Neuro-Symbolic Federated Learning over Heterogeneous Data-Views: A Structured Approach to Distributive EHR Modelling

Soheila Molaei[1,2], Bahareh Fatemi[3], Anshul Thakur[1], Andrew Soltan[4], Fazle Rabbi[3], Andreas L. Opdahl[3], Kim Branson[5], Patrick Schwab[5], Danielle Belgrave[5], David A. Clifton[1,6]

[1]Department of Engineering Science, University of Oxford
[2]CeBAM, Nuffield Department of Medicine, University of Oxford
[3]Department of Information Science and Media Studies, University of Bergen
[4]Department of Oncology, University of Oxford
[5]GlaxoSmithKline, London, UK
[6]Oxford-Suzhou Institute of Advanced Research (OSCAR), Suzhou, China
{soheila.molaei, anshul.thakur, david.clifton}@eng.ox.ac.uk, {bahareh.fatemi, fazle.rabbi, andreas.opdahl}@uib.no, andrew.soltan@oncology.ox.ac.uk, {patrick.x.schwab, kim.m.branson, danielle.x.belgrave}@gsk.ai

Abstract

Federated learning (FL) enables privacy-preserving model training across distributed Electronic Health Records (EHRs), but its deployment remains limited by data-view heterogeneity, where institutions maintain incompatible local schemas. Most existing methods address this by enforcing flat, aligned data views, which require extensive cross-site preprocessing and manual harmonisation that often discards client-specific features, or by projecting inputs into a shared latent space, which sacrifices interpretability. We propose a modelling shift from conventional FL with vectorised inputs to a symbolic, relation-centric framework, where each client organises its EHR data as a structured, type-aware relational graph. This enables client-specific inference without requiring schema alignment and supports FL across heterogeneous data views. To model over these symbolic structures, we introduce an architecture that combines relation-aware message passing with a learnable feature relevance mechanism, jointly enabling accurate local predictions and client-specific interpretability while supporting parameter sharing across clients. Beyond strong performance on three real-world EHR datasets exhibiting data-view heterogeneity, we further show that our framework supports multimodal FL under modality-level heterogeneity. Using MC-MED, a publicly available multimodal emergency department dataset, we demonstrate that our method accommodates clients with partially missing modalities, highlighting its robustness and scalability in real-world clinical settings.

Introduction

Electronic Health Records (EHRs) offer a wealth of clinical insight, yet remain siloed within individual institutions due to strict privacy laws and data governance constraints (Sauer et al. 2022; Tayefi et al. 2021; Rieke et al. 2020). This fragmentation makes it nearly impossible to centralise data for training, limiting the development of clinical models that capture population-level trends and generalise across care settings. Federated Learning (FL) has emerged as a promising approach for training such models collaboratively across distributed EHRs, without requiring data to leave institutional boundaries (McMahan et al. 2017; Soltan et al. 2023; Sheller et al. 2020). Despite its effectiveness in preserving privacy, FL in clinical settings faces a distinctive challenge: *data-view heterogeneity*—the structural misalignment of available features and inconsistencies in the representation, scale, or units of clinical measurements (Molaei et al. 2024; Thakur et al. 2024). Figure 1 illustrates a typical data-view heterogeneity scenario. These structural mismatches, often driven by differences in available clinical services, coding standards and local data collection practices, render many state-of-the-art FL methods inapplicable unless data-views across clients are explicitly aligned. These challenges are particularly pronounced in low- and middle-income countries (LMICs), where healthcare infrastructure varies widely across regions, exacerbating the fragmentation and limiting participation in global FL initiatives (Thakur et al. 2024).

Most clinical FL studies address data-view heterogeneity by manually aligning the feature spaces across clients prior to training. However, this preprocessing step is often labour-intensive and error-prone, requiring expert knowledge of local data schemas. Moreover, to achieve uniformity, features that are missing at some sites are typically discarded altogether, leading to information loss and underutilisation of client-specific data. Such constraints not only limit the expressiveness of the resulting models but may also discourage institutions, particularly those with limited technical resources or non-standard data structures, from participating in FL collaborations.

Beyond manual alignment, a few recent studies have addressed data-view heterogeneity through algorithmic strategies. One class of solutions uses imputation or data augmentation to synthetically complete missing features across clients, enforcing input compatibility during training (Molaei et al. 2024). However, this can introduce artificial noise that distorts clinical signals and undermines model reliability. Another class of methods bypasses alignment entirely by learning shared latent representations across clients (Thakur

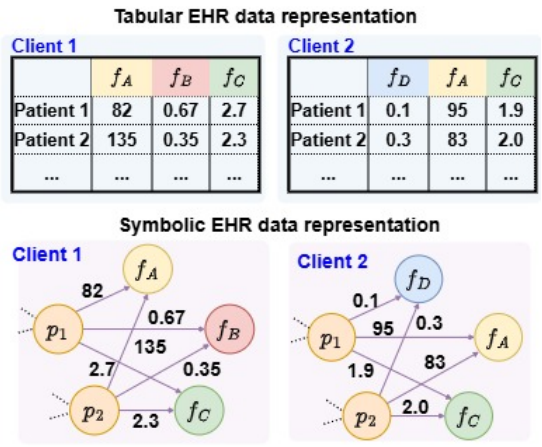

Figure 1: Data-view heterogeneity across clients. *Top:* Each institution maintains incompatible tabular schemas with non-overlapping features. *Bottom:* Symbolic transformation into knowledge graphs.

et al. 2024; Ye et al. 2023). While this abstraction facilitates global training over mismatched schemas, the latent modules are often trained locally, relying on client-specific data to encode clinical semantics, potentially resulting in inconsistent representations, obscured relationships, and reduced interpretability (Ahmad, Eckert, and Teredesai 2018; Allgaier et al. 2023).

To overcome the limitations of existing methods, this paper proposes a shift in how clinical data are represented. Rather than representing EHRs as fixed-length feature vectors, we model them as symbolic structures, specifically, heterogeneous knowledge graphs (Peng et al. 2023; Abu-Salih et al. 2023; Rotmensch et al. 2017) that encode patients and clinical variables as typed entities linked by semantic relationships (Figure 1). This symbolic perspective reflects the inherently relational nature of clinical data, where values such as blood pressure acquire meaning only in context, for example, in relation to medications, co-morbidities, or symptoms. Each client constructs its own knowledge graph from locally available data. While variations in local data lead to differences in graph topology across clients, all graphs conform to a shared ontology of entity and relation types. This unified representation addresses data-view heterogeneity without requiring manual input alignment or schema harmonisation, and can also accommodate modality heterogeneity, where clients may have access to different subsets of modalities, such as structured variables, clinical notes, or physiological signals, depending on local infrastructure.

Building on this foundation, we design a symbolic-neural hybrid architecture tailored to the characteristics of clinical data, which are often incomplete, irregular, and semantically structured. By operating on symbolic graphs, the same graph-based model can naturally adapt to client-specific feature sets and topologies, enabling decentralised training without requiring aligned input spaces or custom architectures. The shared ontology ensures semantic consistency across structurally distinct graphs, allowing global coordination through standard federated optimisation techniques such as Federated Averaging (FedAvg) and Federated SGD (FedSGD). Beyond structural flexibility, the proposed architecture also supports interpretability. The explicit graph schema provides transparency at the type level, and a learnable feature mask highlights the relevance of clinical variables on a per-client basis.

Method	Alignment Strategy	Interpretability
FedAvg	Manual	✗
Hypernet	Implicit	✗
LG-FedAvg	Projection	✗
AGAT-FL	Augmentation	✗
Knowledge Filtering	Projection	✗
NS-FL (Ours)	Semantic (Ontology)	✓

Table 1: Comparison of Neuro-symbolic FL with existing methods in terms of data-view heterogeneity and clinical interpretability.

This paper makes the following key contributions:

- Introduces a symbolic representation of EHRs as heterogeneous knowledge graphs, enabling FL across clients with data-view and modality heterogeneity.
- Proposes a neuro-symbolic architecture that performs relation-aware inference over client-specific graphs via attention-based message passing.
- Incorporates a feature-aware interpretability mechanism using learnable feature masking and graph semantics to identify clinically relevant variables.

Earlier Studies

Most FL frameworks assume consistent input schemas across clients, though some methods tolerate data-view heterogeneity as a by-product of their design (Yang et al. 2019; Zhu et al. 2021). For instance, Local-Global Federated Averaging (LG-FedAvg) addresses model heterogeneity by allowing clients to maintain distinct local architectures while sharing a global classification head (Liang et al. 2020; Ye et al. 2023). This setup incidentally enables projection into a shared latent space without input alignment. However, purely local representation learning often leads to inconsistent abstractions, poor knowledge transfer, and limited generalisation across divergent schemas. Hypernetwork-based FL (Shamsian et al. 2021) offers an alternative by training a central hypernetwork to generate personalised model weights based on each client's data view. While theoretically well-suited to input disparity, it introduces high training complexity, minimal inter-client knowledge sharing, and tight coupling to model architecture.

In contrast, a smaller set of methods are explicitly designed for data-view heterogeneity. The Augmented Graph Attention Network (AGAT) framework combines synthetic feature alignment with graph attention to prioritise informative features during message passing (Molaei et al. 2024). Though effective under schema mismatch, AGAT depends

on heuristic graph construction and synthetic augmentation, which may obscure clinical semantics and limit interpretability. The *knowledge abstraction and filtering* framework instead maps local features into a shared latent space using encoders conditioned on a global trainable knowledge vector (Thakur et al. 2024). This promotes consistency across clients and improves information transfer, but the reliance on learned abstractions may reduce transparency and fidelity to original clinical relationships.

COMPARISON TO THE PROPOSED METHOD: Unlike prior approaches that depend on latent projection or synthetic augmentation, the proposed method enables federated learning across heterogeneous client schemas by leveraging a shared ontology for semantic alignment, while also providing variable-level interpretability to identify clinically relevant features at each client site (see Table 1).

Method

This section presents the proposed neuro-symbolic framework for federated modelling across clients with non-aligned, heterogeneous data views. We first outline how each client represents its local EHR data using a symbolic relational graph and applies neural inference for local predictive modelling (Choi et al. 2017). We then describe how this modelling framework is integrated into a FedSGD-based protocol to support decentralised training.

Neuro-symbolic Modelling

Each federated client (i.e., participating medical institution) represents its local EHR data as a symbolic relational graph and employs a relation-aware neural architecture with built-in interpretability to process this graph for local prediction tasks:

Symbolic Data Modelling: We model clinical data as a symbolic system comprising typed entities and semantic relations, capturing the inherently relational structure of medical knowledge. Each client constructs a local knowledge graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ from its tabular dataset $\mathcal{D}_k$, where $v_p \in \mathcal{V}_k^{\text{patient}}$ denotes patient nodes and $v_f \in \mathcal{V}_k^{\text{feature}}$ denotes clinical feature nodes. The initial representation of each patient node is given by $h_{v_p}^{(0)} = \epsilon_p$, where $\epsilon_p \sim \mathcal{N}(0, I)$ is a random vector in $\mathbb{R}^{d_h}$. The initial representation of each feature node is given by $h_{v_f}^{(0)} = m_f \cdot \text{Emb}(f)$, where f denotes a discrete feature identifier (e.g., `"heart_rate"`), $\text{Emb}(f) \in \mathbb{R}^{d_h}$ is a trainable vector retrieved from a shared embedding table (described later), and $m_f \in [0, 1]$ is a learnable scalar mask.

To define the structure of the local graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$, we specify a lightweight relational schema that governs how nodes are connected. The edge set $\mathcal{E}_k$ comprises typed edges instantiated from the following three core clinical relations:

- `has_feature`(v_p, v_f): indicates that patient v_p exhibits feature f, with edge weights corresponding to the observed value.
- `of_patient`(v_f, v_p): the reverse of `has_feature`, enabling bidirectional information flow between patients and features.
- `similar_to`(v_p, v_p'): connects similar patients based on observed feature vectors using k-nearest neighbours, with weights reflecting similarity scores.

Each client instantiates these relations independently from its own tabular dataset $\mathcal{D}_k$, constructing the local graph $\mathcal{G}_k$ whose edges reflect the available observations. While the relational schema is globally defined and shared across all clients (see Figure 2), the resulting graph structure is client-specific, allowing flexible construction without requiring schema alignment and thereby accommodating data-view heterogeneity. This schema-driven design captures both the relational structure of EHRs and the computational needs of relation-aware GNNs: bidirectional edges between patients and features enable effective message passing, and similarity-based connections act as an inductive bias, encouraging patients with similar profiles to produce similar predictions. Together, these properties improve the robustness and generalisability of local models.

Relation-aware Graph Processing: To learn from the symbolic knowledge graph $\mathcal{G}_k$, we leverage a relation-aware graph neural network (GNN) that performs message passing over typed edges. To reflect differences in predictive signal, the architecture employs a heterogeneous attention mechanism that weighs both relation types and individual edges within each type (Veličković et al. 2017; Schlichtkrull et al. 2018). For example, among feature-related edges, the model may attend more to lymphocyte count than to total bilirubin in the context of COVID-19 prediction. This design enables the model to focus on the most informative relations and neighbours when updating patient representations.

This relation-aware architecture is implemented as a two-layer heterogeneous GNN that performs attention-based relation-specific updates. At each layer ℓ, the embedding of node $u \in \mathcal{V}_k$ is updated as:

$$h_u^{(\ell+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{v \in \mathcal{N}_r(u)} \alpha_{uv}^{(r)} \mathbf{W}^{(r)} h_v^{(\ell)}\right) \tag{1}$$

where $\mathcal{R} = \{\text{has\_feature}, \text{of\_patient}, \text{similar\_to}\}$ is the set of relation types, $\mathcal{N}_r(u)$ denotes the neighbours of u under relation r, $\mathbf{W}^{(r)} \in \mathbb{R}^{d_h \times d_h}$ is a relation-specific transformation, and σ is a nonlinearity (e.g., ReLU). The attention weight $\alpha_{uv}^{(r)}$ quantifies the importance of node v's message to node u under relation r, and is computed as:

$$e_{uv}^{(r)} = \text{LeakyReLU}\left(\mathbf{a}^{(r)\top}\left[\mathbf{W}^{(r)} h_u^{(\ell)} \,\|\, \mathbf{W}^{(r)} h_v^{(\ell)}\right]\right) \tag{2}$$

$$\alpha_{uv}^{(r)} = \frac{\exp(e_{uv}^{(r)})}{\sum_{v' \in \mathcal{N}_r(u)} \exp(e_{uv'}^{(r)})} \tag{3}$$

where $\mathbf{a}^{(r)} \in \mathbb{R}^{2d_h}$ is a relation-specific attention vector and $\|$ denotes concatenation.

After two rounds of message passing, the final embedding $h_{v_p}^{(L)} \in \mathbb{R}^{d_h}$ of each patient node v_p is used for prediction. A linear layer followed by a sigmoid activation produces the predicted outcome:

$$\hat{y}_p = \sigma\left(\mathbf{W}_{\text{out}} h_{v_p}^{(L)} + b\right) \tag{4}$$

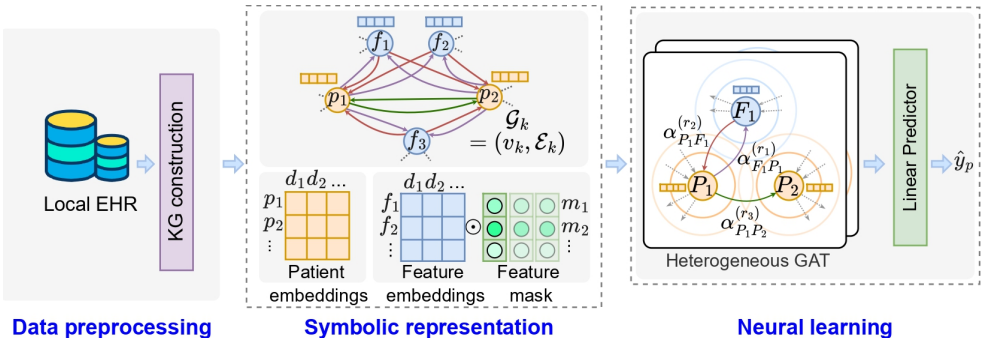

Figure 2: Overview of the proposed neuro-symbolic federated learning framework. Each client k constructs a local heterogeneous knowledge graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ from its private EHR dataset $\mathcal{D}_k$, with patients and clinical features represented as typed entities linked by semantic relations. Feature nodes use masked embeddings modulated by a learnable importance vector $\mathbf{m}$, and patient nodes are updated via relation-aware message passing using a heterogeneous GAT (Wang et al. 2019). The resulting embeddings are used for local outcome prediction.

where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{1\times d_h}$ and $b \in \mathbb{R}$ are learnable parameters. The model is trained end-to-end using binary cross-entropy loss computed over labelled patient nodes, optimising the GNN parameters, output layer, trainable embeddings and feature masks.

Feature-aware Interpretability: To address differences in feature availability and relevance across clients, we leverage the feature mask vector $\mathbf{m} = [m_1, \ldots, m_m]^\top$ introduced during graph construction. Each scalar $m_f \in [0, 1]$ modulates the initial embedding of feature node f, and is optimized during training to reflect the relative importance of that feature in the local context. These learned mask values serve as an intrinsic mechanism for interpretability, enabling the model to highlight which features contribute most to predictions at each site. At inference time, $\mathbf{m}$ can be visualized as a saliency map, offering client-specific insight into model behaviour without relying on post-hoc explanation methods.

Extension to Time-Series Data: The proposed symbolic modelling framework extends naturally to time-series data by treating each time step as a separate snapshot. This yields a sequence of symbolic graphs that capture temporal dynamics across time. At each step, a graph is constructed from observed feature–patient relations.

Federated Training Protocol

We incorporate the proposed neuro-symbolic modelling in a standard FedSGD setup, where server-side and client-side operations can be described as:

Server-side Processing: The server initiates the federated learning process by defining the global model architecture, initializing its parameters θ, and distributing them to all clients at the start of training. Before training begins, the server also constructs a *global feature dictionary* $\mathcal{F}$, which assigns each clinical feature name (e.g., `age`, `BP`, etc.) a unique integer index:

$$\mathcal{F} : \text{feature name} \mapsto \{0, 1, \ldots, |\mathcal{F}| - 1\}. \tag{5}$$

This dictionary is shared with all clients once during initialization and remains fixed throughout training. It provides a consistent indexing scheme for feature nodes, enabling clients to initialise feature node embeddings in a uniform manner, despite differences in local feature availability.

At the beginning of each communication round t, the server transmits the current global model parameters $\theta^{(t)}$ to all participating clients. After receiving local gradient updates $\Delta\theta_i$, the server performs a weighted aggregation, where each client's contribution is proportional to the size of its dataset, and updates the global model as:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \cdot \sum_{i=1}^{N} \frac{n_i}{n_{\text{total}}} \Delta\theta_i, \tag{6}$$

where η is the global learning rate, n_i is the number of samples at client i, and $n_{\text{total}} = \sum_{j=1}^{N} n_j$ is the total number of samples across all clients.

Client-side Processing Every kth client receives the initialised model parameters and the shared feature dictionary $\mathcal{F}$. Before federated training, the client computes its symbolic knowledge graph $\mathcal{G}_k$ from local dataset $\mathcal{D}_k$, as discussed earlier. Using dictionary $\mathcal{F}$, local feature names are mapped to indices for retrieving embeddings from a trainable table of size $|\mathcal{F}| \times d_h$. Only embeddings corresponding to locally observed features are accessed and updated, and each is modulated by a learnable scalar mask $m_f \in [0, 1]$ to enable feature-wise interpretability.

During the tth training round, the client receives the latest model parameters $\theta^{(t)}$ from the server. It initialises its local model as $\theta_k = \theta^{(t)}$ and trains it for multiple epochs over $\mathcal{G}_k$:

$$\theta_k^{'} = \theta_k - \nabla_{\theta_k} \mathcal{L}_k(\theta_k), \tag{7}$$

where $\mathcal{L}_k(\theta)$ is a task-specific objective function defined over the client's graph $\mathcal{G}_k$. Then, the resulting gradient for global model update is computed as: $\Delta\theta_k = \theta^{(t)} - \theta_k^{'}$, which is transmitted to the server at the end of training round.

Theoretical Analysis

This section formalises the guarantees offered by the proposed neuro-symbolic FL framework. We present three the-

orems that jointly establish: (i) invariance to schema heterogeneity, (ii) structural advantage of symbolic modelling, and (iii) generalisation benefits induced by the feature-aware masking mechanism.

Schema Invariance

We first establish that the proposed framework is robust to data-view heterogeneity.

Theorem 1 *Let $\mathcal{F}$ denote the global feature dictionary, and let $\mathcal{F}_k \subseteq \mathcal{F}$ be the subset observed at client k. Each client constructs a symbolic graph G_k from its local data and defines a local loss $\mathcal{L}_k(\theta)$ over predictions $g_\theta(G_k, p)$ for patient nodes $p \in \mathcal{V}_P$, where g_θ is relation-aware GNN parameterised by θ. Assume $\widetilde{G}_k$ is a schema-completed graph obtained by adding isolated nodes, without incident edges, for each unobserved feature $f \in \mathcal{F} \setminus \mathcal{F}_k$. Then the following holds:*

1. ***Forward-pass invariance:***

$$g_\theta(G_k, p) = g_\theta(\widetilde{G}_k, p)$$

2. ***Gradient invariance:***

$$\nabla_\theta \mathcal{L}_k(G_k) = \nabla_\theta \mathcal{L}_k(\widetilde{G}_k), \quad \nabla_\theta \mathcal{L}_k[f] = 0 \quad \forall f \notin \mathcal{F}_k.$$

Therefore, feature disparity does not affect the predictions or parameter updates, and the federated training converges identically without any schema alignment or imputation.

Optimal Risk under Symbolic Modelling

Next, we compare symbolic models to those that operate on vectorised inputs derived from tabular data, without capturing relational structure. Assuming a relational data-generating process, we show that symbolic predictors, which reason over typed graphs, achieve lower population risk under this setting.

Theorem 2 *Let $\mathcal{G}$ denote the space of typed symbolic graphs, and let $\mathcal{H}_{sym}$ be the class of node-level predictors that map $(G, p) \mapsto \hat{y}$, where $G \in \mathcal{G}$ and p is a patient node in G. Let $\mathcal{H}_{flat}$ denote the class of flat models that operate on vectorised representations $\Phi(G, p) \in \mathbb{R}^m$ obtained by flattening the neighbourhood of p in G. Assume the data-generating distribution $\mathcal{D}$ is over triples (G, p, y) with $G \in \mathcal{G}$, $p \in \mathcal{V}_P$, and label $y \in \mathcal{Y}$. Let $\mathcal{R}(h) := \mathbb{E}_{(G,p,y)\sim\mathcal{D}}[\ell(h(G, p), y)]$ denote the expected risk. Then:*

$$\inf_{h \in \mathcal{H}_{sym}} \mathcal{R}(h) \;\leq\; \inf_{h \in \mathcal{H}_{flat}} \mathcal{R}(h). \tag{8}$$

Feature Masking and Generalisation Bounds

Finally, we examine the generalisation impact of learnable feature masks. Beyond interpretability, the mask acts as an information bottleneck that limits overfitting. We provide an information-theoretic bound that connects generalisation error to the mutual information between original and masked features.

Theorem 3 *Let $(G, p, y) \sim \mathcal{D}$, where G is a symbolic graph, $p \in \mathcal{V}_P$ is a patient node, and $y \in \mathcal{Y}$ is its label. Let $x = \phi(G, p) \in \mathbb{R}^d$ denote a feature vector extracted from the neighbourhood of p in G, and let $\tilde{x} = m \odot x$, where $m \in [0, 1]^d$ is a learned or sampled masking vector. Assume f_θ is a deterministic predictor and the loss $\ell(f_\theta(\tilde{x}), y)$ is bounded in $[0, B]$. Then, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$ over n i.i.d. samples:*

$$\mathbb{E}_{(G,p,y)\sim\mathcal{D}}[\ell(f_\theta(\tilde{x}), y)] \;\leq\; \frac{1}{n}\sum_{i=1}^{n} \ell(f_\theta(\tilde{x}_i), y_i) + \sqrt{\frac{2B^2}{n} I(\tilde{X}; X) + \frac{2B^2 \log(1/\delta)}{n}}, \tag{9}$$

where X and $\tilde{X}$ denote the random variables corresponding to x and $\tilde{x}$ respectively, and $I(\tilde{X}; X)$ is the mutual information between them.

Experiments

Datasets

We evaluate the proposed framework on four healthcare datasets:

- CURIAL DATASETS: A collection of four datasets from distinct NHS Trusts, each considered a federated client, containing vital signs and blood test results for COVID-19 prediction (Soltan et al. 2023).
- EICU & MIMIC-III: eICU is a multi-centre ICU database (Pollard et al. 2018; Tang et al. 2020); we select the top 50 hospitals to simulate federated clients for 4-hour shock prediction using time-series data. MIMIC-III (Johnson et al. 2016; Harutyunyan et al. 2019) is used to simulate 15 clients with heterogeneous data schemas for 48-hour ICU mortality prediction.
- MC-MED: A multi-modal emergency department dataset partitioned into three clients with differing modality availability, used for early stroke prediction (Kansal et al. 2025).

The MC-MED dataset exhibits inherent data-view heterogeneity across clients. For the remaining datasets, we simulate it by randomly dropping features at selected clients.

Baselines & Performance Evaluation

We compare our framework against both federated and non-federated baselines. As a non-collaborative reference, each client independently trains a neural network on its local data. Among federated methods, we include FedAvg (McMahan et al. 2017), using a manually aligned feature subset to ensure a consistent input space, as well as heterogeneity-aware approaches, Hypernet (Shamsian et al. 2021), LG-FedAvg (Liang et al. 2020), AGAT (Molaei et al. 2024), and Knowledge Filtering (Thakur et al. 2024), which address schema and representation mismatches via distinct mechanisms. Performance is evaluated using AUROC and AUPRC, averaged across clients and over five runs, with standard deviations shown as error bars.

EVALUATION SCENARIOS: We evaluate the proposed framework alongside baseline methods under two federated settings: (i) *Standard*, which assesses performance in conventional FL scenarios with aligned feature spaces; and (ii)

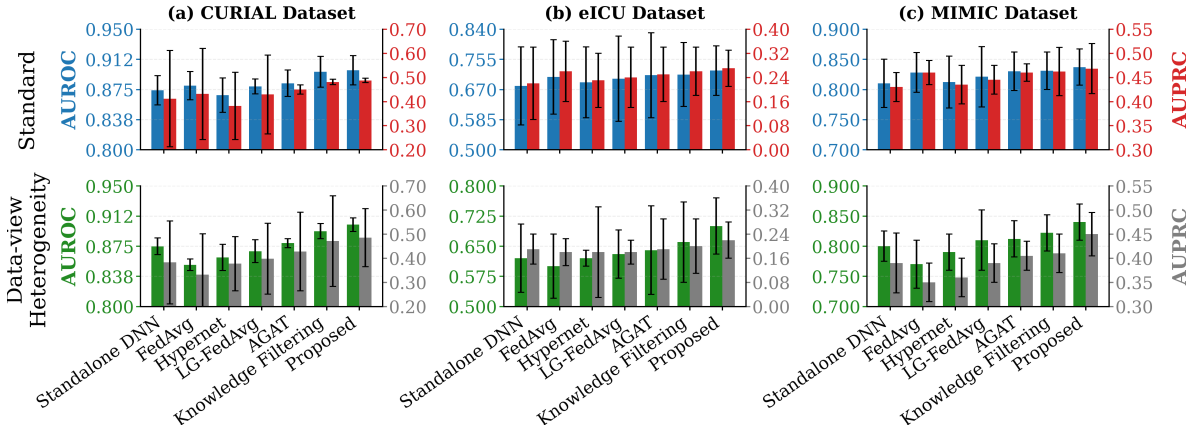

Figure 3: Performance comparison under standard and data-view heterogeneity scenarios across CURIAL, eICU and MIMIC-III datasets.

Data-view Heterogeneity, where clients possess differing local features, testing each method's ability to operate without manual harmonisation.

Results and Analysis

Performance on CURIAL, eICU, and MIMIC-III

The results across the CURIAL, eICU, and MIMIC-III datasets are presented in Figure 3. The proposed framework demonstrates consistently strong performance in both *Standard* and *Data-View Heterogeneity* scenarios. In the *Standard* setting, across all datasets, it achieves marginal yet consistent improvements over all baselines, showing that our symbolic and personalized design is fully compatible with standard FL and performs at least as well as established FL methods in aligned feature spaces.

Under the *Data-View Heterogeneity* scenario, FedAvg exhibits a consistent performance drop, even falling behind standalone models. This degradation is primarily due to feature loss incurred when aligning client data views, providing empirical support for the need for frameworks that explicitly address data-view heterogeneity. In contrast, The proposed method, along with other baselines designed to operate under heterogeneous data views, achieves substantial improvements over FedAvg–highlighting the overall importance of accommodating feature-view variability in federated settings.

Compared to other baselines equipped with mechanisms to handle data-view heterogeneity, the proposed framework maintains performance levels close to the Standard setting, demonstrating strong robustness to sparsity and feature non-overlap. Compared to Knowledge Filtering (strongest baseline), the proposed framework achieves relative improvements of 1.5% in AUROC and 2.1% in AUPRC on CURIAL, 5.0% in AUROC and 22.4% in AUPRC on eICU, and 1.2% in AUROC and 9.5% in AUPRC on MIMIC-III. These gains highlight the robustness of symbolic modelling and relation-aware inference compared to the latent projection and data augmentation strategies employed by existing baselines.

Performance on MC-MED

The MC-MED dataset presents a realistic federated setting in which each client observes a different subset of modalities. Clients are assigned combinations of static features, vital sign time series, and PPG waveforms: Client 1 has access to all three; Client 2 lacks PPG; and Client 3 relies solely on static features. To accommodate modality-specific processing, time series and waveform inputs are first embedded using pre-trained models before constructing symbolic graphs. This structured heterogeneity makes global feature alignment infeasible, as enforcing a common input space would require discarding modality-specific information, effectively reducing the task to a unimodal setting.

Interestingly, the proposed framework yields the largest absolute improvements on clients with fewer available modalities. On Clients 2 and 3, operating with two and one modality, respectively, we observe AUROC gains of 0.18 and 0.09, and AUPRC gains of 0.004 and 0.048 over the strongest baseline (Knowledge Filtering). These gains highlight the model's ability to maintain performance despite reduced input richness, enabled by flexible graph-based modelling that adapts to each client's available modalities. In contrast, Client 1, which has access to all three modalities, already performs well under baseline methods, yielding smaller gains of 0.01 in AUROC and 0.006 in AUPRC (Table 2). While the overall AUPRC remains low due to severe class imbalance, the consistent improvements underscore the framework's effectiveness in handling sparse and disjoint inputs in realistic federated settings.

Feature-Level Interpretability on CURIAL

As shown in Figure 4, our neuro-symbolic framework provides feature-level interpretability by assigning client-specific importance scores to clinical variables. While features such as oxygen delivery device and white cell count are consistently identified as important across multiple sites (BH, OUH, and PUH), each client also exhibits a distinct attribution profile shaped by its local data distribution. For instance, UHB assigns higher importance to alkaline phos-

Client	Metric	Standalone DNN	FedAvg	Hypernet	LG-FedAvg	AGAT	Knowledge Filtering	Proposed
Client 1	AUROC	0.700 ± 0.010	0.714 ± 0.011	0.692 ± 0.009	0.699 ± 0.010	0.789 ± 0.010	0.804 ± 0.009	**0.813 ± 0.004**
	AUPRC	0.011 ± 0.003	0.018 ± 0.002	0.017 ± 0.002	0.010 ± 0.002	0.012 ± 0.002	0.018 ± 0.002	**0.024 ± 0.001**
Client 2	AUROC	0.610 ± 0.015	0.669 ± 0.017	0.698 ± 0.014	0.603 ± 0.015	0.645 ± 0.013	0.542 ± 0.018	**0.721 ± 0.005**
	AUPRC	0.003 ± 0.002	0.004 ± 0.001	0.005 ± 0.001	0.007 ± 0.001	0.006 ± 0.001	0.004 ± 0.001	**0.008 ± 0.001**
Client 3	AUROC	0.710 ± 0.012	0.716 ± 0.014	0.707 ± 0.013	0.722 ± 0.012	0.724 ± 0.011	0.765 ± 0.015	**0.851 ± 0.003**
	AUPRC	0.011 ± 0.004	0.013 ± 0.002	0.012 ± 0.003	0.019 ± 0.004	0.013 ± 0.004	0.010 ± 0.002	**0.059 ± 0.001**
Average	AUROC	0.673 ± 0.009	0.700 ± 0.011	0.699 ± 0.011	0.675 ± 0.011	0.719 ± 0.011	0.704 ± 0.014	**0.795 ± 0.004**
	AUPRC	0.008 ± 0.003	0.012 ± 0.002	0.011 ± 0.002	0.012 ± 0.002	0.011 ± 0.003	0.011 ± 0.001	**0.030 ± 0.001**

Table 2: Performance (mean ± std) of all methods on the MC-MED dataset under modality-level heterogeneity. Each row shows results for a client-metric pair. Bold indicates best-performing method.

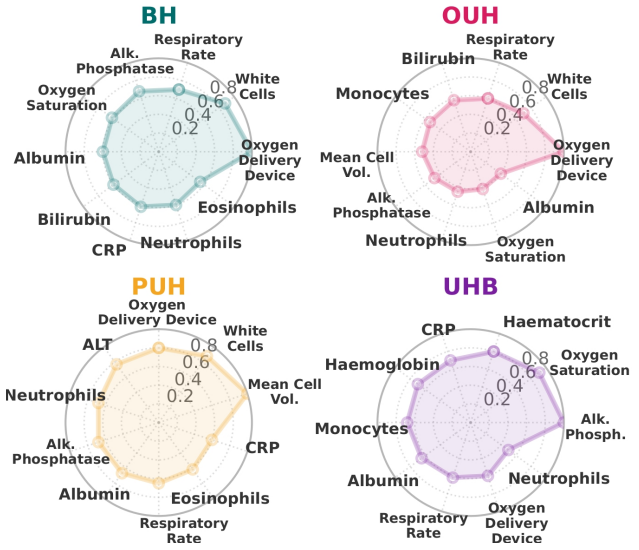

Figure 4: Client-specific radar plots of learned feature importance.

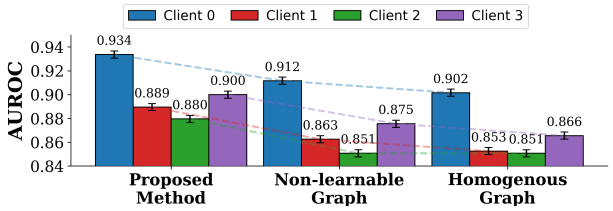

Figure 5: Client-wise validation AUROC comparison across graph types.

phatase, oxygen saturation, and haematocrit, reflecting patterns unique to its patient population or clinical practices. In contrast, PUH presents a more uniformly distributed profile, indicating reliance on a broader set of features. These variations demonstrate the model's ability to personalize inference without sacrificing global consistency. Importantly, the repeated identification of certain variables across sites underscores their shared clinical relevance, while the local variation highlights the framework's capacity to adapt to real-world heterogeneity without requiring centralised schema alignment.

Impact of Symbolic Structure and Learnable Relations

We conduct an ablation study to assess the contribution of symbolic structure and learnable relational inference in our framework by evaluating three graph construction strategies across clients on the CURIAL datasets: (1) our proposed method, which uses heterogeneous knowledge graphs with learnable feature embeddings and relation-specific attention; (2) a non-learnable variant that retains symbolic structure but disables learned embeddings and attention; and (3) a homogeneous graph baseline that flattens the structure by removing semantic typing and treating all nodes and edges as identical. As shown in Figure 5, the proposed method consistently achieves the highest AUROC on all clients. Performance degrades when learnable feature relevance and relation weighting are disabled, highlighting the utility of adaptive message passing. The largest performance drop is observed in the homogeneous graph variant, confirming that preserving symbolic heterogeneity is critical for effective learning in settings with complex, client-specific data schemas. A similar trend is observed for the *MC-MED* dataset.

Conclusion

This paper introduced a neuro-symbolic federated learning framework for robust and interpretable clinical modelling across clients with heterogeneous data-views. Each client's EHR is represented as a typed knowledge graph, enabling relation-aware message passing with learnable feature relevance and federated training without schema alignment. Experiments on four clinical datasets show consistent gains, especially for clients with sparse modalities. The symbolic design preserves clinical semantics, enhances client-specific interpretability, and lowers the barrier to FL participation, especially for resource-constrained sites. While the primary focus was addressing data-view heterogeneity, the framework remains fully compatible with privacy-preserving techniques such as secure aggregation and differential privacy. Future work will explore ontology-guided graph construction and better calibration for imbalanced outcomes to advance equitable and trustworthy clinical AI.

Acknowledgments

DAC was funded by an NIHR Research Professorship; a Royal Academy of Engineering Research Chair; and the InnoHK Hong Kong Centre for Cerebro-cardiovascular Engineering (COCHE); and was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) and the Pandemic Sciences Institute at the University of Oxford.

References

Abu-Salih, B.; Al-Qurishi, M.; Alweshah, M.; Al-Smadi, M.; Alfayez, R.; and Saadeh, H. 2023. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data*, 10(1): 81.

Ahmad, M. A.; Eckert, C.; and Teredesai, A. 2018. Interpretable Machine Learning in Healthcare. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '18, 559–560. New York, NY, USA: Association for Computing Machinery. ISBN 9781450357944.

Allgaier, J.; Mulansky, L.; Draelos, R. L.; and Pryss, R. 2023. How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. *Artificial Intelligence in Medicine*, 143: 102616.

Choi, E.; Bahadori, M. T.; Song, L.; Stewart, W. F.; and Sun, J. 2017. GRAM: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 787–795.

Harutyunyan, H.; Khachatrian, H.; Kale, D. C.; Steeg, G. V.; and Galstyan, A. 2019. Multitask Learning and Benchmarking with Clinical Time Series Data. *Scientific Data*, 6(96): 1–18.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3.

Kansal, A.; Chen, E.; Jin, B. T.; Rajpurkar, P.; and Kim, D. A. 2025. MC-MED, multimodal clinical monitoring in the emergency department. *Scientific Data*, 12(1): 1094.

Liang, P. P.; Liu, T.; Ziyin, L.; Allen, N. B.; Auerbach, R. P.; Brent, D.; Salakhutdinov, R.; and Morency, L.-P. 2020. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Molaei, S.; Thakur, A.; Niknam, G.; Soltan, A.; Zare, H.; and Clifton, D. A. 2024. Federated learning for heterogeneous electronic health records utilising augmented temporal graph attention networks. In *International Conference on Artificial Intelligence and Statistics*, 1342–1350. PMLR.

Peng, C.; Xia, F.; Naseriparsa, M.; and Osborne, F. 2023. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11): 13071–13102.

Pollard, T. J.; Johnson, A. E.; Raffa, J. D.; Celi, L. A.; Mark, R. G.; and Badawi, O. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data*, 5(1): 1–13.

Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H. R.; Albarqouni, S.; Bakas, S.; Galtier, M. N.; Landman, B. A.; Maier-Hein, K.; et al. 2020. The future of digital health with federated learning. *NPJ digital medicine*, 3(1): 119.

Rotmensch, M.; Halpern, Y.; Tlimat, A.; Horng, S.; and Sontag, D. 2017. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1): 5994.

Sauer, C. M.; Chen, L.-C.; Hyland, S. L.; Girbes, A.; Elbers, P.; and Celi, L. A. 2022. Leveraging electronic health records for data science: common pitfalls and how to avoid them. *The Lancet Digital Health*, 4(12): e893–e898.

Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, 593–607. Springer.

Shamsian, A.; Navon, A.; Fetaya, E.; and Chechik, G. 2021. Personalized federated learning using hypernetworks. In *International conference on machine learning*, 9489–9502. PMLR.

Sheller, M. J.; Edwards, B.; Reina, G. A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Colen, R. R.; et al. 2020. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1): 12598.

Soltan, A. A.; Thakur, A.; Yang, J.; Chauhan, A.; D'Cruz, L. G.; Dickson, P.; Soltan, M. A.; Thickett, D. R.; Eyre, D. W.; Zhu, T.; et al. 2023. Scalable federated learning for emergency care using low cost microcomputing: Real-world, privacy preserving development and evaluation of a COVID-19 screening test in UK hospitals. *medRxiv*, 2023–05.

Tang, S.; Davarmanesh, P.; Song, Y.; Koutra, D.; Sjoding, M. W.; and Wiens, J. 2020. Democratizing EHR analyses with FIDDLE: a flexible data-driven preprocessing pipeline for structured clinical data. *Journal of the American Medical Informatics Association*, 27(12): 1921–1934.

Tayefi, M.; et al. 2021. Challenges in implementing machine learning for healthcare: A systematic review. *Journal of Healthcare Informatics*, 5(2): 123–130.

Thakur, A.; Molaei, S.; Nganjimi, P. C.; Liu, F.; Soltan, A.; Schwab, P.; Branson, K.; and Clifton, D. A. 2024. Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare. *npj Digital Medicine*, 7(1): 283.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference*, 2022–2032.

Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2): 1–19.

Ye, M.; Fang, X.; Du, B.; Yuen, P. C.; and Tao, D. 2023. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3): 1–44.

Zhu, H.; Xu, J.; Liu, S.; and Jin, Y. 2021. Federated learning on non-IID data: A survey. *Neurocomputing*, 465: 371–390.