Hypergraph-based Pre-training for Atrial Fibrillation Prediction in Patients with Embolic Stroke of Undetermined Source

Presented at the International Stroke Conference (ISC), 2026

Yuzhang Xie, Yuhua Wu, Ruiyu Wang, Xiao Hu, Carl Yang
Emory University, Atlanta, GA, USA
{yuzhang.xie, gloria.wu, ruiyu.wang, xiao.hu, j.carlyang}@emory.edu

I. Background

Embolic stroke of undetermined source (ESUS) carries a high risk of later atrial fibrillation (AF), yet early detection remains difficult. Common clinical scores (e.g., CHA_2DS_2-VASc) depend on limited variables and often fail across diverse populations. Learning directly from electronic health records (EHRs) is promising but constrained by the small size of ESUS cohorts and the high dimensionality of diagnostic codes. We hypothesized that hypergraph-based pre-training on a larger stroke cohort could yield compact, transferable patient embeddings to improve AF prediction in ESUS.

II. Methods

Diagnostic codes were modeled as a hypergraph, with vertices as diagnostic features and each patient visit as a hyperedge, enabling higher-order co-occurrence to be captured. Two pre-training strategies were applied on a stroke cohort (n = 7,780; 1,735 PSCI cases; 2,595 features; treated at Emory University Hospital in 2012–2021). (1) Supervised hypergraph transformer: trained to predict PSCI via cross-entropy loss, encouraging the encoder to learn predictive clinical structures. (2) Unsupervised contrastive pre-training: for each patient, two stochastic views were generated through code dropout, time-window sampling, and mild hyperedge perturbation. A hypergraph transformer encoded each view, and InfoNCE losses at node, hyperedge, and membership levels promoted agreement and invariance to produce robust embeddings. The pretrained encoders were transferred to the ESUS cohort (n = 510; 107 AF cases; 2015–2023) to generate 32-D patient embeddings, concatenated with 58 baseline features (demographics, biomarkers, ECG measures, and comorbidities). Logistic regression, random forest, and gradient boosting classifiers were trained using five-fold nested cross-validation for unbiased evaluation.

III. Results

Compared to models trained from raw features, hypergraph pre-training improved AUROC by about 5–15% across classifiers and reduced overfitting variance from high feature dimensionality. In ESUS-AF, the best AUROC from hypergraph pipelines reached 0.66, versus 0.45–0.60 for from-scratch models. Performance remained stable when training data were reduced to 20–80%, demonstrating data efficiency of learned embeddings.

IV. Conclusion

Hypergraph-based pre-training produces compact, transferable EHR embeddings that enhance AF prediction in ESUS despite limited data and high feature dimensionality. This approach supports earlier AF detection and targeted monitoring in ESUS patients.

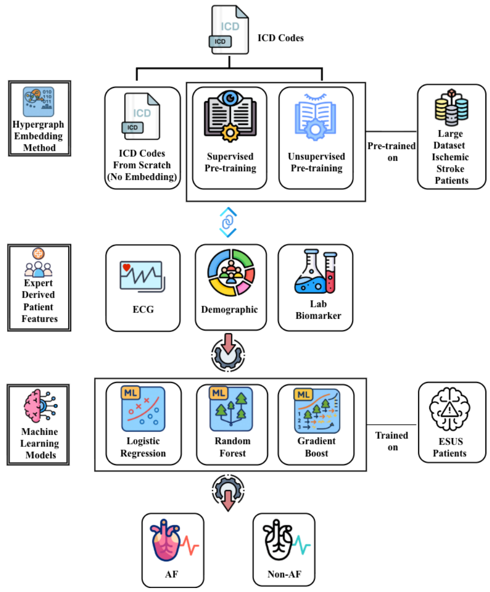

Fig. 1. Overview of our methodology.

Embedding	Model	AUROC	ACC	F1	R	P
From-scratch	LR	0.450± 0.041	0.698± 0.078	0.106± 0.096	0.113± 0.107	0.100± 0.088
From-scratch	RF	0.600± 0.063	0.790± 0.005	0.000± 0.000	0.000± 0.000	0.000± 0.000
From-scratch	GB	0.561± 0.055	0.776± 0.024	0.098± 0.091	0.065± 0.063	0.270± 0.248
Suprevised	LR	0.591± 0.052	0.624± 0.051	0.357± 0.059	0.494± 0.073	0.280± 0.050
Suprevised	RF	0.603± 0.016	**0.794± 0.006**	0.036± 0.044	0.019± 0.023	**0.400± 0.490**
Suprevised	GB	0.566± 0.032	0.751± 0.034	0.096± 0.098	0.074± 0.075	0.140± 0.150
Unsupervised	LR	0.600± 0.067	0.608± 0.035	**0.367± 0.052**	**0.541± 0.073**	0.278± 0.041
Unsupervised	RF	**0.660± 0.059**	0.782± 0.007	0.016± 0.032	0.010± 0.019	0.050± 0.100
Unsupervised	GB	0.599± 0.040	0.773± 0.037	0.165± 0.148	0.119± 0.109	0.297± 0.230

Fig. 2. AF prediction performance comparison. "Embedding Method" describes how we generate the patient representation "ML Model" represents how we train the AF prediction model ("LR" for logistic regression, "RF" for random forest, and "GB" for gradient boosting tree). Values are mean±standard deviation over 5 runs. Bold indicates the best value within the embedding group.

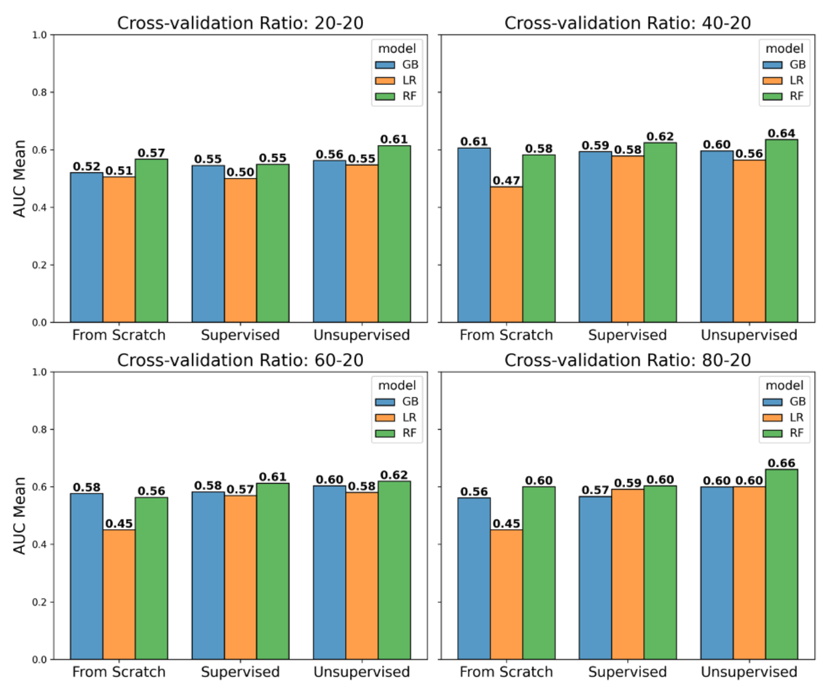

Fig. 3. Impact of training dataset size on model performance.