# Enhanced Atrial Fibrillation Prediction in ESUS Patients with Pre-training and Transfer Learning

Yuhua Wu[1*], Yuzhang Xie[2*], Ruiyu Wang[2], Fadi Nahab[1], Xiao Hu[1†], and Carl Yang[2†]

[1]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta GA, USA
[2]Department of Computer Science, Emory University, Atlanta GA, USA

**Abstract.** Atrial fibrillation (AF) is a major complication following an embolic stroke of an undetermined source (ESUS), increasing the risk of recurrent stroke and mortality. Early AF detection enables timely intervention, but traditional methods, such as clinical scoring systems and loop recorders, have limitations in accuracy, scalability, and cost. Machine learning (ML) offers potential solutions but struggles with data scarcity and high-dimensional medical features. To address these challenges, we propose to leverage pre-training and transfer learning approaches for AF prediction in ESUS patients. We first pre-train hypergraph-based patient embedding models on a large stroke cohort (AI-RESPECT, 7780 samples) to capture key patient features and feature interactions. These learned embeddings are then transferred to a smaller ESUS cohort (510 samples), significantly reducing feature dimensionality while retaining crucial patient information. This enables effective AF prediction using lightweight ML models. Experiments show that our proposed supervised and unsupervised pre-training approaches both outperform traditional ML models applied directly to raw data, improving prediction accuracy and robustness. By leveraging pre-training and transfer learning, our methods effectively mitigate the challenges of small sample sizes and high-dimensional data, offering a scalable and efficient solution for AF prediction in stroke patients.

**Keywords:** Atrial fibrillation · Stroke · Transfer learning · Pre-training.

## 1 Introduction

Embolic stroke of undetermined source (ESUS) is an ischemic stroke subtype defined by the absence of an identifiable cause. It is characterized by a non-lacunar brain infarct without an evident embolic source [2]. Notably, approximately 30% of ESUS patients develop atrial fibrillation (AF) within six months of their initial stroke, despite having no prior history of AF [6]. This post-ESUS AF is marked by rapid, irregular heart rhythms that lead to the formation of embolic clots, thereby increasing the risk of recurrent stroke and hospital readmission

* Equal contribution as co-first authors.
† Corresponding authors: xiao.hu@emory.edu; j.carlyang@emory.edu.

[7]. Moreover, studies indicate that post-ESUS AF significantly contributes to stroke-related mortality, with death rates surpassing those observed in patients with pre-existing AF [3]. Therefore, accurately predicting AF in ESUS patients is crucial, as it enables clinicians to identify individuals at high risk and implement targeted interventions, ultimately improving the quality of post-stroke care and reducing both mortality and morbidity.

Traditional methods for predicting AF include loop recorders and clinical scoring systems such as C2HEST and CHA2DS2-VASc[11][4]. Although clinical scoring systems are accessible and easy to implement, they rely on a limited set of clinician-selected predictors, which may not capture the full complexity of patient data. In contrast, loop recorders offer continuous and detailed cardiac rhythm monitoring, providing a more robust means of detecting AF. However, their use comes with notable limitations, including high costs, invasive procedures, and limited feasibility for widespread implementation, especially in resource-constrained healthcare settings. These limitations highlight the need for more advanced methods that can efficiently integrate diverse and comprehensive patient data for AF prediction.

To address the need for comprehensive patient data in AF prediction, researchers have increasingly utilized International Classification of Diseases (ICD) codes. These standardized diagnostic codes offer extensive, structured information about a patient's medical history, comorbidities, and treatment patterns [13]. ICD codes also enable the identification of meaningful patterns and associations that may not be apparent through traditional clinical scoring systems [5]. These advancements are crucial for facilitating more effective AF detection.

Machine learning (ML) has emerged as a powerful tool for predicting various diseases, including AF. Unlike traditional methods that rely on predefined rules or clinician-selected predictors, ML models excel at processing and analyzing large, complex datasets to uncover patterns beyond human interpretation. Researchers have widely applied ML techniques, such as logistic regression, random forests, and gradient boosting trees, to AF prediction tasks [12][8][1]. These methods have achieved impressive performance for AF prediction. However, these traditional ML methods often face significant data issues when applied to medical datasets, including:

- (1) Sample scarcity: Many medical datasets have limited sample sizes or few labeled examples, especially in rare diseases. This scarcity can make supervised learning impractical, leading to overfitting and reduced generalizability.
- (2) High Dimensionality: Comprehensive patient data, such as ICD codes, often involve thousands of features. This "curse of dimensionality" increases computational complexity and can cause overfitting, particularly when combined with limited sample sizes.

The same challenges arise when predicting AF in our cohort of ESUS patients treated at Emory Healthcare due to the small sample size (510 samples) and high input dimensionality (over 1,000 diagnostic features including ICD). To overcome these challenges, we propose novel methods that leverage pre-training and transfer learning techniques. By pre-training on a larger dataset, AI-RESPECT (7780

samples), with similar features and related labels, we develop generalized models that capture essential patterns using unsupervised and supervised approaches. These pre-trained models are then "transferred" to our smaller ESUS dataset by generating compact representations for the downstream AF prediction task. Our approach effectively mitigates data issues and enables downstream analysis, improving performance even with limited data.

In our experiments, we compare AF prediction performance using pre-training and transfer learning against traditional ML models applied directly to raw input data. Our proposed methods achieve a 5%-15% improvement in AUROC. These findings pave the way for more accurate and scalable clinical solutions - especially for complex, real-world datasets challenged by sample scarcity and high dimensionality.

## 2 Data Description

This study utilizes EHR data from the Emory Healthcare System, incorporating two datasets. The first is the ESUS dataset, a cohort of patients diagnosed with ESUS with labeled outcomes for AF, serving as the target dataset. The second is the AI-RESPECT dataset, a larger cohort of stroke patients labeled for post-stroke cognitive impairment (PSCI), used for pre-training and transfer learning.

The ESUS dataset consists of 510 patients diagnosed between January 1, 2015, and December 13, 2023, among whom 107 developed post-stroke AF as a first occurrence. To ensure data consistency, we apply several inclusion criteria. Inclusion criteria require patients to be at least 18 years old, have no documented stroke within a five-year retrospective window prior to 2015, and have no history of AF before the index stroke.

To enhance model performance, we leverage the AI-RESPECT dataset for pre-training before fine-tuning on the ESUS cohort. This dataset comprises 7,780 stroke patients diagnosed between January 1, 2012, and December 31, 2021, including 1,735 cases of PSCI. Inclusion criteria also require that patients have a stroke diagnosis but no prior history of cognitive impairment.

For the ESUS dataset, we extract 58 baseline features from the literature as candidate predictors, including (I) 6 demographic variables (II) 9 blood biomarkers, (III) 5 echocardiographic features (IV) 21 electrocardiographic (ECG) features, and (V) 17 comorbidities of stroke. To further enrich the dataset, we extract all past diagnoses of ESUS patients, yielding 1,529 diagnostic features, including 990 ICD-based features and 539 related to prescribed medications.

For the AI-RESPECT dataset, baseline features are unavailable, but it includes 2,595 diagnostic features, providing a broader representation of stroke-related medical histories. Across both datasets, 1,494 diagnostic features overlap, covering 97.71% of the ESUS dataset's diagnostic features.

# 3 Methods

## 3.1 Problem Formulation

Let $\mathcal{D}_{\text{ESUS}} = \{(x_i, y_i)\}_{i=1}^{n}$ represent our target dataset of $n = 510$ ESUS patients, where each patient $i$ is characterized by a feature vector $x_i \in \mathbb{R}^d$ and a binary label $y_i \in \{0, 1\}$. Since $n$ is relatively small, we leverage transfer learning to enhance feature representations. Specifically, $x_i$ is constructed using methods such as transfer learning, integrating information from two binary sub-vectors: (1) a baseline clinical feature vector $x_{i,b} \in \mathbb{R}^{d_b}$ and (2) a diagnostic feature vector $x_{i,di} \in \mathbb{R}^{d_{di}}$, where $d_b = 58$ and $d_{di} = 1529$. The objective of this binary classification task is to predict $y_i$, where $y_i = 1$ indicates the presence of AF, and $y_i = 0$ denotes its absence.

## 3.2 Overview

The overall methodology framework is shown in Figure 1. Given the sample scarcity ($n = 510$) and high dimensionality (58 baseline features and 1,529 diagnostic features) of our ESUS patient dataset, we propose novel methods leveraging transfer learning and pre-training techniques. To evaluate different learning strategies for AF prediction, we obtain patient feature representations using three approaches: (I) From-scratch, which directly uses the raw features; (II) Supervised, which pre-trains a model on a larger labeled dataset and then transfers the learned representations to our target dataset; and (III) Unsupervised, which pre-trains a model without using labels. Each representation is then fed into three AF prediction classifiers: Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB).

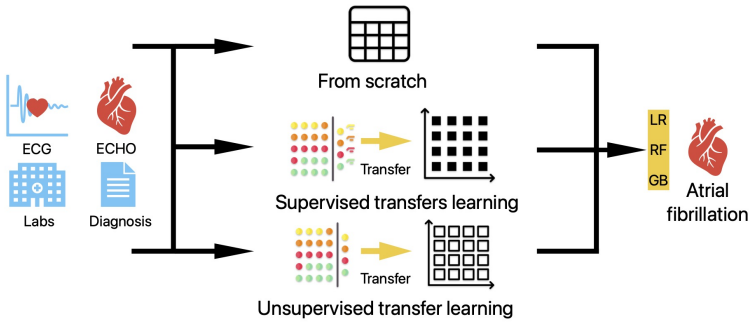


**Fig. 1.** Overview of our methodology.

### 3.3 From Scratch

In this approach, we directly concatenate the baseline feature vector $x_{i,b}$ and the diagnostic feature vector $x_{i,di}$ as feature vector input $x_i$ for patient $i$ to train machine learning models.

$$x_i = x_{i,b} \oplus x_{i,d} \tag{1}$$

This method does not rely on any pre-trained knowledge or external datasets, making it purely data-driven within the target ESUS cohort. While straightforward, this approach often struggles in high-dimensional, small-sample settings due to the risk of overfitting and the inability to leverage external information.

### 3.4 Supervised Method

Our proposed supervised method, leveraging pre-training and transfer learning, involves pre-training a model on a larger external dataset, AI-RESPECT, which contains labeled samples to capture generalizable patterns before fine-tuning on the ESUS cohort. Specifically, the AI-RESPECT dataset is defined as $\mathcal{D}_{\text{AI-RESPECT}} = \{(x_p, y_p)\}_{p=1}^{n_p}$, where $x_p$ consists of 2,595 diagnostic features, $n_p = 7,780$ represents the number of samples, and $y_p$ indicates whether a patient develops PSCI.

The overall methodology leverages the structural similarities between the AI-RESPECT and ESUS datasets—particularly the high overlap in diagnostic features—by pre-training a model $M$ on $\mathcal{D}_{\text{AI-RESPECT}}$. For pre-training, we adopt the state-of-the-art hypergraph transformers method inspired by [15], which formulates PSCI prediction as a hypergraph-based learning problem. A hypergraph $HG = (V, E)$ consists of vertices $V$ and hyperedges $E$, with each hyperedge connecting multiple vertices. In our framework, hyperedges represent patient visits and vertices represent diagnostic features. Diagnostic features $x_{i,di}$ are utilized as the initial node embeddings. Set transformers with a self-attention mechanism are employed to learn informative representations. At layer $l$, the message propagation is defined as

$$x_e^l = f_{V \to E}(x_v^{l-1}), \quad x_v^l = f_{E \to V}(x_e^l), \tag{2}$$

where $x_e^l$ represents the hyperedge embedding, $x_v^l$ represents the node embedding, $f_{V \to E}$ represents the self-attention mechanism within a hyperedge's all connected nodes, and $f_{E \to V}$ represents the self-attention mechanism within a node's all connected hyperedges. The hypergraph transformer model $M$ is optimized under the PSCI prediction task using cross-entropy loss on $\mathcal{D}_{\text{AI-RESPECT}}$ and is transferred to $\mathcal{D}_{\text{ESUS}}$ for generating pre-trained patient representation $x_{i,tr}$ via

$$x_{i,tr} = x_{e,i}^L = M(x_{i,di}), \tag{3}$$

where $M$ denotes the pre-trained model, $x_{i,di}$ denotes the diagnostic features, $x_{e,i}^L$ denotes the final hyperedge embedding corresponding to patient $i$ in the final layer $L$ after training. The final patient representation $x_i$ is constructed

by concatenating the pre-trained 32-D embedding $x_{i,tr}$ with the baseline clinical feature vector $x_{i,b}$:

$$x_i = x_{i,tr} \oplus x_{i,b}. \tag{4}$$

This combined representation $x_i$ is then used as input to machine learning models for the AF prediction task, thereby improving performance and generalizability in the small-sample ESUS cohort.

### 3.5 Unsupervised Method

Our proposed unsupervised method aims to learn patient representations from unlabeled data before fine-tuning them on the ESUS cohort. Specifically, we leverage the dataset $\mathcal{D}_{\text{AI-RESPECT}}$ for pre-training an unsupervised model and also transfer it to the target dataset $\mathcal{D}_{\text{ESUS}}$. Same as the supervised method, we also construct hypergraph $HG = (V, E)$ consisting of vertices $V$ and hyperedges $E$, with hyperedges representing patient visits and vertices representing diagnostic features. Then two key components are introduced:

**Hypergraph View Augmentation (genSim)** [14] generates two augmented hypergraph views, $HG(A)$ and $HG(B)$, by applying node masking and hyperedge selection. For node masking, nodes with high duplication are more likely to be masked, where duplication is measured as $D(v_i) = \frac{\sum_{e \in E\{v_i\}} |e|}{\left|\bigcup_{e \in E\{v_i\}} e\right|}$, with $|e|$ is the number of nodes in hyperedge $e$. The masking probability is $p_{v_i} = \min\left\{\frac{w_{n_{\max}} - w_{v_i}}{w_{n_{\max}} - w_{\text{navg}}} \cdot p_{\text{node}}, p_\tau\right\}$, where $w_{v_i} = \log o(v_i)$ quantifies duplication, and $p_{\text{node}}, p_\tau$ control the masking probability. Hyperedge selection retains, removes, or modifies hyperedges via Gumbel-Softmax sampling [9]. The genSim objective is defined as

$$L_{\text{genSim}} = L_{\text{hyper}} + L_{\text{sim}}. \tag{5}$$

Here,

$$L_{\text{sim}} = MSE(sample_1, sample_2), \tag{6}$$

where $sample_1$ and $sample_2$ being the node features after augmentation.

$$L_{\text{hyper}} = \frac{1}{2}\big(\mathcal{L}(Z_1^T, Z_2^T) + \mathcal{L}(Z_2^T, Z_1^T)\big), \tag{7}$$

where $Z_1$ and $Z_2$ represent the hyperedge embeddings computed from the augmented graphs. This ensures structural consistency while allowing controlled variation between views.

**Hypergraph Triplet Contrastive Learning (Trip)** [10] further enhances hypergraph semantics by applying node feature masking and membership masking to create $G1$ and $G2$, where $\mathbf{h}$ denotes node embeddings, and $\boldsymbol{y}$ represents hyperedge embeddings. Three contrastive objectives are then employed:

*(a) Node-level contrast:* Anchors $\mathbf{h}_{\mathcal{G}_{1,i}}$ with positives $\mathbf{h}_{\mathcal{G}_{2,i}}$ and negatives $\mathbf{h}_{\mathcal{G}_{2,z}}, z \neq i$, optimizing

$$L_n = \frac{1}{2|V|} \sum_{i=1}^{|V|} \big[\ell(\mathbf{h}_{\mathcal{G}_{1,i}}, \mathbf{h}_{\mathcal{M}_{2,i}}) + \ell(\mathbf{h}_{\mathcal{G}_{2,i}}, \mathbf{h}_{\mathcal{G}_{1,i}})\big] \tag{8}$$

*(b) Hyperedge-level contrast:* Anchors $\boldsymbol{y}_{\mathcal{G}_{1,j}}$ with positives $\boldsymbol{y}_{\mathcal{G}_{2,j}}$ and negatives $\boldsymbol{y}_{\mathcal{G}_{2,k}}, k \neq j$, optimizing

$$L_e = \frac{1}{2|E|} \sum_{j=1}^{|E|} \left[ \ell(\boldsymbol{y}_{\mathcal{G}_{1,j}}, \boldsymbol{y}_{\mathcal{G}_{2,j}}) + \ell(\boldsymbol{y}_{\mathcal{G}_{2,j}}, \boldsymbol{y}_{\mathcal{G}_{1,j}}) \right] \tag{9}$$

*(c) Membership-level contrast:* Distinguishes real $v_i \in e_j$ from fake memberships by optimizing

$$L_m = \frac{1}{2K} \sum\nolimits_{i=1}^{|V|} \sum\nolimits_{j=1}^{|E|} \mathbf{1}_{[h_{ij}=1]} \left[ \ell(\mathbf{h}_{\mathcal{M}_{1,i}}, \boldsymbol{y}_{\mathcal{M}_{2,j}}) + \ell(\mathbf{h}_{\mathcal{M}_{2,i}}, \boldsymbol{y}_{\mathcal{M}_{1,j}}) \right] \tag{10}$$

To optimize the pre-training model $M$, we employ *Equal Weights*, leading to a weighted loss

$$L_{\text{total}} = L_{\text{genSim}} + L_n + L_e + L_m. \tag{11}$$

The resulting 32-D unsupervised embeddings for patients $x_{i,tr}$ can be obtained after transferring the pre-trained model $M$ to $\mathcal{D}_{\text{ESUS}}$, the same as the Eq.(3) and (4) for the supervised method.

## 4 Results

### 4.1 Implementation

In the preprocessing stage, missing numerical data are imputed using a SimpleImputer with median, and numerical variables are normalized with MinMaxScaler. We compare three methods for generating ESUS patient representations: From-scratch, Supervised and Unsupervised, and evaluate the performance of the resulting features by training AF prediction models including logistic regression (LR), random forest (RF), and gradient boosting (GB). The dataset is split into 80% for training-validation and 20% for testing. The training configuration sets the maximum iterations to 100 and fixes the random state at 100 for reproducibility. Finally, a five-fold nested cross-validation strategy is employed for robust evaluation, with performance measured using **AUROC** (the primary metric), accuracy, precision, recall, and F1-score.

### 4.2 AF Prediction Performance

The performance of the compared methods is summarized in Table 1. Pre-training-based methods significantly improve AF prediction performance compared to learning from scratch, particularly in AUROC, our primary metric. The worse performance of directly using the raw input features is likely due to overfitting caused by the high dimensionality and small sample size. For example, for "From-scratch + RF" method, the model predicts all test samples as negative class. In contrast, our proposed methods leveraging transfer learning and pre-training effectively transfer the related knowledge to our target dataset, and also reduce the dimensionality, enhancing both model performance and training efficiency.

**Table 1.** AF Prediction Performance Comparison

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

### 4.3 Ablation Studies

To investigate how dataset size impacts the performance of our proposed methods, we train models in varying proportions of the training data (20%, 40%, 60%, and 80%) while holding the test data fixed as the 20% of the whole dataset and recording their AUROC. The results are summarized in Figure 2, indicating that our methods maintain stable performance even when trained with limited data, while models trained from scratch using raw features show a notable decline in performance as the training data size decreases.

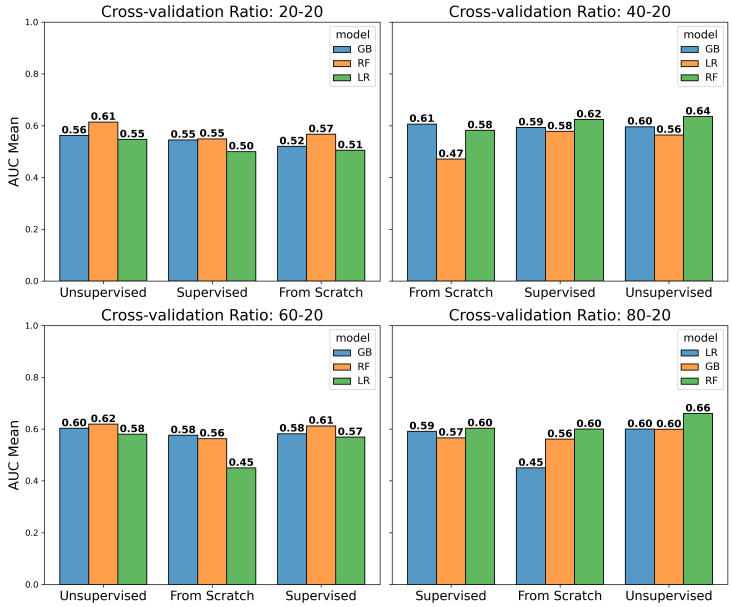


**Fig. 2.** Impact of training dataset size on model performance.

Furthermore, to evaluate the robustness of our pre-training methods, we conduct ablation studies by varying the embedding dimensionality. As shown in Figure 3, models using 32-dimensional embeddings outperform those with 64-dimensional embeddings in terms of AUCROC. This finding validates our choice of generating 32-dimensional pre-trained embeddings, likely due to the benefits of dimensionality reduction in capturing meaningful features.

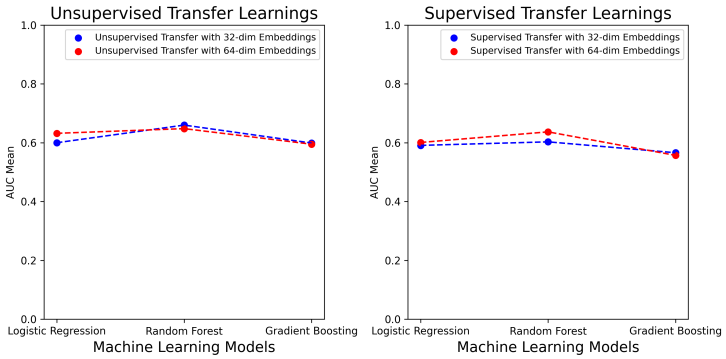


**Fig. 3.** Impact of embedding dimensionality on model performance.

## 5 Conclusion

In this paper, we propose novel methods leveraging pre-training and transfer learning for AF prediction in our ESUS dataset, effectively addressing challenges such as data scarcity and high dimensionality. We employ both supervised and unsupervised approaches, with the unsupervised pre-training yielding the best performance—a significant advantage in healthcare settings where data and labels are limited. Additionally, we incorporate each patient's complete, unbiased diagnostic coded medical history, which is crucial for understanding the cryptogenic nature of ESUS and identifying potential risk factors for stroke onset and subsequent AF development.

Despite these promising findings, our study has limitations. The small dataset increases the risk of overfitting, particularly with complex models, and underscores the need for validation on larger, multi-institutional datasets to ensure robustness and generalizability. Moreover, while embeddings enhance predictive performance, their clinical interpretability remains limited, making it challenging to discern which aspects of patient history most contribute to AF risk. Future research could focus on improving the explainability of these representations and expanding the dataset to include a more diverse patient population, thereby refining embedding techniques and further enhancing prediction performance.

## References


1. Andayeshgar, B., Abdali-Mohammadi, F., Sepahvand, M., Daneshkhah, A., Almasi, A., Salari, N.: Developing graph convolutional networks and mutual information for arrhythmic diagnosis based on multichannel ecg signals. International Journal of Environmental Research and Public Health **19**(17), 10707 (2022)
2. Bhat, A., Mahajan, V., Chen, H.H., Gan, G.C., Pontes-Neto, O.M., Tan, T.C.: Embolic stroke of undetermined source: approaches in risk stratification for cardioembolism. Stroke **52**(12), e820–e836 (2021)
3. Bhatla, A., Borovskiy, Y., Katz, R., Hyman, M.C., Patel, P.J., Arkles, J., Callans, D.J., Chokshi, N., Dixit, S., Epstein, A.E., et al.: Stroke, timing of atrial fibrillation diagnosis, and risk of death. Neurology **96**(12), e1655–e1662 (2021)
4. Chen, L.Y., Norby, F.L., Chamberlain, A.M., MacLehose, R.F., Bengtson, L.G., Lutsey, P.L., Alonso, A.: Cha2ds2-vasc score and stroke prediction in atrial fibrillation in whites, blacks, and hispanics. Stroke **50**(1), 28–33 (2019)
5. Dhingra, L.S., Shen, M., Mangla, A., Khera, R.: Cardiovascular care innovation through data-driven discoveries in the electronic health record. The American Journal of Cardiology **203**, 136–148 (2023)
6. Elsheikh, S., Hill, A., Irving, G., Lip, G.Y., Abdul-Rahim, A.H.: Atrial fibrillation and stroke: State-of-the-art and future directions. Current Problems in Cardiology **49**, 102181 (2024)
7. Essa, H., Hill, A.M., Lip, G.Y.: Atrial fibrillation and stroke. Cardiac electrophysiology clinics **13**(1), 243–255 (2021)
8. Hart, R.G., Catanese, L., Perera, K.S., Ntaios, G., Connolly, S.J.: Embolic stroke of undetermined source: a systematic review and clinical update. Stroke **48**(4), 867–872 (2017)
9. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
10. Lee, D., Shin, K.: I'm me, we're us, and i'm us: Tri-directional contrastive learning on hypergraphs. arXiv preprint arXiv:2206.04739 (2022)
11. Li, Y.G., Pastori, D., Farcomeni, A., Yang, P.S., Jang, E., Joung, B., Wang, Y.T., Guo, Y.T., Lip, G.Y.: A simple clinical risk score (c2hest) for predicting incident atrial fibrillation in asian subjects: derivation in 471,446 chinese subjects, with internal validation and external application in 451,199 korean subjects. Chest **155**(3), 510–518 (2019)
12. Ming, C., Lee, G.J., Teo, Y.H., Teo, Y.N., Toh, E.M., Li, T.Y., Guo, C.Y., Ding, J., Zhou, X., Teoh, H.L., et al.: Machine learning modeling to predict atrial fibrillation detection in embolic stroke of undetermined source patients. Journal of Personalized Medicine **14**(5), 534 (2024)
13. Sarwar, T., Seifollahi, S., Chan, J., Zhang, X., Aksakalli, V., Hudson, I., Verspoor, K., Cavedon, L.: The secondary use of electronic health records for data mining: Data characteristics and challenges. ACM Computing Surveys (CSUR) **55**(2), 1–40 (2022)
14. Song, Y., Gu, Y., Li, T., Qi, J., Liu, Z., Jensen, C.S., Yu, G.: Chgnn: A semi-supervised contrastive hypergraph learning network (2023)
15. Xu, R., Ali, M.K., Ho, J.C., Yang, C.: Hypergraph transformers for ehr-based clinical predictions. AMIA Summits on Translational Science Proceedings **2023**, 582 (2023)