Tạp chí Khoa học và Công nghệ Trường Đại học Công nghệ Đông Á

So sánh hiệu quả các kỹ thuật biểu diễn ảnh hai chiều từ dữ liệu chuỗi thời gian một chiều cho bài toán phân loại bằng mạng nơ-ron tích chập nhẹ

Comparative Evaluation of 2D Image Representations from 1D Time Series for Classification Using Lightweight Convolutional Neural Networks

Trần Nguyên Hoàng[1]*
[1] Khoa Công nghệ thông tin, Trường Đại học Công nghệ Đông Á
*Email: hoangtn@eaut.edu.vn

Abstract

Time series classification is an important task in sensor monitoring, healthcare, industrial diagnosis and financial analytics. Besides models that directly process one-dimensional sequences, transforming time series into two-dimensional images allows convolutional neural networks to exploit spatial patterns from encoded signals. This study does not propose a new time-series imaging transform; instead, it provides a controlled and reproducible comparison of four common two-dimensional representations, namely Gramian Angular Field, Markov Transition Field, Recurrence Plot and short-time Fourier transform images, using the same lightweight convolutional neural network. A 1D-CNN was used as the direct baseline. Experiments were conducted on five public datasets from the UCR Time Series Archive, including Coffee, ECG200, FordA, GunPoint and Wafer. The main experimental setting used an image size of 64 × 64, 50 training epochs, a batch size of 32 and three random seeds. Repeated runs under the same seed were averaged before final aggregation to improve result stability. The results show that no two-dimensional representation was consistently superior across all datasets. The 1D-CNN baseline achieved the highest overall average accuracy and Macro F1-score, reaching 74.57% and 62.70%, respectively. Among the two-dimensional representations, Recurrence Plot obtained the best average performance, with 71.88% accuracy and 59.20% Macro F1-score, and achieved the best results on FordA and Wafer. Although the Light 2D-CNN configurations used fewer trainable parameters than the 1D-CNN baseline, they required longer average training time. These findings indicate that time-series imaging can be useful for datasets with recurrent or dynamical structures; however, representation selection should depend on signal characteristics, model complexity and deployment constraints.

Keywords: *convolutional neural network, Gramian angular field, lightweight model, recurrence plot, time series classification.*

Tóm tắt

Phân loại chuỗi thời gian là bài toán quan trọng trong giám sát cảm biến, y tế, công nghiệp và phân tích dữ liệu tài chính. Bên cạnh hướng xử lý trực tiếp chuỗi một chiều, chuyển đổi chuỗi thời gian thành ảnh hai chiều cho phép mạng nơ-ron tích chập khai thác cấu trúc không gian của tín hiệu. Nghiên cứu này không đề xuất một phép biến đổi ảnh mới, mà tập trung xây dựng một khung so sánh có kiểm soát và có khả năng tái lập giữa bốn kỹ thuật biểu diễn ảnh hai chiều phổ biến, gồm Gramian Angular Field, Markov Transition Field, Recurrence Plot và ảnh phổ thời gian - tần số dựa trên biến đổi Fourier thời gian ngắn. Các biểu diễn ảnh được đưa vào cùng một kiến trúc Light 2D-CNN, trong khi mô hình 1D-CNN được dùng làm đối chứng với chuỗi gốc. Thực nghiệm được tiến hành trên năm bộ dữ liệu thuộc UCR Time Series Archive, gồm Coffee, ECG200, FordA, GunPoint và Wafer. Thiết lập thực nghiệm sử dụng ảnh đầu vào 64 × 64, 50 vòng huấn luyện, kích thước lô 32 và ba giá trị khởi tạo ngẫu

DOI: https://doi.org/10.65153/geb69058

nhiên; các lần chạy lặp trong cùng một seed được lấy trung bình trước khi tổng hợp. Kết quả cho thấy không có kỹ thuật biểu diễn hai chiều nào vượt trội trên mọi bộ dữ liệu. Mô hình 1D-CNN đạt hiệu năng cao nhất với độ chính xác 74,57% và Macro F1-score 62,70%. Trong nhóm biểu diễn hai chiều, Recurrence Plot đạt kết quả tốt nhất với độ chính xác 71,88% và Macro F1-score 59,20%, đồng thời đạt hiệu quả cao nhất trên FordA và Wafer. Mặc dù các mô hình hai chiều nhẹ có số tham số thấp hơn 1D-CNN, thời gian huấn luyện lại dài hơn. Kết quả cho thấy biểu diễn ảnh hai chiều có thể hữu ích với dữ liệu có cấu trúc lặp hoặc động lực học, nhưng cần được lựa chọn theo đặc điểm tín hiệu, độ phức tạp mô hình và ràng buộc triển khai.

Từ khóa: *biểu diễn ảnh hai chiều, chuỗi thời gian, mạng nơ-ron tích chập nhẹ, phân loại, Recurrence Plot.*

1. Giới thiệu

Dữ liệu chuỗi thời gian một chiều xuất hiện phổ biến trong nhiều lĩnh vực như cảm biến, y tế, công nghiệp, mạng máy tính và tài chính. Bài toán phân loại chuỗi thời gian nhằm gán nhãn cho toàn bộ chuỗi dựa trên đặc trưng hình dạng, xu hướng, trạng thái chuyển tiếp hoặc mẫu lặp của tín hiệu. Tuy nhiên, đây là bài toán khó do dữ liệu thường nhiễu, lệch pha, mất cân bằng lớp, số lượng mẫu hạn chế và có sự khác biệt lớn giữa các miền ứng dụng [7], [9].

Trong những năm gần đây, mạng nơ-ron tích chập đã được ứng dụng hiệu quả cho phân loại chuỗi thời gian nhờ khả năng tự động học đặc trưng cục bộ và tổng hợp đặc trưng theo nhiều tầng. Có hai hướng tiếp cận chính: xử lý trực tiếp chuỗi một chiều bằng 1D-CNN hoặc chuyển đổi chuỗi thành ảnh hai chiều để khai thác 2D-CNN. Hướng thứ hai có tiềm năng vì tận dụng được các kỹ thuật xử lý ảnh, nhưng hiệu quả phụ thuộc mạnh vào cách biểu diễn dữ liệu đầu vào.

Một số kỹ thuật biểu diễn ảnh hai chiều tiêu biểu gồm Gramian Angular Field (GAF), Markov Transition Field (MTF), Recurrence Plot (RP) và Short-Time Fourier Transform (STFT). Các kỹ thuật này được lựa chọn vì đại diện cho bốn cách mã hóa bổ sung của chuỗi thời gian: quan hệ hình học giữa các điểm thời gian, xác suất chuyển trạng thái, cấu trúc tái xuất hiện và đặc trưng thời gian - tần số [1], [2], [3]. Các nghiên cứu gần đây tiếp tục cho thấy hướng time-series imaging được ứng dụng trong nhiều bối cảnh cảm biến, y sinh, kỹ thuật và dự báo, nhưng hiệu quả còn phụ thuộc mạnh vào biểu diễn đầu vào, kiến trúc CNN và đặc điểm bộ dữ liệu [10], [12], [13]. Vì vậy, cần có thực nghiệm so sánh trong cùng điều kiện mô hình để đánh giá biểu diễn nào phù hợp hơn với từng loại dữ liệu.

Bên cạnh độ chính xác, chi phí mô hình cũng là yếu tố quan trọng trong các ứng dụng thực tế. Các mô hình sâu hoặc ensemble thường đạt hiệu quả cao nhưng khó triển khai trong môi trường tài nguyên hạn chế. Do đó, nghiên cứu này sử dụng kiến trúc Light 2D-CNN nhằm

đánh giá hiệu quả phân loại của các biểu diễn ảnh hai chiều trong điều kiện mô hình gọn nhẹ, dễ tái lập và có số tham số thấp.

Từ các phân tích trên, bài báo đặt ra câu hỏi nghiên cứu: trong cùng một kiến trúc mạng tích chập nhẹ, các kỹ thuật biểu diễn ảnh hai chiều từ chuỗi thời gian một chiều khác nhau như thế nào về hiệu quả phân loại và chi phí mô hình? Nghiên cứu so sánh bốn kỹ thuật GAF, MTF, RP và STFT trên năm bộ dữ liệu UCR gồm Coffee, ECG200, FordA, GunPoint và Wafer. Các kết quả được đối chiếu với mô hình 1D-CNN xử lý trực tiếp chuỗi gốc.

Bài toán nghiên cứu được xem xét trên tập gồm N mẫu chuỗi thời gian một chiều, mỗi mẫu $X_i = (x_1, x_2, ..., x_t)$ có nhãn y_i thuộc C lớp. Mục tiêu là học hàm ánh xạ từ chuỗi đầu vào sang phân bố xác suất trên các lớp để dự đoán nhãn cho toàn bộ chuỗi. Phạm vi thực nghiệm giới hạn ở chuỗi thời gian đơn biến nhằm kiểm soát rõ tác động của kỹ thuật biểu diễn đầu vào.

Đóng góp chính của bài báo không nằm ở việc đề xuất một phép biến đổi mới, mà ở việc làm rõ giá trị thực nghiệm của các biểu diễn ảnh hai chiều trong một bối cảnh so sánh thống nhất. Cụ thể, bài báo: (i) hệ thống hóa và luận giải vai trò của bốn nhóm biểu diễn GAF, MTF, RP và STFT; (ii) xây dựng khung thực nghiệm so sánh công bằng giữa các biểu diễn hai chiều trên cùng kiến trúc Light 2D-CNN và một mô hình 1D-CNN đối chứng; (iii) đánh giá đồng thời độ chính xác, Macro F1-score, số tham số và thời gian huấn luyện; và (iv) rút ra hàm ý lựa chọn biểu diễn theo đặc điểm tín hiệu thay vì mặc định rằng chuyển đổi sang ảnh hai chiều luôn cải thiện hiệu năng.

2. Phương pháp nghiên cứu

2.1. Phạm vi nghiên cứu và giả thuyết

Nghiên cứu tập trung vào bài toán phân loại chuỗi thời gian đơn biến. Việc giới hạn ở dữ liệu đơn biến giúp kiểm soát rõ hơn ảnh hưởng của kỹ thuật biểu diễn đầu vào, đồng thời bảo đảm tính nhất quán khi so sánh giữa các phương pháp. Các bộ dữ liệu được lựa chọn từ UCR Time Series Archive, một kho dữ liệu chuẩn được sử dụng rộng rãi trong cộng đồng nghiên cứu phân loại chuỗi thời gian [7].

Trong phạm vi bài báo, dữ liệu đầu vào được xử lý theo hai hướng. Hướng thứ nhất giữ nguyên chuỗi thời gian một chiều và đưa trực tiếp vào mô hình 1D-CNN, đóng vai trò đối chứng. Hướng thứ hai chuyển đổi chuỗi thời gian thành ảnh hai chiều bằng GAF, MTF, RP và STFT, sau đó sử dụng cùng kiến trúc Light 2D-CNN để phân loại. Cách thiết kế này giúp tách riêng ảnh hưởng của kỹ thuật biểu diễn khỏi ảnh hưởng của kiến trúc mô hình.

Đối với dữ liệu chuỗi thời gian đa biến, hướng tiếp cận trong bài báo có thể được mở rộng bằng cách chuyển mỗi biến thành một kênh ảnh riêng, ghép nhiều biểu diễn thành ảnh đa kênh hoặc hợp nhất đặc trưng giữa các biến ở các tầng sau của mạng. Các mở rộng này chưa được thực nghiệm trong bài báo hiện tại để giữ thiết kế nghiên cứu tập trung vào tác động của biểu diễn hai chiều đối với chuỗi thời gian một chiều.

Từ phạm vi nghiên cứu trên, bài báo đặt ra ba giả thuyết chính. Thứ nhất, biểu diễn ảnh hai chiều có thể hỗ trợ mô hình 2D-CNN học tốt hơn với dữ liệu có cấu trúc lặp, quan hệ dài hạn hoặc biến đổi tần số rõ rệt. Thứ hai, 1D-CNN có thể là đối chứng mạnh khi thông tin thứ tự cục bộ của chuỗi gốc đóng vai trò quyết định. Thứ ba, kiến trúc tích chập nhẹ, đặc biệt là tích chập tách biệt theo chiều sâu, có thể giảm số tham số, nhưng hiệu quả thực tế cần được đánh giá đồng thời với độ chính xác, Macro F1-score và thời gian huấn luyện.

2.2. Các kỹ thuật biểu diễn ảnh hai chiều từ chuỗi thời gian một chiều

Trong nghiên cứu này, bốn kỹ thuật biểu diễn ảnh hai chiều từ chuỗi thời gian một chiều được sử dụng để so sánh, gồm GAF, MTF, RP và STFT. Việc lựa chọn bốn kỹ thuật này dựa trên ba tiêu chí: mức độ phổ biến trong các nghiên cứu mã hóa chuỗi thời gian thành ảnh, khả năng đại diện cho các nhóm đặc trưng khác nhau của tín hiệu và tính khả thi khi triển khai bằng các thư viện thông dụng. Các kỹ thuật này đều chuyển chuỗi thời gian thành ma trận hoặc ảnh hai chiều, nhưng mỗi phương pháp nhấn mạnh một đặc trưng khác nhau của tín hiệu.

GAF mã hóa chuỗi thời gian đã chuẩn hóa vào hệ tọa độ cực, sau đó xây dựng ma trận Gramian để biểu diễn quan hệ góc giữa các điểm thời gian. Phương pháp này có ưu điểm là bảo toàn thông tin trật tự thời gian và tạo ra cấu trúc ảnh phù hợp cho mạng tích chập hai chiều [1], [2].

MTF biểu diễn chuỗi thời gian thông qua xác suất chuyển trạng thái. Chuỗi được lượng tử hóa thành các trạng thái rời rạc, sau đó ma trận chuyển trạng thái Markov được ánh xạ lên các cặp thời điểm để tạo thành ảnh hai chiều. Biểu diễn này phù hợp với dữ liệu có đặc trưng chuyển trạng thái rõ rệt, chẳng hạn tín hiệu vận hành thiết bị hoặc dữ liệu y sinh [1].

RP mô tả sự tái xuất hiện của các trạng thái trong không gian pha bằng cách đánh giá mức độ tương đồng giữa các vector trạng thái trễ. Do đó, RP phù hợp với các tín hiệu có tính lặp, tính chu kỳ hoặc cấu trúc động lực học phi tuyến [3].

STFT tạo ảnh miền thời gian - tần số bằng cách chia tín hiệu thành các cửa sổ ngắn và phân tích thành phần tần số trong từng cửa sổ. Phương pháp này phù hợp với các tín hiệu có

thành phần dao động thay đổi theo thời gian, như âm thanh, rung động máy móc hoặc tín hiệu điện sinh học. Tuy nhiên, hiệu quả của STFT phụ thuộc vào kích thước cửa sổ, độ chồng lấn, độ phân giải ảnh và cách chuẩn hóa phổ.

Nhìn chung, GAF nhấn mạnh quan hệ hình học giữa các điểm thời gian, MTF tập trung vào quy luật chuyển trạng thái, RP biểu diễn cấu trúc tái xuất hiện của trạng thái, còn STFT làm nổi bật sự biến đổi tần số theo thời gian. Sự khác biệt này là lý do bài báo không lựa chọn một biểu diễn duy nhất, mà so sánh đồng thời bốn hướng mã hóa để xem biểu diễn nào phù hợp hơn với từng loại dữ liệu. Trong cài đặt thực nghiệm, thư viện pyts được sử dụng để tạo các biểu diễn GAF, MTF và RP [8].

Bảng 1. So sánh khái quát các kỹ thuật biểu diễn ảnh hai chiều từ chuỗi thời gian một chiều

Kỹ thuật	Thông tin được nhấn mạnh	Ưu điểm	Hạn chế cần kiểm soát
GAF/GASF/GADF	Quan hệ góc và tương quan giữa các thời điểm	Giữ trật tự thời gian, ảnh có cấu trúc rõ	Nhạy với chuẩn hóa và kích thước ảnh
MTF	Xác suất chuyển trạng thái giữa các mức tín hiệu	Phù hợp dữ liệu có trạng thái chuyển tiếp	Phụ thuộc số mức lượng tử hóa
RP	Mẫu lặp lại và tính tái xuất hiện của trạng thái	Tốt cho tín hiệu chu kỳ hoặc động lực học	Cần chọn ngưỡng và embedding hợp lý
STFT	Năng lượng theo thời gian - tần số	Tốt với tín hiệu dao động, âm thanh, rung động	Phụ thuộc kích thước cửa sổ, độ chồng lấn và độ phân giải

2.3. Kiến trúc mô hình phân loại

2.3.1. Mạng 1D-CNN đối chứng.

Mạng 1D-CNN được sử dụng làm mô hình đối chứng nhằm đánh giá hiệu quả của việc học đặc trưng trực tiếp từ chuỗi thời gian gốc, không qua bước chuyển đổi sang ảnh hai chiều. Mô hình nhận đầu vào là chuỗi một chiều có kích thước T × 1, trong đó T là độ dài chuỗi của từng bộ dữ liệu. Kiến trúc gồm hai khối tích chập một chiều, chuẩn hóa batch, một lớp gộp cực đại, một lớp gộp trung bình toàn cục, dropout và tầng phân loại softmax. Cấu hình chi tiết của mạng 1D-CNN được trình bày trong Bảng 2a.

Bảng 2a. Cấu hình kiến trúc mạng 1D-CNN đối chứng

Tầng/khối	Cấu hình	Mục đích
Input	Chuỗi một chiều kích thước T × 1	Nhận trực tiếp chuỗi thời gian gốc
Conv1D block 1	Conv1D, 32 filters, kernel size 5, padding same, ReLU	Học đặc trưng cục bộ ban đầu trên chuỗi
Batch Normalization 1	BatchNorm sau lớp Conv1D thứ nhất	Ổn định quá trình huấn luyện
Pooling	MaxPooling1D, pool size 2	Giảm chiều chuỗi và chi phí tính toán
Conv1D block 2	Conv1D, 64 filters, kernel size 3, padding same, ReLU	Học đặc trưng mức cao hơn
Batch Normalization 2	BatchNorm sau lớp Conv1D thứ hai	Ổn định phân phối đặc trưng

Tầng/khối	Cấu hình	Mục đích
Global pooling	GlobalAveragePooling1D	Tổng hợp đặc trưng theo toàn bộ chiều thời gian
Regularization	Dropout 0,3	Giảm nguy cơ quá khớp
Classifier	Dense C, Softmax	Phân loại chuỗi vào C lớp
Tổng số tham số	6.914 với C = 2	Đối chiếu với các mô hình Light 2D-CNN

Ghi chú: Với các bộ dữ liệu nhị phân trong nghiên cứu này, C = 2; tổng số tham số của mô hình 1D-CNN là 6.914, gồm 6.722 tham số huấn luyện và 192 tham số không huấn luyện do các lớp Batch Normalization.

Khác với Light 2D-CNN, mô hình 1D-CNN không sử dụng thông tin không gian từ ảnh biểu diễn hai chiều mà học trực tiếp các mẫu cục bộ trên chuỗi gốc. Do đó, mô hình này đóng vai trò đối chứng trực tiếp để đánh giá liệu việc chuyển đổi chuỗi thời gian sang ảnh hai chiều có thực sự cải thiện hiệu quả phân loại trong cùng điều kiện huấn luyện hay không.

Mục tiêu của nghiên cứu không phải là xây dựng mô hình sâu hoặc tối ưu độ chính xác tuyệt đối, mà là đánh giá ảnh hưởng của các kỹ thuật biểu diễn đầu vào đối với bài toán phân loại chuỗi thời gian. Vì vậy, mô hình được thiết kế theo hướng gọn nhẹ, dễ cài đặt, dễ tái lập và có số tham số nhỏ, nhằm bảo đảm việc so sánh giữa các biểu diễn ảnh hai chiều được công bằng hơn.

Nghiên cứu sử dụng mạng nơ-ron tích chập một chiều (1D-CNN) làm mô hình đối chứng và mạng nơ-ron tích chập hai chiều nhẹ (Light 2D-CNN) cho các biểu diễn ảnh hai chiều. Light 2D-CNN nhận đầu vào là ảnh hai chiều được tạo từ chuỗi thời gian một chiều bằng các kỹ thuật Gramian Angular Field, Markov Transition Field, Recurrence Plot và Short-Time Fourier Transform. Kiến trúc gồm một khối tích chập hai chiều chuẩn, hai khối tích chập tách biệt theo chiều sâu, tầng gộp trung bình toàn cục và tầng phân loại softmax. Việc sử dụng tích chập tách biệt theo chiều sâu giúp giảm số tham số nhưng vẫn duy trì khả năng học đặc trưng không gian từ ảnh đầu vào.

Mô hình 1D-CNN xử lý trực tiếp chuỗi gốc mà không qua bước mã hóa ảnh. Các mô hình được huấn luyện trong cùng điều kiện về số epoch, batch size, bộ tối ưu và quy trình đánh giá nhằm hạn chế sai lệch do thiết lập thực nghiệm.

2.3.2. *Mạng Light 2D-CNN cho biểu diễn ảnh hai chiều*

Trong thiết lập chính, kích thước ảnh đầu vào được chuẩn hóa về **64 × 64**. Kích thước này phù hợp với mục tiêu xây dựng mô hình nhẹ, giúp giảm chi phí tính toán, hạn chế quá khớp

và bảo đảm khả năng thực nghiệm trên môi trường tính toán phổ thông. Bảng 2b trình bày kiến trúc tổng quát của mô hình Light 2D-CNN được sử dụng trong nghiên cứu.

Nhìn chung, kiến trúc Light 2D-CNN được lựa chọn nhằm bảo đảm sự cân bằng giữa hiệu quả phân loại và chi phí tính toán. Việc sử dụng các khối tích chập tách biệt theo chiều sâu giúp giảm số tham số, trong khi tầng gộp trung bình toàn cục góp phần hạn chế overfitting và làm mô hình gọn hơn [5], [6]. Do đó, mô hình này phù hợp với mục tiêu của nghiên cứu là đánh giá ảnh hưởng của các kỹ thuật biểu diễn đầu vào trong một điều kiện mạng nhẹ, dễ cài đặt và có khả năng tái lập.

Trong nghiên cứu này, các siêu tham số kiến trúc như số lớp, số block, số bộ lọc, kích thước kernel và tỷ lệ dropout được giữ cố định cho toàn bộ bộ dữ liệu. Cách thiết lập này nhằm ưu tiên mục tiêu so sánh ảnh hưởng của biểu diễn đầu vào trong cùng một điều kiện mô hình, thay vì tối ưu hiệu năng riêng cho từng bộ dữ liệu. Do đó, kết quả phản ánh hiệu quả của các biểu diễn trong một cấu hình Light 2D-CNN thống nhất; việc hiệu chỉnh sâu hơn các siêu tham số kiến trúc được xem là một hướng mở rộng.

Bảng 2b. Kiến trúc đề xuất cho mô hình Light 2D-CNN

Tầng/khối	Cấu hình đề xuất	Mục đích
Input	Ảnh hai chiều kích thước 64 × 64, 1 kênh	Nhận biểu diễn GAF, MTF, RP hoặc ảnh thời gian - tần số
Conv block 1	Conv2D 3 × 3, 16 filters, BatchNorm, ReLU, MaxPooling	Học đặc trưng cục bộ cấp thấp
DS block 2	DepthwiseConv2D 3 × 3 + PointwiseConv2D 1 × 1, 32 filters	Giảm tham số, học đặc trưng trung gian
DS block 3	DepthwiseConv2D 3 × 3 + PointwiseConv2D 1 × 1, 64 filters	Tổng hợp đặc trưng sâu hơn
Classifier	GlobalAveragePooling, Dropout 0,3, Dense C, Softmax	Phân loại C lớp và giảm overfitting

3. Cài đặt thực nghiệm

3.1. Quy trình thực nghiệm

Quy trình thực nghiệm gồm các bước chính: lựa chọn bộ dữ liệu, tiền xử lý dữ liệu, chuyển đổi chuỗi thời gian một chiều thành ảnh hai chiều, huấn luyện mô hình, đánh giá kết quả và phân tích kết quả. Các bộ dữ liệu được chọn từ UCR Time Series Archive nhằm bảo đảm tính chuẩn hóa và khả năng đối chiếu với các nghiên cứu trước về phân loại chuỗi thời gian [7], [9]. Dữ liệu được kiểm tra giá trị thiếu, chuẩn hóa bằng z-score hoặc min-max và đưa về định dạng đầu vào thống nhất cho mô hình.

Ở bước chuyển đổi, mỗi chuỗi thời gian được biểu diễn thành ảnh hai chiều bằng các kỹ thuật GAF, MTF, RP và STFT. Các biểu diễn GAF, MTF và RP dựa trên các phương pháp mã

hóa chuỗi thời gian thành ảnh đã được công bố trước đây [1], [2], [3] và có thể triển khai bằng thư viện pyts [8]. Để bảo đảm so sánh công bằng, toàn bộ ảnh đầu vào được chuẩn hóa về cùng kích thước. Khi độ dài chuỗi lớn hơn kích thước ảnh yêu cầu, nghiên cứu sử dụng Piecewise Aggregate Approximation hoặc lấy mẫu lại để giảm chiều dữ liệu [11].

Sau đó, mô hình Light 2D-CNN được huấn luyện trên từng loại biểu diễn ảnh, trong khi mô hình 1D-CNN được sử dụng làm đối chứng với chuỗi gốc. Kết quả được đánh giá bằng độ chính xác, macro F1-score, số tham số, thời gian huấn luyện. Việc sử dụng đồng thời độ chính xác và macro F1-score giúp phản ánh đầy đủ hơn hiệu quả phân loại, đặc biệt trong trường hợp dữ liệu có khả năng mất cân bằng lớp [9]. Cách đánh giá này cho phép phân tích mô hình không chỉ theo độ chính xác mà còn theo chi phí tính toán và khả năng triển khai thực tế.



Hình 1. Quy trình thực nghiệm so sánh biểu diễn hai chiều cho phân loại chuỗi thời gian bằng mạng tích chập nhẹ

3.2. Bộ dữ liệu thực nghiệm

Thực nghiệm được tiến hành trên năm bộ dữ liệu thuộc UCR Time Series Archive, gồm GunPoint, ECG200, Coffee, FordA và Wafer, đại diện cho các miền dữ liệu chuyển động, y sinh, phổ/cảm biến, công nghiệp và sản xuất bán dẫn [7]. Bảng 3 trình bày thêm tên lớp hoặc ý nghĩa nhãn để làm rõ đặc trưng phân loại của từng bộ dữ liệu.

Bảng 3. Danh sách bộ dữ liệu sử dụng trong thực nghiệm

TT	Bộ dữ liệu	Miền dữ liệu	Số lớp	Độ dài chuỗi	Số mẫu train/test
1	GunPoint	Chuyển động	2	150	50/150
2	ECG200	Tín hiệu y sinh	2	96	100/100
3	Coffee	Phổ/sensor	2	286	28/28
4	FordA	Cảm biến công nghiệp	2	500	3601/1320
5	Wafer	Sản xuất bán dẫn	2	152	1000/6174

Ghi chú: Tên lớp/ý nghĩa nhãn của các bộ dữ liệu lần lượt là GunPoint: rút súng / chỉ tay; ECG200: bình thường / thiếu máu cơ tim; Coffee: Arabica / Robusta; FordA: bình thường / lỗi động cơ; Wafer: bình thường / bất thường.

Các tỷ lệ train/test trong Bảng 3 được giữ nguyên theo phân chia chuẩn của UCR Time Series Archive, không phải là kết quả tự phân tách lại của tác giả. Vì vậy, một số bộ dữ liệu

như GunPoint và Wafer có số mẫu huấn luyện nhỏ hơn đáng kể so với số mẫu kiểm thử. Việc giữ nguyên phân chia chuẩn giúp bảo đảm khả năng so sánh với các nghiên cứu trước, đồng thời tránh nguy cơ rò rỉ dữ liệu khi tái chia tập dữ liệu.

Bộ dữ liệu Coffee có kích thước rất nhỏ, chỉ gồm 28 mẫu huấn luyện và 28 mẫu kiểm thử, nên kết quả trên bộ dữ liệu này cần được diễn giải thận trọng do nguy cơ quá khớp. Để hạn chế rủi ro này, nghiên cứu sử dụng kiến trúc gọn nhẹ, Dropout, số epoch cố định và đánh giá trung bình trên ba seed. Tuy nhiên, bài báo thừa nhận rằng năm bộ dữ liệu hiện tại chưa đủ để khái quát hóa cho toàn bộ bài toán phân loại chuỗi thời gian; các thực nghiệm trên bộ dữ liệu lớn hơn, bao gồm dữ liệu kinh tế và tài chính, sẽ được xem xét trong nghiên cứu tiếp theo.

3.3. Thiết lập thực nghiệm

Thiết lập thực nghiệm sử dụng ảnh đầu vào kích thước 64 × 64, tối đa 50 epoch, batch size 32 và ba seed ngẫu nhiên 42, 2024, 2026. Để bảo đảm tính nhất quán khi tổng hợp kết quả, các lần chạy lặp trong cùng một seed được lấy trung bình trước khi tính giá trị cuối cùng. Các bộ dữ liệu được sử dụng theo phân chia train/test chuẩn do UCR Time Series Archive cung cấp.

3.4. Cấu hình siêu tham số của các biểu diễn hai chiều

Để bảo đảm tính nhất quán trong so sánh, các kỹ thuật biểu diễn ảnh hai chiều được thiết lập với cùng kích thước ảnh đầu ra 64 × 64 cho toàn bộ năm bộ dữ liệu. Các siêu tham số được lựa chọn theo cấu hình thống nhất và có khả năng tái lập, không tối ưu riêng cho từng bộ dữ liệu. Cách thiết lập này giúp hạn chế nguy cơ thiên lệch do tinh chỉnh tham số theo từng tập dữ liệu cụ thể, nhưng đồng thời cũng có nghĩa rằng kết quả thực nghiệm phản ánh hiệu quả của các kỹ thuật trong một cấu hình chung, chưa phải là hiệu năng tối ưu tuyệt đối của từng phương pháp biểu diễn.

Bảng 3a. Cấu hình siêu tham số của các kỹ thuật biểu diễn hai chiều

Kỹ thuật	Hàm/thư viện sử dụng	Siêu tham số chính	Giá trị sử dụng trong thực nghiệm	Xử lý đầu ra
GAF	GramianAngularField từ pyts.image	image_size, method	image_size = 64; method = "summation"	Ảnh được chuyển sang float32 và bổ sung chiều kênh.
MTF	MarkovTransitionField từ pyts.image	image_size, n_bins	image_size = 64; n_bins = 8	Ảnh được chuyển sang float32 và bổ sung chiều kênh.
RP	RecurrencePlot từ pyts.image	Cấu hình RP	Sử dụng cấu hình mặc định của pyts; không đặt riêng threshold, dimension, time_delay trong code.	Ảnh RP được resize về 64 × 64 với anti_aliasing = True, sau đó chuyển sang float32 và bổ sung chiều kênh.
STFT	stft từ scipy.signal	Cấu hình STFT	Sử dụng cấu hình mặc định của scipy.signal.stft; không đặt riêng window_size, overlap, n_fft trong code.	Lấy phổ biên độ abs(zxx), resize về 64 × 64 với anti_aliasing = True, sau đó chuyển sang float32 và bổ sung chiều kênh.

Cách thiết lập trên giúp các biểu diễn hai chiều được so sánh trong cùng một điều kiện đầu vào của mô hình Light 2D-CNN. Tuy nhiên, hiệu năng của MTF, RP và STFT có thể thay đổi khi điều chỉnh các siêu tham số như số mức lượng tử hóa, ngưỡng tái xuất hiện, kích thước cửa sổ hoặc độ chồng lấn. Do đó, kết quả trong nghiên cứu này nên được hiểu là kết quả thu được trong một cấu hình thực nghiệm thống nhất và có khả năng tái lập, chưa phải là kết quả sau khi tối ưu siêu tham số riêng cho từng bộ dữ liệu. Việc khảo sát ảnh hưởng của các siêu tham số này sẽ được xem xét trong các nghiên cứu tiếp theo.

3.5. Môi trường phần cứng và phần mềm

Các thực nghiệm được thực hiện trên Google Colab với GPU NVIDIA Tesla T4 15.360 MiB, CUDA 13.0, CPU Intel Xeon @ 2.00GHz, RAM hệ thống khoảng 12 GiB và Python 3.12.13. Các thư viện chính gồm TensorFlow/Keras, NumPy, scikit-learn, pyts, SciPy và scikit-image. Thời gian huấn luyện trong Bảng 5 được đo trên cùng môi trường thực nghiệm, không bao gồm thời gian tải dữ liệu, tiền xử lý và chuyển đổi chuỗi thời gian thành ảnh hai chiều. Do Google Colab cấp phát tài nguyên động, thời gian huấn luyện có thể thay đổi nhẹ giữa các phiên chạy.

4. Kết quả thực nghiệm và thảo luận

4.1. Kết quả phân loại theo từng bộ dữ liệu

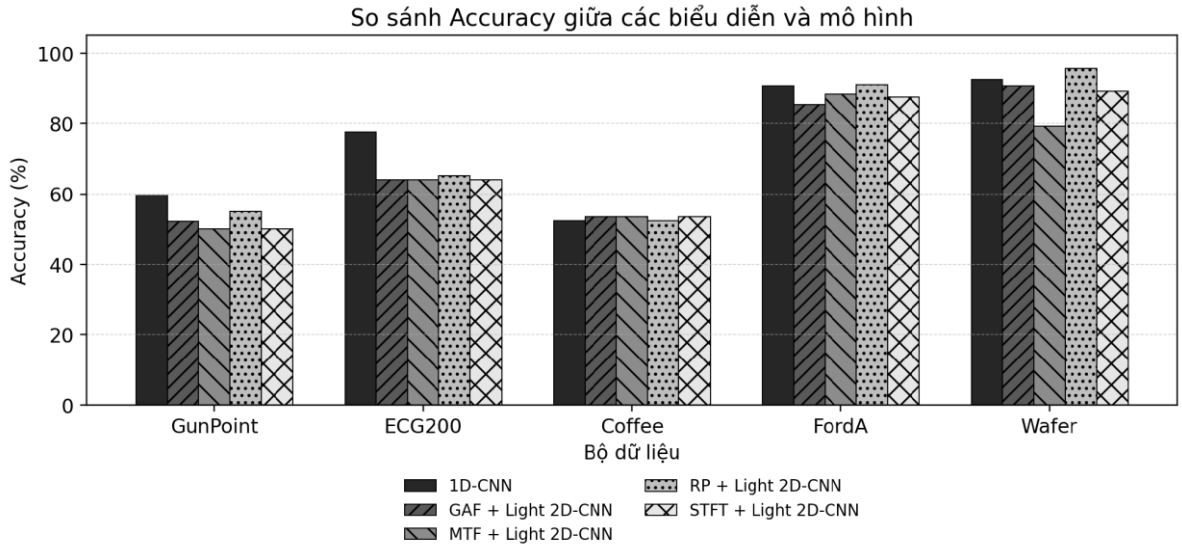
Bảng 4 trình bày kết quả phân loại của các mô hình theo hai chỉ số đánh giá gồm độ chính xác (Acc) và Macro F1-score (F1). Để thuận tiện cho việc đối chiếu kết quả giữa các phương pháp, mỗi mô hình/biểu diễn được chia thành hai cột nhỏ tương ứng với Acc và F1. Các giá trị được tính trung bình trên ba seed sau khi gộp các lần chạy lặp trong cùng một seed. Giá trị tốt nhất trên từng bộ dữ liệu được in đậm.

Bảng 4. Kết quả phân loại theo Accuracy và Macro F1-score (%)

Bộ dữ liệu	1D-CNN		GAF + Light 2D-CNN		MTF + Light 2D-CNN		RP + Light 2D-CNN		STFT + Light 2D-CNN	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
GunPoint	**59.67**	**51.09**	52.17	38.40	50.11	33.76	55.00	43.20	50.11	33.38
ECG200	**77.50**	**73.21**	64.00	39.02	64.00	39.02	65.28	44.75	64.00	39.02
Coffee	52.38	34.35	**53.57**	34.88	**53.57**	34.88	52.38	**36.51**	**53.57**	34.88
FordA	90.72	90.71	85.42	85.39	88.42	88.38	**91.04**	**91.01**	87.54	84.60
Wafer	92.57	64.14	90.74	55.27	79.19	57.04	**95.69**	**80.53**	89.21	47.15

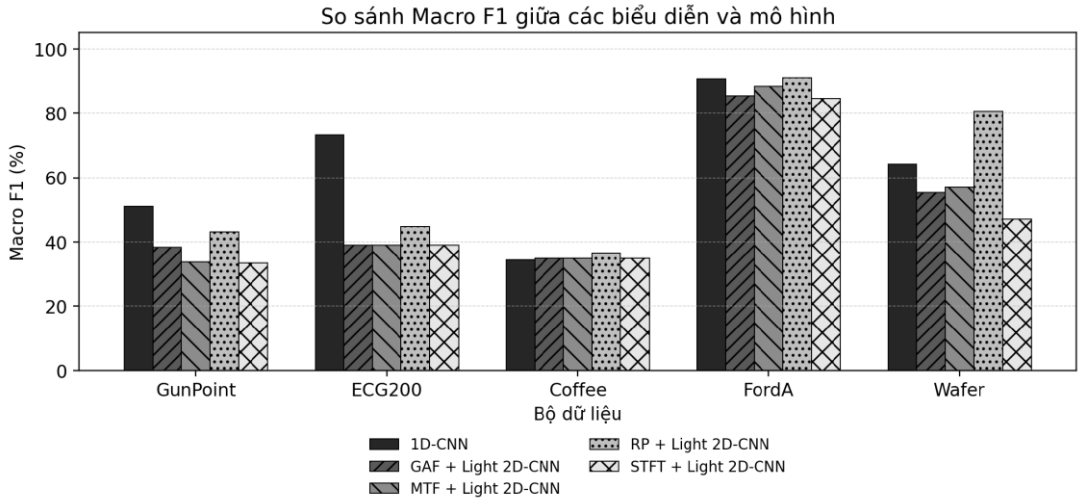
Ghi chú: Acc là Accuracy; F1 là Macro F1-score. Các giá trị in đậm là kết quả tốt nhất theo từng bộ dữ liệu và từng chỉ số đánh giá.

DOI: https://doi.org/10.65153/geb69058



Hình 2. So sánh độ chính xác giữa các mô hình và kỹ thuật biểu diễn trên năm bộ dữ liệu

Bảng 4, Hình 2 và Hình 3 cho thấy hiệu quả phân loại của các mô hình theo hai chỉ số Acc và F1. Nhìn chung, 1D-CNN đạt kết quả tốt trên GunPoint và ECG200, trong khi RP + Light 2D-CNN nổi bật trên FordA và Wafer. Trên Coffee, các phương pháp 2D có chênh lệch nhỏ giữa Acc và F1, cho thấy cần đánh giá đồng thời cả hai chỉ số, đặc biệt khi dữ liệu có khả năng mất cân bằng lớp.



Hình 3. So sánh Macro F1-score giữa các biểu diễn và mô hình trên năm bộ dữ liệu

4.2. So sánh hiệu năng trung bình và chi phí mô hình.

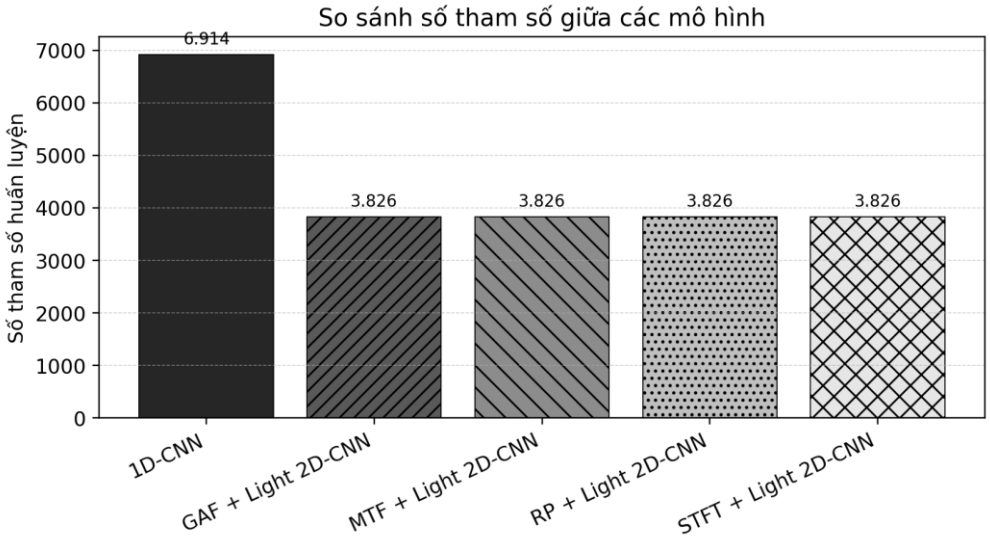
Bảng 5 trình bày hiệu năng trung bình và chi phí huấn luyện của các mô hình. Các mô hình 2D dùng chung kiến trúc Light 2D-CNN nên có cùng 3.826 tham số, thấp hơn 6.914 tham số của 1D-CNN. Tuy nhiên, 1D-CNN đạt độ chính xác và Macro F1-score trung bình cao nhất, đồng thời có thời gian huấn luyện ngắn hơn. Điều này cho thấy, trong thiết lập thực nghiệm

này, số tham số nhỏ hơn không nhất thiết kéo theo chi phí huấn luyện thấp hơn khi đầu vào được biểu diễn dưới dạng ảnh hai chiều.

Bảng 5. Hiệu năng trung bình và chi phí huấn luyện của các mô hình

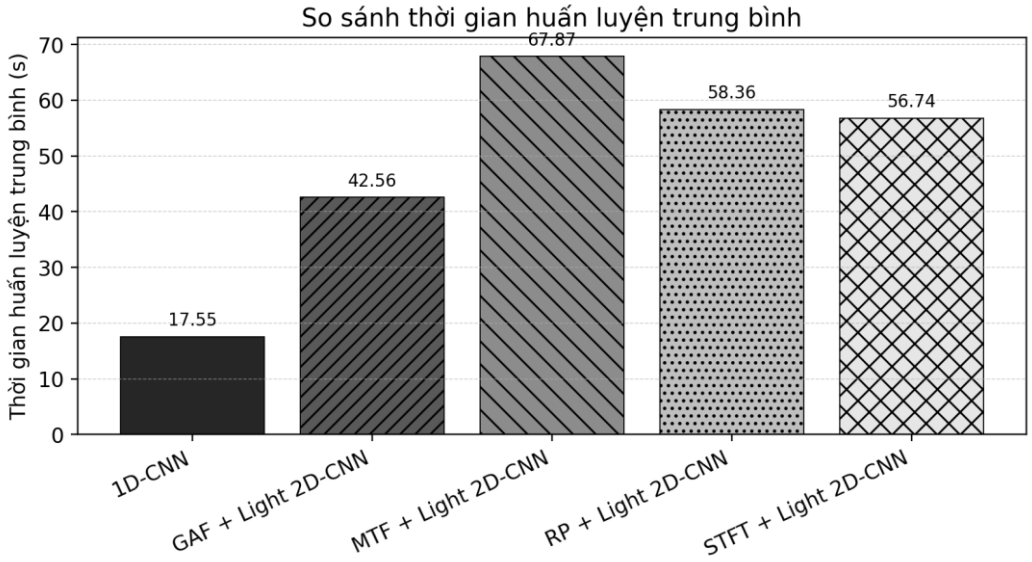
Mô hình/biểu diễn	Tổng số tham số	Độ chính xác trung bình (%)	Macro F1-score trung bình (%)	Thời gian huấn luyện trung bình (s)	Số kết quả tổng hợp
1D-CNN	6914	74.57	62.70	17.55	15
GAF + Light 2D-CNN	3826	69.18	50.59	42.56	15
MTF + Light 2D-CNN	3826	67.06	50.62	67.87	15
RP + Light 2D-CNN	3826	71.88	59.20	58.36	15
STFT + Light 2D-CNN	3826	68.89	47.81	56.74	15

Trong Bảng 5, số lần chạy được tính sau khi tổng hợp theo từng bộ dữ liệu và seed. Với mỗi mô hình/biểu diễn, thực nghiệm được thực hiện trên 5 bộ dữ liệu và 3 seed ngẫu nhiên, tương ứng 15 kết quả dùng để tính trung bình. Các lần chạy lặp trong cùng một seed, nếu có, được lấy trung bình trước khi tổng hợp.



Hình 4. So sánh số tham số trung bình giữa các mô hình

Hình 4 và Hình 5 trực quan hóa sự khác biệt giữa số tham số và thời gian huấn luyện của các mô hình. Mặc dù các mô hình Light 2D-CNN có số tham số thấp hơn 1D-CNN, thời gian huấn luyện trung bình lại cao hơn. Kết quả này cho thấy, trong thiết lập thực nghiệm này, số tham số nhỏ hơn không nhất thiết kéo theo chi phí huấn luyện thấp hơn, đặc biệt khi đầu vào được chuyển đổi thành ảnh hai chiều.



Hình 5. So sánh thời gian huấn luyện trung bình giữa các mô hình

4.3. Phương pháp tốt nhất cho từng bộ dữ liệu

Sau khi so sánh hiệu năng trung bình và chi phí huấn luyện, cần xác định phương pháp đạt kết quả tốt nhất trên từng bộ dữ liệu cụ thể. Bảng 6 trình bày mô hình/biểu diễn có độ chính xác cao nhất tương ứng với từng bộ dữ liệu.

Bảng 6. Phương pháp đạt độ chính xác cao nhất theo từng bộ dữ liệu

Bộ dữ liệu	Phương pháp tốt nhất	Độ chính xác tốt nhất (%)	Macro F1-score tương ứng (%)	Số tham số	Số lần chạy
GunPoint	1D-CNN	59.67	51.09	6914	3
ECG200	1D-CNN	77.50	73.21	6914	3
Coffee	GAF + Light 2D-CNN	53.57	34.88	3826	3
FordA	RP + Light 2D-CNN	91.04	91.01	3826	3
Wafer	RP + Light 2D-CNN	95.69	80.53	3826	3

4.4. Phân tích và thảo luận kết quả

Kết quả thực nghiệm được phân tích theo hiệu quả phân loại và chi phí mô hình. Theo Bảng 5, 1D-CNN đạt kết quả trung bình cao nhất với độ chính xác 74,57% và Macro F1-score 62,70%. Trong nhóm biểu diễn hai chiều, RP đạt hiệu quả tốt nhất với độ chính xác 71,88% và Macro F1-score 59,20%, tiếp theo là GAF, STFT và MTF. Kết quả này cho thấy chuyển đổi chuỗi thời gian sang ảnh hai chiều không luôn cải thiện hiệu năng so với mô hình xử lý trực tiếp chuỗi gốc.

Theo Bảng 6, 1D-CNN đạt kết quả tốt nhất trên GunPoint và ECG200, trong khi RP + Light 2D-CNN nổi bật trên FordA và Wafer. Điều này cho thấy hiệu quả của từng biểu diễn phụ thuộc vào đặc điểm tín hiệu; RP phù hợp với dữ liệu có cấu trúc tái xuất hiện hoặc động lực học rõ rệt [3]. Về chi phí, các mô hình Light 2D-CNN có số tham số thấp hơn 1D-CNN nhưng thời gian huấn luyện trung bình cao hơn. Do đó, lựa chọn biểu diễn đầu vào cần cân nhắc đồng thời độ chính xác, macro F1-score, số tham số, thời gian huấn luyện và đặc điểm dữ liệu. Bảng 7 tổng hợp các hàm ý chính từ kết quả thực nghiệm.

Các kết quả này không nhằm khẳng định một kỹ thuật biểu diễn ảnh hai chiều là tối ưu cho mọi bộ dữ liệu. Ngược lại, sự khác biệt giữa GunPoint, ECG200, Coffee, FordA và Wafer cho thấy lựa chọn biểu diễn cần dựa trên cấu trúc tín hiệu, quy mô dữ liệu, mức mất cân bằng lớp và ràng buộc tính toán. Đặc biệt, với các bộ dữ liệu nhỏ như Coffee, kết quả thực nghiệm nên được xem là chỉ báo ban đầu thay vì bằng chứng khái quát hóa mạnh.

Bảng 7. Hàm ý lựa chọn phương pháp biểu diễn từ kết quả thực nghiệm

Nhận định từ thực nghiệm	Diễn giải	Hàm ý lựa chọn
1D-CNN đạt trung bình cao nhất	Chuỗi gốc vẫn chứa đủ thông tin phân biệt trong nhiều trường hợp	Nên sử dụng 1D-CNN làm mô hình đối chứng bắt buộc
RP tốt nhất trong nhóm biểu diễn hai chiều	RP nổi bật trên FordA và Wafer	Ưu tiên RP khi tín hiệu có cấu trúc lặp hoặc động lực học
Light 2D-CNN có ít tham số hơn	Các mô hình Light 2D-CNN có 3.826 tham số, thấp hơn 1D-CNN	Phù hợp khi ưu tiên mô hình gọn nhẹ
Thời gian huấn luyện của mô hình hai chiều cao hơn	Ảnh hai chiều làm tăng chi phí xử lý dù số tham số thấp	Cần cân nhắc giữa số tham số và thời gian huấn luyện
Độ chính xác và macro F1-score có chênh lệch	Một số bộ dữ liệu có khả năng mất cân bằng lớp	Không nên chỉ dựa vào độ chính xác khi đánh giá

5. Kết luận

Bài báo đã xây dựng và thực nghiệm một khung so sánh các kỹ thuật biểu diễn ảnh hai chiều từ chuỗi thời gian một chiều cho bài toán phân loại bằng mạng nơ-ron tích chập nhẹ. Các biểu diễn được khảo sát gồm GAF, MTF, RP và STFT; trong đó Light 2D-CNN được sử dụng cho các ảnh hai chiều, còn 1D-CNN được dùng làm mô hình đối chứng trực tiếp.

Kết quả thực nghiệm cho thấy 1D-CNN đạt hiệu năng trung bình cao nhất trên các bộ dữ liệu khảo sát. Trong nhóm biểu diễn hai chiều, RP cho kết quả tốt nhất, đặc biệt trên FordA và Wafer. Điều này cho thấy chuyển đổi chuỗi thời gian sang ảnh hai chiều có thể hữu ích trong một số trường hợp, nhưng hiệu quả phụ thuộc vào đặc điểm tín hiệu và không nên được khái quát hóa cho mọi bài toán phân loại chuỗi thời gian. Kết quả cũng cần được hiểu trong giới hạn

của năm bộ dữ liệu UCR được khảo sát và phân chia train/test chuẩn của từng bộ dữ liệu; riêng Coffee có kích thước nhỏ nên chỉ cung cấp bằng chứng tham khảo ban đầu.

Hướng nghiên cứu tiếp theo là mở rộng thực nghiệm trên nhiều bộ dữ liệu hơn, đặc biệt các bộ dữ liệu lớn và đa lĩnh vực như kinh tế, tài chính, công nghiệp và y sinh; bổ sung kiểm định thống kê; báo cáo thời gian suy luận trên mỗi mẫu; và đánh giá ảnh hưởng của các tham số biểu diễn như kích thước ảnh, số mức lượng tử hóa trong MTF, ngưỡng trong RP và cấu hình biến đổi thời gian - tần số. Đồng thời, cần khảo sát siêu tham số kiến trúc Light 2D-CNN như số lớp, số block, số bộ lọc, kích thước kernel và dropout. Ngoài ra, nghiên cứu có thể được mở rộng sang chuỗi thời gian đa biến thông qua biểu diễn ảnh đa kênh hoặc cơ chế hợp nhất đặc trưng giữa các biến.

Tài liệu tham khảo

1. Wang Z., Oates T. (2015). Imaging Time-Series to Improve Classification and Imputation, Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, tr. 3939–3945.
2. Wang Z., Oates T. (2015). Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks, Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.
3. Marwan N., Romano M. C., Thiel M., Kurths J. (2007). Recurrence plots for the analysis of complex systems, Physics Reports, 438(5–6), tr. 237–329. DOI: 10.1016/j.physrep.2006.11.001.
4. Ismail Fawaz H., Lucas B., Forestier G., Pelletier C., Schmidt D. F., Weber J., Webb G. I., Idoumghar L., Muller P.-A., Petitjean F. (2020). InceptionTime: Finding AlexNet for Time Series Classification, Data Mining and Knowledge Discovery, 34(6), tr. 1936–1962. DOI: 10.1007/s10618-020-00710-y.
5. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv preprint arXiv:1704.04861.
6. Chollet F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, tr. 1800–1807. DOI: 10.1109/CVPR.2017.195.
7. Dau H. A., Bagnall A., Kamgar K., Yeh C.-C. M., Zhu Y., Gharghabi S., Ratanamahatana C. A., Keogh E. (2019). The UCR Time Series Archive, IEEE/CAA Journal of Automatica Sinica, 6(6), tr. 1293–1305. DOI: 10.1109/JAS.2019.1911747.
8. Faouzi J., Janati H. (2020). pyts: A Python Package for Time Series Classification, Journal of Machine Learning Research, 21(46), tr. 1–6.
9. Bagnall A., Lines J., Bostrom A., Large J., Keogh E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, Data Mining and Knowledge Discovery, 31(3), tr. 606–660. DOI: 10.1007/s10618-016-0483-9.

10. Al Joumaa H., Al-Shrouf L., Jelali M. (2025). Time Series-to-Image Encoding for Classification Using Convolutional Neural Networks: A Novel and Robust Approach, Machine Learning and Knowledge Extraction, 7(4), Article 155. DOI: 10.3390/make7040155.

11. Keogh E., Chakrabarti K., Pazzani M., Mehrotra S. (2001). Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases, Knowledge and Information Systems, 3(3), tr. 263–286. DOI: 10.1007/PL00011669.

12. Gross J., Buettner R., Baumgartl H. (2022). Benchmarking Transfer Learning Strategies in Time-Series Imaging: Recommendations for Analyzing Raw Sensor Data, IEEE Access, 10, tr. 16977–16991. DOI: 10.1109/ACCESS.2022.3148711.

13. Semenoglou A.-A., Spiliotis E., Assimakopoulos V. (2023). Image-based time series forecasting: A deep convolutional neural network approach, Neural Networks, 157, tr. 39–53. DOI: 10.1016/j.neunet.2022.10.006.

THÔNG TIN TÁC GIẢ

So sánh hiệu quả các kỹ thuật biểu diễn ảnh hai chiều từ dữ liệu chuỗi thời gian một chiều cho bài toán phân loại bằng mạng nơ-ron tích chập nhẹ

Comparative Evaluation of 2D Image Representations from 1D Time Series for Classification Using Lightweight Convolutional Neural Networks

Họ và tên tác giả : Trần Nguyên Hoàng

Đơn vị công tác: Khoa Công nghệ thông tin, Trường Đại học công nghệ Đông Á

Email: hoangtn@eaut.edu.vn

ORCID : Chưa có / Không áp dụng

Tác giả liên hệ (Corresponding Author):

Họ và tên: **Trần Nguyên Hoàng**

Email: hoangtn@eaut.edu.vn

Số điện thoại: 0983277689

Lời cảm ơn

Tác giả trân trọng cảm ơn Khoa Công nghệ thông tin, Trường Đại học Công nghệ Đông Á đã hỗ trợ định hướng học thuật trong quá trình xây dựng bản thảo nghiên cứu này.

Tuyên bố về nguồn tài trợ

Nghiên cứu này chưa nhận tài trợ từ bất kỳ tổ chức tài trợ công, thương mại hoặc phi lợi nhuận nào.

Tuyên bố về xung đột lợi ích

Tác giả tuyên bố không có xung đột lợi ích liên quan đến nghiên cứu này.