Multi-modal Predictive Model for MACE Risk Estimation in Patients with Migraine

***Abstract*—Clinical research has indicated a link between migraine related characteristics and risk of adverse cardiovascular and cerebrovascular events. We propose a fusion predictive model with graph convolution as backbone, for estimating risk of such events among migraine patients using their neurology consultation notes and history of prescription medication. We propensity matched our dataset on all known cardiovascular disease risk factors so that our model can learn relationships between migraine related features and cardiovascular disease. Our fusion based risk prediction model outperformed all comparative models in terms of prediction accuracy. Post-hoc analysis shed light onto migraine characteristics related to vascular disease risk.**

***Index Terms*—Graph neural networks, Multi-modal modeling, MACE risk estimation**

I. Introduction

Migraine is a common neurological disorder characterized by episodic severe headache which are accompanied by autonomic nervous system dysfunction [1], [2]. In the United States, the 1-year prevalence of migraine is approximately 18% in women and 6% in men. Approximately 28 million Americans report severe and disabling migraine [3]. Several clinical research studies have been performed to study the link between specific migraine characteristics and medication and major adverse cardiovascular events (MACE) [4]–[8]. Some efforts have been made to include migraine while predicting risk of cardiovascular disease and stroke [6], [8]. We present a machine learning based approach to estimate risk of MACE among migraine patients using migraine-related features.

We propose a predictive model that can process free-text clinical notes and tabular information about patient's medication history to predict risk of future MACE outcome. Clinical notes included all information recorded during a migraine patient's visit to neurology department. We argue that clinical notes where clinicians record their impressions of important migraine characteristics of any patient, e.g., severity, frequency or presence of aura and related symptoms like photophobia and phonophobia comprehensively capture migraine-related clinical features. Otherwise, migraine related features may need to be obtained from questionnaires provided to migraine patients which may require additional time and effort from clinical experts. Medications are routinely recorded in electronic health records (EHR) in tabular form. Technical challenge lied in fusion of these two heterogeneous modalities, i.e., free-text and tabular data.

We are leveraging graph convolutional neural networks (GNNs) as the backbone for MACE risk prediction model to model the topological structure between patients. In recent years, graph convolutional modeling has proven to be very effective for a number of machine learning problems where individual samples form an informative 'neighborhood' [9]–[14]. Graph-based modeling allows the model to perform two-fold learning, i.e., the model not only learns from features of the an instance (a node in a graph) but also from features of 'similar' instances. Definition of similarity is left upto the model designers. In contrast, traditional convolutional neural network can process data elements only in reference to their spatial neighborhood. GNNs have better capabilities of handling missing data and display have higher data modality agnosticism [11], [13]. GNNs have been widely applied for predictive modeling for autism spectrum disorder (ASD) and Alzheimer with patients as nodes and 'neighborhood' defined by demographically similar patients [9]. In addition, GNNs have been used to predict clinical events like discharge from hospital, mortality, readmission to the hospital, etc., with clinical similarity used to define 'neighborhoods' of patients [10]–[13]. In this work, we defined graph structure with migraine patients as nodes and neighborhoods defined by similarity among patients based on their history of prescribed medications. Clinical notes for each patient were used to design node feature vectors. MACE risk was estimated for each node/patient as target label while allowing the model to learn from the notes of the corresponding patient as well as notes of patients with similar medication prescription history. Thus, MACE risk was estimated by leveraging two-fold learning properties of GNN for fusion to two heterogeneous clinical data elements, i.e., clinical notes and medications' record.

II. Methodology

A. Cohort Selection

We identified 1107 patients being treated at Institute X for migraine with MACE (cardiovascular events of stroke, myocardial infarction and hemorrhage) outcomes recorded after their diagnosis of migraine. These patients formed our case set which was propensity matched with an equal number of control samples (patients with no recorded MACE after migraine diagnosis) based on the following factors; i) age, ii) gender, iii) diabetes, iv) hypertension, v) hyperlipidemia, vi) atrial fibrillation. Propensity matching factors were curated from recorded ICD-9/10 codes in EHR after verification of selected codes with clinicians. Figure 1 shows details of cohort selection. 20% patients were randomly sampled from this cohort and held-out from training process to serve as test set for the risk prediction models.

American Heart Association

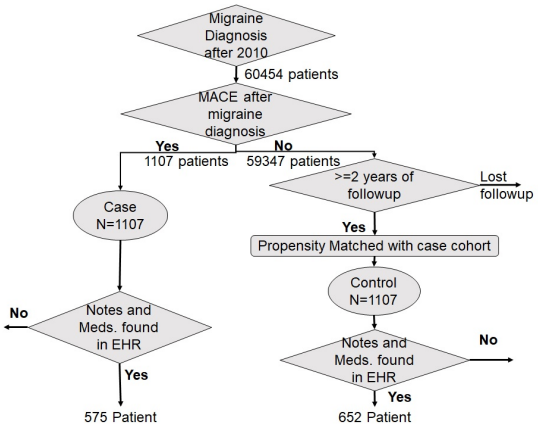


Fig. 1: Cohort selection

B. Clinical Data Elements

1) Clinical Notes: Since we focused on identifying potential risk factors for MACE among migraine patients, we decided to incorporate clinical notes written by neurologists during their consultation with migraine patients. We argue that these notes allow our model to incorporate experts' clinical observations individual to each patient, hence, more comprehensive clinical insight compared to relying on a preselected set of clinical features. Neurologist may note distinctive clinical characteristics of each patient, in addition to recording standard clinical features like frequency and severity of headaches.

We collated all neurology notes of a patient and used a regular expression based framework to identify sentences related to migraine and migraine related symptoms. This step curtailed the length of free-text information for each patient, making it suitable for processing through large language models (LLM). We used BioGPT [15] to generate embeddings of concatenated clinical text for each patient. BioGPT was chosen as it was designed for the domains of medicine and biology compared to general purpose LLMs like GPT.

2) Tabular Electronic Health Records: In addition to clinical notes, we had access EHR of migraine patients including their demographic features, diagnosis, procedures, and medications. Demographic features like age and diagnosis and procedures related to cardiovascular disease represent obvious risk factors for future MACE. We aimed at identifying migraine-related risk factors specifically. Therefore, we curated a propensity matched dataset where case and control samples were matched on obvious MACE risk factors. Since the use of some migraine medications, like triptans [16], could theoretically increase the risk of MACE amongst individuals with cardiovascular disease, we decided to incorporate medications in our predictive modeling. Therefore, we decided to incorporate medications record in our predictive modeling.

While there may be thousands of different medications, they are categorized into 48 therapeutic classes based on their function. For each patient, we generated 48-dimensional feature vector where each element represented count of medication prescribed to the patient under the corresponding therapeutic class.

Any patients in the propensity matched set which had either of the two chosen clinical data elements were dropped.

C. Graph Neural Network

GNN architecture can process an input graph $G(N, E)$ where N denotes the set of nodes and E denotes the set of edges. In our case, i^{th} patient was represented by i^{th} node, and has two feature vectors; i) node feature vector (n_i) formed by BioGPT-based embedding generated from parsed clinical notes of i^{th} patient, and ii) edge feature vector (v_i) formed as $48-$dimensional medication therapeutic class records. An edge between patient i and j was decided based on similarity function $Sim(e_i, e_j)$ based on their edge feature vectors (medication records). GNN model learned to generate embeddings for i^{th} patients by manipulating nodes features of this node (n_i), and 'messages' received from nodes in its edge-connected neighbor $(\eta(i))$.

In essence, 'messages' passed to i^{th} nodes are features of the nodes $n_j, \forall n_j \in \eta(i)$. GNN model, through its training process, learns the function parameters to manipulate features of the i^{th} node (n_i) as well as 'messages' being received from its neighborhood. Hence, GNN is capable of two-fold learning. The model learns from the features of i^{th} node directly, and implicitly learns from information used for edge formation (edge features e_i), through incorporation of 'messages' from edge-connected nodes $(n_j, \forall n_j \in \eta(i))$.

At $k+1^{th}$ graph convolutional layer, the following describes the process of generating embedding of i^{th} node (n_i^{k+1})

$$n_{N_i}^{k+1} = aggregate(n_j^k, \forall v_j \in \eta(i)) \tag{1}$$

$$n_i^{k+1} = \sigma(W_{k+1} \times \phi(n_i^k, n_{N(i)}^{k+1})) \tag{2}$$

Our case represented a supervised learning scenario where class label regarding future MACE (True or False) were present during training for each node. In this case, node/patient embedding generated by graph convolutional layers was used to predict target label $\hat{y}_i$ for i^{th} node/patient as

$$\hat{y}_i = \sigma(W_{fc} \times n_i^{k+1}) \tag{3}$$

Through back-propagation of binary cross entropy loss based on groundtruth labels (y) and predicted labels $(\hat{y})$, weight matrices of K graph convolutional layers $W_k \forall k \in K$ and classification layer W_{fc} were optimized.

Our modeling scheme was inspired by SAGE (SAmple and aggreGatE) graph convolution network (GraphSAGE) [17]. A sample aggregate function was optimized to collect and manipulate 'messages' from the neighborhood of i^{th} nodes and its own node features n_i to generate embedding of i^{th} node/patient. At the time of inference, this optimized aggregate function was used to collect and process messages from neighboring nodes in potentially un-seen graph structures. Thus, our model can inductively reason to assign labels to unseen nodes in unseen graph structures.

Model	Sensitivity	Specificity
Medication-only Model	0.75	0.39
Clinical note-only Model	0.64	0.56
Late fusion	0.58	0.74
GNN-based fusion model	0.64	0.67

TABLE I: Performance for prediction of future MACE by single-modality and fusion models

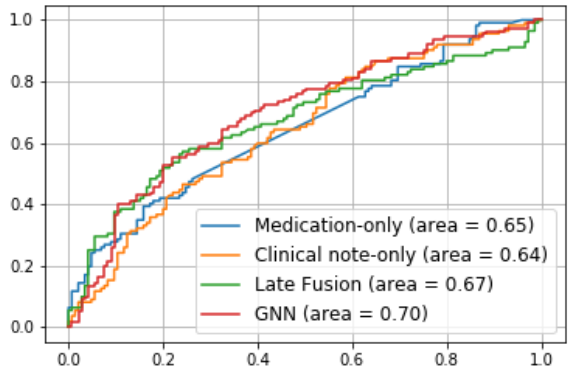


Fig. 2: Receiver operating characteristics curves

III. Results

A. Comparative Models

As we propose a GNN based multi-modal fusion model for MACE risk estimation, intuitive baseline or comparative models include single modality models using the same clinical data elements and traditional fusion model. The list of comparative models is as follows.

- Medication-only: random forest classifier to predict MACE risk using only medications for each patient
- Clinical note-only: classifier on top of BioGPT backbone; BioGPT backbone was frozen before training the classification layers.
- Late Fusion: logistic regression based meta-learner for fusing MACE risk estimates from medication-only and clinical-not only model

Table I shows comparative performance of these models and the proposed GNN based model in terms of sensitivity and specificity. While either sensitivity or specificity can be increased at the cost of the other, area under the ROC incorporate the trade-off between the two. ROC plots with AUC values are shown in Figure 2. Figure 3 shows confusion matrix for the labels predicted by our GNN based model.

IV. Discussion

Quantitative analysis (Table I) indicated that the medications-only model was highly sensitive (identified large fraction of patient with high risk for MACE) but suffered from low specificity (generated many false positive labels, i.e., over-diagnosed patients with high risk for MACE). The late fusion model showed the opposite trend

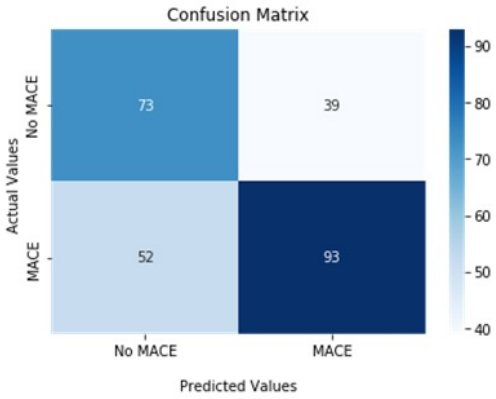


Fig. 3: Confusion matrix for the GNN-based model

(high specificity, low sensitivity). GNN model displayed balanced performance in terms of sensitivity and specificity. It outperformed the notes-only model in terms of specificity while achieving equivalent sensitivity. ROC curves (Figure 2) established the superiority of GNN based model as it achieved the best AUROC. All well-known risk factors for MACE were balanced in our propensity matched cohort with the aim to allow the model to learn correlations between migraine related clinical features and MACE. This may explain moderate performance of even the best performing model indicating only moderately strong correlation of migraine related features with MACE.

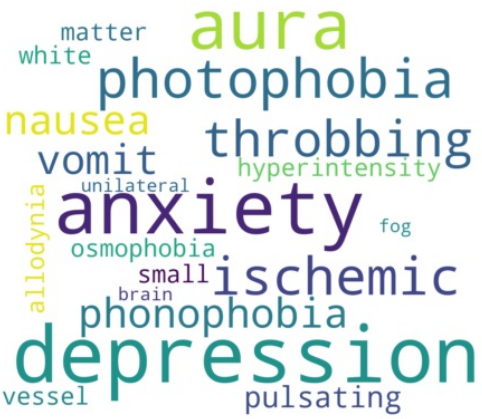
A. Post-hoc Investigation of Risk Factors

Our cohort selection through propensity matching was focused on making sure that the dataset was balanced in terms of known risk factors for MACE including demographic factors like age and commodities like hypertension and atrial fibrillation. We wanted our model to focus on clinical characteristics that were of interest for migraine experts like neurologists. These characteristics were provided as input to the model through clinical notes' text.

As a post-hoc investigation of migraine-related clinical features and their association with MACE risk, we parsed clinical text of true positive (patients with future MACE correctly identified by the model as high-risk for MACE) and true negative (patient with no recorded future MACE correctly identified the model as low-risk for MACE) samples for identification of positive mention of expert-identified migraine-related features. Regular expression-based parsing was applied to filter our text with negated mention of these clinical features. The list of clinical features is as follows; *'photophobia', 'phonophobia', 'osmophobia', 'nausea', 'vomit', 'anxiety', 'aura', 'depression', 'white matter hyperintensity', 'patent foramen', 'small vessel', 'ischemic', 'allodynia', 'pulsating', 'throbbing', 'bilateral headache', 'unilateral headache', 'bilateral migraine', 'unilateral migraine', 'brain fog'*.

Figure 4 shows word clouds generated from clinical text of true positive and true negative samples after parsing the text to preserve only non-negated mention of expert-defined clinical

(a) High-risk for MACE (true positive samples)



(b) Low-risk for MACE (true negative samples)

Fig. 4: Word clouds generated by clinical text of low- and high-risk patient for MACE

features. It appears that most of the migraine-characteristics like 'unilateral' vs. 'bilateral' headache', 'pulsating' or 'throbbing' headache, and migraine-related symptoms like 'photophobia', 'phonophobia' and 'osmophobia' have almost similar representation in both low- and high-risk patients sets. Situation is similar with associated disorders like 'anxiety' and 'depression'. These mental health disorders also seem equally prevalent in both high and low risk groups. Interestingly, migraine with aura which is generally considered a risk factor for stroke [4] did not hold high association with MACE in our cohort. However, 'white matter hyperintensity' and 'small vessel ischemic' disease held higher prevalence in the high risk group as compared to low risk group. This post hoc investigation was aimed at directing future clinical investigation beyond known cardiac disease risk factors, especially for migraine patients.

V. Conclusion

We designed and evaluated a MACE risk estimation model for migraine patients using unique multi-modal fusion modeling architecture based on graph convolution. Our fusion modeling scheme allowed us to incorporate important observations recorded by the neurologists' in clinical notes and medication history of migraine patients for risk estimation. Our model outperformed all comparative baseline models including single modality models and traditional fusion models. Post-hoc analysis provided insights into expert-defined clinical characteristics set and risk of MACE.

References

[1] P. J. Goadsby, R. B. Lipton, and M. D. Ferrari, "Migraine—current understanding and treatment," *New England journal of medicine*, vol. 346, no. 4, pp. 257–270, 2002.

[2] S. D. Silberstein, "Recent developments in migraine," *The Lancet*, vol. 372, no. 9647, pp. 1369–1371, 2008.

[3] R. B. Lipton, W. F. Stewart, S. Diamond, M. L. Diamond, and M. Reed, "Prevalence and burden of migraine in the united states: data from the american migraine study ii," *Headache: The Journal of Head and Face Pain*, vol. 41, no. 7, pp. 646–657, 2001.

[4] T. Kurth, P. M. Rist, P. M. Ridker, G. Kotler, V. Bubes, and J. E. Buring, "Association of migraine with aura and other risk factors with incident cardiovascular disease in women," *JAMA*, vol. 323, no. 22, pp. 2281–2289, 2020.

[5] T. Kurth, J. M. Gaziano, N. R. Cook, G. Logroscino, H.-C. Diener, and J. E. Buring, "Migraine and risk of cardiovascular disease in women," *Jama*, vol. 296, no. 3, pp. 283–291, 2006.

[6] L. R. Øie, T. Kurth, S. Gulati, and D. W. Dodick, "Migraine and risk of stroke," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 91, no. 6, pp. 593–604, 2020.

[7] E. C. McKinley, C. L. Lay, R. S. Rosenson, L. Chen, V. Chia, L. D. Colantonio, P. Muntner, R. Urman, and M. E. Farkouh, "Risk for ischemic stroke and coronary heart disease associated with migraine and migraine medication among older adults," *The journal of headache and pain*, vol. 22, pp. 1–13, 2021.

[8] P. M. Rist, J. E. Buring, N. R. Cook, and T. Kurth, "Contribution of migraine to cardiovascular disease risk prediction," *Journal of the American College of Cardiology*, vol. 81, no. 23, pp. 2246–2254, 2023.

[9] M. Cao, M. Yang, C. Qin, X. Zhu, Y. Chen, J. Wang, and T. Liu, "Using deepgcn to identify the autism spectrum disorder from multi-site resting-state data," *Biomedical Signal Processing and Control*, vol. 70, p. 103015, 2021.

[10] A. Tariq, L. Lancaster, P. Elugunti, E. Siebeneck, K. Noe, B. Borah, J. Moriarty, I. Banerjee, and B. N. Patel, "Graph convolutional network-based fusion model to predict risk of hospital acquired infections," *Journal of the American Medical Informatics Association*, vol. 30, no. 6, pp. 1056–1067, 2023.

[11] S. Tang, A. Tariq, J. A. Dunnmon, U. Sharma, P. Elugunti, D. L. Rubin, B. N. Patel, and I. Banerjee, "Predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 2071–2082, 2023.

[12] A. Tariq, L. Su, B. Patel, and I. Banerjee, "Prediction of transfusion among in-patient population using temporal pattern based clinical similarity graphs," in *AMIA Annual Symposium Proceedings*, vol. 2023. American Medical Informatics Association, 2023, p. 679.

[13] A. Tariq, S. Tang, H. Sakhi, L. A. Celi, J. M. Newsome, D. L. Rubin, H. Trivedi, J. W. Gichoya, and I. Banerjee, "Fusion of imaging and non-imaging data for disease trajectory prediction for coronavirus disease 2019 patients," *Journal of Medical Imaging*, vol. 10, no. 3, pp. 034 004–034 004, 2023.

[14] B. Pfeifer, A. Saranti, and A. Holzinger, "Gnn-subnet: disease subnetwork detection with explainable graph neural networks," *Bioinformatics*, vol. 38, no. Supplement_2, pp. ii120–ii126, 2022.

[15] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, "Biogpt: generative pre-trained transformer for biomedical text generation and mining," *Briefings in bioinformatics*, vol. 23, no. 6, p. bbac409, 2022.

[16] V. Albieri, T. S. Olsen, and K. K. Andersen, "Risk of stroke in migraineurs using triptans. associations with age, sex, stroke severity and subtype," *EBioMedicine*, vol. 6, pp. 199–205, 2016.

[17] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.