CHẨN ĐOÁN BỆNH ĐỘNG MẠCH VÀNH DỰA TRÊN ẢNH CỰC SPECT SỬ DỤNG MẠNG NƠ-RON ĐỒ THỊ

Bùi Ngọc Tùng [1*], Nguyễn Chí Thành[2], Nguyễn Thành Trung[3], Lại Phú Minh[3], Nguyễn Hoàng Việt[4], Phùng Vũ Hạnh My[5]

Tóm tắt: *Bệnh động mạch vành đang là một trong những nguyên nhân hàng đầu gây tử vong trên thế giới. Hiện nay, chụp xạ hình tưới máu cơ tim (MPI) bằng máy chụp cắt lớp phát xạ đơn photon (SPECT) là phương pháp tiên tiến đang được sử dụng để phát hiện và chẩn đoán các tổn thương cơ tim. Việc xây dựng các mô hình học máy phục vụ hỗ trợ chẩn đoán bệnh là cần thiết nhằm tiết kiệm thời gian và gia tăng độ chính xác trong chẩn đoán. Trong bài báo này, nhóm tác giả đề xuất mạng nơ-ron đồ thị giúp hỗ trợ chẩn đoán các tổn thương cơ tim. Dữ liệu ảnh được biểu diễn thành đồ thị sử dụng thuật toán phân cụm pixel. Đồ thị được đưa vào mạng để huấn luyện, trích xuất đặc trưng và đưa ra kết quả chẩn đoán. Kết quả, các mô hình thử nghiệm lần lượt cho độ chính xác như sau: GCN: 79.34%; GatedGCN 78.45%; GAT: 76.42%; MoNet: 80.1%*

Từ khóa: Chụp cắt lớp phát xạ đơn photon (SPECT); Bệnh động mạch vành, xạ hình tưới máu cơ tim (MPI); Mạng nơ-ron đồ thị (GNN).

1. ĐẶT VẤN ĐỀ

Theo số liệu của WHO năm 2015, bệnh động mạch vành gây ra cái chết của khoảng 8,8 triệu người, chiếm 15,5% tổng số ca tử vong trên toàn cầu. Tại Việt Nam, các bệnh lý về tim mạch, chủ yếu là bệnh động mạch vành, đã cướp đi sinh mạng của 200.000 người mỗi năm, chiếm ¼ tổng số ca tử vong. Trong khi đó, việc chẩn đoán bệnh của các bác sĩ đòi hỏi nhiều thời gian để phân tích, hội chẩn và phụ thuộc vào những yếu tố như chất lượng hình ảnh, kinh nghiệm.

Các mạng nơ-ron học sâu đang nổi lên là một phương pháp hiệu quả để phát hiện và chẩn đoán bệnh dựa trên hình ảnh [1-3]. Trong chẩn đoán bệnh động mạch vành, học sâu cũng mang lại những kết quả ấn tượng [1, 4]. Mạng nơ-ron tích chập trong nghiên cứu [4] có độ chính xác chẩn đoán là 82,3% trên bộ dữ liệu bao gồm 1638 ảnh cực.

Trong bài báo này, nhóm tác giả đề xuất một phương pháp sử dụng mạng nơ-ron đồ thị nhằm phân loại ảnh cực SPECT. Thay vì sử dụng trực tiếp các ảnh cực để huấn luyện cho mô hình, ảnh cực sẽ được biến đổi thành đồ thị và những đồ thị này chính là đầu vào giúp mô hình học và cập nhật tham số.

2. PHÂN LOẠI ẢNH CỰC SPECT BẰNG MẠNG NƠ-RON ĐỒ THỊ

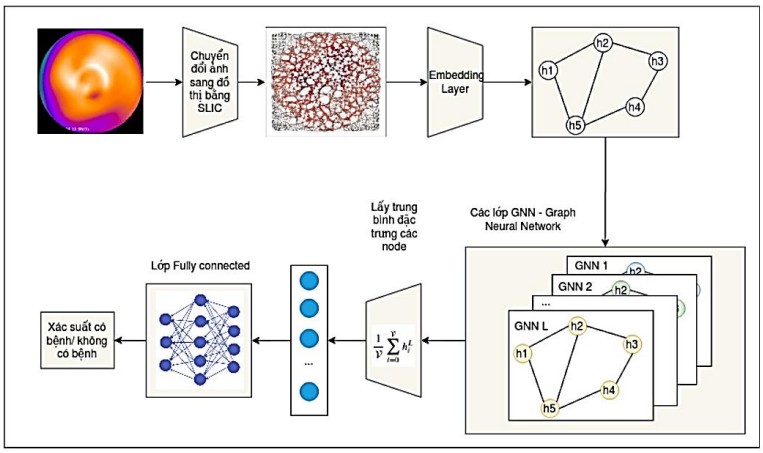
2.1. Mạng nơ-ron đồ thị (Graph Neural Network)

2.1.1. Giới thiệu

Các mạng nơ-ron đồ thị (Graph Neural Network) ra đời giúp giải quyết những bài toán mà dữ liệu đầu vào của nó được biểu diễn dưới dạng các đồ thị và đang thu hút được nhiều sự chú ý bởi tiềm năng của nó đã được chứng minh khi giải quyết những bài toán thuộc nhiều lĩnh vực hóa học, vật lý, khoa học xã hội,... [5, 6].

Dữ liệu dạng ảnh cũng có thể được biểu diễn thành dạng các đồ thị, đơn giản nhất là coi mỗi pixel của ảnh là một nút của đồ thị và các pixel gần nhau được nối với nhau coi như các cạnh. Qua đó có thể ứng dụng các mạng nơ-ron đồ thị cho bài toán phân loại ảnh. Các loại mạng nơ-ron đồ thị được lựa chọn cho bài toán phân loại ảnh cực bao gồm: GCN [7], MoNet [8], GAT [9] và GatedGCN [10] đây là các mạng nơ-ron đồ thị mà đặc trưng ở một nút sẽ được cập nhật dựa trên các nút hàng xóm của nút đó.

Mô hình áp dụng mạng nơ-ron đồ thị cho bài toán phân loại ảnh cực được cho ở hình 1.



Hình 1. *Mô hình tổng quan áp dụng mạng nơ-ron đồ thị cho phân loại ảnh cực.*

2.1.2. Kiến trúc mạng nơ-ron đồ thị cho phân loại ảnh cực

Dữ liệu đầu vào của mạng là đồ thị G với số nút V và tập các cạnh E mà mỗi nút i của nó sẽ được biểu diễn bằng một vec-tơ đặc trưng $\alpha_i \in \mathbb{R}^{a \times 1}$ và mỗi cạnh từ nút i đến nút j được biểu diễn bởi một vec-tơ $\beta_{ij} \in \mathbb{R}^{b \times 1}$. Các đặc trưng α_i và β_{ij} sẽ được ánh xạ sang các đặc trưng ẩn $h_i^{l=0}$ và $e_{ij}^{l=0}$ thông qua một tầng *embedding*. Cụ thể là:

$$h_i^0 = U^0 \alpha_i + u^0 \quad (1)$$

$$e_{ij}^0 = V^0 \beta_{ij} + v^0 \quad (2)$$

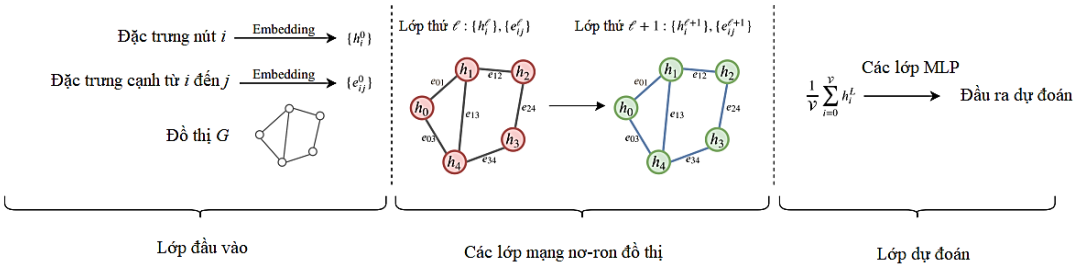
Trong đó, $U^0 \in \mathbb{R}^{d \times a}$, $V^0 \in \mathbb{R}^{d \times b}$ và $u^0, v^0 \in \mathbb{R}^d$ lần lượt các tham số của tầng *embedding*. Các tham số này sẽ được cập nhật dần thông qua quá trình huấn luyện của mạng. Giá trị đầu ra của tầng *embedding* này sẽ được đưa vào các lớp của mạng nơ-ron đồ thị.

Gọi h_i^l là vec-tơ đặc trưng ở tầng nơ-ron đồ thị l của nút thứ i. Đối với các mạng nơ-ron đồ thị được sử dụng trong nghiên cứu này thì đặc trưng này khi đi qua tầng $l+1$ sẽ được cập nhật thành h_i^{l+1} theo công thức tổng quát như sau:

$$h_i^{l+1} = f(h_i^l, \{h_j^l : i \to j\}) \quad (3)$$

Trong đó, $\{i \to j\}$ là tập hợp các nút j có kết nối đến nút i. Tùy thuộc vào các loại mạng nơ-ron đồ thị khác nhau mà dạng hàm f cũng sẽ khác nhau.

Kiến trúc của mạng nơ-ron đồ thị sử trong trong bài toán phân loại ảnh cực được cho ở hình 2.



Hình 2. *Kiến trúc tổng quan của mô hình được sử dụng trong bài toán phân loại ảnh cực [11].*

Sau khi qua L tầng mạng nơ-ron đồ thị, các đặc trưng của các nút ở tầng L – tầng nơ-ron đồ thị cuối cùng, sẽ được kết hợp qua công thức:

$$\frac{1}{V}\sum_{i=0}^{V} h_i^L \quad (4)$$

Kết quả sẽ được cho đi qua một lớp nơ-ron thông thường (multi layer perceptron) để cho ra kết quả dự đoán cuối cùng là xác suất có bệnh và không có bệnh động mạch vành.

3. MÔ PHỎNG, TÍNH TOÁN, THẢO LUẬN

3.1. Dữ liệu thí nghiệm

Bộ dữ liệu gồm 2011 ảnh cực từ các ca chụp SPECT tim được lấy ở *Khoa Y học hạt nhân - Bệnh viện TƯQĐ 108*. Nhãn của dữ liệu được lấy từ kết quả hội chẩn của các bác sĩ. Tập dữ liệu sẽ được chia làm 3 tập con: Train, Validation và Test với số lượng cụ thể từng tập như sau:

Bảng 1. *Thống kê các tập dữ liệu.*

	Số mẫu	**Không có bệnh mạch vành**	**Có bệnh mạch vành**
Train	1648	659	989
Validation	100	46	54
Test	263	99	164
Tổng số	**2011**	**804**	**1207**

Các mô hình thử nghiệm sẽ lần lượt được huấn luyện trên tập Train và kiểm định trên tập Validation cho đến khi không có sự cải thiện trên tập này nữa thì kết thúc. Mô hình kết quả của quá trình huấn luyện sẽ được kiểm tra độ chính xác trên tập Test.

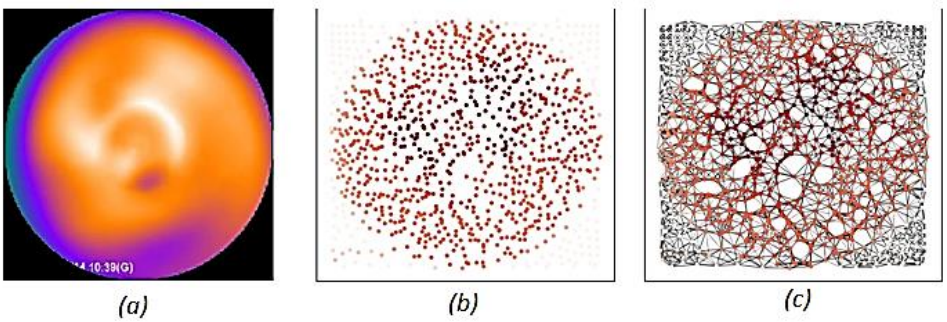
3.2. Chuẩn bị thực nghiệm

Ảnh cực sẽ được chuyển đổi sang dạng đồ thị bằng cách sử dụng các super-pixel. Mỗi super-pixel biểu diễn cho một vùng nhỏ trên ảnh mà các pixel ở trong vùng đó có giá trị độ sáng xấp xỉ nhau. Đầu tiên, các super-pixel ở mỗi ảnh sẽ được tính ra bằng thuật toán SLIC [12], mỗi super-pixel này sẽ được coi là một nút của đồ thị. Ma trận trọng số W của đồ thị sẽ được xây dựng tương tự với thuật toán k-nearest neighbor, với các phần tử được xác định như sau:

$$W_{ij}^{k-NN} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_x^2}\right) \quad (5)$$

Trong đó: x_i, x_j lần lượt là các tọa độ hai chiều trên ảnh của các super-pixel thứ *i* và thứ *j*; σ_x được gọi là tham số tỉ lệ, tham số này được xác định cho từng nút và được tính bằng trung bình khoảng cách của *k* nút gần nhất với nút đó.

Sau khi tính ra ma trận kề, mỗi nút sẽ tạo liên kết với *k* nút gần nó nhất tính theo các giá trị W_{ij}^{k-NN}, vì thế ma trận kề sẽ được cập nhật lại sao cho mỗi nút chỉ liên kết với *k* nút gần nó nhất, các nút còn lại sẽ có trọng số bằng 0. Ví dụ về ảnh cực được chuyển sang dạng đồ thị được cho ở hình 3.



Hình 3. *Ảnh cực được chuyển sang dạng đồ thị bằng các super-pixel: (a) Ảnh gốc; (b) Ảnh chỉ gồm các nút của đồ thị (các super-pixel); (c) Ảnh gồm các các nút và các cạnh của đồ thị*

Đặc trưng của mỗi nút sẽ bao gồm các giá trị màu RGB và tọa độ của super-pixel ứng với nút đó. Đặc trưng của cạnh từ nút i đến nút j sẽ là sẽ là giá trị W_{ij}^{k-NN} .

3.3. Kết quả thử nghiệm và bình luận

Các mô hình được đánh giá dựa trên độ chính xác:

$$\text{Độ } chính\ xác = \frac{Số\ mẫu\ dự\ liệu\ \text{được}\ phân\ loại\ \text{đúng}}{Tổng\ số\ mẫu\ dữ\ liệu} \times 100 \quad (6)$$

Mỗi mô hình sẽ được huấn luyện lặp lại 3 lần, giá trị độ chính xác cuối cùng cho mô hình sẽ được lấy bằng trung bình trên 3 lần huấn luyện này. Bên cạnh đó, chúng tôi cũng tiến hành thử nghiệm với một số kiến trúc dựa trên mạng tích chập (Convolutional Neural Network-CNN) như: VGG11, Inception-v3, Resnet152, Efficientnet-B3 để so sánh. Kết quả độ chính xác được cho ở bảng 2.

Bảng 2. *Kết quả độ chính xác của các mô hình thử nghiệm.*

Mô hình	Độ chính xác
GCN	79.34
GatedGCN	78.45
GAT	76.42
MoNet	**80.10**
VGG11	79.18
InceptionV3	58.94
Resnet-152	77.42
EfficientNet-B3	75.37

Có thể thấy được hiệu năng của các mạng đồ thị không kém hơn so với các mạng tích chập với độ chính xác của cao nhất của mô hình mạng nơ-ron đồ thị đạt được là 80.10% khi sử dụng mạng MoNet. Kết quả này mở ra một con đường mới cho bài toán phân loại ảnh cực, thay vì áp dụng các mạng nơ-ron tích chập thường thấy ở các bài toán phân loại ảnh thì mạng nơ-ron đồ thị cũng là một hướng đi có khả năng cải thiện kết quả của bài toán phân loại ảnh cực.

4. KẾT LUẬN

Từ những kết quả thực nghiệm trong bài báo cho thấy, việc áp dụng các mạng nơ-ron đồ thị cho chẩn đoán bệnh mạch vành dựa trên ảnh SPECT tưới máu cơ tim đem lại các kết quả khả quan. Độ chính xác khi phân loại các ảnh cực của mạng nơ-ron đồ thị có thể so sánh được với các mạng học sâu phổ biến thường được áp dụng cho các bài toán phân loại ảnh.

Nghiên cứu trong bài báo đưa ra một hướng tiếp cận mới cho bài toán phân loại ảnh cực, chuyển từ phân loại ảnh sang phân loại đồ thị với khả năng cho kết quả tương đương hay có thể là tốt hơn phương pháp sử dụng các mạng nơ-ron tích chập.

Trong tương lai, chúng tôi sẽ tiếp tục nghiên cứu để cải thiện hiệu năng của mạng nơ-ron đồ thị cho bài toán phân loại ảnh cực. Các hướng cải thiện sẽ tập trung vào việc thử nghiệm các cách thức xây dựng đồ thị từ ảnh, các kiến trúc mạng nơ-ron đồ thị khác nhau, các lớp pooling cho mạng nơ-ron đồ thị.

TÀI LIỆU THAM KHẢO

[1]. LeCun Y, Bengio Y, Hinton G, "*Deep learning*," Nature, 521 (2015), pp. 436-444, doi: 10.1038/nature14539.

[2]. Shen D, Wu G, Suk H-I, "*Deep learning in medical image analysis*," Annu Rev Biomed Eng, 19 (2017), pp: 221–248, doi: 10.1146/annurev-bioeng-071516-044442.

[3]. Andree Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, "*Derma-tologist-level classification of skin cancer with deep neural networks,*"

Nature, 542 (2017), pp: 115–118, doi: 10.1038/nature21056.01.
[4]. Julian Betancur et. al., "*Deep Learning for Prediction of Obstructive Disease from Fast Myocardial Perfusion SPECT,*" JACC: Cardiovascular Imaging, vol. 11(11) (2018), pp. 1654-1663.
[5]. Duvenaud, David, et al, "*Convolutional networks on graphs for learning molecular fingerprints,"* arXiv preprint arXiv:1509.09292 (2015).
[6]. Cranmer, Miles D., et al, "*Learning symbolic physics with graph networks,"* arXiv preprint arXiv:1909.05862 (2019).
[7]. Kipf, Thomas N., and Max Welling, "*Semi-supervised classification with graph convolutional networks,"* arXiv preprint arXiv:1609.02907 (2016).
[8]. Monti F, Boscaini D, Masci J, Rodola E, Svoboda J, Bronstein MM, "*Geometric deep learning on graphs and manifolds using mixture model CNNs,*" In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (2017).
[9]. Veličković, Petar, et al, "*Graph attention networks,*" arXiv preprint arXiv:1710.10903 (2017).
[10]. Bresson, Xavier, and Thomas Laurent, "*Residual gated graph convnets*," arXiv preprint arXiv:1711.07553 (2017).
[11]. Dwivedi, Vijay Prakash, et al, "*Benchmarking graph neural networks,*" arXiv preprint arXiv:2003.00982 (2020).
[12]. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S, "*SLIC superpixels compared to state-of-the-art superpixel methods,*" *IEEE Trans Pattern Anal Mach Intell*, (2012) ;34(11):2274-2282.

ABSTRACT

DETECTING CORONARY ARTERY DISEASE ON POLAR IMAGES USING GRAPH NEURAL NETWORKS

Nowadays, myocardial perfusion imaging (MPI) with single-photon emission tomography (SPECT) is an advanced method used to detect coronary artery disease (CAD). However, analyzing and diagnosing CAD from SPECT images are a difficult, time-cosuming job and depends heavily on the experience of the doctors. In this paper, we propose a method using graph neural networks on polar maps from SPECT image to detect CAD automatically. Polar maps are transformed into graph using pixel clustering algorithm. Graphs are then fed into the network for training, extracting features and predicting the labels. The experimental models in turn give the following accuracy: GCN: 79.34%; GatedGCN 78.45%; GAT: 76.42%; MoNet: 80.1%

Keywords: Single-photon emission tomography (SPECT); Coronary artery disease; Myocardial perfusion imaging (MPI); Graph neural networks.

Nhận bài ngày 11 tháng 4 năm 2021
Hoàn thiện ngày 25 tháng 8 năm 2021
Chấp nhận đăng ngày 05 tháng 9 năm 2021

Địa chỉ: [1]Trường Đại học Công nghệ, Đại học quốc gia Hà Nội;
[2]Viện Công nghệ thông tin/Viện KH-CN quân sự;
[3]Khoa Trang bị/Bệnh viện Trung ương quân đội 108;
[4]Viện Điện tử - Viễn thông/Trường Đại học Bách khoa Hà Nội;
[5]Trường THPT chuyên Đại học Sư phạm.
*Email: bntung1193@gmail.com (Corresponding author).