

MÔ HÌNH HỌC SÂU TÍCH HỢP 3D FUZZY POOLING CHO PHÂN ĐOẠN CÁC CẤU TRÚC TIM

A DEEP LEARNING MODEL INTEGRATING 3D FUZZY POOLING FOR SEGMENTING CARDIAC STRUCTURES

NGUYỄN VĂN HÙNG^a, PHAN ANH CANG

Trường Đại học Sư phạm Kỹ thuật Vĩnh Long

^aTác giả liên hệ: 23904001@st.vlute.edu.vn

Nhận bài(Received): 05/7/2025; Phản biện (Reviewed):21/8/2025; Chấp nhận(Accepted):22/9/2025

TÓM TẮT

Phân đoạn các cấu trúc tim trên ảnh 3D MRI đóng vai trò quan trọng trong chẩn đoán bệnh tim. Tuy nhiên, các kỹ thuật pooling truyền thống như Max hoặc Average pooling dễ làm mất thông tin quan trọng. Bên cạnh đó, phần lớn nghiên cứu chỉ tập trung vào độ chính xác phân đoạn trên hình ảnh của các mô hình máy học mà thiếu diễn giải kết quả. Trong nghiên cứu này, chúng tôi đề xuất mô hình học sâu 3DMFL-Net tích hợp 3D fuzzy pooling và kỹ thuật Explainable AI (XAI) để phân đoạn các cấu trúc tim bao gồm tâm thất trái, cơ tim và tâm thất phải từ ảnh 3D MRI trục ngắn, đồng thời cung cấp diễn giải kết quả phân đoạn. 3D fuzzy pooling sử dụng logic mờ và hàm thành viên Gaussian để giảm kích thước bản đồ đặc trưng mà vẫn giữ lại thông tin thiết yếu. XAI được áp dụng thông qua 3D Grad-CAM để trực quan hóa vùng ảnh quan trọng trong phân đoạn. Thử nghiệm trên tập dữ liệu M&Ms-2020 đa trung tâm cho thấy mô hình đạt hệ số Dice trung bình 71.1%.

Từ khóa: Bệnh tim mạch, học sâu mờ, 3D fuzzy pooling, 3DMFL-Net.

ABSTRACT

Segmentation of cardiac structures in 3D MRI plays a crucial role in the diagnosis of heart diseases. However, traditional pooling techniques such as Max or Average pooling often result in the loss of important information. Moreover, most studies focus solely on the segmentation accuracy of machine learning models, lacking interpretability of the results. In this study, we propose a deep learning model, 3DMFL-Net, which integrates 3D fuzzy pooling and Explainable AI (XAI) techniques to segment key cardiac structures, including the left ventricle, myocardium, and right ventricle, from short-axis 3D MRI images while providing interpretability of the segmentation results. The 3D fuzzy pooling technique employs fuzzy logic and Gaussian membership functions to downsample feature maps while preserving essential information. XAI is applied using 3D Grad-CAM to visualize the important regions involved in the segmentation process. Experiments on the multi-center M&Ms-2020 dataset show that the model achieved an average Dice score of 71.1%.

Keywords: Cardiovascular disease, Fuzzy deep learning, 3D fuzzy pooling, 3DMFL-Net.

1. GIỚI THIỆU

1.1. Giới thiệu bài toán

Theo tổ chức Y tế Thế giới (WHO) [1]

bệnh tim mạch là nhóm bệnh gây tử vong hàng đầu trên toàn cầu, mỗi năm có khoảng 17,9 triệu người tử vong. Do đó, phân đoạn

chính xác các cấu trúc tim gồm tâm thất trái (LV - Left Ventricle), cơ tim tâm thất trái (MYO - Myocardium) và tâm thất phải (RV - Right Ventricle) trong ảnh 3D MRI tim trực ngắn là bước quan trọng để trích xuất các thông số sinh lý nhằm phân loại bệnh tim. Tuy nhiên, quá trình này mất nhiều thời gian khi thực hiện thủ công. Các mô hình học sâu hiện nay tạo ra nhiều bản đồ đặc trưng kèm theo bước giảm kích thước không gian bằng phương pháp Max pooling hoặc Average pooling. Quá trình này có thể làm mất các thông tin quan trọng [2], trong khi phương pháp fuzzy pooling có thể xử lý tốt đối với dữ liệu y tế không chắc chắn trong vùng pooling [3]. Bên cạnh đó, phần lớn nghiên cứu hiện nay chỉ tập trung vào độ chính xác phân đoạn trên hình ảnh của các mô hình máy học mà thiếu diễn giải kết quả phân đoạn dựa trên các đặc trưng hình ảnh. Đây là yếu tố quan trọng trong ứng dụng lâm sàng. Để giải quyết các vấn đề trên, chúng tôi đề xuất mô hình học sâu mờ 3DMFL-Net (3D Max Fuzzy Light U-Net) để phân đoạn các cấu trúc tim như LV, MYO và RV từ ảnh 3D MRI tim trực ngắn tại hai thời điểm cuối kỳ tâm trương (ED - End-Diastolic) và cuối kỳ tâm thu (ES - End-Systolic). Mô hình được tích hợp phương pháp đề xuất 3D fuzzy pooling nhằm xác định chính xác các đặc trưng tại cửa sổ pooling để giảm kích thước các bản đồ đặc trưng nhưng vẫn giữ được những thông tin quan trọng. Để diễn giải kết quả, chúng tôi áp dụng XAI với phương pháp 3D Grad-CAM nhằm tạo bản đồ nhiệt hiển thị các vùng ảnh quan trọng trong quá trình phân đoạn của 3DMFL-Net.

1.2. Các nghiên cứu liên quan

Nhóm nghiên cứu Qianying Liu et al [4] giới thiệu CS-UNet - mô hình kết hợp giữa mạng tích chập và Transformer theo kiến trúc Encoder-Decoder. Dữ liệu đầu vào được trích xuất đặc trưng qua các lớp

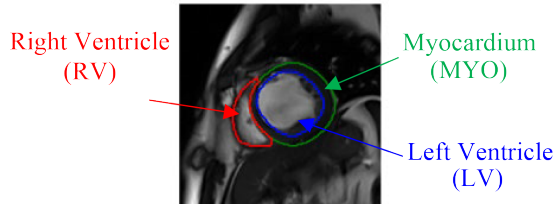
tích chập, sau đó học mỗi liên kết qua các khối Convolutional Swin Transformer. Mô hình đạt hệ số Dice 95.4% cho vùng LV, 89.5% cho vùng MYO và 89.2% cho vùng RV trên bộ dữ liệu ACDC-2017. Nhóm nghiên cứu Lei Li et al [5] giới thiệu mô hình STDGNs kết hợp thông tin không gian và hình dạng để cải thiện độ chính xác trong phân đoạn ba vùng chính của tim gồm LV, MYO và RV. Kết quả thực nghiệm trên tập dữ liệu M&Ms-2020 cho thấy mô hình đạt hệ số Dice 76.6% đối với vùng LV, 71.5% đối với vùng MYO và 63.6% đối với vùng RV. Italo Francyles Santos da Silva et al [6] đã đề xuất kiến trúc mạng EAIS-Net để phân đoạn vùng LV, MYO, đồng thời phát triển IRAX-Net nhằm phân đoạn vùng RV. Phương pháp này được đánh giá trên tập dữ liệu M&Ms-2020 và đạt hệ số Dice lần lượt là 83.3% cho vùng LV, 70.8% cho vùng MYO và 74.6% cho vùng RV. Nhóm nghiên cứu Jiapeng Li et al [7] đã đề xuất một khung Unsupervised Domain Adaptation hai giai đoạn nhằm phân đoạn các cấu trúc tim trên ảnh MRI, bao gồm LV, MYO và RV. Thực nghiệm trên tập dữ liệu M&Ms-2020 đạt hệ số Dice lần lượt là 79.0% cho vùng LV, 88.1% cho vùng MYO và 86.3% cho vùng RV.

2. CƠ SỞ LÝ THUYẾT

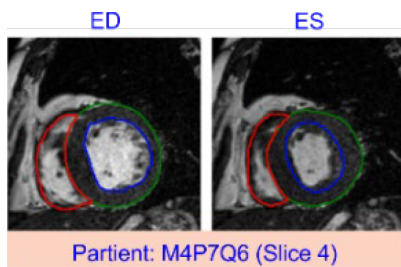
2.1. Các cấu trúc giải phẫu trên ảnh MRI tim trực ngắn.

Cấu trúc giải phẫu trên ảnh MRI tim trực ngắn bao gồm ba vùng quan trọng là LV, MYO và RV. Hình 1 trình bày ba cấu trúc quan trọng gồm LV, MYO và RV ở lát cắt thứ 5 tại thời điểm ES của bệnh nhân S1S3Z7 trong tập dữ liệu M&Ms-2020 [8]. Chu kỳ tim có hai giai đoạn gồm tâm thu khi các tâm thất co lại để đẩy máu vào mạch máu và tâm trương khi tâm thất được đổ đầy máu [9]. Cuối kỳ tâm thu (ES - End-Systolic) là thời điểm mà thể tích máu trong tâm thất đạt mức thấp nhất, trong khi cuối kỳ tâm trương

(ED - End-Diastolic) là thời điểm thể tích máu trong tâm thất đạt mức cao nhất. Hình 2 trình bày cấu trúc LV, MYO và RV ở lát cắt thứ 4 tại ED và ES của bệnh nhân M4P7Q6 trong tập dữ liệu M&Ms-2020 [8].



Hình 1. Cấu trúc LV, MYO và RV ở lát cắt thứ 5 tại thời điểm ES của bệnh nhân S1S3Z7 trong tập dữ liệu M&Ms-2020.



Hình 2. Cấu trúc LV, MYO và RV ở lát cắt thứ 4 tại thời điểm ED và ES của bệnh nhân M4P7Q6 trong tập dữ liệu M&Ms-2020.

2.2. Các độ đo đánh giá mô hình

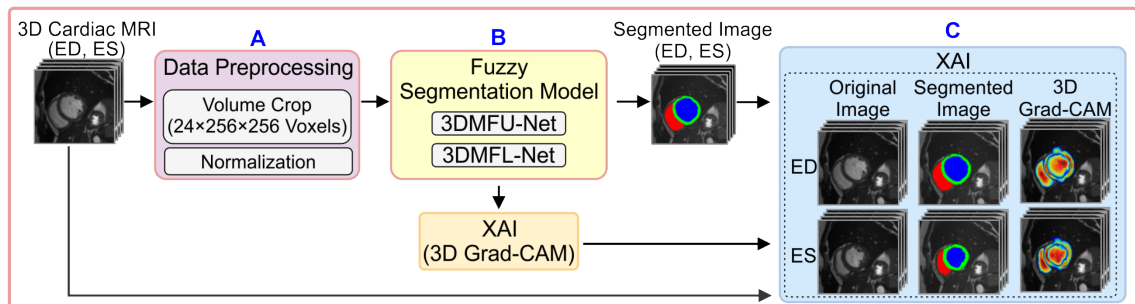
Hệ số Dice được sử dụng để đánh giá mức độ tương đồng giữa vùng phân đoạn và vùng thực tế trong ảnh, được xác định tại công thức (1). Hàm mất mát Cross-Entropy Loss (LCE) đóng vai trò quan trọng trong quá trình tối ưu hóa các tham số, nhằm giảm sự sai lệch dự đoán so với thực tế. Hàm mất mát LCE được xác định tại công thức (2).

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

$$LCE = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (2)$$

Trong đó, A là tập hợp các pixel thuộc lớp được phân đoạn, B là tập hợp các pixel thuộc lớp thực tế trong ảnh, N là tổng số lượng pixel trên ảnh, $p_{i,c} \in [0,1]$ là xác suất dự đoán điểm ảnh i thuộc lớp c, $y_{i,c} \in \{0,1\}$ là nhãn thực tế của điểm ảnh i.

3. PHƯƠNG PHÁP ĐỀ XUẤT

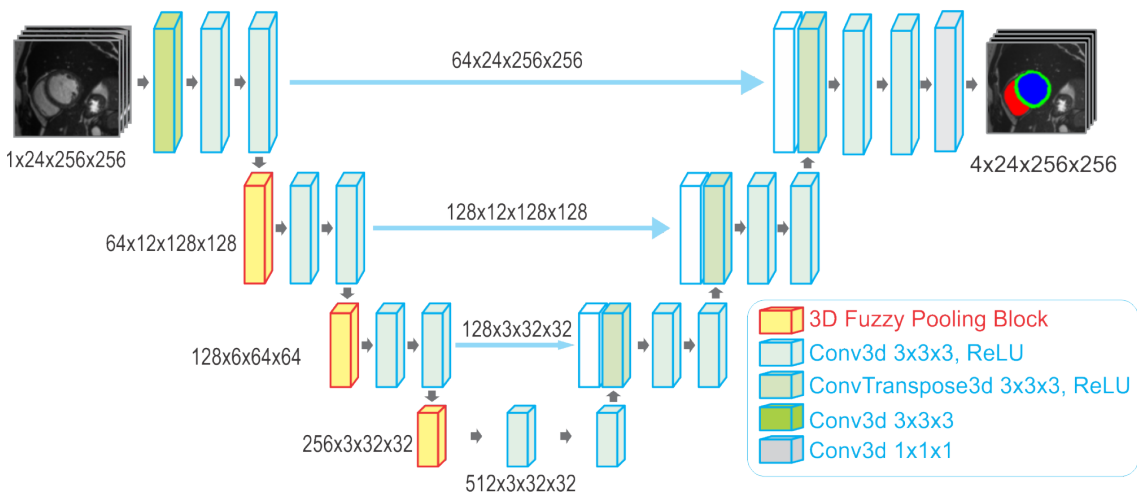


Hình 3. Mô hình tổng quát phân đoạn các cấu trúc tim trên ảnh 3DMRI trực tuyến.

Trong nghiên cứu này, chúng tôi đề xuất phương pháp phân đoạn ảnh 3D MRI tim trực tuyến và giải thích kết quả bằng XAI, được minh họa chi tiết trong Hình 3. Cụ thể, giai đoạn đầu thực hiện tiền xử lý nhằm trích xuất vùng quan tâm (ROI) nhằm giới hạn không gian xử lý vào khu vực chứa các cấu trúc tim quan trọng, được minh họa phần A trong Hình 3. Tiếp theo, vùng ROI thu được, được sử dụng làm đầu vào cho giai đoạn thứ hai, trong đó

tiến hành phân đoạn các vùng LV, MYO và RV bằng mô hình học sâu mờ đề xuất. Quá trình huấn luyện được thực hiện trên hai mô hình học sâu mờ để lựa chọn mô hình có hiệu suất tối ưu, gồm 3DMFU-Net và 3DMFL-Net, được minh họa phần B trong Hình 3. Hai mô hình 3DMFU-Net và 3DMFL-Net được phát triển dựa trên kiến trúc lần lượt 3D U-Net [10] và Light U-Net [11]. Cuối cùng là diễn giải kết quả, được minh họa phần C trong Hình 3.

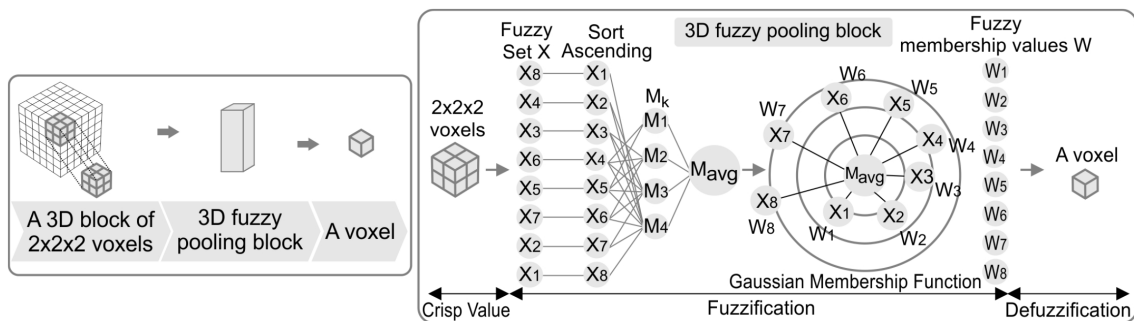
3.1. Mô hình mạng 3D Max Fuzzy U-Net (3DMFU-Net) đề xuất



Hình 4. Minh họa mô hình mạng học sâu mờ 3DMFU-Net đề xuất.

Chúng tôi đề xuất mô hình mạng học sâu mờ 3DMFU-Net được cải tiến của mạng 3D U-Net [10], nhằm phân đoạn các vùng LV, MYO và RV trên ảnh 3D MRI tim trực tuyến. 3DMFU-Net gồm 2 nhánh chính Encoder, Decoder và được minh họa chi tiết trong Hình 4. Encoder thực hiện trích xuất các đặc trưng trên ảnh thông qua việc tạo ra 512 bản đồ đặc trưng. Cụ thể, Encoder nhận đầu vào là ảnh 3D MRI có kích thước 1x24x256x256 (số kênh, chiều sâu, chiều cao, chiều rộng) voxels. Encoder bắt đầu với một lớp tích chập, tiếp theo là ba tầng tích chập, mỗi tầng gồm hai lớp tích

chập liên tiếp với bộ lọc 3x3x3 voxels và hàm ReLU. Sau mỗi tầng tích chập, chúng tôi sử dụng phương pháp 3D fuzzy pooling thay thế Max pooling nhằm giảm kích thước không gian của các bản đồ đặc trưng, nhưng vẫn giữ lại những thông tin quan trọng. Cuối nhánh Encoder là một lớp tích chập sâu nhằm tăng cường trích xuất đặc trưng. Phương pháp 3D fuzzy pooling được minh họa trong Hình 5. Nhánh Decoder sử dụng các phép tích chập Deconvolution với stride bằng 2 để tăng dần kích thước không gian ảnh và giảm số lượng bản đồ đặc trưng.



Hình 5. Phương pháp 3D fuzzy pooling đề xuất.

Cụ thể, phương pháp 3D fuzzy pooling sử dụng lý thuyết tập mờ trong fuzzy logic để xác định mức độ thuộc về của các đặc

trung trong cửa sổ pooling. Phương pháp 3D fuzzy pooling áp dụng hàm thành viên Gaussian được xác định tại công thức (6),

một hàm thành viên trong lý thuyết tập mờ, để xác định giá trị độ thuộc về của từng phần tử từ 0 đến 1, với giá trị càng gần 1 thì phần tử đó càng thuộc về tập mờ và ngược lại. Cụ thể, hàm thành viên Gaussian có hình dạng giống một chiếc chuông, với giá trị cực đại bằng 1 tại trung tâm và giảm dần khi phần tử đầu vào cách xa trung tâm. Độ rộng của chiếc chuông được quyết định bởi tham số độ lệch chuẩn: khi giá trị này nhỏ, đường cong trở nên hẹp và tập trung, trong khi giá trị lớn tạo ra đường cong trải rộng hơn. Chúng tôi lựa chọn hàm Gaussian vì đặc tính liên tục và ít tham số, giúp mô tả tốt sự chuyển tiếp mờ và tăng tính ổn định của mô hình khi xử lý dữ liệu nhiễu. So với Gaussian, hàm tam giác có ưu điểm đơn giản và dễ tính toán hơn, tuy nhiên nó không trơn tại các điểm góc, dẫn đến sự thay đổi đột ngột trong mức độ thành viên và làm giảm khả năng biểu diễn sự chuyển tiếp mượt mà.

Phương pháp 3D fuzzy pooling có ba giai đoạn chính gồm xác định giá trị đầu vào, xác định giá trị mờ và giải mờ. Giai đoạn 1, xác định giá trị đầu vào (Crisp value). Để xác định giá trị đầu vào, từ một khối 3D pooling có kích thước $2 \times 2 \times 2$ voxels trong bản đồ đặc trưng, thực hiện trích xuất tám giá trị này. Quá trình này cho phép thu thập các giá trị cần thiết trong vùng pooling, từ đó cung cấp giá trị đầu vào cho các bước xử lý tiếp theo. Giai đoạn 2, xác định giá trị mờ (Fuzzification). Mỗi giá trị trong vùng 3D pooling sẽ được gán một ký hiệu tương ứng từ x_1 đến x_8 . Gọi $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ là tập hợp các giá trị được trích xuất trong giai đoạn 1 và xem tập hợp X là một tập mờ (fuzzy set). Các giá trị trong tập mờ X được sắp

xếp theo thứ tự tăng dần. Việc sắp xếp này giúp làm rõ xu hướng chung của dữ liệu và giảm ảnh hưởng của nhiễu. Thực hiện xác định giá trị trung bình tổng thể (m_{avg}) của tập mờ X . Cụ thể, m_{avg} được xác định theo công thức (4) là trung bình của các giá trị trung bình cục bộ m_k được xác định dựa trên công thức (3). Sau đó, sử dụng hàm thành viên Gaussian để xác định giá trị mờ cho tập mờ X , được xác định tại công thức (6). Tương ứng với các giá trị trong tập mờ X , tập $W = (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8)$, biểu thị giá trị mức độ thành viên của các phần tử trong X . Giai đoạn 3, giải mờ (Defuzzification). Giải mờ là quá trình sử dụng tập W và tập mờ X để ước lượng giá trị đầu ra của của phép pooling ($P_{pooling}$). $P_{pooling}$ đại diện cho mức độ thuộc về của các phần tử trong tập X , được xác định tại công thức (7).

$$m_k(X) = \frac{1}{2k} \sum_{i=q-k+1}^{q+k} X_i \quad (3)$$

$$m_{avg} = \frac{1}{q} \sum_{k=1}^q m_k \quad (4)$$

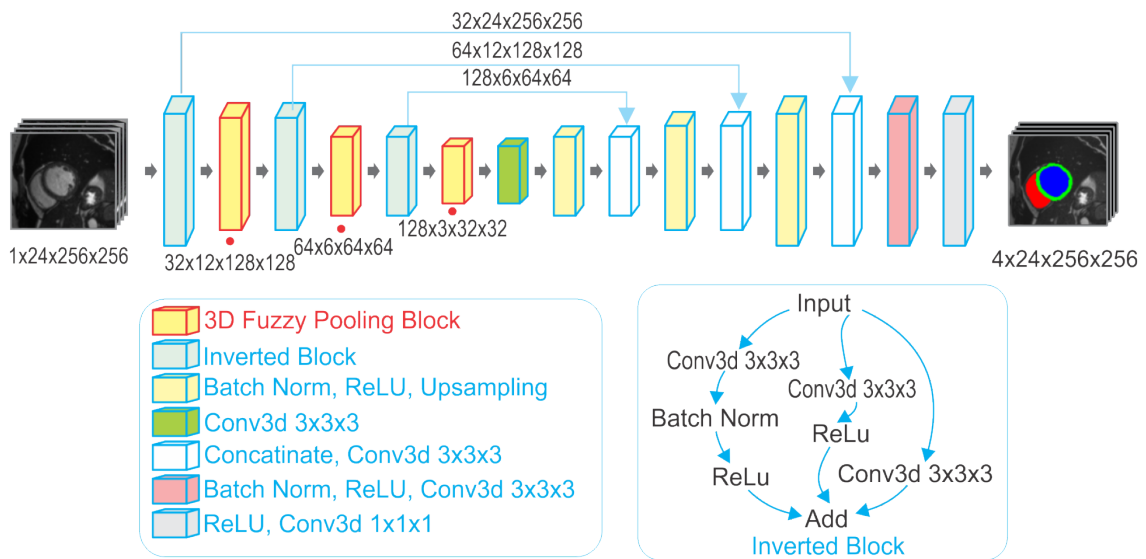
$$\sigma_X = \sum_{i=1}^n (|X_i - m_{avg}|) \quad (5)$$

$$w_i = e^{-\frac{(X_i - m_{avg})^2}{2\sigma_X^2}} \quad (6)$$

$$P_{pooling} = \max(w_i X_i) \quad (7)$$

Trong đó, n là số phần tử trong tập mờ X , $q = n/2$ là một hằng số dùng để xác định phạm vi số lượng phần tử của tập mờ X để tính m_k , k là số lần tính giá trị m_k với điều kiện $1 \leq k \leq q$, x_i là giá trị của các phần tử thứ i trong tập mờ X . w_i là giá trị mức độ thành viên của phần tử thứ i trong tập mờ X , i là vị trí của các phần tử trong tập mờ X , σ_X là phương sai của tập mờ X . Phương sai đo mức độ phân tán của các đặc trưng xung quanh giá trị m_{avg} .

3.2. Mô hình mạng 3D Max Fuzzy Light U-Net (3DMFL-Net) đề xuất.



Hình 6. Minh họa mô hình mạng học sâu mờ 3DMFL-Net đề xuất.

Chúng tôi đề xuất mô hình học sâu mờ 3DMFL-Net được cải tiến từ 3D Light U-Net [11], nhằm phân đoạn các vùng LV, MYO và RV trên ảnh 3D MRI tim trực ngắn. 3DMFL-Net gồm 2 nhánh chính Encoder và Decoder được minh họa chi tiết trong Hình 6. Encoder có số lượng bản đồ đặc trưng tăng từ 1 lên 256, trong khi kích thước không gian của bản đồ đặc trưng giảm dần từ $24 \times 256 \times 256$ xuống còn $3 \times 32 \times 32$ voxels. Encoder gồm ba khối Inverted và một lớp tích chập cuối sử dụng bộ lọc $3 \times 3 \times 3$ voxels. Mỗi khối Inverted gồm ba nhánh tích chập song song: nhánh thứ nhất tích chập với bộ lọc $3 \times 3 \times 3$ voxels kết hợp Batch Normalization và hàm kích hoạt ReLU; nhánh thứ hai sử dụng tích chập với bộ lọc $3 \times 3 \times 3$ voxels và ReLU và nhánh thứ ba chỉ thực hiện tích chập với bộ lọc $3 \times 3 \times 3$ voxels. Đầu ra của khối Inverted là tổng các bản đồ đặc trưng từ ba nhánh. Sau ba khối Inverted, mô hình áp dụng phương pháp 3D fuzzy pooling (được trình bày trong mô hình 3DMFU-Net) thay cho Max pooling để giảm kích thước không gian nhưng vẫn giữ lại thông

tin quan trọng. Nhánh Decoder sử dụng lớp tích chập và upsampling với stride bằng 2 để tăng kích thước không gian và giảm số lượng bản đồ đặc trưng.

3.3. Diễn giải kết quả phân đoạn

Chúng tôi đề xuất mô hình học sâu mờ kết hợp XAI với phương pháp 3D Grad-CAM để diễn giải kết quả phân đoạn, được minh họa phần C trong Hình 3. Kết quả diễn giải bao gồm: a) hiển thị ảnh 3D MRI tim trực ngắn tại hai thời điểm sinh lý quan trọng là ED và ES; b) hiển thị kết quả phân đoạn với các vùng phân đoạn được hiển thị và đánh dấu bằng màu sắc rõ ràng để hỗ trợ trực quan bao gồm vùng LV màu xanh lam, vùng MYO màu xanh lá và vùng RV màu đỏ; c) hiển thị bản đồ nhiệt bằng phương pháp 3D Grad-CAM được tính toán theo Deepa Darshini Gunashekar et al [12]. Các bản đồ nhiệt này được mã hóa màu theo dải liên tục từ xanh đến đỏ, trong đó các vùng màu đỏ biểu thị khu vực có mức độ ảnh hưởng cao đến quyết định của mô hình, màu xanh thể hiện các vùng ít được chú ý.

4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1. Môi trường cài đặt và tập dữ liệu

Môi trường cài đặt sử dụng Google Colab với A100 GPU; Ram GPU là 40 GB; bộ nhớ lưu trữ là 235.7 GB. Mô hình được xây dựng và huấn luyện bằng thư viện PyTorch.

Tập dữ liệu M&Ms-2020 [8] có 345 bệnh nhân, mỗi bệnh nhân có 2 ảnh 3D MRI tim trực ngắn ở thời điểm ED và ES. M&Ms-2020 có nhiều bệnh nhân được chụp từ thiết bị MRI của bốn nhà cung cấp là Vendor A, Vendor B, Vendor C, Vendor D và được mô tả trong Bảng 1.

Bảng 1. Số lượng ảnh 3D MRI tim trực ngắn trong tập dữ liệu thực nghiệm M&Ms-2020.

| | Vendor A | | Vendor B | | Vendor C | | Vendor D | | Total |
|------------|----------|----|----------|----|----------|----|----------|----|-------|
| | ED | ES | ED | ES | ED | ES | ED | ES | |
| Training | 75 | 75 | 75 | 75 | 25 | 25 | 0 | 0 | 690 |
| Validation | 4 | 4 | 10 | 10 | 10 | 10 | 10 | 10 | |
| Testing | 16 | 16 | 40 | 40 | 40 | 40 | 40 | 40 | |

4.2. Tham số huấn luyện mô hình

Chúng tôi đề xuất các tham số huấn luyện mô hình được trình bày chi tiết trong Bảng 2.

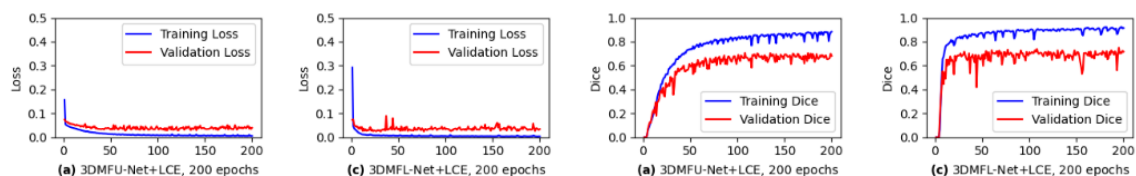
Bảng 2. Các tham số huấn luyện mô hình

| Proposed Model | Epoch | Learning rate | Loss | Num classes | Batch size |
|----------------|-------|---------------|---------|-------------|------------|
| 3DMFU-Net | 200 | 1e-4 | LCE (2) | 4 | 2 |
| 3DMFL-Net | 200 | 1e-4 | LCE (2) | 4 | 2 |

4.3. Kết quả huấn luyện

Hình 7 trình bày sự thay đổi giá trị Loss và hệ số Dice trong suốt quá trình huấn luyện. Cụ thể: (a) 3DMFU-Net đạt giá trị loss thấp nhất là 0.005 trên tập training và 0.034 trên tập validation tại epoch thứ 178; (b) 3DMFL-Net thể hiện xu hướng giảm

loss đều đặn và ổn định, đạt mức thấp nhất là 0.004 trên tập training và 0.026 trên tập validation tại epoch thứ 195; (c) 3DMFU-Net đạt hệ số Dice cao nhất là 87.4% trên tập training và 70.0% trên tập validation; (d) 3DMFL-Net đạt hệ số Dice lần lượt là 90.4% trên tập training và 74.9% trên tập validation.



Hình 7. Biểu đồ thể hiện giá trị Loss và hệ số Dice trong quá trình huấn luyện của hai mô hình 3DMFU-Net và 3DMFL-Net.

Để làm rõ hơn về thời gian xử lý và độ phức tạp tính toán, chúng tôi đã tiến hành so sánh hai mô hình đề xuất. Cụ thể, thời gian huấn luyện của 3DMFL-Net là 296 phút, trong khi 3DMFU-Net cần 373 phút trên cùng môi trường tính toán. Bên cạnh đó, 3DMFL-Net đạt validation Dice 74.9%, cao hơn so với 70% của 3DMFU-Net. Kết quả này cho thấy 3DMFL-Net không chỉ rút ngắn đáng kể thời gian huấn luyện mà còn đạt hiệu suất phân đoạn cao hơn, chứng tỏ mô hình mang lại sự cân bằng tốt giữa độ chính xác và chi phí tính toán, từ đó có tiềm năng ứng dụng cao hơn trong thực tế lâm sàng.

4.4. Kết quả kiểm thử

Thứ nhất, để đánh giá hiệu suất của hai mô hình 3DMFU-Net và 3DMFL-Net, chúng tôi tiến hành so sánh độ chính xác phân đoạn ảnh 3D MRI tim trực ngực từ các nhà cung cấp thiết bị khác nhau (như đã trình bày trong phần 4.1). Cụ thể, bảng 3 trình bày hệ số Dice (%) cho ba vùng

giải phẫu chính gồm LV, MYO, RV tương ứng với hai mô hình đề xuất. Kết quả cho thấy mô hình 3DMFL-Net đạt hệ số Dice trung bình là 71.1% trên tập kiểm thử, cao hơn 6.2% so với 3DMFU-Net với giá trị 64.9%. Cụ thể, hiệu suất mô hình 3DMFL-Net cao nhất ở thiết bị từ nhà cung cấp B với 80.2%, tiếp theo là thiết bị từ nhà cung cấp A đạt 73.8%, thiết bị từ nhà cung cấp C đạt 68.2% và thấp nhất ở thiết bị từ nhà cung cấp D với 62.2%. Đáng chú ý, tác giả tập dữ liệu M&Ms-2020 đã thiết kế sao cho ảnh 3D MRI tim trực ngực được chụp từ thiết bị của nhà cung cấp D chỉ xuất hiện trong tập testing và validation, hoàn toàn không có trong training, từ đó mô hình phải khái quát hóa trên dữ liệu chưa từng thấy trong quá trình huấn luyện. Điều này vừa phản ánh sự khác biệt giữa thiết bị của nhiều nhà cung cấp, vừa chứng minh tính ổn định, khả năng khái quát và tiềm năng ứng dụng của mô hình trong môi trường lâm sàng đa thiết bị.

Bảng 3. Hệ số Dice (%) giữa 3DMFU-Net và 3DMFL-Net từ nhiều nhà cung cấp thiết bị.

| Proposed Model | Vendor | LV | MYO | RV | Mean |
|----------------|--------|------|------|------|------|
| 3DMFU-Net | A | 79.9 | 65.7 | 65.2 | 70.3 |
| | B | 81.1 | 74.5 | 74.5 | 76.7 |
| | C | 71.8 | 60.4 | 55.4 | 62.5 |
| | D | 59.0 | 42.7 | 47.9 | 49.9 |
| | Mean | 73.0 | 60.8 | 60.8 | 64.9 |
| 3DMFL-Net | A | 81.8 | 68.8 | 70.8 | 73.8 |
| | B | 84.6 | 76.6 | 79.5 | 80.2 |
| | C | 75.3 | 65.6 | 63.8 | 68.2 |
| | D | 69.5 | 56.2 | 60.8 | 62.2 |
| | Mean | 77.8 | 66.8 | 68.7 | 71.1 |

Thứ hai, chúng tôi đánh giá hiệu quả của phương pháp 3D fuzzy pooling bằng

cách so sánh với Max pooling và Average pooling. Ba phương pháp pooling được

chúng tôi huấn luyện trên cùng mô hình 3DMFL-Net. Kết quả từ Bảng 4 cho thấy khi áp dụng phương pháp 3D fuzzy pooling, mô hình 3DMFL-Net đạt hiệu suất

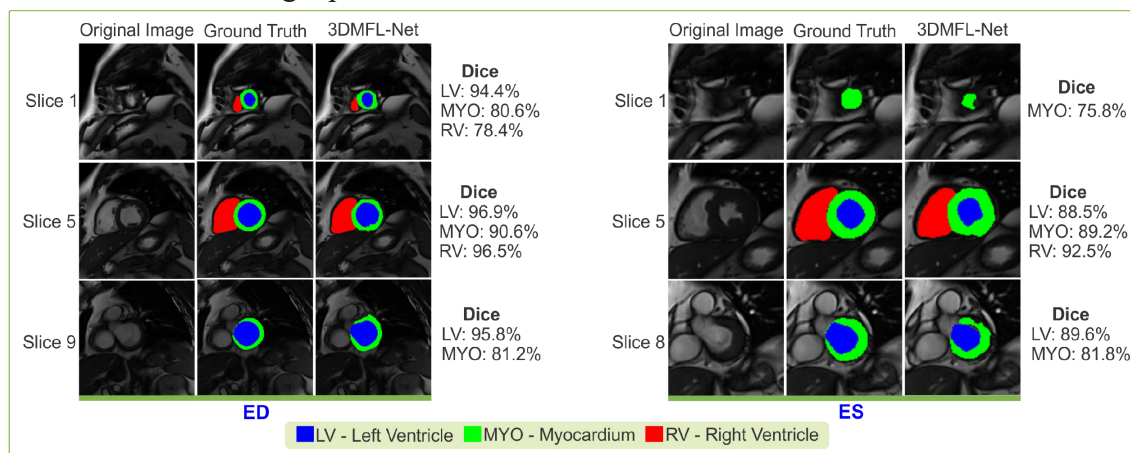
phân đoạn cao nhất với hệ số Dice trung bình là 71.1%, vượt trội so với khi sử dụng Max pooling là 70.2% và Average pooling là 69.3%.

Bảng 4. Hệ số Dice (%) các phương pháp pooling trên cùng mô hình 3DMFL-Net.

| Proposed Model | Pooling method | LV | MYO | RV | Mean |
|----------------|------------------|------|------|------|------|
| 3DMFL-Net | Average pooling | 76,9 | 64,0 | 67,0 | 69,3 |
| | Max pooling | 78,0 | 66,4 | 66,2 | 70,2 |
| | 3D Fuzzy pooling | 77,8 | 66,8 | 68,7 | 71,1 |

Hình 8 minh họa kết quả phân đoạn của 3DMFL-Net trên bệnh nhân I5L3S2 tại thời điểm ED và ES được chụp từ thiết bị của Vendor B, trong tập dữ liệu M&Ms-

2020. Ảnh theo trục dọc ED có 9 lát cắt, ES có 8 lát cắt, giúp minh họa kết quả phân đoạn theo chiều sâu của ảnh.



Hình 8. Ảnh phân đoạn trên từng lát cắt của bệnh nhân I5L3S2 tại thời điểm ED và ES trong tập dữ liệu M&Ms-2020.

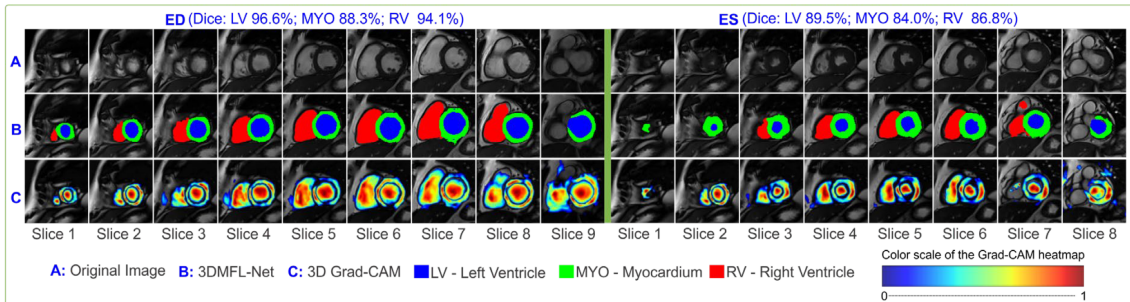
4.5. Diễn giải kết quả phân đoạn bằng 3D Grad-CAM

Chúng tôi trình bày quá trình diễn giải kết quả của mô hình 3DMFL-Net đối với bệnh nhân I5L3S2 trong tập kiểm thử M&Ms-2020 và minh họa trong hình 9. Ngoài hiển thị các lát cắt trên ảnh gốc, ảnh phân đoạn tại hai thời điểm sinh lý quan trọng là ED và ES cũng được hiển thị với ba cấu trúc chính là LV màu xanh lam, MYO màu xanh lá và RV màu đỏ, cho thấy sự tương ứng cao với cấu trúc trên

ảnh gốc. Ngoài ra, XAI bằng phương pháp 3D Grad-CAM tạo ra các bản đồ nhiệt thể hiện mức độ chú ý của mô hình trên từng lát cắt MRI. Bản đồ nhiệt được mã hóa từ xanh đến đỏ, trong đó vùng đỏ cho thấy khu vực có ảnh hưởng lớn đến quyết định của mô hình, còn vùng xanh thể hiện khu vực ít được chú ý. Việc chồng bản đồ nhiệt lên ảnh gốc giúp bác sĩ quan sát trực quan vùng tập trung của mô hình dọc theo chiều sâu ảnh. Các lát cắt tại hai thời điểm ED và ES được trình bày riêng biệt, cho phép bác sĩ so sánh sự thay đổi theo chu kỳ tim –

một yếu tố quan trọng trong đánh giá chức năng tim. Cách trực quan hóa này không chỉ hỗ trợ kiểm chứng quyết định phân đoạn mà còn giúp phát hiện các trường hợp mô hình tập trung sai lệch, từ đó tăng

độ tin cậy và tính diễn giải trong ứng dụng lâm sàng. Điều này đặc biệt quan trọng khi đưa các mô hình trí tuệ nhân tạo vào hỗ trợ chẩn đoán và điều trị.



Hình 9. Diễn giải kết quả phân đoạn bằng 3D Grad-CAM của bệnh nhân I5L3S2 tại hai thời điểm ED và ES trong tập kiểm thử M&Ms-2020.

4.6 So sánh đánh giá và thảo luận.

Bảng 5 so sánh kết quả phân đoạn giữa mô hình 3DMFL-Net và các nghiên cứu hiện có. Trong đó, mô hình 3DMFL-Net đạt hệ số Dice trung bình 71,1% trên tập kiểm thử. Kết quả này cao hơn kết quả của Lei Li et al [5] đạt hệ số Dice trung bình 70,5%. Bên cạnh đó, mô hình đề xuất vẫn cho thấy hệ số Dice tương đương so với

phương pháp của Italo Francyles Santos da Silva et al [6] đạt 76,2%. Đáng chú ý, mô hình đề xuất 3DMFL-Net có ưu điểm nổi bật là khả năng diễn giải kết quả phân đoạn bằng XAI với phương pháp 3D Grad-CAM trong khi hầu hết các phương pháp khác đều thiếu cơ chế diễn giải kết quả phân đoạn, ảnh hưởng độ tin cậy trong ứng dụng lâm sàng.

Bảng 5. Kết quả phân đoạn của mô hình đề xuất cùng các nghiên cứu liên quan.

| Authors | Methods | Mean Dice (%) | XAI |
|---|--------------------|---------------|-------------|
| Proposed | 3DMFV-Net | 71,1 | 3D Grad-CAM |
| Lei Li et al [5] | STDGNs | 70,5 | - |
| Italo Francyles Santos da Silva et al [6] | EAIS-Net, IRAX-Net | 76,2 | - |

5. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đề xuất mô hình học sâu mờ 3DMFL-Net kết hợp phương pháp XAI nhằm phân đoạn các vùng tâm thất trái, cơ tim và tâm thất phải, đồng thời diễn giải kết quả phân đoạn. Các đóng góp quan trọng của nghiên cứu bao gồm: (1) đề xuất phương pháp 3D fuzzy pooling với hàm thành viên Gaussian nhằm

giảm kích thước các bản đồ đặc trưng nhưng vẫn giữ các đặc trưng quan trọng, điều đó cho thấy có nhiều ưu điểm hơn các phương pháp Max pooling và Average pooling trong các mô hình học sâu; (2) phát triển mô hình Light UNet từ 2D sang 3D, đồng thời tích hợp phương pháp 3D fuzzy pooling để cải thiện khả năng phân đoạn trong không gian ba chiều; (3) đề xuất mô hình phân đoạn

mờ 3DMFL-Net kết hợp XAI nhằm hỗ trợ bác sĩ trong việc phân đoạn các cấu trúc tim, đồng thời cung cấp diễn giải dựa trên đặc trưng hình ảnh. Thực nghiệm trên tập dữ liệu M&Ms-2020 đa trung tâm cho thấy mô hình đạt hệ số Dice trung bình 71.1%, khẳng định tính ổn định và khả năng khái quát hóa trong môi trường lâm sàng đa thiết bị. Nghiên cứu đạt được nhiều kết quả khả quan nhưng vẫn tồn tại một số hạn chế. Thứ nhất, mô hình mới chỉ thực hiện phân đoạn và dự đoán trên ảnh 3D MRI tim trực ngắn, chưa khai thác ảnh trực dài. Thứ hai,

dữ liệu thực nghiệm chủ yếu dựa trên tập M&Ms-2020 và độ sâu ảnh 3D MRI tim trực ngắn chỉ giới hạn ở 24 lát cắt, nên khả năng tổng quát và xử lý các ảnh có chiều sâu lớn hơn còn hạn chế. Trong tương lai, chúng tôi sẽ mở rộng nghiên cứu trên tập M&Ms-2 2021 để đánh giá tính tổng quát trên cả hai loại ảnh trực ngắn và trực dài, đồng thời phát triển phương pháp 3D fuzzy pooling với hàm thành viên hình thang, kết hợp nhiều mô hình theo fuzzy SUNEGO và triển khai trong môi trường xử lý song song nhằm đáp ứng yêu cầu về dữ liệu lớn.

TÀI LIỆU THAM KHẢO

- [1] WHO, “Cardiovascular diseases,” [Online]. Available: <https://www.who.int/health-topics/cardiovascular-diseases>. [Accessed 10 12 2024].
- [2] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai and D. Rueckert, “Deep Learning for Cardiac Image Segmentation: A Review,” *Frontiers in Cardiovascular Medicine*, vol. 7, 2020.
- [3] Riandini, E. M. Yuniarno, K. E. Purnama, M. Aritsugi and M. H. Purnomo, “Investigating diverse loss functions for myocardium ring segmentation in Cardiac Magnetic Resonance images using fuzzy pooling,” *Array*, vol. 26, p. 100382, 2025.
- [4] Q. Liu, C. Kaul, C. Anagnostopoulos, R. M. Smith and F. Deligianni, “Optimizing Vision Transformers for Medical Image Segmentation,” 26 10 2022. [Online]. Available: <https://arxiv.org/pdf/2210.08066>.
- [5] L. Li, V.A. Zimmer, W. Ding, F. Wu, L. Huang, J.A. Schnabel and X. Zhuang, “Random Style Transfer based Domain Generalization Networks Integrating Shape and Spatial Information,” 3 9 2020. [Online]. Available: <https://arxiv.org/pdf/2008.12205>.
- [6] I. F. S. d. Silva, A. C. Silva, A. C. d. Paiva, M. Gattass and A. M. Cunha, “A Multi-Stage Automatic Method Based on a Combination of Fully Convolutional Networks for Cardiac Segmentation in Short-Axis MRI,” *Applied Sciences*, vol. 17, p. 7352, 2024.
- [7] J. Li, Y. Lv, L. Xu and L. Qi, “An Unsupervised Domain Adaptation Model Based on Multi-Level Joint Alignment for Multi-Modal Cardiac Image Segmentation,” *Procedia Computer Science*, vol. 226, pp. 106-112, 2025.
- [8] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martín-Isla, A. Sojoudi and P. M. Full, “Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, 2021.

- [9] S. I. Flores-Alonso, B. Tovar-Corona and R. Luna-García, “Deep Learning Algorithm for Heart Valve Diseases,” *Applied Sciences*, vol. 12 (8), p. 3780, 2022.
- [10] X. Feng, N. J. Tustison, S. H. Patel and C. H. Meyer, “Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features,” *Frontiers in Computational Neuroscience*, vol. 14, 8 4 2020.
- [11] M. Irshad, M. Yasmin, M. I. Sharif, M. Rashid, M. I. Sharif and S. Kadry, “A Novel Light U-Net Model for Left Ventricle Segmentation Using MRI,” *Mathematics*, vol. 11, 2023.
- [12] D. D. Gunashekar, L. Bielak, L. Hägele and B. Oerther, “Explainable AI for CNN-based prostate tumor segmentation in multi-parametric MRI correlated to whole mount histopathology,” *Radiation Oncology*, 2022.