto an Archimedean spiral timeline to reduce temporal completion to lower-order tensor factorization while preserving relational properties, whereas TCompoundE [30] composes relation- and time-specific translation and scaling operations. We adopt these scoring-function-based architectures in TSKG-ICU because each preserves a direct correspondence with a non-temporal counterpart, allowing the empirical contribution of the temporal component itself to be isolated. Beyond these lineages the field has recently diversified considerably: the broader TKGE field spans GNN and autoregressive snapshot models, continuous-time temporal point processes, hypercomplex and non-Euclidean geometries [31], as well as inductive methods designed for unseen entities [32, 33]. These methods are, however, developed and benchmarked almost entirely on historical and geopolitical event corpora, where time is absolute and the graph is open-world and incomplete, assumptions that do not transfer to clinical data.

Taken together, existing literature leaves a clear gap on both sides: ICU readmission models capture temporal dynamics but ignore clinical semantics; clinical KGs capture semantics but omit temporal evolution; and TKGE architectures, which could in principle unify the two, rest on temporal assumptions incompatible with ICU data. TSKG-ICU addresses this gap by encoding ICU events relative to admission within an ontology-grounded temporal knowledge graph, learning representations in which the timing and clinical meaning of events are optimized jointly rather than modelled separately.

Methods

In TSKG-ICU we formulate ICU readmission prediction as binary classification at the ICU stay level, where the task is to predict 30-day unplanned readmission following discharge. TSKG-ICU achieves this through three methodological stages (Overview on Fig. 1). First, we encode all clinical events from each ICU stay as a temporal–semantic knowledge graph, anchoring events to relative time (hours since admission) whilst linking them to biomedical ontology concepts. This creates a shared representational space where temporal position within the ICU timeline becomes directly comparable regardless of total length of stay. Second, we apply state-of-the-art TKGEs to learn patient representations that capture latent clinical and temporal dependencies. Third, we use these embeddings combined with demographic features as input to supervised learning models for ICU readmission prediction.

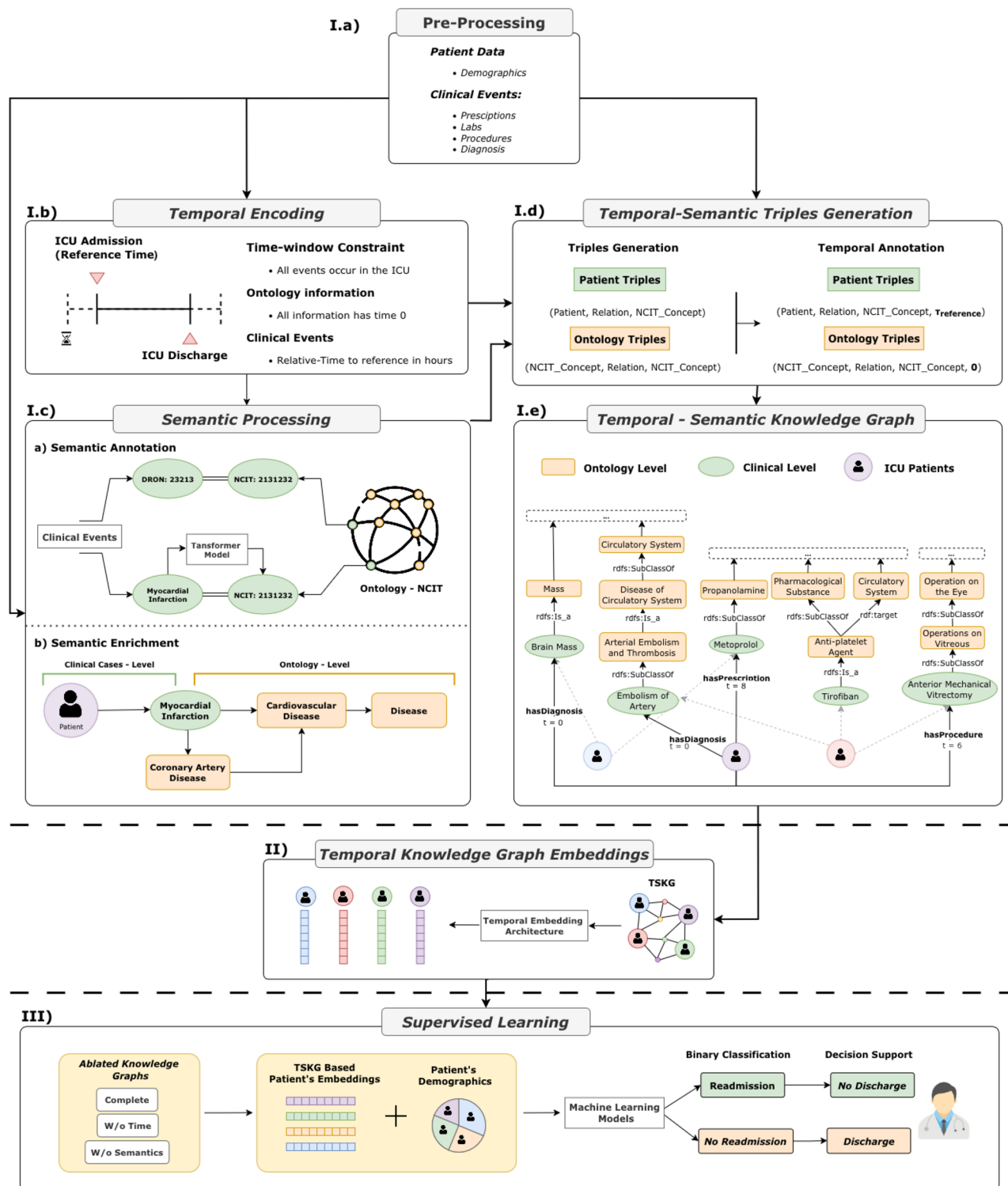


Fig. 1 Overview of the temporal-semantic modelling workflow for ICU readmission prediction. The framework comprises: (1) construction of a temporal-semantic knowledge graph through data extraction, semantic annotation and enrichment, temporal encoding, and temporal-semantic triple generation; (2) learning temporal knowledge-graph embeddings that capture latent temporal-semantic structure in patient trajectories; and (3) supervised prediction using patient embeddings combined with demographic features

Data and pre-processing

Data source and extraction

We employed the MIMIC-III database [5], an open-access collection of de-identified EHR data covering more than 60,000 admissions at the Beth Israel Deaconess Medical Center between 2001 and 2012. MIMIC-III provides detailed, multimodal clinical information suitable for constructing TSKGs. It includes structured coded diagnoses and procedures (International Classification of Diseases 9th Modification (ICD-9)), prescriptions (The Drug Ontology (DRON)), and laboratory tests (coded with Logical Observation Identifiers Names and Codes (LOINC)), as well as unstructured free-text notes of admission diagnosis, linked by patient and admission identifiers.

From the database, we extracted information from five main domains relevant to ICU trajectories. Structured coded events included: (i) diagnoses (`DIAGNOSES_ICD`, *ICD-9*), (ii) procedures (`PROCEDURES_ICD`, *ICD-9*), (iii) prescriptions (`PRESCRIPTIONS`, *NDC*), and (iv) laboratory tests (`LABEVENTS`, *LOINC*). In addition, unstructured free-text admission diagnoses (v) were extracted from the `ADMISSIONS` table and incorporated after standardization and annotation. Each clinical event was associated with a timestamp, with prescription records modelled as time intervals using their start and end times. Notably, the temporal resolution in MIMIC-III frequently reaches the minute level, contrasting with the day- or year-scale timestamps commonly assumed in temporal graph models [26–28]. Additional temporal anchors, including ICU admission (`Admission_Time`), discharge (`Discharge_Time`), and date of death (`DOD`), together with demographic descriptors such as age, biological sex, ethnicity, and insurance type, were extracted from the `ADMISSIONS` table to complement graph-based features.

Cohort definition and exclusion criteria

Following established approaches [7–9], we evaluate TSKG-ICU on a large adult ICU cohort obtained from the MIMIC-III database, focusing on the prediction of unplanned ICU readmission in accordance with our problem formulation. ICU stays were the unit of analysis and were treated as bounded episodes defined by admission (`Admission_Time`) and discharge (`Discharge_Time`) timestamps. Each stay was labelled as positive if followed within 30 days by unplanned ICU readmission, post-discharge death, or transfer to another ICU under equivalent criteria, consistent with established clinical definitions. All remaining stays were treated as negative instances. Importantly, because MIMIC-III is a single-center database, the negative class likely includes unobserved positives. Readmissions occurring at other hospitals are not captured in the database, resulting in a form of outcome under-ascertainment common to single-centre EHR studies. While this bias cannot be quantified directly from the data itself as out-of-network readmissions are, by definition, unobserved, and MIMIC-III's de-identification prevent access to patient's information, it is a shared limitation with prior work using the same cohort and outcome definition [7–9].

We excluded pediatric stays (under 18 years of age) since patients are managed under distinct care protocols. Additionally, stays ending in in-ICU death were excluded as individuals who die during their ICU stay are, by definition, not eligible for readmission. For records with missing death dates (`DOD`), survival was assumed unless readmission criteria were met. Following previously established guidelines [23], clinical events were strictly restricted to the window between ICU `Admission_Time` and ICU `Discharge_Time`, ensuring that the graph captures solely the clinically relevant, temporally bounded events associated with each stay.

After applying these criteria, the final cohort comprises 53,330 ICU stays, including 13,605 readmissions (25.5%) and 39,725 non-readmissions (74.5%). This level of class imbalance is consistent with prior ICU readmission studies [7, 9, 34–37] and reflects real-world clinical populations, presenting a challenging and clinically realistic setting.

Data preprocessing and cleaning

Although the data includes both structured terminological codes and unstructured clinical text, raw EHR entries are subject to inconsistencies, formatting errors, and variability in clinical documentation. To mitigate these issues prior to TSKG construction, we applied targeted preprocessing and normalisation steps. For structured data, empty entries were removed, and malformed codes were corrected to comply with terminology specifications. For free-text fields, non-informative or malformed entries (e.g., short, non-informative symbols) were excluded, whilst incomplete but clinically meaningful text was retained. When multiple clinical concepts were recorded within a single entry, they were separated to allow independent annotation.

Initially, the cohort included 31,979,551 ICU events across diagnoses, prescriptions, laboratory results, and procedures. After applying all exclusion criteria and preprocessing steps, the volume was refined to 3,110,254 events distributed across five entity types: 84,746 initial diagnoses, 1,098,849 drug prescriptions, 1,868,665 laboratory events, 44,782 final diagnoses, and 13,212 procedures. Following temporal annotation (described in Section “[Temporal annotation](#)”), the total number of time-stamped triples increased to 11,207,832. Table 1 summarises the entity distribution across sequential data preparation stages. These normalisation steps ensure consistent alignment between raw EHR data and ontology concepts, providing the necessary foundation for TSKG construction and embedding learning.

Temporal-semantic KG construction

Our TSKG-ICU framework builds on and extends a previously proposed methodology to represent patient trajectories as TKGs [23]. Whilst the original framework, applied to general hospital admissions, supported multiple KG variants, we adapt it here to construct a unified temporal–semantic graph for ICU data.

Semantic annotation

Semantic annotation serves as the foundation for constructing the TSKG. All diagnoses, procedures, laboratory tests, and prescriptions are mapped to National Cancer Institute Thesaurus (NCIt) concepts, which provide broad biomedical coverage and a coherent semantic space for heterogeneous EHR data.

For coded entries (ICD-9 diagnoses and procedures, LOINC laboratory tests, and NDC drug codes), annotation is performed using a string-based matching approach between clinical code descriptions provided in MIMIC-III and NCIt concept labels and synonyms. This process relies on lexical normalisation (e.g., case folding and punctuation removal) and exact or high-confidence partial matches, quantified by a normalized similarity score

$$\text{sim}(s_1, s_2) = \frac{2M(s_1, s_2)}{|s_1| + |s_2|} \quad (1)$$

Table 1 Entity distribution across sequential data preparation stages used in temporal–semantic graph construction

Entity Type	Clean MIMIC-III	ICU Window	Temporally Annotated
Initial Diagnoses	84,746	84,746	84,740
Drug Prescriptions	4,151,991	1,098,849	1,672,073
Laboratory Events	26,851,670	1,868,665	9,393,025
Final Diagnoses	651,048	44,782	44,782
Procedures	240,096	13,212	13,212
Total	31,979,551	3,110,254	11,207,832

The clean MIMIC-III dataset includes all processed clinical events, ICU window restricts these to the time span of ICU stays, and temporally annotated corresponds to processed temporal annotations

where $M(s_1, s_2)$ is the length of the longest matching subsequence between strings s_1 and s_2 , and matches are accepted when $\text{sim}(s_1, s_2) > \tau$ (with $\tau=0.75$ in our implementation). Because these entries already rely on standardised vocabularies, direct label matching provides an efficient and reliable mechanism for alignment [38].

For the semantic annotation of free-text elements, such as admission diagnoses, we employ a language model-based concept matching. We use a transformer matching approach, which provides robust coverage in the presence of variable clinical language and achieves high performance [39, 40]. Each free-text entry t is encoded using the pre-trained sentence transformer `multi-qa-mpnet-base-dot-v1` (fine-tuned version is available at [HuggingFace](#)), and cosine similarity is computed against similar representations $f(c)$ of NCIt concepts c . Let $f(\cdot)$ denote the embedding function; we compute

$$\text{sim}(t, c) = \frac{f(t) \cdot f(c)}{\|f(t)\| \|f(c)\|} \quad (2)$$

and select the most similar concept

$$c^*(t) = \arg \max_{c \in \mathcal{C}} \text{sim}(t, c) \quad (3)$$

subject to $\text{sim}(t, c^*(t)) \geq \tau$, where τ is a predefined similarity threshold. The most similar concept exceeding this threshold is selected as the mapped ontology class. This process captures semantic similarity beyond surface-level string overlap, resulting in improved coverage of clinically relevant concepts (83.5% of NCIt concepts can be mapped with lexical similarity, whereas the transformer-based approach increases this to 96.2%).

Ontology Layer integration

Ontology classes and their hierarchical relations (*subClassOf*, *equivalentClass*) are incorporated into the graph following the semantic enrichment procedure described in the base method [23]. However, while the original framework implicitly linked ontology entities to patient-level data, it did not explicitly formalise how these two layers interact. In TSKG-ICU, we make this connection explicit by adopting OWL punning [41], allowing ontology classes to function simultaneously as categories and as graph individuals. This allows the same node to participate in patient-level assertions and ontology-level relations, ensuring a coherent alignment between the semantic information and the clinical facts represented in the graph. This clarification preserves the semantic depth of the original approach while establishing a more explicit and consistent integration of ontology knowledge with patient trajectories.

Triples generation

In TSKG-ICU, triple generation is explicitly restricted to the ICU stay window as stated in Section “[Data preprocessing and cleaning](#)”. Under this constraint, patient-level triples (t_{patient}) take the general form:

$$t_{\text{patient}} = (s, r, c)$$

where s is the patient’s ICU stay, r is a relation (*hasInitialDiagnosis*, *hasFinalDiagnosis*, *hasProcedure*, *hasLab* or *hasPrescription*) and c is the NCIt concept describing an event during the stay. These triples collectively encode the diagnostic, procedural, laboratory, and pharmacological events that define the clinical course of each ICU episode. In addition to patient-level triples, the graph includes ontology-level triples (t_{ontology}) derived from NCIt, which represent the hierarchical structure of biomedical concepts. These take the form:

$$t_{\text{ontology}} = (c_i, \text{relation}, c_j)$$

where *relation* corresponds to either *subClassOf* or *equivalentClass* and c_i and c_j correspond to NCIt classes. Together, the patient-level and ontology-level triples provide the factual backbone of the temporal–semantic graph prior to temporal annotation.

Temporal annotation

A key innovation is the introduction of a uniform temporal encoding strategy tailored to Temporal-Semantic Knowledge Graph Embedding (TSKGE) models. This is essential for supporting the direct comparison of patient trajectories. Without temporal encoding, embeddings either lose temporal information entirely or anchor to absolute time, preventing the identification of similar patterns across patients with different temporal trajectories. All event timestamps are normalised relative to a reference time τ_{ref} , which corresponds to the admission time (Admission_Time = 0) and is discretised into hourly bins. While this discretization inevitably discards exact event ordering and co-occurrence patterns within a single hour, it balances temporal expressiveness and computational efficiency. Interval events are represented by their start times, while ontology-level edges receive time 0 to reflect their static nature. This temporal layer is a novel contribution specifically designed to support downstream time-aware embedding architectures, ensuring that temporal ordering and event proximity are consistently represented across patients. The final formulation of the triples is as follows:

$$t_{\text{patient}} = (s, r, c, \tau_{\text{ref}}) \mid t_{\text{ontology}} = (c_i, r, c_j, 0) \quad (4)$$

In Fig. 2 we illustrate how a patient’s trajectory with nine events is normalised and grouped into distinct hourly timestamps. Most importantly, the figure demonstrates how temporal decomposition captures repeated events

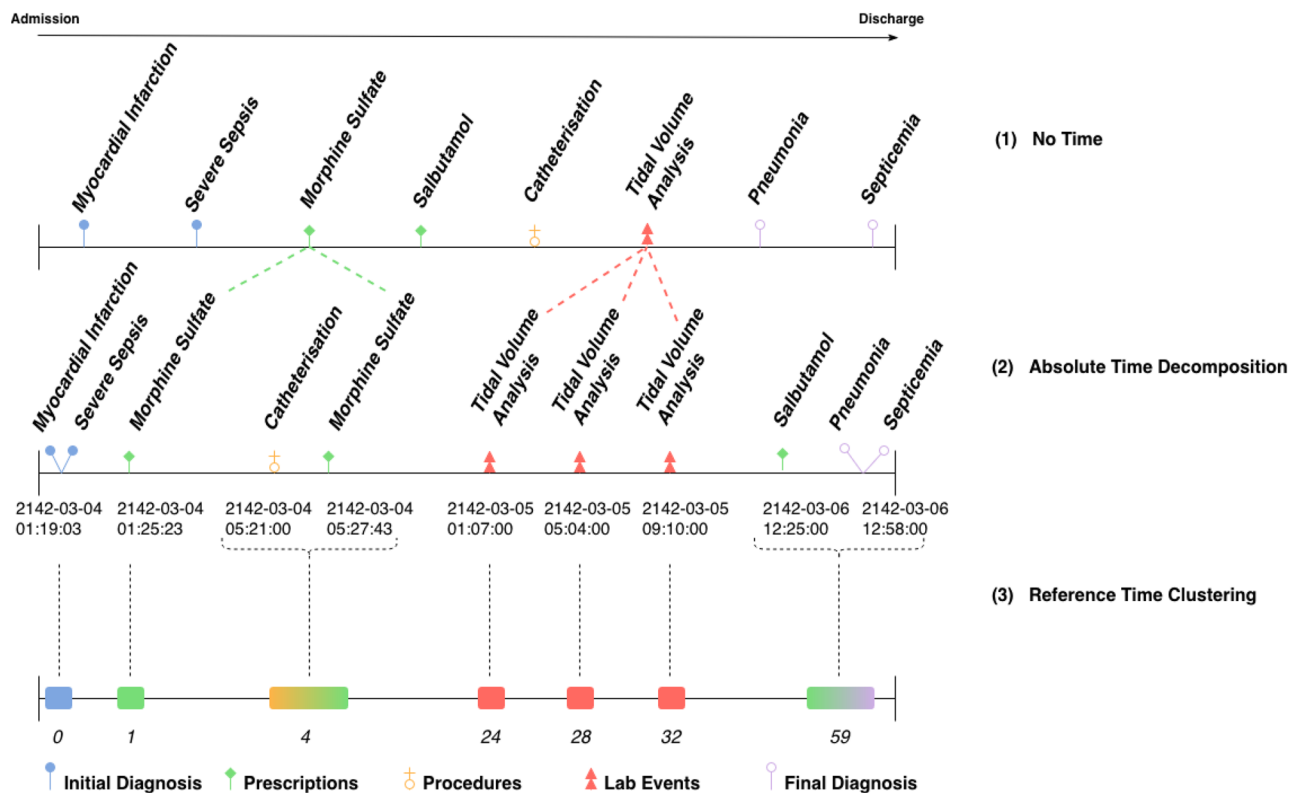


Fig. 2 Temporal encoding strategies for event annotation. Comparison of three representations for clinical event timing: (1) without temporal information, (2) original timestamps, and (3) relative timestamps. The figure illustrates how temporal encoding influences event granularity, graph size, and trajectory understanding within the TKG

(e.g., multiple administrations of salbutamol) that would otherwise collapse into a single patient event without temporal information.

Final TSKG characteristics

The complete TSKG-ICU construction process yielded a temporal–semantic knowledge graph with over 11 million triples representing more than 60,000 admissions and over 200,000 biomedical concepts extracted from the NCIIt ontology. Each triple corresponds either to a patient event (e.g., *patient - has_Diagnosis→Myocardial Infarction*) or to an ontology-level statement (e.g., *Myocardial Infarction - subclassOf→Cardiovascular Disease*). Semantic annotation successfully connected 3,015,508 coded entities directly mapped from their original MIMIC vocabularies to NCIIt (e.g., *Metoprolol | DRON: 74,277,825→NCIt: C61845*), whilst the remaining 84,746 entities from clinical notes were annotated using language model-based concept matching (e.g., *"Acute Myocardial Infarction"→NCIt: C35204*). Semantic enrichment expanded the KG beyond direct patient annotations by integrating the full NCIIt biomedical ontology. As shown in Table 2, this enrichment introduced 232,157 new ontology nodes and extended the total number of relation types from 5 to 7, connecting patient-level entities to their biomedical contexts through relations such as *subclassOf* and *equivalentClass*. Figure 1 (I.c) illustrates how this enrichment provides conceptual organization, enabling the simultaneous representation of factual patient-level events and higher-level semantic relationships between clinical concepts.

Temporal annotation, using relative time encoding with hourly bins referenced to ICU admission, increased patient events from 3,110,254 to 11,434,158 time-stamped triples. The resulting TSKG unifies ICU patients within a shared representational space, enabling comparison across both semantic and temporal dimensions. In Fig. 3 we present a representative sub-graph of the TSKG, illustrating how patient stays, ontology structure, and relative temporal annotations are jointly encoded within the unified temporal–semantic framework.

Embedding learning

The next stage of TSKG-ICU consists in learning low-dimensional vector representations that encode the temporal ordering and semantic structure captured in the TSKG. TKGE models are trained to reconstruct the time-stamped, ontology-enriched triples, projecting entities and relations into a continuous space where both *what* occurred and *when* it occurred are jointly preserved.

We place particular emphasis on the ComplEx family of embeddings [28, 42], which is well-suited to clinical KGs due to its ability to model asymmetric and multi-relational patterns using complex-valued vector representations. These properties make ComplEx and its temporal extensions especially appropriate for ICU data, where clinical events involve directional and non-commutative relations (e.g., medications administered after a diagnosis, procedures performed following a laboratory abnormality). Temporal variants such as T-ComplEx and, especially, TNT-ComplEx extend this formulation through temporal tensor factorization and time-dependent regularization, enabling the model to capture evolving relational structure over the course of an ICU stay.

Table 2 Structural characteristics of the temporal-semantic knowledge graph

	Before Semantic Enrichment	After Semantic Enrichment
Patient Admissions	66,540	66,540
Ontology Classes	0	232,157
Properties	5	7
Triples w/o time*	3,110,254	3,342,411
Triples w/ time*	11,207,832	11,434,158

The table reports the number of entities, properties, and triples in the temporal KG prior to adding ontology classes and semantic relations, and after full temporal–semantic construction. *The total triple count is presented both with and without temporal information

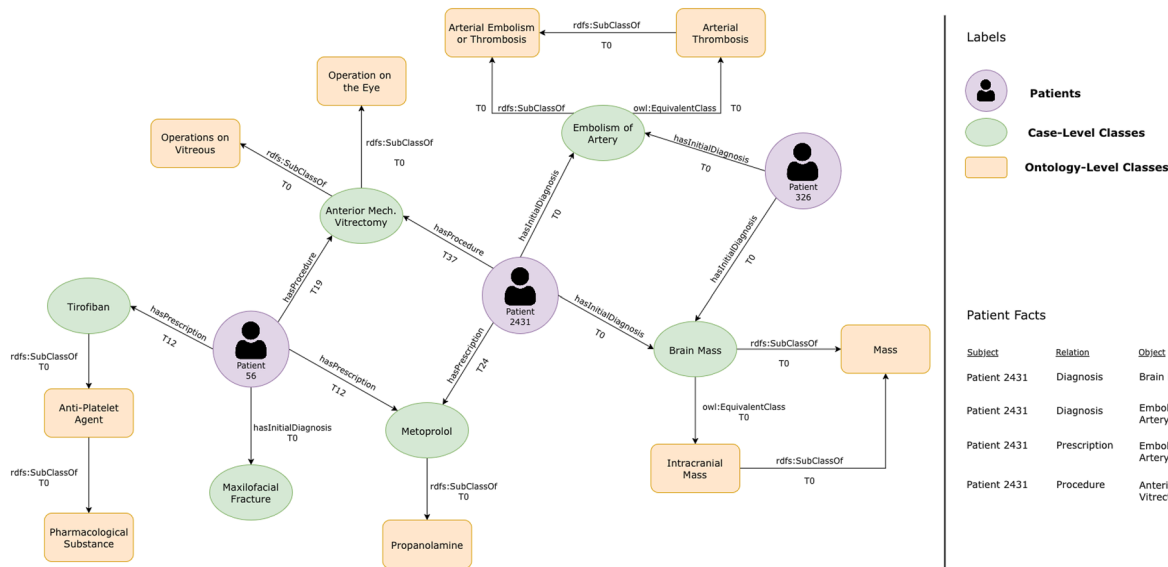


Fig. 3 Example sub-graph from the TSKG. Each green node represents an individual ICU stay linked to ontology concepts (purple) through clinically defined relations. Higher-level ontology connections (orange) arise from semantic enrichment, while temporal annotations are expressed as relative hours since admission, illustrating conversion from absolute timestamps to a unified temporal reference frame

Complex [42] models static KGs by subjects, relations and objects in a complex-valued vector space, enabling the representation of asymmetric relations, where the subject and object factor vectors can share the parameters of values but are complex conjugates of each other. For a triple (s, r, o) , the scoring function is defined as:

$$f_{\text{Complex}}(s, r, o) = \text{Re}(\mathbf{s}, \mathbf{r}, \bar{\mathbf{o}}) \quad (5)$$

where $\mathbf{s}, \mathbf{r}, \mathbf{o} \in \mathbb{C}^d$ are the subject, relation, and object, respectively, $\bar{\mathbf{o}}$ denotes the complex conjugate of the concept embedding, and $\text{Re}(\cdot)$ extracts the real part. This formulation captures rich semantic structure and relational asymmetry but assumes all facts are time-invariant.

T-Complex [28] extends Complex by introducing time as an additional multiplicative factor in a four-way tensor factorization. In this setting, each temporal fact (s, r, c, τ) corresponds to an entry in a fourth-order tensor indexed by subject, relation, object, and time, respectively. The scoring function is defined as:

$$f_{\text{TComplex}}(s, r, o, \tau) = \text{Re}(\langle \mathbf{s}, \mathbf{r}, \bar{\mathbf{o}}, \mathbf{t}_\tau \rangle) \quad (6)$$

where $\mathbf{s}, \mathbf{r}, \mathbf{o}, \mathbf{t}_\tau \in \mathbb{C}^d$ and $\mathbf{t}_\tau$ is the embedding associated with time τ . In this formulation, time is treated as a fourth tensor factor that modulates relational interactions, allowing relations to vary across discrete timestamps while remaining embedded in a shared temporal space.

TNT-Complex [28] further generalises the factorization by decomposing the temporal tensor into two coupled components, enabling selective temporal modelling in heterogeneous KGs where only a subset of relations evolves over time. Each relation r is factorised into a static component $\mathbf{r}$ and a temporal component $\mathbf{r}^t$. The resulting scoring function is:

$$f_{\text{TNTComplex}}(s, r, o, \tau) = \text{Re}(\langle \mathbf{s}, \mathbf{r}^t \odot \mathbf{t}_\tau + \mathbf{r}, \bar{\mathbf{o}} \rangle) \quad (7)$$

where $\odot$ denotes element-wise multiplication. This formulation can be interpreted as the sum of the two tensor factorizations: a static semantic tensor and a temporal interaction tensor. By separating persistent relational

semantics from time-dependent effects, TNT-ComplEx supports expressive modelling of irregular, overlapping, and long-range temporal patterns without imposing temporal dynamics on relations that do not require them.

Model training setup

For contextual comparison, we additionally trained representative models from translational (TransE, T-TransE, TA-TransE) [26, 27, 43] and bilinear (DistMult, TA-DistMult) [27, 44] families. These serve as contrastive baselines, allowing us to assess how architectural assumptions (translation-based, symmetric, or asymmetric) influence temporal–semantic patient modelling.

To ensure this comparison also reflects recent advances in TKGE research, we additionally included two more recent models. TCompoundE [30], which extends the translational paradigm; and TeAST [29], which works with tensor-factorization. All models have publicly available implementations at [OpenKE-GitHub](#), [RENet-GitHub](#), [Tkbc-GitHub](#), [TCompoundE-GitHub](#) and [TeAST-GitHub](#).

All embedding models were trained with a shared embedding dimension ($d = 300$), following readmission prediction KGE literature [7, 8], alongside a fixed batch size ($b = 512$) and maximum epoch budget ($E = 750$), with early stopping on training loss stabilization). Fixing these core capacity-related hyperparameters isolates architectural and training-objective differences from confounding effects like model capacity or training duration. Crucially, maintaining a consistent embedding dimension (d) enables a direct comparison with prior works using the same baseline for downstream prediction. Models belonging to the same base KGE family were trained with an identical optimizer, learning rate, and loss function, following standard reimplementation practice, so that performance differences primarily reflect their temporal modelling choices rather than unequal optimisation settings or model-specific tuning. The full per-model training configuration, including optimizer, learning rate, loss function, and regularisation scheme for every evaluated architecture, is reported in Additional file 1 : Table S1.

Supervised learning

In TSKG-ICU, ICU readmission prediction is a binary node classification problem, consistent with established clinical criteria [7–9]. As mentioned before, each node corresponds to an individual adult ICU stay, and the label indicates whether the patient experiences an unplanned readmission within 30 days of discharge. Positive cases correspond to ICU stays followed within 30 days by unplanned ICU readmission, post-discharge death, or transfer to another ICU under equivalent conditions, all in the MIMIC-III network. All remaining stays are treated as negative instances.

For all stays, demographic information (age, sex, ethnicity, insurance type) was extracted as features for downstream machine learning models. The final representation of each ICU stay combines demographic descriptors with the 300-dimensional embedding produced by the TKGE model. Age was retained as a continuous variable, while biological sex, ethnicity, and insurance type were expanded into one-hot representations (48 binary indicators in total). When concatenated with the clinical embedding, this yielded a 349-dimensional feature vector for each stay.

To assess the predictive value of the learned representations, we trained three supervised learning models: Logistic Regression (LR) [45], Random Forests (RF) [46], and XGBoost (XGB) [47]. These models were selected because they provide a robust and interpretable evaluation of how well the TSKGEs support downstream prediction. Moreover, all three classifiers are widely adopted in ICU readmission and clinical risk prediction studies [7, 10, 48–51], allowing the comparison with prior work. For cross-validation, the dataset was partitioned into stratified 10-fold splits to support the training and evaluation of the supervised models, ensuring that patients never appear simultaneously in training and test sets to prevent leakage. Classifier hyper-parameters were selected via an inner 5-fold validation to ensure consistent optimization across models, and performance was evaluated on the corresponding held-out test split.

Model performance was evaluated using six complementary metrics: accuracy, area under the receiver operating characteristic curve (ROC-AUC), area under the precision–recall curve (PR-AUC), precision, sensitivity (recall), specificity and F1 score. Because standard accuracy can be misleading in imbalanced clinical settings, ROC-AUC and PR-AUC serve as primary solutions. ROC-AUC quantifies the ability to rank patients by relative risk across all classification thresholds, a critical property in imbalanced and cost-sensitive clinical settings. PR-AUC offers a complementary perspective by focusing specifically on positive class performance. Precision and sensitivity further disentangle predictive behaviour, particularly on clinical settings, by measuring true positive correctness and retrieval coverage, respectively, with F1 score summarizing their harmonic mean. Differences across model outputs were evaluated using the Kruskal–Wallis test, with Dunn’s post-hoc comparisons and Bonferroni correction applied where appropriate ($p < 0.05$). All reported results correspond to mean $\pm$ standard deviation across cross-validation folds.

All experiments were conducted on a single workstation with 12 CPU cores, 128 GB RAM, and an NVIDIA RTX 4090 GPU (24 GB VRAM; capped at 14 GB per job). Temporal embedding training required approximately 34 to 48 hours per model, bounded by early stopping upon loss stabilization, whereas static baselines took only 2 to 8 hours. Supervised classifier training across 10-fold cross-validation required far less compute: 10 to 20 minutes for LR and 3.4 to 3.6 hours for XGB and RF. This asymmetry reflects the underlying scale: tensor factorization models like TNT-ComplEx [28] train over $N = 11,434,158$ TSKG triples at $\mathcal{O}(N \cdot d)$, whereas downstream classifiers operate on only $n = 53,330$ ICU stays with a fixed $p = 349$ dimensional feature vector. Because N is over 200 times larger than n , embedding training dominates overall computational cost.

Ablation studies

To isolate the separate contributions of temporal and semantic information, we introduce two controlled ablations of the full temporal–semantic graph: *w/o time*, temporal annotations removed, semantic structure preserved; and *w/o semantic*, semantic structure removed, temporal annotations preserved. These are implemented with ComplEx and TNT-ComplEx respectively, as the static and temporal members of the same architectural family.

These ablations are not alternative modelling paradigms; rather, they are systematic degradations introduced solely for analysis, enabling an empirical assessment of how temporal and semantic information influence the learned embedding representations and downstream predictive performance. Both variants follow the exact same construction and embedding training protocol, hyper-parameter settings, and supervised learning workflow as the full TSKG-ICU, and are evaluated across all three classifiers (LR, RF, and XGB), ensuring that any observed differences arise from representational changes rather than differences in model training or evaluation. Predictions are thresholded at the default decision boundary of 0.5 throughout.

Beyond downstream classification, we assess representation quality through link prediction on held-out patient-level triples, reporting Mean Reciprocal Rank (MRR) and Hits@ k ($k \in \{1, 3, 10\}$) across the *complete* graph and its ablated variants. Following the standard filtered evaluation protocol [28], candidate entities are ranked after removing known true entities, while shared ontology-level triples are excluded.

To complement this evaluation, we also analyse model behaviour directly at the patient trajectory level, examining individual ICU stays in which the *complete* and ablated configurations diverge in their predictions.

Results

Predicting ICU readmission with temporal semantics

To evaluate the effectiveness of temporal–semantic knowledge graph embeddings, we compare multiple TKGE architectures against a demographic-only baseline across three supervised learning models. This first analysis addresses two questions: do temporal–semantic representations improve prediction over conventional features,

and which embedding architectures best leverage the relational and temporal structure necessary for accurate readmission prediction.

Table 3 summarises the performance of the three classifiers (LR, RF, and XGB) across baseline features and multiple TKGE architectures. Given the class imbalance in this cohort, we report PR-AUC and F1-score alongside ROC-AUC, as ROC-AUC alone can be insensitive to majority-class overestimation in imbalanced settings. In all classifiers, TSKGE-based models outperform the baseline, indicating that temporal and semantic information extracted from the TSKG contributes substantially to improving ICU readmission prediction. The baseline model exhibit modest performance (ROC-AUC $\approx$ 0.60, PR-AUC $\approx$ 0.34, F1 $\approx$ 0.42), reflecting the limitations of raw EHR features. All KG-based embeddings exceed this baseline, demonstrating that structured temporal–semantic representations capture clinically meaningful patterns that are missing from conventional inputs.

Among the architectures evaluated, TNT-ComplEx is consistently the strongest embedding within every classifier, achieving the highest ROC-AUC, PR-AUC, and F1-score in each case (Table 3). This advantage holds under PR-AUC and F1, metrics sensitive to majority-class overestimation, confirms it is not an artefact of ROC-AUC alone. It also holds against more recent architectures trained under the same pipeline: TeAST and TCompoundE achieve competitive results, particularly TCompoundE under XGB (ROC-AUC 0.783, PR-AUC 0.632), but neither surpasses TNT-ComplEx on any classifier. Like its static counterpart, ComplEx, TNT-ComplEx leverages asymmetric relations effectively, which is crucial for handling clinical KGs, where relations are often asymmetric and directed. It is also capable of capturing higher-order interactions among entities, relations, and time simultaneously, which is a crucial feature when modelling the complex clinical phenomena that underlie ICU stays. Considering its consistent performance across classifiers, we focused on the TNT-ComplEx model for the subsequent studies.

Table 3 ICU readmission prediction performance

ML	Embedding	Acc.	AUROC	AUPRC	Prec.	Spec.	Sens.	F1
LR	T-TransE	0.748 $\pm$ 0.002	0.652 $\pm$ 0.010	0.390 $\pm$ 0.011	0.566 $\pm$ 0.037	0.987 $\pm$ 0.001	0.051 $\pm$ 0.006	0.094 $\pm$ 0.01
	TA-TransE	0.762 $\pm$ 0.003	0.701 $\pm$ 0.008	0.491 $\pm$ 0.006	0.660 $\pm$ 0.024	0.975 $\pm$ 0.002	0.140 $\pm$ 0.012	0.231 $\pm$ 0.017
	TA-DistMult	0.791 $\pm$ 0.005	0.760 $\pm$ 0.007	0.584 $\pm$ 0.012	0.682 $\pm$ 0.02	0.946 $\pm$ 0.004	0.341 $\pm$ 0.012	0.455 $\pm$ 0.013
	T-ComplEx	0.806 $\pm$ 0.002	0.776 $\pm$ 0.009	0.624 $\pm$ 0.011	0.730 $\pm$ 0.011	0.952 $\pm$ 0.003	0.383 $\pm$ 0.009	0.502 $\pm$ 0.007
	TNT-ComplEx	0.810 $\pm$ 0.004	0.786 $\pm$ 0.009	0.641 $\pm$ 0.01	0.742 $\pm$ 0.014	0.953 $\pm$ 0.003	0.395 $\pm$ 0.016	0.515 $\pm$ 0.015
	TeAST	0.799 $\pm$ 0.003	0.774 $\pm$ 0.007	0.603 $\pm$ 0.008	0.721 $\pm$ 0.015	0.954 $\pm$ 0.004	0.349 $\pm$ 0.006	0.470 $\pm$ 0.004
	TcompoundE	0.804 $\pm$ 0.003	0.778 $\pm$ 0.008	0.621 $\pm$ 0.008	0.731 $\pm$ 0.013	0.954 $\pm$ 0.003	0.368 $\pm$ 0.011	0.490 $\pm$ 0.01
RF	T-TransE	0.688 $\pm$ 0.007	0.647 $\pm$ 0.009	0.380 $\pm$ 0.014	0.388 $\pm$ 0.014	0.791 $\pm$ 0.007	0.388 $\pm$ 0.023	0.388 $\pm$ 0.018
	TA-TransE	0.696 $\pm$ 0.006	0.688 $\pm$ 0.006	0.432 $\pm$ 0.008	0.419 $\pm$ 0.007	0.765 $\pm$ 0.011	0.495 $\pm$ 0.024	0.454 $\pm$ 0.006
	TA-DistMult	0.742 $\pm$ 0.004	0.752 $\pm$ 0.006	0.538 $\pm$ 0.010	0.495 $\pm$ 0.007	0.811 $\pm$ 0.009	0.541 $\pm$ 0.016	0.517 $\pm$ 0.006
	T-ComplEx	0.756 $\pm$ 0.003	0.757 $\pm$ 0.006	0.563 $\pm$ 0.009	0.521 $\pm$ 0.008	0.833 $\pm$ 0.005	0.529 $\pm$ 0.011	0.525 $\pm$ 0.007
	TNT-ComplEx	0.761 $\pm$ 0.004	0.770 $\pm$ 0.005	0.589 $\pm$ 0.008	0.530 $\pm$ 0.008	0.835 $\pm$ 0.005	0.545 $\pm$ 0.013	0.538 $\pm$ 0.009
	TeAST	0.741 $\pm$ 0.002	0.754 $\pm$ 0.006	0.542 $\pm$ 0.013	0.493 $\pm$ 0.004	0.808 $\pm$ 0.005	0.544 $\pm$ 0.012	0.517 $\pm$ 0.006
	TcompoundE	0.718 $\pm$ 0.006	0.702 $\pm$ 0.008	0.467 $\pm$ 0.007	0.449 $\pm$ 0.011	0.808 $\pm$ 0.009	0.456 $\pm$ 0.017	0.452 $\pm$ 0.011
XGB	T-TransE	0.494 $\pm$ 0.007	0.643 $\pm$ 0.009	0.389 $\pm$ 0.013	0.315 $\pm$ 0.016	0.431 $\pm$ 0.069	0.759 $\pm$ 0.057	0.444 $\pm$ 0.006
	TA-TransE	0.638 $\pm$ 0.037	0.697 $\pm$ 0.009	0.470 $\pm$ 0.010	0.381 $\pm$ 0.024	0.636 $\pm$ 0.072	0.645 $\pm$ 0.067	0.476 $\pm$ 0.008
	TA-DistMult	0.726 $\pm$ 0.004	0.765 $\pm$ 0.007	0.586 $\pm$ 0.010	0.471 $\pm$ 0.006	0.764 $\pm$ 0.007	0.617 $\pm$ 0.015	0.535 $\pm$ 0.008
	T-ComplEx	0.748 $\pm$ 0.007	0.773 $\pm$ 0.010	0.616 $\pm$ 0.010	0.505 $\pm$ 0.012	0.798 $\pm$ 0.011	0.601 $\pm$ 0.014	0.549 $\pm$ 0.009
	TNT-ComplEx	0.758 $\pm$ 0.008	0.785 $\pm$ 0.007	0.635 $\pm$ 0.007	0.522 $\pm$ 0.015	0.809 $\pm$ 0.016	0.608 $\pm$ 0.02	0.561 $\pm$ 0.008
	TeAST	0.740 $\pm$ 0.009	0.776 $\pm$ 0.006	0.606 $\pm$ 0.007	0.494 $\pm$ 0.012	0.780 $\pm$ 0.019	0.625 $\pm$ 0.027	0.551 $\pm$ 0.008
	TcompoundE	0.753 $\pm$ 0.007	0.783 $\pm$ 0.007	0.632 $\pm$ 0.011	0.515 $\pm$ 0.012	0.803 $\pm$ 0.017	0.608 $\pm$ 0.026	0.557 $\pm$ 0.009
	Baseline	0.537 $\pm$ 0.010	0.603 $\pm$ 0.008	0.335 $\pm$ 0.011	0.307 $\pm$ 0.007	0.498 $\pm$ 0.016	0.650 $\pm$ 0.020	0.417 $\pm$ 0.010

mean and standard deviation (10-fold cross-validation) of accuracy (acc.), ROC-AUC (AUROC), PR-AUC (AUPRC), precision (prec.), specificity (spec.), sensitivity or recall (sens.), and F1-score, for Logistic Regression (LR), Random forest (RF), and XGBoost (XGB) classifiers trained on TKG embeddings. For all metrics, higher values indicate better performance. Bold values denote the best value per metric within each classifier

Across the remaining embedding families, TransE- and DistMult-based temporal models showed limited ability to capture the complexity of clinical trajectories. Temporal TransE variants delivered the weakest performance, reflecting the difficulty of modelling heterogeneous and asymmetric clinical relations through simple vector translations. Notably, TA-TransE outperformed T-TransE, indicating that the specific mechanism used to incorporate time (LSTMs) to encode temporal sequence rather than direct translation, has a measurable impact on model effectiveness. TA-DistMult achieved more competitive results overall, yet its inherent symmetry assumption restricts its ability to represent directed and semantically layered interactions. Although both families benefited from incorporating temporal information and improved over the baseline, they were ultimately less effective at leveraging temporal–semantic structure than the ComplEx-based models.

Extending beyond the two questions posed above, no single classifier dominates, and the differences between them are clinically consequential. LR attains the highest accuracy (0.810), ROC-AUC (0.786), PR-AUC (0.641), and precision (0.742), but the lowest sensitivity (0.395), recovering fewer than two in five true readmissions. XGB achieves near-identical ranking quality (ROC-AUC 0.785, PR-AUC 0.635) yet substantially higher sensitivity (0.608) and F1 (0.561), trading precision for recall. RF sits between the two on sensitivity (0.545) and F1 (0.538), with weaker ranking performance overall (ROC-AUC 0.770, PR-AUC 0.589). Since PR-AUC is threshold-independent and effectively equivalent for LR and XGB, the divergence in sensitivity reflects where each classifier’s predicted probabilities fall relative to the default 0.5 decision boundary rather than a difference in what the embeddings encode.

This distinction carries direct clinical weight, and it is one that global metrics alone cannot express. ROC-AUC and PR-AUC summarise performance across all operating points and are therefore silent on the errors a deployed model would actually make at a given threshold. In readmission prediction the two error types are not equivalent: a false positive delays a discharge and consumes observation resources, whereas a false negative discharges a patient who subsequently deteriorates and is readmitted, incurring precisely the mortality and prolonged-stay risks that motivate the task. This asymmetry favours over-flagging, and therefore favours sensitivity over precision. On this criterion XGB is the stronger choice and LR the weaker, despite their near-identical overall performance. We therefore retain all three classifiers throughout the ablation and patient-level analyses that follow, rather than designating a single reference model, since the determining factor is tolerance for each error type rather than any property of the embeddings.

Ablation studies

The results of the ablation study presented in Table 4 highlight the contributions of temporal and semantic information in our ICU readmission approach. Removing the temporal component (“w/o time”), which corresponds to employing a KG without temporal annotations on edges, leads to the largest drop in performance across all three

Table 4 Ablation studies

ML	Model	Acc.	AUROC	AUPRC	Prec.	Spec.	Sens.	F1
LR	w/o time	0.791 ±0.004	0.760 ±0.008	0.553 ±0.01	0.686 ±0.017	0.948 ±0.005	0.334 ±0.013	0.449 ±0.012
	w/o semantics	0.807 ±0.003	0.782 ±0.008	0.628 ±0.013	0.730 ±0.016	0.951 ±0.004	0.385 ±0.014	0.504 ±0.012
	complete	0.810 ±0.004	0.786 ±0.009	0.641 ±0.011	0.742 ±0.014	0.953 ±0.003	0.395 ±0.016	0.515 ±0.015
RF	w/o time	0.776 ±0.005	0.757 ±0.006	0.560 ±0.009	0.572 ±0.013	0.878 ±0.006	0.477 ±0.011	0.520 ±0.010
	w/o semantics	0.752 ±0.003	0.763 ±0.006	0.571 ±0.006	0.514 ±0.007	0.825 ±0.007	0.540 ±0.012	0.527 ±0.006
	complete	0.761 ±0.004	0.770 ±0.005	0.589 ±0.008	0.530 ±0.008	0.835 ±0.005	0.545 ±0.013	0.538 ±0.009
XGB	w/o time	0.732 ±0.005	0.768 ±0.005	0.578 ±0.009	0.481 ±0.008	0.769 ±0.009	0.623 ±0.018	0.543 ±0.009
	w/o semantics	0.753 ±0.006	0.781 ±0.008	0.625 ±0.007	0.513 ±0.009	0.803 ±0.014	0.607 ±0.027	0.556 ±0.011
	complete	0.758 ±0.008	0.785 ±0.007	0.635 ±0.007	0.522 ±0.015	0.809 ±0.016	0.608 ±0.020	0.561 ±0.008

Performance comparison of *complete* TSKG-ICU versus ablated configurations across all three classifiers, reporting accuracy (acc.), ROC-AUC (AUROC), PR-AUC (AUPRC), precision (prec.), specificity (spec.), sensitivity (Sens.), and F1-score. Mean ± standard deviation over 10-fold cross-validation. Bold values denote the best value per metric within each classifier

classifiers, with ROC-AUC falling from 0.786 to 0.760 under LR, 0.770 to 0.757 under RF, and 0.785 to 0.768 under XGB. This demonstrates that modelling the temporal evolution of facts is crucial for capturing dynamic patterns in the knowledge graph. Excluding semantic information (“*w/o semantics*”), which corresponds to using a knowledge graph without the ontological enrichment, results in a smaller but consistent decline (ROC-AUC 0.782, 0.763, and 0.781 respectively). This indicates that ontological context provides complementary information that enhances relational reasoning. The full approach, which integrates both temporal and semantic features in the KG, achieves the best performance, illustrating that the synergy between temporal dynamics and ontological knowledge is essential for accurate prediction.

Under PR-AUC, the metric most sensitive to the class imbalance in this cohort, the same pattern is more pronounced: removing time costs between 0.029 and 0.088 (on RF and LR respectively), while removing semantics costs between 0.010 and 0.018. The clinical consequence is carried by sensitivity and precision: under LR and RF, the *complete* configuration recovers 18 and 14% more true readmissions than *w/o time*, at comparable or better precision. Since a missed readmission carries greater risk than a false alarm, this is where the temporal component has its clearest effect. This ordering is consistent across all three classifiers, indicating that it stems from the representations themselves rather than from a particular learning algorithm.

The ROC and precision–recall curves presented in Fig. 4 reinforce these results: across all three classifiers, the *complete* and *w/o-semantics* curves track closely together and well above the *w/o-time* curve, which in turn remains clearly above the demographic-only baseline. The separation is more pronounced in the precision–recall panels than in the ROC panels, consistent with the larger PR-AUC deltas reported above.

Figure 5 presents the corresponding confusion matrices at the default decision threshold across true positives, false negatives, and false positives. Removing semantic information yields a consistent but modest performance drop across all classifiers. The *complete* graph simultaneously improves sensitivity and precision across the board, identifying between 19 and 131 additional readmissions while reducing false positives by 69 to 373 depending on the model. The gain is largest for LR, where true positives rise from 5,236 to 5,367 and false positives fall from 1,939 to 1,870, and smallest for XGB, where the 19 additional readmissions identified (8,251 to 8,270) are accompanied by 245 fewer false positives (7,839 to 7,594). RF sits between them on detection (7,344 to 7,414) but shows the largest reduction in false positives (6,937 to 6,564). Semantic enrichment therefore improves the ensemble classifiers by suppressing incorrect predictions, and the linear classifier by recovering readmissions it would otherwise miss. In contrast, removing temporal information affects far more patients. For LR and RF, temporal ablation severely impacts sensitivity, causing 829 and 928 readmissions to go undetected, respectively. For XGB, temporal ablation yields a marginal gain of 207 detected readmissions, but at the cost of 1,571 additional false alarms. While the specific directional trade-off depends on the classifier, its relative magnitude remains consistent: temporal ablation displaces roughly an order of magnitude more patients than semantic ablation in every case. This confirms temporal structure as the primary contributor to predictive performance, with semantic enrichment serving as a consistent but secondary refinement.

These performance characteristics also differentiate the classifiers. Out of 13,605 total readmissions, the *complete* configuration under LR identifies 5,367 readmissions (misses 61%) with 1,870 false positives. By contrast, RF identifies 7,414 (misses 45%) with 6,564 false positives, and XGB identifies 8,270 (misses 39%) with 7,594 false positives. LR is the most precise (74% precision across 7,237 alerts), but misses three out of five true readmissions. XGB operates at a lower precision (52% across 15,864 alerts) but detects substantially more positive cases. Transitioning from LR to XGB captures 2,903 additional readmitted patients at a cost of 5,724 additional false positives, roughly two false alarms for every additional true positive identified. In a critical care setting, this trade-off heavily favors sensitivity. As previously mentioned, missed readmissions carry severe risks, whereas false positives typically prompt continued monitoring. Consequently, reducing missed readmissions by 35% represents a far more clinically meaningful gain. Furthermore, this sensitivity gain comes with virtually no loss in global discrimination quality, as both ROC-AUC (0.785 vs. 0.786) and PR-AUC (0.635 vs. 0.641) remain effectively identical between the models. On this basis, XGB represents the more clinically appropriate configuration for deployment.

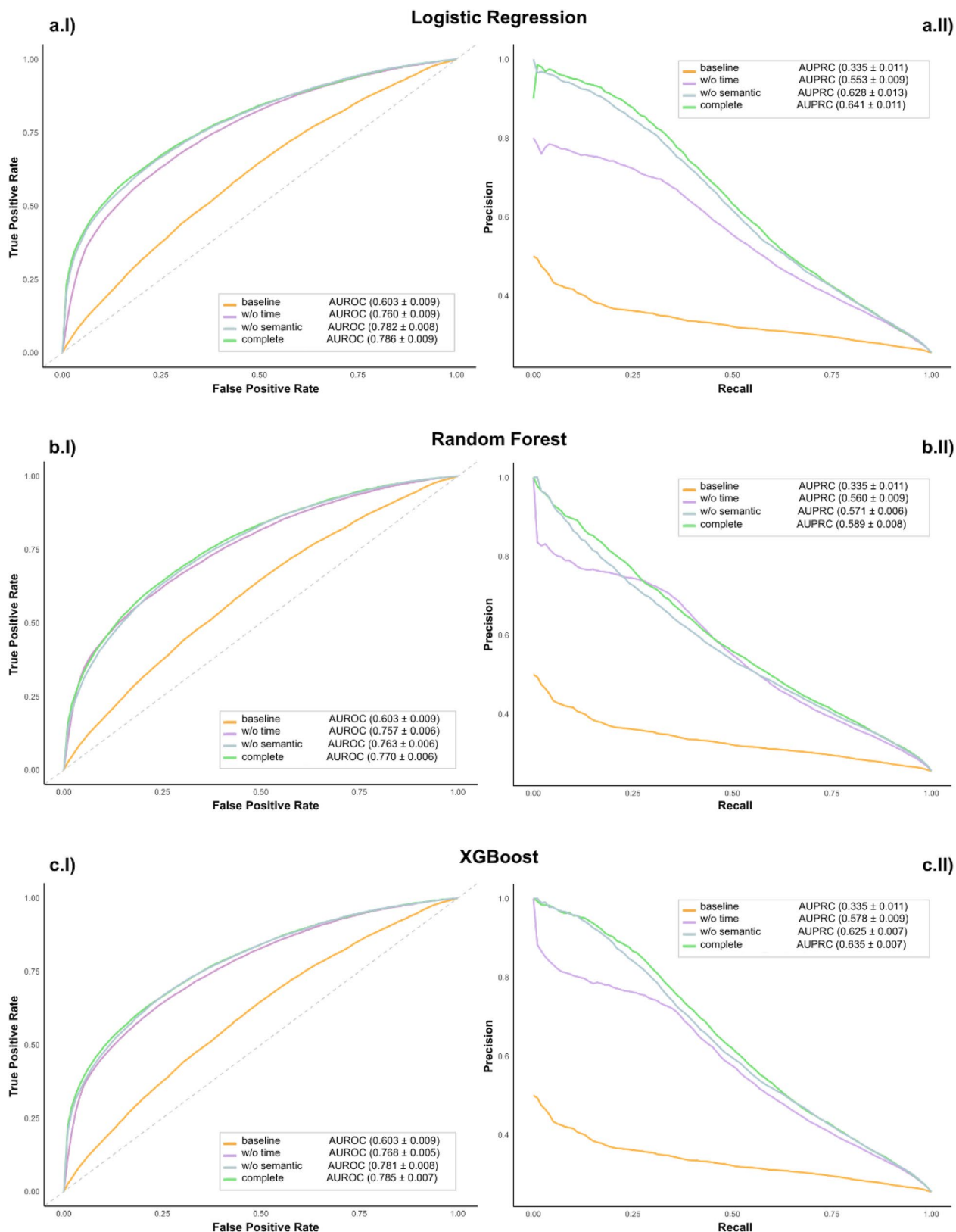


Fig. 4 Predictive performance across ablated knowledge-graph configurations and classifiers. ROC curves and precision–recall curves for **(a)** Logistic Regression, **(b)** Random Forest, and **(c)** XGBoost, each comparing four settings: the demographic-only baseline, *w/o time*, *w/o semantic*, and the *complete* temporal–semantic KG

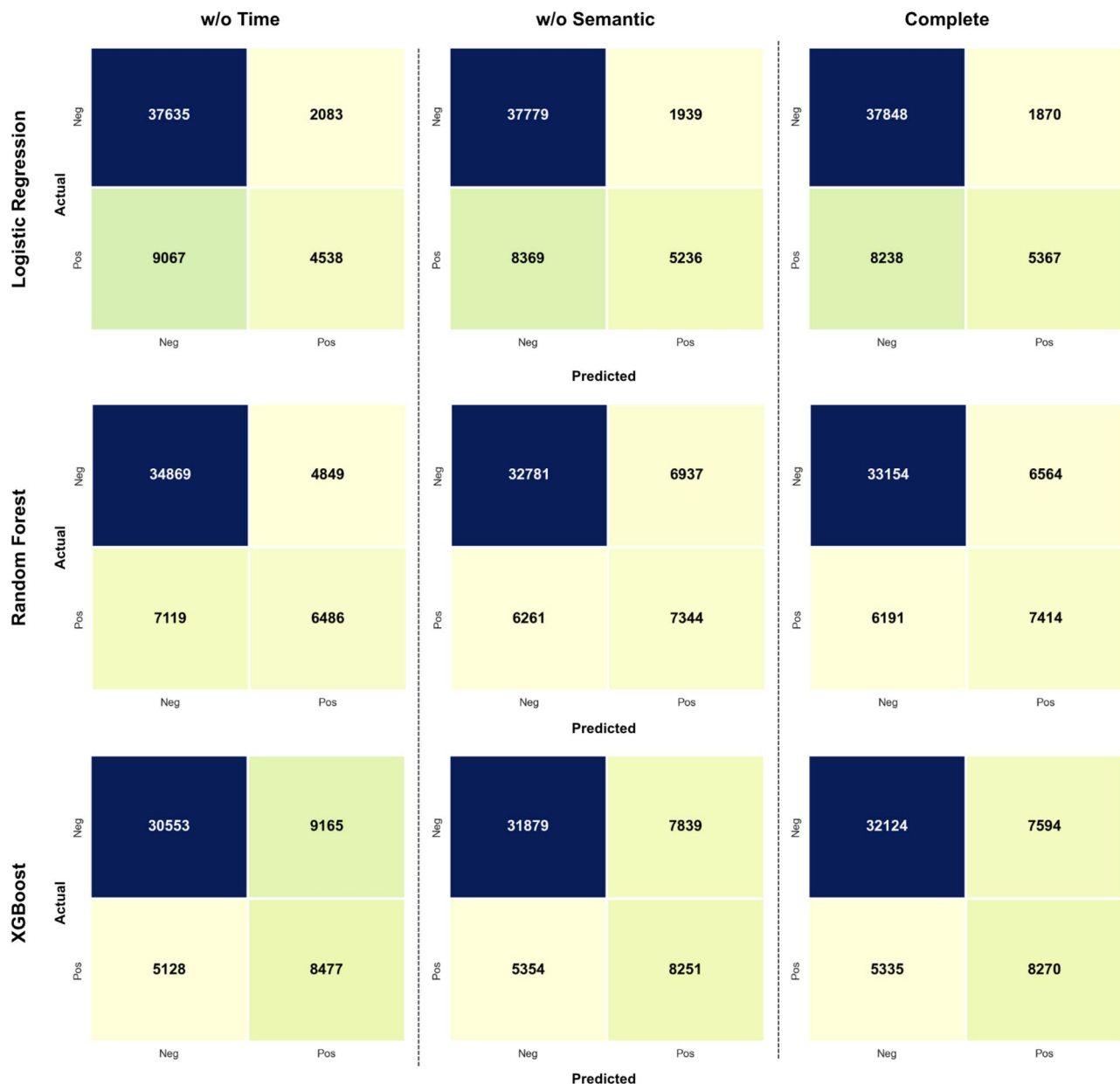


Fig. 5 Confusion matrices across ablated knowledge-graph configurations and classifiers. counts at the default decision threshold ($\tau = 0.5$) for the *w/o time*, *w/o semantic*, and *complete* configurations (columns), under Logistic Regression, XGBoost, and Random Forest (rows), summed across the 10 cross-validation folds. Each matrix reports true negatives and false positives (upper row) and false negatives and true positives (lower row), for a cohort of 13,605 readmissions and 39,718 non-readmissions. Colour intensity scales with cell count within each matrix

Intrinsic KGE performance

To complement the downstream evaluation, we examine how temporal and semantic structure affect the intrinsic quality of the learned representations. Table 5 reports filtered link prediction performance on held-out patient-level triples. The *complete* and *w/o time* configurations achieve comparable MRR values of 0.943 and 0.945, respectively, whereas removing semantic structure reduces MRR to 0.877. This reduction is concentrated at the highest rank: Hits@1 decreases from 0.890 for the *complete* graph to 0.760 without semantic relations,

Table 5 Intrinsic link prediction performance across ablated configurations

Model	MRR	Hits@1	Hits@3	Hits@10
w/o time	0.945	0.914	0.978	0.981
w/o semantic	0.877	0.760	0.996	0.997
complete	0.943	0.890	0.998	0.998

Mean Reciprocal rank (MRR) and Hits@ k ($k \in \{1, 3, 10\}$) on held-out patient-level triples, using a filtered ranking protocol. Higher values indicate better-ranked, more internally consistent embeddings

although almost all correct entities remain within the top three predictions. Semantic structure therefore primarily improves the embeddings' ability to identify the correct entity as the highest-ranked candidate rather than their broader retrieval capacity.

Removing temporal relations produces the highest Hits@1 (0.914), while the *complete* graph achieves the highest Hits@3 and Hits@10. Thus, temporal information does not improve the reconstruction of individual triples, for which the additional temporal relations may introduce alternative plausible candidates. This does not imply that temporal structure is uninformative for readmission prediction. Intrinsic link prediction evaluates how faithfully embeddings reconstruct local graph structure, whereas the downstream task evaluates whether the resulting patient representations discriminate between clinically distinct outcomes. Indeed, the downstream results in Table 3 show that the temporally expressive TNT-ComplEx embeddings consistently provide the strongest predictive performance across classifiers, achieving the highest ROC-AUC and PR-AUC with LR, RF, and XGB. The two evaluations are therefore complementary: semantic structure improves the internal consistency of the embedding space, while temporal structure contributes information that is particularly valuable when patient representations are used for clinical prediction.

Patient-level insights

While aggregate analyses characterize overall model behaviour, they do not convey how representational differences manifest within individual patient trajectories. To provide an interpretable, patient-level perspective, we examine two representative cases. We specifically isolate true readmissions correctly identified only by the *complete* model and missed by both ablated variants across all three classifiers. Positive cases are prioritized because they constitute the primary clinical target: identifying patients who will return. Selecting from this unique subset directly tests our central claim: that temporal and semantic information are synergistic, such that neither dimension alone is sufficient to recover the correct prediction.

For each case, we reconstruct the patient's ICU timeline from the TSKG, reporting key clinical events alongside their relative time since admission (Fig. 6, panels a), and compare the predicted readmission probability produced by each model variant (Fig. 6, panels b).

Case 1 (Patient 998): This patient represents a true 30-day readmission correctly identified only by the *complete* model ($p = 0.77$), whereas both ablated variants fell below the decision threshold (*w/o time*: $p = 0.49$; *w/o semantics*: $p = 0.47$). The patient's stay presents no obvious primary diagnostic escalation, opening with coronary artery disease and closing with an unrelated diagnosis of gastroesophageal reflux. Instead, clinical deterioration appears as a subtle trajectory shift: an initial cardiac presentation marked by antiplatelet therapy (clopidogrel escalated with eptifibatide at hour 7) rapidly pivots to renal failure. From hour 30 through hour 73, recurring creatinine and blood urea nitrogen panels lead directly to the initiation of hemodialysis. Without temporal ordering, the time-ablated variant cannot distinguish an isolated anomalous laboratory result from a sustained 40-hour trajectory. Conversely, the semantic-ablated variant retains temporal sequence but underestimates risk, failing to connect creatinine, blood urea nitrogen, and hemodialysis as progressive manifestations of renal failure. By capturing both the temporal progression and semantic linkages across this cardiac-to-renal shift, the *complete* model successfully identifies an otherwise elusive high-risk trajectory.

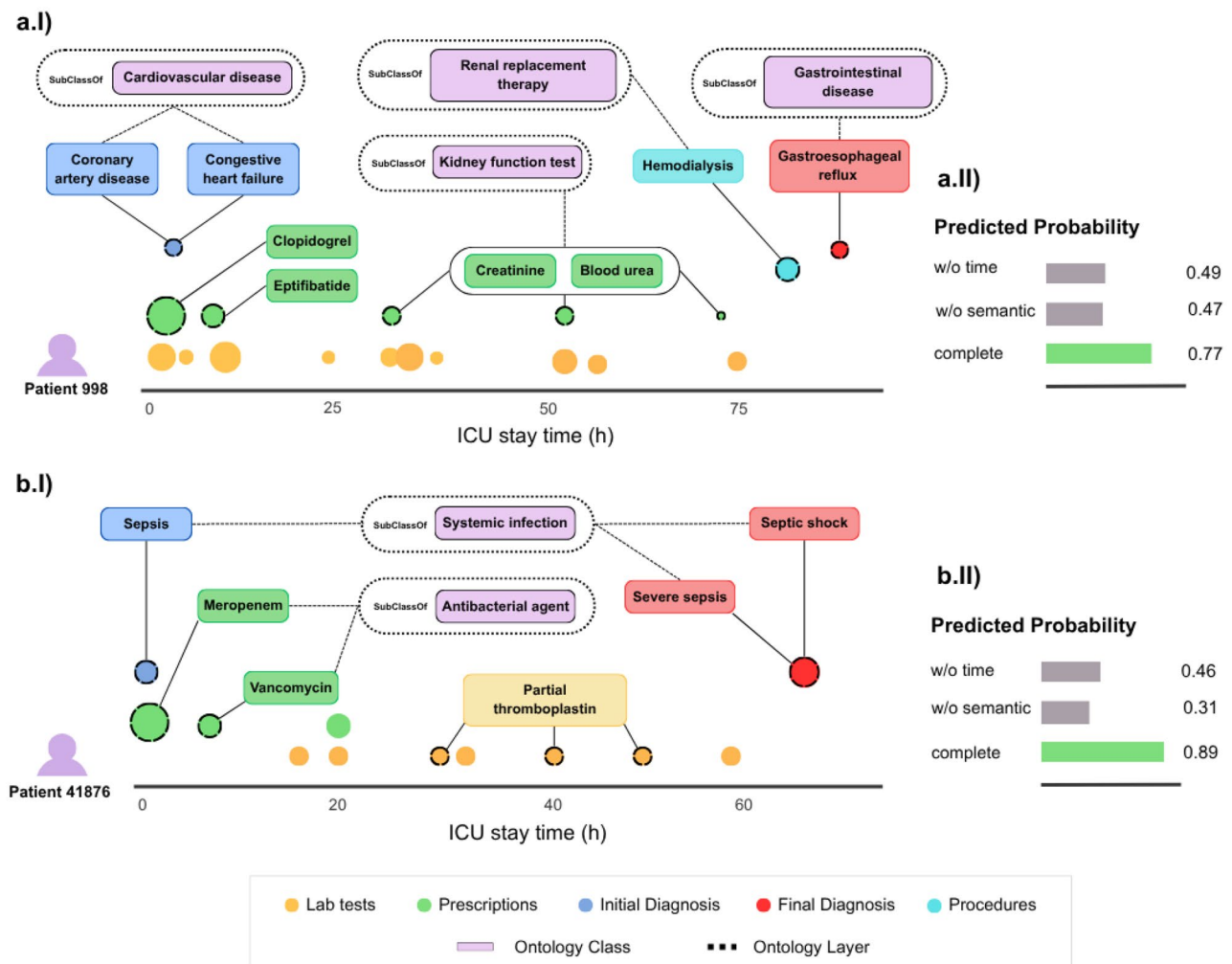


Fig. 6 Illustrative comparison of two true 30-day readmissions correctly identified only by the *complete* model. (a.I, b.I) condensed timelines of key clinical events for case 1 (patient 998) and case 2 (patient 41,876), relative to hours since admission, showing initial and final diagnoses, prescriptions, laboratory events, and procedures. Marker size reflects the number of concurrent events at each time point, and selected ontology classes are drawn above the events they group. (a.II, b.II) predicted readmission probability for each patient under the ablated variants (*w/o time*, *w/o semantics*) and the *complete* model. Both cases are illustrative rather than representative of typical model behaviour across the cohort

Case 2 (Patient 41,976): This patient was likewise a true 30-day readmission correctly identified only by the *complete* model ($p = 0.89$), while both ablated variants underestimated risk (*w/o time*: $p = 0.46$; *w/o semantics*: $p = 0.31$). The patient's stay demonstrates a clear diagnostic escalation: an initial diagnosis of sepsis treated with meropenem was escalated with vancomycin at hour 7, alongside sustained anticoagulation monitoring through hour 58. Fentanyl was administered at admission and again at hour 31, a recurring analgesic response reflecting ongoing pain management across a deteriorating course that culminated by hour 71 in severe sepsis and septic shock. Without temporal ordering, the time-ablated variant cannot detect diagnosis escalation, nor can it recognize recurring pain management as part of an evolving clinical course, effectively collapsing a worsening sequence into an unordered set of facts. Conversely, the semantic-ablated variant retains temporal sequence but fails to appreciate risk, lacking the ontological structure required to link sepsis, severe sepsis, and septic shock as progressive, increasingly severe manifestations of the same underlying disease process.

Table 6 Performance comparison with prominent ICU readmission methods

Work	AUROC	AUPRC	F1	Prec.	Spec.	Sens.
Niu et al. [12]	0.73	0.67	0.64	N/R	0.77	0.59
Barbieri et al. [11]	0.74	0.32	0.38	N/R	0.70	0.68
Sheetrit et al. [9]	0.75	0.30	0.37	0.27	0.58	N/R
Pakbin et al. [10]	0.75	N/R	0.37	N/R	N/R	N/R
Pei et al. [13]	0.79	0.38	0.42	N/R	0.71	0.74
Lin et al. [7]	0.79	0.51	N/R	0.37	N/R	0.74
TSKG-ICU	0.79	0.64	0.56	0.52	0.81	0.61

Performance of TSKG-ICU relative to widely used machine-learning and deep-learning approaches evaluated on MIMIC-III, reported using ROC-AUC (AUROC), PR-AUC (AUPRC), F1-score, precision (prec.), specificity (spec.), and sensitivity (Sens.) where available. TSKG-ICU results correspond to the XGB configuration. Bold values denote the best-performing method per metric. N/R: not reported

Both these patients showcase how the *complete* model correctly identifies risk across different clinical trajectories where the *w/o time* and *w/o semantics* variants each miss a different aspect of the pattern. These interpretations are illustrative, not causal claims, and the two cases were selected specifically because only the *complete* model succeeded on them; general model performance remains established by the aggregate results in Table 4.

Comparative performance of TSKG-ICU

To contextualize the performance of the proposed TSKG-ICU, we compare it against reported results of prominent machine-learning and deep-learning models for ICU readmission prediction [6]. Given inherent variations in cohort definitions and feature availability, this comparison positions our results within the broader literature rather than serving as a strict head-to-head benchmark. Consistent with established field standards, ROC-AUC is used as the primary metric; PR-AUC and F1-score, precision, sensitivity and specificity, are also reported where available to provide a more complete picture, though some works do not report these metrics (denoted N/R). All metrics are reported as decimals to two places, as source studies use varying precision and formats. We report TSKG-ICU under XGB here, as the configuration whose operating point best reflects the clinical priorities established above, and note that ROC-AUC and PR-AUC are effectively identical across our three classifiers.

Table 6 summarises the predictive performance of six selected approaches chosen for their similar conditions: all were applied to MIMIC-III, address 30-day ICU readmission, operate under highly imbalanced class distributions, and incorporate demographic and structured clinical data. Furthermore, approaches leveraging temporal or longitudinal modelling were prioritised, as capturing patient trajectories is central to the readmission task and directly relevant to the proposed method.

Across the selected works, reported ROC-AUC values range from 0.73 to 0.79. TSKG-ICU achieves the top ROC-AUC (0.79), in line with the best state-of-the-art approaches, while also bringing methodological advantages in how patient trajectories are represented. We note, however, that Niu et al. [12] report higher PR-AUC (0.67) and F1-score (0.64) than TSKG-ICU (0.64, 0.56), despite a lower ROC-AUC (0.73). TSKG-ICU in turn achieves the highest precision (0.52) and specificity (0.81) among the compared works. Lin et al. [7] and Pei et al. [13] report higher sensitivity (0.74 and 0.74 against 0.60), though at substantially lower precision where reported (0.37 for Lin), indicating a more permissive operating point rather than better discrimination. As with the ROC-AUC comparison, differences in cohort definition, feature availability, and decision threshold limit direct interpretation. Still, these differences show that matching the top ROC-AUC does not guarantee the best performance on every metric: TSKG-ICU is competitive throughout and strongest on precision and specificity, but not uniformly superior.

Many referenced approaches also restrict modelling to fixed pre-discharge subsets (e.g., the final 24–48 hours) [7, 10] or rely on task-specific temporal abstractions, such as ODE-based encoders [11–13] or GRU-based [9], that do not embed patient trajectories into a shared temporal reference space, biasing predictions toward late-stage events and limiting comparability across patients with different lengths of stay. In contrast, TSKG-ICU leverages

the entire ICU timeline through relative temporal encoding, aligning all patient trajectories to a common temporal reference and enabling direct comparison of disease progression and care patterns across stays. In addition, ontology grounding places clinical events within a shared semantic space, ensuring that heterogeneous patient records are represented in a semantically comparable manner. Together, these design choices allow TSKG-ICU to achieve high predictive performance while providing a more general, temporally fair, and semantically aligned representation of ICU trajectories than prior methods.

Discussion

Conventional ICU readmission models either capture temporal dynamics through sequential processing or encode clinical relationships through knowledge graphs, but rarely both. TSKG-ICU was designed to address this gap by integrating these dimensions within a unified temporal–semantic representation. Our approach yielded significant predictive improvements, and ablation studies quantified their contributions consistently across three classifiers: temporal modelling provides the primary signal, whilst semantic enrichment refines predictions, and their combination surpasses either dimension alone. These gains also hold for PR-AUC, which is more sensitive to the class imbalance inherent to ICU readmission data: temporal modelling similarly provides the primary signal, while semantic enrichment offers a smaller, complementary refinement. The consistent ordering across all three classifiers confirms that these contributions are not an artefact of ROC-AUC or of an interaction with a particular learning algorithm, but instead reflect the information encoded by the representations.

A key finding from our evaluation is that temporal and semantic information must be encoded jointly for effective predictions, when compared to using each dimension alone. Our framework, TSKG-ICU, consistently enables effective ICU readmission prediction; however, the strongest performance is achieved only when time and semantics are integrated within an expressive embedding architecture. The ablation study highlights this interplay clearly: whilst removing temporal or semantic information still yields competitive results over non-embedding baselines, neither ablated configuration surpasses the full temporal–semantic formulation on ROC-AUC, PR-AUC or F1, confirming that predictive gains stem from their interaction. This interaction is driven primarily by temporal modelling, a pattern clear when examining individual contributions. Although all temporal–semantic embedding architectures benefit from access to time-stamped data, their effectiveness depends on how temporal information is encoded. In particular, TNT-Complex achieves the most reliable predictive performance, indicating that modelling time as a continuous relational transformation better allows embeddings to capture the overlapping and irregular event patterns of ICU trajectories. Simpler temporal encodings result in weaker gains despite using the same data, showing that temporal information is not inherently beneficial; its value emerges only when supported by sufficient representational expressiveness.

Semantic information, in turn, operates as a complementary signal that strengthens temporally grounded representations. Its contribution is small but consistent: for every classifier, the *complete* graph is simultaneously more sensitive and more precise than its semantically ablated counterpart, recovering additional readmissions while producing fewer false-positive predictions. By contrast, temporal ablation changes the predictions of roughly an order of magnitude more patients, with the direction and magnitude of this displacement varying across classifiers. This pattern indicates that temporal modelling is the primary factor aligning models on true readmission cases. Semantic enrichment further refines this alignment by adding contextual structure to clinical events, improving consistency among temporally informed predictions. A similar tendency is observed for negative cases, although the very high overall agreement reduces the visibility of architecture-dependent differences. Semantic enrichment therefore does not compete with temporal structure for the same predictive signal; instead, it refines a representation that temporal encoding has already organised. This helps explain why its effect is more readily detected through precision-sensitive metrics such as PR-AUC than through ROC-AUC alone. Together, these results indicate that semantic information is most effective when it amplifies a temporally grounded representation, rather than acting as an independent driver of prediction, reinforcing the value of TSKG representations.

These global patterns are also reflected in individual clinical trajectories. Our patient-level analysis shows that the ablated variants fail in distinct and clinically interpretable ways: temporal ordering and semantic structure are both required to differentiate routine early interventions from critical clinical escalation and to recognise concurrent, multi-system findings as a coherent pattern of high collective severity.

While TSKGEs demonstrate strong predictive performance, their implementation in real-time clinical settings still offers room for major development and refinement. At the data and methodological level, the use of the single-centre MIMIC-III database means that readmissions occurring outside the contributing hospital network are unobserved and consequently recorded as negative instances, as discussed in Section “[Cohort definition and exclusion criteria](#)”. Although this is a standard limitation of single-centre cohorts and is shared by comparable studies, it introduces outcome-label uncertainty and suggests that the reported performance may represent a conservative estimate of absolute clinical utility. Moreover, many existing TKGE methods are transductive, meaning that generating representations for previously unseen patients or newly recorded events typically requires model update. In dynamic clinical environments, where patient data accumulate continuously and risk assessments must adapt accordingly, developing more flexible representation strategies is essential to further enhance operational integration. Emerging approaches, such as incremental training paradigms and inductive temporal graph methods [32, 33], provide promising avenues toward greater adaptability. Finally, the choice of temporal representation also introduces trade-offs. Hourly temporal discretisation preserves the relative ordering and proximity of events substantially better than a static representation, but cannot distinguish between multiple events occurring within the same hour. Similarly, laboratory tests and vital signs are represented as ontological concepts, overlooking numerical values and preventing the model from learning from them.

Future work should therefore address three interconnected architectural developments. First, inductive or incremental architectures should support the dynamic generation of representations for unseen patients and newly recorded events without requiring full graph retraining. Second, finer event-driven temporal resolutions or explicit modelling of temporal latency could recover clinical sequences that are obscured by hourly discretisation. Finally, future clinical temporal–semantic graphs should move beyond purely categorical concepts to jointly represent continuous-valued clinical observations [52]. Progress in this area would enable KG-based systems to evolve into continuously updating decision-support tools, strengthening their impact in real-world critical care settings.

Conclusion

TSKG-ICU establishes a high-performing and conceptually robust framework for ICU readmission prediction by unifying temporal dynamics and biomedical semantics within a single representation. By encoding clinical events at relative time points and grounding them in ontology-linked concepts, the model produces comparable patient embeddings that faithfully capture trajectories of deterioration, intervention, and recovery. This unified temporal–semantic design enables reliable identification of high-risk patients and provides a clinically meaningful foundation for monitoring and triage.

Our evaluation further demonstrates that architectural design is critical: asymmetric, complex-valued embeddings and explicit temporal decomposition significantly improve performance in directional clinical knowledge graphs. These findings clarify which representational properties are necessary to model ICU dynamics effectively.

To translate this framework into continuously operating decision-support systems, future work must advance inductive scalability, improve robustness to irregularly sampled data, and align more tightly with ICU workflows. With these developments, temporal–semantic knowledge graphs can enable AI systems that reason not only about what happens to patients, but when and why events unfold—supporting more adaptive and clinically integrated critical care decision-making.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/>

s12911-026-03856-9.

Acknowledgements Not applicable. Editorial Policies for: BMC journals: <https://www.biomedcentral.com/getpublished/editorial-policies>.

Author contributions RC, AST and CP designed the study. RC implemented the software and ran the experiments. RC, AST and CP analyzed and discussed the results. RC and CP wrote the manuscript. All authors have read, revised and approved the manuscript.

Funding This work was supported by FCT through the fellowship 2022.10769.BD, and the LASIGE Research Unit, ref. UID/000408/2025 (DOI: <https://doi.org/10.54499/UID/00408/2025>). It was also partially supported by the project TrustAI 4Sci CPCA-IAC/AV/594790/2023 (DOI: <https://doi.org/10.54499/CPCA-IAC/AV/594790/2023>), and the CancerScan project which received funding from the European Union's Horizon Europe Research and Innovation Action (EIC Pathfinder Open) under grant agreement No. 101,186,829. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Innovation Council and SMEs Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data availability The data used in this application (MIMIC-III) is publicly available on PhysioNet at <https://mimic.physionet.org>. Source code for our application is available from GitHub (<https://github.com/liseda-lab/TSKG-ICU.git>) under an Apache-2.0 licence. Guides for other external resources are also provided on the same GitHub. The MIMIC-III code is available at: <https://doi.org/10.5281/zenodo.821872>. The Knowledge Graph construction base code is available at <https://github.com/liseda-lab/clinical-temporal-kg> (GPL-3.0 license).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Declaration of generative AI and AI-assisted technologies in the writing process During the preparation of this work, the author(s) used genAI for copy-editing and to improve readability. After using this tool/service, the authors reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Alloubani A, Saadeh E, Alsalthi I, Alkhateeb H. Factors influencing unplanned icu readmissions: a retrospective cohort study. *Int J Africa Nurs Sci*. 2024;20, 100758. <https://doi.org/10.1016/j.ijans.2024.100758>.
2. Ponzoni CR, Corrêa TD, Filho RR, Serpa Neto A, Assunção MSC, Pardini A, et al. Readmission to the intensive care unit: incidence, risk factors, resource use, and outcomes. A retrospective cohort study. *Ann Am Thorac Soc*. 2017;14(8):1312–19. <https://doi.org/10.1513/AnnalsATS.201611-851OC>.
3. Jensen P, Jensen L, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13:395–405.
4. Goldstein B, Navar A, Pencina M, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inf Assoc*, ocw042. 2016;24.

5. Johnson AE, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):1–9. <https://doi.org/10.1038/sdata.2016.35>.
6. Fathy W, Emeriaud G, Cheriet F. A comprehensive review of ICU readmission prediction models: from statistical methods to deep learning approaches. *Artif Intell Med*. 2025;165, 103126 (<https://doi.org/10.1016/j.artmed.2025.103126>. <https://www.sciencedirect.com/science/article/pii/S0933365725000612>).
7. Lin Y-W, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS ONE*. 2019;14, e0218942.
8. Carvalho RM, Oliveira D, Pesquita C. Knowledge graph embeddings for ICU readmission prediction. *Bmc Med Inf Decis*. 2023;23:12.
9. Shee-trit E, Brief M, Elisha O. Predicting unplanned readmissions in the intensive care unit: a multimodality evaluation. *Sci Rep*. 2023;13, 15426.
10. Pakbin A, et al. Prediction of ICU readmissions using data at patient discharge. 2018 40th annual international conference of the IEEE engineering in medicine and biology society. 2018(EMBC). 4932–35.
11. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. *Sci Rep*. 2020;10(1):1111. <https://doi.org/10.1038/s41598-020-58053-z>.
12. Niu K, Pei S, Peng X, Zeng J, Zhang K. Intensive care unit readmission prediction with correlation enhanced multi-task learning. *Comput Electr Eng*. 2023;110, 108780. <https://doi.org/10.1016/j.compeleceng.2023.108780>.
13. Pei S, Niu K, Peng X, Zeng J. Readmission prediction with knowledge graph attention and RNN-based ordinary differential equations. *International conference on knowledge science, engineering and management*. 2021. 559–70.
14. Auer S, et al. Towards a knowledge graph for science. *Proceedings of the 8th international conference on web intelligence, mining and semantics*. 2018;(1–6).
15. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng*. 2017;PP:1–1.
16. Ehrlinger L, Wöß W. Towards a definition of knowledge graphs. *Semant (Posters, Demos, Success)*. 2016;48(2).
17. Kiryakov A, Popov B, Terziev I, Manov D, Ognyanoff D. Semantic annotation, indexing, and retrieval. *Web Semant: Sci, Serv Agents World Wide Web*. 2004;2:49–79.
18. Jovanović J, Bagheri E. Semantic annotation in biomedicine: the current landscape. *J Biomed Semant*. 2017;8:1–18.
19. Beaulieu-Jones B, Orzechowski P, Moore J. Mapping patient trajectories using longitudinal extraction and deep learning in the MIMIC-III critical care database. *Pac Symp Biocomputing, Pac Symp Biocomputing*. 2018;23:123–32.
20. Lu Q, Nguyen TH, Dou D. Predicting patient readmission risk from medical text via knowledge graph enhanced multiview graph convolution, 1990–1994. New York, NY, USA: Association for Computing Machinery; 2021. <https://doi.org/10.1145/3404835.3463062>.
21. Zhang J, Liang S, Sheng Y, Shao J. Temporal knowledge graph representation learning with local and global evolutions. *Knowl-Based Syst*. 2022;251, 109234. <https://doi.org/10.1016/j.knsys.2022.109234>.
22. Cai B, et al. Temporal knowledge graph completion: a survey. *arXiv preprint arXiv:2201.08236*;2022.
23. Carvalho RM, Teixeira AS, Pesquita C. Building a temporal knowledge graph for electronic health records. *CEUR Workshop Proceedings*. 2024.
24. Schrodt J, Dudchenko A, Knaup-Gregori P, Ganzinger M. Graph-representation of patient data: a systematic literature review. *J Med Syst*. 2020;44:86.
25. Kazemi SM, et al. Representation learning for dynamic graphs: a survey. *J Mach Learn Res*. 2020;21:2648–720.
26. Leblay J, Chekol MW. Deriving validity time in knowledge graph. *Companion proceedings of the the web conference 2018*. 2018. 1771–76.
27. García-Durán A, Dumančić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion. *arXiv preprint arXiv:1809.03202*. 2018.
28. Lacroix T, Obozinski G, Usunier N. Tensor decompositions for temporal knowledge base completion. *arXiv preprint arXiv:2004.04926*. 2020.
29. Li J, Su X, Gao G. TeAST: temporal knowledge graph embedding via archimedean spiral timeline. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2023. (15460–74). <https://aclanthology.org/2023.acl-long.862/>.
30. Ying R, et al. Simple but effective compound geometric operations for temporal knowledge graph completion. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2024(11074–86).
31. Cai L, et al. A survey on temporal knowledge graph: representation learning and applications. *arXiv preprint arXiv:2403.04782*. 2024.
32. Liu J, et al. Towards continual knowledge graph embedding via incremental distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024;8759–68.
33. Dong H, Wang P, Xiao M, Ning Z, Wang P, Zhou Y. Temporal inductive path neural network for temporal knowledge graph reasoning. *Artif Intel*. 2024;329, 104085. <https://doi.org/10.1016/j.artint.2024.104085>.

34. Xue Y, Klabjan D, Luo Y. Predicting icu readmission using grouped physiological and medication trends. *Artif Intell Med.* 2019;95:27–37. <https://doi.org/10.1016/j.artmed.2018.08.004>.
35. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking deep learning architectures for predicting readmission to the icu and describing patients-at-risk. *Sci Rep.* 2020;10:1111. <https://doi.org/10.1038/s41598-020-58053-z>.
36. Zebin T, Chaussalet TJ. Design and implementation of a deep recurrent model for prediction of readmission in urgent care using electronic health records. 2019 IEEE conference on computational intelligence in bioinformatics and computational biology. 2019. CIBCB). 1–5.
37. Junqueira ARB, Mirza F, Baig MM. A machine learning model for predicting icu readmissions and key risk factors: analysis from a longitudinal health records. *Health Technol (Berl).* 2019;9(3):297–309. <https://doi.org/10.1007/s12553-019-00329-0>.
38. Sherertz D, Tuttle M, Blois M, Erlbaum M. Intervocabulary mapping within the umls: the role of lexical matching. *Proceedings of the Annual Symposium on Computer Application in Medical Care.* 1988. 201.
39. He Y, Chen J, Antonyrajah D, Horrocks I. Biomedical ontology alignment with BERT. *Proceedings of the 16th International Workshop on Ontology Matching Co-Located with the 20th International Semantic Web Conference (ISWC 2021).* 2021. 1–12.
40. He Y, Chen J, Antonyrajah D, Horrocks I. Bertmap: a bert-based ontology alignment system. *Proceedings of the AAAI Conference on Artificial Intelligence.* 2022. 5684–91.
41. Golbreich C, Wallace EK, Patel-Schneider PF. Owl 2 web ontology language new features and rationale. W3C Work Draft, W3C (June 2009). 2009. <http://www.w3.org/TR/2009/WD-owl2-new-features-20090611>.
42. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. *International conference on machine learning.* 2016. 2071–80.
43. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. *Adv Neural Inf Process Syst.* 2013;26.
44. Yang B, Yih W-T, He X, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575.* 2014.
45. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied logistic regression.* John Wiley & Sons; 2013.
46. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
47. Chen T. Xgboost: a scalable tree boosting system. *Cornell University.* 2016.
48. Alghatani K, Ammar N, Rezgui A, Shaban-Nejad A. Precision clinical medicine through machine learning: using high and low quantile ranges of vital signs for risk stratification of icu patients. *IEEE Access.* 2022;10:52418–30. <https://doi.org/10.1109/ACCESS.2022.3175304>.
49. Moerschbacher A, He Z. Building prediction models for 30-day readmissions among icu patients using both structured and unstructured data in electronic health records. 2023 IEEE International Conference on Bioinformatics and Biomedicine. 2023. BIBM. 4368–4373.
50. Fathy W, Emeriaud G, Cheriet F. Prediction of icu readmission using lightgbm classifier. 2023 IEEE 20th Int Symp Biomed Imag (ISBI). 2023;1–4.
51. González-Nóvoa JA, Campanioni S, Busto L, Fariña J, Rodríguez-Andina JJ, Vila D, et al. Improving intensive care unit early readmission prediction using optimized and explainable machine learning. *Int J Environ Res Public Health.* 2023;20:3455. <https://doi.org/10.3390/ijerph20043455>.
52. Gesese GA, Biswas R, Alam M, Sack H. A survey on knowledge graph embeddings with literals: which model links better literal-ly? *Semant Web.* 2021;12:617–47.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ricardo M. S. Carvalho¹ · Andreia Reia Sofia Teixeira^{1,2,3,4} · Catia Pesquita^{1,2}

✉ Ricardo M. S. Carvalho
rmscarvalho@fc.ul.pt

¹ LASIGE, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

² BRAN LAB, Network Science Institute, Northeastern University London, London, UK

³ Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA

⁴ Kent and Medway School, Canterbury, UK