DOI: 10.15625/vap.2022.0256

ỨNG DỤNG MẠNG GAN CÓ ĐIỀU KIỆN PHÁT HIỆN VÙNG BẤT THƯỜNG TRÊN ẢNH MRI NÃO

Trần Nguyễn Minh Thư[1], Bùi Thanh Liêm[1], Lê Minh Lợi[1,2], Phù Trí Nghĩa[2], Phạm Nguyên Khang[1]

[1]Khoa Công nghệ Thông tin và Truyền thông, Đại học Cần Thơ

[2]Trường Đại học Y Dược Cần Thơ

tnmthu@ctu.edu.vn, liemkg1234@gmail.com, leminhloi@gmail.com, ptnghia@ctump.edu.vn, pnkhang@ctu.edu.vn

TÓM TẮT: *Ngày nay, việc áp dụng thị giác máy tính để phát hiện và chẩn đoán bất thường bên trong não đang được áp dụng rộng rãi. Trong bài báo này, mô hình phân lớp (EfficienNetB7) kết hợp với mô hình mạng GAN có điều kiện (Pix2Pix) được sử dụng để phát hiện vùng bất thường trên ảnh MRI não. Mô hình đề xuất được thực nghiệm và đánh giá trên tập dữ liệu LGG kết hợp với tập dữ liệu được các bác sĩ tại trường Đại học Y Dược Cần Thơ thu thập và gán nhãn. Để đánh giá chính xác hiệu quả của các mô hình, các lát cắt sử dụng để huấn luyện cũng như đánh giá mô hình được phân chia theo bệnh nhân. Mô hình đề xuất đạt hiệu quả trung bình như sau: Accuracy là 84,5 % (phân chia dữ liệu huấn luyện và kiểm tra theo ảnh) và 71,2 % (phân chia dữ liệu huấn luyện và kiểm tra theo bệnh nhân), F1 là 70,1 % (phân chia dữ liệu huấn luyện và kiểm tra theo ảnh), F1 là 58,6 %, (phân chia dữ liệu huấn luyện và kiểm tra theo bệnh nhân).*

Từ khóa: *Mô hình học sâu, phát hiện bất thường, ảnh MRI não.*

I. GIỚI THIỆU

Hình ảnh chụp cắt lớp sọ não MRI là tập hợp nhiều ảnh chụp cắt ngang hoặc cắt dọc sọ não để phản ánh cấu trúc sọ và mô não bên trong [1]. Sọ não là xương, các mô não gồm chất xám, nước,… mỗi thành phần này có khả năng hấp thu năng lượng tia X đi qua khác nhau từ đó tạo nên ảnh với thể hiện và màu khác nhau. Các bác sĩ chẩn đoán hình ảnh dựa vào cơ chế tạo ảnh này để khảo sát với mục tiêu là tìm ra vùng bất thường trên MRI. Mỗi bệnh nhân khi chụp MRI được kỹ thuật viên chụp nhiều lần, mỗi lần từng với các điều chỉnh cường độ tín hiệu khác nhau tạo nên nhiều bộ MRI khác nhau.

Nhiều phương pháp học sâu đang được nghiên cứu và áp dụng để hỗ trợ các bác sĩ xác định tự động và nhanh chóng vùng bất thường trên MRI não [2] [3] [4] [5] [6] [7]. Mô hình mạng DNN đã được áp dụng để phát hiện vùng bất thường trên MRI não trong nghiên cứu của Mohammad Havaei và cộng sự [4]. Các tác giả đã đề xuất kiến trúc "TwoPathCNN" cho phép huấn luyện 2 giai đoạn nhằm khám phá được đặc trưng cục bộ (LocalPathCNN) cũng như đặc trưng toàn cục (GlobalPathCNN) để phát hiện được vùng bất thường trên ảnh. Mô hình đề xuất thực nghiệm được đánh giá trên tập dữ liệu BRATS năm 2013 của 30 bệnh nhân cho tập huấn luyện; 10 bệnh nhân cho tập kiểm tra và 25 bệnh nhân cho tập dữ liệu bảng xếp hạng. Kết quả đánh giá chỉ số Dice, Specificity và Sensitivity của mô hình TwoPathCNN đạt lần lượt là 0,85, 0,93 và 0,80 cho vùng khối u hoàn toàn; tốc độ của mô hình đề xuất cũng nhanh hơn 30 lần so với các nghiên cứu mà tác giả đã thực hiện so sánh, đánh giá.

Một cách tiếp cận khác để phát hiện khối u là sử dụng các mô hình sinh GAN. Mô hình SegAN, một biến thể của GAN được áp dụng để phát hiện vùng bất thường trên ảnh MRI não [6]. Mô hình "Generator-Discriminator" của mạng GAN được thay thế bởi "Segmentor-Critic". Mô hình đề xuất được thực nghiệm và đánh giá trên tập dữ liệu BRATS 2015. Đây là tập dữ liệu ảnh bộ não 3D có kích thước 240x240x155 pixel. Kết quả đạt được với chỉ số Precision đạt tỷ lệ 92 %, chỉ số Sensitivity đạt tỷ lệ 80 % và chỉ số Dice (F1) đạt tỷ lệ 85 % khi dự đoán trên toàn bộ khối u (Whole tumor). Mô hình U-Net cũng được sử dụng để dự đoán vùng bất thường trên ảnh MRI não [5]. Mô hình được huấn luyện trên tập dữ liệu LGG[1] (110 bệnh nhân với 3929 MRI não) và đánh giá trên tập dữ liệu DICOM được thu thập tại bệnh viện trường Đại học Y Dược Cần Thơ (10 bệnh nhân với 212 ảnh MRI não). Các bác sĩ tại bệnh viện Trường Đại học Y Dược Cần Thơ dựa vào kết quả dự đoán của mô hình, kiểm tra và đánh giá kết quả của 212 ảnh của 10 bệnh nhân đạt độ chính xác 85,5 %.

Mô hình Pix2Pix được đề xuất vào năm 2018 bởi nhóm nghiên cứu trí tuệ nhân tạo (AI) của Viện Đại học California tại Berkeley [8]. Mục tiêu của mô hình là tạo ra các hình ảnh có cùng kích thước với ảnh đầu vào, nhưng có sự biến đổi các thuộc tính bên trong ảnh để phục vụ cho bài toán cần giải quyết. Mô hình Pix2Pix thuộc nhóm mô hình mạng sinh đối nghịch có điều kiện (cGAN) [9], được sử dụng để giải quyết các bài toán dịch ảnh trong nhiều lĩnh vực như: ảnh trắng đen thành ảnh màu, ảnh vệ tinh thành ảnh bản đồ, ảnh ban ngày thành ban đêm,… Đối với tập dữ liệu Cityspaces[2] gồm các ảnh đường phố cùng với nhãn phân đoạn tương ứng với 30 lớp của nhiều nhóm đối tượng khác nhau, trong đó nhóm đối tượng có kích thước lớn bao gồm công trình, cây cối, bầu trời, lề đường, phương tiện và các đối tượng có kích thước nhỏ gồm có con người, biển báo. Tập dữ liệu gồm 25000 hình ảnh đường phố được trích từ video quay đường phố của 50 thành phố khác nhau, trong đó có 5000 hình ảnh được gán nhãn chất lượng cao và 20000

[1] https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation

[2] https://www.cityscapes-dataset.com/dataset-overview/#features

ảnh gán nhãn với chất lượng thấp hơn. Kết quả đạt được với độ chính xác trung bình là 36 % khi đánh giá độ chính xác trên từng lớp, và đạt tỷ lệ 29 % với chỉ số IoU.

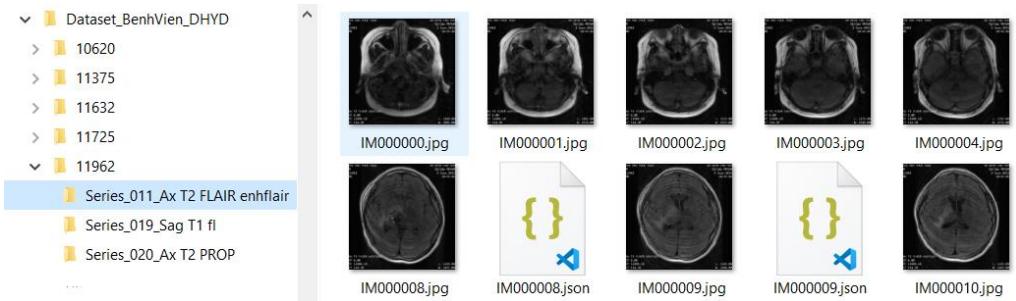
Dựa vào những kết quả thực nghiệm khả quan của mô hình Pix2Pix trên các lĩnh vực khác nhau, trong nghiên cứu này mô hình Pix2Pix được áp dụng để phát hiện các vùng bất thường trên MRI não. Tập dữ liệu được thu thập và gán nhãn trước khi dự đoán bởi các bác sĩ tại Bệnh viện Trường Đại học Y Dược Cần Thơ. Phần tiếp theo của bài viết được tổ chức như sau: mô hình phát hiện vùng bất thường trên MRI não được trình bày trong phần hai; kết quả thực nghiệm được giới thiệu ở phần ba; và cuối cùng là kết luận và hướng phát triển của nghiên cứu.

II. PHÁT HIỆN VÙNG BẤT THƯỜNG TRÊN MRI NÃO

A. Dữ liệu MRI não

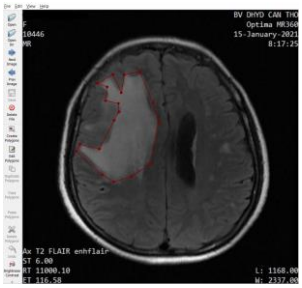
Cộng hưởng từ (MRI: magnetic resonance imaging) là kỹ thuật tạo hình cắt lớp sử dụng từ trường và sóng radio dựa trên nguyên tử Hydrogen(H) [1]. Bản chất của MRI là việc tạo ra một từ trường mạnh tác động lên các nguyên tử hydrogen trong cơ thể. Các momen từ hạt nhân trong nguyên tử sẽ di chuyển và sắp xếp theo chiều dọc hay chiều ngay theo xu hướng cân bằng từ trường bệnh trong và bên ngoài cơ thể. Sau khi đạt được trạng thái cân bằng, từ trường dùng để kích hoạt sẽ được tắt, các hạt nhân proton trong nguyên tử sẽ phóng thích năng lượng để về trạng thái ban đầu. Thời gian cần thiết cho 63 % vectơ từ hạt nhân khôi phục theo chiều dọc gọi là thời gian T1. Thời gian cần thiết để cho 63 % vectơ từ hạt nhân khôi phục theo chiều ngang gọi là thời gian T2. Dựa vào cường độ phát ra tính hiệu vô tuyến trên một đơn vị mô khi từ trường giải phóng năng lượng để xác định thang màu từ đen tới trắng (màu trắng là cường độ tín hiệu cao, màu đen là không có tín hiệu hoặc tính hiệu thấp). Cường độ tín hiệu của một loại mô phụ thuộc vào thời gian khôi phục lại từ tính T1 và T2, mật độ proton của nó. Bằng cách điều chỉnh thời gian chụp ảnh của T1 và T2, ta thu được các tương phản ảnh tương ứng với một đặc tính mô riêng biệt cũng như tạo ra các chuỗi xung khác nhau. Các chuỗi xung thường sử dụng là DWI (Diffusion-weighted Imaging), FLAIR (Fluid Attenuated Inversion Recovery), STIR (Short Time Inversion Recovery). Trong nghiên cứu này, chuỗi xung FLAIR được sử dụng vì hình ảnh thu được từ chuỗi xung FLAIR có thể phát hiện các tổn thương khác nhau như chảy máu, viêm não, xơ hóa mảng (MS) [1].

Tập dữ liệu MRI thu thập và gán nhãn bởi các bác sĩ tại bệnh viện Trường Đại học Y Dược Cần Thơ được sử dụng để đánh giá mô hình đề xuất. Mỗi bệnh nhân sẽ có một hoặc nhiều tập tin nhỏ lưu trữ theo mã số bệnh nhân có chuỗi xung khác nhau (T1, T2, T2FLAIR) và trong mỗi chuỗi xung sẽ chứa các MRI não cùng với mặt nạ tương ứng (ảnh gán vùng bất thường nếu ảnh đó có tổn thương) như Hình 1.



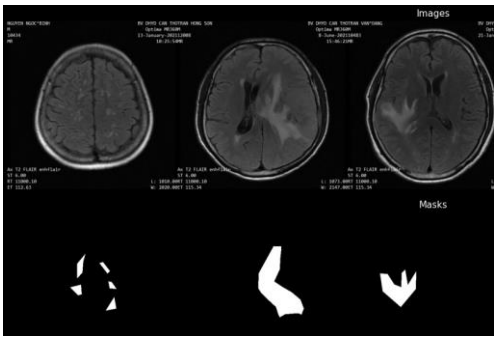
Hình 1. Cấu trúc lưu trữ tập dữ liệu MRI Brain DHYDCT

Các MRI não của 139 bệnh nhân được trích xuất từ file DICOM và chuyển sang định dạng JPG với kích cỡ ảnh là 512x512 trên thang màu xám. Các bác sĩ sử dụng công cụ labelme[3] để gán nhãn các vùng bất thường trên các lát cắt của ảnh MRI. Hình 2 là kết quả định vị vùng bất thường bởi bác sĩ trên 1 lát cắt, ví vùng bất thường sau khi được định vị sẽ lưu lại với định dạng JSON, bên trong file sẽ chứa thông tin nguồn của ảnh và vị trí các điểm của khối u. Thông tin vùng bất thường này được lưu thành 1 một tập tin ảnh gọi là mặt nạ chứa vùng bất thường. Tập dữ liệu thu thập được gồm 2812 MRI não trong đó có 604 ảnh có chứa vùng bất thường. Hình 3 thể hiện một số lát cắt và mặt nạ lưu vùng bất thường tương ứng được định vị bởi bác sĩ.



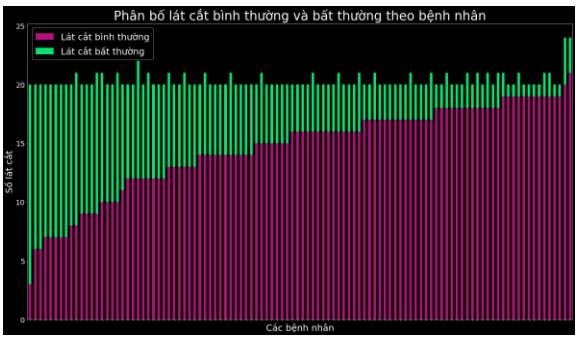
Hình 2. Vùng bất thường được gán nhãn bởi bác sĩ với công cụ labelme

[3] https://github.com/wkentaro/labelme



Hình 3. MRI não và mặt nạ chứa vùng bất thường tương ứng

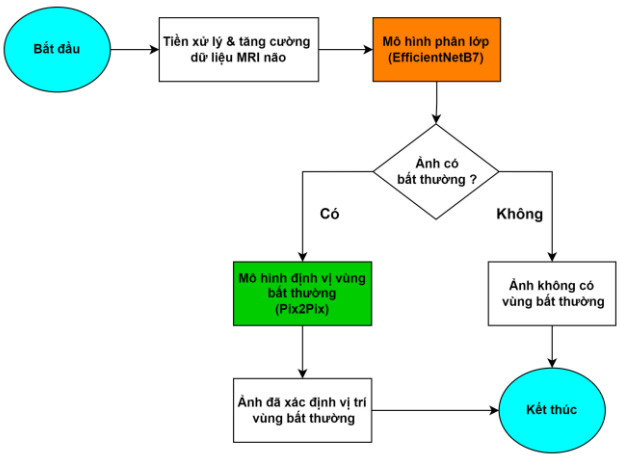
Nhằm tăng độ chính xác của quá trình huấn luyện và đánh giá mô hình, các ảnh MRI của bệnh nhân không có vùng bất thường sẽ được loại bỏ. Tập dữ liệu thực nghiệm gồm 2155 ảnh MRI não của 106 bệnh nhân, trong đó có 1551 ảnh bình thường và 604 ảnh bất thường được biểu diễn như Hình 4. Màu hồng tương ứng số lát cắt bình thường, màu xanh tương ứng với số lát cắt bất thường, tỷ lệ ảnh thu được của mỗi bệnh nhân dao động từ 20 đến 25 lát cắt, số lát cắt bất thường trung bình khoảng 5-7 lát cắt. Để tăng độ chính xác của mô hình huấn luyện, bên cạnh tập dữ liệu thu thập được, tập dữ liệu LGG chứa các MRI não cùng với mặt nạ phân đoạn bất thường FLAIR cũng được bổ sung vào quá trình huấn luyện. Tập dữ liệu hình ảnh của 110 bệnh nhân với 3929 ảnh trong đó có 1373 ảnh có chứa vùng khối u, 2437 ảnh không chứa vùng khối u và 119 ảnh trống (không chứa thông tin). Bên cạnh việc bổ sung tập dữ liệu LGG, các phương pháp tăng cường dữ liệu như xoay ảnh góc 90 độ, tạo ảnh đối xứng theo chiều ngang và dọc.



Hình 4. Phân bố lát cắt bình thường và bất thường theo bệnh nhân

B. Mô hình phát hiện vùng bất thường trên MRI não

Mô hình phát hiện phát hiện vùng bất thường trên ảnh MRI não đề xuất gồm 2 bước như Hình 5. Mô hình EfficientNetB7 [10] được sử dụng để phát hiện lát cắt có chứa vùng bất thường hay không. Nếu ảnh có chứa vùng bất thường thì sử dụng mô hình Pix2Pix để định vị vùng bất thường trên trong MRI não.



Hình 5. Sơ đồ tổng quát của bài toán phát hiện vùng bất thường

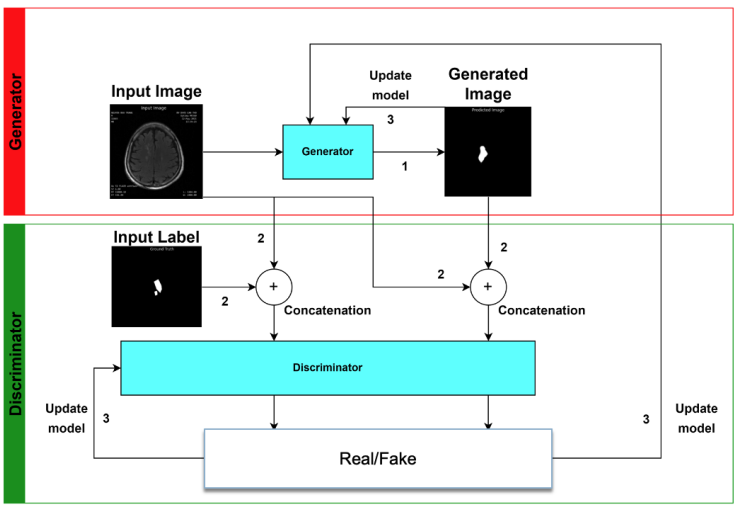
- **Mô hình phân lớp EfficientNetB7**

EfficientNet là một mô hình mạng nơron tích chập (Convolutional Neural Networks - CNN) được giới thiệu vào năm 2019 bởi các nhóm nghiên cứu của Google [10]. Với phương pháp "mở rộng kết hợp" (compound scaling), mô hình mạng EfficientNet đã cải thiện được độ chính xác và cải thiện được hiệu suất cũng như giảm kích cỡ của mô hình.

Khi thiết kế một mô hình mạng, cần phải thực nghiệm các tham số trên để tăng kích thước cho mô hình nhằm tăng độ chính xác, tuy nhiên quá trình thực nghiệm tốn rất nhiều thời gian và thường là tạo ra mô hình mạng chưa phải là tối ưu nhất. Vì vậy, tác giả đã tối ưu công việc này bằng cách xây dựng phương pháp "mở rộng kết hợp" nhằm tối ưu cả ba tham số độ rộng, độ sâu và độ phân giải sao cho mô hình đạt độ chính xác cao nhất và hiệu suất tốt nhất. Để đạt được độ chính xác cao nhất khi huấn luyện mô hình phân lớp, bài báo này sẽ sử dụng mô hình mạng EfficientNetB7. Các tham số được sử dụng cho mô hình phân lớp: hàm lỗi "Binary CrossEntropy" và "Epoch"=30.

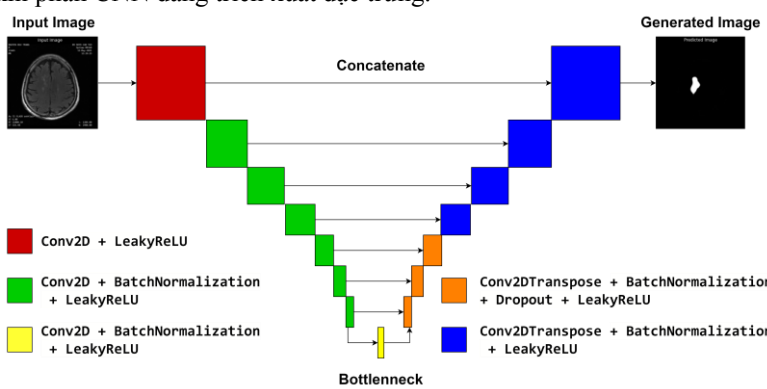
- **Mô hình Pix2Pix dùng để dự đoán vị trí bất thường**

Mô hình Pix2Pix dùng để dự đoán vị trí bất thường trên MRI não là mạng GAN có điều kiện (cGAN) gồm mô hình sinh (Generator) và mô hình phân biệt (Discriminator) được tổ chức như Hình 6. Đầu vào của mô hình sinh là ảnh MRI não và đầu ra của mô hình sinh là ảnh chứa vị trí và hình dạng bất thường tương ứng với ảnh trước đó. Đầu vào của mô hình phân biệt là một cặp ảnh: ảnh chứa vùng bất thường tạo bởi mô hình sinh và 1 ảnh chứa vùng bất thường gán nhãn bởi bác sĩ. Mô hình phân biệt cố gắng phân loại cặp ảnh đó, nếu kết quả là ảnh giống như gán nhãn bởi bác sĩ thì đó là thật.



Hình 6. Mô hình Pix2Pix phát hiện vùng bất thường trên MRI não

Mô hình sinh: Với mục đích tạo ra ảnh chứa vị trí và hình dạng của khối u, mô hình U-Net [11] được áp dụng. Mô hình có kiến trúc mô tả như Hình 7 gồm hai phần đối xứng nhau là mạng tích chập - mạng giải chập (CNN-DNN). Mạng CNN với mục tiêu trích xuất các đặc trưng từ hình ảnh ban đầu và mạng DNN với mục đích tạo ra ảnh chứa vị trí khối u từ các đặc trưng trước đó, ở giữa có thêm một khối Bottleneck để kết nối hai phần với nhau. Ngoài ra, để giữ các đặc trưng không bị mất đi sau khi trích lọc các nối kết tương ứng giữa các khối với nhau của hai phần giúp cho các đặc trưng được giữ lại khi phần CNN đang trích xuất đặc trưng.



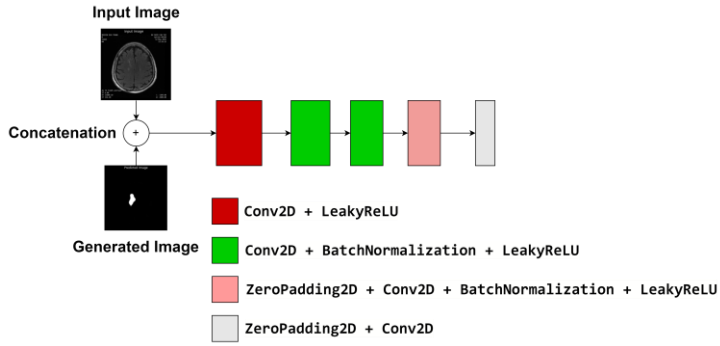
Hình 7. Kiến trúc U-Net bên trong mô hình sinh

Cấu trúc của phần CNN gồm 7 khối, trong mỗi khối có tầng chính là tầng tích chập hai chiều với nhiệm vụ trích lọc đặc trưng từ hình ảnh với kích cỡ của bộ lọc là 4×4 và độ dịch chuyển của bộ lọc là 2×2. Với mỗi lần thực nghiệm phép tích chập thì kích cỡ của ma trận sẽ giảm đi hai lần tương ứng. Ngoài ra, mạng CNN còn sử dụng thêm các tầng BatchNormalization để chuẩn hóa các giá trị trong ma trận về vùng phân bố là 0 nhằm giúp mô hình tránh hiện tượng học vẹt và sử dụng hàm kích hoạt LeakyReLU nhằm giảm giá trị âm. Cấu trúc của mạng DNN cũng gồm 7 khối, trong mỗi khối có tầng chính là tầng giải chập hai chiều với tính năng tạo ra hình ảnh từ các đặc trưng đã tìm được với bộ lọc là 4×4 và độ dịch chuyển bộ lọc 2×2. Ngoài ra trong DNN cũng sử dụng các loại tầng BatchNormalization, LeakyReLU và sử dụng thêm tầng Dropout để lọc bớt một phần trọng số bên trong mô hình, giúp mô hình tránh hiện tượng học vẹt.

Hàm mất mát của mô hình Generator trong mạng Pix2Pix sẽ giống với mô hình sinh trong mạng GAN có điều kiện (cGAN) [8]. Dựa theo thực nghiệm của tác giả Pix2Pix, hàm mất mát của cGAN học được các chi tiết khung ảnh nhiều hơn, tuy nhiên để mô hình học được các chi tiết nhỏ thì tác giả sử dụng thêm hàm khoảng cách. Trong thực nghiệm, khoảng cách Euclid (L2) cho kết quả tốt hơn khoảng cách Manhattan (L1) được hiển thị trong công thức (1):

$$\mathcal{L}_G^{pix2pix}(G,D) = \mathcal{L}_G^{cGAN}(G,D) + \lambda.\mathcal{L}_{L2}(G) = E_x\left[\log\left(1 - D\left(x, G(x)\right)\right)\right] + \lambda.E_{x,y}\left[\left|\left|y - G(x)\right|\right|_2\right] \quad (1)$$

Mô hình phân biệt: Mục tiêu của mô hình phân biệt là phải xác định được hai loại ảnh: nếu ảnh nhận được là mặt nạ chứa vùng bất thường được bác sĩ gán nhãn thì phân lớp là thật và ngược lại nếu là mặt nạ vùng bất thường được tạo từ mô hình sinh thì phân lớp là giả. Kiến trúc của mô hình phân biệt tổ chức như Hình 8, có phần đầu là mạng CNN để trích xuất đặc trưng từ hình ảnh, phần cuối của mô hình sẽ sử dụng lý thuyết PatchGAN được giới thiệu trong nghiên cứu [8] để phân lớp từng vùng nhỏ trong hình ảnh thay vì phân lớp cho cả hình ảnh.



Hình 8. Kiến trúc của mô hình phân biệt

Hàm mất mát của mô hình phân biệt được sử dụng là hàm Binary-Cross Entropy để tính xác suất phân lớp cho hai loại ảnh được đưa vào, giống với hàm mất mát trong mô hình cGAN được thể hiện trong công thức (2):

$$\mathcal{L}_D^{pix2pix}(G,D) = \mathcal{L}_D^{cGAN}(G,D) = E_{x,y}\left[\log D(x,y) + \log(1 - D\left(x, G(x)\right))\right] \quad (2)$$

Với cấu trúc của các mô hình sinh và mô hình phân biệt như đã trình bày, các thông số sử dụng để huấn luyện mô hình được hiển thị như Bảng 1.

Bảng 1. Bộ tham số sử dụng để phát hiện vùng bất thường với mô hình Pix2Pix

Bộ tham số	Giá trị
Step	20000
Batch_size	64
Generator_lr	$2 * 10^{-4}$
Discriminator_lr	$2 * 10^{-4}$
Loss Distance	Euclid (L2)
Lambda	200

III. KẾT QUẢ THỰC NGHIỆM

Thông thường, các nghiên cứu phân chia dữ liệu dựa trên ảnh để huấn luyện và đánh giá mô hình. Tuy nhiên, cách phân chia theo bệnh nhân sẽ giống với việc áp dụng trong thực tế hơn. Việc sử dụng một vài lát cắt của 1 bệnh nhân đem đi huấn luyện và một số lát cắt khác của cùng 1 bệnh nhân được dùng để đánh giá thì mô hình sẽ tốt hơn khi triển khai thực tế. Vì vậy, chúng tôi thử nghiệm với cả 2 phương pháp phân chia. Phân chia dữ liệu theo ảnh: tập huấn luyện gồm tất cả ảnh của tập dữ liệu LGG và thêm khoảng 1379 ảnh của tập dữ liệu Trường ĐHYDCT (chiếm 65 %) trong đó có trung bình 386 ảnh bất thường; tập "validation" có khoảng 438 ảnh (chiếm 15 %) với khoảng 98 ảnh bất thường; tập đánh giá với 552 ảnh (chiếm 20 %) với khoảng 121 ảnh bất thường. Phân chia theo bệnh nhân: tập huấn luyện tất cả ảnh của tập dữ liệu LGG và ảnh của 76 bệnh nhân thu thập từ tập dữ liệu của Trường ĐHYDCT (số ảnh dao động từ 1443-1545 chứa trung bình 423 ảnh bất thường), tập "validation": ảnh MRI của 10 bệnh nhân từ tập dữ liệu của ĐHYDCT (số ảnh dao động từ 202-206 chứa trung bình 62 ảnh bất thường), tập đánh giá: ảnh của 20 bệnh nhân (số ảnh dao động từ 406-410 chứa trung bình 119 ảnh bất thường).

Để đánh giá hiệu của các mô hình đề xuất, chỉ số "Accuracy", "Sensitivity", "Specificity", "F1" được sử dụng. Trong đó, TP thể hiện số dự đoán đúng vị trí cho ảnh bất thường, TN thể hiện số dự đoán đúng cho ảnh bình thường, FP thể hiện số dự đoán sai cho ảnh bất thường, FN thể hiện số dự đoán sai cho ảnh bình thường.

(1) Accuracy: Thể hiện khả năng dự đoán của mô hình, được tính bằng tỷ lệ giữa số ảnh dự đoán đúng so với tổng số ảnh bất thường và ảnh bình thường.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(2) Sensitivity (Recall): Thể hiện khả năng tìm các ảnh có bất thường, được tính bằng tỷ lệ giữa số ảnh dự đoán đúng có bất thường với tổng số ảnh có bất thường ban đầu.

$$Sensitivity, Recall = \frac{TP}{TP + FN}$$

(3) Specificity: Thể hiện khả năng tìm các ảnh bình thường, được tính bằng tỷ lệ giữa số ảnh dự đoán đúng bình thường với tổng số ảnh bình thường ban đầu.

$$Specificity = \frac{TN}{TN + FP}$$

(4) F1: Thể hiện khả năng trung hòa giữa việc dự đoán và tìm điểm bất thường của mô hình, được tính bằng tỷ lệ hai lần tích với tổng của hai chỉ số Precison và Recall.

$$F1 = 2 * \frac{Precison * Recall}{Precison + Recall}$$

Ngoài ra để đánh giá độ chính xác đối với đầu ra là hình ảnh, cần các chỉ số đánh giá trên từng pixel để đánh giá được độ bao phủ cũng như độ trùng lắp giữa kết quả dự đoán của mô hình và kết quả của bác sĩ, bao gồm:

(1) mPrecision: Thể hiện khả năng bao phủ trung bình vùng bất thường khi dự đoán vị trí bất thường bên trong ảnh, được tính bằng số pixel dự đoán đúng bất thường chia cho tổng số pixel bất thường mà mô hình dự đoán bên trong ảnh.

$$mPrecison = \sum_{1}^{n} \frac{1}{n} * \frac{Số\ pixel\ dự\ đoán\ đúng}{Số\ pixel\ bất\ thường\ trong\ ảnh\ dự\ đoán}$$

(2) mIoU: Thể hiện độ trùng lắp giữa vùng bất thường mà mô hình dự đoán so với vùng bất thường được đánh giá bởi các bác sĩ, được tính bằng số pixel dự đoán đúng chia cho phần gộp hai tập hợp pixel.

$$mIoU = \sum_{1}^{n} \frac{1}{n} * \frac{Số\ pixel\ dự\ đoán\ đúng}{Số\ pixel\ bất\ thường\ trong\ ảnh\ dự\ đoán \cup Số\ pixel\ bất\ thường\ trong\ ảnh\ thật}$$

A. Đánh giá mô hình phân lớp ảnh bất thường hay bình thường trên ảnh MRI não

Mô hình đề xuất định vị vùng bất thường gồm hai giai đoạn, giai đoạn xác định ảnh có bất thường được thử nghiệm và đánh giá trên 3 mô hình ResNet50, InceptionV3 và EfficientNetB7. Kết quả dự đoán ảnh có bất thường hay không có bất thường được hiển thị như Bảng 2 khi áp dụng phương pháp phân chia tập dữ liệu theo ảnh. Mô hình EfficientNetB7 cho kết quả dự đoán tốt nhất với chỉ số Sensitivity đạt tỷ lệ 80,2 % và tỷ lệ Specificity đạt 84,8 %.

Bảng 2. Kết quả phân lớp bất thường với của các mô hình CNN với giá trị random_state=1

Model	Số lát cắt bất thường	Số lát cắt bình thường	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	F1
ResNet50	121	310	40	49	261	81	69,8 %	33,1 %	84,2 %	38,1 %
InceptionV3	121	310	58	67	243	63	69,8 %	47,9 %	78,4 %	47,2 %
EfficientNetB7	121	310	97	47	263	24	83,5 %	80,2 %	84,8 %	73,2 %

B. Đánh giá hiệu quả của mô hình dự đoán vùng bất thường trên MRI não

Để có được kết quả tổng quan, các mô hình được huấn luyện 5 lần, mỗi lần huấn luyện sẽ sử dụng MRI não của các bệnh nhân khác nhau hay các ảnh khác nhau tùy cách phân chia dữ liệu. Với cách phân chia theo hình ảnh (Bảng 3), kết quả cho thấy độ chính xác tổng thể tốt hơn so với phân chia theo bệnh nhân (0). Tuy nhiên, thực tế áp dụng thì việc huấn luyện và đánh giá theo bệnh nhân là cần thiết và đánh giá đúng hiệu quả của mô hình khi triển khai thực tiễn.

Bảng 3. Kết quả dự đoán vùng bất thường trên MRI não với phương pháp phân chia theo ảnh bằng mô hình đề xuất

Random State	Số lát cắt bất thường	Số lát cắt bình thường	TP (IoU >0)	FP	TN	FN	Class				Pixel	
							Accuracy	Sensitivity	Specificity	F1	mPrecision	mIoU
1	121	310	84	26	284	37	85,4 %	69,4 %	91,6 %	72,7 %	60,4 %	53,4 %
10	120	311	79	24	287	41	84,9 %	65,8 %	92,3 %	70,9 %	59,6 %	49,5 %
20	107	324	67	27	297	40	84,4 %	62,6 %	91,7 %	66,7 %	50,9 %	44,2 %
50	129	302	83	26	276	46	83,3 %	64,3 %	91,4 %	69,7 %	54,7 %	46,6 %
100	126	305	82	24	281	44	84,2 %	65,1 %	92,1 %	70,7 %	52,0 %	46,0 %
Trung bình	120,6	310,4	79	25,4	285	41,6	84,5 %	65,5 %	91,8 %	70,1 %	55,5 %	47,9 %

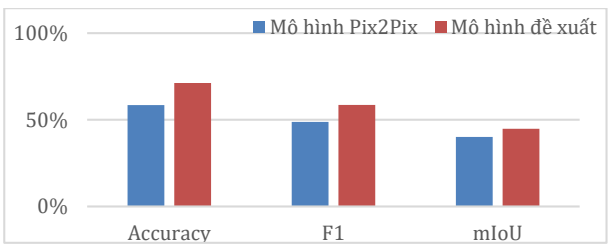
Bảng 4. Kết quả dự đoán vùng bất thường trên MRI não với phương pháp phân chia theo bệnh nhân bằng mô hình đề xuất

STT	Số lát cắt bất thường	Số lát cắt bình thường	TP (IoU >0)	FP	TN	FN	Class				Pixel	
							Accuracy	Sensitivity	Specificity	F1	mPrecision	mIoU
Lần 1	110	296	80	129	167	30	60,8%	72,7%	56,4%	50,2%	61,9%	48,6%
Lần 2	104	306	60	47	259	44	77,8%	57,7%	84,6%	56,9%	57,1%	45,6%
Lần 3	117	290	80	146	144	37	55,0%	68,4%	49,7%	46,6%	50,0%	41,3%
Lần 4	129	279	91	33	246	38	82,6%	70,5%	88,2%	71,9%	52,2%	45,2%
Lần 5	136	272	86	34	238	50	79,4%	63,2%	87,5%	67,2%	53,4%	43,5%
Trung bình	119,2	288,6	79,4	77,8	210,8	39,8	71,2%	66,5%	73,3%	58,6%	54,9%	44,8%

Bên cạnh việc đánh giá hiệu quả của mô hình đề xuất, chúng tôi cũng tổ chức so sánh kết quả đạt được với việc chỉ sử dụng mô hình Pix2Pix, bỏ qua bước phân lớp ảnh có bất thường hay không. Kết quả 0, Bảng 5 và Hình 9 cho thấy mô hình phân cấp đề xuất thật sự hiệu quả hơn khi xác định ảnh không có bất thường và mô hình đề xuất bao phủ vùng bất thường bên trong MRI não tốt hơn so với mô hình Pix2Pix thuần túy. Độ chính xác trung bình của mô hình đề xuất với Accuracy = 71,2 %, Specificity=73,3 %, F1=58,6 %, mPrecision = 54,9 % và hoàn toàn tốt hơn so với mô hình định vị vùng bất thường thuần túy: Accuracy = 58,5 %, Specificity=54,7 %, F1=48,8 %, mPrecision = 45,8 %.

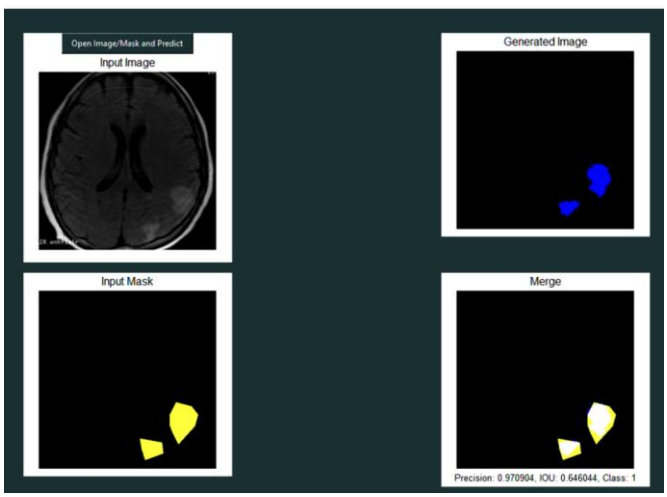
Bảng 5. Kết quả dự đoán vùng bất thường trên MRI não với phương pháp phân chia theo bệnh nhận trên mô hình Pix2Pix khi không sử dụng mô hình phân lớp

STT	Số lát cắt bất thường	Số lát cắt bình thường	TP (IoU >0)	FP	TN	FN	Class				Pixel	
							Accuracy	Sensitivity	Specificity	F1	mPrecision	mIoU
Lần 1	110	296	79	129	167	31	60,6 %	71,8 %	56,4 %	49,7 %	54,9 %	47,3 %
Lần 2	104	306	65	158	148	39	52,0 %	62,5 %	48,4 %	39,8 %	44,2 %	38,5 %
Lần 3	117	290	72	122	168	45	59,0 %	61,5 %	57,9 %	46,3 %	46,0 %	41,5 %
Lần 4	129	279	98	128	151	31	61,0 %	76,0 %	54,1 %	55,2 %	44,7 %	37,7 %
Lần 5	136	272	91	118	154	45	60,0 %	66,9 %	56,6 %	52,8 %	39,2 %	35,7 %
Trung bình	119,2	288,6	81	131	157,6	38,2	58,5 %	67,7 %	54,7 %	48,8 %	45,8 %	40,1 %



Hình 9. So sánh mô hình đề xuất và mô hình Pix2Pix thuần tuý

Mô hình dự đoán vùng bất thường được xây dựng như hình 10, từ ảnh MRI cung cấp, mô hình có thể phát hiện các vùng bất thường như ảnh "Generated image". Để so sánh đánh giá, mặt nạ vùng bất thường đã khoanh vùng bởi bác sĩ cũng được hiển thị trong ảnh "input image" và kết quả so sánh ảnh dự đoán được vùng bất thường và vùng bất thường thực tế được hiên thị trong ảnh "Merge" có kèm thông tin "precision" và "IOU".



Hình 10. Kết quả dự đoán vùng bất thường bởi mô hình đề xuất

IV. KẾT LUẬN

Việc phát hiện kịp thời khối u hỗ trợ các bác sĩ trong quá trình chẩn đoán và điều trị cho bệnh nhân được thực hiện hiệu quả trong tình trạng các bệnh viện luôn quá tải là rất cần thiết. Trong nghiên cứu này, mô hình phân lớp (EfficientNetB7) kết hợp mô mình phân vùng bất thường (Pix2Pix) được thực nghiệm và đánh giá. Mô hình đề xuất được huấn luyện trên tập dữ liệu LGG của 110 bệnh nhân với 3929 ảnh MRI (2437 ảnh không chưa vùng bất thường, 1373 ảnh có chứa vùng bất thường) kết hợp với tập dữ liệu đã được thu thập và gán nhãn của 106 bệnh nhân (1551 ảnh bình thường và 604 ảnh bất thường) tại Bệnh viện Trường Đại học Y Dược Cần Thơ. Kết quả cho thấy việc kết hợp mô hình phân loại trước khi xác định vùng bất thường mang lại hiệu quả cao hơn sử dụng trực tiếp mô hình định vị vùng bất thường. Mô hình Pix2Pix kết hợp mô hình phân lớp EfficientNetB7 đạt hiệu quả cao nhất khi phân chia dữ liệu huấn luyện và kiểm tra theo ảnh: Accuracy=82,6 %, Sensitivity =70,5 %, Specificity=88,2 %, F1=71,9 %, mPrecision = 52,2 %, mIoU = 45,2 %. Tập dữ liệu thu thập thực tế tại bệnh viện cũng cho thấy tỉ lệ lát cắt chứa vùng bất thường thấp hơn so với tập dữ liệu LGG. Phương pháp đánh giá theo bệnh nhân hay theo ảnh cũng làm thay đổi kết quả đánh giá hiệu quả của giải thuật. Kết quả của mô hình đề xuất cũng cho thấy vị trí vùng bất thường được phát hiện tốt, tuy nhiên tính theo pixel trùng khớp giữa vùng bất thường dự đoán bởi mô hình và vùng bất thường gán nhãn bởi bác sĩ chưa cao. Kết quả gán nhãn cũng phụ thuộc rất nhiều vào kinh nghiệm của bác sĩ.

TÀI LIỆU THAM KHẢO

[1] Trần Đức Quang, "Nguyên lý và kỹ thuật cộng hưởng từ", Đại học Quốc gia Thành phố Hồ Chí Minh, 2008, pp. 71-73.

[2] Holger R. Roth, Chen Shen, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, Kensaku Mori, "Deep learning and its application to medical image segmentation", *Med Imaging Technol*, pp. 63-71, 2018.

[3] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, Clara I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis", *Med Image Anal*, pp. 60-88, 2017.

[4] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle, "Brain Tumor Segmentation with Deep Neural Networks", *Medical Image Analusis*, pp. 18-31, 2017.

[5] Lê Minh Lợi, Trần Nguyễn Minh Thư, Hồ Trọng Nguyễn, Phạm Nguyên Khang, "Ứng dụng mô hình U-Net phát hiện vùng bất thường trên ảnh MRI não", *DOI: 10.15625/vap.2020.00225*, 2020.

[6] Yuan Xue, Tao Xu, Han Zhang, Rodney Long, Xiaolei Huang, "SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation", *CoRR*, 2017.

[7] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger Roth, Daguang Xu, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images", *ARXIV.2201.01266*, 2022.

[8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks", *CoRR*, 2016.

[9] Mehdi Mirza, Simon Osindero, "Conditional Generative Adversarial Nets", *CoRR*, 2014.

[10] Mingxing Tan, Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", *CoRR*, 2019.

[11] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *CoRR*, 2015.

APPLYING CGAN MODEL FOR DETECTING ABNORMAL AREAS ON BRAIN MRI IMAGES

Tran Nguyen Minh Thu, Bui Thanh Liem, Le Minh Loi, Phu Tri Nghia, Pham Nguyen Khang

ABSTRACT: *Currently, the computer vision to detect and diagnose abnormalities of brain is being widely applied. In this paper, the classification model (EfficienNetB7) combined with the conditional GAN model (Pix2Pix) is used to detect abnormal areas on brain MRI images. The proposed model is evaluated on the LGG dataset combined with the data set collected and labeled by doctors at Can Tho University of Medicine and Pharmacy. To accurately evaluate the effectiveness of the models, the slices of MRI images used for training as well as evaluating the model were divided by patient. The proposed model has an average effectiveness as accuracy =84.5 (the data is divided into train and test by slices), accuracy =71.2 (the data is divided into train and test by patiens, F1= 58.61 % (the data is divided into train and test by slices).*