CHẨN ĐOÁN LỖI MÁY CÔNG NGHIỆP DỰA TRÊN PHÂN TÍCH TÍN HIỆU ÂM THANH SỬ DỤNG MÔ HÌNH ĐƯỢC HUẤN LUYỆN TRƯỚC

Nguyễn Đức Hoàng[1]

TÓM TẮT

Chẩn đoán lỗi máy công nghiệp là một bài toán quan trọng để đảm bảo hiệu suất cho toàn hệ thống sản xuất, song các phương pháp truyền thống dựa trên cảm biến vật lý thường thiếu linh hoạt và khó mở rộng. Bài báo này đề xuất một hướng tiếp cận mới dựa trên phân tích tín hiệu âm thanh, sử dụng mô hình WavLM được huấn luyện trước để trích đặc trưng bậc cao trực tiếp từ âm thanh thô. Kiểm chứng trên bộ dữ liệu MIMII cho thấy kiến trúc đề xuất có hiệu năng tốt so với phương pháp dùng mô hình DAE gốc, đặc biệt trong môi trường nhiễu cao. Nghiên cứu đã khẳng định tiềm năng to lớn của việc ứng dụng các mô hình âm thanh được huấn luyện trước cho bài toán chẩn đoán lỗi máy trong môi trường công nghiệp, đồng thời gợi mở các hướng cải tiến trong tương lai như sử dụng kiến trúc lai và chiến lược tinh chỉnh mô hình theo miền ứng dụng.

Từ khóa: *Bộ dữ liệu MIMII, chẩn đoán lỗi máy công nghiệp, mô hình được huấn luyện trước, phân tích tín hiệu, mô hình được huấn luyện trước, WavLM, Transfer Learning.*

1. Giới thiệu

Trong các môi trường công nghiệp hiện đại, độ tin cậy và hoạt động liên tục của máy công nghiệp đóng vai trò quan trọng trong việc đảm bảo năng suất, hiệu quả chi phí và an toàn lao động. Các sự cố máy công nghiệp bất ngờ không chỉ dẫn đến thời gian ngừng hoạt động tốn kém mà còn gây ra rủi ro cho người vận hành và tính ổn định của hệ thống. Do đó, chẩn đoán lỗi máy công nghiệp đã trở thành một lĩnh vực nghiên cứu quan trọng, nhằm phát hiện và phân loại lỗi ở giai đoạn sớm để cho phép bảo trì kịp thời và tránh các sự cố đáng tiếc.

Các phương pháp chẩn đoán lỗi truyền thống chủ yếu dựa vào phân tích rung động, chụp ảnh nhiệt và phân tích đặc trưng dòng điện động cơ [1, 2]. Phân tích rung động được coi là phương pháp tốt nhất trong việc xác định tình trạng máy công nghiệp, trong khi phân tích đặc trưng dòng điện động cơ được công nhận rộng rãi như là phương pháp tiêu chuẩn để phát hiện lỗi động cơ. Mặc dù các phương pháp này đã chứng minh hiệu quả trong nhiều tình huống, chúng thường yêu cầu lắp đặt cảm biến trực tiếp, phần cứng chuyên dụng và môi trường được kiểm soát. Gần đây, việc tận dụng tín hiệu âm thanh-có sẵn, không can thiệp vào máy công nghiệp và giàu thông tin - như một giải pháp thay thế cho giám sát tình trạng máy công nghiệp đã thu hút sự quan tâm của các nhà nghiên cứu.

Âm thanh phát ra từ máy công nghiệp có thể phản ánh trạng thái cơ học của các thành phần như ổ bi, bánh răng và động cơ. Gần đây, chẩn đoán lỗi cơ khí dựa trên phân tích tín hiệu âm thanh đã dần trở thành một chủ đề nghiên cứu nóng trong lĩnh vực chẩn đoán lỗi cơ khí, do các ưu điểm về độ chính xác phát hiện cao, khả năng tổng quát hóa

tốt, đo lường không nhúng và chi phí thấp [3]. Cụ thể, các nghiên cứu đã chứng minh hiệu quả của phương pháp này trên nhiều loại máy công nghiệp khác nhau bao gồm van điện từ, máy bơm, quạt làm mát, ray trượt và hộp số. Các cảm biến âm thanh có khả năng nắm bắt chính xác đặc trưng lỗi, từ đó cải thiện độ chính xác trong phát hiện các bất thường so với hoạt động bình thường [4]. Các âm thanh bất thường có thể phản ánh nhiều dạng hư hỏng khác nhau như nhiễm bẩn, rò rỉ, mất cân bằng quay, hỏng ray,… khiến chẩn đoán lỗi dựa trên âm thanh trở thành một giải pháp đầy hứa hẹn và hiệu quả về chi phí, đặc biệt trong các môi trường công nghiệp có độ ồn lớn hoặc khó tiếp cận.

Đồng thời, các giải pháp học máy với sự phát triển vượt bậc hiện nay đã cách mạng hóa lĩnh vực nhận dạng mẫu và xử lý tín hiệu, cho phép các mô hình tự động học các cấp độ biểu diễn từ dữ liệu thô [5]. Với khả năng trích xuất các đặc trưng có tính đại diện cao, các mô hình học sâu với cấu trúc nhiều lớp ẩn đã cho thấy hiệu suất vượt trội trong chẩn đoán và giám sát lỗi hệ thống máy móc. Đặc biệt, các mô hình được huấn luyện trước như WavLM [6] đã cho thấy hiệu suất đáng chú ý trong các nhiệm vụ xử lý liên quan đến giọng nói và âm thanh bằng cách học từ các tập dữ liệu lớn không nhãn. WavLM được huấn luyện trên dữ liệu âm thanh đa dạng thông qua phương pháp học tự giám sát. Mô hình này có khả năng nắm bắt các đặc trưng ngữ cảnh và thời gian phong phú, đồng thời thể hiện tính tổng quát hóa tốt cho nhiều dạng âm thanh khác nhau chứ không chỉ cho xử lý giọng nói, mà còn cho phân loại âm thanh, nhận dạng âm thanh môi trường và giám sát tình trạng máy công nghiệp…

Nghiên cứu này đề xuất một phương pháp cho chẩn đoán lỗi máy công nghiệp dựa trên phân tích tín hiệu âm thanh sử dụng mô hình được huấn luyện trước WavLM. Khác với các phương pháp trích xuất đặc trưng truyền thống, giải pháp nghiên cứu đề xuất tận dụng sức mạnh của kỹ thuật transfer learning để trích xuất các đặc trưng âm thanh cấp cao từ các bản ghi âm thô của máy công nghiệp. Các vector đặc trưng này sau đó được sử dụng cho mô hình Autoencoder (học cách nén và tái tạo đặc trưng âm thanh trích xuất từ WavLM) để phát hiện bất thường. Các đóng góp chính của nghiên cứu này như sau:

1. Tận dụng khả năng của mô hình được huấn luyện trước WavLM để bỏ qua các kỹ thuật trích đặc trưng như xây dựng mel-spectrogram hoặc trích xuất hệ số MFCC, từ đó đơn giản hóa quy trình chẩn đoán. Giải pháp đề xuất sử dụng mô hình WavLM thay cho mô hình DAE (Deep Autoencoder) như trong giải pháp gốc [7]. Triển khai kiến trúc với mô hình WavLM: phát triển mã nguồn gốc (https://github.com/MIMII-hitachi/mimii_baseline) để tích hợp mô hình WavLM và sử dụng mô hình Autoencoder tương tự mô hình DAE để phát hiện bất thường.

2. Đánh giá giải pháp trên tập dữ liệu MIMII (https://zenodo.org/records/3384388). Kết quả cho thấy kiến trúc sử dụng mô hình WavLM đạt được hiệu suất AUC tốt hơn so với kết quả thu được khi sử dụng với mã nguồn gốc (dùng mã nguồn baseline.py và giữ nguyên các thông số như trong baseline.yaml).

Phần còn lại của bài báo được tổ chức như sau: Phần 2 giới thiệu các công trình liên quan trong chẩn đoán lỗi máy công nghiệp, Phần 3 mô tả chi tiết giải pháp đề xuất, Phần 4 trình bày thực nghiệm, các phân tích và thảo luận, Phần 5 kết luận bài báo và đề xuất các hướng nghiên cứu trong tương lai.

2. Các công trình liên quan

Các phương pháp dựa trên phân tích tín hiệu rung động, hình ảnh nhiệt và phân tích đặc trưng dòng điện động cơ yêu cầu phải lắp đặt cảm biến vật lý trực tiếp lên máy công nghiệp và thực hiện tiền xử lý tín hiệu phức tạp, làm hạn chế tính linh hoạt và khả năng mở rộng trong các môi trường công nghiệp phức tạp hoặc khắc nghiệt [1].

Để tự động hóa quá trình chẩn đoán, các kỹ thuật học máy gồm Support Vector Machines (SVM), k-Nearest Neighbors (k-NN) và Gaussian Mixture Models (GMM),… đã được áp dụng để phân loại các mẫu lỗi dựa trên các đặc trưng trích xuất từ miền thời gian, tần số, hoặc biểu diễn thời gian-tần số của các tín hiệu thu được từ máy công nghiệp [2]. Tuy nhiên, các phương pháp này phụ thuộc nhiều vào kiến thức chuyên sâu để chọn giải pháp trích đặc trưng, cơ chế phân loại phù hợp và nhạy cảm với nhiễu cũng như sự biến đổi của môi trường.

Những công bố gần đây đã khám phá việc sử dụng âm thanh truyền qua không khí như một phương thức không can thiệp để giám sát máy công nghiệp, mang lại giải pháp dễ tiếp cận và linh hoạt hơn. Các phương pháp dựa trên âm thanh thường sử dụng các đặc trưng như MFCC, phổ log-mel và đặc trưng chroma để biểu diễn các đặc trưng âm thanh máy công nghiệp [8]. Mã nguồn gốc sử dụng mô hình DAE là một giải pháp được trích dẫn nhiều trong nghiên cứu phát hiện các mẫu âm thanh bình thường (normal) và bất thường (abnomal) liên quan đến lỗi máy công nghiệp [7]. Mặc dù hiệu quả, phương pháp này vẫn phụ thuộc vào các đặc trưng như MFCC, phổ log-mel, vốn có thể không tổng quát hóa tốt trên các loại máy công nghiệp hoặc điều kiện âm thanh khác nhau.

Kỹ thuật transfer learning sử dụng các mô hình âm thanh được huấn luyện trước đã nổi lên như một hướng tiếp cận đầy triển vọng trong phân loại âm thanh và phát hiện bất thường. Các mô hình như wav2vec 2.0 [9], HuBERT [10] và WavLM [6] tận dụng học tự giám sát để nắm bắt các biểu diễn cấp cao về thời gian và ngữ nghĩa từ dữ liệu dạng sóng âm thanh thô. Các mô hình này được huấn luyện trên các tập dữ liệu không gắn nhãn quy mô lớn và có thể được tinh chỉnh hoặc sử dụng như các bộ trích đặc trưng cho các bước xử lý tiếp theo.

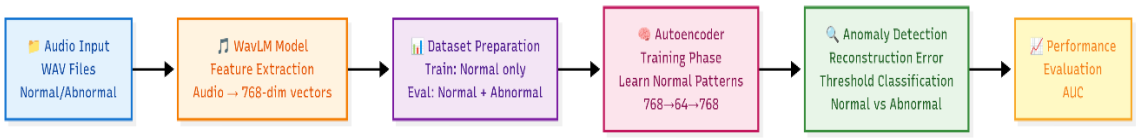
Trong đó, WavLM thể hiện hiệu suất vượt trội trong các bước xử lý liên quan đến tiếng nói và âm thanh tổng quát nhờ khả năng mô hình hóa ngữ cảnh nâng cao và hiệu quả đối với nhiễu. Tuy nhiên, ứng dụng của WavLM trong các lĩnh vực không phải tiếng nói, như phân loại âm thanh môi trường hoặc giám sát trạng thái máy công nghiệp, vẫn còn ít được khám phá. Một số công trình trước đây đã khám phá việc ứng dụng những mô hình tự giám sát như wav2vec hoặc HuBERT cho bài toán phát hiện âm thanh bất

thường, tuy vậy, việc tích hợp mô hình WavLM vào hệ thống chẩn đoán lỗi máy công nghiệp vẫn chưa được nghiên cứu sâu và có các công bố.

3. Giải pháp đề xuất

3.1. Kiến trúc hệ thống

Kiến trúc hệ thống dựa trên ý tưởng chính là thay thế kiến trúc phát hiện bất thường dựa trên mô hình DAE (dùng các đặc trưng phổ log-mel) bằng một kiến trúc được xây dựng dựa trên các đặc trưng âm thanh bậc cao (high-level) và có khả năng tổng quát hóa tốt hơn của mô hình WavLM. Tổng quan về kiến trúc hệ thống được minh họa trong Hình 1.



Hình 1. *Kiến trúc hệ thống*

3.2. Mô tả quá trình xử lý

3.2.1. Trích xuất đặc trưng WavLM

Mô hình WavLM-base [6] huấn luyện trên 94.000 giờ dữ liệu âm thanh không gắn nhãn với mục tiêu là học được các biểu diễn (representations) phổ quát từ âm thanh, không chỉ chứa thông tin về nội dung lời nói mà còn về danh tính người nói, cảm xúc và các đặc điểm khác. Với một dạng sóng đầu vào $x(t)$, được lấy mẫu ở tần số 16 kHz, mô hình WavLM (phiên bản WavLM-base) được cấu hình "đóng băng" - frozen (đóng băng tất cả các trọng số của mô hình và chỉ huấn luyện một lớp tuyến tính cuối cùng) tạo ra một vector đặc trưng có chiều cố định $\boldsymbol{z} \in \mathbb{R}^d$, trong đó $d = 768$. Công thức được biểu diễn như sau:

$$\boldsymbol{z} = \mathrm{WavLM}\big(x(t)\big), \quad \boldsymbol{z} \in \mathbb{R}^{768} \qquad (1)$$

Các đặc trưng này được trích xuất trực tiếp từ các file âm thanh thô, không yêu cầu các phép biến đổi phổ.

3.2.2. Quá trình huấn luyện

Chuẩn bị dữ liệu

Dataset: Dữ liệu được xử lý bởi lớp *OptimizedWavLMPipelineDataset*:

Âm thanh được tải bằng *librosa.load* với tần số lấy mẫu 16000 Hz; Đặc trưng WavLM được trích xuất từ mô hình WavLM (microsoft/wavlm-base), được tải sẵn và đặt ở chế độ eval (không huấn luyện); Đặc trưng được lưu vào bộ nhớ cache (file HDF5) để tối ưu hóa tốc độ xử lý; Dataset bao gồm danh sách file âm thanh và nhãn (0 cho bình thường, 1 cho bất thường).

DataLoader: Dữ liệu huấn luyện được chia thành tập huấn luyện (train_loader) và tập xác thực (val_loader) với tỷ lệ chia được xác định bởi validation_split trong file cấu hình .YAML; Tập đánh giá (eval_loader) được tạo từ các file đánh giá; Kích thước batch được điều chỉnh động bởi lớp AdaptiveBatchSizer dựa trên bộ nhớ GPU và kích thước đầu vào.

Mô hình Autoencoder: Mô hình bao gồm hai thành phần chính:

- Bộ mã hóa (Encoder): Một chuỗi các lớp mạng nơ-ron (nn.Linear) có nhiệm vụ nén đặc trưng đầu vào (ví dụ: từ 768 chiều của mô hình WavLM-base) xuống một không gian nhỏ hơn (64 chiều). Mỗi tầng tuyến tính (Nén dữ liệu: input_dim -> 256 -> 128 -> 64) được theo sau bởi hàm kích hoạt ReLU (nn.ReLU(True)).

- Bộ giải mã (Decoder): Một chuỗi các lớp mạng nơ-ron khác có nhiệm vụ tái tạo lại đặc trưng gốc từ phiên bản đã được nén. Tương tự, mỗi tầng tuyến tính (Tái tạo dữ liệu: 64 -> 128 -> 256 -> input_dim) cũng sử dụng ReLU làm hàm kích hoạt.

Quy trình huấn luyện

Hàm mất mát: Sử dụng Mean Squared Error (MSE) để đo lường sự khác biệt giữa đầu vào và đầu ra tái tạo.

Tối ưu hóa: Sử dụng thuật toán Adam với tốc độ học (learning_rate) được cấu hình trong file .YAML; Kỹ thuật Mixed Precision Training (torch.amp.autocast) được áp dụng để tăng tốc và giảm sử dụng bộ nhớ trên GPU; GradScaler được sử dụng để xử lý gradient trong quá trình huấn luyện.

Quy trình từng epoch

Giai đoạn huấn luyện: Mô hình ở chế độ train; Với mỗi batch: Xóa gradient cũ (optimizer.zero_grad()), tính toán đầu ra tái tạo và hàm mất mát MSE, tính gradient và cập nhật tham số mô hình bằng scaler.step(optimizer), ghi lại mất mát trung bình trên tập huấn luyện.

Giai đoạn xác thực: Mô hình ở chế độ eval, tính mất mát trên tập xác thực (không tính gradient), so sánh mất mát xác thực với giá trị tốt nhất (best_val_loss).

Early Stopping: Nếu mất mát xác thực không cải thiện sau patience epoch (mặc định là 5), quá trình huấn luyện dừng sớm, mô hình tốt nhất (dựa trên mất mát xác thực thấp nhất) được lưu vào file .pth.

3.2.3. Quá trình đánh giá

Tính toán điểm số bất thường

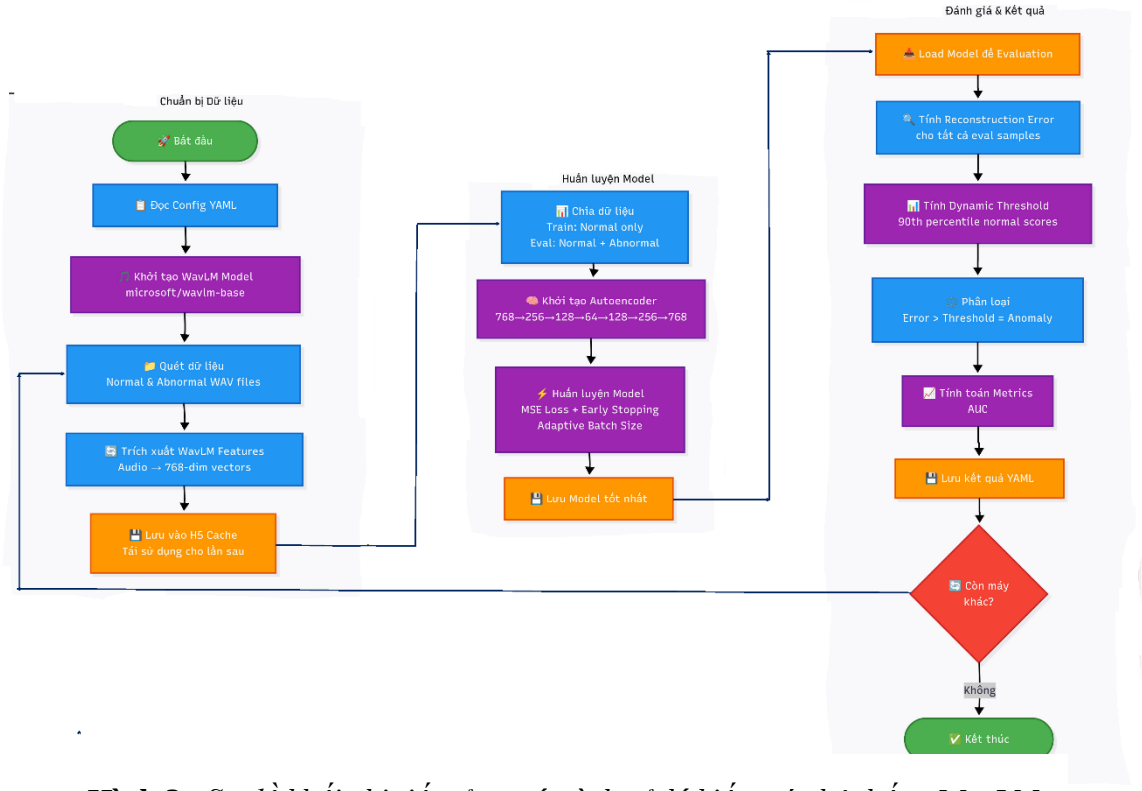
Mô hình tốt nhất được tải lại (model.load_state_dict); Với mỗi mẫu trong tập đánh giá (eval_loader): Tính lỗi tái tạo bằng cách lấy trung bình bình phương sai lệch giữa đầu vào và đầu ra ((model(features) - features) ** 2), các điểm số lỗi này (y_pred_scores) được thu thập để đánh giá.

Xác định ngưỡng động

Điểm ngưỡng (dynamic_threshold) được tính dựa trên bách phân vị thứ 90 của điểm số lỗi trên các mẫu bình thường (nhãn 0); Nếu không có mẫu bình thường, ngưỡng được đặt bằng giá trị trung bình của tất cả điểm số.

Gán nhãn dự đoán: Mẫu có điểm số lỗi lớn hơn ngưỡng được gán nhãn bất thường (1), ngược lại là bình thường (0).

Chỉ số đánh giá: Hiệu suất của cả hai mô hình được đánh giá bằng chỉ số AUC (Area Under the Curve), đo lường khả năng phân biệt giữa điều kiện máy bình thường và bất thường. Kết quả AUC được tính riêng cho từng ID Machine của tập dữ liệu MIMII ở cả ba mức SNR để đánh giá hiệu quả trong các môi trường nhiễu khác nhau.



Hình 2. *Sơ đồ khối chi tiết của quá trình xử lý kiến trúc hệ thống WavLM*

4. Kết quả thực nghiệm

Mô tả tập dữ liệu

Tập dữ liệu MIMII [7] cho phép các nhà nghiên cứu phát triển, huấn luyện và đánh giá các mô hình có khả năng: *(a) Phát hiện bất thường (Anomaly Detection)*: Xác định khi nào âm thanh của một máy công nghiệp có dấu hiệu khác biệt so với hoạt động bình thường; *(b) Phân loại lỗi*: Phân biệt các loại lỗi khác nhau dựa trên đặc điểm âm thanh; *(c) Dự báo để bảo trì:* Dự báo các hỏng hóc tiềm ẩn trước khi chúng xảy ra, cho phép can thiệp kịp thời.

Phiên bản tập dữ liệu MIMII công khai được sử dụng trong bài báo này bao gồm các bản ghi âm từ bốn loại máy công nghiệp phổ biến, mỗi loại có nhiều mẫu khác nhau để đảm bảo tính đa dạng: *(a) Van (Valves)*: Ghi lại âm thanh của các van điện từ khi chúng đóng và mở; *(b) Máy bơm (Pumps):* Âm thanh từ các máy bơm nước hoạt động liên tục; *(c) Quạt (Fans):* Tiếng ồn từ các loại quạt công nghiệp dùng để lưu thông không khí; *(d) Thanh trượt (Slide rails):* Âm thanh phát ra từ các hệ thống trượt tuyến tính khi di chuyển.

Tập MIMII có 48 ID Machine cho 4 loại máy, mỗi loại máy được tổ chức theo 04 ID Machine (id_00, id_02, id_04, id_06) và ba mức tỷ lệ tín hiệu trên nhiễu (SNR): -6 dB, 0 dB, và 6 dB, phản ánh các mức độ nhiễu nền khác nhau. Các bản ghi âm thanh ở điều kiện bình thường được sử dụng để huấn luyện, trong khi bộ đánh giá bao gồm một tập cân bằng giữa dữ liệu âm thanh bình thường và bất thường. Tất cả file âm thanh được lấy mẫu lại ở tần số 16 kHz để đảm bảo tương thích với mô hình WavLM.

Cấu hình thực nghiệm

Kiến trúc dùng mô hình DAE: Đây là kiến trúc được cung cấp trong phiên bản chính thức dùng cho tập dữ liệu MIMII (https://github.com/MIMII-hitachi/mimii_baseline), sử dụng mô hình DAE với kiến trúc mạng kết nối đầy đủ (fully-connected) gồm các lớp (64-64-8-64-64). Đặc trưng log-mel spectrogram được trích xuất từ file âm thanh với các tham số: 64 băng tần mel (n_mels: 64), 5 khung thời gian (frames: 5), kích thước FFT là 1024 (n_fft: 1024), bước nhảy 512 (hop_length: 512), và công suất là 2.0 (power: 2.0). Mô hình được triển khai bằng thư viện Keras, huấn luyện với thuật toán tối ưu Adam (tốc độ học 0.001), kích thước lô (batch_size) là 512 và 50 vòng lặp (epochs) huấn luyện. Các tham số của kiến trúc này được giữ nguyên như trong mã nguồn gốc (chưa tinh chỉnh, tối ưu tham số), kết quả thu được khi chạy kiến trúc này được sử dụng cho mục tiêu so sánh, đánh giá.

Kiến trúc dùng mô hình WavLM: phát triển từ mã nguồn gốc (baseline.py) để ứng dụng mô hình WavLM cho trích các đặc trưng từ các file âm thanh thô. WavLM được sử dụng như một bộ trích đặc trưng, sinh ra các vector đặc trưng có chiều dài 768. Các cấu trúc và tham số của kiến trúc này sau phần xử lý trích đặc trưng từ mô hình WavLM được sử dụng như trong kiến trúc dùng mô hình DAE.

Chỉ số đánh giá: Hiệu suất của cả hai mô hình được đánh giá bằng chỉ số AUC (Area Under the Curve), đo lường khả năng phân biệt giữa điều kiện máy bình thường và bất thường. Kết quả AUC được tính riêng cho từng ID Machine của tập dữ liệu MIMII ở cả ba mức SNR để đánh giá hiệu quả trong các môi trường nhiễu khác nhau.

Kết quả

Các giá trị AUC thu được sau khi chạy các kiến trúc dùng mô hình DAE và kiến trúc dùng mô hình WavLM cho tập dữ liệu MIMII được trình bày trong Bảng 1. Bảng cũng bao gồm độ chênh lệch AUC (DAE AUC – WavLM AUC) và mô hình có hiệu suất cao hơn trong từng điều kiện. Kết quả AUC đối với tập dữ liệu MIMII gồm 48 ID

Machine cho thấy kiến trúc dựa trên mô hình WavLM tốt hơn DAE trong 26/22 ID Machine. Bảng 2 là kết quả AUC trung bình của các loại máy và hiệu suất trung bình tổng thể của 2 mô hình. Khi xét trên toàn bộ dữ liệu, mô hình WavLM có hiệu suất trung bình cao hơn so với DAE. Cụ thể, WavLM đạt AUC trung bình là 0.7456, trong khi DAE đạt AUC trung bình là 0.7303. Kết quả này cho thấy về tổng thể mô hình WavLM tốt hơn DAE.

Bảng 1. *Kết quả AUC của tập dữ liệu MIMII*

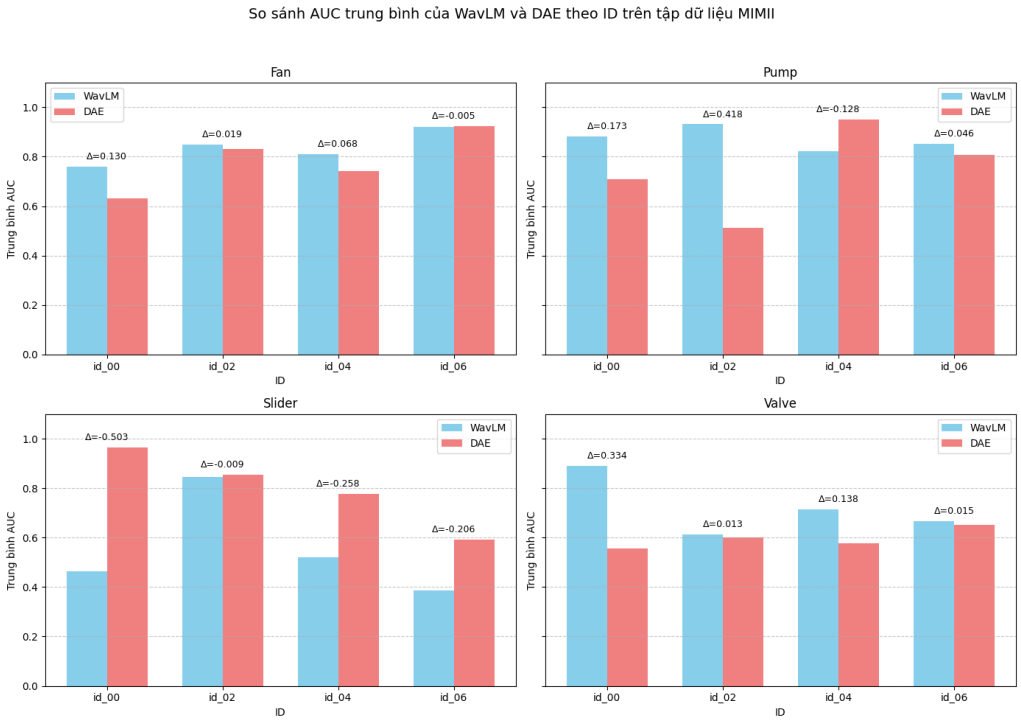
Machine	ID	SNR	AUC WavLM	AUC DAE	Chênh lệch	Mô hình tốt hơn
fan	id_00	-6 dB	**0.6458**	0.5689	0.0769	WavLM
fan	id_00	0 dB	**0.7441**	0.6225	0.1216	WavLM
fan	id_00	6 dB	**0.8896**	0.6988	0.1908	WavLM
fan	id_02	-6 dB	**0.7263**	0.6598	0.0665	WavLM
fan	id_02	0 dB	**0.8560**	0.8512	0.0048	WavLM
fan	id_02	6 dB	0.9665	**0.9805**	-0.0139	DAE
fan	id_04	-6 dB	**0.6479**	0.5917	0.0563	WavLM
fan	id_04	0 dB	**0.8331**	0.7140	0.1191	WavLM
fan	id_04	6 dB	**0.9523**	0.9228	0.0294	WavLM
fan	id_06	-6 dB	**0.7987**	0.7885	0.0102	WavLM
fan	id_06	0 dB	0.9623	**0.9867**	-0.0244	DAE
fan	id_06	6 dB	0.9973	**0.9983**	-0.0010	DAE
pump	id_00	-6 dB	**0.7992**	0.5610	0.2383	WavLM
pump	id_00	0 dB	**0.8951**	0.7938	0.1013	WavLM
pump	id_00	6 dB	**0.9511**	0.7728	0.1783	WavLM
pump	id_02	-6 dB	**0.8270**	0.5571	0.2699	WavLM
pump	id_02	0 dB	**0.9704**	0.4971	0.4733	WavLM
pump	id_02	6 dB	**0.9968**	0.4854	0.5113	WavLM
pump	id_04	-6 dB	0.8430	**0.9174**	-0.0744	DAE
pump	id_04	0 dB	0.8205	**0.9462**	-0.1257	DAE
pump	id_04	6 dB	0.8065	**0.9895**	-0.1830	DAE
pump	id_06	-6 dB	0.6693	**0.6811**	-0.0118	DAE
pump	id_06	0 dB	**0.9109**	0.8520	0.0589	WavLM
pump	id_06	6 dB	**0.9766**	0.8854	0.0912	WavLM
slider	id_00	-6 dB	0.4859	**0.9268**	-0.4409	DAE
slider	id_00	0 dB	0.3921	**0.9810**	-0.5889	DAE
slider	id_00	6 dB	0.5092	**0.9896**	-0.4804	DAE
slider	id_02	-6 dB	0.6955	**0.7716**	-0.0761	DAE
slider	id_02	0 dB	**0.9162**	0.8544	0.0619	WavLM
slider	id_02	6 dB	0.9211	**0.9356**	-0.0144	DAE
slider	id_04	-6 dB	0.5567	**0.6644**	-0.1078	DAE
slider	id_04	0 dB	0.4682	**0.7437**	-0.2755	DAE
slider	id_04	6 dB	0.5321	**0.9233**	-0.3911	DAE
slider	id_06	-6 dB	0.3496	**0.5270**	-0.1774	DAE
slider	id_06	0 dB	0.3014	**0.5383**	-0.2370	DAE

slider	id_06	6 dB	0.5118	**0.7148**	-0.2030	DAE
valve	id_00	-6 dB	**0.7418**	0.2938	0.4480	WavLM
valve	id_00	0 dB	**0.9376**	0.5880	0.3496	WavLM
valve	id_00	6 dB	**0.9918**	0.7878	0.2040	WavLM
valve	id_02	-6 dB	**0.6779**	0.5681	0.1099	WavLM
valve	id_02	0 dB	**0.5767**	0.5699	0.0068	WavLM
valve	id_02	6 dB	0.5881	**0.6664**	-0.0783	DAE
valve	id_04	-6 dB	0.4754	**0.5742**	-0.0988	DAE
valve	id_04	0 dB	**0.7785**	0.4868	0.2917	WavLM
valve	id_04	6 dB	**0.8901**	0.6686	0.2215	WavLM
valve	id_06	-6 dB	**0.5916**	0.5369	0.0547	WavLM
valve	id_06	0 dB	0.6753	**0.6785**	-0.0033	DAE
valve	id_06	6 dB	0.7360	0.7408	-0.0048	DAE

Bảng 2. *Kết quả AUC trung bình theo các máy của tập dữ liệu MIMII*

Machine	AUC WavLM	AUC DAE	Chênh lệch	Mô hình tốt hơn
fan	**0.8350**	0.7820	0.0530	WavLM
pump	**0.8722**	0.7449	0.1273	WavLM
slider	0.5533	**0.7975**	-0.2442	DAE
valve	**0.7217**	0.5967	0.1251	WavLM
Trung bình	**0.7456**	0.7303		**WavLM**

Kết quả AUC theo từng loại máy thể hiện trong Hình 3.



Hình 3. *So sánh kết quả AUC trung bình theo loại máy*

Nhận xét:

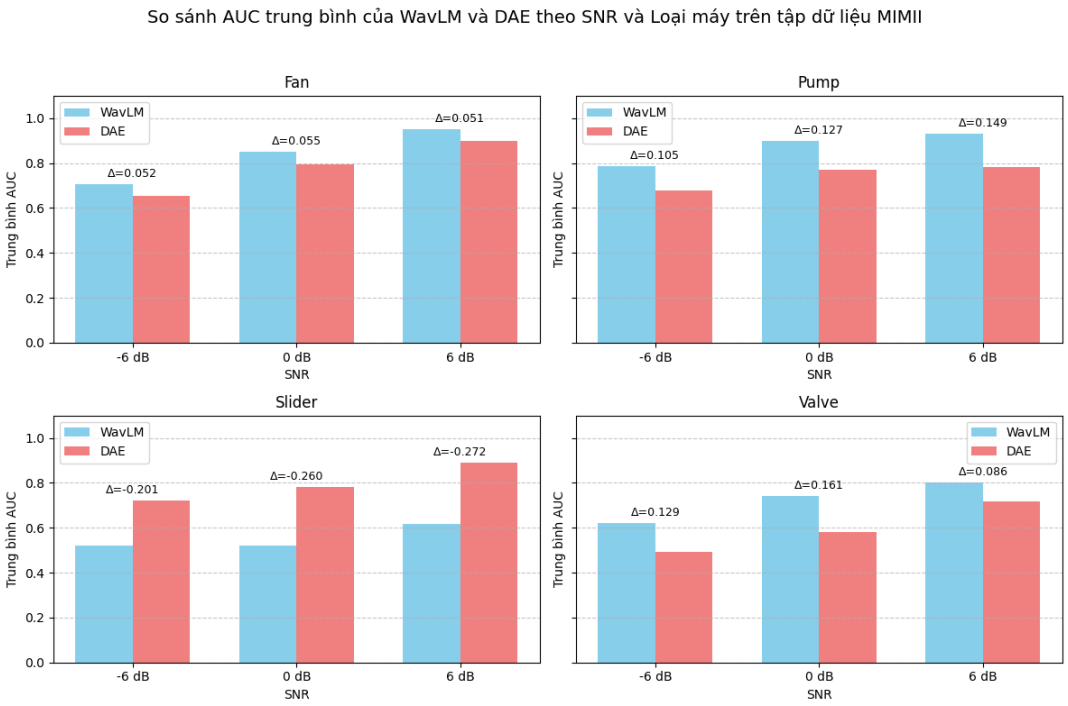
- Fan: WavLM chiếm ưu thế ở hầu hết các trường hợp, đặc biệt ở SNR thấp (-6 dB). Tuy nhiên, DAE có hiệu suất cao hơn ở một số trường hợp với SNR cao (6 dB) cho id_02 và id_06.

- Pump: WavLM có hiệu suất vượt trội ở phần lớn các trường hợp, đặc biệt với id_02 (chênh lệch lên đến 0.4733 - 0.5113). Tuy nhiên, DAE hoạt động tốt hơn ở id_04 trên cả ba mức SNR, cho thấy sự bất ổn định của WavLM với model này.

- Slider: DAE vượt trội rõ rệt so với WavLM trên loại máy slider, đặc biệt ở SNR thấp (-6 dB) và trung bình (0 dB). WavLM chỉ tốt hơn ở một trường hợp (id_02, 0 dB).

- Vavle: WavLM thường tốt hơn, đặc biệt ở id_00 (chênh lệch lớn: 0.2040–0.4480). Tuy nhiên, DAE có hiệu suất tốt hơn ở một số trường hợp, đặc biệt với id_04 (-6 dB) và id_06 (0 dB, 6 dB).

Kết quả AUC theo SNR thể hiện trong Hình 4:



Hình 4. *So sánh kết quả AUC trung bình theo SNR*

Nhận xét:

- Mức -6 dB (nhiễu cao): Trong điều kiện nhiễu cao, hiệu suất của cả hai mô hình thường thấp hơn so với các mức SNR cao hơn. WavLM thường có xu hướng vượt trội hơn DAE ở mức SNR này cho các loại máy như Fan và Pump, và Valve (ngoại trừ một số ID cụ thể). Tuy nhiên, DAE vẫn mạnh hơn ở Slider.

- Mức 0 dB (nhiễu vừa): Khi chất lượng tín hiệu được cải thiện, AUC của cả hai mô hình đều có xu hướng tăng. WavLM vẫn duy trì sự vượt trội ở nhiều trường hợp cho

Fan, Pump và Valve, nhưng DAE tiếp tục khẳng định ưu thế cho Slider và bắt đầu vượt trội cho một số ID cụ thể của Pump và Valve.

- Mức 6 dB (nhiễu thấp): Ở mức SNR cao nhất (ít nhiễu nhất), cả hai mô hình đều đạt được hiệu suất cao nhất, với AUC gần 1.0 trong nhiều trường hợp. Tại mức này, sự cạnh tranh giữa hai mô hình trở nên rõ rệt hơn, với DAE thể hiện tốt hơn ở một số ID Fan và Valve, và vẫn giữ vững vị thế cho Slider.

Nhận xét chung

- Hiệu suất tăng theo SNR: Khi SNR tăng (tức là tín hiệu sạch hơn), hiệu suất AUC của cả hai mô hình đều có xu hướng tăng, điều này là đúng như kỳ vọng vì việc phát hiện sự bất thường trở nên dễ dàng hơn khi nhiễu giảm.

- Phụ thuộc vào loại máy: Sự lựa chọn mô hình tốt hơn phụ thuộc rất nhiều vào loại máy. WavLM là lựa chọn mạnh mẽ cho các thiết bị quay (Fan, Pump) và Vavle, trong khi DAE là lựa chọn rõ ràng cho Slider.

- Không có mô hình chiến thắng tuyệt đối: Không có mô hình nào vượt trội hoàn toàn trong mọi điều kiện. Hiệu suất tối ưu đạt được dựa trên sự kết hợp giữa mô hình, loại máy và mức độ nhiễu.

- Chi phí tính toán cao: Quá trình trích xuất vector đặc trưng từ mô hình WavLM tốn thời gian xử lý hơn so với việc tính toán log-mel spectrogram.

- Khác biệt miền ứng dụng: WavLM chủ yếu được huấn luyện trên dữ liệu giọng nói, khả năng mô hình nắm bắt các đặc điểm âm thanh chuyên biệt của dữ liệu máy có thể bị hạn chế, dẫn đến hiệu suất không đồng đều so với các đặc trưng phổ được thiết kế riêng cho lĩnh vực này như trong mô hình DAE. Cần các thử nghiệm để tinh chỉnh mô hình WavLM giúp nắm bắt tốt hơn các đặc điểm âm thanh chuyên biệt của máy công nghiệp.

5. Kết luận và hướng nghiên cứu

Nghiên cứu này đã đề xuất một kiến trúc dựa trên mô hình mới cho bài toán chẩn đoán lỗi máy công nghiệp dựa trên âm thanh, thay thế kiến trúc sử dụng DAE bằng cách tiếp cận phân loại sử dụng vector đặc trưng từ mô hình WavLM đã được huấn luyện trước. Các kết quả thực nghiệm trên tập dữ liệu MIMII cho thấy mô hình WavLM đạt hiệu suất tốt hơn so với DAE trong 3 loại máy (Fan và Pump, và Valve). Tuy nhiên, mô hình DAE vẫn chiếm ưu thế trong 1 loại máy (Slider), cho thấy tính chất bổ sung lẫn nhau giữa hai hướng tiếp cận.

Hướng nghiên cứu tiếp theo của bài báo này sẽ mở rộng phạm vi nghiên cứu đối với các tập dữ liệu khác trong các "Thách thức DCASE" [11]. Trong đó, kiến trúc hệ thống sẽ được nghiên cứu theo hướng tổng quát hóa mô hình hoặc tinh chỉnh các mô hình riêng biệt cho từng loại máy gồm: *(a) Mô hình lai (hybrid):* Kết hợp vector đặc trưng từ WavLM với các đặc trưng phổ truyền thống (như log-mel spectrogram) để tận

dụng đồng thời ưu điểm của các mô hình; *(b) Fine-tuning WavLM theo miền ứng dụng:* Tinh chỉnh mô hình WavLM biểu diễn tốt hơn đặc điểm âm thanh chuyên biệt của các máy công nghiệp để tăng hiệu suất phát hiện lỗi.

TÀI LIỆU THAM KHẢO

[1]. Ghazali, M. H. M., & Rahiman, W. (2021). Vibration analysis for machine monitoring and diagnosis: a systematic review. *Shock and Vibration*, *2021*, 9469318.

[2]. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587.

[3]. Tang, L., Tian, H., Huang, H., Shi, S., & Ji, Q. (2023). A survey of mechanical fault diagnosis based on audio signal analysis. *Measurement*, *220*, 113294.

[4]. Senanayaka, A., Lee, P., Lee, N., Dickerson, C., Netchaev, A., & Mun, S. (2024). Enhancing the accuracy of machinery fault diagnosis through fault source isolation of complex mixture of industrial sound signals. The International Journal of Advanced Manufacturing Technology, *133*(11), 5627-5642.

[5]. Lou, C., Atoui, M. A., & Li, X. (2024). Recent deep learning models for diagnosis and health monitoring: A review of research works and future challenges. *Transactions of the Institute of Measurement and Control*, *46*(14), 2833-2870.

[6]. Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., ... & Wei, F. (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1505-1518.

[7]. Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (1909). MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. arXiv 2019. *arXiv preprint arXiv:1909.09347*.

[8]. Zhou, Z., et al. (2016). "Deep learning for anomaly detection: A survey." *arXiv preprint arXiv:1611.05247*.

[9]. Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, *33*, 12449-12460.

[10]. Hsu, W.-N., et al. (2021). "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units." *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3451-3460.

INDUSTRIAL MACHINERY FAULT DIAGNOSIS BASED ON AUDIO SIGNAL ANALYSIS USING A PRETRAINED MODEL

Nguyen Duc Hoang[1]

ABSTRACT

Industrial machinery fault diagnosis represents a critical challenge to ensure optimal performance across manufacturing systems, yet conventional approaches relying on physical sensors often lack flexibility and scalability. This paper proposes a novel approach based on acoustic signal analysis, leveraging pre-trained WavLM models to extract high-level features directly from raw audio signals. Experimental validation on the MIMII dataset demonstrates that the proposed architecture achieves superior performance compared to conventional Deep Autoencoder (DAE) methods, particularly in high-noise environments. This study has demonstrated the substantial potential of applying pre-trained audio models to machine fault diagnosis in industrial environments, while suggesting promising directions for future improvements, including hybrid architectures and domain-specific model fine-tuning strategies.

Keywords: *MIMII Dataset, industrial Machine Fault Diagnosis, fault Diagnosis, signal Analysis, pretrained Model, WavLM, transfer Learning.*

[1]Trường Đại học Phạm Văn Đồng; Email: duchoang@pdu.edu.vn