

KHẢO SÁT ẢNH HƯỞNG CỦA HÀM MẤT MÁT ĐẾN HIỆU QUẢ PHÂN LỚP TRÊN TẬP DỮ LIỆU MẤT CÂN BẰNG

Phạm Thanh An^{1,2}, Hoàng Văn Dũng³, Lê Văn Tường Lân^{4*}

¹Khoa CNTT -Trường Đại học Khoa học, Đại học Huế

²Trường Đại học Ngân hàng Tp.HCM

³Khoa Công nghệ thông tin - Trường Đại học Công nghệ Kỹ thuật Tp.HCM

⁴Ban Đào tạo và Công tác sinh viên, Đại học Huế

*Email: lvtlan@hueuni.edu.vn

Ngày nhận bài: 14/6/2026; ngày hoàn thành phản biện: 30/6/2026; ngày duyệt đăng: 30/6/2026

TÓM TẮT

Hàm mất mát cross-entropy đang được sử dụng rộng rãi như một tiêu chuẩn cho bài toán phân lớp. Hàm mất mát này có xu hướng ưu tiên các lớp chiếm ưu thế về số lượng, dẫn đến sự thiên lệch trong quá trình học khi tập dữ liệu bị mất cân bằng. Bài báo này trình bày một khảo sát và đánh giá toàn diện ảnh hưởng của ba hàm mất mát tiêu biểu bao gồm Cross-Entropy (CE), Weighted Cross-Entropy (WCE), và Focal Loss (FL) đến hiệu quả phân lớp tập dữ liệu mất cân bằng. 6 mô hình tiêu biểu đã được huấn luyện trước được chọn để thực nghiệm. Tập dữ liệu thực nghiệm gồm 3.064 ảnh cộng hưởng từ bệnh u não, được phân thành 3 lớp bệnh lý. Tập dữ liệu được chia ngẫu nhiên theo tỷ lệ 80:10:10 theo bệnh nhân, tương ứng với tập huấn luyện, tập xác thực và tập kiểm thử. Kết quả thực nghiệm cho thấy hiệu quả phân lớp bệnh u não khi áp dụng hàm mất mát FL đạt kết quả tốt nhất trên 5 trong số 6 mô hình khảo sát. Hàm WCE đạt hiệu quả kém hơn nhiều so với khi sử dụng hàm mất mát FL.

Từ khóa: Deep learning, U não, Hàm mất mát, Cross-Entropy, Focal Loss.

1. GIỚI THIỆU

Trong lĩnh vực học sâu và thị giác máy tính, hàm mất mát là thành phần cốt lõi quyết định hướng tối ưu của quá trình học và ảnh hưởng trực tiếp đến hiệu năng của mô hình. Các kiến trúc học sâu khác nhau, bao gồm mạng nơ-ron tích chập (CNN) và Vision Transformer, thường áp dụng các hàm mất mát chuyên biệt được xây dựng tương

thích với cơ chế xử lý thông tin và mục tiêu tối ưu của từng mô hình. Việc lựa chọn hàm mất mát phù hợp đóng vai trò quan trọng trong việc cải thiện khả năng học đặc trưng, tăng độ chính xác dự đoán và nâng cao chất lượng của dữ liệu hoặc hình ảnh được tạo sinh [1].

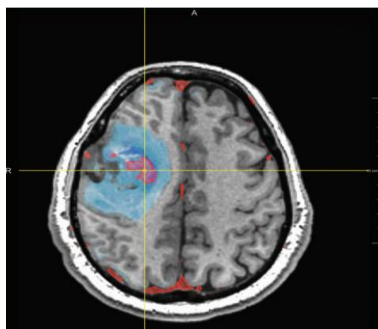
Trong các mô hình học sâu, việc lựa chọn hàm mất mát đóng vai trò quyết định đối với quá trình tối ưu hóa và chất lượng dự đoán của mô hình. Hàm mất mát không chỉ định lượng mức độ sai khác giữa đầu ra dự đoán và giá trị mục tiêu mà còn định hướng quá trình cập nhật tham số mô hình. Các hàm mất mát khác nhau được xây dựng để giải quyết những thách thức riêng biệt, chẳng hạn như tăng độ nhạy đối với các mẫu khó phân loại, xử lý dữ liệu mất cân bằng hoặc giảm ảnh hưởng của các giá trị ngoại lệ. Do đó, việc lựa chọn và thiết kế hàm mất mát phù hợp là điều kiện tiên quyết để phát triển các mô hình có khả năng học hiệu quả, khái quát hóa tốt và duy trì hiệu năng ổn định trên dữ liệu thực tế. Xu hướng nghiên cứu hiện nay còn tập trung vào việc kết hợp nhiều hàm mất mát trong một khuôn khổ huấn luyện thống nhất nhằm khai thác đồng thời các ưu điểm của từng hàm, qua đó cải thiện độ chính xác và tính mạnh mẽ của mô hình [1].

Trong các nhiệm vụ phân loại hình ảnh y tế với phân bố dữ liệu mất cân bằng, việc giảm thiểu hàm mất mát cross-entropy có thể tạo ra sự thiên lệch trong quá trình học. Điều này xuất phát từ hai nguyên nhân chính: (i) hàm mất mát gán mức đóng góp tương đương cho mọi lớp mà không xét đến sự khác biệt về tần suất xuất hiện; và (ii) quá trình tối ưu hóa bị chi phối bởi lớp đa số, khiến mô hình có xu hướng phân loại phần lớn các mẫu đầu vào vào lớp bình thường chiếm ưu thế [2].

Trong những năm gần đây, hình ảnh y tế số đã trở nên ngày càng phổ biến và đóng vai trò thiết yếu trong việc phát hiện, chẩn đoán nhiều loại bệnh lý. Ngoài ứng dụng lâm sàng, hình ảnh y tế còn được sử dụng rộng rãi trong đào tạo và nghiên cứu. Các hệ thống chẩn đoán hỗ trợ bằng máy tính (Computer-Aided Diagnosis - CAD) phát triển dựa trên công nghệ trí tuệ nhân tạo (AI), đặc biệt là học sâu (deep learning), đã góp phần nâng cao độ chính xác và rút ngắn thời gian chẩn đoán cho các bác sĩ chẩn đoán hình ảnh [3].

U não là một trong những dạng ung thư nguy hiểm nhất với tỷ lệ tử vong cao. Bệnh xuất hiện ở cả người lớn và trẻ em, đồng thời bao gồm nhiều loại khác nhau dựa trên vị trí, cấu trúc và đặc điểm hình thái của khối u. Việc phân loại chính xác u não là yêu cầu thiết yếu trong thực hành lâm sàng, vì những sai sót trong chẩn đoán có thể dẫn đến các quyết định điều trị không phù hợp và gây ra những hậu quả nghiêm trọng đối với bệnh nhân [4]. Trong số các khối u não, những dạng phổ biến nhất bao gồm u di căn nội sọ từ các bệnh ung thư hệ thống, u màng não và u tế bào thần kinh đệm, đặc biệt là u nguyên bào thần kinh đệm. Các tổn thương di căn trong hệ thần kinh trung ương có thể phát triển tại nhiều vị trí khác nhau dọc theo trục thần kinh, dẫn đến sự đa dạng về

biểu hiện lâm sàng và mức độ tiến triển của bệnh. Vì vậy, việc điều trị các trường hợp này thường yêu cầu cách tiếp cận đa chuyên khoa với sự tham gia của các lĩnh vực phẫu thuật thần kinh, xạ trị và điều trị nội khoa ung thư nhằm nâng cao hiệu quả điều trị và cải thiện tiên lượng cho bệnh nhân [5].



Hình 1. Hình ảnh khối u não [6]

Trong lĩnh vực chẩn đoán và điều trị u não, hình ảnh học y tế đóng vai trò then chốt. Những tiến bộ trong công nghệ hình ảnh liên tục được nghiên cứu và cải tiến nhằm nâng cao hiệu quả hỗ trợ bác sĩ trong việc chẩn đoán và chăm sóc bệnh nhân. Các kỹ thuật hình ảnh y học như chụp cắt lớp vi tính (CT), chụp cắt lớp phát xạ positron (PET), và cộng hưởng từ (MRI) thường được sử dụng để phát hiện và phân loại u não. Khác với hình ảnh thông thường, hình ảnh y tế phản ánh sự phân bố của các đặc tính vật lý trong cơ thể con người, giúp làm rõ các đặc điểm mà các phương pháp không xâm lấn khác không thể phát hiện [7], [8].

Quá trình chẩn đoán khối u não không phải lúc nào cũng rõ ràng và có thể gặp nhiều khó khăn do sự chồng lấp về đặc điểm hình ảnh và mô bệnh học giữa các khối u và các bệnh lý thần kinh không phải u. Khối u não được phân loại thành lành tính hoặc ác tính dựa trên các yếu tố như vị trí giải phẫu, giai đoạn tiến triển, đặc điểm mô học và tốc độ phát triển của khối u. Khối u lành tính thường có tốc độ tăng trưởng chậm, ranh giới rõ ràng và hiếm khi xâm lấn mô lân cận; các ví dụ điển hình bao gồm u tuyến yên và u màng não. Ngược lại, khối u ác tính có xu hướng phát triển nhanh, xâm lấn mô não hoặc tủy sống lân cận, và thường không có ranh giới rõ ràng, điển hình là u thần kinh đệm [9].

2. CÁC NGHIÊN CỨU LIÊN QUAN

Học sâu, đặc biệt là Mạng nơ-ron tích chập (CNN) [10], [11] là một phương pháp rất phổ biến để phân lớp và phân đoạn khối u não trong hình ảnh y tế như MRI. Điều này có thể giúp bác sĩ chẩn đoán bệnh nhân chính xác và nhanh hơn. Sự phát triển nhanh chóng của công nghệ học sâu (Deep Learning - DL) đã tạo ra những ảnh hưởng sâu rộng đối với lĩnh vực phân tích hình ảnh y khoa. Nhiều nghiên cứu đã tập trung vào việc đề

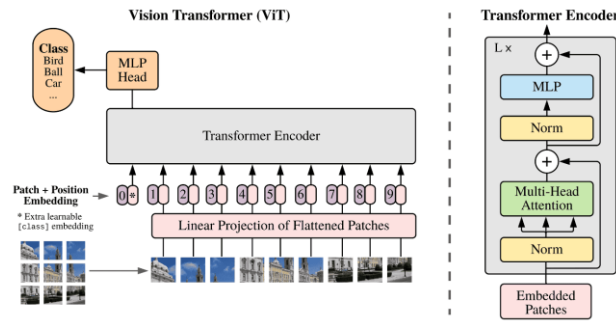
xuất các phương pháp tiên tiến dựa trên DL nhằm nâng cao độ chính xác trong chẩn đoán, cũng như hỗ trợ hiệu quả cho công tác sàng lọc.

Saleh và cộng sự đã tiến hành huấn luyện và đánh giá bộ dữ liệu u não bằng 5 [12] mô hình học sâu được huấn luyện trước gồm Xception, ResNet50, InceptionV3, VGG16 và MobileNet. Kết quả thực nghiệm trên tập kiểm tra độc lập cho thấy các mô hình đạt chỉ số F1 lần lượt là 98,75%, 98,50%, 98,00%, 97,50% và 97,25%. Hiệu suất phân loại cao của các mô hình này cho thấy tiềm năng hỗ trợ phát hiện sớm khối u não với độ tin cậy cao, từ đó góp phần nâng cao hiệu quả chẩn đoán và can thiệp điều trị trước khi bệnh tiến triển và gây ra các biến chứng nghiêm trọng về thể chất.

Nassar và cộng sự [13] đề xuất một mô hình lai dựa trên chiến lược bỏ phiếu đa số được đề xuất nhằm thực hiện phân loại ba loại u não phổ biến gồm u màng não, u tế bào thần kinh đệm và u tuyến yên. Nghiên cứu sử dụng năm kiến trúc CNN gồm GoogleNet, AlexNet, ShuffleNet, SqueezeNet và NASNet-Mobile. Mỗi mô hình được tinh chỉnh trên dữ liệu u não để khai thác các đặc trưng biểu diễn khác nhau của hình ảnh cộng hưởng từ. Quyết định phân loại cuối cùng được đưa ra thông qua việc tổng hợp kết quả dự đoán của năm mô hình, với mục tiêu nâng cao độ chính xác, độ tin cậy và tính nhất quán của hệ thống phân loại. Phương pháp đề xuất đạt độ chính xác 99,31%, góp phần hỗ trợ các bác sĩ chẩn đoán hình ảnh trong việc đưa ra nhận định lâm sàng chính xác hơn.

Cơ chế chú ý [14] là một tiến bộ nền tảng trong lĩnh vực trí tuệ nhân tạo và học máy, góp phần định hình lại hiệu quả và khả năng biểu diễn của các mô hình học sâu. Được lấy cảm hứng từ cơ chế tập trung có chọn lọc của hệ thần kinh con người, cơ chế này đã trở thành một thành phần cốt lõi trong nhiều ứng dụng hiện đại, đặc biệt là trong xử lý ngôn ngữ tự nhiên, thị giác máy tính và các lĩnh vực liên quan khác. Cơ chế chú ý nhằm xác định và lượng hóa mức độ quan trọng tương đối của các thành phần khác nhau trong chuỗi đầu vào đối với ngữ cảnh tổng thể [14].

Trong những năm gần đây, việc dịch chuyển các kiến trúc dựa trên cơ chế tự chú ý (Self-Attention) sang lĩnh vực thị giác máy tính đã thiết lập những cột mốc đột phá mới. Vision Transformer (ViT) [15] đã đánh dấu một bước chuyển dịch mang tính bước ngoặt trong lĩnh vực thị giác máy tính, thách thức vị thế thống trị lâu năm của mạng nơ-ron tích chập. Thay vì sử dụng các bộ lọc trích xuất cục bộ, ViT tiếp cận dữ liệu hình ảnh bằng cách phân tách một bức ảnh hai chiều thành một chuỗi các phân mảnh độc lập không chồng lấp (gọi là các bản vá - *patches*), tương tự như cấu trúc các mã báo (*tokens*) trong xử lý ngôn ngữ tự nhiên. Các bản vá này sau đó được làm phẳng, nhúng tuyến tính cùng với mã vị trí (*position embeddings*) và đưa vào các khối Transformer Encoder tiêu chuẩn để tính toán cơ chế tự chú ý toàn cục. Mặc dù đạt được nhiều kết quả, ViT bị giới hạn bởi độ phức tạp bậc 2, $O(N^2)$, gây khó khăn trong việc xử lý ảnh có kích thước lớn.



Hình 2. Kiến trúc ViT [15]

Một số mô hình mới được đề xuất, nhằm giải quyết hạn chế của ViT. Trong số đó, Swin Transformer [16] và MaxViT [17] là hai đại diện tiên phong, giải quyết xuất sắc bài toán tối ưu hóa chi phí tính toán và nâng cao khả năng khai thác đặc trưng không gian trên các dữ liệu ảnh y tế phức tạp.

Swin Transformer đã khắc phục giới hạn về chi phí tính toán bậc hai của mô hình ViT chuẩn bằng cách giới thiệu cơ chế chú ý cục bộ trong các cửa sổ không gian cố định, kết hợp cùng kỹ thuật dịch chuyển cửa sổ (Shifted Windows) ở các tầng liên tiếp để duy trì sự giao thoa thông tin toàn cục. Mặt khác, MaxViT (Multi-Axis Vision Transformer) lại đại diện cho xu hướng kiến trúc Hybrid toàn diện khi tích hợp các khối chú ý đa trục (Multi-Axis Self-Attention - Max-SA) song song với các tầng tích chập MBConv của CNN. Bằng cách phân tách quá trình tính toán thành Block Attention (chú ý cục bộ vùng lân cận) và Grid Attention (chú ý toàn cục thưa thớt dạng lưới), MaxViT sở hữu năng lực vượt trội trong việc vừa trích xuất các đặc trưng hình học mịn ở vùng biên tổn thương, vừa thấu hiểu bối cảnh giải phẫu tổng thể của các cấu trúc não bộ.

Tummala và cộng sự (2022) [18] sử dụng mô hình ViT được huấn luyện trước trên ImageNet (B/16, B/32, L/16 và L/32) và tinh chỉnh cho nhiệm vụ phân loại 3064 lát cắt MRI tăng cường độ tương phản với u màng não, u tế bào thần kinh đệm và u tuyến yên. độ chính xác kiểm tra tổng thể là 98,2% ở độ phân giải 384×384 đối với L/32.

Hàm mất mát Cross-Entropy tiêu chuẩn [19] tối ưu hóa phân phối xác suất dựa trên lý thuyết thông tin nhưng gặp hạn chế lớn khi dữ liệu mất cân bằng. Để khắc phục, Weighted Cross-Entropy (WCE) [20] giới thiệu cơ chế gán trọng số tĩnh theo lớp nhằm nâng cao hình phạt cho các lớp thiểu số. Gần đây, năm 2017 Lin và các cộng sự [21] đã đề xuất Focal Loss, một bước đột phá giúp điều biến động hình phạt theo độ khó của từng mẫu dữ liệu, mở ra hướng đi mới cho việc tối ưu các mô hình học sâu.

Hàm mất mát cross-entropy (CE) là một trong những loại hàm mất mát được sử dụng rộng rãi nhất để huấn luyện mạng nơ-ron cho bài toán phân lớp ảnh.

$$CE = - \sum_{i=1}^C y_i * \log(p_i) \quad (1)$$

$$WCE = - \sum_{i=1}^C w_i * y_i * \log(p_i) \quad (2)$$

$$FL = - \sum_{i=1}^C \alpha_i * (1 - p_i)^{\gamma} * y_i * \log(p_i) \quad (3)$$

C: Số lớp tập dữ liệu

p_i : xác suất mô hình dự đoán cho lớp thứ i

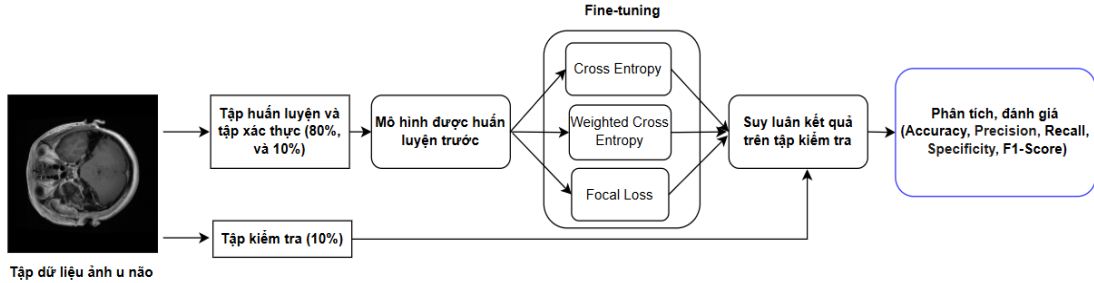
w_i trong hàm WCE: Là trọng số tĩnh được gán cho lớp thứ i. Tham số w_i được tính toán trước khi huấn luyện theo công thức nghịch đảo tần suất mẫu, $w_i = \frac{N_{total}}{C * N_i}$. Theo đó, lớp nào có số lượng mẫu N_i nhỏ, trọng số w_i sẽ rất lớn.

α_i là hệ số cân bằng lớp trong hàm FL: Tương tự như trong hàm WCE, đây là trọng số tĩnh gán cho lớp i nhằm kiểm soát sự mất cân bằng số lượng mẫu giữa các lớp. Giá trị thường nằm trong khoảng [0, 1] và tổng các α_i của các lớp bằng 1.

γ là hệ số tập trung trong hàm FL: là siêu tham số điều khiển tốc độ triệt tiêu lỗi của các mẫu dễ, theo các tác giả đề xuất tối ưu khi $\gamma = 2$.

3. PHƯƠNG PHÁP VÀ DỮ LIỆU THỰC NGHIỆM

Chúng tôi tiến hành huấn luyện tinh chỉnh 6 mô hình bao gồm ResNet-50 [22], DenseNet-201 [23], ConvNeXt-T [24], ViT/B-16, SWin-T, MaxViT-T. Các mô hình được huấn luyện trước trên tập ImageNet_1K. Chi tiết phương pháp thực nghiệm và đánh giá được mô tả ở hình 3.



Hình 3. Sơ đồ phương pháp đề xuất

Bộ dữ liệu Figshare [25] được sử dụng trong nghiên cứu này được xây dựng từ các ảnh cộng hưởng từ (MRI) não thu thập trong quá trình chẩn đoán u não tại Bệnh viện Nam Phương và Bệnh viện Đa khoa Đại học Y Thiên Tân (Trung Quốc) trong giai đoạn 2005–2010. Bộ dữ liệu bao gồm 3.064 ảnh MRI não tăng cường tương phản T1 của 233 bệnh nhân, kích thước 512×512. Dữ liệu được công bố lần đầu vào năm 2015 và đã được cập nhật đến phiên bản 8 vào năm 2024; phiên bản mới nhất này được lựa chọn cho nghiên cứu hiện tại. Bộ dữ liệu gồm ba loại khối u não khác nhau: u màng não với 708 ảnh, u tế bào thần kinh đệm với 1.426 ảnh và u tuyến yên với 930 ảnh. Các ảnh được ghi nhận ở ba mặt phẳng chụp khác nhau, bao gồm mặt phẳng dọc (1.025 ảnh), mặt

phẳng ngang (994 ảnh) và mặt phẳng đứng (1.045 ảnh). Một số hình ảnh minh họa đại diện cho từng loại u và từng mặt phẳng chụp được trình bày trong Hình 4.

Tập dữ liệu được phân chia ngẫu nhiên theo bệnh nhân, sử dụng random seed = 42 nhằm đảm bảo khả năng tái lập, thành tập huấn luyện, tập xác thực và tập kiểm tra độc lập, với tỷ lệ 80:10:10 tương ứng, đảm bảo mẫu của cùng bệnh nhân chỉ nằm trong tập duy nhất. Chi tiết phân bố tập dữ liệu được thể hiện ở Bảng 1, Bảng 2.

Trọng số lớp w_i và α_i trong hàm Weighted Cross Entropy, Focal Loss được tính theo công thức (4) và (5) như sau:

$$w_i = \frac{N}{C \times N_i} \quad (4)$$

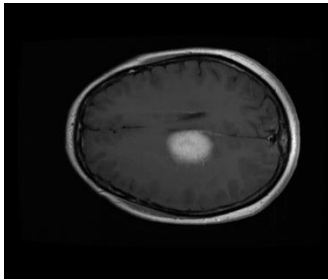
$$\alpha_i = \frac{1/N_i}{\sum_{j=1}^C 1/N_j} \quad (5)$$

Trong đó:

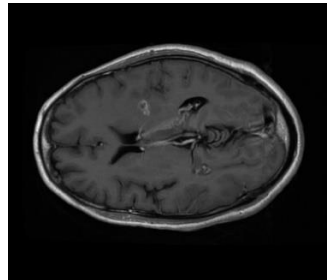
N là tổng số lượng mẫu của toàn bộ tập huấn luyện

C là số lớp

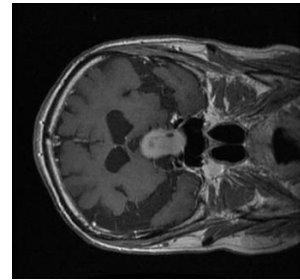
N_i là số mẫu lớp i



(a) U màng não
(Meningioma)



(b) U thần kinh đệm (Glioma)



(c) U tuyến yên (Pituitary tumor)

Hình 4. Hình ảnh tập dữ liệu u não MRI

Bảng 1. Phân bố tập dữ liệu U não

Lớp bệnh lý	Tổng số hình ảnh	Tập huấn luyện	Tập xác thực	Tập kiểm tra
U thần kinh đệm	1426	1168	129	129
U màng não	708	535	83	90
U tuyến yên	930	750	98	82
Tổng số (N)	3064	2453	310	301

Bảng 2. Chi tiết phân chia tập dữ liệu theo bệnh nhân

Lớp bệnh lý	Tổng số bệnh nhân	Tập huấn luyện	Tập xác thực	Tập kiểm tra
U thần kinh đệm	89	71	10	8
U màng não	82	65	7	10
U tuyến yên	62	50	6	6
Tổng số bệnh nhân	233	186	23	24

4. KẾT QUẢ THỰC NGHIỆM VÀ THẢO LUẬN

Trong nghiên cứu này, các độ đo được tính toán từ ma trận nhầm lẫn. Với lớp thứ i , các giá trị tương ứng: dương tính thực sự (TP_i), âm tính thực sự (TN_i), dương tính giả (FP_i), và âm tính giả (FN_i). Các độ đo được định nghĩa cụ thể như sau:

$$\text{Accuracy (ACC)} = \frac{\sum_{i=1}^C TP_i}{N} \quad (6)$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (7)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (8)$$

$$\text{Specificity}_i = \frac{TN_i}{TN_i + FP_i} \quad (9)$$

$$\text{F1-Score}_i (F1_i) = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (10)$$

Trong đó C là số lớp, i đề cập đến lớp thứ i . Khác với phương pháp lấy trung bình vĩ mô coi mọi lớp có vai trò bằng nhau bất kể số lượng mẫu, phương pháp trung bình trọng số (Weighted Average) tính toán các chỉ số cho từng lớp độc lập, sau đó tính trung bình tổng thể bằng cách gán trọng số dựa trên cấu trúc tần suất mẫu của từng lớp trong tập thử nghiệm.

$\text{Weight}(i) = \frac{N_i}{N_{\text{total}}}$, Với N_i là số lượng mẫu trong tập kiểm tra lớp i , N_{total} là tổng số lượng mẫu trong tập kiểm thử.

$$\text{Weighted-Metric} = \sum_{i=1}^C \text{Weight}(i) * \text{Metric}(i), \text{Metric}(i) \text{ là các độ đo của lớp } i \quad (11)$$

Các độ đo Precision, Recall, Specificity, F1-Score của các mô hình khảo sát được tính bằng công thức trung bình có trọng số của độ đo các lớp, theo công thức (11).

Bảng 3. Tham số huấn luyện tinh chỉnh

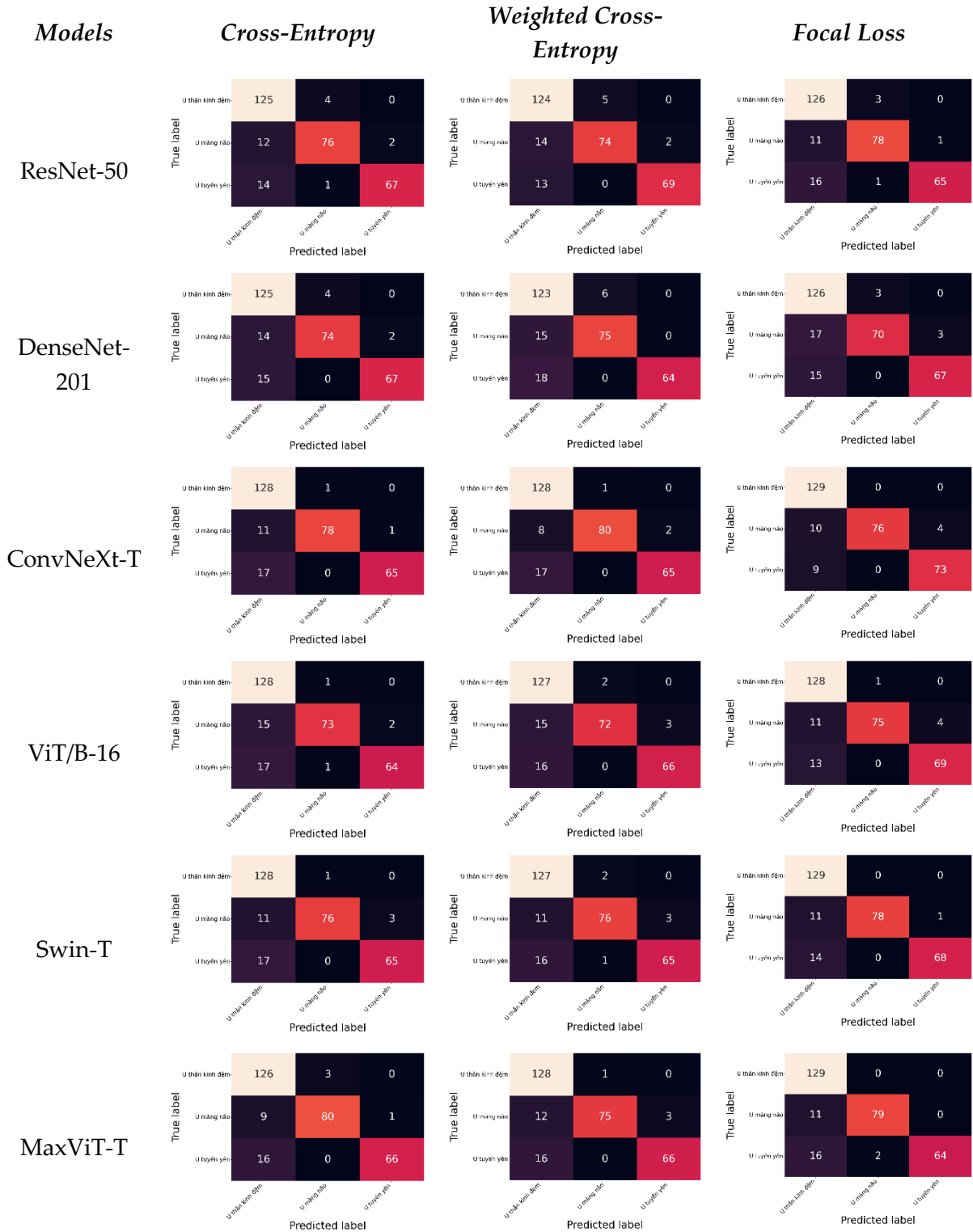
Số epoch huấn luyện/mô hình	300
Batch size	32

Learning rate	2e-05
Trình tối ưu	AdamW, weight_decay=1e-04
Trình thay đổi tốc độ học	ExponentialLR, decay = 0.995
Environment OS, GPU	Ubuntu 22.04, RTX 3080, Pytorch
Kích thước ảnh đầu vào	224×224
Hệ số gamma γ của hàm FL	2

Để tăng khả năng khái quát hóa và hạn chế hiện tượng quá khớp, trong quá trình huấn luyện chúng tôi áp dụng chiến lược tăng cường dữ liệu trực tuyến (Online Data Augmentation) thông qua thư viện torchvision.Transforms của PyTorch. Các phép biến đổi như xoay ngẫu nhiên trong phạm vi $\pm 30^\circ$, lật ngang (Horizontal Flip) và điều chỉnh màu sắc (Color Jitter) được thực hiện động trong quá trình nạp dữ liệu cho từng batch huấn luyện. Nhờ đó, ở mỗi epoch, mô hình được tiếp cận với các biến thể dữ liệu khác nhau nhưng vẫn bảo toàn các đặc trưng ngữ nghĩa quan trọng của ảnh. Phương pháp này giúp gia tăng tính đa dạng của dữ liệu huấn luyện, giảm nguy cơ quá khớp, đồng thời tiết kiệm tài nguyên lưu trữ và nâng cao hiệu quả tính toán thông qua cơ chế nạp dữ liệu đa luồng.

Mặc dù các phương thức ảnh y tế thường tuân theo các mặt phẳng giải phẫu và định hướng chuẩn, việc áp dụng biên độ xoay giới hạn ($\leq 30^\circ$) mô phỏng một cách hợp lý sự biến thiên trong thực tế lâm sàng, chẳng hạn như sự lệch góc do thao tác của kỹ thuật viên, vị trí đặt đầu dò hoặc tư thế của bệnh nhân. Đồng thời, phép lật ngang giúp mô hình đạt được tính bất biến đối xứng đối với các đặc trưng bệnh lý độc lập với tính đối xứng hai bên của cơ thể. Theo kết quả nghiên cứu đánh giá [26], các phép biến đổi này tập trung duy trì tính toàn vẹn của các cấu trúc vi mô và đặc trưng tổn thương cục bộ, yếu tố cốt lõi cho nhiệm vụ phân lớp – mà không làm thay đổi hay biến dạng bản chất giải phẫu quan trọng của ảnh.

Để đảm bảo tính tương thích hoàn toàn về mặt phân phối đặc trưng, toàn bộ ảnh thuộc tập huấn luyện và tập xác thực, tập kiểm tra đều được đưa về kích thước chuẩn 224×224 pixel, sau đó tiến hành chuẩn hóa kênh màu độc lập dựa trên các giá trị kỳ vọng (Mean) và độ lệch chuẩn (Standard Deviation) theo tiêu chuẩn tập dữ liệu ImageNet.



Hình 5. Ma trận nhầm lẫn trên tập kiểm tra các mô hình tương ứng với các hàm mất mát

Bảng 4. Chỉ số hiệu quả phân lớp trung bình có trọng số trên tập kiểm tra độc lập

Model	Loss Function	Accuracy (%)	Precision (%)	Recall/ Sensitivity (%)	Specificity (%)	F1-Score (%)
ResNet-50	CE	89,04	89,99	89,04	92,56	89,02
	WCE	88,70	89,68	88,70	92,32	88,72
	FL	89,37	90,57	89,37	92,58	89,35
DenseNet-201	CE	88,37	89,61	88,37	91,96	88,38
	WCE	87,04	88,72	87,04	90,93	87,10
	FL	87,38	88,92	87,38	91,23	87,33
ConvNeXt-T	CE	90,03	91,52	90,03	92,76	90,03
	WCE	90,70	91,81	90,70	93,38	90,65
	FL	92,36	93,08	92,36	94,77	92,31
ViT/B-16	CE	88,04	89,81	88,04	91,49	87,98
	WCE	88,04	89,60	88,04	91,62	88,00
	FL	90,37	91,35	90,37	93,38	90,32
Swin-T	CE	89,37	90,72	89,37	92,51	89,32
	WCE	89,04	90,15	89,04	92,47	88,97
	FL	91,36	92,65	91,36	93,65	91,37
MaxViT-T	CE	90,37	91,42	90,37	93,22	90,36
	WCE	89,37	90,73	89,37	92,51	89,33
	FL	90,37	91,84	90,37	92,99	90,31

Kết quả ở Bảng 4, cho thấy đối với bài toán phân loại hình ảnh y tế, các vùng đặc trưng nền thường chiếm phần lớn và dễ bị nhận diện sai lệch. Hàm FL tập trung tài nguyên tính toán và điều chỉnh gradient vào các mẫu khó hoặc các vùng tổn thương tiềm cận ranh giới, từ đó giúp mô hình học được các biểu diễn đặc trưng sắc nét và tổng quát hơn so với CE tiêu chuẩn. MaxViT duy trì F1-Score ổn định ở mức 89,33% - 90,36% trên cả 3 hàm mất mát. Tính ổn định tương đối này do kiến trúc lai kết hợp giữa Grid Attention và Block Attention, giúp mô hình tự cân bằng và trích xuất đặc trưng rất đa dạng ngay từ tầng mạng cơ sở, khiến độ nhạy của nó đối với sự thay đổi của hàm mất mát thấp hơn so với các kiến trúc thuần túy khác. Cả Swin-T và ViT/B16 đều cho thấy mức bứt phá F1-Score rất lớn khi chuyển từ hàm CE sang FL (tăng từ 2,0% đến 2,34%).

Bản chất của cơ chế Self-Attention yêu cầu dữ liệu và gradient đủ mạnh để học được mối quan hệ phi tuyến giữa các patch ảnh. Khi kết hợp với FL, quá trình tính toán hệ số chú ý được định hướng để tập trung vào các vùng ngữ cảnh quan trọng thay vì bị nhiễu bởi các khu vực nền. Mô hình ConvNeXt-T đạt kết quả cao nhất (F1-Score = 92,31% với hàm FL) nhờ sự kết hợp tinh tế giữa ưu điểm của CNN và Transformer. Chỉ số đo lường hiệu quả phân lớp của các mô hình được trình bày ở bảng 4, được tính toán theo phương pháp trung bình có trọng số theo công thức 11, nên giá trị Accuracy, Recall là như nhau. Chúng tôi, bổ sung dữ liệu các chỉ số ở bảng 5 được theo phương pháp trung bình vĩ mô của các độ đo hiệu quả phân lớp.

Bảng 5. Chỉ số trung bình vĩ mô hiệu quả phân lớp trên tập kiểm tra độc lập

Model	Loss Function	Accuracy (%)	Precision (%)	Recall/ Sensitivity (%)	Specificity (%)	F1-Score (%)
ResNet-50	CE	89,04	91,24	87,68	93,87	88,97
	WCE	88,70	90,99	87,50	93,67	88,78
	FL	89,37	91,99	87,87	93,98	89,30
DenseNet-201	CE	88,37	91,05	86,94	93,44	88,39
	WCE	87,04	90,48	85,58	92,66	87,24
	FL	87,38	90,45	85,72	92,87	87,28
ConvNeXt-T	CE	90,03	93,09	88,39	94,26	89,99
	WCE	90,70	93,15	89,13	94,69	90,53
	FL	92,36	93,99	91,16	95,71	92,18
ViT/B-16	CE	88,04	91,43	86,13	93,18	87,85
	WCE	88,04	91,11	86,31	93,22	87,91
	FL	90,37	92,47	88,90	94,58	90,17
Swin-T	CE	89,37	92,11	87,65	93,96	89,17
	WCE	89,04	91,42	87,39	93,84	88,79
	FL	91,36	94,11	89,86	95,00	91,36
MaxViT-T	CE	90,37	92,78	89,02	94,53	90,36
	WCE	89,37	92,13	87,68	93,96	89,20
	FL	90,37	93,41	88,61	94,45	90,20

Hàm FL thể hiện sự áp đảo rõ rệt trên hầu hết các kiến trúc, như ConvNeXt-T, ViT/B-16, Swin-T, ResNet-50. FL giúp nâng cao độ nhạy và F1-Score vượt trội nhờ tập trung xử lý các mẫu khó phân loại hoặc tình trạng mất cân bằng lớp. Cross-Entropy mang lại kết quả ổn định và cân bằng ở mức trung bình khá. Đối với MaxViT-T, CE lại cho F1-Score tốt nhất (90,36%), nhỉnh hơn WCE và FL trên kiến trúc này.

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Hàm mất mát là yếu tố chính định hướng cho quá trình huấn luyện các mô hình học sâu, nó quyết định việc mô hình điều chỉnh các trọng số trong quá trình huấn luyện.

Qua kết quả thực nghiệm đánh giá 3 hàm mất mát đối với 6 mô hình học sâu, được sử dụng cho nhiệm vụ phân lớp trên tập dữ liệu mất cân bằng với 3 lớp bệnh lý, hiệu quả phân lớp của hàm FL vượt trội rõ rệt so với hàm CE tiêu chuẩn trên cả 5 mô hình, riêng mô hình DenseNet-201, kết quả khi áp dụng hàm CE đạt hiệu quả cao hơn FL. Cùng một mô hình, phương pháp huấn luyện, nhưng khi áp dụng các hàm mất mát phù hợp, hiệu quả của mô hình có sự cách biệt rõ rệt.

WCE chỉ tối ưu hóa dựa trên tần suất xuất hiện của lớp chứ không căn cứ vào độ khó của đặc trưng theo từng mẫu. Do đó, WCE không giải quyết được triệt để vấn đề nhiễu ngữ cảnh trong ảnh y tế và dẫn đến hiện tượng phạt quá (over-penalty), làm giảm độ chính xác tổng thể so với hàm mất mát CE và FL.

Trên cơ sở khảo sát, đánh giá 3 hàm mất mát tiêu biểu, với phương pháp phân chia tập dữ liệu nghiên cứu thành 3 tập: tập huấn luyện, tập xác thực và tập kiểm tra lưu trữ độc lập theo bệnh nhân đã cho thấy hiệu suất của FL luôn ổn định. Trong nghiên cứu tiếp theo, cần đánh giá trên bộ dữ liệu lớn hơn để có kết quả khách quan và đáng tin cậy hơn.

TÀI LIỆU THAM KHẢO

- [1] O. Elharrouss *et al.*, (2025) Loss functions in deep learning: a comprehensive review, *arXiv e-prints*, p. arXiv: 2504.04242, .
- [2] S. Rajaraman, G. Zamzmi, and S. K. Antani, (2021), Novel loss functions for ensemble-based medical image classification, *Plos one*, vol. 16, no. 12, p. e0261307
- [3] H. P. Chan, L. M. Hadjiiski, and R. K. Samala (2020) Computer-aided diagnosis in the era of deep learning, *Medical physics*, vol. 47, no. 5, pp. e218-e227 .
- [4] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, (2021) Deep CNN for brain tumor classification, *Neural processing letters*, vol. 53, no. 1, pp. 671-700.
- [5] J. R. McFaline-Figueroa and E. Q. Lee, (2018) Brain tumors, *The American journal of medicine*, vol. 131, no. 8, pp. 874-882,

- [6] F. Ritter *et al.*, (2011), Medical image analysis," *IEEE pulse*, vol. 2, no. 6, pp. 60-70, .
- [7] T. Chen *et al.*, (2023), A computer-aided diagnosis system for brain tumors based on artificial intelligence algorithms, *Frontiers in Neuroscience*, vol. 17, p. 1120781.
- [8] B. Sistaninejhad, H. Rasi, and P. Nayeri, (2012), A review paper about deep learning for medical image analysis, *Computational and Mathematical Methods in Medicine*, vol. 2023, no. 1, p. 7091301, 2023.
- [9] K. Herholz, K.-J. Langen, C. Schiepers, and J. M. Mountz, (2012), Brain tumors, presented at the Seminars in nuclear medicine.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, (1998), Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, (2015), Deep learning, *nature communications*, vol. 521, no. 7553, pp. 436-444.
- [12] A. Saleh, R. Sukaik, and S. S. Abu-Naser, (2020), Brain tumor classification using deep learning, presented at the 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech).
- [13] S. E. Nassar, I. Yasser, H. M. Amer, and M. A. Mohamed, (2024), A robust MRI-based brain tumor classification via a hybrid deep learning technique: SE Nassar *et al*, *The Journal of Supercomputing*, vol. 80, no. 2, pp. 2403-2427.
- [14] A. Vaswani *et al.*, (2017), "Attention is all you need," *Advances in neural information processing systems*, vol. 30.
- [15] A. Dosovitskiy *et al.*, (2021), An image is worth 16x16 words: Transformers for image recognition at scale, presented at the 9th International Conference on Learning Representations ICLR 2021.
- [16] Z. Liu *et al.*, (2021), Swin transformer: Hierarchical vision transformer using shifted windows, presented at the Proceedings of the IEEE/CVF international conference on computer vision 2021.
- [17] Z. Tu *et al.*, (2022), Maxvit: Multi-axis vision transformer," presented at the European conference on computer vision.
- [18] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, (2022), Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling, *Current Oncology*, vol. 29, no. 10, pp. 7498-7511.
- [19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, (1986) Learning representations by back-propagating errors, *Nature*, vol. 323, no. 6088, pp. 533-536.
- [20] Y. S. Aurelio, G. M. De Almeida, C. L. de Castro, and A. P. Braga, (2019), Learning from imbalanced data sets with weighted cross-entropy function, *Neural processing letters*, vol. 50, no. 2, pp. 1937-1949.
- [21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, (2017), Focal loss for dense object detection, presented at the Proceedings of the IEEE international conference on computer vision.

- [22] K. He, X. Zhang, S. Ren, and J. Sun, (2016), Deep residual learning for image recognition, presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- [23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, (2017), Densely connected convolutional networks, presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- [24] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, (2022), A convnet for the 2020s, presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- [25] J. Cheng, (2017), brain tumor dataset. figshare. Dataset.,.
- [26] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, (2018), Differential data augmentation techniques for medical imaging classification tasks, presented at the AMIA annual symposium proceedings.

IMPACT OF LOSS FUNCTIONS ON CLASSIFICATION PERFORMANCE FOR IMBALANCED DATASETS

Pham Thanh An^{1,2}, Hoang Van Dung³, Le Van Tuong Lan^{4*}

¹Faculty of Information Technology, University of Sciences, Hue University

²Ho Chi Minh University of Banking

³Faculty of Information Technology, HCMC University of Technology and Engineering

⁴Training and Student Affairs Department, Hue University

*Email: lvtlan@hueuni.edu.vn

ABSTRACT

The cross-entropy (CE) loss function has been widely adopted as the standard objective for classification tasks. However, CE tends to favor majority classes, resulting in biased learning when trained on imbalanced datasets. This paper presents a comprehensive investigation of the impact of three representative loss functions, namely Cross-Entropy (CE), Weighted Cross-Entropy (WCE), and Focal Loss (FL), on the classification performance of imbalanced datasets. Six representative pre-trained deep learning models were selected for experimental evaluation. The experiments were conducted on a dataset comprising 3,064 magnetic resonance imaging (MRI) scans of brain tumors, categorized into three diagnostic classes. The dataset was randomly split at an 80:10:10 ratio by patient, corresponding to the training, validation, and test sets. Experimental results demonstrate that brain tumor classification performance using the FL loss function was superior across five of the six models evaluated. The WCE function performed significantly worse than the FL loss function.

Keywords: Deep learning, Brain tumor, Loss Function, Cross-Entropy, Focal Loss.