

Hệ số Cepstral tần số Mel hiệp đồng và Biến đổi Fourier thời gian ngắn để tăng cường phát hiện trạng thái ong bằng cách sử dụng máy học

Thi-Thu-Hồng Phan¹ · 

Khoa Trí tuệ nhân tạo, Đại học FPT, Đà Nẵng, Việt Nam

Tóm tắt. Nghiên cứu này điều tra tiềm năng của phân tích âm thanh để phát hiện trạng thái của ong trong tổ ong, một thách thức quan trọng đối với người nuôi ong. Chúng tôi đề xuất một phương pháp mới kết hợp Hệ số Cepstral tần số Mel (MFCC) và Biến đổi Fourier thời gian ngắn (STFT), tạo ra dữ liệu đại diện nhiều thông tin hơn để phân loại trạng thái của ong. Thử nghiệm trên một tập dữ liệu thực tế chứng minh rằng bộ phân loại Rừng ngẫu nhiên sử dụng các tính năng hiệp lực được trích xuất từ MFCC và STFT hoạt động tốt hơn các mô hình chỉ dựa trên các tính năng MFCC hoặc STFT, cải thiện đáng kể độ chính xác phân loại (lên đến 87,2%). Phương pháp tích hợp này mang lại lợi thế trong việc nắm bắt cả chi tiết quang phổ (MFCC) và thông tin thời gian (STFT) có khả năng dẫn đến cải thiện độ chính xác phân loại để phát hiện trạng thái của ong. Những phát hiện này đóng góp những hiểu biết có giá trị để phát triển các hệ thống giám sát sức khỏe đàn ong mạnh mẽ.

Từ khóa: Tiếng ong kêu-Không có Nữ hoàng-Trạng thái của ong-STFT-MFCC-Phương pháp học máy-Tính năng kết hợp.

1 Giới thiệu

Ong mật là loài thụ phấn thiết yếu, đóng vai trò quan trọng trong việc duy trì sự cân bằng của hệ sinh thái và tăng năng suất nông nghiệp. Khả năng kiếm ăn khéo léo của chúng hỗ trợ sự sinh sản của thực vật, bao gồm các loại cây trồng chính, và các sản phẩm từ tổ ong của chúng, chẳng hạn như mật ong và sáp ong, đáp ứng nhiều nhu cầu khác nhau của con người. Hơn nữa, quần thể ong mật đóng vai trò là chỉ số về sức khỏe môi trường nói chung, nhấn mạnh tầm quan trọng của chúng và sự cần thiết của các biện pháp bảo tồn.

Việc duy trì đàn ong khỏe mạnh là rất quan trọng đối với người nuôi ong. Họ phải đối mặt với những mối đe dọa liên tục như đàn ong bay, ve Varroa và nghiêm trọng hơn là ong chúa mất tích. Theo truyền thống, người nuôi ong dựa vào việc kiểm tra đàn ong thường xuyên để theo dõi đàn ong của họ. Điều này bao gồm việc mở tổ ong để kiểm tra trực quan tình trạng sức khỏe của đàn ong, xác nhận sự hiện diện của ong chúa và đánh giá tình trạng sức khỏe tổng thể của ong, v.v. Tuy nhiên, những cuộc kiểm tra này tốn nhiều thời gian, có khả năng gây rối loạn hành vi của ong và gây hại cho ong chúa. Điều này đòi hỏi một cách tiếp cận không xâm lấn để theo dõi đàn ong và tạo điều kiện cho các biện pháp can thiệp. Để giải quyết những hạn chế này, nhiều nghiên cứu đã khám phá các phương pháp dựa trên công nghệ tiên tiến [3, 8]. Internet of

¹ Tác giả liên hệ: hongptt11@fe.edu.vn

Vạn vật (IoT) và trí tuệ nhân tạo đóng vai trò quan trọng trong cách tiếp cận mới này. Các tổ ong được trang bị cảm biến có thể liên tục thu thập dữ liệu về nhiều yếu tố môi trường khác nhau như nhiệt độ và độ ẩm, và quan trọng nhất là các hệ thống này có thể thu được âm thanh của ong để phân tích thêm.

Phân tích âm thanh, kết hợp với các thuật toán học máy, hứa hẹn rất lớn cho việc giám sát tổ ong. Nghiên cứu đã chỉ ra rằng phương pháp này có hiệu quả trong việc phát hiện trạng thái ong trong đàn ong, với nhiều thuật toán được sử dụng bằng cách sử dụng âm thanh làm dữ liệu đầu vào. Ví dụ, Ruvunga et al. (2021) [13] đã đạt được độ chính xác 92% khi sử dụng mạng LSTM với hệ số cepstral tần số Mel (MFCC) để phát hiện ong không có ong chúa trong tổ ong. Tương tự như vậy, Nolasco et al. (2019) đã khám phá cả Máy vectơ hỗ trợ (SVM) và Mạng nơ-ron tích chập (CNN) với MFCC, đạt được độ chính xác cao (lên đến 94% khi bổ sung Biến đổi Hilbert Huang (HHT)). Trong [9], Phan et al. đã nghiên cứu các kỹ thuật điều chỉnh MFCC và siêu tham số mới để nâng cao hiệu suất nhận dạng âm thanh của ong. Truong et al. [14] đã phát triển một phương pháp dựa trên các mô hình học sâu để xác định âm thanh ong vo ve từ các âm thanh khác như tiếng ồn hoặc tiếng đé kêu. Barbisan và Riente [1] đạt được độ chính xác cao (lên đến 98,8%) khi sử dụng cả Mạng nơ-ron (NN) và Máy vectơ hỗ trợ (SVM) với 20 đặc điểm MFCC. Các nghiên cứu sâu hơn đã khám phá các phương pháp thay thế để phát hiện sự hiện diện của ong chúa, Fourer và cộng sự đã sử dụng STFT với Mạng nơ-ron tích chập (CNN), đạt được độ chính xác 96%. Tương tự như vậy, Ho và cộng sự (2023) đã sử dụng MFCC để trích xuất đặc điểm trên một tập dữ liệu thực, đạt độ chính xác cao nhất là 91,75%.

Trong [12], Rustam et al. tập trung vào việc phân loại âm thanh của ong thành ba loại: Bee, NoBee và NoQueen. Để đạt được mục tiêu này, các tác giả đã sử dụng nhiều kỹ thuật lựa chọn tính năng và thuật toán học máy. Họ đạt được độ chính xác cao hơn một chút với K-Nearest Neighbors (KNN) ở mức 0,83 so với 0,82 với Random Forest (RF). Tuy nhiên, hiệu suất của các mô hình này vẫn chưa tối ưu trên tập dữ liệu của chúng. Điều này đặt ra câu hỏi về sự kém hiệu quả của chúng và những cách tiềm năng để cải thiện hiệu quả của chúng. Thách thức này thúc đẩy chúng tôi tìm ra câu trả lời. Cụ thể, chúng tôi đề xuất một phương pháp mới kết hợp Hệ số Cepstral tần số Mel (MFCC) và Biến đổi Fourier thời gian ngắn (STFT), tạo ra dữ liệu đại diện nhiều thông tin hơn để phát hiện trạng thái của ong trong tổ ong. Kết quả thử nghiệm cho thấy các kỹ thuật này có thể cải thiện đáng kể độ chính xác của các phương pháp học máy trong việc nhận dạng trạng thái của ong trong tổ ong.

Phần còn lại của bài báo được cấu trúc như sau. Phần 2 cung cấp mô tả ngắn gọn về các phương pháp được sử dụng trong nghiên cứu. Phần 3 trình bày các thí nghiệm đã tiến hành, kết quả thu được và các thảo luận có liên quan. Phần 4 đưa ra các nhận xét kết luận tóm tắt các phát hiện chính và các hướng tiềm năng trong tương lai.

2 Phương pháp

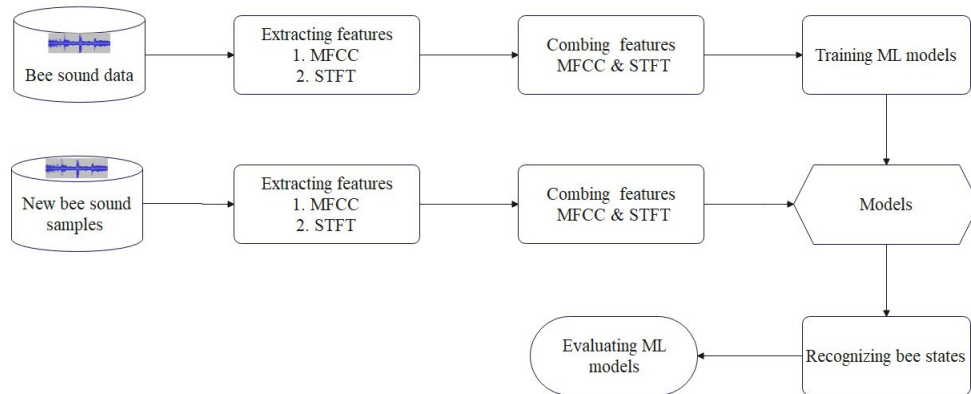
Bài báo này đề xuất một phương pháp phân loại các trạng thái tổ ong khác nhau thông qua phân tích âm thanh như hình 1. Chúng tôi tận dụng sự kết hợp của STFT và MFCC cho

trích xuất tính năng để nâng cao hiệu suất của các phương pháp ML cho nhiệm vụ phát hiện này.

Quá trình bắt đầu bằng việc thu thập các mẫu âm thanh từ tổ ong. Các mẫu này bao gồm nhiều hoạt động khác nhau của ong, nhằm mục đích thu thập các chữ ký âm thanh liên quan đến ba trạng thái riêng biệt: Sự hiện diện của ong, Không có sự hiện diện của ong và Không có sự hiện diện của ong chúa.

Các mẫu âm thanh thu thập được sau đó được xử lý để trích xuất các đặc điểm phù hợp với phân loại trạng thái ong. Công trình này sử dụng hai kỹ thuật riêng biệt: STFT và MFCC. Mỗi phương pháp cung cấp những hiểu biết có giá trị về dữ liệu âm thanh – STFT tiết lộ phân phối thời gian-tần số và MFCC nắm bắt được lớp phổ liên quan đến thính giác của con người. Sau khi trích xuất, các đặc điểm STFT và MFCC được kết hợp, tạo ra một biểu diễn toàn diện của dữ liệu âm thanh.

Những tính năng kết hợp này sau đó được đưa vào các mô hình học máy để phân loại trạng thái ong. Để đảm bảo khám phá toàn diện dữ liệu và xác định phương pháp tiếp cận phù hợp nhất, chúng tôi nghiên cứu sáu thuật toán học máy khác nhau: RF, ET (Extra Trees), XGBoost (eXtreme Gradient Boosting), SVM, KNN và LR. Mỗi thuật toán đều có điểm mạnh riêng, cho phép phân tích dữ liệu âm thanh của tổ ong một cách mạnh mẽ. Khám phá này có tiềm năng đạt được kết quả phân loại trạng thái ong chính xác nhất.



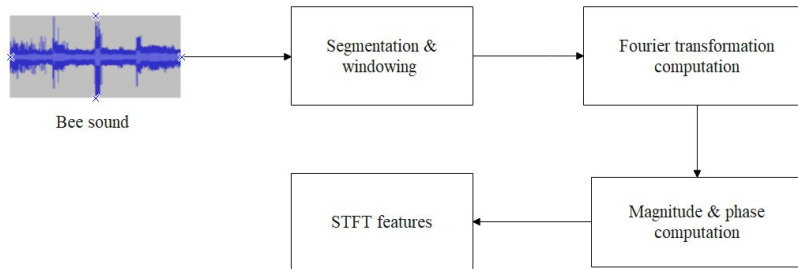
Hình 1. Tổng quan về phương pháp tiếp cận được đề xuất để nhận biết trạng thái của ong trong tổ ong

2.1 Phương pháp trích xuất đặc điểm

a) Biến đổi Fourier thời gian ngắn (STFT)

STFT là một phương pháp cơ bản trong xử lý tín hiệu, được sử dụng trong các lĩnh vực như phân tích âm thanh, nhận dạng giọng nói và kỹ thuật y sinh [6]. Nó cung cấp biểu diễn tần số thời gian, cho phép kiểm tra sự thay đổi phổ

đặc điểm. Không giống như phân tích Fourier chuẩn, phù hợp với tín hiệu dừng, STFT thực hiện biến đổi Fourier cục bộ trên các cửa sổ nhỏ chồng chéo, khiến nó hiệu quả đối với tín hiệu động. Hình 2 mô tả các bước liên quan đến việc tính toán STFT:



Hình 2. Sơ đồ phương pháp STFT

Chia tín hiệu thành các đoạn thời gian ngắn: Tín hiệu được chia thành các đoạn thời gian ngắn, gọi là khung, với độ dài cố định. Độ dài khung xác định độ phân giải của STFT theo thời gian. Độ dài khung ngắn hơn sẽ cung cấp độ phân giải thời gian tốt hơn nhưng độ phân giải tần số thấp hơn, trong khi độ dài khung dài hơn sẽ cung cấp độ phân giải tần số tốt hơn nhưng độ phân giải thời gian thấp hơn.

Áp dụng hàm cửa sổ cho mỗi phân đoạn: Một hàm cửa sổ được áp dụng cho mỗi phân đoạn để giảm hiện tượng nhiễu do tín hiệu bị cắt ngắn đột ngột. Các hàm cửa sổ phổ biến bao gồm cửa sổ Hanning, Hamming và Gaussian. Việc lựa chọn hàm cửa sổ có thể ảnh hưởng đến hình dạng của các đỉnh STFT.

Tính toán biến đổi Fourier của mỗi phân đoạn: Biến đổi Fourier được tính toán cho từng phân đoạn để có được phổ tần số của nó. Biến đổi Fourier phân tích tín hiệu thành các thành phần hình sin cấu thành, tiết lộ nội dung tần số của tín hiệu trong phân đoạn đó.

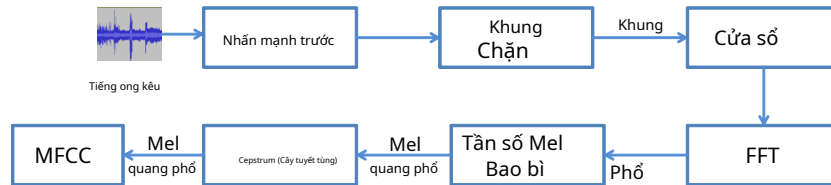
Kết hợp các kết quả: Độ lớn và pha của mỗi phân đoạn trong phép biến đổi Fourier được kết hợp để tạo thành STFT. Độ lớn biểu thị cường độ của từng thành phần tần số, trong khi pha biểu thị độ trễ thời gian của từng thành phần tần số.

Hiển thị STFT: STFT có thể được hình dung như một phổ tần số thời gian, biểu diễn độ lớn theo cường độ màu. Phổ này cho thấy nội dung tần số của tín hiệu thay đổi như thế nào theo thời gian.

b) Hệ số Cepstral tần số Mel (MFCC)

MFCC là một phương pháp mạnh mẽ để trích xuất các tính năng từ tín hiệu âm thanh. Kỹ thuật này bao gồm việc chia băng tần thành các băng tần con trên thang MEL và sau đó áp dụng Biến đổi Cosine rời rạc (DCT) để trích xuất Cepstral

Hệ số. Các bước chính của MFCC, như thể hiện trong Hình 3, bao gồm nhấn mạnh trước, đóng khung, tạo cửa sổ, áp dụng Biến đổi Fourier rời rạc/nhanh (DFT/FFT), biến dạng tần số Mel và tính toán cepstrum (DCT ngược).



Hình 3.Sơ đồ thuật toán MFCC

Nhấn mạnh trước: Giai đoạn đầu của MFCC, trước khi nhấn mạnh, khuếch đại năng lượng ở tần số cao hơn [10]. Điều này liên quan đến việc truyền tín hiệu qua bộ lọc thông cao bậc nhất để giảm tiếng ồn trong quá trình thu âm thanh.

Chặn khung và tạo cửa sổ: MFCC hoạt động trên các khoảng thời gian ngắn, tính của dữ liệu âm thanh bằng cách chia tín hiệu thành các khung chồng chéo. Mỗi khung chứa nhiều mẫu âm thanh với một số chồng chéo giữa các khung liên tiếp. Cửa sổ Hanning hoặc Hamming thường được áp dụng [11] để tăng cường sóng hài, làm mịn các cạnh và giảm hiệu ứng cạnh trong quá trình tính toán DFT/FFT.

Áp dụng Biến đổi Fourier nhanh (FFT): Trong bước này, mỗi khung cửa sổ được chuyển đổi thành phổ biên độ bằng FFT, nhanh chóng tính toán Biến đổi Fourier rời rạc (DFT). Quá trình này chuyển đổi từng khung từ miền thời gian sang miền tần số.

Gói Mel-Frequency: Phổ Mel thu được bằng cách áp dụng bộ lọc Mel vào phổ công suất (được lấy từ FFT của tín hiệu). Các bộ lọc Mel này mô phỏng thính giác của con người, thu năng lượng trong các dải tần số cụ thể có liên quan đến nhận thức âm thanh của chúng ta.

Cepstrum: Phổ công suất theo thang Mel sau đó được chuyển đổi sang miền thời gian bằng cách sử dụng Biến đổi Cosin rời rạc (DCT) để thu được Hệ số Cepstral tần số Mel (MFCC).

Trong nghiên cứu này, chúng tôi sử dụng ba tham số trích xuất tính năng khác nhau để trích xuất các tính năng MFCC từ các mẫu âm thanh: 20 tính năng phù hợp với phương pháp được áp dụng bởi [1] và 40 và 80 tính năng tương ứng với phương pháp trích xuất tính năng được sử dụng bởi [9].

c) Kết hợp các tính năng MFCC và STFT

Nghiên cứu này đề xuất một phương pháp mới để tăng cường biểu diễn đặc điểm cho phân loại trạng thái ong trong tổ ong. Chúng tôi tận dụng thế mạnh bổ sung của hai kỹ thuật trích xuất đặc điểm: Hệ số Cepstral tần số Mel và Biến đổi Fourier thời gian ngắn. MFCC rất giỏi trong việc nắm bắt các đặc điểm nhận thức của âm thanh, mô phỏng thính giác của con người. Điều này làm cho nó đặc biệt hiệu quả

để biểu diễn các tín hiệu có nội dung cao độ nổi bật, như tiếng kêu của ong. Ngoài ra, MFCC cung cấp khả năng giảm chiều, giảm độ phức tạp của tính toán. Mặt khác, STFT cung cấp biểu diễn tần số thời gian chi tiết của tín hiệu. Điều này cho phép nó nắm bắt các biến thể tinh tế về tần số và các sự kiện thoáng qua có thể rất quan trọng để phân biệt trạng thái của ong trong tổ ong. Đặc điểm này làm cho STFT phù hợp để phân tích các tín hiệu không tĩnh như âm thanh của tổ ong, thường biểu hiện các mẫu tần số phức tạp.

Bằng cách kết hợp các tính năng này một cách chiến lược, chúng tôi hướng đến mục tiêu tạo ra một biểu diễn thông tin hơn về dữ liệu âm thanh của tổ ong. Biểu diễn phong phú này được cấu trúc như sau:

$$stft_1, stft_2, \dots, stft_{80}, mfcc_1, mfcc_2, \dots, mfcc_{40}$$

Các tính năng mới kết hợp cả khía cạnh nhận thức và thời gian của âm thanh, có khả năng cải thiện độ chính xác phân loại trong việc xác định các trạng thái khác nhau của ong trong đàn.

2.2 Mô hình học máy

K-Láng giềng gần nhất (KNN) là một thuật toán học có giám sát phổ biến được sử dụng cho cả nhiệm vụ phân loại và hồi quy [5]. Nó gán nhãn cho một mẫu mới dựa trên nhãn của nó K những người hàng xóm gần nhất trong tập huấn luyện. Nhãn lớp thường được xác định bằng phiếu bầu đa số giữa những người hàng xóm này, với những người hàng xóm gần hơn có khả năng có ảnh hưởng lớn hơn. KNN sử dụng một phép đo khoảng cách, thường là khoảng cách Euclidean, để đo khoảng cách giữa các điểm dữ liệu. Phương pháp này có thể dễ dàng thích ứng với dữ liệu mới mà không cần đào tạo lại mô hình. Tuy nhiên, KNN hoạt động kém trong không gian nhiều chiều và nhạy cảm với nhiễu và dữ liệu bị thiếu trong tập huấn luyện.

Máy vectơ hỗ trợ (SVM) là một thuật toán học có giám sát mạnh mẽ được sử dụng cho các tác vụ phân loại và hồi quy [7]. Nó nhằm mục đích tìm siêu phẳng tối ưu trong không gian nhiều chiều phân tách các lớp khác nhau với biên độ lớn nhất. Siêu phẳng này được xác định bởi các vectơ hỗ trợ, là các điểm dữ liệu quan trọng tối đa hóa sự tách lớp. SVM có thể xử lý dữ liệu không thể tách rời tuyến tính bằng cách sử dụng thủ thuật hạt nhân, ánh xạ các tính năng đầu vào vào các không gian nhiều chiều hơn để phân loại phi tuyến tính. Trong nghiên cứu này, chúng tôi chỉ sử dụng hạt nhân hàm cơ sở xuyên tâm (RBF) do hiệu quả đã được chứng minh của nó trong nhiều nghiên cứu và ấn phẩm học thuật khác nhau.

Hồi quy logistic (LR) là một phương pháp được sử dụng rộng rãi chủ yếu cho các nhiệm vụ phân loại. Nó hoạt động tốt nhất khi mối quan hệ giữa các biến độc lập và biến phụ thuộc (thường là nhị phân) có thể được mô hình hóa tuyến tính. LR đặc biệt hữu ích để ước tính xác suất xảy ra một sự kiện. Các hệ số do LR tạo ra cho biết mức độ mà mỗi biến độc lập ảnh hưởng đến khả năng xảy ra của một kết quả cụ thể. Tính đơn giản và hiệu quả của nó khiến nó trở thành lựa chọn phổ biến cho các vấn đề phân loại nhị phân, trong đó việc hiểu được ảnh hưởng của các biến là điều cần thiết.

Rừng ngẫu nhiên (RF) là một thuật toán học máy mạnh mẽ kết hợp nhiều cây quyết định để cải thiện độ chính xác và ngăn ngừa tình trạng quá khớp [2].

Mỗi cây trong rừng được tạo ra bằng cách sử dụng một tập hợp con ngẫu nhiên của dữ liệu đào tạo (khởi động), thúc đẩy sự đa dạng giữa các cây. Các cây phân loại hoặc dự đoán kết quả độc lập bằng cách tìm các điểm phân chia tối ưu, thường sử dụng tạp chất Gini làm số liệu. Đối với các tác vụ hồi quy, các dự đoán của tất cả các cây được tính trung bình, trong khi đối với các tác vụ phân loại, dự đoán cuối cùng được xác định bằng cách bỏ phiếu đa số.

Cây bổ sung (ET) hoặc Extremely Randomized Trees, là một thuật toán học tập tổng hợp tương tự như Random Forest (RF) nhưng có một số điểm khác biệt chính. ET sử dụng toàn bộ mẫu gốc, giúp giảm độ lệch. Ngoài ra, ET đưa vào tính ngẫu nhiên bằng cách chọn các điểm chia ngẫu nhiên thay vì tính toán tối ưu cục bộ bằng các số liệu như tạp chất Gini hoặc entropy. Việc chọn ngẫu nhiên các điểm chia này làm tăng tính đa dạng và giảm tương quan giữa các cây, nâng cao hiệu quả của thuật toán. ET được biết đến với tốc độ và khả năng giảm thiểu tình trạng quá khớp, khiến nó phù hợp với các tập dữ liệu lớn có nhiều tính năng.

Tăng cường XG (XGB) hoặc Extreme Gradient Boosting, là một triển khai hiệu quả cao của cây quyết định tăng cường độ dốc được thiết kế để tăng tốc độ và hiệu suất [4]. Nó sử dụng cây tăng cường song song để làm mịn tổn thất khi đào tạo và áp dụng chính quy hóa, kết hợp các điểm mạnh của nhiều cây cơ sở để có kết quả tối ưu. Thuật toán sửa các lỗi trước đó, học theo cách lặp đi lặp lại để cải thiện hiệu suất. Để giảm tình trạng quá khớp và tăng tốc đào tạo, XGBoost kết hợp các kỹ thuật ngẫu nhiên hóa, chẳng hạn như chọn các mẫu con để xây dựng cây và chọn các tính năng ở nhiều cấp độ khác nhau. Nó cũng sử dụng các phần trăm để kiểm tra một tập hợp con các phân tách ứng viên, tăng tốc đáng kể quá trình trong khi vẫn duy trì độ chính xác và làm cho nó hiệu quả đối với nhiều vấn đề khoa học dữ liệu khác nhau.

3 Thí nghiệm

3.1 Mô tả dữ liệu

Để đánh giá phương pháp tiếp cận được đề xuất, chúng tôi sử dụng tập dữ liệu trong nghiên cứu trước đây [12]. Tập dữ liệu này, bao gồm 13.792 mẫu, cung cấp những hiểu biết có giá trị về hoạt động của tổ ong thông qua các bản ghi âm, được phân loại thành ba lớp:

Tiếng ong: Âm thanh phát ra từ hoạt động bình thường của ong trong tổ.

NoBee: biểu thị âm thanh xung quanh, chỉ ra những thời điểm có tiếng ồn bên ngoài. Khi ong chúa già, bị bệnh hoặc chết, ong thợ có thể bắt đầu nuôi ong chúa mới trong vòng 12-24 giờ, dẫn đến hoạt động của đàn ong và mức độ âm thanh giảm. Trong những trường hợp như vậy, chỉ phát hiện ra tiếng ồn nền hoặc tiếng ồn xung quanh, cho thấy đàn ong có khả năng sụp đổ hoặc đang nuôi ong chúa mới.

NoQueen: Âm thanh liên quan đến những đàn ong không có ong chúa, có khả năng báo hiệu sự sụp đổ của đàn ong (do ong thợ không hoạt động), ong chúa đang nuôi (dẫn đến giảm hoạt động) hoặc các hiện tượng bất thường khác.

Chúng tôi chia tập dữ liệu này thành các tập huấn luyện và thử nghiệm với tỷ lệ 80:20 (bảng 1) tiến hành và đánh giá các thí nghiệm.

Bảng 1. Phân phối mẫu cho các danh mục Bee, NoBee và NoQueen

Lớp học	# Mẫu tàu hỏa (80%)	Kiểm tra (20%)
Con ong	5473	4378
Không cóBee	3458	2766
Không cóNữ hoàng	4861	3889

3.2 Kết quả và thảo luận

Hiệu suất của các phương pháp ML sử dụng các tính năng riêng lẻ

a) Tính năng MFCC

Bảng 2 trình bày hiệu suất của các mô hình học máy khác nhau trong việc xác định trạng thái ong (Bee, NoBee và NoQueen) bằng MFCC với các tính năng trích xuất khác nhau: 20 tính năng, 40 tính năng và 80 tính năng. Các mô hình được xem xét bao gồm KNN, SVM, LR, RF, ET và XGB. Bảng này chứng minh tác động của số lượng tính năng MFCC đến hiệu suất của các mô hình học máy. Nhìn chung, một mô hình được đào tạo với nhiều tính năng có liên quan hơn có xu hướng đạt được độ chính xác cao hơn, nhưng việc thêm quá nhiều tính năng hoặc tính năng không liên quan có thể dẫn đến quá khớp và giảm hiệu suất. Điều này cho thấy rằng việc trích xuất thông tin phổ chi tiết hơn từ các bản ghi âm có thể nâng cao khả năng phân biệt giữa các trạng thái ong khác nhau. RF cho thấy sự cải thiện đáng kể nhất về độ chính xác với số lượng tính năng ngày càng tăng, đạt đỉnh là 83,74% với 40 tính năng. Điều này cho thấy RF rất nhạy cảm với việc lựa chọn tính năng và có thể được hưởng lợi từ việc tối ưu hóa giữa 20 và 40 tính năng. ET có xu hướng tương tự như RF, với độ chính xác cao nhất là 84,05% ở 40 đặc điểm và giảm nhẹ ở 80 đặc điểm.

Trong [12], Rustam et al. đã sử dụng kết hợp các kỹ thuật lựa chọn tính năng (PCA, Chi-squared và SVD) được áp dụng cho một tập hợp lớn gồm 1740 tính năng MFCC, đạt được độ chính xác cao nhất là 83% để phát hiện ong bằng cách sử dụng bộ phân loại RF và KNN. Trong nghiên cứu của chúng tôi, chúng tôi sử dụng một phương pháp đơn giản hơn là trích xuất trực tiếp một tập hợp nhỏ hơn các tính năng MFCC (40 tính năng) có thể đạt được độ chính xác tương đương hoặc thậm chí tốt hơn với 84,05%.

Bảng 2. Hiệu suất của các phương pháp ML để xác định trạng thái của ong bằng cách sử dụng ba bộ tính năng MFCC khác nhau (%)

Người mẫu	20 tính năng	40 tính năng	80 tính năng
KNN	79,46	81,49	82,41
SVM	66,05	68,63	69,53
LỚP 1	69,45	75,33	75,64
Tần số vô tuyến	81,73	83,74	83,54
và	82,21	84,05	83,18
XGB	80,96	82,58	83,33

b) Các tính năng của STFT

Bảng 3 trình bày độ chính xác thử nghiệm đạt được bởi các mô hình học máy khác nhau trong việc xác định trạng thái của ong bằng cách sử dụng các tính năng Biến đổi Fourier thời gian ngắn (STFT) được trích xuất bằng hai tập tính năng khác nhau: 80 tính năng và 257 tính năng. Nhìn chung, tất cả các mô hình ngoại trừ SVM đều cho thấy độ chính xác tăng nhẹ với nhiều tính năng hơn (từ 80 đến 257 tính năng). Điều này cho thấy rằng việc nắm bắt phạm vi thông tin tần số rộng hơn có thể nâng cao khả năng phân biệt giữa các trạng thái khác nhau của ong. KNN cho thấy sự cải thiện đáng kể nhất (1,83%), tiếp theo là RF (0,56%) và ET (0,31%). LR cũng cho thấy sự cải thiện đáng chú ý (2,88%). Phù hợp với phân tích trước đó khi sử dụng các tính năng MFCC, XGBoost một lần nữa đạt được độ chính xác cao nhất trên cả hai số lượng tính năng (85,40% với 80 tính năng và 85,72% với 257 tính năng). Điều này củng cố thêm tính mạnh mẽ và hiệu quả của nó trong việc xác định trạng thái của ong.

Khi so sánh hiệu suất của các mô hình ML tổng hợp (RF, ET và XGB) sử dụng các tính năng STFT và các tính năng MFCC, rõ ràng là khi sử dụng các tính năng STFT, ngay cả với 80 tính năng, chúng vẫn vượt trội hơn hiệu suất tốt nhất đạt được khi sử dụng các tính năng MFCC (84,05%). Điều này chỉ ra rằng các tính năng STFT hiệu quả hơn trong việc nắm bắt thông tin có liên quan để nhận dạng trạng thái của ong so với các tính năng MFCC trong trường hợp cụ thể này.

Bảng 3. Hiệu suất của các phương pháp ML để xác định trạng thái của ong bằng cách sử dụng hai bộ tính năng STFT (%)

Người mẫu	80 tính năng	257 tính năng
KNN	81,66	83,49
SVM	75,01	75,91
LỚP 1	68,41	71,29
Tần số vô tuyến	84,75	85,31
và	84,10	84,41
XGB	85,40	85,72

Hiệu suất của các phương pháp ML sử dụng các tính năng kết hợp

Các phân tích trước đây đã chứng minh tính hiệu quả của cả đặc điểm MFCC và STFT trong việc nhận dạng trạng thái của ong. Đặc điểm MFCC nổi trội trong việc nắm bắt các đặc điểm nhận thức của âm thanh, trong khi đặc điểm STFT cung cấp biểu diễn chi tiết về phổ tần số. Để khai thác thế mạnh của cả đặc điểm MFCC và STFT, chúng tôi đề xuất một phương pháp hợp nhất đặc điểm kết hợp các đặc điểm được trích xuất từ cả hai phương pháp. Chúng tôi chọn kết hợp MFCC với 80 đặc điểm STFT sau khi đánh giá sự đánh đổi giữa hiệu suất và hiệu quả tính toán. Trong khi sử dụng 257 đặc điểm STFT mang lại độ chính xác cao hơn một chút, thì sự gia tăng đáng kể các đặc điểm cũng dẫn đến hiệu suất tính toán cao hơn

chi phí. Để đạt được sự cân bằng giữa hiệu suất và hiệu quả, chúng tôi đã chọn 80 tính năng STFT.

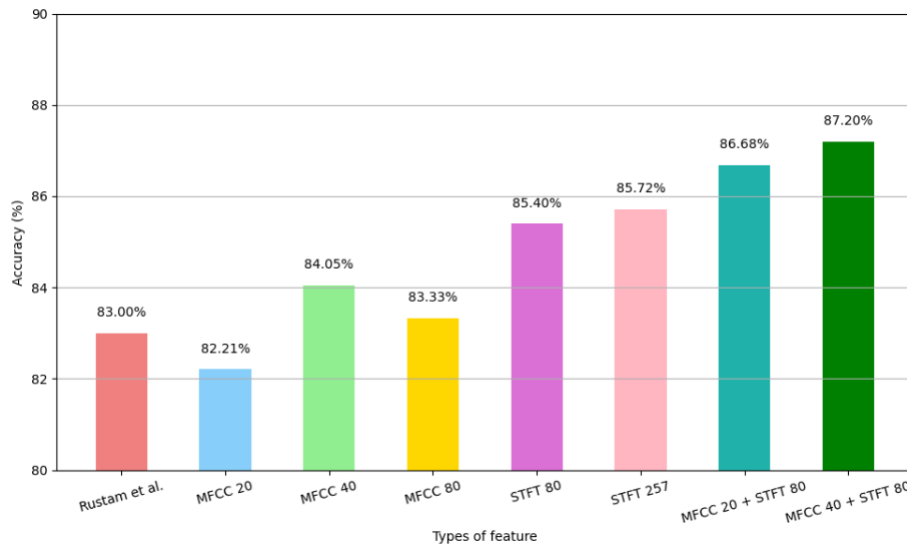
Bảng 4 trình bày độ chính xác đạt được bởi các mô hình học máy khác nhau trong việc xác định trạng thái của ong bằng cách sử dụng kết hợp 80 tính năng STFT và hai bộ tính năng MFCC khác nhau: 20 tính năng và 40 tính năng. Tất cả các mô hình đều cho thấy sự cải thiện về độ chính xác khi sử dụng 40 tính năng MFCC so với 20 tính năng MFCC kết hợp với 80 tính năng STFT. Điều này cho thấy rằng việc kết hợp nhiều nội dung tần số nhận thức hơn của âm thanh được thu thập bởi các tính năng MFCC bổ sung có thể nâng cao hiệu suất của hầu hết các mô hình. RF và ET cho thấy sự cải thiện đáng kể nhất về độ chính xác với 40 tính năng MFCC, với mức tăng lần lượt là 1,18% và 0,53%. XGBoost cho thấy sự cải thiện tối thiểu về độ chính xác (0,27%) với 40 tính năng MFCC so với 20 tính năng. Điều này cho thấy rằng XGBoost có thể đã sử dụng hiệu quả thông tin do cả tính năng STFT và 20 MFCC cung cấp. KNN và SVM thể hiện mức tăng độ chính xác vừa phải với 40 đặc điểm MFCC, cho thấy chúng có thể hưởng lợi từ thông tin bổ sung nhưng ở mức độ ít hơn so với các mô hình như RF và ET. Hồi quy logistic (LR) cho thấy sự cải thiện đáng kể về độ chính xác (3,84%) với 40 đặc điểm MFCC, cho thấy nó sử dụng thông tin thời gian bổ sung một cách hiệu quả.

Hình 4 cung cấp một minh họa rõ ràng và súc tích về cách hợp nhất tính năng tận dụng thế mạnh của các kỹ thuật trích xuất tính năng khác nhau như MFCC và STFT. Bằng cách kết hợp thông tin nhận thức được thu thập bởi các tính năng MFCC với thông tin tần số được thu thập bởi các tính năng STFT, mô hình có được sự hiểu biết sâu sắc hơn về dữ liệu âm thanh, dẫn đến phân loại trạng thái ong chính xác hơn. Sự kết hợp của MFCC (20 hoặc 40 tính năng) với các tính năng STFT (80) đạt đến điểm cao nhất trên biểu đồ. Điều này nhấn mạnh về mặt trực quan rằng việc kết hợp các tính năng sẽ thu được biểu diễn toàn diện hơn về âm thanh của tổ ong, dẫn đến cải thiện hiệu suất mô hình trong việc xác định trạng thái ong.

Trong nghiên cứu trước đây [12], độ chính xác cao nhất đạt được là khoảng 83%. Bằng cách sử dụng hợp nhất tính năng, phương pháp tiếp cận mới đã đẩy độ chính xác lên 87,20%, thể hiện bước tiến đáng kể. Sự cải tiến này cho thấy sự kết hợp của các tính năng MFCC và STFT nắm bắt được nhiều thông tin có liên quan hơn về âm thanh của tổ ong, cho phép mô hình phân biệt tốt hơn giữa các trạng thái khác nhau của ong.

Bảng 4. Hiệu suất của các phương pháp ML để xác định trạng thái của ong bằng 80 STFT với các bộ tính năng MFCC khác nhau (%)

Người mẫu	80 tính năng STFT & 20 MFCC	Tính năng 80 STFT & 40 MFCC
KNN	80,14	82,75
SVM	65,97	68,85
LỚP 1	75,28	79,12
Tần số vô tuyến	86,01	87,19
và	86,44	86,97
XGB	86,68	86,95



Hình 4. So sánh độ chính xác tốt nhất bằng cách sử dụng các loại tính năng khác nhau

4 Kết luận

Bài báo này đề xuất một phương pháp mới để cải thiện khả năng nhận dạng trạng thái của ong bằng cách tạo ra dữ liệu đại diện nhiều thông tin hơn thông qua việc hợp nhất các đặc điểm MFCC và STFT. Chúng tôi trích xuất và phân tích cả hai đặc điểm MFCC và STFT riêng lẻ, sau đó kết hợp chúng để tạo ra dữ liệu đại diện mới bao gồm một tập hợp thông tin phong phú hơn về âm thanh của tổ ong. Kết quả thử nghiệm cho thấy mô hình RF đạt được độ chính xác đáng chú ý là 87,2%, vượt qua kết quả tốt nhất trước đó với biên độ đáng kể là 4,2% [12]. Những phát hiện này ủng hộ mạnh mẽ hiệu quả của phương pháp được đề xuất trong việc tạo ra dữ liệu đại diện giúp nâng cao đáng kể hiệu suất nhận dạng trạng thái của ong. Mặc dù kết quả này thể hiện sự cải thiện đáng kể, nhưng vẫn chưa đạt được mục tiêu cuối cùng là đạt được khả năng nhận dạng trạng thái của ong có độ chính xác và độ tin cậy cao. Nhận thấy nhu cầu tiến bộ hơn nữa, chúng tôi đề xuất nghiên cứu ứng dụng các kỹ thuật học sâu tiên tiến, chẳng hạn như học chuyển giao và chế độ RegNet, để giải quyết thách thức này.

Tài liệu tham khảo

1. Barbisan, L., Riente, F.: Khung học máy để phát hiện âm thanh sự hiện diện của ong chúa. *Acta Acustica* (2023)
2. Breiman, L.: Rừng ngẫu nhiên. *Học máy* 45, 5–32 (2001)

3. Cecchi, S., Spinsante, S., Terenzi, A., Orcioni, S.: Hệ thống đo lường dựa trên cảm biến thông minh để giám sát tổ ong tiên tiến. *Cảm biến*20(9), 2726 (tháng 5 năm 2020)
4. Chen, T., Guestrin, C.: Xgboost: Hệ thống tăng cường cây có thể mở rộng. Trong: *Biên bản báo cáo hội nghị quốc tế lần thứ 22 của acm sigkdd về khám phá kiến thức và khai thác dữ liệu*. trang 785–794 (2016)
5. Cover, T., Hart, P.: Phân loại mẫu hàng xóm gần nhất. *IEEE Transactions on Information Theory* (1967)
6. Durak Ata, L., Arikan, O.: Biến đổi Fourier thời gian ngắn: Hai tính chất cơ bản và một triển khai tối ưu. *Xử lý tín hiệu, Giao dịch IEEE về* 51,1231 – 1242 (06 2003). <https://doi.org/10.1109/TSP.2003.810293>
7. Evgeniou, T., Pontil, M.: Máy vectơ hỗ trợ: Lý thuyết và ứng dụng. *Học máy và ứng dụng của nó, Bài giảng nâng cao* (2001)
8. Liao, Y., McGuirk, A., Biggs, B., Chaudhuri, A., Langlois, A., Deters, V.: Giám sát tổ ong không xâm lấn thông qua dữ liệu âm thanh sử dụng SAS@Xử lý luồng sự kiện và SAS@Viya®. *Diễn đàn toàn cầu SAS* trang 24 (2020)
9. Phan, TTH, Nguyễn-Doan, Đ., Nguyễn-Hữu, D., Nguyễn-Văn, H., Phạm-Hồng, T.: Nghiên cứu về các đặc điểm hệ số cepstral tần số mel mới và kỹ thuật điều chỉnh siêu tham số để nhận dạng âm thanh của ong. *Soft Computing* (2022). <https://doi.org/10.1007/s00500-022-07596-6>
10. Picone, J.: Kỹ thuật mô hình hóa tín hiệu trong nhận dạng giọng nói. *Biên bản báo cáo của IEEE* 81(9), 1215–1247 (tháng 9 năm 1993)
11. Rabiner, L.: Hướng dẫn về các mô hình Markov ẩn và các ứng dụng được chọn trong nhận dạng giọng nói. *Biên bản của IEEE*77(2), 257–286 (tháng 2 năm 1989). <https://doi.org/10.1109/5.18626>, *Tên hội nghị: Biên bản báo cáo của IEEE*
12. Rustam, F., Zahid Sharif, M., Aljedaani, W., Lee, E., Ashraf, I.: Phát hiện ong trong tổ ong bằng cách sử dụng các đặc điểm chọn lọc từ dữ liệu âm thanh. *Công cụ và ứng dụng đa phương tiện*82(5), 7095–7112 (2023). <https://doi.org/10.1007/s11042-023-15192-5>
13. Ruvinga, S., Hunter, GJ, Duran, O., Nebel, JC: Sử dụng mạng LSTM để xác định “tình trạng không có ong chúa” trong tổ ong mật từ tín hiệu âm thanh. Trong: *Hội nghị quốc tế lần thứ 17 về môi trường thông minh (IE) năm 2021*. tr. 1–4. *IEEE* (tháng 6 năm 2021). <https://doi.org/10.1109/IE51775.2021.9486575>
14. Trường, TH, Nguyễn, HD, Mai, TQA, Nguyễn, HL, Đặng, TNM, Phan, TTH: Một phương pháp tiếp cận dựa trên học sâu để nhận dạng âm thanh của ong. *Tin học sinh thái*78,102274 (2023). <https://doi.org/10.1016/j.ecoinf.2023.102274>