

Phân Tích Âm Thanh Phổi Sử Dụng Phương Pháp Học Máy - Một Bước Tiến Mới Trong Kỹ Thuật Chẩn Đoán Bệnh Hô Hấp

Nguyễn Thị Kim Trúc*, Trần Thị Minh Dung*, Cao Nguyễn Khoa Nam[†],
Nguyễn Hữu Hoàng[‡], Nguyễn Văn Sĩ[‡], Lê Khắc Bảo[‡]

*Khoa Điện, Trường Đại học Bách khoa Đà Nẵng,

[†]Khoa Điện, Trường Đại học Sư phạm Kỹ thuật Đà Nẵng,

[‡]Đại học Y Dược Thành phố Hồ Chí Minh

Email: {ntktruc, tmdung}@dut.udn.vn, caonam@gmail.com, drhuuhoang@gmail.com,
{si.nguyen, lekhaobao}@ump.edu.vn

Tóm tắt nội dung—Phân tích âm thanh phổi sử dụng các phương pháp tính đã và đang mang lại lợi ích cho việc hỗ trợ chẩn đoán bệnh lý hô hấp sử dụng máy tính, lưu trữ dữ liệu và theo dõi tình trạng sức khỏe của cá nhân trong lĩnh vực chăm sóc sức khỏe. Trong đó, các hướng nghiên cứu về phát hiện các loại âm thanh bất thường của phổi cũng như phân loại bệnh hô hấp thu hút nhiều sự quan tâm của các nhà nghiên cứu. Bài báo này tổng hợp các hướng tiếp cận khác nhau của các hệ thống phân tích âm thanh phổi sử dụng phương pháp học máy cũng như phương pháp học sâu. Chúng tôi sẽ cung cấp tổng quan lý thuyết một cách có cấu trúc các hệ thống phân tích âm thanh phổi từ xử lý dữ liệu, cụ thể là xử lý tín hiệu âm thanh, trích xuất đặc trưng và tăng cường dữ liệu cho đến mô hình hóa dữ liệu thông qua các kiến trúc mạng nơ-ron và các cơ chế học. Bên cạnh đó, một số thuận lợi cũng như thách thức của việc ứng dụng thực tế của các hệ thống phân tích âm thanh phổi này cũng được giới thiệu một cách ngắn gọn.

Keywords-học máy; học sâu; phân tích âm thanh phổi; phân loại bệnh hô hấp; phân loại âm thanh phổi

I. GIỚI THIỆU CHUNG

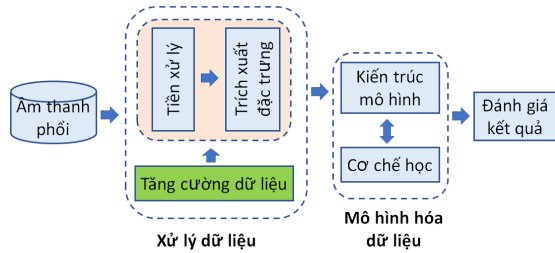
Các bệnh hô hấp hiện nay đang là một trong những nguyên nhân chính dẫn đến tử vong cho con người. Theo Tổ chức Y tế Thế giới (WHO), năm bệnh hô hấp hàng đầu bao gồm hen suyễn, bệnh tắc nghẽn phổi mạn tính (chronic obstructive pulmonary disease - COPD), nhiễm trùng cấp tính đường hô hấp dưới (acute lower respiratory tract infections), ung thư phổi và viêm phổi gây ra nhiều vấn đề nghiêm trọng đến sức khỏe con người cũng như hơn bảy triệu trường hợp tử vong mỗi năm trên toàn thế giới. Thêm vào đó, theo như [1] các bệnh hô hấp hàng đầu ngoại trừ ung thư phổi cũng tăng đáng kể trong suốt đại dịch Covid-19. Triệu chứng

của các bệnh hô hấp này có đặc điểm giống nhau cao và những âm thanh bất thường trong khi hít thở có thể là nhân tố quan trọng có thể tìm thấy trong suốt quá trình chẩn đoán [2]. Bởi vì những hậu quả nghiêm trọng của các bệnh hô hấp, việc đưa ra một phương pháp chẩn đoán sớm và chính xác các loại bệnh này thực sự cấp thiết.

Các âm thanh (tiếng) phổi được xem là các chỉ thị liên quan của sức khỏe hô hấp [3]. Âm thanh phổi được chia làm hai loại bao gồm tiếng phổi bình thường và tiếng phổi bất thường [2]. Tiếng phổi bình thường được nghe thấy khi phổi không có bất thường hay rối loạn. Tiếng phổi bất thường là những âm thanh đặc trưng xuất hiện trên nền tiếng phổi bình thường. Tiếng phổi bất thường có thể chia thành hai loại là liên tục như tiếng khò khè (wheezes) và không liên tục như tiếng ran nổ (fine crackles), tiếng ran ầm (coarse crackles), ...

Các phương pháp thính chẩn truyền thống sử dụng ống nghe có các đặc điểm sau: tiếng phổi được đánh giá dựa vào kinh nghiệm của bác sĩ, không thể cung cấp việc giám sát một cách liên tục và yêu cầu cần có một chuyên gia đã được huấn luyện. Ngoài ra, đặc điểm của tiếng phổi nằm trong dải tần số thấp, nơi thính giác của con người có độ nhạy hạn chế và dễ bị nhiễu [4]. Các phương pháp tính toán cho phân tích âm thanh phổi (Computational lung sound analysis - CLSA) đang được phát triển để có thể tự động nhận dạng và phân loại các tiếng phổi bất thường nhằm hỗ trợ cho chẩn đoán các bệnh hay các tình trạng hô hấp. CLSA sử dụng các thiết bị ghi âm số, kỹ thuật xử lý tín hiệu, và các thuật toán học máy. Chính vì vậy, các hệ thống CLSA mang lại lợi ích cho việc chẩn đoán hỗ trợ bởi máy tính, lưu trữ số và giám sát trong chăm sóc sức khỏe [2].

Các hệ thống CLSA có hai tác vụ nhận dạng/



Hình 1: Các bước xử lý của một hệ thống phân loại âm thanh phổi điển hình.

phân loại chính: (i) phân loại tiếng phổi bất thường (adventitious lung sound classification - ALSC) và phân loại bệnh hô hấp (respiratory disease classification - RDC). Phân loại tiếng phổi bất thường có nhiệm vụ: nhận dạng các sự kiện bất thường của tiếng phổi như là tiếng ran (crackles), tiếng khò khè (wheezes) hay các loại tiếng phổi bất thường khác, hoặc nhận dạng tiếng phổi bình thường và tiếng phổi bất thường [5]–[8]. Phân loại bệnh hô hấp cũng bao gồm các loại khác nhau như phân loại hai nhóm gồm khỏe mạnh và bệnh [4], hay phân loại ba nhóm gồm khỏe mạnh, bệnh mạn tính và bệnh không mạn tính [5], [6], hoặc phân loại nhiều bệnh riêng biệt như hen suyễn, bệnh tắc nghẽn phổi mạn tính, nhiễm trùng cấp tính đường hô hấp dưới, ung thư phổi và viêm phổi [9], [10]. Các hệ thống này được đánh giá trên một vài cơ sở dữ liệu không được công khai như tập dữ liệu đa kênh [11] và được công khai như cơ sở dữ liệu bệnh viện đại học Abdullah 2020 [12], HF_Lung_V1 [13] hay cơ sở dữ liệu truy cập mở phổ biến nhất hiện nay ICBHI 2017 [3].

Trong bài báo này, chúng tôi tổng hợp một cách hệ thống cho phân loại âm thanh phổi được minh họa ở Hình 1. Âm thanh phổi được xử lý bởi các kỹ thuật xử lý tín hiệu như tiền xử lý tín hiệu, trích xuất đặc trưng và tăng cường dữ liệu (xem Mục II). Các đặc trưng trích xuất được đưa vào kiến trúc mạng nơ-ron và được xử lý với các cơ chế học khác nhau để phân loại (xem Mục III).

Như một phần bổ sung của bài báo này, các bài báo tổng quan lý thuyết như [2], [14] và [5] cung cấp nhiều tham chiếu tài liệu tham khảo cho các nội dung được mô tả.

II. XỬ LÝ DỮ LIỆU

Các tín hiệu âm thanh phổi ghi âm được là các tín hiệu số. Đặc điểm của bản ghi âm thanh này phụ thuộc rất lớn vào các thiết bị ghi âm như sử

dụng đơn, đa ống nghe hay micro điện tử. Trong phần này, chúng tôi sẽ khái quát (i) các kỹ thuật tiền xử lý tín hiệu, (ii) các đặc trưng được sử dụng trong các hệ thống phân loại âm thanh phổi và (iii) các phương pháp tăng cường dữ liệu.

A. Tiền xử lý tín hiệu

Tín hiệu âm thanh phổi thu được bằng cách sử dụng các thiết bị ghi âm ở các điều kiện khác nhau, vì vậy chúng cần được tiền xử lý để có cùng đặc điểm. Các bước tiền xử lý tín hiệu bao gồm: (i) lọc nhiễu, (ii) tái lấy mẫu tín hiệu, (iii) chuẩn hóa biên độ, (iv) phân đoạn và điền đầy phân đoạn.

Tuy nhiên, hiện nay có nhiều hệ thống phân loại âm thanh phổi dùng phương pháp học sâu có thể bỏ qua các bước lọc nhiễu và chuẩn hóa biên độ nhờ vào khả năng trích xuất đặc trưng bậc cao và chuẩn hóa dữ liệu của kiến trúc mạng nơ-ron sâu (DNNs) [5], [15]. Thêm vào đó có một số hệ thống tách và điền đầy hoặc thay đổi kích thước của các tín hiệu sau khi được biến đổi sang dạng 2D như kỹ thuật xử lý ảnh [16], thay vì xử lý với các tín hiệu âm thanh 1D.

1) *Lọc nhiễu*: Thường được áp dụng để loại bỏ những thông tin không liên quan như tiếng tim, nhiều nền hay các nhiễu gây ra bởi thiết bị ghi âm. Các bộ lọc thông dải từ 15Hz đến 1800 Hz thường được sử dụng cả trong những hệ thống thông thường và các hệ thống sử dụng học sâu.

2) *Tái lấy mẫu*: Tín hiệu âm thanh phổi được ghi âm với dải tần số lấy mẫu rộng [3]. Vì các tín hiệu phổi bất thường thường nằm trong dải 2 kHz, tần số lấy mẫu là 4 kHz hoặc cao hơn. Để thuận tiện cho các bước xử lý tiếp theo thì tần số lấy mẫu của tất cả dữ liệu thường phải giống nhau.

3) *Chuẩn hóa biên độ*: Thường được sử dụng cho mỗi tín hiệu nhằm làm cho các tín hiệu có giá trị nằm trong khoảng -1 đến 1. Việc này sẽ bù lại sự khác nhau về cường độ của dữ liệu được thu thập từ các nguồn khác nhau mà vẫn có thể đảm bảo giữ được những đặc điểm thống kê quan trọng của âm thanh hô hấp.

4) *Phân đoạn và điền đầy phân đoạn*: Các hệ thống phân loại tiếng phổi dựa trên học sâu sử dụng các mô hình mạng nơ-ron tích chập (Convolutional neural networks - CNNs) thường đòi hỏi một kích thước giống nhau của dữ liệu đầu vào. Thêm vào đó, các tập dữ liệu tiếng phổi có chu kỳ hô hấp cũng như chiều dài của đoạn ghi âm có độ dài khác nhau. Chính vì vậy, các tín hiệu âm thanh thường được chia tách có hoặc không có trùng lặp (overlap) thành

những phân đoạn với chiều dài cố định và chúng cần được điền đầy tín hiệu để đạt được chiều dài đó. Hiện nay, có nhiều phương pháp khác nhau để thực hiện như điền đầy phân đoạn bằng các mẫu 0 (zero padding [17]), điền đầy sử dụng các mẫu trong cùng bản ghi âm (sample padding) [5], hay điền đầy thông minh (smart padding) [18]. Theo các nghiên cứu [5], [18], các hệ thống sử dụng các mẫu padding và smart padding có độ chính xác cao hơn so với hệ thống sử dụng zero padding với cùng thông số cài đặt của hệ thống.

B. Trích xuất đặc trưng

Trích xuất đặc trưng là một bước quan trọng đối với các thuật toán trong kỹ thuật nhận dạng mẫu (pattern recognition) và phân loại. Bước này tăng cường thông tin cho việc phân loại. Theo như các cách tiếp cận khác nhau của các hệ thống phân loại âm thanh phổ, có thể phân chia theo hai hướng: (i) đặc trưng dành cho các bộ phân loại thông thường (conventional classifiers) (ii) đặc trưng là các biểu diễn 2D trong miền thời gian - tần số cho phương pháp học sâu.

1) *Đặc trưng dành cho các bộ phân loại thông thường*: Báo cáo tổng quan [2] đã tổng hợp một cách có hệ thống các hệ thống phát hiện và phân loại tiếng phổ bất thường được giới thiệu giữa các năm 1938 và 2016, hầu hết các thuật toán dựa vào việc nhận dạng mẫu và độ chính xác của chúng phụ thuộc phần lớn vào các đặc trưng được trích xuất thủ công. Cụ thể, các phương pháp trích xuất đặc trưng được sử dụng để tạo ra các vector đặc trưng mà các dữ liệu cùng một nhóm sẽ được phân bố gần nhau trong không gian đặc trưng, đồng thời cực đại khoảng cách với các dữ liệu của các nhóm khác. Tập hợp các đặc trưng thủ công này thường bao gồm các đặc trưng được trích xuất trong miền thời gian, trong miền tần số như cepstral, đặc trưng của spectral (cụ thể: tâm spectral, độ sáng, entropy, độ lệch và kurtosis, flux), đặc trưng về giai điệu (cụ thể: đỉnh và sắc phổ), các hệ số MFCCs - mel frequency cepstral coefficients, các hệ số dự đoán tuyến tính (LPCs - linear prediction coefficients), các hệ số cepstral dự đoán tuyến tính (LPCCs - linear prediction cepstral coefficients) hay các đặc trưng của biến đổi wavelet rời rạc hoặc liên tục (DWT/CWT). Những đặc trưng này vẫn tiếp tục được khai thác trong một số nghiên cứu gần đây) [8], [14],

Bên cạnh đó, đối với các phương pháp này việc lựa chọn đặc trưng cũng được áp dụng để loại bỏ

các đặc trưng không liên quan và giữ lại những đặc trưng có ý nghĩa trước khi được đưa vào các bộ phân loại. Nhờ đó hiệu suất của bộ phân loại được cải thiện cả về độ chính xác lẫn chi phí tính toán.

2) *Biểu diễn 2D trong miền tần số - thời gian dành cho thuật toán học sâu*: Hiện nay, các loại biểu diễn 2D được sử dụng như spectrogram của các phép biến đổi Fourier thời gian ngắn (short-time Fourier transform - STFT), phép biến đổi hằng số Q (constant - Q transform), phép biến đổi Stockwell (S-transform) hoặc scalogram của phép biến đổi wavelet [5], [6], [18]. Ngoài ra, các bộ lọc được thiết kế - bắt chước theo hệ thống thính giác của con người - như mel và gammatone cũng được sử dụng nhằm tạo ra các biểu diễn 2D như mel spectrogram, gammatonegram hay MFCCs. Các dải tần số của bộ lọc được lựa chọn cũng góp phần đáng kể trong việc giảm số lượng các đặc trưng trong mỗi khung thời gian. Bên cạnh các phương pháp biến đổi này, phương pháp phân tích chế độ thực nghiệm - empirical mode decomposition (EMD) cũng được sử dụng để phân tích tín hiệu âm thanh thành các hàm chế độ nội tại - intrinsic mode functions (IMFs), chúng được xử lý thêm sau đó để trở thành những biểu diễn thị giác 2D của tín hiệu [8], [9]. Hiện nay, nhằm cung cấp thêm nhiều thông tin cho các hệ thống phân loại âm thanh phổ, nhiều đặc trưng thị giác 2D được khai thác cùng lúc. Chúng có thể được nối với nhau để tạo thành một ma trận 2D [11] hay có thể được sử dụng như các đầu vào 2D cho các hệ thống đa đầu vào của DNNs [19], hoặc dùng làm dữ liệu đầu vào cho các hệ thống tổng hợp (ensemble) của các mô hình DNNs sử dụng một dữ liệu đầu vào [15].

C. Phương pháp tăng cường dữ liệu

Các tập dữ liệu của âm thanh phổ thường khá giới hạn về số lượng bệnh nhân và các bản ghi âm. Hơn nữa, các nhóm dữ liệu (ví dụ tiếng phổ bình thường, tiếng khò khè và tiếng ran) trong các tập dữ liệu cũng thường không cân bằng. Chính vì những đặc điểm bất lợi này làm cho việc đánh giá hiệu suất của các thuật toán hoặc bị quá thấp hoặc quá cao, cũng như khả năng tổng quát hóa của mô hình bị hạn chế. Để khắc phục những nhược điểm đó, phương pháp tăng cường dữ liệu đang là một giải pháp phổ biến và hiệu quả được giới thiệu trong các hệ thống phân loại âm thanh phổ gần đây. Phương pháp này giúp cân bằng dữ liệu cho các nhóm và đặc biệt có thể tăng cường tính đa dạng cho tập dữ liệu dùng để huấn luyện mô hình. Chúng ta có thể

chia kỹ thuật tăng cường dữ liệu thành hai hướng chính: (i) trong miền thời gian, nghĩa là tác động vào các tín hiệu âm thanh ở dạng tín hiệu 1D, (ii) trong miền thời gian - tần số, nghĩa là tác động vào các biểu diễn 2D của tín hiệu.

1) *Miền thời gian*: Các dữ liệu mới cho tập huấn luyện được tạo ra trong miền thời gian. Cụ thể, các phương pháp biến đổi được thực hiện trên các tín hiệu âm thanh như kéo dài thời gian (time stretching) bằng cách thay đổi tốc độ và chiều dài tín hiệu mà không làm ảnh hưởng đến cao độ của tín hiệu, hoặc dịch cao độ, dịch biên độ, điều chỉnh tốc độ, ... Ngoài ra, phương pháp smart padding [18] cũng được sử dụng cho mục đích tăng cường dữ liệu.

2) *Miền thời gian - tần số*: Các dữ liệu mới cho tập huấn luyện được tạo ra bằng các cách phổ biến sau. Kỹ thuật xáo trộn chiều dài đường âm thanh (Vocal tract length perturbation - VTLP), đây là một kỹ thuật phổ biến trong lĩnh vực nhận dạng giọng nói (speech recognition). Nó được áp dụng để tăng cường dữ liệu bằng cách tạo ra các mẫu dữ liệu 2D mel spectrogram mới thông qua việc làm thay đổi dải tần số của bộ lọc mel, nhờ đó không làm phá hủy thông tin có ích của dữ liệu ban đầu [4], [5]. Ngoài ra, vì các tín hiệu âm thanh được chuyển đổi thành các biểu diễn 2D nên được xem như là ảnh 2D cho các hệ thống thị giác máy tính, nên tăng cường dữ liệu có thể sử dụng những phương pháp tương tự, cụ thể tăng cường dữ liệu bằng cách trộn dữ liệu (mixup data augmentation) bất theo phương ngang hoặc phương thẳng đứng, xén ảnh ngẫu nhiên hoặc biến đổi màu sắc [5], [15], [18]. Đặc biệt, các dẫn xuất của phương pháp mạng sáng tạo đối nghịch - generative adversarial network (GAN) cũng được áp dụng cho các spectrogram [7]. Điều đáng lưu ý là vì những hạn chế của các tập dữ liệu âm thanh phổ biến nên một vài kỹ thuật tăng cường dữ liệu thường được sử dụng kết hợp với nhau trong các nghiên cứu gần đây nhằm tạo ra tính đa dạng cũng như số lượng hữu hiệu của dữ liệu phục vụ cho việc huấn luyện mô hình kiến trúc mạng.

III. MÔ HÌNH HÓA DỮ LIỆU

A. Học máy

Hai hướng chính sử dụng học máy trong các hệ thống phân loại âm thanh phổ biến: (i) các bộ phân loại thông thường dành cho các đặc trưng được trích xuất thủ công và (ii) mạng nơ-ron sâu được giới thiệu cụ thể như sau:

1) *Bộ phân loại thông thường*: Thường được sử dụng kết hợp các đặc trưng được trích xuất thủ công trong các hệ thống phân loại âm thanh phổ biến. Chúng được tổng hợp trong [2] bao gồm các thuật toán như sau: Random forest (RF), Support vector machines (SVM), Multi-layer perceptrons (MLPs), k-nearest neighbors (k-NN), Gaussian mixture models (GMMs), Decision tree, Bayes rule-based classifiers, Self organised map (SOM), Linear discriminant analysis (LDA), K-mean clustering, Logistic regression.

2) *Kiến trúc học sâu*: Thường sử dụng các biểu diễn 2D được giới thiệu ở Mục C làm dữ liệu đầu vào. Các kiến trúc mạng nơ-ron này thực hiện trích xuất đặc trưng bậc cao kết hợp với chức năng phân loại. Một số kiến trúc mạng nơ-ron được sử dụng như sau:

CNNs: các tác vụ phân loại của âm thanh phổ biến và bệnh hô hấp chủ yếu sử dụng các kiến trúc mạng nơ-ron của CNNs với đầu vào là các dữ liệu 2D. Mô hình CNN thường được khai thác cho các đoạn âm thanh phổ biến dài như chu kỳ hít thở, khi đó tín hiệu âm thanh này có thể chứa một vài sự kiện âm thanh bất thường. Các mô hình CNN được sử dụng theo hai cách như sau:

- **Mô hình kiến trúc CNN tiêu chuẩn**: nhiều kiến trúc khác nhau của CNNs được báo cáo trên tập dữ liệu hình ảnh lớn nhất hiện nay - ImageNet - được tái sử dụng cho các hệ thống phân loại âm thanh phổ biến, cụ thể như VGGs, Resnets, AlexNets, InceptionNets, Googlenets hay Mobilenets. Các mô hình kiến trúc này thường được sử dụng cho các hệ thống áp dụng học chuyển tiếp - transfer learning.
- **Mô hình kiến trúc CNN tùy biến**: có một số mô hình kiến trúc CNN được tự định nghĩa và được khai thác trên tập dữ liệu âm thanh phổ biến. Bên cạnh đó, một số mô hình kiến trúc nhiều đầu vào của CNN [19] hay sử dụng các lớp pooling song song [20] cũng được sử dụng nhằm tận dụng hiệu quả các biểu diễn đặc trưng khác nhau. Tuy nhiên, những mô hình kiến trúc này thường công kênh và chi phí tính toán cao. Vì vậy, các mô hình kiến trúc CNN hiệu quả về tài nguyên (resource - efficiency CNNs) với mô hình kiến trúc nhỏ như CNN có bộ trọng số nhẹ (lightweight CNNs) [9], các lớp tích chập theo chiều sâu (depth-wise convolutional layers) hay lượng tử hóa trọng số của mô hình (weight quantization) [10] được sử dụng. Bên cạnh đó, CNNs 1D cũng được

sử dụng như trong nghiên cứu [12].

RNNs: Một số hệ thống phân loại âm thanh phổ biến sử dụng các phiên bản xuất khác nhau của mạng nơ-ron hồi qui như mạng nơ-ron chứa đơn vị chức năng hồi quy có kiểm soát (gated recurrent units - GRUs), mạng nơ-ron có các đơn vị nhớ ngắn hạn (long short-term memories - LSTMs) hay các phiên bản hai chiều mở rộng của chúng như BiGRUs, BiLSTMs. Mạng nơ-ron RNNs có thể giải quyết các tác vụ liên quan đến phát hiện/phân loại các sự kiện âm thanh phổ biến bất thường và các tác vụ phân loại âm thanh phổ biến bất thường cũng như các bệnh hô hấp [7].

Các thuật toán lai: có một vài hướng nghiên cứu như sau.

- **Mô hình lai sử dụng mạng nơ-ron sâu:** đây là sự kết hợp của các mô hình CNNs với hoặc là các phiên bản khác nhau của RNNs [12] hay một cơ chế đặc biệt như tập hợp các chuyên gia (mixture of experts - MoEs) [6].
- **Mô hình học máy kết hợp:** đây là mô hình được phát triển thông qua sự kết hợp giữa học sâu và các hệ thống phân loại âm thanh phổ biến thông thường. Ví dụ, các đặc trưng thủ công được đưa vào MLPs [11]. Một hướng tiếp cận ngược lại cũng được khai thác, cụ thể là các bộ phân loại thông thường được sử dụng để phân loại cho các đặc trưng được huấn luyện từ các kiến trúc CNNs (embedding) [20].
- **Ensembles:** một số hệ thống phân loại âm thanh phổ biến sử dụng nhiều mô hình khác nhau được phối hợp với nhau gọi là ensemble. Nguyên tắc của các thuật toán ensemble là cho phép dự đoán đầu ra của các bộ phân loại được kết hợp lại với nhau nhằm cải thiện độ chính xác [6]. Tuy nhiên, trong các ứng dụng thực tế ensemble không được lựa chọn bởi thuật toán này đòi hỏi các nhu cầu về tính toán cũng như tài nguyên tính toán lớn.

B. Cơ chế học

Học chuyển tiếp - transfer learning: Học chuyển tiếp là một giải pháp hiệu quả đối với việc cải thiện hiệu suất của các hệ thống phân loại âm thanh hiện đại. Bởi vì các hệ thống này có thể tiết kiệm được thời gian và sử dụng kiến thức đã được huấn luyện một cách hiệu quả từ tập dữ liệu nguồn. Cụ thể, nó có thể cải thiện khả năng làm việc của mô hình thông qua việc truyền những kiến thức đã học từ một số phần hoặc tất cả các phần của mô hình DNNs đã được huấn luyện trước đó trên tập

dữ liệu nguồn (ví dụ ImageNet, ICBHI 2017, Audio set). Khi sử dụng mô hình này cho tập dữ liệu đích, có hai cách tiếp cận khác nhau: (i) kiến trúc từ các lớp trên cùng của mô hình - phần biểu diễn của mô hình (representation part), sẽ không thay đổi trong suốt quá trình huấn luyện kế tiếp trên tập dữ liệu đích, ta nói các trọng số của phần biểu diễn trong mô hình bị "đóng băng". Trong trường hợp này, học chuyển tiếp đóng vai trò của bộ trích xuất đặc trưng bậc cao [18], [19]; (ii) tất cả hay một phần mô hình đã được huấn luyện sẽ được hiệu chỉnh trong suốt quá trình huấn luyện kế tiếp trên tập dữ liệu đích của âm thanh phổ biến bất thường [5]. Trong các nghiên cứu phân loại âm thanh phổ biến gần đây sử dụng phương pháp học chuyển tiếp, mô hình kiến trúc ResNets được huấn luyện trên tập dữ liệu hình ảnh ImageNet được sử dụng phổ biến.

Bên cạnh đó, phương pháp đồng hiệu chỉnh và chuẩn hóa ngẫu nhiên là các phiên bản khác của học chuyển tiếp cũng được khai thác và cải thiện được đáng kể hiệu suất của hệ thống [5]. Thêm vào đó, phương pháp chưng cất kiến thức (knowledge distillation), một cơ chế trò - thầy (student - teacher scheme) được áp dụng cho phân loại bệnh hô hấp cũng được sử dụng [6].

IV. KẾT LUẬN

Sự ra đời của phương pháp học máy và học sâu đã cải thiện được hiệu suất của các hệ thống phân loại âm thanh phổ biến hiện đại như sau. (i) Các kiến trúc mạng nơ-ron sâu (DNNs) mạnh được khai thác một cách rộng rãi để trích xuất các đặc trưng bậc cao. (ii) Học chuyển tiếp (transfer learning) là một kỹ thuật hiệu quả trong việc giải quyết các hạn chế về số lượng của dữ liệu trong các tập dữ liệu thông qua việc khai thác kiến thức đã được học từ tập dữ liệu cùng hay khác lĩnh vực trước đó. (iii) Kỹ thuật tăng cường dữ liệu (data augmentation) ngày càng được ứng dụng trong các hệ thống phân loại âm thanh nhằm cải thiện tính đa dạng và số lượng dữ liệu cũng như cân bằng các nhóm dữ liệu của các tập dữ liệu âm thanh phổ biến. Điều này giúp giải quyết tình trạng quá khớp (overfitting) của các kiến trúc mạng nơ-ron sâu.

Bài báo này cung cấp một cái nhìn đầy đủ của các nghiên cứu gần đây của các hệ thống phân loại âm thanh phổ biến sử dụng phương pháp tính. Có hai hướng tiếp cận quan trọng như phân loại các âm thanh phổ biến bất thường và bệnh hô hấp. Chúng tôi tổng hợp và phân loại một cách có cấu trúc các chủ đề từ xử lý dữ liệu như tiền xử lý tín hiệu âm

thanh, trích xuất đặc trưng và tăng cường dữ liệu cho đến việc mô hình hóa dữ liệu như là các kiến trúc mạng nơ-ron và các cơ chế học. Song song với những tiến bộ gần đây trong khoa học kỹ thuật, một thách thức đặt ra cho các mô hình phân loại âm thanh phổi là áp dụng chúng cho các hệ thống chẩn đoán trong thực tế bởi việc thiết lập lâm sàng gặp nhiều bất lợi như nhiễu, chất lượng tín hiệu không tốt. Tuy nhiên, chúng ta có thể thấy rằng việc kết hợp của các mạng nơ-ron sâu với phương pháp học chuyển tiếp và tăng cường dữ liệu mở ra nhiều con đường đầy triển vọng cho các hệ thống phân loại âm thanh phổi.

TÀI LIỆU

- [1] M.T. Barbosa *et al.*, “The “big five” lung diseases in covid-19 pandemic—a google trends analysis,” *Pulmonology*, vol. 27, no. 1, pp. 71–72, 2021.
- [2] R. X. A. Pramono *et al.*, “Automatic adventitious respiratory sound analysis: A systematic review,” *PLoS one*, vol. 12, no. 5, p. e0177926, 2017.
- [3] B. M. Rocha *et al.*, “A respiratory sound database for the development of automated classification,” in *Precision Medicine Powered by pHealth and Connected Health*. Springer, 2018, vol. 66, pp. 33–37.
- [4] M. Aykanat *et al.*, “Classification of lung sounds using convolutional neural networks,” *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, pp. 1–9, 2017.
- [5] T. Nguyen and F. Pernkopf, “Lung sound classification using co-tuning and stochastic normalization,” *IEEE Transactions on Biomedical Engineering*, 2022.
- [6] L. Pham *et al.*, “Cnn-moe based framework for classification of respiratory anomalies and lung disease detection,” *IEEE journal of biomedical and health informatics*, vol. 25, no. 8, pp. 2938–2947, 2021.
- [7] K. Kochetov and A. Filchenkov, “Generative adversarial networks for respiratory sound augmentation,” in *Proc. of 2020 ICCRIS*, 2020, pp. 106–111.
- [8] S. I. Khan and R. B. Pachori, “Automated classification of lung sound signals based on empirical mode decomposition,” *Expert Systems with Applications*, vol. 184, p. 115456, 2021.
- [9] S. B. Shuvo *et al.*, “A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram,” *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [10] J. Acharya and A. Basu, “Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning,” *IEEE trans. on biomedical circuits and systems*, vol. 14, no. 3, pp. 535–544, 2020.
- [11] E. Messner *et al.*, “Multi-channel lung sound classification with convolutional recurrent neural networks,” *Computers in Biology and Medicine*, vol. 122, p. 103831, 2020.
- [12] M. Fraiwan *et al.*, “Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2021.
- [13] F. S. Hsu *et al.*, “Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—hf_lung_v1,” *PLoS one*, vol. 16, no. 7, p. e0254134, 2021.
- [14] R. X. A. Pramono *et al.*, “Evaluation of features for classification of wheezes and normal respiratory sounds,” *PLoS one*, vol. 14, no. 3, p. e0213659, 2019.
- [15] L. Pham *et al.*, “Inception-based network and multi-spectrogram ensemble applied to predict respiratory anomalies and lung diseases,” in *Proc. 2021 EMBC*. IEEE, 2021, pp. 253–256.
- [16] S. Gupta *et al.*, “Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning,” *Biomedical Signal Processing and Control*, vol. 70, p. 102947, 2021.
- [17] S. Y. Jung *et al.*, “Efficiently classifying lung sounds through depthwise separable cnn models with fused stft and mfcc features,” *Diagnostics*, vol. 11, no. 4, p. 732, 2021.
- [18] S. Gairola *et al.*, “Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting,” in *Proc. 2021 EMBC*. IEEE, 2021, pp. 527–530.
- [19] T. Nguyen and F. Pernkopf, “Crackle detection in lung sounds using transfer learning and multi-input convolutional neural networks,” in *Proc. 2021 EMBC*. IEEE, 2021, pp. 80–83.
- [20] F. Demir *et al.*, “Classification of lung sounds with cnn model using parallel pooling structure,” *IEEE Access*, vol. 8, pp. 105 376–105 383, 2020.