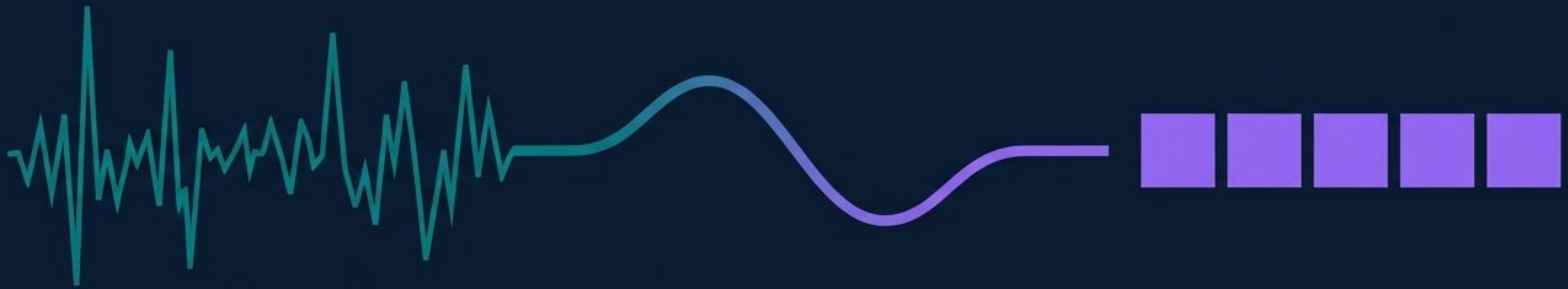
Can we generate portable representations for clinical time series data using LLMs?

Record2Vec: A Summarize-then-Embed Pipeline for ICU Data



Zongliang Ji, Yifei Sun, Andre Amaral,
Anna Goldenberg, Rahul G. Krishnan

University of Toronto, Vector Institute,
Sunnybrook Health Sciences Centre

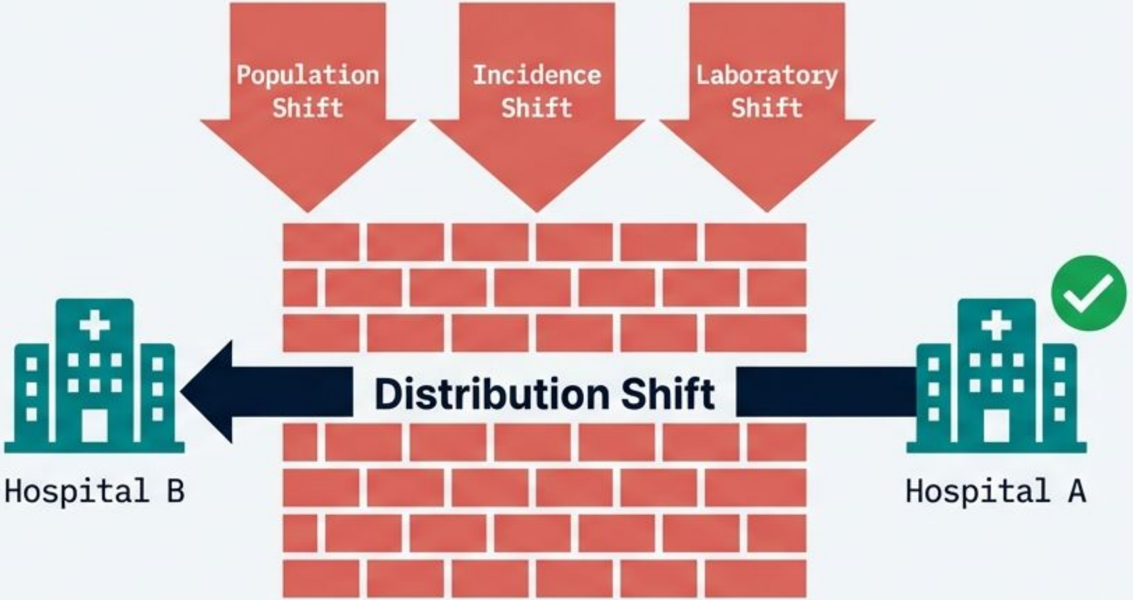
ICLR 2026

Clinical machine learning is trapped in a one site at a time deployment loop

Models built at Hospital A degrade rapidly when moved to Hospital B.

The culprits: Changing case mixes, sensor schemas, and missingness patterns.

The result: Endless recalibration, delayed validation, and stunted patient impact.



Human doctors survive distribution shift through semantic handoffs

Traditional ML

- ✅ Accessible
- ❌ Semantics
- ❌ Portable

Human Doctor

- ✅ Semantics
- ✅ Clinical
- ⚠️ Scalable / Busy

LLM (Record2Vec)

- ✅ Accessible
- ✅ Semantics
- ✅ Portable

Portable input representations enable portable models.

Instead of forcing heterogeneous hospitals into a shared syntactic format, we map them into a shared semantic space.

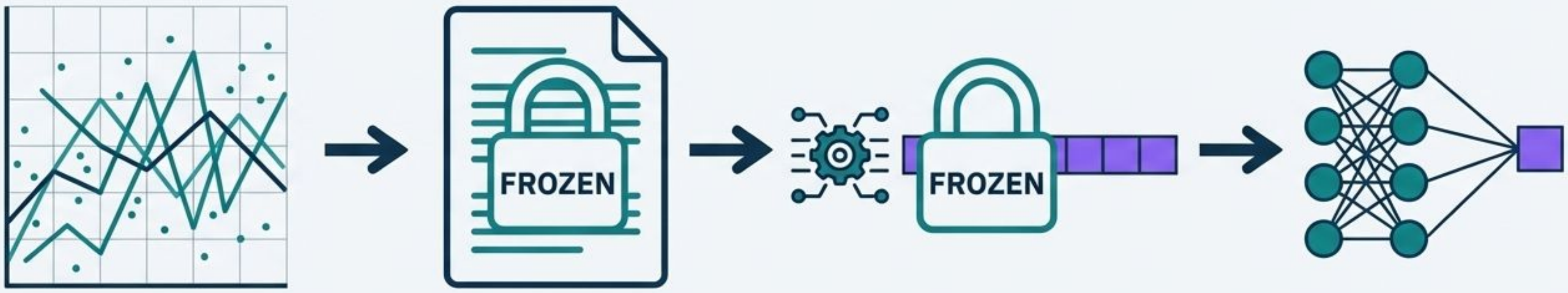
The Record2Vec Pipeline: Summarize-then-Embed

1. Irregular Patient Record
48-hour messy time-series.

2. LLM Summarizer
Gemini / MedGemma / Llama

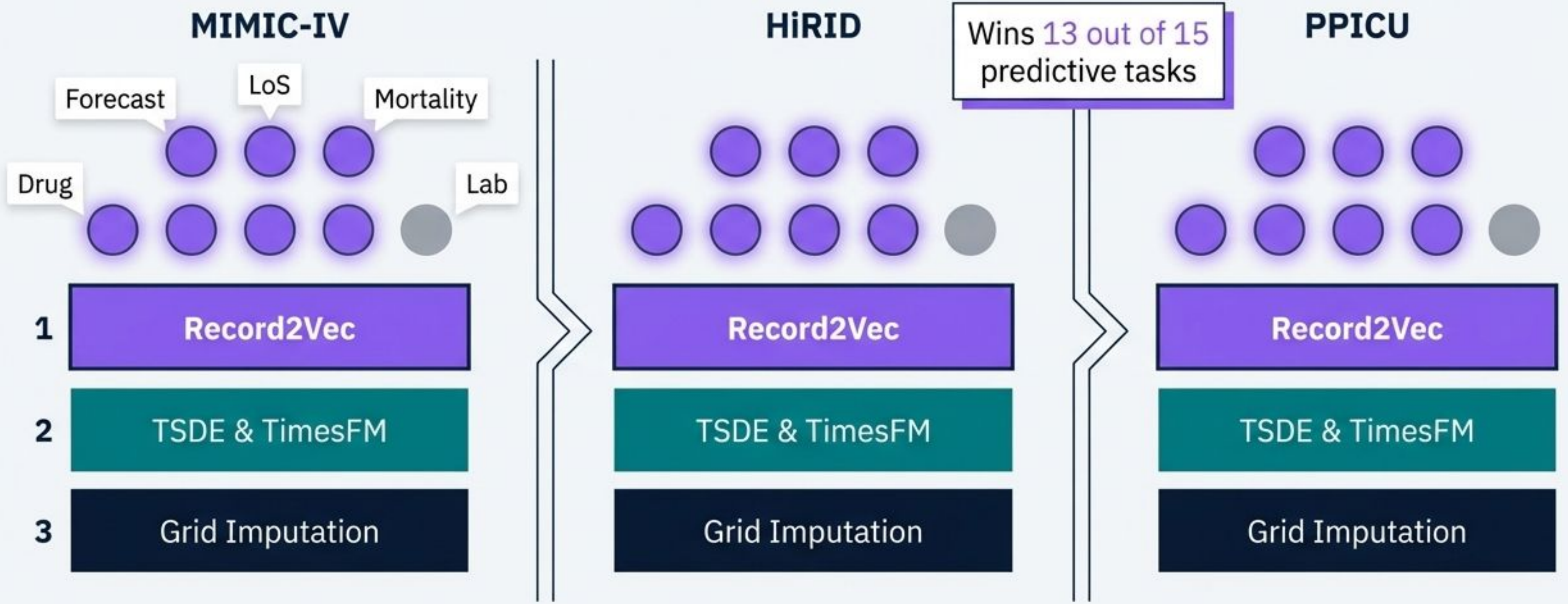
3. Text Encoder
Qwen3

4. Shared Predictor
Standard downstream heads.



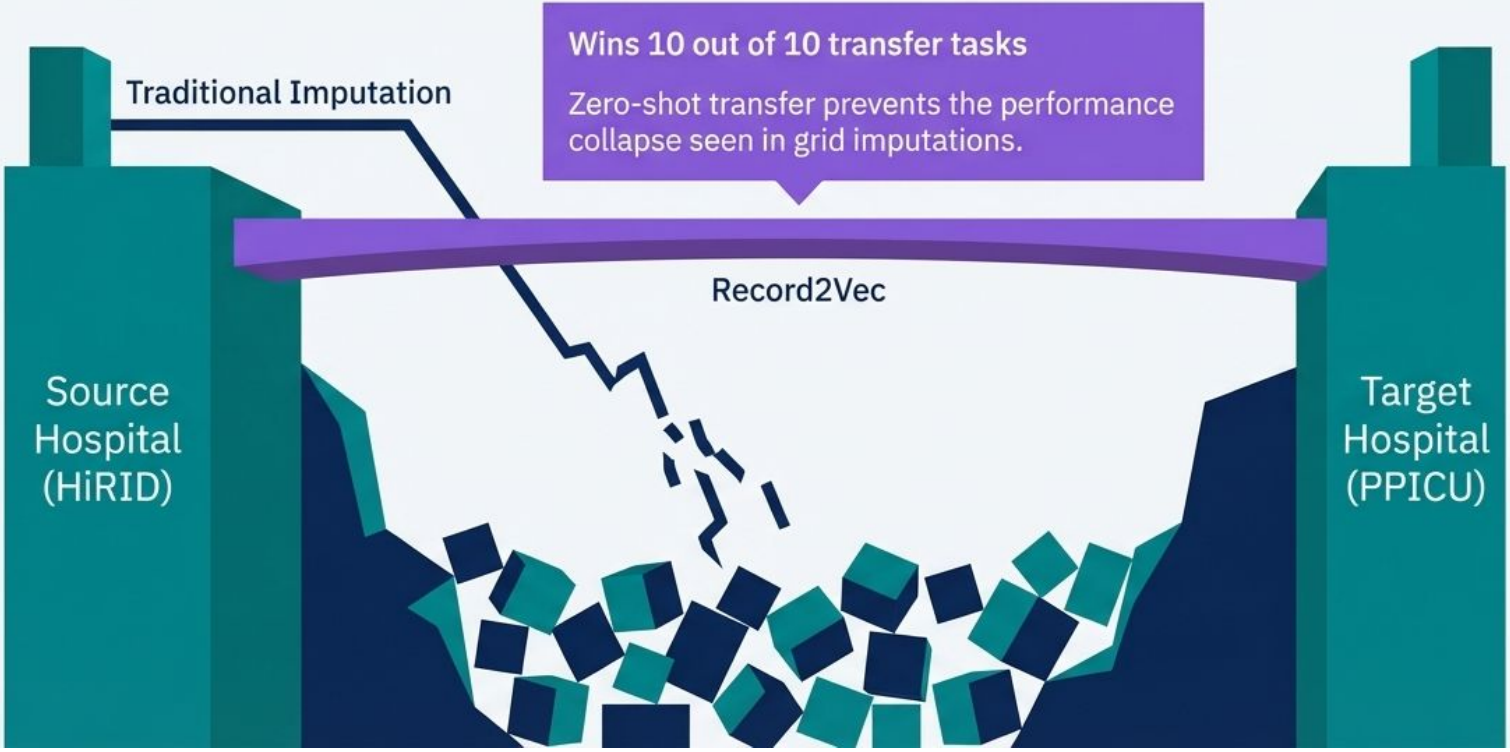
Record2Vec maintains competitive in-distribution accuracy

Adding a language layer preserves critical clinical signals for forecasting and classification, outperforming traditional imputation and competing with foundation models.



Summarization unlocks unprecedented cross-site portability

Natural language aligns heterogeneous units, names, and sampling habits into a shared clinical language before embedding.

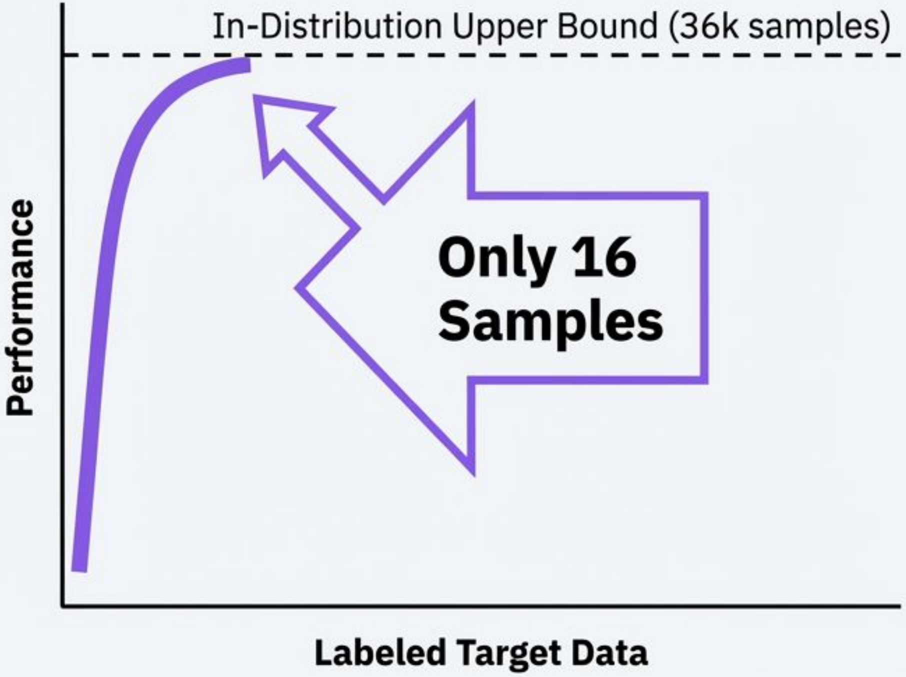


Finetuning with just 16 patient samples recovers target distribution performance

Ideal for smaller healthcare centers.

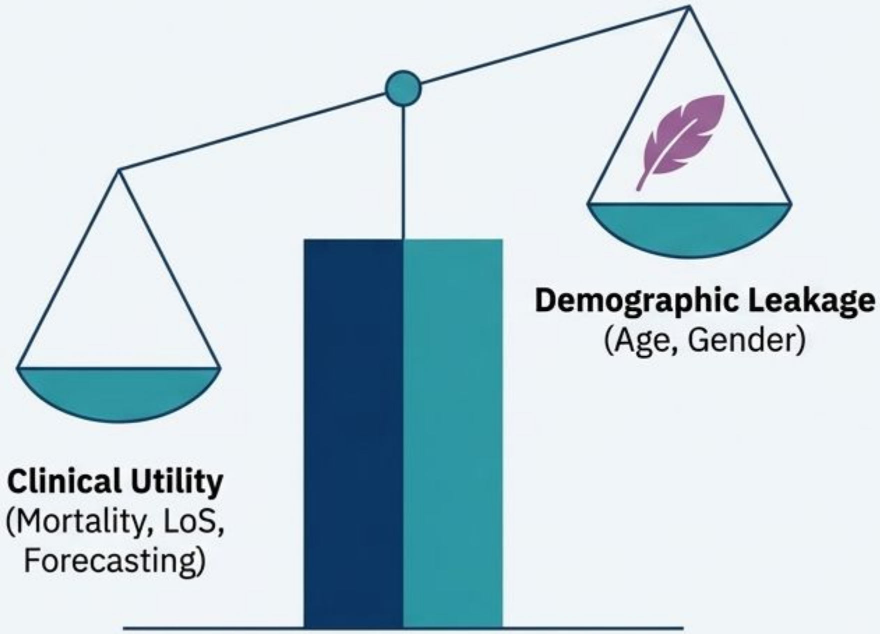
Record2Vec leverages source pre-training and requires minimal labeled target data to calibrate task boundaries.

Outperforms all competing methods under the same tight supervision budget.



High clinical utility does not come at the cost of increased demographic leakage

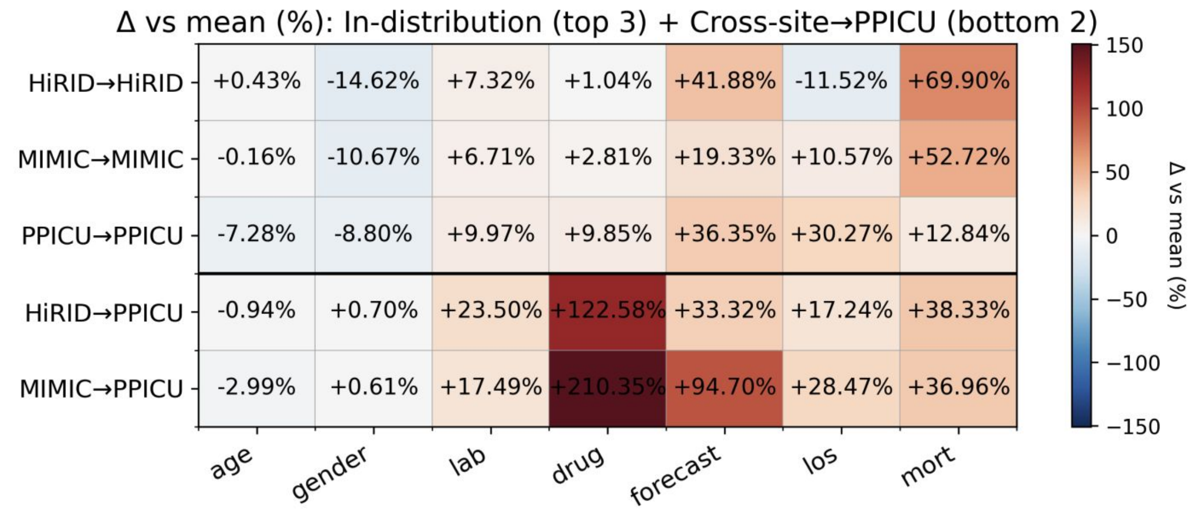
The language transformation layer emphasizes clinical states and trends, deliberately attenuating signals tied to age and gender.



Gender recoverability collapses to baseline chance across all methods.

- Age prediction error is on par with classical imputation baselines.

What information obtained in embeddings?



Record2Vec captures semantically meaningful structure that improves in-distribution performance and, importantly, encodes cross-site invariant patterns, while attenuating demographic attributes.

The Three Pillars of the Semantic Bridge

1. Semantic Abstraction: Normalizes units and resolves synonyms automatically, creating cross-site invariant patterns.

2. Stable Interface: Fixed-length vectors completely ignore missingness patterns that instantly break grid imputations.

3. Biological Relevance: Captures clinical trajectory and patient state, not just numeric correlation.

Rethinking healthcare ML: from site-specific engineering to portable semantics

Open Review Link

- LLMs serve as a powerful information transformation layer for numeric data.
- Record2Vec drastically reduces per-site engineering and enables robust zero/few-shot transfer.
- By focusing on portable inputs, we can move clinical ML out of the lab and into hospitals worldwide.