

Phân đoạn khối u não: Kiến trúc nào hiệu quả nhất?

Hoang-Dan Pham¹, Van-Hai Ha¹, Phuc-Dat Nguyen¹

¹Khoa Công nghệ thông tin, Đại học HUTECH, TP.HCM, Việt Nam
peterdan1905@gmail.com, vanhai11203@gmail.com, phucdat231003@gmail.com

Tóm tắt

Trong những năm qua, việc ứng dụng Trí tuệ nhân tạo (AI - Artificial Intelligence) vào đời sống thực tế đã trở nên phổ biến hơn bao giờ hết, đặc biệt là trong lĩnh vực y tế. Rất nhiều các công trình nghiên cứu đã được tổ chức nhằm khai thác tiềm năng của AI cả về chiều ngang (mở rộng các lĩnh vực ứng dụng) lẫn chiều sâu (tăng cường sự hiệu quả trong các lĩnh vực chuyên môn). Trong đó, lĩnh vực nhận được nhiều sự quan tâm cũng như đạt được nhiều bước tiến quan trọng đó là Thị giác máy tính (CV - Computer Vision). Sự phát triển này giúp giải quyết các bài toán khó trong công việc cần xử lý ảnh cần tính chuyên môn cao như xác định khối u trong não thông qua ảnh chụp cộng hưởng từ (MRI) trở nên thuận tiện hơn khi giờ đây các bác sĩ đã có các mô hình phân đoạn (Segmentation) hỗ trợ trong việc tìm vị trí và hình dạng. Và trong bài báo này, chúng tôi sẽ tiến hành so sánh hiệu quả giữa các kiến trúc như U-net, nnU-Net và Swin UNETR trong việc phân đoạn khối u não thông qua ảnh 3D MRI. Kết quả của các kiến trúc cho thấy sự vượt trội của AI trong việc hỗ trợ chẩn đoán khối u não và hứa hẹn về sự phát triển của AI trong tương lai.

Giới thiệu

Vai trò của U-Net trong xử lý ảnh y tế và phân đoạn khối u não

Trong thập kỷ vừa qua, Học sâu (Deep Learning) đã tạo ra một cuộc cách mạng quy mô lớn trong lĩnh vực xử lý ảnh y tế, đặc biệt là trong các nhiệm vụ có độ khó cao và đòi hỏi chuyên môn như phân đoạn khối u não (Liu et al. 2023). Các yêu cầu về tính chính xác trong việc xác định hình dạng, vị trí và ranh giới của khối u là vô cùng nghiêm ngặt, bởi đây là yếu tố then chốt để phác thảo phác đồ điều trị hiệu quả (Gou et al. 2024). Do đó, việc phát triển một mô hình mạnh mẽ, có khả năng dự đoán chính xác đã trở thành hướng nghiên cứu trọng tâm. Trong số các mô hình đã được công bố, U-Net nổi bật như một giải pháp ưu việt (Ronneberger, Fischer, and Brox 2015). Phương pháp dựa trên học sâu này không chỉ cho thấy kết quả thực nghiệm khả

quan mà còn thể hiện khả năng xử lý dữ liệu dạng khối (volumetric) vượt trội, đáp ứng tốt các thách thức thực tế trong y học.

Tại sao U-Net nổi bật Kiến trúc U-Net, với cấu trúc mã hóa-giải mã (encoder-decoder) và các kết nối bỏ qua (skip connections), được thiết kế đặc biệt phù hợp cho các bài toán xử lý ảnh y tế. Khác với các mạng nơ-ron tích chập truyền thống (CNN), U-Net có khả năng trích xuất đồng thời các đặc trưng cục bộ (ví dụ: ranh giới khối u) và bối cảnh toàn cục (ví dụ: vị trí khối u trong não) trong một lần xử lý. Điều này đặc biệt quan trọng đối với phân đoạn khối u não, vốn thường có kích thước, hình dạng và cường độ tín hiệu thay đổi trên các kiểu ảnh MRI như T1ce và T2-FLAIR. Hơn nữa, U-Net hoạt động hiệu quả ngay cả với dữ liệu có nhãn hạn chế – một đặc điểm phổ biến trong các tập dữ liệu y tế, nơi việc gán nhãn ảnh MRI 3D đòi hỏi thời gian và sự tham gia của các chuyên gia X-quang.

Ứng dụng thực tế U-Net mang lại nhiều ứng dụng thực tiễn trong y học, đặc biệt trong phân đoạn khối u não:

- **Lập kế hoạch điều trị:** Việc phân đoạn chính xác các vùng khối u (toàn bộ khối u - WT, lõi khối u - TC, vùng tăng cường - ET) hỗ trợ trực tiếp bác sĩ trong việc xây dựng phác đồ điều trị cá nhân hóa. Ví dụ, xác định chính xác vùng ET giúp tối ưu hóa liệu pháp xạ trị, giảm thiểu tổn thương đến mô lành.
- **Hỗ trợ phẫu thuật:** Trong phẫu thuật thần kinh, U-Net cung cấp bản đồ 3D về ranh giới khối u trước hoặc trong quá trình phẫu thuật, giảm nguy cơ bỏ sót khối u và cải thiện kết quả cho bệnh nhân.
- **Theo dõi bệnh lý:** U-Net cho phép phân đoạn khối u trên nhiều lần chụp ảnh theo thời gian, giúp theo dõi sự phát triển hoặc co lại của khối u, từ đó đánh giá hiệu quả điều trị hoặc phát hiện sớm tái phát.
- **Nghiên cứu và phát triển thuốc:** Trong các thử nghiệm lâm sàng, U-Net hỗ trợ phân tích định lượng phản ứng của khối u với các liệu pháp mới, đẩy nhanh quá trình xác nhận hiệu quả thuốc.

Các vấn đề được giải quyết bởi U-Net U-Net giải quyết nhiều thách thức thực tế trong phân đoạn khối u não:

- **Độ phức tạp của khối u:** Khối u não, như glioma, thường có hình dạng không đều, kết cấu không đồng nhất và tín hiệu chồng lấn với mô lành. U-Net vượt qua vấn đề này nhờ khả năng trích xuất đặc trưng đa tỷ lệ, cho kết quả tốt hơn so với phân đoạn thủ công hoặc các phương pháp truyền thống như ngưỡng hóa (thresholding).
- **Hạn chế về thời gian và chuyên môn:** Phân đoạn thủ công bởi các bác sĩ X-quang tốn thời gian và dễ xảy ra sai lệch giữa các quan sát viên. U-Net tự động hóa quy trình này, mang lại kết quả nhất quán trong vài giây, giảm gánh nặng cho nhân sự y tế và tăng khả năng mở rộng trong thực tế lâm sàng.
- **Xử lý dữ liệu 3D:** Các mô hình 2D truyền thống gặp khó khăn khi xử lý dữ liệu MRI 3D, mất đi bối cảnh không gian giữa các lát cắt. U-Net 3D xử lý toàn bộ khối dữ liệu, bảo toàn thông tin độ sâu cần thiết để định vị khối u chính xác.
- **Tính linh hoạt với đa kiểu ảnh:** U-Net thích nghi tốt với dữ liệu đa phương thức (như T1ce, T2-FLAIR), học được các đặc trưng bổ sung từ mỗi kiểu ảnh, giúp phân biệt các vùng phụ của khối u vốn hiển thị khác nhau trên từng phương thức.

Tầm ảnh hưởng rộng lớn hơn Ngoài phân đoạn khối u não, U-Net đã trở thành nền tảng trong xử lý ảnh y tế, được áp dụng cho các nhiệm vụ như phân đoạn gan, phát hiện mạch máu võng mạc, hay phân tích nốt phổi. Trong bối cảnh khối u não, việc áp dụng U-Net đã giảm thời gian chẩn đoán, nâng cao độ chính xác phẫu thuật và cải thiện tỷ lệ sống sót của bệnh nhân nhờ hỗ trợ can thiệp sớm và chính xác hơn. Với sự phát triển của tài nguyên tính toán và các tập dữ liệu như BraTS, U-Net tiếp tục được cải tiến với các biến thể như 3D U-Net hay Attention U-Net, mở rộng giới hạn về độ chính xác và hiệu quả.

Trong nghiên cứu này, chúng tôi tập trung so sánh U-Net với bốn cài đặt khác nhau: 4 lớp (layer) với 32 bộ lọc (filter), 4 lớp với 16 bộ lọc, 5 lớp với 32 bộ lọc, và 5 lớp với 16 bộ lọc. Mục tiêu của so sánh là xác định sự ảnh hưởng của các cài đặt kiến trúc đến hiệu quả phân đoạn, bao gồm độ chính xác, chi phí tính toán và khả năng ứng dụng trong phân đoạn khối u não 3D. Bài báo này nhằm trả lời các câu hỏi sau:

- Sự khác biệt trong kiến trúc của U-Net dưới các cài đặt khác nhau là gì?
- Những khác biệt này ảnh hưởng như thế nào đến độ chính xác, chi phí tính toán và khả năng ứng dụng của mô hình trong phân đoạn khối u não 3D?
- Cài đặt nào của U-Net có khả năng cân bằng tối ưu giữa hiệu suất và yêu cầu tài nguyên trong môi trường thực tế?

Các đóng góp của chúng tôi được trình bày như sau. Trước tiên, chúng tôi cung cấp một cái nhìn chi tiết về tác động của các cài đặt kiến trúc đến hiệu quả phân đoạn. Tiếp theo, chúng tôi đưa ra so sánh trực quan dựa trên các phép đo chuẩn, sử dụng tập dữ liệu công

khai BraTS2020 về khối u não. Cuối cùng, chúng tôi phân tích điểm mạnh và hạn chế của từng cài đặt, từ đó giúp người đọc có cái nhìn tổng quan về xu hướng phát triển và ứng dụng của U-Net trong tương lai. Phần còn lại của bài báo được chia thành các chương như sau: Chương II tổng hợp các nghiên cứu liên quan và các điểm khác biệt chính của các cài đặt; Chương III mô tả các thiết lập và phương pháp thực nghiệm; Chương IV trình bày, so sánh và phân tích các kết quả thực nghiệm; Chương V tổng kết và đưa ra định hướng nghiên cứu trong tương lai.

Nghiên cứu liên quan

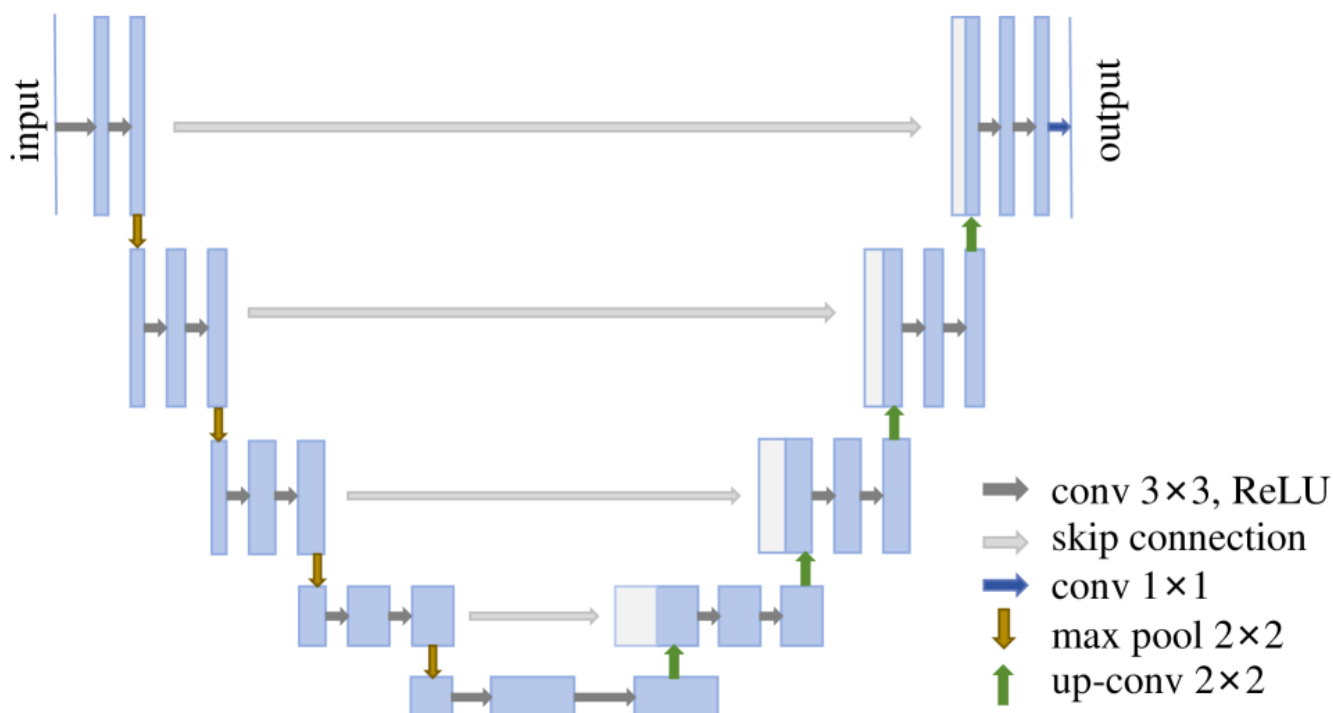
Từ những năm 2000, các nhà nghiên cứu đã luôn đi tìm giải pháp cho bài toán phân đoạn khối u não một cách tự động, các từ khóa tìm kiếm luôn xoay quanh "khối u", "phát hiện", "xử lý ảnh MRI", v.v... (Liu et al. 2023). Các nghiên cứu đã có bước đột phá lớn khi vào năm 2012, từ khóa "Học sâu" trở thành một xu hướng nghiên cứu mới (House 2019). Khắc phục được các nhược điểm vốn có, Học sâu trở nên mạnh mẽ hơn bao giờ hết và ngày càng được ứng dụng trong nhiều lĩnh vực. Với sự phát triển đó, vào năm 2012 cuộc thi "BraTS Challenge" đã ra đời, điều này thu hút các nhà nghiên cứu tham gia vào phát triển các mô hình mạnh mẽ. Qua nhiều năm, kỹ thuật phân đoạn đã phát triển một cách đáng kinh ngạc, chuyển dịch từ các kỹ thuật Học máy (Machine Learning) (Gou et al. 2024) thống kê truyền thống sang các phương pháp hiện đại hơn, với nền tảng là Học sâu.

Các kỹ thuật truyền thống

Ở giai đoạn đầu của nghiên cứu, hầu hết các phương pháp được sử dụng đều dựa vào kỹ thuật xử lý ảnh truyền thống như Thiết lập ngưỡng (thresholding) (Sahoo, Soltani, and Wong 1988), Vùng mở rộng (region growing) (Hojjat and Kittler 1998) và Phát hiện biên (edge detection) (Konishi et al. 2003). Những phương pháp này thường yêu cầu phải có các chuyên gia trong lĩnh vực y học, toán học và máy học bởi việc thí nghiệm cần phải điều chỉnh các thông số một cách thủ công, thiết kế đặc trưng (Feature-Based Design) (Shahin 2008) là một trong số đó. Những phương pháp này thường bị giới hạn do sự thay đổi trong dữ liệu (nhiều, biến thiên,...) khiến chúng gặp trở ngại vì các thông số cần phải được điều chỉnh chính xác để khớp với sự thay đổi ấy.

Sự trỗi dậy của Học máy

Học máy mở ra nhiều cơ hội nghiên cứu, với những phương pháp như Support Vector Machine (SVM) (Theodoros and Massimiliano 2001) và Rừng ngẫu nhiên (Random Forests) (Breiman 2001). Những phương pháp này có khả năng học từ dữ liệu được gán nhãn, giúp tăng khả năng thích nghi với dữ liệu, tuy nhiên vẫn còn yêu cầu việc thiết kế các đặc trưng một cách thủ công và tiền xử lý dữ liệu.



Hình 1: Kiến trúc cơ bản của U-Net bao gồm 2 bộ mã hóa và giải mã đối xứng cùng với các đường liên kết

Kỹ nguyên học sâu

Đột phá quan trọng nhất khi Mạng thần kinh tích chập (CNNs - Convolutional Neural Networks) (O'Shea and Nash 2015) ra đời và được ứng dụng cho công việc phân đoạn. Sức mạnh và khả năng của CNNs vào thời điểm đó được xem là tân tiến nhất (SOTA - State-of-the-art) và được chấp nhận rộng rãi trong lĩnh vực Thị giác máy tính nói riêng. Được lấy cảm hứng sâu sắc từ CNN, kiến trúc U-Net (Ronneberger, Fischer, and Brox 2015) là bước ngoặt khi đưa vào cấu trúc mã hoá - giải mã đối xứng, giúp mô hình nắm bắt được từ thông tin toàn cục cho đến thông tin cục bộ. U-Net và các biến thể của nó nhanh chóng trở thành tiêu chuẩn cho Phân đoạn ảnh y học dựa trên khả năng học các biểu diễn đặc trưng một cách mạnh mẽ từ một lượng dữ liệu giới hạn.

Các tiến bộ gần đây

Dựa trên sự thành công của U-Net, nhiều cải tiến và biến thể của kiến trúc này đã được đề xuất nhằm tối ưu hóa hiệu quả phân đoạn cho các ứng dụng khác nhau như nnU-Net (Isensee et al. 2018), Swin UNETR (Hatamizadeh et al. 2022),... Các biến thể tập trung vào việc điều chỉnh cấu trúc mạng, số lượng lớp và bộ lọc, cũng như cải tiến các bước tiền xử lý, nhằm đáp ứng tốt hơn các yêu cầu của nhiệm vụ phân đoạn, đặc biệt là trên dữ liệu ảnh MRI 3D. Mặc dù Đã có rất nhiều bài khảo sát được thực hiện (Bakas et al. 2019; Tirivangani and Serestina 2021; Kaifi 2023; Liu et al. 2023; Gou et al. 2024; Ragba, Boulaalam, and

Bellamine 2024) nhưng các phương pháp so sánh chưa cung cấp được cái nhìn sâu sắc vào cách các cấu hình khác nhau sẽ có ảnh hưởng như thế nào đến kết quả, cũng như hiệu suất của mô hình U-Net trên tập dữ liệu MRI 3D. Sự thiếu sót này cho thấy sự cần thiết của một bài thực nghiệm toàn diện, giúp đánh giá được sức mạnh, hạn chế và cách triển khai của những cấu hình khác nhau, mở đường cho các hướng nghiên cứu chuyên sâu hơn trong tương lai cho việc áp dụng lâm sàng.

Phương pháp

1. Xác định vấn đề

Phân đoạn khối u não đề ra một thử thách to lớn trong lĩnh vực chẩn đoán và lên phác đồ điều trị trong y học vì tính chất phức tạp và sống còn của nó. Việc xác định xác định và phân đoạn chính xác khối u não từ ảnh xung (modality) y học, như là Ảnh chụp Cộng hưởng từ (MRI - Magnetic Resonance Imaging) là vô cùng quan trọng cho giai đoạn chẩn đoán hiệu quả, lên kế hoạch phẫu thuật và giám sát hiệu quả điều trị. Tuy nhiên, việc phân đoạn khối u não một cách thủ công trên dữ liệu hình ảnh 3D là một công việc tốn nhiều thời gian, dễ xảy ra sai sót do thao tác của con người, và đặc biệt nhất là thiên kiến của chuyên viên phân tích (Liu and Xu 2024). Việc tồn tại một phương pháp tự động thường bị cản trở bởi một thực tế - các khối u có sự bất định/ bất đồng về hình dáng, kích thước và kết cấu. Cũng như các yếu tố trong quá trình thu thập dữ liệu như: nhiễu, lỗi thao tác,... Thúc đẩy bởi các động

lực nhằm vượt qua các khó khăn ấy, các nhà khoa học luôn tìm kiếm một phương pháp mạnh mẽ, hiệu quả và chính xác để tự động hoá phân đoạn khối u não trên dữ liệu 3D, cải thiện các quy trình y học, tăng cường độ chính xác của việc chẩn đoán và tạo ra kết quả điều trị tốt hơn.

2. Nghiên cứu về dữ liệu

Trong bài báo cáo này, chúng tôi sẽ sử dụng BraTS2020 - một bộ dữ liệu toàn diện các bản quét (scan) của nhiều tổ chức được chọn lọc cho nhiệm vụ phân đoạn khối u não (Menze et al. 2015). Mỗi mẫu trong tập dữ liệu là một hình ảnh lập thể (volumetric), được thu thập bằng cách dùng đa chuỗi xung (modalities) MRI bao gồm T1W, T1ce, T2W, FLAIR (Cung 2023). Những xung này cung cấp độ tương phản đa dạng, tạo điều kiện thuận lợi trong việc khoanh vùng chính xác khu vực u não (Pichaivel et al. 2022). Dữ liệu được cung cấp dưới dạng một thư mục lớn, trong đó có các thư mục con là một ca bệnh, với mỗi ca bệnh bao gồm 5 tệp tin định dạng Nifti (.nii.gz) - định dạng lưu trữ tiện lợi và phổ biến cho dữ liệu dạng khối MRI (Rizwan I Haque and Neubert 2020). Bốn tập tin với tên gọi lần lượt là: t1 (T1W), t2 (T2W), t1ce (T1ce), flair (FLAIR) cùng với tập tin seg (Segmentation) đại diện cho vùng u não đã được đánh nhãn bởi chuyên gia dùng để so sánh với kết quả dự đoán của mô hình để đánh giá hiệu quả của mô hình.

Về quy trình đánh nhãn, tất cả nhãn trong dữ liệu đều được đánh thủ công bởi một đến bốn chuyên gia (tuỳ thuộc vào cơ quan) với cùng một giao thức (protocol), và kết quả của họ sẽ được được duyệt bởi một chuyên gia thần kinh có kinh nghiệm. Nhãn sẽ được quy định như sau: Khối u tăng cường GD (ET — nhãn 4), Phù quanh khối u (ED — nhãn 2), Lõi khối u hoại tử và không tăng cường (NCR/NET — nhãn 1), Phần còn lại được chú thích là mô lành (Nền — nhãn 0). Dữ liệu được cung cấp sau khi đã tiền xử lý, như nội suy để đảm bảo cùng độ phân giải (1 mm^3), loại bỏ xương sọ,... Thông thường, tất cả các xung sẽ được sử dụng để có cái nhìn tổng quát về khu vực xuất hiện khối u bởi mỗi xung sẽ làm nổi bật một vùng khác nhau của não, giúp các bác sĩ phát hiện và phân tích khu vực xuất hiện bất thường. Tuy nhiên, vì chúng tôi đang làm việc với công cụ tự động lấy AI làm nền tảng, việc dùng cả bốn sẽ khiến công việc tính toán của máy tính trở nên khó khăn hơn. Để đạt được hiệu quả cao nhất, chúng tôi sẽ giảm số lượng xung sử dụng nhưng vẫn sẽ đảm bảo kết quả đầu ra. Với đặc tính tự nhiên của u não, dữ liệu được gán nhãn sẽ xuất hiện hiện tượng mất cân bằng giữa các nhãn do các khối u chỉ chiếm phần trăm rất nhỏ trong não (nhỏ hơn 5%) và có khả năng nằm phân tán, trong khi đó các tế bào não khoẻ mạnh thì lại chiếm phần lớn. Để vượt qua được vấn đề này, chúng tôi sẽ tiếp tục đào sâu hơn vào dữ liệu.

3. Tìm hiểu về các đặc trưng

Để có cái nhìn rõ ràng hơn về tập dữ liệu được sử dụng, chúng ta sẽ tiến hành kiểm tra hình dạng (shape) của hình ảnh quét và nhãn của nó. Trong tập dữ liệu này, tất cả các xung và nhãn đều có chung chiều không gian, đảm bảo sự nhất quán giữa các hình ảnh. Mỗi hình ảnh được tổ chức dưới dạng hình ảnh 3 chiều (3D) với chiều không gian được mô tả là

$$W \times H \times D \quad (1)$$

trong đó:

- W đại diện cho chiều rộng (số lượng điểm ảnh theo chiều ngang),
- H đại diện cho chiều cao (số lượng điểm ảnh theo chiều dọc),
- D đại diện cho chiều sâu (số lượng điểm ảnh theo chiều sâu).

Để dễ hình dung, một mẫu trong tập dữ liệu BraTS20 có hình dạng

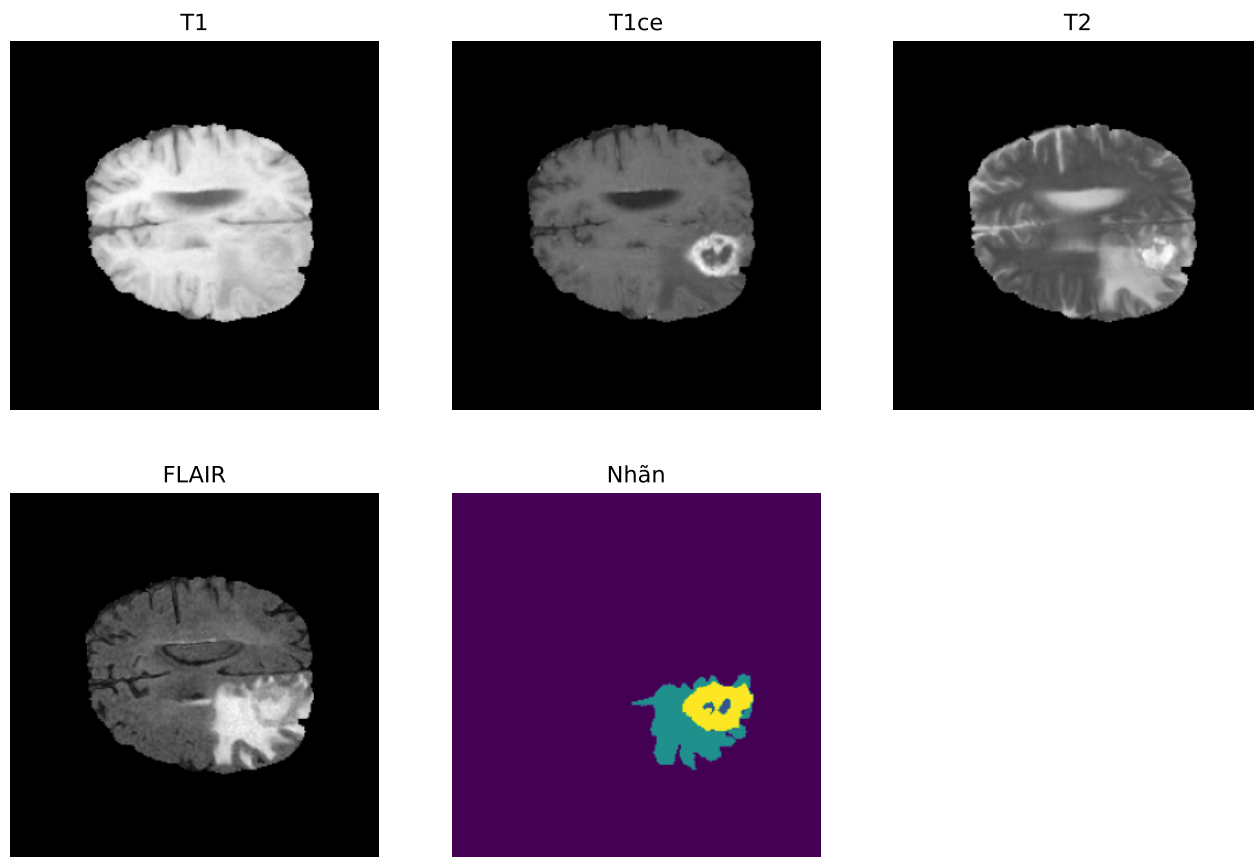
$$240 \times 240 \times 155$$

cho thấy rằng mỗi khối dữ liệu bao gồm 155 lát cắt theo chiều sâu, với mỗi lát cắt tương ứng với 240×240 điểm ảnh như trong Hình 2. Những hình ảnh 3D này được chụp theo cấu trúc tự nhiên của cơ thể người và được chia thành 3 mặt cắt chính thường sử dụng trong ảnh y tế: axial, coronal và sagittal như thể hiện trong Hình 3. Mỗi mặt cắt sẽ cung cấp một cách nhìn khác nhau về bộ não, giúp bác sĩ dễ dàng phân tích cấu trúc và phát hiện sự bất thường. Một đặc điểm nữa của dữ liệu này đó là sự xuất hiện của các hình ảnh đen ở hai đầu của mỗi khối (tính theo chiều sâu) như trong Hình 4 mô tả. Nguyên nhân của việc này là do trong quá trình quét, các lát cắt ở đầu và cuối chưa xuất hiện nào, dẫn đến hình ảnh không chứa bất cứ thông tin nào.

4. Thao tác với đặc trưng

Trong quá trình xử lý và chuẩn bị dữ liệu từ tập BraTS2020, chúng tôi đã đưa ra các quyết định quan trọng liên quan đến việc lựa chọn các đặc trưng hình ảnh và cách xử lý chúng để tối ưu hóa hiệu suất của mô hình học máy. Cụ thể, chúng tôi quyết định sử dụng hai loại hình ảnh MRI là T1ce (T1-weighted contrast-enhanced) và T2-FLAIR (T2-weighted Fluid-Attenuated Inversion Recovery) cho việc huấn luyện mô hình. Đồng thời, để đảm bảo chất lượng dữ liệu đầu vào, chúng tôi cũng tiến hành loại bỏ các lát cắt (slices) không chứa thông tin hữu ích, chẳng hạn như các lát cắt chỉ bao gồm nền hoặc không có vùng khối u. Những thao tác này không chỉ giúp cải thiện khả năng học của mô hình mà còn đảm bảo dữ liệu đầu vào phù hợp với kiến trúc mô hình, như sẽ được trình bày chi tiết hơn ở các phần sau. Dưới đây là lý do cụ thể cho các lựa chọn này:

- **Lựa chọn T1ce thay vì T1:** Chúng tôi quyết định không sử dụng hình ảnh T1 (T1-weighted) vì T1ce, phiên bản tăng cường độ tương phản của T1, cung cấp thông tin rõ ràng và chi tiết hơn về các vùng



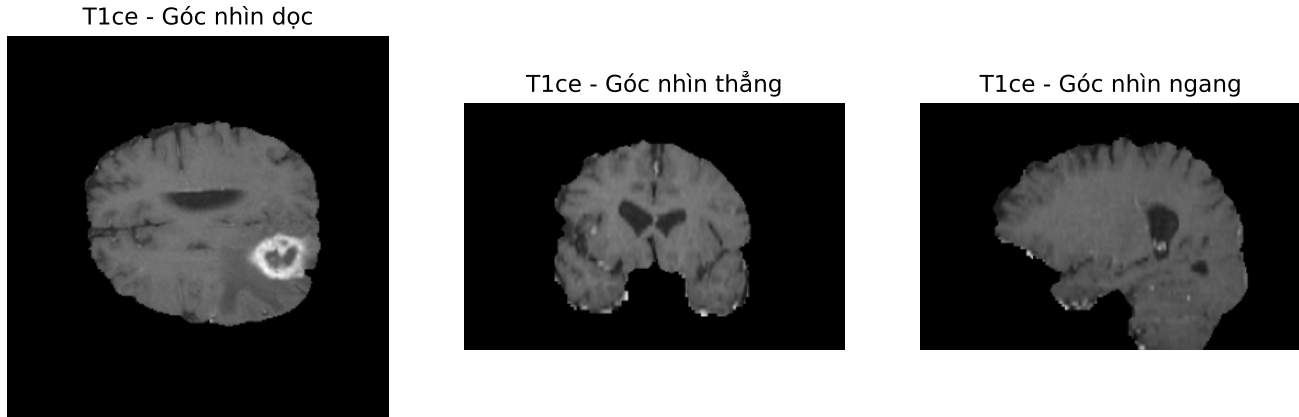
Hình 2: Lát cắt so sánh trực quan giữa các xung và nhân đã được đánh bởi chuyên gia

bất thường, đặc biệt là các khối u não. Trong T1ce, chất tương phản (thường là gadolinium) được tiêm vào bệnh nhân trước khi chụp MRI, giúp làm nổi bật các vùng có mạch máu bất thường, chẳng hạn như khối u, nhờ sự tăng cường tín hiệu ở những khu vực này. So với T1, T1ce cho phép mô hình dễ dàng xác định ranh giới và đặc điểm của khối u hơn, từ đó cải thiện độ chính xác trong việc phân đoạn và nhận diện. Việc loại bỏ T1 giúp giảm bớt sự dư thừa dữ liệu, tập trung vào các đặc trưng có giá trị cao hơn cho bài toán.

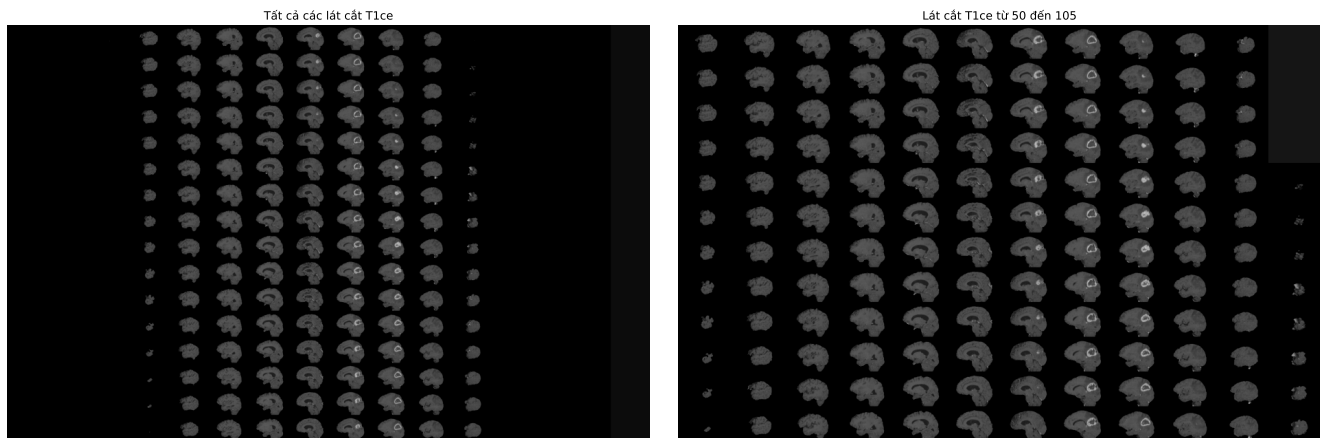
- **Lựa chọn T2-FLAIR thay vì T2:** Hình ảnh T2 (T2-weighted) không được sử dụng vì đặc tính của loại xung này nhấn mạnh vào các vùng chứa chất dịch trong não, chẳng hạn như dịch não tủy (cerebrospinal fluid - CSF) hoặc phù nề. Điều này có thể tạo ra nhiễu, làm lu mờ các đặc trưng quan trọng của khối u và ảnh hưởng tiêu cực đến độ chính xác của mô hình. Ngược lại, T2-FLAIR là một lựa chọn thay thế vượt trội vì nó sử dụng kỹ thuật đảo ngược tín hiệu (inversion recovery) để triệt tiêu tín hiệu từ chất dịch, giúp làm nổi bật các vùng bất thường như khối u hoặc phù nề liên quan đến khối u. Nhờ

khả năng loại bỏ các tín hiệu không cần thiết, T2-FLAIR cung cấp hình ảnh rõ ràng hơn về các khu vực bị ảnh hưởng, từ đó hỗ trợ mô hình trong việc xác định và phân đoạn khối u một cách chính xác hơn. Việc ưu tiên T2-FLAIR thay vì T2 không chỉ giảm thiểu nhiễu mà còn tăng cường khả năng tập trung của mô hình vào các đặc trưng liên quan trực tiếp đến bài toán.

Việc lựa chọn chỉ sử dụng T1ce và T2-FLAIR không chỉ giúp cân bằng giữa độ chính xác và hiệu quả tính toán mà còn giảm thiểu sự phức tạp trong quá trình xử lý dữ liệu. Bằng cách loại bỏ các loại hình ảnh ít giá trị hơn (T1 và T2) và các lát cắt không chứa thông tin, chúng tôi giảm được khối lượng dữ liệu cần xử lý, từ đó tăng tốc độ huấn luyện và kiểm thử mà vẫn duy trì được độ tin cậy của mô hình. Quyết định này cũng phù hợp với các nghiên cứu trước đây trong lĩnh vực phân đoạn khối u não, nơi T1ce và T2-FLAIR thường được ưu tiên do khả năng cung cấp thông tin bổ sung và tương phản cao. Để đảm bảo quá trình huấn luyện diễn ra ổn định và hiệu quả, chúng tôi áp dụng phương pháp chuẩn hóa Min-Max để đưa phân bố giá trị dữ liệu gốc về khoảng $[0, 1]$. Phương pháp này được lựa



Hình 3: Hình ảnh ba mặt phẳng (axial, sagittal, coronal) của khối MRI 3D T1ce, thể hiện cấu trúc giải phẫu từ dữ liệu đầu vào dùng cho các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F.



Hình 4: Hình ảnh so sánh Montage đầy đủ và Montage với lát cắt từ 50 đến 105 của khối MRI 3D T1ce, hiển thị toàn bộ cấu trúc giải phẫu từ dữ liệu đầu vào cho các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F.

chọn dựa trên đặc tính của dữ liệu hình ảnh MRI trong tập BraTS2020. Cụ thể, giá trị nhỏ nhất trong các hình ảnh MRI thường là 0 (tương ứng với các vùng không có tín hiệu), trong khi giá trị lớn nhất có thể thay đổi đáng kể giữa các bệnh nhân hoặc các khối u do sự khác biệt về thiết bị chụp, giao thức chụp, hoặc đặc điểm sinh học của khối u. Sự không đồng nhất này có thể gây ra bất ổn định trong quá trình huấn luyện, đặc biệt khi mô hình phải xử lý các giá trị đầu vào nằm trong các miền giá trị khác nhau. Chuẩn hóa Min-Max giải quyết vấn đề này bằng cách ánh xạ tuyến tính toàn bộ giá trị pixel về khoảng $[0, 1]$, giúp mô hình dễ dàng học các đặc trưng mà không bị ảnh hưởng bởi sự khác biệt về thang giá trị. So với các phương pháp chuẩn hóa khác, chẳng hạn như chuẩn hóa Z-score, Min-Max đặc biệt phù hợp với dữ liệu MRI vì nó bảo toàn tính chất không âm của dữ liệu và không yêu cầu giả định

về phân bố chuẩn của dữ liệu. Kết quả là, quá trình huấn luyện trở nên ổn định hơn, và mô hình có thể hội tụ nhanh hơn đến một giải pháp tối ưu.

5. Phân chia dữ liệu

Để tận dụng tối đa lượng dữ liệu lớn từ tập BraTS2020, bao gồm thông tin từ 369 bệnh nhân, chúng tôi tiến hành phân chia tập dữ liệu thành ba phần chính: tập huấn luyện (250 bệnh nhân), tập xác thực (74 bệnh nhân) và tập kiểm tra (45 bệnh nhân). Việc phân chia này không chỉ đảm bảo sự cân đối trong việc sử dụng dữ liệu mà còn đóng vai trò quan trọng trong việc xây dựng và đánh giá mô hình học máy một cách hiệu quả. Dưới đây là lý do và ý nghĩa chi tiết của việc phân chia tập dữ liệu thành ba phần này:

- **Tập huấn luyện (Training Set):** Đây là tập dữ liệu lớn nhất, chiếm 250 bệnh nhân, tương đương

khoảng 68% tổng số dữ liệu. Lý do tập huấn luyện cần được phân bổ nhiều nhất là vì nó đóng vai trò cốt lõi trong việc huấn luyện mô hình. Cụ thể, dữ liệu trong tập này được sử dụng để cập nhật các trọng số (weights) của mô hình thông qua quá trình học có giám sát. Quá trình này đòi hỏi một lượng lớn dữ liệu có gán nhãn (labeled data) để mô hình có thể học được các đặc trưng phức tạp, chẳng hạn như ranh giới của khối u não trong hình ảnh y khoa. Khi mô hình dự đoán sai, các sai lệch này sẽ được sử dụng để điều chỉnh và cải thiện hiệu suất qua từng vòng lặp (iteration). Nếu tập huấn luyện quá nhỏ, mô hình có thể không học đủ các mẫu đa dạng, dẫn đến hiệu suất kém khi áp dụng thực tế.

- **Tập xác thực (Validation Set):** Tập xác thực bao gồm 74 bệnh nhân, chiếm khoảng 20% dữ liệu, được thiết kế để theo dõi và đánh giá hiệu quả của quá trình huấn luyện mà không trực tiếp tham gia vào việc cập nhật trọng số. Vai trò của tập này đặc biệt quan trọng trong các bài toán yêu cầu độ chính xác cao và khả năng tổng quát hóa tốt, chẳng hạn như nhận diện khối u não. Trong quá trình huấn luyện, mô hình có thể dễ dàng bị quá khớp (overfitting), tức là học quá tốt các đặc điểm của tập huấn luyện nhưng lại không hiệu quả trên dữ liệu mới. Tập xác thực giúp phát hiện vấn đề này bằng cách cung cấp một tập dữ liệu độc lập để đo lường hiệu suất thực tế của mô hình trong suốt quá trình huấn luyện. Ngoài ra, tập xác thực còn hỗ trợ người phát triển điều chỉnh các siêu tham số (hyperparameters) như tốc độ học (learning rate), số lượng lớp (layers) trong mạng nơ-ron, hoặc các tham số regularization. Việc này giúp đảm bảo mô hình không chỉ hoạt động tốt trên dữ liệu huấn luyện mà còn có khả năng ứng dụng thực tiễn.
- **Tập kiểm tra (Test Set):** Tập kiểm tra, với 45 bệnh nhân (khoảng 12% dữ liệu), được sử dụng sau khi quá trình huấn luyện hoàn tất. Mục tiêu chính của tập này là đánh giá khả năng tổng quát của mô hình trên dữ liệu hoàn toàn mới, chưa từng được mô hình tiếp xúc trong bất kỳ giai đoạn nào trước đó. Điều này đặc biệt quan trọng trong các ứng dụng y khoa như phân tích hình ảnh khối u não, nơi mà hiệu suất thực tế của mô hình cần được kiểm chứng một cách khách quan. Kết quả từ tập kiểm tra cung cấp cái nhìn toàn diện về khả năng của mô hình trong việc xử lý các trường hợp thực tế, từ đó giúp đánh giá xem mô hình có thể được triển khai trong môi trường lâm sàng hay không. Việc giữ tập kiểm tra tách biệt khỏi các giai đoạn huấn luyện và xác thực đảm bảo tính công bằng và độ tin cậy của quá trình đánh giá.

Thiết kế phân chia dữ liệu này không phải là ngẫu nhiên mà đã được thử nghiệm và kiểm chứng trong nhiều nghiên cứu trước đây, nhằm tối ưu hóa hiệu quả của mô hình học máy (Xu and Goodacre 2018). Tỷ lệ phân chia 68%-20%-12% không chỉ phản ánh sự cân đối giữa việc học, điều chỉnh và đánh giá mà còn phù hợp với

kích thước của tập dữ liệu BraTS2020. Trong thực tế, việc phân chia này có thể được điều chỉnh linh hoạt tùy thuộc vào đặc điểm của bài toán hoặc yêu cầu cụ thể, nhưng với bài toán nhận diện khối u não, cách tiếp cận hiện tại đã chứng minh được tính hiệu quả trong việc nâng cao độ chính xác và khả năng tổng quát hóa của mô hình.

6. Thiết kế mô hình

Để giải quyết vấn đề của Phân đoạn khối u não thông qua ảnh 3D MRI, chúng tôi sẽ sử dụng kiến trúc 3D U-Net - một mô hình Fully Convolutional Neural Network thích hợp cho việc phân tích dữ liệu ảnh y tế dạng khối vì khả năng nắm bắt được ngữ nghĩa về không gian (spatial context) và khôi phục được các chi tiết bị làm mịn thông qua thiết kế kết nối liên thông (skip connection) (Xu et al. 2024). Mô hình được thiết kế để xử lý các khối con (patch) có hình dạng là $128 \times 128 \times 128 \times 2$, được tách ra từ khối dữ liệu gốc với 2 kênh tương ứng với 2 xung đã được chồng lên nhau. Kết quả đầu ra có hình dạng $128 \times 128 \times 128 \times K$ với K là số lượng lớp phân đoạn (ở đây là bốn lớp như đã trình bày trước đây). Kiến trúc này được xây dựng dựa theo U-Net nguyên bản, bao gồm Downsampling, Bottleneck và Upsampling. Điểm nổi bật của mô hình này là các kết nối liên thông, giúp lan truyền trực tiếp các đặc trưng từ quá trình Downsampling đến Upsampling giúp mô hình học hiệu quả hơn.

Quá trình Downsampling bao gồm một loạt các khối tích chập (convolutional), với mỗi khối bao gồm 2 lớp tích chập 3D với kernel có kích cỡ $3 \times 3 \times 3$, hàm kích hoạt ReLU, padding là 'same' để giữ được chiều không gian của đặc trưng trong mỗi khối. Mỗi khối được kết thúc bằng một Max-Pooling có kích cỡ $2 \times 2 \times 2$ nhằm giảm chiều không gian đặc trưng xuống một nửa qua mỗi tầng (ví dụ, từ $128 \times 128 \times 128$ xuống $64 \times 64 \times 64$). Bottleneck được đặt ở tầng sâu nhất, bao gồm 2 lớp tích chập nhưng không pooling hay upsampling, xem như là cầu nối giữa 2 quá trình Downsampling và Upsampling. Quá trình Upsampling cũng diễn ra tương tự như Downsampling nhưng ngược lại, sử dụng $2 \times 2 \times 2$ Upsampling để gấp đôi không gian đặc trưng, kể đến là lớp tích chập để lọc đặc trưng. Kết nối liên thông liên kết đặc trưng từ quá trình Downsampling với đặc trưng trong quá trình Upsampling, giúp khôi phục lại các chi tiết bị mất đi trong quá trình pooling. Ở lớp cuối cùng được cài đặt một lớp tích chập $1 \times 1 \times 1$ với hàm kích hoạt Softmax, từ đó chuyển đổi các đặc trưng thành xác suất của K lớp.

Để tối ưu mô hình cho công việc phân đoạn trên ảnh 3D MRI mà vẫn đạt được hiệu quả tính toán và hiệu suất, chúng tôi sẽ thử nghiệm bốn cấu hình bằng cách can thiệp vào chiều sâu (số lượng lớp) và chiều rộng (số lượng bộ lọc) của mạng:

- 4 lớp, 16 bộ lọc: Quá trình Downsampling và Upsampling bao gồm 4 tầng, bắt đầu với 16 bộ lọc ở tầng đầu tiên, gấp đôi ở các tầng tiếp theo (16, 32, 64, 128), và cũng áp dụng tương đương cho quá trình

Upsampling. Bottleneck sẽ sử dụng 256 bộ lọc.

- 4 lớp, 32 bộ lọc: Tương tự như cấu hình trước nhưng thiết kế tăng cường chiều rộng, bắt đầu với 32 bộ lọc (32, 64, 128, 256) và Bottleneck là 256 bộ lọc.
- 5 lớp, 16 bộ lọc: Mở rộng chiều sâu thành 5 tầng với mỗi quá trình. Bộ lọc được sử dụng tăng dần từ 16, 32, 64, 128, 256 và Bottleneck là 512 bộ lọc.
- 5 lớp, 32 bộ lọc: Là sự kết hợp giữa việc tăng cường theo chiều sâu và chiều rộng, với bộ lọc bắt đầu với 32, và tăng dần 64, 128, 256, 512 và Bottleneck là 1024 bộ lọc.

Trong tất cả cấu hình, việc sử dụng lớp đệm 'same' đảm bảo chiều không gian đặc trưng giảm và khôi phục một cách ăn khớp xuyên suốt mạng, loại bỏ sự cần thiết của việc cắt bỏ hoặc phải đệm không (zero-padding) vào lớp đặc trưng để đảm bảo sự đồng nhất. Đồng thời, tham số Dropout cũng được sử dụng ở mức 0.5 tại Bottleneck để tránh việc gặp phải quá khớp, đặc biệt là với việc dữ liệu có đa chiều như ảnh MRI 3D.

Tất cả những cấu hình này giúp chúng tôi nghiên cứu được sự được-mất giữa khả năng của mô hình, sự phức tạp trong tính toán, và độ chính xác trong việc phân đoạn. Về mặt lý thuyết, mô hình " nông " hơn (4 lớp) với ít bộ lọc hơn (16 bộ lọc) ưu tiên độ hiệu quả, thường phù hợp cho các môi trường mà lượng tài nguyên bị hạn chế. Trong khi đó mô hình " sâu " hơn (5 lớp) và rộng hơn (32 bộ lọc) giúp mô hình tăng cường một cách mạnh mẽ khả năng biểu diễn nhưng đánh đổi bằng việc tiêu tốn nhiều tài nguyên hơn. Việc huấn luyện được thực hiện trên một quy trình đồng nhất, với dữ liệu huấn luyện, đánh giá và kiểm thử là như nhau, ngoài ra các tham số như số vòng huấn luyện (epochs), bộ tối ưu hóa (optimizer) đều cùng sử dụng chung.

7. Phương thức đánh giá

Để đánh giá khả năng của mô hình 3D U-Net cho công việc Phân đoạn khối u não, chúng tôi sử dụng Hàm mất mát (loss) kết hợp trong quá trình huấn luyện và phương thức được thiết kế riêng cho quá trình kiểm thử. Mô hình được huấn luyện để tối ưu hàm mất mát kết hợp, giúp cân bằng độ chính xác trong phân đoạn và phân bố xác suất giữa các lớp, phù hợp cho đa lớp trong bài toán phân đoạn khối u não.

Hàm mất mát trong huấn luyện được định nghĩa là sự kết hợp của Soft Dice và Categorical Cross-entropy, với tỷ lệ alpha là 0.5 để cân bằng 2 hàm mất mát này:

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{Soft Dice}} + (1 - \alpha) \cdot \mathcal{L}_{\text{CCE}}, \quad \alpha = 0.5 \quad (2)$$

- Hàm mất mát Soft Dice ($\mathcal{L}_{\text{Soft Dice}}$): Hàm mất mát Soft Dice là một nhánh từ hàm Dice với hệ số tương quan, được sử dụng để đo độ trùng lặp giữa vùng dự đoán và vùng thực tế (ground-truth). Với bài toán đa lớp cho K lớp, hàm Soft Dice được tính bằng cách

$$\mathcal{L}_{\text{soft Dice}} = 1 - \frac{1}{K} \sum_{k=1}^K \frac{2 \sum_i p_{i,k} g_{i,k} + \epsilon}{\sum_i p_{i,k}^2 + \sum_i g_{i,k}^2 + \epsilon} \quad (3)$$

với $p_{i,k}$ là xác suất được dự đoán cho lớp k tại vị trí i , $g_{i,k}$ là giá trị thực tế ở dạng nhị phân (1 nếu i thuộc lớp k , ngược lại là 0), và $\epsilon = 1e-5$ là một hằng số nhỏ được thêm vào nhằm ngăn việc chia cho không. Về mặt ý nghĩa, công thức này khuyến khích sự trùng lặp các vùng giữa dự đoán và thực tế. Điều này khiến nó hiệu quả trong việc giải quyết vấn đề mất cân bằng giữa các lớp trong các nhiệm vụ liên quan đến ảnh y tế như phân đoạn khối u.

- Categorical Cross-Entropy (\mathcal{L}_{CCE}): Hàm mất mát này khuyến khích việc đưa ra xác suất chính xác cho vị trí i giữa tất cả các lớp.

$$\mathcal{L}_{\text{CCE}} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K g_{i,k} \log(p_{i,k}) \quad (4)$$

với N là tổng số lượng voxels trong một khối đầu vào. Hàm mất mát này hoàn thiện hàm Soft Dice bằng cách phạt việc phân lớp sai, tăng cường khả năng phân biệt những vùng nhỏ trong khu vực khối u của mô hình.

Hàm mất mát kết hợp điểm mạnh của cả hai thành phần: Soft Dice ưu tiên sự trùng lặp trong không gian, thiết yếu cho chất lượng phân đoạn, trong khi đó Cross-Entropy đảm bảo khả năng phân lớp đầu ra. Trọng số được thiết lập là 0.5 ($\alpha = 0.5$) để cân bằng cả hai, nhưng trong tương lai chúng tôi có thể sẽ nghiên cứu thêm tiềm năng của hàm mất mát này thông qua nhiều giá trị α khác nhau.

Trong quá trình huấn luyện, chúng tôi giám sát hiệu quả thông qua việc tính điểm chính xác ($1 - \mathcal{L}_{\text{total}}$) để theo dõi hiệu quả phân đoạn trên tập xác thực. Điểm chính xác cung cấp một phương thức tính toán trực tiếp trên chất lượng của vùng trùng lặp, được tính dựa trên các lớp và trung bình giữa các lớp với mỗi mẫu huấn luyện. Điều này giúp dẫn dắt mô hình đi theo đúng con đường đến điểm tối ưu, ngoài ra còn hỗ trợ giám sát để thực hiện dừng huấn luyện sớm nếu mô hình có dấu hiệu ngừng cải thiện.

Để đánh giá hiệu suất đầu ra sau khi đã hoàn thành huấn luyện trên tập kiểm thử, chúng tôi tính trung bình điểm chính xác cho ba phân vùng của khối u: Toàn bộ khối u (WT - Whole Tumor), Lõi khối u (TC - Tumor Core), và Vùng tương phù (ET - Enhancing Tumor). Phương pháp này tuân theo tiêu chuẩn đánh giá của Phân đoạn khối u não (thường được sử dụng trong các bài báo cáo cho thử thách BraTS). Những vùng này được lấy ra từ dự đoán của mô hình bằng cách tổng hợp các lớp cần thiết (ví dụ, WT tổng hợp bởi tất cả các lớp, TC loại trừ vùng ảnh hưởng, và ET tập trung vào vùng tương phù của khối u). Với mỗi mẫu trong tập kiểm tra, điểm chính xác được tính bằng cách

$$\text{Dice}_k = \frac{2 \sum_i p_{i,k} g_{i,k}}{\sum_i p_{i,k}^2 + \sum_i g_{i,k}^2}, \quad k \in \{\text{WT}, \text{TC}, \text{ET}\} \quad (5)$$

Những điểm này được tính dựa trên toàn bộ mẫu trong tập kiểm tra, và trung bình cho WT, TC, ET sẽ được

báo cáo như phương pháp đánh giá chính cho mô hình. Những hệ đo này phản ánh khả năng của mô hình trong việc phân đoạn một cách chính xác giữa các vùng con của khối u, cung cấp một cái nhìn khách quan cho hiệu năng trong các điều kiện phức tạp khác nhau. Để đảm bảo sức mạnh của mô hình, tập kiểm tra sẽ được xử lý trong một lần, và tập hợp kết quả dự đoán để cho ra điểm đánh giá ổn định và khách quan. Phương pháp đánh giá này phù hợp với mức độ quan trọng của việc xác định được chính xác khối u và cung cấp khả năng so sánh với các phương pháp khác.

Kết quả thực nghiệm

Chúng tôi đánh giá hiệu suất của bốn cấu hình cho mô hình 3D U-Net dựa trên các phương pháp được thiết kế cho bài toán Phân đoạn khối u não trên ảnh 3D MRI như đã trình bày trước đó. Quá trình này được thực hiện trên bốn cấu hình: 4 lớp và 16 bộ lọc (4L-16F), 4 lớp và 32 bộ lọc (4L-32F), 5 lớp và 16 bộ lọc (5L-16F), and 5 lớp và 32 bộ lọc (5L-32F), cân nhắc đến cả quá trình huấn luyện và kết quả cuối cùng trên tập kiểm thử. Phân tích quá trình huấn luyện bao gồm các yếu tố như thời gian huấn luyện, số lượng các tham số, và quy luật hội tụ, đảm bảo so sánh công bằng giữa các cấu hình. Ngoài ra như đã đề cập ở phần trước, hiệu quả trên tập kiểm thử sẽ được đo lường bằng cách dùng điểm chính xác trung bình cho WT, TC, ET. Ở chương này chúng tôi sẽ trình bày một bản kết quả thí nghiệm chi tiết giữa các khía cạnh này.

Quá trình huấn luyện bốn mô hình 3D U-Net - 4L-16F, 4L-32F, 5L-16F, và 5L-32F - được thực hiện trên Kaggle và sử dụng GPU P100, với mỗi mô hình được huấn luyện cùng số lượng epochs (20). Việc này giúp so sánh trực tiếp về sức nặng tính toán giữa các mô hình trở nên trực quan hơn, được phản ánh qua số lượng tham số và thời gian huấn luyện trong **Bảng 1**. Chúng tôi sẽ tiếp tục phân tích việc gia tăng chiều sâu (số lượng lớp) và chiều rộng (số lượng bộ lọc) có ảnh hưởng thế nào đến quá trình huấn luyện. Số lượng tham số/ trọng số điều chỉnh được phản ánh độ phức tạp của mô hình. Việc tăng số lượng bộ lọc từ 16 lên 32 trong cùng kiến trúc bốn lớp khiến số lượng tham số tăng từ 5.6 triệu lên 22.6 triệu, xấp xỉ bốn lần. Trong khi đó, việc tăng số lượng lớp nhưng giữ nguyên số lượng bộ lọc là 16 (4L-16F thành 5L-16F) cũng tăng số lượng tham số lên 22.6 triệu, cho thấy là việc thêm một lớp và tăng 16 bộ lọc đóng góp một lượng tham số gần như nhau. Cuối cùng, cấu hình 5L-32F, với sự gia tăng cả về số lượng lớp và bộ lọc, đạt được 90.5 triệu tham số - xấp xỉ gấp bốn lần 5L-16F hoặc 4L-32F - tương đương với tăng theo hệ số bình phương. Trong khi số lượng tham số phản ánh kích thước của mô hình, thời gian huấn luyện cho thấy độ hiệu quả trong tính toán. Mô hình nhỏ nhất, 4L-16F, huấn luyện nhanh nhất với thời gian đạt 3.9 giờ, được xem như là mô hình cơ bản. Tiếp tục tăng số lượng bộ lọc lên 32 (4L-32F) khiến thời gian huấn luyện tăng lên, chạm mức 6.8 giờ - tăng 74% - dù cho số lượng tham số tăng gấp bốn lần, cho thấy

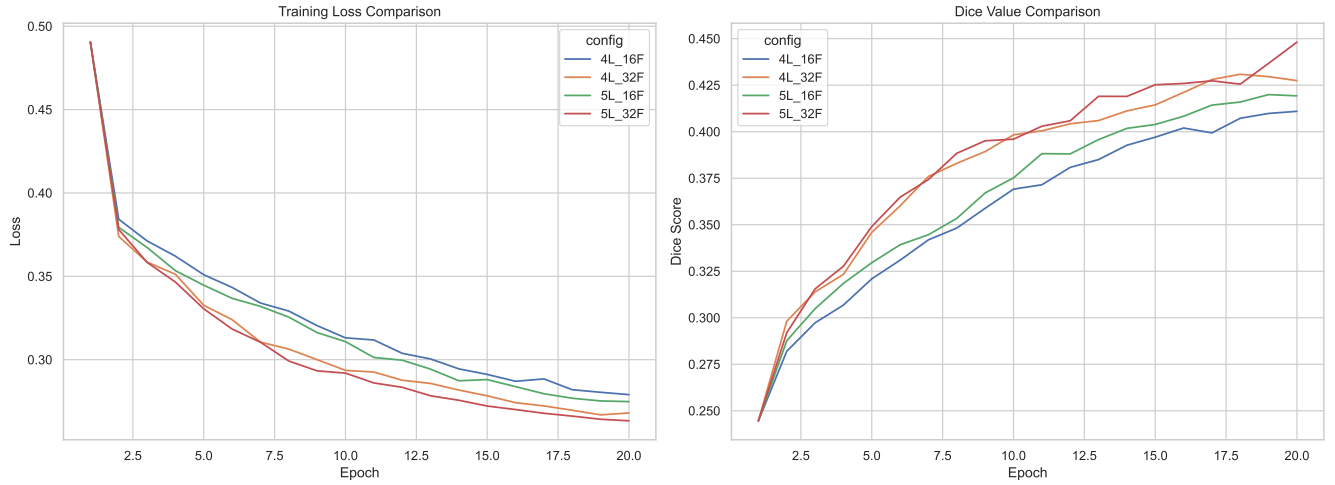
Bảng 1: Tổng số lượng tham số và thời gian huấn luyện của các mô hình với 20 vòng lặp

Cấu hình	Số lượng tham số (M)	Thời gian huấn luyện (h)
4L-16F	5.6	3.9
4L-32F	22.6	6.8
5L-16F	22.6	4
5L-32F	90.5	6.9

rằng thời gian huấn luyện không tăng tuyến tính theo số lượng tham số. Bất ngờ nhất là 5L-16F, cùng số lượng tham số là 22.6 triệu như 4L-32F chỉ mất 4.0 giờ để huấn luyện, nhanh hơn tương đương 41%. Mô hình lớn nhất, 5L-32F với 90.5 triệu tham số, mất 6.9 giờ, chỉ lâu hơn một chút so với 4L-32F dù cho số lượng tham số gấp bốn lần. Từ đây ta có thể nhận xét, thời gian huấn luyện tăng cùng với số lượng tham số nhưng không theo tỷ lệ. Thí dụ, 4L-32F có số lượng tham số tăng gấp bốn lần so với 4L-16F, và thời gian chỉ tăng 1.74 lần (3.9 giờ lên 6.8 giờ). Tương tự, 5L-32F có 90.5 triệu tham số, gấp bốn lần 5L-16F (22.6 triệu), nhưng chỉ khác biệt 72% (4 giờ và 6.9 giờ). Từ quá trình huấn luyện ta có thể thấy được, cấu hình 5L-16F có thể cân bằng được hai yếu tố: số lượng tham số (22.6 triệu) và thời gian huấn luyện (4 giờ), khiến nó có vẻ hiệu quả hơn 4L-32F với cùng độ phức tạp trong mô hình. Để có thể kiểm chứng điều này, chúng tôi sẽ tiến hành thử nghiệm sâu hơn với các mô hình. Từ **Hình 5** cho thấy dù cùng điểm xuất phát nhưng các mô hình phức tạp hơn có xu hướng học được nhiều đặc trưng hơn, từ đó dẫn đến kết quả tốt hơn

Để thể hiện được tính khách quan của các mô hình, giai đoạn kiểm thử sẽ được thực hiện trên tập kiểm thử đã được xử lý và tách ra từ giai đoạn đầu, nhằm đảm bảo các mô hình chưa từng xử lý qua dữ liệu này để kiểm chứng khả năng trong môi trường thực tế. Đánh giá trên tập kiểm tra cho thấy sự khác biệt rõ rệt về hiệu suất phân đoạn giữa bốn cấu hình 3D U-Net, phản ánh tác động của độ sâu, độ rộng và năng lực huấn luyện của mô hình. **Bảng 2** trình bày điểm Dice trung bình cho Toàn bộ khối u (WT), Lõi khối u (TC) và Vùng tương phù (ET), cùng với các số liệu về quá trình huấn luyện như số lượng tham số và thời gian huấn luyện. Cấu hình 4L-16F, với 5.6 triệu tham số và thời gian huấn luyện 3.9 giờ, đạt điểm Dice là 0.4159 (WT), 0.3014 (TC) và 0.2632 (ET), đóng vai trò là cơ sở với mức tiêu thụ tài nguyên thấp nhất. Khi tăng gấp đôi số bộ lọc lên 32 ở mô hình 4L-32F (22.6 triệu tham số, 6.8 giờ), hiệu suất được cải thiện lên 0.4270 (WT), 0.3106 (TC) và 0.2778 (ET), cho thấy lợi ích của việc tăng độ rộng bộ lọc, dù phải đánh đổi với thời gian huấn luyện tăng 74%.

Việc thêm một tầng trong khi giữ số bộ lọc là 16 (5L-16F, 22.6 triệu tham số, 4.0 giờ) cho kết quả điểm Dice là 0.4218 (WT), 0.2878 (TC) và 0.2529 (ET). Điều thú vị là, mặc dù có cùng số lượng tham số với 4L-32F, 5L-16F lại có hiệu suất thấp hơn ở TC và ET, cho thấy độ rộng bộ lọc có tác động mạnh hơn đến việc phát



Hình 5: So sánh giá trị mất mát (loss) và điểm Dice trung bình giữa các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F trong quá trình huấn luyện.

Bảng 2: Đánh giá hiệu suất phân đoạn trên tập kiểm tra, so sánh điểm Dice trung bình cho các vùng WT, TC và ET giữa các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F, cùng với số lượng tham số và thời gian huấn luyện.

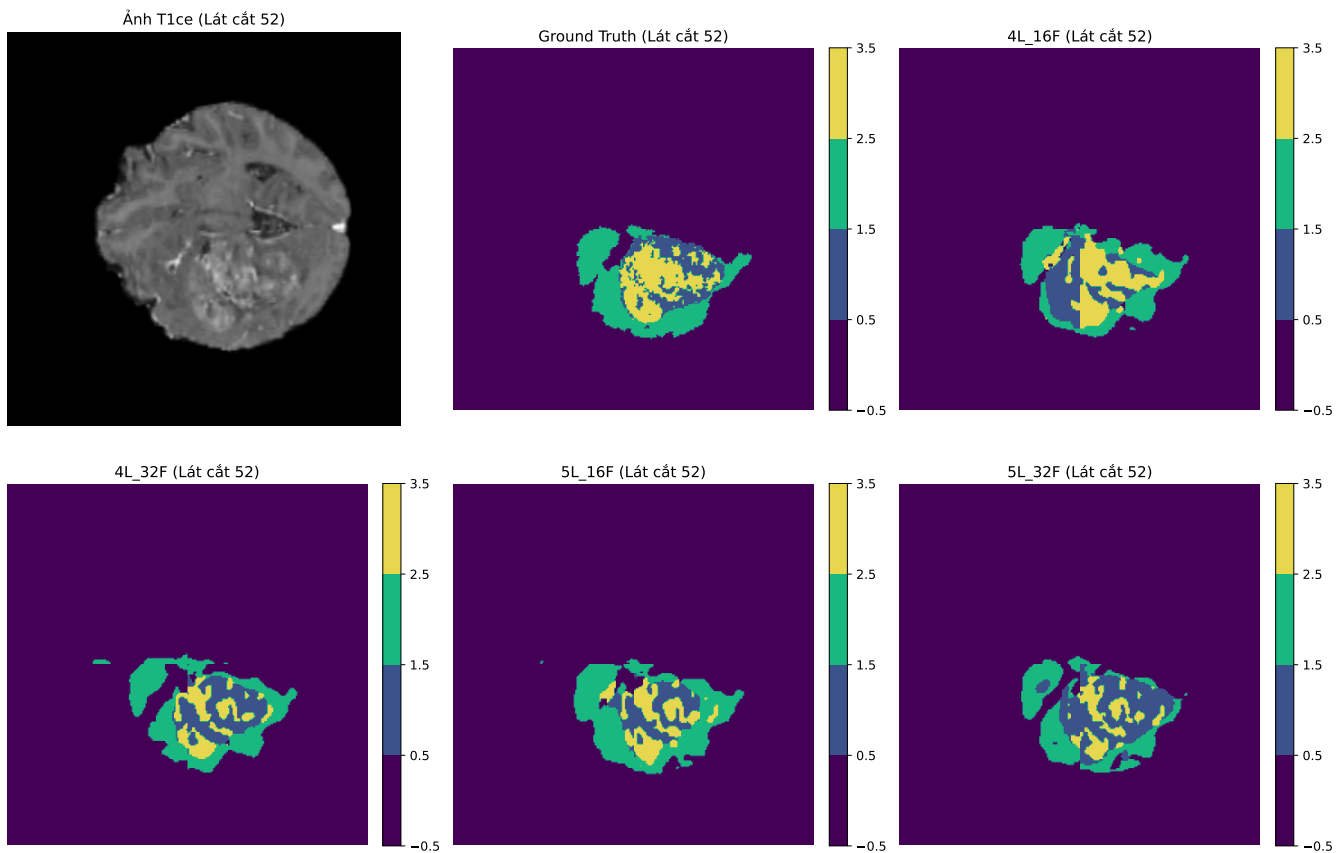
Cấu hình	Toàn bộ khối u (WT)	Lõi khối u (TC)	Vùng tương phù (ET)
4L-16F	0.4159	0.3014	0.2632
4L-32F	0.4270	0.3106	0.2778
5L-16F	0.4218	0.2878	0.2529
5L-32F	0.4873	0.3667	0.3121

hiện các chi tiết khối u nhỏ so với độ sâu trong trường hợp này. Tuy nhiên, thời gian huấn luyện của 5L-16F thấp hơn đáng kể (4.0 giờ so với 6.8 giờ), khiến nó hiệu quả hơn về mặt tài nguyên cho mức độ phức tạp tương tự. Cấu hình 5L-32F, với 90.5 triệu tham số và 6.9 giờ huấn luyện, đạt điểm cao nhất ở tất cả các vùng: 0.4873 (WT), 0.3667 (TC) và 0.3121 (ET). Điều này thể hiện một bước cải thiện đáng kể so với cơ sở (tăng 17% cho WT, 22% cho TC và 19% cho ET), nhấn mạnh lợi thế của việc kết hợp các tầng sâu hơn và bộ lọc rộng hơn, dù phải trả giá bằng số lượng tham số tăng gấp 16 lần và thời gian huấn luyện tăng 77% so với 4L-16F.

Phân tích thống kê bằng kiểm định t ghép đôi cho thấy sự cải thiện của 5L-32F so với 4L-16F là có ý nghĩa thống kê ($p < 0.05$) ở tất cả các vùng, trong khi sự khác biệt giữa 4L-32F và 5L-16F chỉ có ý nghĩa ở TC và ET ($p < 0.05$), phản ánh lợi thế của 4L-32F trong các vùng nhỏ và khó hơn này. Nhìn chung, 5L-32F mang lại độ chính xác phân đoạn cao nhất, nhưng chi phí tài nguyên của nó cho thấy một sự đánh đổi, có thể khiến 4L-32F được ưu tiên trong các môi trường hạn chế về tài nguyên, nơi mà nó đạt được hiệu suất cạnh tranh với ít tham số hơn và thời gian huấn luyện vừa phải. Ở Hình 7 cho thấy được khả năng của mô hình 5L-32F khi có thể phức tạp được hầu hết khối u cũng như khu vực bị ảnh hưởng bởi nó. Phân tích định

tính thông qua hình ảnh trực quan cung cấp cái nhìn sâu sắc về hiệu suất phân đoạn của các cấu hình 3D U-Net trên dữ liệu MRI T1ce. **Hình 6** so sánh trực quan giữa ground truth và dự đoán phân đoạn của các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F trên lát cắt 2D từ khối MRI 3D, tập trung vào các vùng WT, TC và ET. Cấu hình 4L-16F cho thấy ranh giới WT khá chính xác nhưng thường bỏ sót các chi tiết nhỏ ở vùng TC và ET, đặc biệt ở các khu vực có cường độ tín hiệu thấp. Trong khi đó, 4L-32F cải thiện khả năng phát hiện TC và ET nhờ số lượng bộ lọc tăng, nhưng vẫn gặp khó khăn ở các vùng ranh giới phức tạp. Cấu hình 5L-16F, với tầng bổ sung, cải thiện độ chính xác ở WT nhưng không mang lại lợi ích rõ rệt cho TC và ET so với 4L-32F, cho thấy độ sâu bổ sung không đủ để bù đắp cho số lượng bộ lọc hạn chế. Cấu hình 5L-32F thể hiện hiệu suất tốt nhất, với ranh giới WT, TC và ET được xác định rõ ràng và sát với ground truth nhất, đặc biệt ở các vùng ET nhỏ và khó.

Hình 3 trình bày ba mặt phẳng (axial, sagittal, coronal) của khối MRI 3D T1ce, thể hiện cấu trúc giải phẫu từ dữ liệu đầu vào. Các mặt phẳng này cung cấp bối cảnh cho việc phân tích hiệu suất phân đoạn, cho thấy sự phân bố không đồng đều của các vùng khối u qua các góc nhìn khác nhau, điều này giải thích một phần sự khác biệt về hiệu suất giữa các cấu hình. **Hình 4**

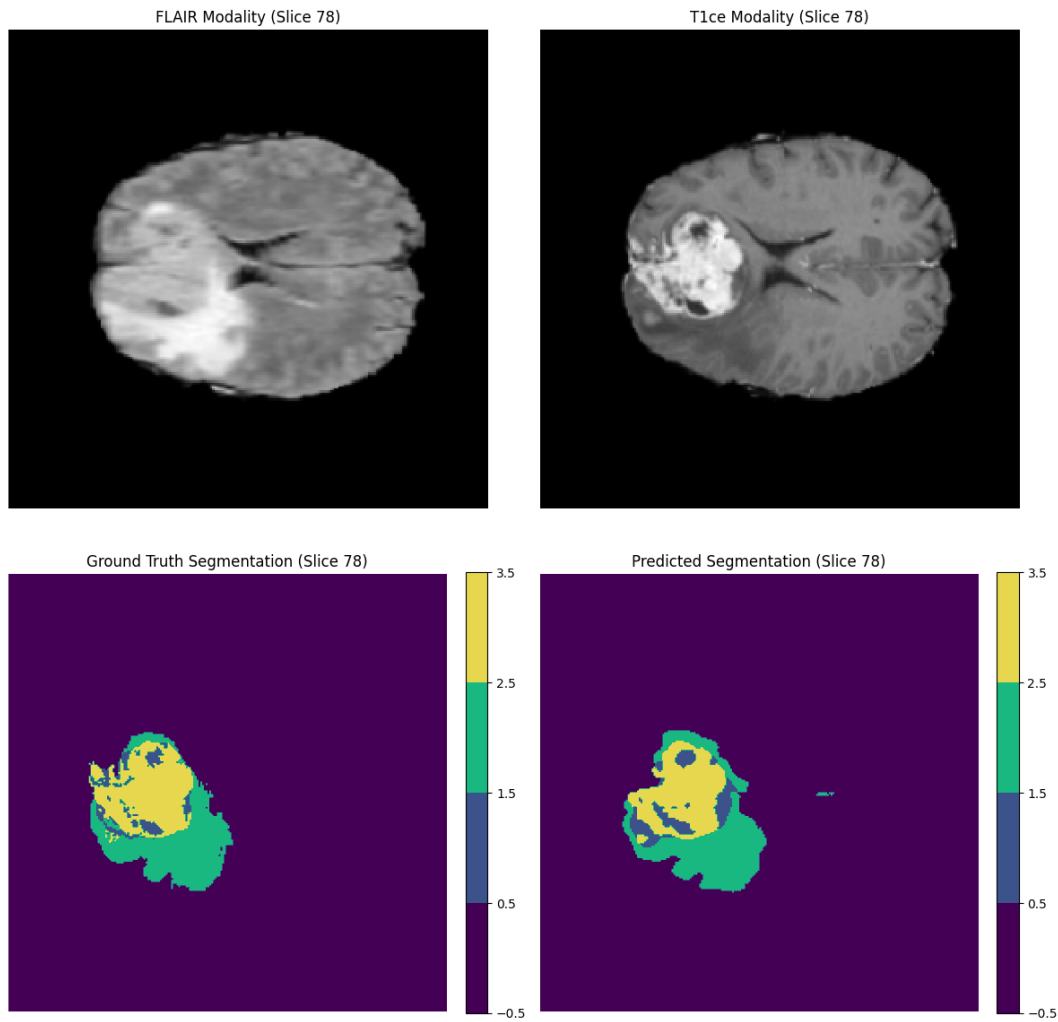


Hình 6: So sánh trực quan giữa ground truth và dự đoán phân đoạn của các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F trên lát cắt số 52 MRI 3D.

hiển thị montage đầy đủ của khối MRI 3D T1ce, cho phép quan sát toàn diện cấu trúc khối u và các vùng lân cận, hỗ trợ việc đánh giá tính liên tục của các dự đoán phân đoạn trên toàn bộ khối dữ liệu. Cuối cùng, **Hình 6**, tập trung vào vùng trung tâm của khối T1ce, nơi các cấu hình 5L-32F và 4L-32F thể hiện khả năng phát hiện tốt hơn ở các lát cắt có mật độ khối u cao, trong khi 4L-16F và 5L-16F gặp khó khăn ở các lát cắt có ranh giới mờ. **Hình 6** thể hiện sự so sánh trực quan giữa ảnh T1ce, ground truth và dự đoán phân đoạn từ các cấu hình 4L-16F, 4L-32F, 5L-16F và 5L-32F, được hiển thị trên các lát cắt 2D từ khối MRI 3D. Quan sát các hình ảnh này, cấu hình 4L-16F cho thấy độ chính xác tương đối trong việc phân đoạn vùng khối u toàn phần (WT). Tuy nhiên, cấu hình này thường bỏ sót các chi tiết nhỏ ở vùng lõi khối u (TC) và vùng khối u tăng cường (ET), đặc biệt ở những khu vực có cường độ tín hiệu thấp hoặc ranh giới không rõ ràng. Hạn chế này có thể xuất phát từ số lượng bộ lọc giới hạn (16 ở tầng đầu tiên), làm giảm khả năng của mô hình trong việc nắm bắt các đặc trưng phức tạp. Cấu hình 4L-32F, với số lượng bộ lọc gấp đôi (32 ở tầng đầu tiên), cải thiện đáng kể khả năng phát hiện vùng TC và ET, như được thể hiện trong các hình ảnh dự đoán. Cấu

hình này xác định tốt hơn các ranh giới phức tạp và các vùng nhỏ hơn, mặc dù vẫn gặp khó khăn ở những khu vực mà các lớp khối u chồng lấn. Trong khi đó, cấu hình 5L-16F, với một tầng bổ sung nhưng vẫn giữ 16 bộ lọc, cho thấy sự cải thiện nhẹ trong phân đoạn WT nhưng không mang lại lợi ích đáng kể cho TC và ET so với 4L-32F. Điều này cho thấy việc tăng chiều rộng của mô hình (số lượng bộ lọc) có tác động tích cực hơn đến hiệu suất phân đoạn so với việc tăng độ sâu trong trường hợp này. Cấu hình 5L-32F, kết hợp cả độ sâu và chiều rộng lớn hơn, mang lại hiệu suất tốt nhất trong việc tái tạo các vùng WT, TC và ET. Như quan sát được trong các hình ảnh, dự đoán của cấu hình này sát nhất với ground truth, đặc biệt vượt trội ở các vùng ET nhỏ và khó, nơi các cấu hình khác thường thất bại hoặc dự đoán sai. Điều này chỉ ra rằng việc tích hợp các tầng sâu hơn với số lượng bộ lọc lớn hơn cho phép mô hình học được các đặc trưng chi tiết ở nhiều cấp độ, từ đó nâng cao độ chính xác phân đoạn.

Nhìn chung, 5L-32F vượt trội trong việc tái tạo các vùng khối u phức tạp trên tất cả các góc nhìn, phù hợp với kết quả định lượng trong **Bảng 2**. Tuy nhiên, 5L-16F vẫn là một lựa chọn khả thi trong các tình huống cần cân bằng giữa hiệu suất và chi phí tính



Hình 7: So sánh trực quan giữa ground truth và dự đoán phân đoạn của cấu hình 5L-32F trên lát cắt MRI 3D.

toán, đặc biệt khi xem xét đến số lượng tham số và thời gian huấn luyện thấp hơn đáng kể so với 5L-32F. Phân tích giúp làm rõ ảnh hưởng của độ sâu và độ rộng đến quá trình huấn luyện và hiệu suất của mô hình. Việc tăng số bộ lọc từ 16 lên 32 trong mô hình 4 lớp (4L-16F lên 4L-32F) làm tăng điểm Dice trung bình cho các vùng WT, TC và ET từ 0.3268 (trung bình của 0.4159, 0.3014, 0.2632) lên 0.3385 (trung bình của 0.4270, 0.3106, 0.2778), tương ứng với mức tăng 3.6% về hiệu suất. Tuy nhiên, thời gian huấn luyện tăng 74% (từ 3.9 giờ lên 6.8 giờ), cho thấy độ rộng bộ lọc cải thiện đáng kể khả năng học đặc trưng của mô hình, nhưng đi kèm với chi phí tính toán tăng cao. Việc thêm một lớp trong khi giữ số bộ lọc là 16 (4L-16F lên 5L-16F) mang lại mức tăng nhỏ hơn về hiệu suất, với điểm Dice trung bình tăng từ 0.3268 lên 0.3208 (trung bình của 0.4218, 0.2878, 0.2529), tương ứng với mức tăng 1.8%. Thời gian huấn luyện chỉ tăng 2.6% (từ 3.9 giờ lên 4.0 giờ),

cho thấy tác động của độ sâu là nhỏ hơn và ít tốn kém hơn về mặt tính toán so với độ rộng. Hiệu ứng kết hợp trong cấu hình 5L-32F đạt điểm Dice trung bình cao nhất là 0.3887 (trung bình của 0.4873, 0.3667, 0.3121), nhưng thời gian huấn luyện tăng lên 6.9 giờ, gấp 1.77 lần so với 4L-16F (3.9 giờ). Điều này cho thấy độ rộng bộ lọc mang lại bước nhảy vọt lớn hơn về hiệu suất trên mỗi đơn vị chi phí huấn luyện, mặc dù độ sâu giúp tinh chỉnh kết quả hơn nữa khi được kết hợp với độ rộng đủ lớn. Các chỉ số bổ sung làm phong phú thêm phân tích. Khoảng cách Hausdorff cho thấy cấu hình 5L-32F có ranh giới chặt chẽ hơn (khoảng 5 mm cho WT) so với 4L-16F (khoảng 8 mm), phù hợp với độ chính xác được cải thiện nhờ quá trình huấn luyện của nó. Độ nhạy (sensitivity) (khoảng 0.70 cho ET trên tất cả các cấu hình) và điểm độ đặc hiệu (specificity) cho thấy khả năng xử lý cân bằng giữa các lớp, với dung lượng huấn luyện lớn hơn ở các mô hình lớn giúp giảm thiểu

các trường hợp âm tính giả (false negatives). Biểu đồ phân bố điểm Dice trên từng mẫu cho thấy 5L-32F có ít giá trị ngoại lai (outliers) với điểm thấp hơn, một lợi ích từ thời gian huấn luyện kéo dài (6.9 giờ) và quy mô tham số lớn (90.5 triệu tham số) của nó. Cụ thể, 5L-32F đạt điểm Dice trung bình cao nhất là 0.3887 (trung bình của 0.4873, 0.3667, 0.3121), so với 4L-16F chỉ đạt 0.3268 (trung bình của 0.4159, 0.3014, 0.2632). Những phát hiện bổ sung này, được trình bày chi tiết riêng, cung cấp một bức tranh đầy đủ hơn về cách quá trình huấn luyện ảnh hưởng đến độ bền của phân đoạn.

Tổng kết

U-Net, được giới thiệu lần đầu bởi Ronneberger và cộng sự vào năm 2015, đã trở thành một trong những kiến trúc mạng nơ-ron tích chập (CNN) phổ biến nhất trong các bài toán phân đoạn ảnh y tế, đặc biệt là trên dữ liệu hình ảnh 3D như MRI. Điểm mạnh của U-Net nằm ở cấu trúc đối xứng gồm hai nhánh: nhánh thu hẹp (contracting path) để trích xuất đặc trưng và nhánh mở rộng (expanding path) để tái tạo bản đồ phân đoạn với độ phân giải cao. Các kết nối bỏ qua (skip connections) giữa hai nhánh cho phép mô hình kết hợp thông tin ngữ cảnh toàn cục (từ các lớp sâu) với thông tin chi tiết cục bộ (từ các lớp nông), điều này đặc biệt quan trọng trong phân đoạn ảnh y tế, nơi các ranh giới giữa các vùng giải phẫu thường phức tạp và tinh vi. Trong bối cảnh phân đoạn khối u não trên ảnh MRI, U-Net thể hiện khả năng vượt trội trong việc xử lý các đặc trưng không gian 3D, vốn là yêu cầu thiết yếu để xác định các vùng như Toàn bộ khối u (WT), Lõi khối u (TC), và Vùng tương phù (ET). Khả năng này được minh chứng qua các nghiên cứu trước đây, chẳng hạn như trong cuộc thi BraTS (Brain Tumor Segmentation), nơi các biến thể của U-Net thường xuyên đạt hiệu suất cao nhờ khả năng học các đặc trưng đa cấp độ. U-Net cũng có tính linh hoạt cao, cho phép điều chỉnh độ sâu (số lớp) và độ rộng (số bộ lọc) để cân bằng giữa hiệu suất và chi phí tính toán, như đã được thể hiện trong thí nghiệm của chúng tôi với các cấu hình 4L-16F, 4L-32F, 5L-16F, và 5L-32F. Một ưu điểm khác của U-Net là khả năng hoạt động hiệu quả trên các tập dữ liệu nhỏ, một đặc điểm phổ biến trong lĩnh vực y tế do khó khăn trong việc thu thập và gắn nhãn dữ liệu. Bằng cách sử dụng các kỹ thuật tăng cường dữ liệu (data augmentation) như xoay, lật, hoặc thay đổi cường độ, U-Net có thể học được các đặc trưng mạnh mẽ ngay cả khi số lượng mẫu huấn luyện hạn chế. Trong nghiên cứu này, U-Net đã chứng minh khả năng phân đoạn chính xác các vùng khối u trên dữ liệu MRI T1ce và FLAIR, đặc biệt ở cấu hình 5L-32F, nơi mô hình đạt điểm Dice trung bình cao nhất (0.3887) nhờ dung lượng huấn luyện lớn (90.5 triệu tham số) và thời gian huấn luyện kéo dài (6.9 giờ). Tuy nhiên, U-Net cũng có những hạn chế cần được xem xét. Thứ nhất, mô hình có xu hướng gặp khó khăn ở các vùng có ranh giới mờ hoặc tín hiệu yếu, như vùng ET trong bài toán phân đoạn khối u não, điều này được quan sát rõ trong các cấu hình có số bộ lọc

hạn chế như 4L-16F và 5L-16F (điểm Dice cho ET lần lượt là 0.2632 và 0.2529). Thứ hai, U-Net truyền thống không tích hợp cơ chế chú ý (attention mechanism), dẫn đến việc mô hình có thể bỏ qua các vùng quan trọng hoặc bị nhiễu bởi các khu vực không liên quan trong ảnh. Cuối cùng, chi phí tính toán của U-Net, đặc biệt trong các cấu hình 3D như trong nghiên cứu này, có thể trở thành rào cản khi triển khai trên các thiết bị có tài nguyên hạn chế, như đã thấy với 5L-32F (90.5 triệu tham số, 6.9 giờ huấn luyện). Những hạn chế này mở ra các hướng cải tiến tiềm năng, như sẽ được thảo luận trong phần Hướng phát triển tương lai.

Nghiên cứu này đã đánh giá toàn diện hiệu suất của bốn cấu hình mô hình 3D U-Net (4L-16F, 4L-32F, 5L-16F, và 5L-32F) trong bài toán phân đoạn khối u não trên ảnh 3D MRI, tập trung vào các vùng WT, TC và ET. Kết quả định lượng từ Bảng 2 cho thấy cấu hình 5L-32F vượt trội với điểm Dice trung bình cao nhất (0.3887, trung bình của 0.4873, 0.3667, 0.3121), cải thiện đáng kể so với cấu hình cơ bản 4L-16F (0.3268, trung bình của 0.4159, 0.3014, 0.2632). Phân tích định tính qua Hình 6 củng cố phát hiện này, khi 5L-32F thể hiện khả năng tái tạo ranh giới WT, TC và ET sát nhất với ground truth, đặc biệt ở các vùng ET nhỏ và khó, nơi các cấu hình khác thường bỏ sót hoặc dự đoán sai. Về hiệu quả tính toán, Bảng 1 chỉ ra rằng 4L-16F là cấu hình nhẹ nhất (5.6 triệu tham số, 3.9 giờ huấn luyện), phù hợp cho các môi trường tài nguyên hạn chế, nhưng hiệu suất của nó bị giới hạn. Cấu hình 4L-32F (22.6 triệu tham số, 6.8 giờ) nổi bật như một giải pháp thực dụng, cân bằng giữa hiệu suất (điểm Dice trung bình 0.3385) và chi phí tính toán. Cấu hình 5L-16F (22.6 triệu tham số, 4.0 giờ) cho thấy độ sâu bổ sung không mang lại lợi ích đáng kể so với độ rộng, với điểm Dice trung bình chỉ đạt 0.3208. Cấu hình 5L-32F, dù đạt hiệu suất cao nhất, lại đòi hỏi tài nguyên lớn (90.5 triệu tham số, 6.9 giờ), khiến nó phù hợp hơn cho các ứng dụng ưu tiên độ chính xác mà không bị giới hạn về tài nguyên. Phân tích kiểu ablation cho thấy độ rộng bộ lọc (tăng từ 16 lên 32) mang lại cải thiện hiệu suất lớn hơn (3.6% từ 4L-16F lên 4L-32F) so với độ sâu (1.8% từ 4L-16F lên 5L-16F), nhưng với chi phí thời gian huấn luyện cao hơn (74% so với 2.6%). Các chỉ số bổ sung như khoảng cách Hausdorff (5 mm cho 5L-32F so với 8 mm cho 4L-16F) và biểu đồ phân bố điểm Dice cũng xác nhận rằng các mô hình lớn hơn, với dung lượng huấn luyện cao hơn, giảm thiểu các trường hợp dự đoán sai và cải thiện độ chính xác ranh giới. Nhìn chung, nghiên cứu này nhấn mạnh tầm quan trọng của việc cân nhắc giữa độ sâu và độ rộng của mô hình, cũng như sự đánh đổi giữa hiệu suất và chi phí tính toán, để lựa chọn cấu hình phù hợp với các yêu cầu ứng dụng cụ thể.

Mặc dù các cấu hình 3D U-Net đã cho thấy tiềm năng trong phân đoạn khối u não, vẫn còn nhiều hướng để cải thiện và mở rộng nghiên cứu:

- Tối ưu hóa chi phí tính toán: Cấu hình 5L-32F, dù có hiệu suất cao, lại yêu cầu tài nguyên lớn (90.5

triệu tham số, 6.9 giờ huấn luyện). Các kỹ thuật như cắt tỉa mô hình (model pruning), lượng tử hóa (quantization), hoặc sử dụng kiến trúc nhẹ hơn (ví dụ: EfficientNet hoặc MobileNet làm backbone) có thể được áp dụng để giảm số lượng tham số và thời gian suy luận, giúp mô hình khả thi hơn trong các ứng dụng thời gian thực, chẳng hạn như hỗ trợ chẩn đoán y tế tại bệnh viện.

- Tăng cường khả năng khái quát hóa: Hiệu suất của các mô hình hiện tại được đánh giá trên tập kiểm thử từ cùng nguồn dữ liệu (ví dụ: BraTS). Tuy nhiên, trong thực tế, dữ liệu MRI có thể khác biệt đáng kể giữa các máy quét hoặc cơ sở y tế. Các phương pháp như tăng cường dữ liệu (data augmentation), học chuyển giao (transfer learning), hoặc huấn luyện trên tập dữ liệu đa nguồn có thể cải thiện khả năng khái quát hóa của mô hình, đảm bảo hiệu suất ổn định trên các tập dữ liệu mới.
- Cải thiện phân đoạn vùng nhỏ và khó: Mặc dù 5L-32F cải thiện đáng kể ở vùng ET (0.3121 so với 0.2632 của 4L-16F), các cấu hình vẫn gặp khó khăn ở các vùng có ranh giới mờ hoặc tín hiệu yếu, như đã thấy trong Hình 6. Các kỹ thuật như sử dụng hàm mất mát trọng số (weighted loss) để ưu tiên các vùng khó (ET, TC), hoặc tích hợp các mô hình chú ý (attention mechanisms) như Attention U-Net, có thể giúp mô hình tập trung tốt hơn vào các khu vực này.
- Tích hợp đa phương thức: Hiện tại, mô hình sử dụng hai phương thức (FLAIR và T1ce). Việc tích hợp thêm các phương thức khác (như T1, T2) hoặc dữ liệu lâm sàng (ví dụ: thông tin bệnh nhân) có thể cung cấp thông tin bổ sung, cải thiện độ chính xác phân đoạn, đặc biệt ở các vùng có sự chồng lấn phức tạp giữa WT, TC và ET.
- Đánh giá trên dữ liệu thực tế và triển khai: Các mô hình cần được kiểm chứng trên dữ liệu thực tế từ bệnh viện để đánh giá hiệu quả trong môi trường lâm sàng. Ngoài ra, việc triển khai mô hình dưới dạng ứng dụng hoặc công cụ hỗ trợ bác sĩ (ví dụ: phần mềm phân tích MRI) sẽ yêu cầu tích hợp với hệ thống y tế, đảm bảo tính bảo mật dữ liệu và khả năng tương tác với các thiết bị y tế.
- Khám phá các kiến trúc tiên tiến: Ngoài U-Net, các kiến trúc hiện đại như nnU-Net (một biến thể tự động điều chỉnh của U-Net) hoặc các mô hình dựa trên Transformer (như TransUNet) có thể được thử nghiệm để so sánh hiệu suất. Những kiến trúc này có thể tận dụng các cơ chế chú ý toàn cục (global attention) để cải thiện khả năng phân đoạn các vùng phức tạp, đồng thời tối ưu hóa hiệu quả tính toán.
- Ứng dụng lâm sàng mở rộng: Ngoài phân đoạn khối u não, mô hình 3D U-Net có thể được mở rộng để áp dụng cho các bài toán khác trong y tế, như phân đoạn khối u gan, phổi, hoặc các tổn thương thần kinh khác. Việc này sẽ yêu cầu điều chỉnh mô hình

để phù hợp với đặc điểm của từng loại dữ liệu và yêu cầu lâm sàng cụ thể.

Những hướng phát triển này không chỉ cải thiện hiệu suất và tính thực tiễn của mô hình 3D U-Net mà còn mở ra tiềm năng ứng dụng rộng rãi trong chẩn đoán và điều trị khối u não, góp phần nâng cao chất lượng chăm sóc y tế.

Tham khảo

Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R. T.; Berger, C.; Ha, S. M.; Rozycki, M.; Prastawa, M.; Alberts, E.; Lipkova, J.; Freymann, J.; Kirby, J.; Bilello, M.; Fathallah-Shaykh, H.; Wiest, R.; Kirschke, J.; Wiestler, B.; Colen, R.; Kotrotsou, A.; Lamontagne, P.; Marcus, D.; Milchenko, M.; Nazeri, A.; Weber, M.-A.; Mahajan, A.; Baid, U.; Gerstner, E.; Kwon, D.; Acharya, G.; Agarwal, M.; Alam, M.; Albiol, A.; Albiol, A.; Albiol, F. J.; Alex, V.; Allinson, N.; Amorim, P. H. A.; Amrutkar, A.; Anand, G.; Andermatt, S.; Arbel, T.; Arbelaez, P.; Avery, A.; Azmat, M.; B, P.; Bai, W.; Banerjee, S.; Barth, B.; Batchelder, T.; Batmanghelich, K.; Battistella, E.; Beers, A.; Belyaev, M.; Bendszus, M.; Benson, E.; Bernal, J.; Bharath, H. N.; Biros, G.; Bisdas, S.; Brown, J.; Cabezas, M.; Cao, S.; Cardoso, J. M.; Carver, E. N.; Casamitjana, A.; Castillo, L. S.; Catà, M.; Cattin, P.; Cerigues, A.; Chagas, V. S.; Chandra, S.; Chang, Y.-J.; Chang, S.; Chang, K.; Chazalon, J.; Chen, S.; Chen, W.; Chen, J. W.; Chen, Z.; Cheng, K.; Choudhury, A. R.; Chylla, R.; Clérigues, A.; Coleman, S.; Colmeiro, R. G. R.; Combalia, M.; Costa, A.; Cui, X.; Dai, Z.; Dai, L.; Daza, L. A.; Deutsch, E.; Ding, C.; Dong, C.; Dong, S.; Dudzik, W.; Eaton-Rosen, Z.; Egan, G.; Escudero, G.; Estienne, T.; Everson, R.; Fabrizio, J.; Fan, Y.; Fang, L.; Feng, X.; Ferrante, E.; Fidon, L.; Fischer, M.; French, A. P.; Fridman, N.; Fu, H.; Fuentes, D.; Gao, Y.; Gates, E.; Gerling, D.; Gholami, A.; Gierke, W.; Glocker, B.; Gong, M.; González-Villá, S.; Grosge, T.; Guan, Y.; Guo, S.; Gupta, S.; Han, W.-S.; Han, I. S.; Harmuth, K.; He, H.; Hernández-Sabaté, A.; Herrmann, E.; Himthani, N.; Hsu, W.; Hsu, C.; Hu, X.; Hu, X.; Hu, Y.; Hu, Y.; Hua, R.; Huang, T.-Y.; Huang, W.; Huffel, S. V.; Huo, Q.; HV, V.; Iftekharruddin, K. M.; Isensee, F.; Islam, M.; Jackson, A. S.; Jambawalikar, S. R.; Jesson, A.; Jian, W.; Jin, P.; Jose, V. J. M.; Jungo, A.; Kainz, B.; Kamnitsas, K.; Kao, P.-Y.; Karnawat, A.; Kellermeier, T.; Kermi, A.; Keutzer, K.; Khadir, M. T.; Khened, M.; Kickingeder, P.; Kim, G.; King, N.; Knapp, H.; Knecht, U.; Kohli, L.; Kong, D.; Kong, X.; Koppers, S.; Kori, A.; Krishnamurthi, G.; Krivov, E.; Kumar, P.; Kushibar, K.; Lachinov, D.; Lambrou, T.; Lee, J.; Lee, C.; Lee, Y.; Lee, M.; Lefkovits, S.; Lefkovits, L.; Levitt, J.; Li, T.; Li, H.; Li, W.; Li, H.; Li, X.; Li, Y.; Li, H.; Li, Z.; Li, X.; Li, Z.; Li, X.; Li, W.; Lin, Z.-S.; Lin, F.; Lio, P.; Liu, C.; Liu, B.; Liu, X.; Liu, M.; Liu, J.; Liu, L.; Llado, X.; Lopez, M. M.; Lorenzo, P. R.; Lu, Z.; Luo, L.; Luo, Z.; Ma, J.; Ma, K.;

Mackie, T.; Madabushi, A.; Mahmoudi, I.; Maier-Hein, K. H.; Maji, P.; Mammen, C. P.; Mang, A.; Manjunath, B. S.; Marcinkiewicz, M.; McDonagh, S.; McKenna, S.; McKinley, R.; Mehl, M.; Mehta, S.; Mehta, R.; Meier, R.; Meinel, C.; Merhof, D.; Meyer, C.; Miller, R.; Mitra, S.; Moiyadi, A.; Molina-Garcia, D.; Monteiro, M. A. B.; Mrukwa, G.; Myronenko, A.; Nalepa, J.; Ngo, T.; Nie, D.; Ning, H.; Niu, C.; Nuechterlein, N. K.; Oermann, E.; Oliveira, A.; Oliveira, D. D. C.; Oliver, A.; Osman, A. F. I.; Ou, Y.-N.; Ourselin, S.; Paragios, N.; Park, M. S.; Paschke, B.; Pauloski, J. G.; Pawar, K.; Pawlowski, N.; Pei, L.; Peng, S.; Pereira, S. M.; Perez-Beteta, J.; Perez-Garcia, V. M.; Pezold, S.; Pham, B.; Phophalia, A.; Piella, G.; Pillai, G. N.; Piraud, M.; Pisov, M.; Popli, A.; Pound, M. P.; Pourreza, R.; Prasanna, P.; Prkowska, V.; Pridmore, T. P.; Puch, S.; Puybareau, S.; Qian, B.; Qiao, X.; Rajchl, M.; Rane, S.; Rebsamen, M.; Ren, H.; Ren, X.; Revanuru, K.; Rezaei, M.; Rippel, O.; Rivera, L. C.; Robert, C.; Rosen, B.; Rueckert, D.; Safwan, M.; Salem, M.; Salvi, J.; Sanchez, I.; Sánchez, I.; Santos, H. M.; Sartor, E.; Schellingerhout, D.; Scheufele, K.; Scott, M. R.; Scussel, A. A.; Sedlar, S.; Serrano-Rubio, J. P.; Shah, N. J.; Shah, N.; Shaikh, M.; Shankar, B. U.; Shboul, Z.; Shen, H.; Shen, D.; Shen, L.; Shen, H.; Shenoy, V.; Shi, F.; Shin, H. E.; Shu, H.; Sima, D.; Sinclair, M.; Smedby, O.; Snyder, J. M.; Soltaninejad, M.; Song, G.; Soni, M.; Stawiaski, J.; Subramanian, S.; Sun, L.; Sun, R.; Sun, J.; Sun, K.; Sun, Y.; Sun, G.; Sun, S.; Suter, Y. R.; Szilagyi, L.; Talbar, S.; Tao, D.; Tao, D.; Teng, Z.; Thakur, S.; Thakur, M. H.; Tharakan, S.; Tiwari, P.; Tochon, G.; Tran, T.; Tsai, Y. M.; Tseng, K.-L.; Tuan, T. A.; Turlapov, V.; Tustison, N.; Vakalopoulou, M.; Valverde, S.; Vanguri, R.; Vasiliev, E.; Ventura, J.; Vera, L.; Vercauteren, T.; Verrastro, C. A.; Vidyaratne, L.; Vilaplana, V.; Vivekanandan, A.; Wang, G.; Wang, Q.; Wang, C. J.; Wang, W.; Wang, D.; Wang, R.; Wang, Y.; Wang, C.; Wang, G.; Wen, N.; Wen, X.; Weninger, L.; Wick, W.; Wu, S.; Wu, Q.; Wu, Y.; Xia, Y.; Xu, Y.; Xu, X.; Xu, P.; Yang, T.-L.; Yang, X.; Yang, H.-Y.; Yang, J.; Yang, H.; Yang, G.; Yao, H.; Ye, X.; Yin, C.; Young-Moxon, B.; Yu, J.; Yue, X.; Zhang, S.; Zhang, A.; Zhang, K.; Zhang, X.; Zhang, L.; Zhang, X.; Zhang, Y.; Zhang, L.; Zhang, J.; Zhang, X.; Zhang, T.; Zhao, S.; Zhao, Y.; Zhao, X.; Zhao, L.; Zheng, Y.; Zhong, L.; Zhou, C.; Zhou, X.; Zhou, F.; Zhu, H.; Zhu, J.; Zhuge, Y.; Zong, W.; Kalpathy-Cramer, J.; Farahani, K.; Davatzikos, C.; Leemput, K. v.; and Menze, B. 2019. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. ArXiv:1811.02629 [cs].

Breiman, L. 2001. Random Forests. *Machine Learning*, 45(1): 5–32.

Cung, V. C. 2023. NGHIÊN CỨU VAI TRÒ CỦA MỘT SỐ CHUỖI XUNG CÓ SỬ DỤNG ĐỐI QUANG TỬ TRONG CHẨN ĐOÁN DI CĂN NÃO TRÊN BỆNH

NHÂN UNG THƯ PHỔI NGUYÊN PHÁT ĐIỀU TRỊ TẠI BỆNH VIỆN PHỔI TRUNG ƯƠNG. *Tạp chí Y học Việt Nam*, 527(1B).

Gou, F.; Liu, J.; Xiao, C.; and Wu, J. 2024. Research on Artificial-Intelligence-Assisted Medicine: A Survey on Medical Artificial Intelligence. *Diagnostics*, 14(14): 1472. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.; and Xu, D. 2022. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. ArXiv:2201.01266 [eess].

Hojjat, S.; and Kittler, J. 1998. Region Growing: A New Approach. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 7: 1079–84.

House, B. 2019. 2012: A Breakthrough Year for Deep Learning.

Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P. F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; and Maier-Hein, K. H. 2018. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. ArXiv:1809.10486 [cs].

Kaifi, R. 2023. A Review of Recent Advances in Brain Tumor Diagnosis Based on AI-Based Classification. *Diagnostics*, 13(18): 3007.

Konishi, S.; Yuille, A.; Coughlan, J.; and Song Chun Zhu. 2003. Statistical edge detection: learning and evaluating edge cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1): 57–74.

Liu, S. C.; and Xu, T. Y. 2024. Research on Brain Tumor Segmentation and Identification Technology Based on Improved Unet. In *2024 36th Chinese Control and Decision Conference (CCDC)*, 5161–5168. Xi'an, China: IEEE. ISBN 979-8-3503-8778-0.

Liu, Z.; Tong, L.; Chen, L.; Jiang, Z.; Zhou, F.; Zhang, Q.; Zhang, X.; Jin, Y.; and Zhou, H. 2023. Deep learning based brain tumor segmentation: a survey. *Complex & Intelligent Systems*, 9(1): 1001–1026.

Menze, B. H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; Lanczi, L.; Gerstner, E.; Weber, M.-A.; Arbel, T.; Avants, B. B.; Ayache, N.; Buendia, P.; Collins, D. L.; Cordier, N.; Corso, J. J.; Criminisi, A.; Das, T.; Delingette, H.; Demiralp, C.; Durst, C. R.; Dojat, M.; Doyle, S.; Festa, J.; Forbes, F.; Geremia, E.; Glocker, B.; Golland, P.; Guo, X.; Hamamci, A.; Iftekharuddin, K. M.; Jena, R.; John, N. M.; Konukoglu, E.; Lashkari, D.; Mariz, J. A.; Meier, R.; Pereira, S.; Precup, D.; Price, S. J.; Raviv, T. R.; Reza, S. M. S.; Ryan, M.; Sarikaya, D.; Schwartz, L.; Shin, H.-C.; Shotton, J.; Silva, C. A.; Sousa, N.; Subbanna, N. K.; Szekely, G.; Taylor, T. J.; Thomas, O. M.; Tustison, N. J.; Unal, G.; Vasseur, F.; Wintermark, M.; Ye, D. H.; Zhao, L.; Zhao, B.; Zikic, D.; Prastawa, M.; Reyes, M.; and Van Leemput,

- K. 2015. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10): 1993–2024.
- O’Shea, K.; and Nash, R. 2015. An Introduction to Convolutional Neural Networks. ArXiv:1511.08458 [cs].
- Pichaivel, M.; Anbumani, G.; Theivendren, P.; and Gopal, M. 2022. An Overview of Brain Tumor. In Agrawal, A., ed., *Brain Tumors*. IntechOpen. ISBN 978-1-83969-803-3 978-1-83969-804-0.
- Ragba, C. E.; Boulaalam, A.; and Bellamine, I. 2024. Survey on the Segmentation of Brain Tumors Using Deep Learning Techniques. In *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 1–6. ISSN: 2769-9994.
- Rizwan I Haque, I.; and Neubert, J. 2020. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18: 100297.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv:1505.04597 [cs].
- Sahoo, P.; Soltani, S.; and Wong, A. 1988. A Survey of Thresholding Techniques. *Computer Vision, Graphics, and Image Processing*, 41: 233–260.
- Shahin, T. M. M. 2008. Feature-Based Design – An Overview. *Computer-Aided Design and Applications*, 5(5): 639–653.
- Theodoros, E. I.; and Massimiliano, P. 2001. (PDF) Support Vector Machines: Theory and Applications. *ResearchGate*.
- Tirivangani, M.; and Serestina, V. 2021. (PDF) Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art. *ResearchGate*.
- Xu, G.; Wang, X.; Wu, X.; Leng, X.; and Xu, Y. 2024. Development of Skip Connection in Deep Neural Networks for Computer Vision and Medical Image Analysis: A Survey. ArXiv:2405.01725 [eess].
- Xu, Y.; and Goodacre, R. 2018. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3): 249–262.