GIẢI PHÁP KẾT HỢP VISION TRANSFORMER VÀ ACTIVE LEARNING TRONG PHÁT HIỆN VÀ PHÂN LOẠI TỔN THƯƠNG GAN

COMBINED SOLUTION VISION TRANSFORMER AND ACTIVE LEARNING IN DETECTION AND CLASSIFICATION LIVER LESIONS

HỒ CHÍ HƯNG[1,a], PHAN THƯỢNG CANG[2]
[1] *Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Kỹ thuật Vĩnh Long*
[2] *Trường Công nghệ Thông tin và Truyền thông, Đại học Cần Thơ*
[a]*Tác giả lên hệ: hunghc@vlute.edu.vn*

Nhận bài (Received): 21/8/2024; Phản biện (Reviewed): 21/9/2024; Chấp nhận (Accepted): 9/10/2024

TÓM TẮT

Gan là cơ quan quan trọng trong cơ thể người, hầu hết các bệnh lý và các tổn thương ở gan thường khó phát hiện sớm do thiếu triệu chứng rõ ràng. Điều này dẫn đến nguy cơ biến chứng nặng, đặc biệt là ung thư gan, một trong những loại ung thư gây tử vong cao nhất toàn cầu. Bài báo đề xuất sử dụng các mô hình máy học như DenseNet-121, VGG-16 và ViT để phát hiện và phân loại tổn thương gan trên 2008 ảnh CT cắt lớp qua các thì arterial, delay, plain và venous. Các tổn thương được phân loại gồm nang gan, u mạch máu và ung thư biểu mô tế bào gan, nhằm nâng cao hiệu quả tầm soát và chẩn đoán sớm. Kết quả cho thấy mô hình ViT đạt độ chính xác lên đến 0.99 với thời gian huấn luyện ngắn. Ngoài ra, bài báo cũng chỉ ra rằng quá trình gán nhãn dữ liệu thủ công gặp nhiều thách thức lớn, bao gồm việc đòi hỏi một lượng lớn nhân lực có chuyên môn cao, tiêu tốn nhiều thời gian và chi phí. Bên cạnh đó, bài báo đề xuất sử dụng phương pháp học chủ động (active learning), nhằm tự động hóa một phần quy trình gán nhãn giúp giảm thiểu nguồn lực, tiết kiệm thời gian, chi phí và đảm bảo tính nhất quán cũng như chất lượng cho dữ liệu.

Từ khóa: DenseNet-121, VGG-16, ViT, Active Learning

ABTRACT

The liver is an essential organ in the human body, and most liver diseases and lesions are often difficult to detect early due to the lack of clear symptoms. This leads to a high risk of severe complications, particularly liver cancer, one of the leading causes of cancer-related deaths globally. This paper proposes using machine learning models such as DenseNet-121, VGG-16, and ViT to detect and classify liver lesions on 2008 CT scan images across arterial, delay, plain, and venous phases. The lesions are categorized into liver cysts, hemangiomas, and hepatocellular carcinoma, aiming to improve the efficiency of screening and early diagnosis. The results show that the ViT model achieved an accuracy of up to 0.99 with a short training time. Additionally, the paper highlights the major challenges of manual data labeling, which requires a significant amount of skilled

labor, consumes time, and incurs high costs. Furthermore, the paper suggests the use of active learning to automate part of the labeling process, reducing labor requirements, saving time and costs, while ensuring consistency and data quality.

Keywords: *DenseNet-121, VGG-16, ViT, Active Learning*

1. GIỚI THIỆU

1.1. Giới thiệu bài toán

Với 905.677 ca mắc, ung thư gan đứng thứ 6 trong các loại ung thư phổ biến và thứ 3 về tỷ lệ tử vong toàn cầu với 830.180 ca tử vong theo báo cáo năm 2020 của GLOBOCAN [12], một dự án của Cơ quan Nghiên cứu Ung thư Quốc tế. Cũng theo số liệu thống kê từ GLOBOCAN 2020 cho thấy, mỗi năm Việt Nam có 26.418 ca mắc ung thư gan mới, trong đó có tới 20.256 số ca phát hiện ở nam giới chiếm tỷ lệ 77%. Số ca tử vong do ung thư gan luôn dẫn đầu với 25.272 ca, chiếm 21,9% tổng số ca tử vong do ung thư. Trong đó có khoảng 80% số người mắc ung thư gan phát hiện ra bệnh ở giai đoạn nặng. Việc đưa ra giải pháp để phát hiện và chẩn đoán sớm các tổn thương ở gan thông qua các kỹ thuật máy học, học sâu là cần thiết để nhanh chóng phát hiện, kịp thời đưa ra hướng điều trị đúng đắn giúp giảm thiểu số ca mắc và số ca tử vong của các bệnh lý về gan đặc biệt là ung thư gan ở Việt Nam nói riêng và thế giới nói chung.

1.2. Các nghiên cứu liên quan

Một số công trình nghiên cứu về máy học trong dự đoán và phát hiện bệnh lý về gan tiêu biểu như: Phan Anh Cang và cộng sự [2] sử dụng các mạng Faster R-CNN, R-FCN và SSD để phát hiện u gan trên ảnh CT, đạt độ chính xác 95,1%; Weibin Wang và cộng sự [19] đã đề xuất việc phân loại tổn thương gan khu trú bằng phương pháp học chuyển giao transfer learning và tinh chỉnh fine tuning trên mô hình ResNet, độ chính xác đạt 91,2%; Ahmad và cộng sự [15] kết hợp ANFIS để chẩn đoán viêm gan, đạt 96,17%; Samuel Budd và cộng sự [8] khảo sát các phương pháp active learning và Human-in-the-Loop để cải thiện hiệu suất mô hình trong phân tích hình ảnh y tế.

2. CÁC CÔNG VIỆC LIÊN QUAN

2.1. Dấu hiệu nhận biết tổn thương gan

Tổn thương gan là bất thường tế bào trong gan, thường lành tính nhưng có thể gây ung thư. Các tổn thương này hiếm khi gây triệu chứng cho đến khi phát triển lớn như đau bụng, buồn nôn và thay đổi màu sắc nước tiểu hoặc phân [1]. Việc phát hiện và điều trị kịp thời thông qua các kỹ thuật hình ảnh như siêu âm, CT scan và MRI là rất quan trọng. **Hình 1** minh họa các dạng tổn thương gan phổ biến thông qua ảnh CT bao gồm u nang gan (hepatic cysts), u mạch máu (hepatic hemangioma) và ung thư biểu mô tế bào (hepatocellular carcinoma).

hepatic cysts ***hepatic hemangioma*** ***hepatocellular carcinoma***

Hình 1. Các dạng tổn thương ở gan

2.2. Mô hình máy học và phương pháp được dùng trong huấn luyện

Mạng DenseNet-121 (Densely Connected Convolutional Networks) [9] là cải tiến của ResNet với cấu trúc đơn giản và ít tham số hơn [14]. Mỗi lớp trong DenseNet kết nối trực tiếp với các lớp trước đó, tối ưu hóa luồng thông tin và gradient, giúp huấn luyện dễ dàng hơn, đặc biệt trên tập dữ liệu nhỏ [11]. DenseNet-121 đã chứng minh hiệu quả trong việc chẩn đoán các bệnh dựa trên hình ảnh y tế [17][7] và trong các lĩnh vực khác như phân loại hình ảnh công nghiệp với độ chính xác vượt trội. Mô hình này có khả năng tinh chỉnh đa cấp độ, tạo ra nhiều mạng chuyên biệt [6].

Mạng VGG-16 (Visual Geometry Group 16-layer network) [21] là một mạng nơ-ron tích chập sâu với 16 lớp, gồm 13 lớp tích chập và 3 lớp kết nối đầy đủ. Mạng này sử dụng các bộ lọc nhỏ kích thước 3x3 với stride 1 và padding, đảm bảo giữ nguyên kích thước của đầu vào qua các lớp tích chập. Các lớp tích chập này được theo sau bởi các lớp pooling (lớp giảm mẫu) kích thước 2x2, giúp giảm kích thước không gian của đặc trưng, giảm thiểu số lượng tham số và tính toán. VGG-16 cho phép học các đặc trưng phức tạp từ dữ liệu [3].

Mạng ViT (Vision Transformer) [20] là phiên bản của Transformer dành cho xử lý hình ảnh, đã cho thấy hiệu suất vượt trội trong phân loại hình ảnh so với các CNN truyền thống [10][4]. Sự khác biệt chính là việc sử dụng các lớp tích chập để trích xuất đặc trưng hình ảnh thay vì lớp nhúng được sử dụng trong Transformer gốc. Không giống như các CNN truyền thống sử dụng các phép tích chập không gian để trích xuất các đặc trưng từ hình ảnh, các mô hình Vision Transformer (ViT) sử dụng các cơ chế tự chú ý để nắm bắt các mối quan hệ giữa các vùng khác nhau của một hình ảnh có thể cải thiện hiệu suất [5].

Active learning [13] là phương pháp giảm công sức gán nhãn bằng cách hỗ trợ gán nhãn bán tự động. Thay vì huấn luyện trên lượng lớn dữ liệu, nó sử dụng chiến lược lấy mẫu để chọn các dữ liệu chứa nhiều thông tin, giúp tăng tốc độ hội tụ và giảm số lượng dữ liệu cần gán nhãn. Kết quả lý thuyết cho thấy chiến lược chọn lọc hiệu quả có thể giảm đáng kể công sức gán nhãn so với chọn ngẫu nhiên. Tuy nhiên, trong phát hiện lỗi, các chiến lược hiện tại còn gặp nhiều thách thức [18][16]. Trong bài nghiên cứu này, nhóm nghiên cứu sẽ sử dụng thuật toán "random selection" để chọn mẫu ngẫu nhiên, phù hợp với dữ liệu nhỏ và tăng độ chính xác hơn so với các thuật toán khác như "least confidence", "highest entropy", "least margin".

2.3. Phương pháp đánh giá mô hình

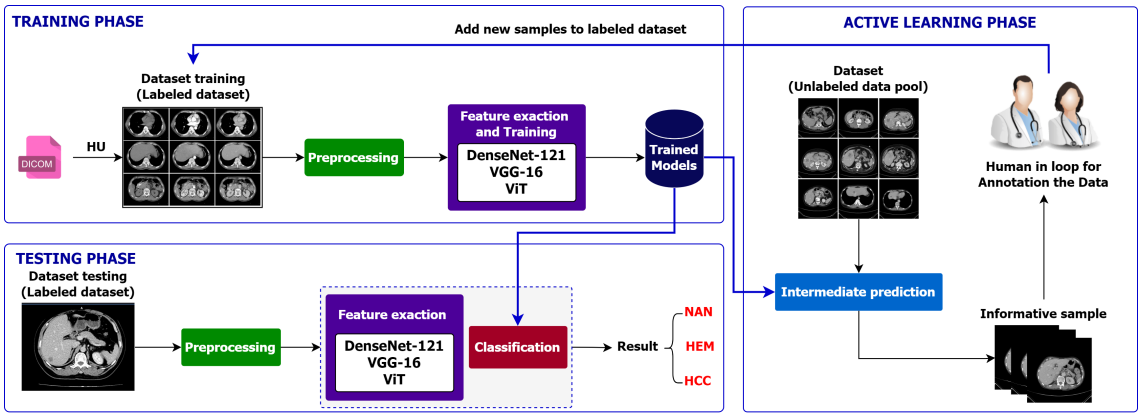
Hàm tính toán độ mất mát (loss) được sử dụng trong các bài toán phân loại nhiều lớp được trình bày chi tiết trong công thức (1). Trong đó, y_{true} thể hiện cho các nhãn thực tế, y_{pred} thể hiện cho nhãn dự đoán từ mô hình, N là số lượng mẫu có trong mô hình, $y_{pred}[i,j]$ biểu thị giá trị logic của mẫu thứ i và lớp thứ j.

$$Loss(y_{true}, y_{pred}) = -\frac{1}{N}\sum_{i=1}^{N} log\left(\frac{\exp(y_{pred}[i,y_{true}[i]])}{\sum_j exp(y_{pred}[i,j])}\right) \quad (1)$$

Hàm tính toán độ chính xác (accuracy) của mô hình được sử dụng trong các bài toán phân loại nhiều với được biểu diễn dưới dạng phương trình (2). Trong đó, N là số lượng mẫu có trong mô hình, $\hat{y}_i$ là nhãn dự đoán cho mẫu thứ i, y_i, là nhãn thực tế tương ứng cho mẫu thứ i, $1(\cdot)$ là điều kiện, bằng 1 nếu điều kiện bên trong là đúng và bằng 0 nếu ngược lại.

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} 1(\hat{y}_i = y_i) \quad (2)$$

3. PHƯƠNG PHÁP ĐỀ XUẤT



Hình 2. Mô hình đề xuất để phát hiện và phân loại bệnh gan

3.1. Giai đoạn huấn luyện

3.1.1. Chuẩn bị dữ liệu và tiền xử lý ảnh

Trong giai đoạn chuẩn bị dữ liệu, sau khi tập dữ liệu ảnh CT được chuẩn hóa từ ảnh DICOM sang ảnh kỹ thuật số bằng kỹ thuật HU (Hounsfield), bộ dữ liệu đã được gán nhãn sẽ được chia thành 4 tập nhỏ, tương ứng với các thì arterial, delay, plain và venous. Mỗi tập dữ liệu nhỏ sẽ được chia thành tập huấn luyện và tập kiểm thử. Các tập huấn luyện và kiểm thử này sẽ chứa 3 lớp: nang gan (NAN), u mạch máu (HEM) và ung thư biểu mô tế bào (HCC). Nhóm nghiên cứu sẽ áp dụng một số kỹ thuật nhằm tăng cường dữ liệu ảnh cho quá trình huấn luyện như: rotation, width shift, height shift, zoom, horizontal flip.

3.1.2. Rút trích đặc trưng và quá trình huấn luyện

Sau khi tiền xử lý dữ liệu ảnh CT, nhóm sử dụng ba mô hình DenseNet-121, VGG-16 và ViT để trích xuất đặc trưng và huấn luyện trên ba loại tổn thương. Các mô hình được huấn luyện trong cùng môi trường với thông số giống nhau và khi loss không giảm đáng kể, quá trình huấn luyện kết thúc. Sau đó, nhóm sẽ kiểm thử và so sánh hiệu quả các mô hình.

3.2. Giai đoạn kiểm thử

Trong giai đoạn kiểm thử, dữ liệu được tiền xử lý và phân loại để xác định bệnh gan. Các đặc trưng được trích xuất tương tự như trong giai đoạn huấn luyện. Mỗi hình ảnh sẽ được xử lý qua mô hình đã huấn luyện để tự động phát hiện bệnh và dự đoán nhãn tổn thương, nhằm chọn kết quả chính xác và nâng cao độ chính xác của quá trình phân loại.

3.3. Giai đoạn active learning

Nhóm nghiên cứu đề xuất tích hợp active learning để tối ưu hóa gán nhãn và hỗ trợ phát hiện, phân loại tổn thương gan. Với 500 ảnh, trong đó 150 ảnh đã được gán nhãn, mô hình sẽ được huấn luyện và thử nghiệm qua các thuật toán chọn mẫu như random selection, least confidence, highest entropy và least margin. Từ 350 ảnh chưa gán nhãn, mô hình sẽ tạo dự đoán trung gian, được chuyên gia xác nhận. Nếu chính xác, các ảnh này sẽ bổ sung vào tập gán nhãn, cải thiện dữ liệu huấn luyện. Quy trình lặp lại cho đến khi đạt hiệu suất mong muốn, giảm thiểu công sức và chi phí.

4. KẾT QUẢ THỰC NGHIỆM

4.1. Môi trường cài đặt và tập dữ liệu thực nghiệm

Quá trình huấn luyện và kiểm thử mô hình được thực hiện trên môi trường Linux với CPU 6 lõi, RAM 64 GB, GPU NVIDIA Tesla V100 16GB, TensorFlow GPU 2.10 và Python 3.10. Tập dữ liệu gồm 2008 ảnh CT từ một bệnh viện ở Cần Thơ, với kích thước 512x512 pixel. Bộ dữ liệu chứa các ảnh CT tương ứng với bốn thì và ba loại tổn thương: nang gan (NAN) với 316 ảnh, u mạch máu (HEM) với 504 ảnh và ung thư biểu mô tế bào gan (HCC) chiếm hơn 50% với 1188 ảnh.

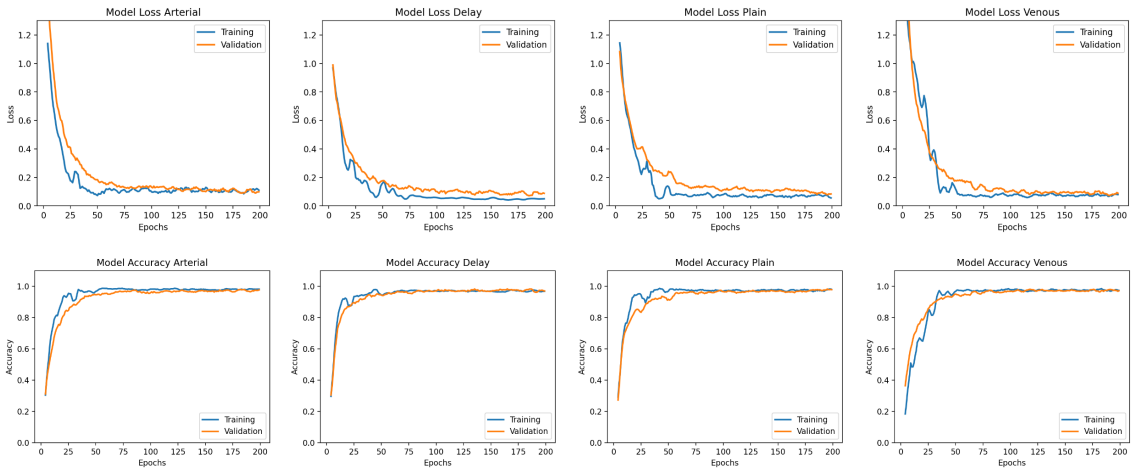
4.2. Kịch bản huấn luyện

Bảng 1. Tổng quan kịch bản huấn luyện

Kịch bản	Mô hình sử dụng	Tập dữ liệu	batch size	num classes	learning rate	num steps
1	DenseNet-121	Arterial	16	3	0.001	200
		Delay	16	3	0.001	200
		Plain	16	3	0.001	200
		Venous	16	3	0.001	200
2	VGG-16	Arterial	16	3	0.001	200
		Delay	16	3	0.001	200
		Plain	16	3	0.001	200
		Venous	16	3	0.001	200
3	ViT	Arterial	16	3	0.001	200
		Delay	16	3	0.001	200
		Plain	16	3	0.001	200
		Venous	16	3	0.001	200

4.3. Kết quả quá trình huấn luyện

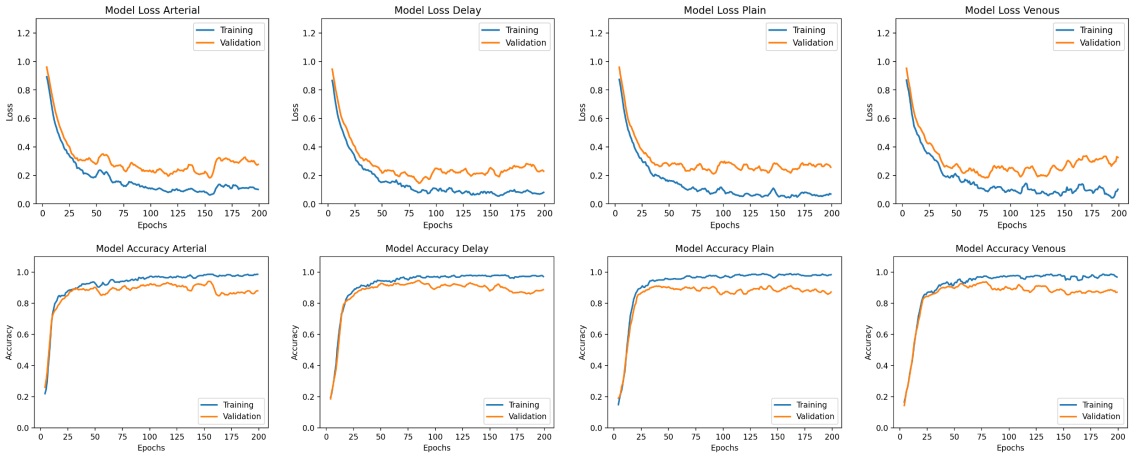
a. Kịch bản 1



Hình 3. Biểu đồ minh họa độ mất mát (loss) và độ chính xác (accuracy) của kịch bản 1 (DenseNet-121)

Hình 3 minh họa biểu đồ loss và accuracy của mạng DenseNet-121 dựa trên các tập dữ liệu arterial, plain, delay và venous. Giá trị loss giảm đáng kể trong 50 epochs đầu và ổn định sau đó, việc huấn luyện dừng ở 200 epochs với các giá trị thu được là 0.09, 0.07, 0.11 và 0.08. Về accuracy, các giá trị tăng mạnh trong 25 epochs đầu và ổn định, với kết quả thu được là 0.98, 0.97, 0.95 và 0.97.

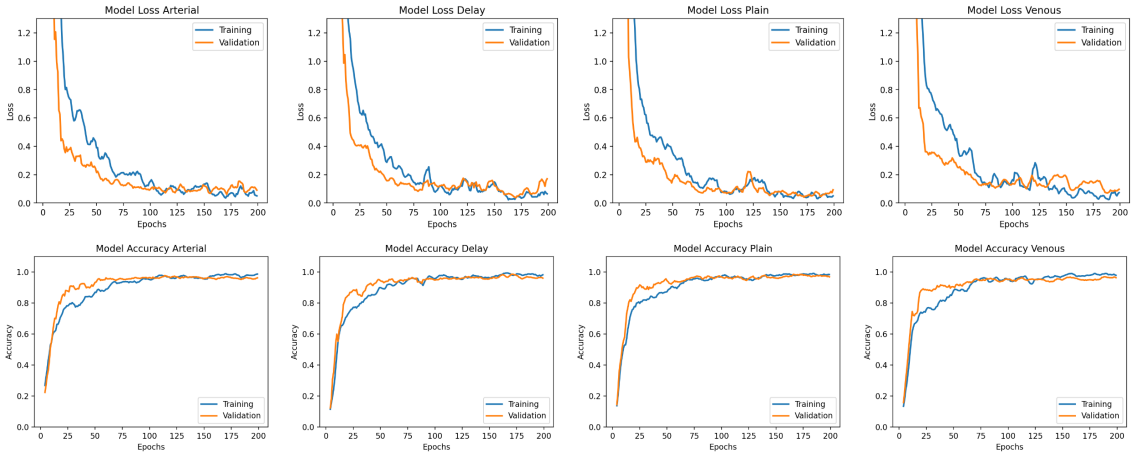
b. Kịch bản 2



Hình 4. Biểu đồ minh họa độ mất mát (loss) và độ chính xác (accuracy) của kịch bản 2 (VGG-16)

Hình 4 minh họa biểu đồ loss và accuracy của mạng VGG-16. Dù giá trị training tốt, nhưng giá trị validation chưa đạt hiệu quả cao. Loss giảm đáng kể trong 100 epochs đầu và ổn định sau đó, khi chúng tôi dừng huấn luyện ở 200 epochs với các giá trị loss cuối cùng thu được là 0.06, 0.09, 0.15 và 0.13. Về accuracy, các giá trị tăng mạnh trong 25 epochs đầu và ổn định, với kết quả cuối cùng thu được là 0.96, 0.95, 0.94 và 0.95.

c. Kịch bản 3

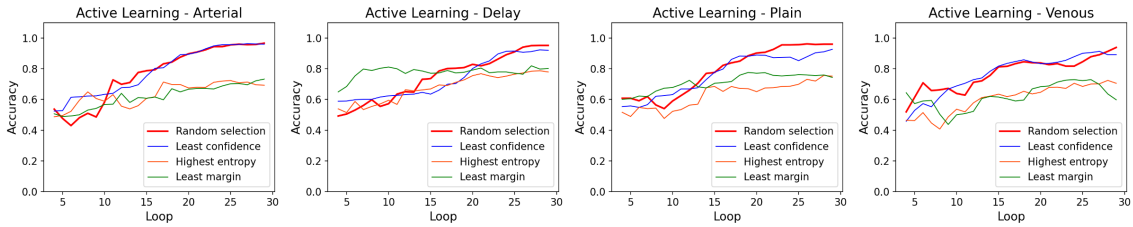


Hình 5. Biểu đồ minh họa độ mất mát (loss) và độ chính xác (accuracy) của kịch bản 3 (ViT)

Hình 5 cho thấy sự tiến triển rõ rệt của mạng Vision Transformer trong kịch bản 3. Với các tập dữ liệu arterial, plain, delay và venous, mô hình đạt độ chính xác lần lượt là 0.99, 0.98, 0.96 và 0.98. Trong 100 epochs đầu, loss giảm đáng kể, dù tập plain

và venous có chút không ổn định. Đến epochs 130, loss ổn định và giảm đáng kể. Sau 200 epochs, khi loss không giảm thêm, nhóm dừng huấn luyện. Kết quả cuối cùng cho loss lần lượt là 0.03, 0.06, 0.12 và 0.07.

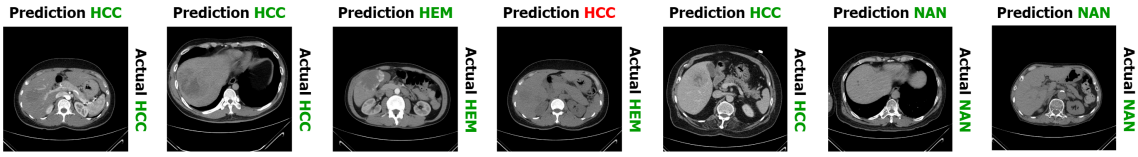
4.4. Kết hợp ViT với active learning



Hình 6. Kết quả thực nghiệm giải pháp kết hợp mô hình mạng ViT và active learning

Sau khi thực nghiệm, nhóm nghiên cứu chọn mô hình ViT kết hợp với active learning qua 30 vòng lặp, mỗi vòng gồm 10 epochs và thêm 5 ảnh vào dữ liệu như đã trình bày ở phần 3.3. Kết quả từ hình 6 cho thấy random selection đạt hiệu suất cao nhất với độ chính xác 0.95-0.98, vượt trội hơn các thuật toán khác. Least confidence có độ chính xác 0.91-0.97, trong khi highest entropy và least margin đạt 0.70-0.83.. Với kết quả khả quan, nhóm nghiên cứu kết luận random selection là giải pháp hiệu quả nhất cho việc gán nhãn tự động trong mô hình Vision Transformer.

4.5. Kết quả phân loại bệnh gan



Hình 7. Một số hình ảnh kết quả phân loại bệnh gan

5. ĐÁNH GIÁ VÀ SO SÁNH CÁC KỊCH BẢN HUẤN LUYỆN

Qua quá trình thực nghiệm trên bốn tập dữ liệu cho ba kịch bản đề xuất. **Biểu đồ 1.a** cho thấy độ chính xác dao động từ 0.94 đến 0.99, với tập arterial ổn định và đạt cao nhất (0.99). Trong khi đó, tập delay, plain và venous dao động nhiều hơn, với độ chính xác thấp nhất 0.94 ở kịch bản 2. **Biểu đồ 1.b** cho thấy độ mất mát ở kịch bản 3 thấp nhất, đặc biệt arterial chỉ 0.03, trong khi kịch bản 2 có mức mất mát cao nhất (plain là 0.15). **Biểu đồ 1.c** thể hiện thời gian huấn luyện, với kịch bản 3 chỉ mất 56 phút tổng cộng nhờ ViT chia dữ liệu thành patch, giảm thiểu thời gian huấn luyện. Kịch bản 3 được đánh giá hiệu quả nhất, cân bằng giữa thời gian và hiệu suất. Vì mô hình mạng ViT sẽ chia nhỏ các dữ liệu ra thành các patch để đem đi huấn luyện giúp giảm thiểu thời gian cũng như dò tìm đặc trưng trên ảnh, thông qua đó cho thấy, kịch bản 3 không chỉ hiệu quả về thời gian huấn luyện mà còn vượt trội về độ chính xác và hiệu suất, do đó nhóm đánh giá đây là mô hình mạng hiệu quả nhất trong việc cân bằng giữa thời gian huấn luyện và hiệu suất mô hình.

6. KẾT LUẬN

Trong nghiên cứu này, nhóm nghiên cứu đã đề xuất và áp dụng các mô hình máy học DenseNet-121, VGG-16 và ViT để nhận dạng và phân loại tổn thương gan dựa trên ảnh CT và đưa ra giải pháp active learning để gán nhãn tự động dữ liệu. Mục tiêu là phát hiện ba loại tổn thương gan và đề xuất ứng dụng phương pháp active

learning vào bài toán nhằm giải quyết vấn đề gán nhãn dữ liệu tự động giúp giảm thời gian, chi phí và tăng cường dữ liệu một cách bán tự động cho các chuyên gia về chẩn đoán. Kết quả thực nghiệm cho thấy cả ba mô hình mạng đều đạt độ chính xác cao, tuy nhiên, mô hình ViT nổi bật hơn với khả năng phát hiện và phân loại tổn thương vượt trội, đạt độ chính xác lên đến 0.99 và độ mất mát chỉ 0.03. Điều này khẳng định tiềm năng của ViT trong việc nâng cao hiệu quả tầm soát và chẩn đoán sớm tổn thương gan, góp phần cải thiện sức khỏe và chất lượng cuộc sống của bệnh nhân. Tuy nhiên, trong nghiên cứu này còn hạn chế về thời gian huấn luyện đối với các mạng CNN và số lượng dữ liệu ảnh trong tập huấn luyện và kiểm thử còn khá ít. Mặc dù phương pháp active learning có thể hỗ trợ đội ngũ chuyên gia trong việc gán nhãn tự động cho tập dữ liệu chưa được gán nhãn nhưng nó cũng có những hạn chế. Các thuật toán active learning có thể gặp khó khăn trong việc chọn lựa mẫu dữ liệu cần gán nhãn một cách hiệu quả khi đối mặt với dữ liệu phức tạp và đa dạng. Thêm vào đó, việc triển khai active learning đòi hỏi sự hiểu biết sâu về các thuật toán và có thể dẫn đến chi phí ban đầu cao cho việc phát triển và duy trì hệ thống. Cuối cùng, dù tự động hóa quy trình gán nhãn là một mục tiêu lý tưởng nhưng sự can thiệp của con người vẫn cần thiết để đảm bảo chất lượng và tính chính xác của dữ liệu. Trong tương lai, nhóm sẽ tiếp tục nghiên cứu đề tài này với các mô hình mạng khác như MobileNet, EfficientNet, ANFIS kết hợp với một số kỹ thuật về tăng cường dữ liệu ảnh để so sánh và đánh giá nhằm tìm ra mô hình tối ưu nhất cho bài toán này. Bên cạnh đó, nhóm cũng sẽ áp dụng bài toán vào môi trường xử lý phân tán và song song dữ liệu lớn để xử lý lượng dữ liệu ngày càng tăng khi áp dụng phương pháp active learning.

TÀI LIỆU THAM KHẢO

[1]. Ashish Sharma; Shivaraj Nagalli, "Chronic Liver Disease," StatPearls Publishing, 2024.

[2]. Anh-Cang Phan, Hung-Phi Cao, Thanh-Ngoan Trieu, Thuong-Cang Phan, "Improving liver lesions classification on CT/MRI images based on Hounsfield Units attenuation and deep learning," Gene Expression Patterns, vol. 119289, p. 47, 2023.

[3]. Ramadhan, Muhammet Baykara, "A Novel Approach to Detect COVID-19: Enhanced Deep Learning Models with Convolutional Neural Networks," Applied Sciences, vol. 12, no. 18, 2022.

[4]. Xia et al., "Vision transformer with deformable attention," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 4794–4803, 2022.

[5]. Bhojanapalli et al., "Understanding Robustness of Transformers for Image Classification," In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10-17, 2021.

[6]. Najmul Hasan, Yukun Bao, Ashadullah Shawon and Yanmei Huang , "DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image," SN Computer Science, vol. 2, no. 389, 2021.

[7]. Sarkar, Hazra, Das, "Classification of colorectal cancer histology images using image reconstruction and modified DenseNet," CICBA, 2021.

[8]. Samuel Budd, Robinson and Bernhard Kainz, "A survey on active learning and human-in-the-loop deep learning for medical image analysis," Medical Image Analysis, vol. 71, 2021.

[9]. Gao, Chen, Niu and Plaza, "Recognition and mapping of landslide using a fully convolutional densenet and influencing factors," IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens, vol. 14, pp. 7881-7894, 2021.

[10]. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," Proceedings of the IEEE/CVF International Conference on Computer Vision, p. 10012–10022, 2021.

[11]. Wang et al.,"A novel LiDAR data classification algorithm combined capsnet with ResNet," Sensors, vol. 20, 2020.

[12]. Hyuna Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," GLOBOCAN, 2020.

[13]. Xiaoming Lv, Fajie Duan, Jia-Jia Jiang Xiao Fu and Lin Gan, "Deep Active Learning for Surface Defect Detection," Sensors 2020, 2020.

[14]. Deng, Jiang, Lan, Huang and Luo, "Image captioning using densenet network and adaptive attention," Signal Process. Image Commun, vol. 85, 2020.

[15]. Ahmad et al., "Intelligent hepatitis diagnosis using adaptive neuro-fuzzy inference system and information gain method," Soft Computing, pp. 10931-10938, 2019.

[16]. Fan et al., "Uncertainty metric in model-based eddy current inversion using the adaptive Monte Carlo method," Measurement, vol. 137, pp. 323-331, 2019.

[17]. Guo, Xu, Zhang, "Interstitial lung disease classification using improved DenseNet," Multimed Tools Appl, p. 78(21):30615–26, 2019.

[18]. Hu et al., "Pattern deep region learning for crack detection in thermography diagnosis system," Metals , vol. 8, p. 612, 2018.

[19]. Weibin Wang et al., "Classification of Focal Liver Lesions Using Deep Learning with Fine-Tuning," DMIP, 2018.

[20]. Ashish Vaswani et al., "Attention is all you need," Advances in neural information processing systems, , vol. 30, 2017.

[21]. Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 3rd International Conference on Learning Representations (ICLR 2015), pp. 1-14, 2015.