

CHẨN ĐOÁN BỆNH ALZHEIMER BẰNG KỸ THUẬT VISION TRANSFORMER

DIAGNOSIS OF ALZHEIMER'S DISEASE VISION TRANSFORMER ENGINEERING DIPLOMA

PHAN ANH CANG¹, TRẦN THỊ THU NGÂN^{2,a}

¹Trường Đại học Sư phạm Kỹ thuật Vĩnh Long

²Học viên cao học, Trường Đại Học Sư Phạm Kỹ Thuật Vĩnh Long,

^aTác giả liên hệ: ngantran.dl@gmail.com

Nhận bài(Received): 25/3/2024; Phản biện (Reviewed):20/4/2024; Chấp nhận (Accepted):23/04/2024

TÓM TẮT

Alzheimer là loại bệnh được chẩn đoán phổ biến nhất và là một trong bảy nguyên nhân gây tử vong hàng đầu thế giới. Khi được chẩn đoán sớm, bệnh nhân có thể tiếp cận với các phương pháp điều trị sẵn có, từ đó giúp bệnh nhân cải thiện nhận thức và nâng cao chất lượng cuộc sống. Chụp MRI não hiện nay là phương pháp hữu hiệu để xác định và chẩn đoán sớm bệnh Alzheimer. Trong bài báo này, chúng tôi đề xuất kỹ thuật phân loại dựa trên kiến trúc mạng Inception V3, Xception, DenseNet 121, Vision Transformer để phân loại ảnh bệnh Alzheimer là bình thường, nhẹ, rất nhẹ hoặc trung bình. Qua kết quả thực nghiệm cho thấy độ chính xác của mô hình đề xuất đạt 96,7% trên tập dữ liệu augmented-alzheimer-mri-dataset-v2 [1].

Từ khóa: Vision Transformer, Deep learning, Alzheimer, MRI.

ABSTRACT

Alzheimer's is the most commonly diagnosed disease and one of the seven leading causes of death in the world. When diagnosed early, patients can access available treatments. From there, it helps patients improve cognition and improve quality of life. Brain MRI is currently an effective method to identify and diagnose Alzheimer's disease early. In this paper, we propose a classification technique based on the network architecture of Inception V3, Xception, DenseNet 121, Vision Transformer to classify Alzheimer's disease images as normal, mild, very mild or moderate. Experimental results show that the accuracy of the proposed model is up to 96.7% on the augmented-alzheimer-mri-dataset-v2 [1] dataset.

Keywords: Vision Transformer, Deep learning, Alzheimer's, MRI

1. MỞ ĐẦU

1.1. Giới thiệu bài toán

Alzheimer [2] là một hội chứng thần kinh thoái hóa nghiêm trọng của não được đặc trưng bởi mất trí nhớ và suy giảm nhận

thức, cuối cùng là khả năng thực hiện các nhiệm vụ đơn giản nhất. Theo thống kê của WHO [3], năm 2019 có 55,2 triệu người mắc hội chứng sa sút trí tuệ trên toàn cầu, dự đoán đến năm 2030 con số này sẽ tăng lên 78 triệu người, đây cũng là nguyên

nhân một trong các nguyên nhân tử vong hàng đầu trên toàn thế giới. Triệu chứng của Alzheimer không hiện rõ từ đầu, điều này làm cho việc phát hiện và chẩn đoán bệnh trở nên khó khăn. Việc xây dựng một hệ thống hỗ trợ việc chẩn đoán bệnh Alzheimer sẽ tạo nhiều thuận lợi hơn cho các bác sĩ trong quá trình phân tích và chẩn đoán thông qua hình ảnh MRI não để có giải pháp điều trị kịp thời vô cùng cần thiết. Alzheimer không có cách chữa trị hoàn toàn, các biện pháp chữa trị sẽ chỉ đóng vai trò trì hoãn.

1.2. Những nghiên cứu liên quan

Menagadevi và cộng sự [4] đã phát triển một hệ thống chẩn đoán có sự trợ giúp của máy tính để phát hiện AD. Trước tiên, nghiên cứu tiền xử lý trên hình ảnh MRI đầu vào để nâng cao chất lượng hình ảnh, tiếp đến thực hiện phân đoạn trên các hình ảnh được xử lý để thu được vùng quan tâm. Sau đó trích xuất các tính năng bằng mô hình bộ mã hóa tự động đa cấp và cuối cùng sử dụng các bộ phân loại riêng biệt để phân loại. Nghiên cứu đạt độ chính xác tổng thể là 96,88% khi sử dụng bộ phân loại KNN và 98,97% khi sử dụng bộ phân loại ELM cho nhiệm vụ phân loại nhị phân. Sharma và cộng sự [5] đã trình bày một phương thức kết hợp có tên là “HTML” dựa trên phương pháp tiếp cận AI để phát hiện AD từ hình ảnh MRI. Các tác giả thực hiện tiền xử lý trên các hình ảnh MRI, sau đó cung cấp song song những hình ảnh được xử lý này vào hai mô hình được đào tạo trước như DenseNet201 và DenseNet121. Tiếp đến, tiến hành phân loại bằng các bộ phân loại riêng biệt cho từng mô hình được đào tạo trước. Cuối cùng, kết hợp đầu ra cho mỗi bộ phân loại bằng cách sử dụng chiến lược bỏ phiếu, kết quả độ chính xác tổng thể là 91,75% cho nhiệm vụ đa phân loại. Hazarika và cộng sự [6] đã trình bày

một cách tiếp cận để phân loại AD bằng cách sử dụng mạng lưới thần kinh sâu và MRI. Phương pháp này xử lý trước các lần quét MRI, trích xuất các đặc điểm từ các hình ảnh não được phân đoạn bằng cách sử dụng kết hợp các CNN 2D và 3D, đồng thời phân loại các lần quét thành AD và không AD bằng mạng thần kinh được kết nối đầy đủ. Các tác giả đã đạt được kết quả đầy hứa hẹn với độ chính xác 95,34%. Hu và cộng sự [7] đã giới thiệu một nghiên cứu dài hạn về MCI bằng cách sử dụng MRI cấu trúc não (sMRI) làm dấu ấn sinh học chính. Mô hình VGG-Tswinformer kết hợp CNN và Transformer dựa trên VGG-16 để trích xuất và mã hóa các đặc điểm từ các hình ảnh sMRI theo chiều dọc, đồng thời mô hình này sử dụng các cơ chế chú ý theo thời gian và cửa sổ trượt để tích hợp các đặc điểm không gian ở xa và cục bộ để dự đoán tiến trình MCI. Họ đạt được độ chính xác 77,20% cho nhiệm vụ phân loại nhị phân.

1.3. Dấu hiệu của não về bệnh Alzheimer

Bộ não con người khỏe mạnh chứa hàng chục tỷ tế bào thần kinh - các tế bào chuyên biệt xử lý và truyền thông tin qua các tín hiệu điện và hóa học. Bệnh Alzheimer gây ra sự chết dần của các tế bào não. Điều này là kết quả của hai cấu trúc bất thường trong não [8]: mảng amyloid, là những khối protein tích tụ bên ngoài tế bào và đám rối sợi thần kinh là những khối protein bị thay đổi bên trong tế bào. Phác thảo về bộ não của một người mắc bệnh Alzheimer cho thấy nhiều tế bào thần kinh bắt đầu co lại và cuối cùng biến mất.

1.4. Các mạng nơ-ron trích xuất đặc trưng

a) *Mạng InceptionV3*: InceptionV3 [9] sử dụng cấu trúc “Inception” độc đáo, kết

hợp các bộ lọc kích thước khác nhau để trích xuất đa dạng đặc điểm hình ảnh hiệu quả. Nhờ kiến trúc tối ưu, InceptionV3 đạt hiệu quả tính toán cao, tiết kiệm thời gian và tài nguyên.

b) Mạng Xception: Xception [10] sử dụng mô-đun độc đáo, tách biệt phép toán tích chập theo chiều sâu và chiều rộng, giúp giảm số lượng tham số và tăng hiệu quả. Ngoài ra, Xception có thể được điều chỉnh dễ dàng cho nhiều nhiệm vụ khác nhau và đạt hiệu suất cao so với các mạng nơ-ron khác trên nhiều tập dữ liệu tiêu chuẩn.

c) DenseNet 121: DenseNet 121 [11] sử dụng kết nối dày đặc giữa các lớp, giúp tăng cường khả năng truyền ngược tín hiệu và cải thiện hiệu quả học tập. Nhờ kết nối tối ưu, DenseNet 121 đạt hiệu quả cao với số lượng tham số ít hơn so với các mạng nơ-ron khác, do đó giúp tiết kiệm tài nguyên và thời gian huấn luyện.

d) Mạng Vision Transformer: Vision Transformer [12] sử dụng kiến trúc Transformer dựa trên sự chú ý thay vì tích chập truyền thống, cho phép chia nhỏ hình ảnh thành các “patch” nhỏ. Các lớp Transformer tiếp tục xử lý thông tin, trích

xuất đặc điểm và đưa ra dự đoán cho nhiệm vụ cụ thể. Vision Transformer có thể được điều chỉnh dễ dàng cho nhiều nhiệm vụ khác nhau và có khả năng mở rộng cho các tập dữ liệu lớn. Đây là một mạng học sâu đột phá với nhiều ưu điểm và tiềm năng ứng dụng rộng rãi trong lĩnh vực thị giác máy tính.

e) Các độ đo đánh giá mô hình: Accuracy [13] được đánh giá dựa trên các giá trị như true positives (TP), false positives (FP), true negatives (TN) và false negatives (FN) được mô tả theo phương trình (1). Hàm categorical cross-entropy loss $L(C,P)$ [14] với i là nhãn dữ liệu, C là giá trị thực tế và P là giá trị dự đoán. Hàm $L(C,P)$ được mô tả theo phương trình (2).

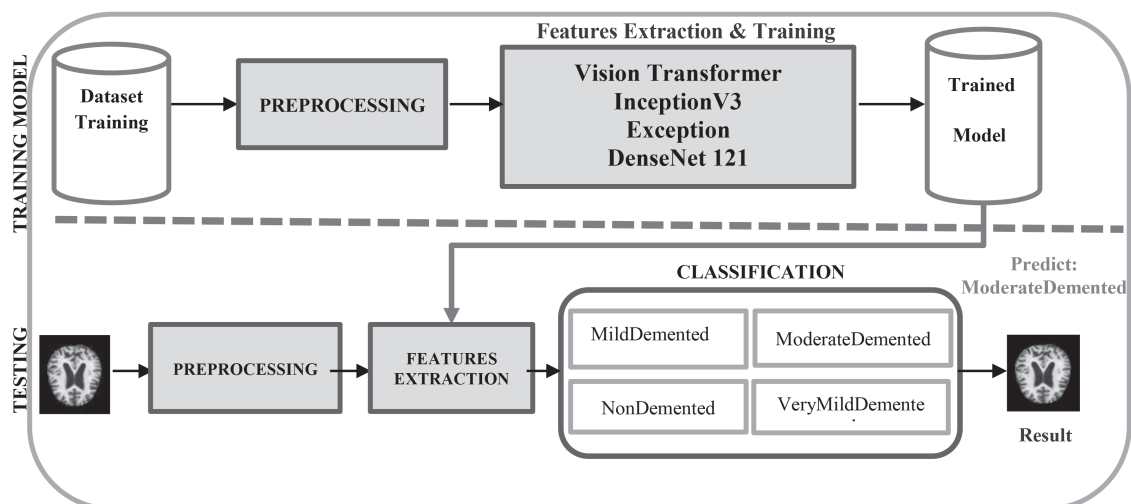
$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (1)$$

$$L(C, P) = - \sum_i C_i \log(P_i) \quad (2)$$

2. KẾT QUẢ NGHIÊN CỨU

2.1 Phương pháp đề xuất

Mô hình đề xuất gồm 2 giai đoạn là huấn luyện và kiểm thử được thể hiện ở hình 1:



Hình 1. Mô hình đề xuất phát hiện và phân loại bệnh Alzheimer

2.1.1. Đào tạo mô hình

a) Tiền xử lý và chuẩn bị dữ liệu:

Chuẩn hoá tất cả hình ảnh thành kích thước 224×224 pixel, đây là kích thước phù hợp để tạo bản vá từ hình ảnh đầu vào. Sau đó, tiến hành tăng cường dữ liệu bằng cách thay đổi góc quay hình ảnh (xoay ảnh) và lật ảnh.

b) Rút trích đặc trưng và huấn luyện:

Đầu tiên tiến hành rút trích đặc trưng đặc điểm các loại bệnh Alzheimer dựa trên ảnh MRI não, sau khi được tiền xử lý và huấn luyện bằng các kiến trúc mạng Inception V3, Xception, DenseNet 121 và Vision Transformer. Đối với Vision Transformer, hình ảnh đầu vào sau khi tiền xử lý được cắt thành các patch đưa qua tầng liner projection với kích thước mỗi patch là 16×16 [15], mỗi patch sẽ gán vị trí và

nhãn cụ thể. Tiếp đến đưa qua transform encoder để tiến hành rút trích đặc trưng mã hóa, phân loại bệnh. Kết quả của quá trình này sẽ là ảnh được phân loại thuộc vào nhãn nào.

2.1.2 Kiểm thử

Ở giai đoạn kiểm thử, hình ảnh đầu vào được chuẩn hoá thành kích thước 224×224 , sau đó tiến hành đưa qua mô hình đã được huấn luyện để tiến hành dự đoán. Kết quả kiểm thử sẽ đưa ra một trong bốn kết luận là Mild Demented, Very Mild Demented, Moderate Demented hay Non Demented.

2.2 Kết quả thực nghiệm

2.2.1. Các kịch bản áp dụng

Trong nghiên cứu này, chúng tôi thực hiện đề xuất 4 kịch bản với các tham số trong Bảng 1.

Bảng 1. Các kịch bản được đề xuất và các tham số huấn luyện

Kịch bản	Kiến trúc mạng	Mạng rút trích đặc trưng và phân loại	Số bước huấn luyện	Num classes	Learning rate	Kích thước ảnh	Batch size
1	InceptionV3	Default	1000	4	0.0001	224×224	256
2	Xception	Default	1000	4	0.0001	224×224	256
3	DenseNet 121	Default	1000	4	0.0001	224×224	256
4	Vision Transformer	Default	1000	4	0.0001	224×224	256

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

a) Môi trường cài đặt:

Phương pháp đề xuất được cài đặt trên nền tảng Kaggle với thư viện hỗ trợ đào tạo mô hình mạng sử dụng là Tensorflow và

Keras với ngôn ngữ lập trình Python.

b) Tập dữ liệu thử nghiệm:

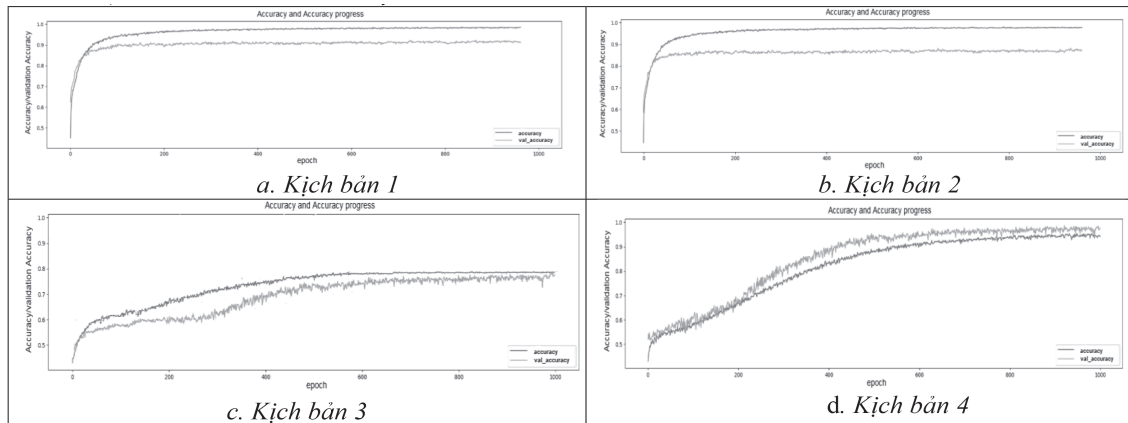
Trong nghiên cứu này, chúng tôi sử dụng bộ dữ liệu về bệnh Alzheimer [16], là bộ dữ liệu có sẵn trên kaggle. Dữ liệu bao gồm 6400 ảnh, trong đó: Mild Demented

(Mất trí nhớ nhẹ) 896 ảnh, Moderate Demented (Mất trí nhớ vừa phải) 64 ảnh, Non Demented (Không mất trí nhớ) 3200 ảnh và Very Mild Demented (Mất trí nhớ

rất nhẹ) 2240 ảnh.

2.2.3. Các kết quả thu được

a) Độ chính xác – Accuracy:

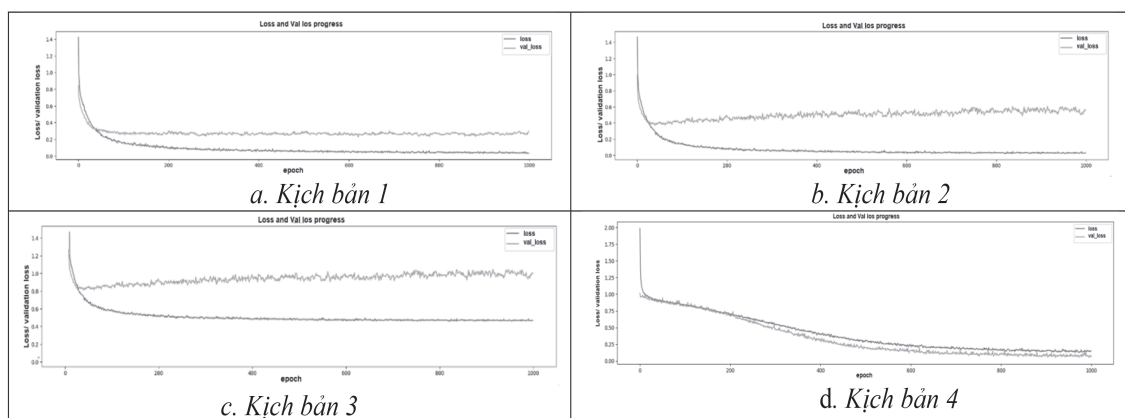


Hình 2. Độ đo Accuracy của các kịch bản trong giai đoạn huấn luyện.

Hình 2 biểu diễn độ chính xác (accuracy) của các kịch bản trong giai đoạn huấn luyện. Kịch bản 1 đạt độ chính xác 95.21%, kịch bản 2 có độ chính xác là 94.47%, kịch bản 3 có độ chính xác là 79%, kịch bản 4 độ chính xác cao hơn so với các kịch bản còn lại với độ chính xác là 96.79%. Mặt khác, khi quan sát kịch bản 1, 2 và 3 ta thấy rằng, giá trị val

accuracy cho kết quả thấp hơn giá trị train accuracy, ngược lại kịch bản 4 lại cho kết quả tối ưu hơn so với các kịch bản trên khi giá trị val accuracy cao hơn so với giá trị train accuracy. Điều này cho thấy mô hình trong kịch bản 4 có khả năng dự đoán tốt hơn các mô hình trong ba kịch bản còn lại.

b) Độ mất mát – Loss:



Hình 3. Độ đo Loss của các kịch bản trong giai đoạn huấn luyện

Hình 3 biểu diễn giá trị mất mát (loss) của các kịch bản trong giai đoạn huấn luyện. Giá trị val loss của các kịch bản lần

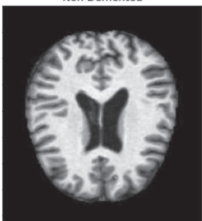
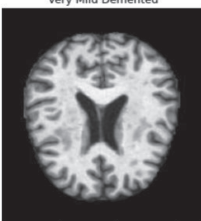
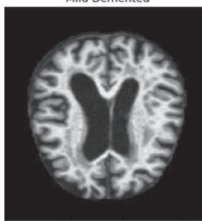
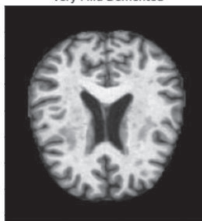
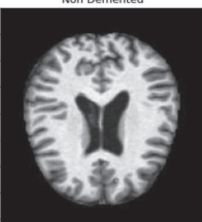
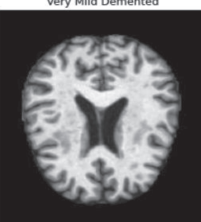
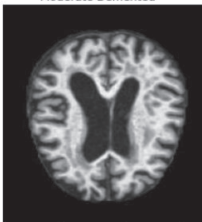
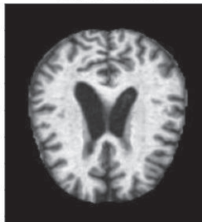
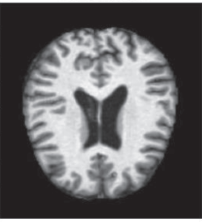
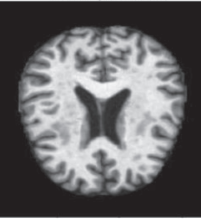
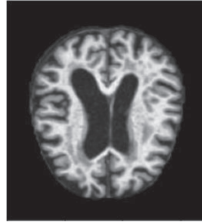
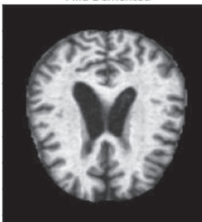
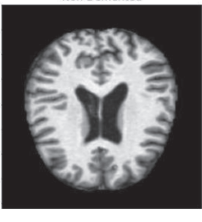
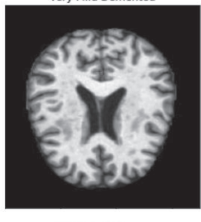
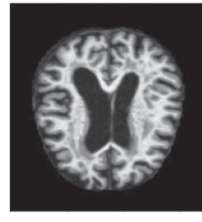
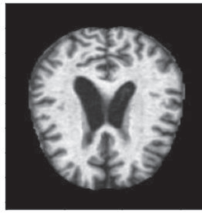
lượt là kịch bản 1 là 0.3645, kịch bản 2 là 0.5779, kịch bản 3 là 0.9769, kịch bản 4 là 0.0996. Ở kịch bản 1, 2 và 3, giá trị

val_loss cho kết quả cao hơn giá trị train loss. Ngược lại, kịch bản 4 lại cho kết quả tối ưu hơn so với các kịch bản trên vì giá trị val loss thấp hơn giá trị train loss. Điều

này cho thấy mô hình trong kịch bản 4 cho kết quả tốt hơn các mô hình trong các kịch bản còn lại.

c) Một số hình ảnh kết quả phân loại:

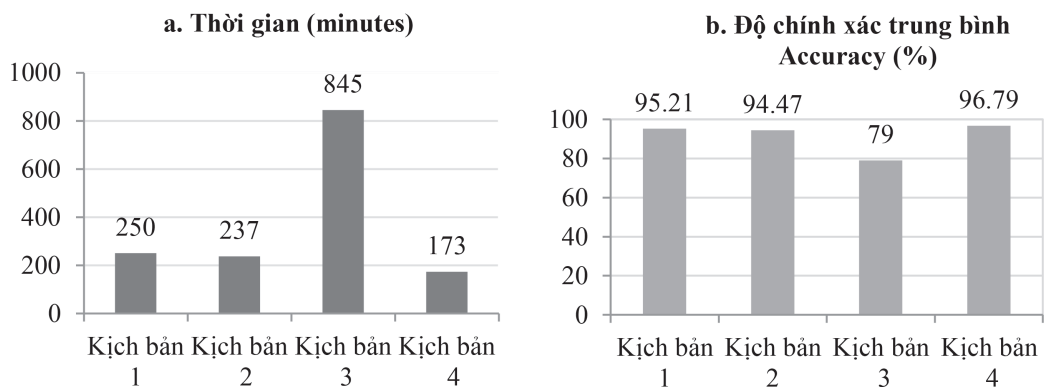
Bảng 2. Một số kết quả trên tập dữ liệu kiểm thử

	Actual: Non Demented	Actual: Very Mild Demented	Actual: Mild Demented	Actual: Moderate Demented
Kịch bản 1	Non Demented  Accuracy: 85.39%	Very Mild Demented  Accuracy: 85.39%	Mild Demented  Accuracy: 83.35%	Very Mild Demented  Accuracy: 70.05%
Kịch bản 2	Non Demented  Accuracy: 75.94%	Very Mild Demented  Accuracy: 75.94%	Moderate Demented  Accuracy: 75.94%	Very Mild Demented  Accuracy: 58.33%
Kịch bản 3	Mild Demented  Accuracy: 60.08%	Very Mild Demented  Accuracy: 60.08%	Very Mild Demented  Accuracy: 40.01%	Mild Demented  Accuracy: 50.00%
Kịch bản 4	Non Demented  Accuracy: 99%	Very Mild Demented  Accuracy: 99%	Mild Demented  Accuracy: 99%	Moderate Demented  Accuracy: 96.94%

Bảng 2 minh họa một số hình ảnh thực nghiệm của các kịch bản đã được đề xuất. Kịch bản 1 với mô hình Inception V3 dự đoán đúng nhãn thực tế là Non Demented, Very Mild Demented, Mild Demented, kịch bản 2 dự đoán đúng các nhãn Non

Demented, Very Mild Demented. Riêng kịch bản 3 chỉ dự đoán đúng nhãn Very Mild Demented. Trong khi đó kịch bản 4 dự đoán chính xác tất cả các nhãn. Kết quả cho thấy kịch bản 4 hoạt động tốt hơn hẳn các kịch bản còn lại.

d) So sánh, đánh giá các mô hình phân loại bệnh



Hình 4. Thời gian đào tạo và độ chính xác của các kịch bản

Hình 4a biểu diễn thời gian đào tạo của các kịch bản được đề xuất. Ở kịch bản 1 là 250 phút, kịch bản 2 là 237 phút, thời gian đào tạo khá tương đồng. Kịch bản 3 tốn khá nhiều thời gian để đào tạo 845 phút. Trong khi đó, kịch bản 4 tốn ít thời gian đào tạo nhất với 173 phút. Hình 4b biểu diễn độ chính xác của các kịch bản đề xuất lần lượt là kịch bản 1 là 95.21%, kịch bản 2 là 94.47%, kịch bản 3 là 79%, kịch bản

4 là 96.79%. Qua kết quả trên, có thể thấy mô hình 4 cho kết quả đào tạo nhanh nhất trong các kịch bản được đề xuất, bên cạnh đó, độ chính xác kịch bản 4 cũng cao hơn so với các kịch bản còn lại. Điều này cho thấy, kịch bản 4 mang lại kết quả dự đoán tốt hơn so với các kịch bản còn lại.

2.2.4. Đánh giá với các thực nghiệm gần đây

	Author	Classes	Dataset Source	No. of Image	Technique	Accuracy (%)
Mô hình 1	Menagadevi et al	4	ADNI	6400	DEMNET	95.23
Mô hình 2	S. Sharma et al	4	Kaggle	6200	DenseNet-121, Densenet-201, SVM, Naive base and XGBoost	91.75
Mô hình 3	R. A. Hazarika et al	4	ADNI	5000	LeNet AlexNet	95.23
Mô hình 4	Z. Hu et al	2	ADNI	-	VGG-16 TSwinformer	77.2
Mô hình 5	Trần Thị Thu Ngân	4	Kaggle	6400	Vision Transformer	96.7

Dựa vào bảng số liệu, mô hình 5 có hiệu suất cao và thậm chí vượt trội so với 4 mô hình còn lại trong bảng dữ liệu, đặc biệt là trong việc phân loại bệnh AD. Tóm

lại, mô hình Vision Transformer cho thấy hiệu suất tốt và cung cấp kết quả ấn tượng trong việc phân loại bệnh AD trên hình ảnh MRI là bình thường, nhẹ, rất nhẹ hoặc

trung bình, vượt trội so với nhiều mô hình khác trong bảng dữ liệu.

3. KẾT LUẬN

Trong bài báo này, chúng tôi đề xuất phương pháp phát hiện và chẩn đoán bệnh Alzheimer dựa trên các mạng rút trích đặc trưng như Inception V3, Xception, Densenet 121 và Vision Transformer. Ngoài ra, chúng tôi cũng mang lại một số kết quả so sánh giữa các mô hình, điều này góp phần trong việc đánh giá và lựa chọn

được phương pháp phù hợp để có thể xây dựng hệ thống phân loại nhằm hỗ trợ bác sĩ có thể đưa ra chẩn đoán một cách nhanh chóng để có thể tiến hành điều trị kịp thời cho bệnh nhân. Kết quả thực nghiệm cho thấy mô hình phân loại bệnh Alzheimer dựa trên mạng Vision Transformer đạt hiệu suất tốt hơn về mặt thời gian dự đoán với độ chính xác cao hơn so với các mạng học sâu khác. Mặt khác, độ chính xác của mô hình đề xuất trong quá trình đào tạo lên đến 96,7%.

TÀI LIỆU THAM KHẢO

- [1] <https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset-v2>.
- [2] A. K. Desai and G. T. Grossberg, “Diagnosis and Treatment of Alzheimer’s Disease”, *Neurology*, vol. 64 (Suppl. 3), p. S34–S39, 2005.
- [3] Tổ chức Y tế Thế giới. <https://www.who.int/publications/i/item/9789240033245>. [Truy cập: 11/ 11/ 2023]
- [4] M. Menagadevi and S. Mangai and N. Madian and D. Thiagarajan, “Automated Prediction System for Alzheimer Detection Based on Deep Residual Autoencoder and Support Vector Machine”, *Optik*, vol. 272, p. 170212, 2023.
- [5] S. Sharma and S. Gupta and D. Gupta and S. R. Nayak, “HTLML: Hybrid AI Based Model for Detection of Alzheimer’s Disease”, *Diagnostics*, vol. 12, p. 1833, 2022.
- [6] R. A. Hazarika and A. K. Maji and D. Kandar and E. Jasinska and P. Krejci and Z. Leonowicz and M. Jasinski, “An Approach for Classification of Alzheimer’s Disease Using Deep Neural Network and Brain Magnetic Resonance Imaging (MRI)”, *Electronics*, vol. 12, p. 676, 2023.
- [7] Z. Hu and Z. Wang and Y. Jin and W. Hou, “VGG-TSwinformer: Transformer - Based Deep Learning Model for Early Alzheimer’s Disease Prediction”, *Computer Methods Programs Biomed*, vol. 229, p. 107291, 2023.
- [8] <https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease> [Truy cập: 11/ 07/ 2023].
- [9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna (2016), “Rethinking the Inception Architecture for Computer Vision,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
- [10] François Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”,

- arXiv: 1610.02357v3, <https://doi.org/10.48550/arXiv.1610.02357>.
- [11] G.Huang and Z. Liu and Weinberger,... Q. Kilian, “Densely Connected Convolutional Networks”, *arXiv.org*, arxiv.org/abs/1608.06993, 2016.
- [12] M. Menagadevi and S. Mangai and N. Madian and D. Thiagarajan, “Automated Prediction System for Alzheimer Detection Based on Deep Residual Autoencoder and Support Vector Machine”, *Optik*, vol. 272, p. 170212, 2023.
- [13] Rethinking Model Scaling for Convolutional Neural Networks” của Mingxing Tan và Quoc V. Le, arxiv.org/abs/1905.11946.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition”, IEEE Computer Vision and Pattern Recognition 2016, doi: 10.1109/CVPR.2016.90.
- [15] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *arXiv.org*, arXiv.1409.1556, 2014.
- [16] Iqbal Z, Luo D, Henry P, Kazemifar S, Rozario T, Yan Y, Westover K, Lu W, Nguyen D, Long T, Wang J, Choy H, Jiang S. (2018), “Accurate real time localization tracking in a clinical environment using Bluetooth Low Energy and deep learning”, PLoS One. 2018 Oct 11;13(10):e0205392. doi: 10.1371/journal.pone.0205392.
- [17] Gulli, A.; Kapoor, A.; Pal, S. (2019), “Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API”, Packt Publishing Ltd.: Birmingham, UK, 2019.
- [18] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv 2020, arXiv:2010.11929.