HỌC VIỆN NÔNG NGHIỆP VIỆT NAM

HOÀNG THỊ HƯƠNG

NGHIÊN CỨU CÁC KỸ THUẬT HỌC MÁY ÁP DỤNG CHO BÀI TOÁN NHẬN DẠNG ĐỐI TƯỢNG DỰA TRÊN ÂM THANH

Ngành:	Công nghệ thông tin
Mã số:	8 48 02 01
Người hướng dẫn:	TS. Phan Thị Thu Hồng
	TS. Phạm Quang Dũng

NHÀ XUẤT BẢN HỌC VIỆN NÔNG NGHIỆP – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi, các kết quả nghiên cứu được trình bày trong đề án là trung thực, khách quan và chưa từng dùng để bảo vệ lấy bất kỳ học vị nào.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện đề án đã được cám ơn, các thông tin trích dẫn trong đề án này đều được chỉ rõ nguồn gốc.

Hà Nội, ngày tháng năm 2024

Tác giả đề án

Hoàng Thị Hương

LỜI CẢM ƠN

Trong suốt thời gian học tập, nghiên cứu và hoàn thành đề án, tôi đã nhận được sự hướng dẫn, chỉ bảo tận tình của các thầy cô giáo, sự giúp đỡ, động viên của bạn bè, đồng nghiệp và gia đình.

Nhân dịp hoàn thành đề án, cho phép tôi được bày tỏ lòng kính trọng và biết ơn sâu sắc tới TS. Phan Thị Thu Hồng và TS. Phạm Quang Dũng đã tận tình hướng dẫn, dành nhiều công sức, thời gian và tạo điều kiện cho tôi trong suốt quá trình học tập và thực hiện đề tài.

Tôi xin bày tỏ lòng biết ơn chân thành tới Ban Giám đốc, Ban Quản lý đào tạo, Khoa Công nghệ thông tin - Học viện Nông nghiệp Việt Nam, Bộ môn Mạng và hệ thống thông tin đã tận tình giúp đỡ tôi trong quá trình học tập, thực hiện đề tài và hoàn thành đề án.

Tôi xin chân thành cảm ơn tập thể lãnh đạo, cán bộ viên chức Trung tâm nghiên cứu ong và nuôi ong nhiệt đới, Học viện Nông nghiệp Việt Nam đã giúp đỡ và tạo điều kiện cho tôi trong suốt quá trình thực hiện đề tài.

Xin chân thành cảm ơn gia đình, người thân, bạn bè, đồng nghiệp đã tạo mọi điều kiện thuận lợi và giúp đỡ tôi về mọi mặt, động viên khuyến khích tôi hoàn thành đề án.

Tôi xin cảm ơn, đề tài "Nghiên cứu ứng dụng công nghệ của công nghiệp 4.0 vào quản lý sản xuất sản phẩm mật ong phục vụ xuất khẩu và tiêu dùng trong nước", số hiệu đề tài KC-4.0-20/19-25 thuộc chương trình Khoa học và Công nghệ trọng điểm cấp nhà nước "Hỗ trợ nghiên cứu, phát triển và ứng dụng công nghệ của công nghệ 4.0", mã số KC 4.0/19-25.

Mặc dù có nhiều cố gắng nhưng do thời gian có hạn và kiến thức còn hạn chế nên không tránh khỏi những sai sót. Vì vậy, tôi rất mong nhận được sự góp ý, hướng dẫn của các thầy cô giáo để đề án của tôi được hoàn thiện hơn.

Tôi xin chân thành cảm ơn!

Hà Nội, ngày 10 tháng 01 năm 2024

Tác giả đề án

Hoàng Thị Hương

MỤC LỤC

Lời cam đoan i
Lời cảm ơn ii
Mục lục iii
Danh mục viết tắt v
Danh mục bảng vi
Danh mục hình vii
Phần 1. Mở đầu 1
1.1. Tính cấp thiết của đề tài 1
1.1.1. Ý nghĩa khoa học 1
1.1.2. Ý nghĩa thực tiễn 1
1.2. Cơ sở khoa học 4
1.3. Mục tiêu nghiên cứu 5
1.3.1. Mục tiêu tổng quát 5
1.3.2. Mục tiêu cụ thể 5
1.4. Đối tượng và phạm vi nghiên cứu 5
1.4.1. Đối tượng nghiên cứu 5
1.4.2. Phạm vi nghiên cứu 5
Phần 2. Tổng quan tài liệu 6
2.1. Tình hình nghiên cứu trong nước 6
2.2. Tình hình nghiên cứu ngoài nước 6
Phần 3. Phương pháp nghiên cứu và xử lý số liệu 10
3.1. Địa điểm nghiên cứu 10
3.2. Thời gian nghiên cứu 10
3.3. Đối tượng, phạm vi nghiên cứu 10
3.4. Nội dung nghiên cứu 10
3.5. Phương pháp nghiên cứu 10
Phần 4. Kết quả và thảo luận 11
4.1. Xử lí dữ liệu âm thanh 13
4.1.1. Tổng quan về xử lý dữ liệu âm thanh 13
4.1.2. Âm thanh và biểu diễn âm thanh đối với học máy 13

4.1.3. Hệ số quang phổ tần số Mel (MFCC) 15
4.2. Các phương pháp học máy 19
4.2.1. Học máy 20
4.2.2. Bài toán phân loại 21
4.2.3. Một số kĩ thuật học máy cho bài toán phân loại 21
4.3. Ứng dụng các kỹ thuật học máy cho bài toán nhận dạng đối tượng dựa trên âm thanh 32
4.3.1. Mô tả dữ liệu và tiền xử lí dữ liệu 32
4.3.2. Trích xuất đặc trưng MFCC 34
4.3.3. Cài đặt thử nghiệm 35
4.4. Thảo luận 42
Phần 5. Kết luận và kiến nghị 45
5.1. Kết luận 45
5.2. Kiến nghị 46
Tài liệu tham khảo 47
Phụ lục 50

DANH MỤC VIẾT TẮT

Từ viết tắt	Nghĩa tiếng Việt
AI	Trí tuệ nhân tạo
CNN	Mạng thần kinh tích chập
DCT	Biến đổi Cosin rời rạc
DFT	Biến đổi Fourier rời rạc
DT	Cây quyết định
FFT	Biến đổi Fourier nhanh
FT	Biến đổi Fourier
IoT	Kết nối vạn vật
k_NN	k- láng giềng gần nhất
LR	Hồi quy logistis
MFCC	Hệ số tần số quang phổ Mel
ML	Học máy
RF	Rừng ngẫu nghiên
SVM	Máy véc tơ hỗ trợ

DANH MỤC BẢNG

Bảng 4.1. Ưu điểm và nhược điểm chính của các mô hình học máy. 31
Bảng 4.2. Ma trận nhầm lẫn đối với phân lớp nhị phân 36
Bảng 4.3. Kết quả phân lớp của các mô hình với 39 đặc trưng MFCC_BT 37
Bảng 4.4. Độ chính xác của các mô hình khi chạy với kịch bản 13 đặc trưng MFCC 39
Bảng 4.5 Thời gian thực hiện của các mô hình với 13 đặc trưng MFCC (tính theo giây) 39
Bảng 4.6. Độ chính xác của các mô hình khi chạy với kịch bản 26 đặc trưng MFCC 40
Bảng 4.7. Thời gian thực hiện của các mô hình với 26 đặc trưng MFCC (tính theo giây) 41
Bảng 4.8. Độ chính xác của các mô hình khi chạy với kịch bản 39 đặc trưng MFCC 41
Bảng 4.9. Thời gian thực hiện của các mô hình với 39 đặc trưng MFCC (tính theo giây) 42

DANH MỤC HÌNH

Hình 3.1. Dữ liệu âm thanh ong thu được theo thời gian.....11
Hình 4.1. Biểu diễn trong miền thời gian của giọng nói, tiếng đàn và âm thanh xe cứu hỏa.....14
Hình 4.2. Biến đổi Fourier đưa miền thời gian (t) về miền tần số (f).....15
Hình 4.3. Khung làm việc của MFCC.....16
Hình 4.4. Mối liên hệ giữa AI, học máy và học sâu (deep learning).....19
Hình 4.5. Cấu trúc của phân lớp rừng ngẫu nhiên.....24
Hình 4.6. Minh họa phân lớp tuyến tính với SVM.....26
Hình 4.7. Minh họa hồi quy logistic và hồi quy tuyến tính.....28
Hình 4.8. Mô phỏng phân lớp k_NN.....30
Hình 4.9. Số lượng file âm thanh ong được sử dụng để phân lớp.....32
Hình 4.10. Ảnh phổ của một mẫu âm thanh ong ở trạng thái bình thường.....33
Hình 4.11. Ảnh phổ của một mẫu âm thanh ong chia đàn.....33
Hình 4.12. Sơ đồ thực hiện nhận dạng đối tượng dựa trên âm thanh.....34
Hình 4.13 Phân bố dữ liệu của một mẫu 39 MFCC.....35
Hình 4.14. Độ chính xác của các mô hình với 13 đặc trưng MFCCs.....39
Hình 4.15. Độ chính xác của các mô hình với 26 đặc trưng MFCCs.....40
Hình 4.16. Độ chính xác của các mô hình với 39 đặc trưng MFCCs.....42

TRÍCH YẾU LUẬN VĂN

Tên tác giả: Hoàng Thị Hương

Tên Luận văn: ***Nghiên cứu các kỹ thuật học máy áp dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh.***

Ngành: Công nghệ Thông tin **Mã số**: 8 48 02 01

Tên cơ sở đào tạo: Học viện Nông nghiệp Việt Nam

Mục đích nghiên cứu

Tìm hiểu một số thuật toán trích chọn đặc trưng âm thanh, các thuật toán học máy có giám sát và ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong.

Luận văn đi sâu vào các vấn đề chính như:

- Tìm hiểu về dữ liệu âm thanh, đặc trưng âm thanh
- Tìm hiểu thuật toán trích chọn đặc trưng âm thanh: MFCCS (Mel-frequency cepstral coefficients)
- Nghiên cứu một số thuật toán học máy có giám sát như: cây quyết định, máy véc tơ hỗ trợ (SVM), rừng ngẫu nhiên (RF), mô hình hồi quy logistic, k-láng giềng gần nhất (k-NN).
- Ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong: Tiền xử lý; Trích chọn đặc trưng; Phân lớp

Phương pháp nghiên cứu

Luận văn sử dụng các phương pháp nghiên cứu như:

- Phương pháp lý thuyết: Tham khảo các tài liệu đã được công bố về xử lý tín hiệu âm thanh, các thuật toán học máy cơ bản, các phương pháp lựa chọn đặc trưng.
- Phương pháp tham khảo ý kiến: Tham khảo ý kiến của giáo viên hướng dẫn.
- Phương pháp thực nghiệm: Cài đặt kiểm thử một số thuật toán học máy cho bài toán nhận dạng đối tượng sử dụng âm thanh
- Phương pháp đánh giá và đối sánh: Nhận xét, đánh giá, so sánh các kết quả thực nghiệm.

Kết quả chính và kết luận

Sau khi thực hiện sau luận văn này, tôi đã có cái nhìn tổng quan về dữ liệu âm thanh, đặc trưng âm thanh, tìm hiểu về bài toán phân lớp. Đồng thời luận văn cũng nghiên cứu các thuật toán học máy áp dụng cho bài toán phân lớp điển hình như cây

quyết định, rừng ngẫu nhiên, máy học vector hỗ trợ hồi quy logistic, và k- láng giềng gần nhất.

Bên cạnh đó, luận văn cũng tìm hiểu, nghiên cứu giải thuật lựa chọn đặc trưng MFCC. Sau đó tôi áp dụng vào bài toán nhận dạng đối tượng dựa trên âm thanh. Cụ thể đối tượng trong nghiên cứu này là dữ liệu âm thanh về chia đàn tự nhiên ở ong và tình trạng thiếu chúa của ong. Bộ dữ liệu âm thanh này được thu tại Trung tâm nghiên cứu ong và nuôi ong nhiệt đới, Học viện Nông nghiệp Việt Nam.

Ngoài ra, nghiên cứu này cũng xét các kịch bản là chỉ lấy 13 đặc trưng MFCC đầu tiên của các file âm thanh; kịch bản nữa là lấy 13 đặc trưng đầu tiên và 13 đạo hàm cấp một của chúng là 26 đặc trưng MFCC của các tệp âm thanh và kịch bản thứ 3 là lấy 39 đặc trưng âm thanh. Các kịch bản được thực hiện bốn trường hợp lấy đại diện cho mỗi đặc trưng MFCC là: lấy phần tử đầu tiên của M (đây chính là trường hợp mặc định khi xác định MFCC trong thư viện librosa cho tập các tệp âm thanh), lấy phần tử nhỏ nhất (min) trên véc tơ M, lấy phần tử lớn nhất (max) của M, lấy giá trị trung bình (mean) của véc tơ M.

Kết quả chỉ ra trường hợp lấy giá trị trung bình làm phần tử đại diện cho mỗi đặc trưng MFCC trên các file âm thanh cho kết quả tốt nhất. Độ chính xác đạt trên 99.7%. Chúng tôi đề nghị mô hình hồi quy logistic với 26 đặc trưng MFCCs và lấy đại diện là phần tử trung bình để phân lớp cho dữ liệu này. Kết quả này cũng là một sự gợi ý cho việc lấy đại diện cho mỗi đặc trưng MFCC cho trích chọn đặc trưng trong việc phân loại âm thanh, cụ thể ở đây là âm thanh ong.

Bên cạnh đó, luận văn cũng còn một số hạn chế như sau:

+ Chưa thử nghiệm được với các đặc trưng cụ thể của các file âm thanh như: trọng tâm của phổ, độ rộng của phổ, năng lượng tín hiệu trung bình của file âm thanh,...

+ Chưa chạy được nhiều mô hình phân lớp khác như các mô hình học sâu: CNN, RNN, ANN.

+ Chưa thực hiện được chuyển âm thanh sang phổ hình ảnh để chạy các mô hình nơ-ron. Trong khi chạy các file ảnh mới là thế mạnh của các mạng nơ-ron.

Trong tương lai, chúng tôi sẽ tìm hiểu thêm các phương pháp trích chọn đặc trưng khác như mô hình phân lớp khác như các mô hình học sâu: CNN, RNN, ANN và thực hiện thử nghiệm với các đặc trưng cụ thể của các file âm thanh như: trọng tâm của phổ, độ rộng của phổ, năng lượng tín hiệu trung bình của file âm thanh,...Tìm hiểu thêm việc thực hiện được chuyển âm thanh sang phổ hình ảnh để chạy các mô hình nơ-ron.

THESIS ABSTRACT

Master candidate: Hoang Thi Huong

Thesis title: *Research machine learning techniques applied to sound-based object recognition problems.*

Major: **Information Technology** **Code**: 8 48 02 01

Education organization: Vietnam National University of Agriculture

Research Objectives

Learn some sound feature extraction algorithms, supervised machine learning algorithms and apply them to the problem of detecting natural swarming in bees.

The thesis delves into the followings:

- Learn about audio data and audio characteristics
- Learn the audio feature extraction algorithm: MFCCS (Mel-frequency cepstral coefficients)
- Research some supervised machine learning algorithms such as: decision trees, support vector machines (SVM), random forests (RF), logistic regression models, k-nearest neighbors (k-NN).
- Application to the problem of detecting natural swarming in bees: Preprocessing; Feature extraction; Classification

Materials and Methods

The thesis uses research methods such as

- Theoretical methods: Refer to published documents on audio signal processing, basic machine learning algorithms, and feature selection methods.
- Consultation method: Consult with instructors.
- Experimental method: Setting up and testing some machine learning algorithms for the problem of object recognition using sound
- Evaluation and comparison method: Comment, evaluate, compare experimental results.

Main results and conclusions

After completing this thesis, I have learned an overview of audio data, audio characteristics, and learned about the classification problem. At the same time, the thesis also researches machine learning algorithms applied to typical classification problems such as decision trees, random forests, logistic regression support vector machines, and k-nearest neighbors.

- Besides, the thesis also explores and researches the MFCC feature selection algorithm. Then I applied it to the problem of object recognition based

on sound. Specifically, the object in this study is acoustic data on natural swarming in bees and the lack of a queen in bees. This sound data set was collected at the Center for Bee Research and Tropical Beekeeping, Vietnam National University of Agriculture.

- In addition, this study also considers the scenario of only taking the first 13 MFCC features of audio files; Another scenario is to take the first 13 features and their 13 first derivatives as 26 MFCC features of the audio files and the third scenario is to get 39 audio features. The scenarios that implement the four cases that represent each MFCC feature are: get the first element of M (this is the default case when defining MFCC in the librosa library for a set of audio files), Get the smallest element (min) on vector M, get the largest element (max) of M, get the average value (mean) of vector M.

- The results show that taking the average value as the representative element for each MFCC feature on audio files gives the best results. Accuracy reaches over 99.7%. We propose a logistic regression model with 26 MFCCs features and take the average element as representative to classify this data. This result is also a suggestion for taking a representative for each MFCC feature for feature extraction in sound classification, specifically here the bee sound.

Besides, the thesis also has some limitations as follows:

+ Not tested with specific characteristics of audio files such as: center of spectrum, width of spectrum, average signal energy of audio files...

+ Can’t run many other classification models such as deep learning models: CNN, RNN, ANN.

+ It has not been possible to convert audio to the image spectrum to run neural models. While running image files is the strength of neural networks.

In the future, I will study more about other feature extraction methods such as other classification models such as deep learning models: CNN, RNN, ANN and perform experiments with specific features of the files. sound such as: the center of the spectrum, the width of the spectrum, the average signal energy of the audio files...Study more about implementing audio-to-visual spectroscopy to run neural models.

PHẦN 1. MỞ ĐẦU

1.1. TÍNH CẤP THIẾT CỦA ĐỀ TÀI

1.1.1. Ý nghĩa khoa học

Trong những năm gần đây, các phương pháp máy học (ML) đã được áp dụng rộng rãi vào nhiều lĩnh vực khác nhau và tạo nên những tiến bộ đáng kể trong các hệ thống dự báo, giám sát… Các hệ thống này cung cấp các giải pháp hiệu quả về chi phí với hiệu suất cao. Trong các ứng dụng giám sát dựa vào dữ liệu âm thanh các mô hình ML được xây dựng dựa trên dữ liệu (data driven) do đó có khả năng phản ánh chính xác hơn các quan hệ giữa các biến dự báo (attributes/features) và các sự kiện (target variable). Điều này cho phép hệ thống đưa ra được những cảnh báo kịp thời đến người sử dụng.

Ý nghĩa khoa học của đề án:

- Tìm hiểu được phương pháp trích xuất đặc trưng âm thanh phổ biến là hệ số tần số quang phổ Mel (MFCC), ứng dụng cho dữ liệu nhận dạng âm thanh.
- Xây dựng các kịch bản lấy đặc trưng âm thanh trên MFCCs.
- Tìm hiểu và ứng dụng các mô hình học máy có giám sát và ứng dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh.

1.1.2. Ý nghĩa thực tiễn

Trong ngành nông nghiệp, ong mật (tên khoa học Apis mellifera) được coi là một trong những loài côn trùng quan trọng nhất. Chúng ta có thể dễ dàng đề cập đến nhiều đóng góp tuyệt vời của ong mật trong nhiều lĩnh vực như y tế, kinh tế và công nghiệp nông nghiệp. Đây là loài côn trùng cực kỳ có giá trị này và là nhân tố chính trong quá trình thụ phấn của cây. Khoảng 73% cây trồng trên thế giới phụ thuộc vào những loài ong này. Giám sát sức khỏe đàn ong đóng vai trò then chốt trong quá trình nuôi ong. Trong thực tế, công việc này thường được thực hiện bằng phương pháp thủ công. Tuy nhiên, phương pháp truyền thống dẫn đến nhiều yếu tố khiến việc kiểm tra không đạt hiệu quả cao vì phương pháp này tốn nhiều thời gian và phụ thuộc chủ yếu vào kinh nghiệm và kiến thức của người nuôi ong. Do đó, các phương pháp dựa trên công nghệ hiện đại (ML - IoT) được khuyến khích với hy vọng giảm tác động xấu đến đàn ong, tiết kiệm thời gian và đưa ra những cảnh báo kịp thời cho người nuôi ong mà không cần kiểm

tra xâm lấn tổ ong. Nhiệm vụ cơ bản đầu tiên của bất kỳ công nghệ giám sát tổ ong dựa trên âm thanh nào là nhận ra âm thanh của ong và phân biệt chúng với âm thanh không phải của ong có thể thu được. Các âm thanh không phải của ong thường liên quan đến môi trường và các sự kiện xảy ra trong môi trường xung quanh tổ ong như âm thanh đô thị, mưa hoặc các động vật khác như tiếng dế.

Một khi hệ thống không phân biệt được âm thanh của ong với những âm thanh không phải của ong, tất cả các công việc liên quan đến phân tích dữ liệu cho các vấn đề cụ thể sau đó cũng sẽ thất bại.

Bài toán chia đàn tự nhiên ở ong: Ong chia đàn tự nhiên là hiện tượng ong chúa cũ và khoảng ½ số quân bay đi hình thành một tổ mới. Đàn ở lại gọi là đàn gốc sẽ do ong chúa mới điều phối đàn ong. Đây là hình thức sinh sản của loài ong. Hiện tượng chia đàn tự nhiên sẽ làm giảm năng suất mật. Vì vậy, người nuôi ong cần biết biểu hiện của đàn ong sắp chia đàn, từ đó có cách phòng số lượng của mật ong giảm dẫn đến thiệt hại kinh tế cho người nuôi ong. Một số hiện tượng của ong chia đàn có thể không được nhận ra bởi những người mới nuôi ong chưa có nhiều kinh nghiệm. Hơn nữa, người nuôi ong cũng không thể theo dõi liên tục các tổ ong để phát hiện kịp thời các vấn đề liên quan đến tình trạng đàn ong mật bằng phương pháp thủ công này. Việc tháo dỡ, mở nắp thùng ong, nhấc cầu ong để kiểm tra bên trong tạo ra sự căng thẳng, hoảng sợ ở ong, hoặc thậm chí làm gián đoạn vòng đời của đàn ong. Chính các yếu tố này là động lực để tìm ra các phương pháp dựa trên công nghệ hiện đại để giảm bớt những nhược điểm của phương pháp truyền thống trong việc theo dõi đàn ong. Ý tưởng chung của cách tiếp cận mới này là sử dụng một hệ thống thiết bị điện tử gắn vào mỗi tổ ong để thu thập dữ liệu. Dữ liệu thu thập được sau đó được phân tích và xử lý để thu được thông tin quan trọng cho phép người nuôi ong theo dõi tổ ong từ xa mà không làm ảnh hưởng đến các đàn ong mật (Du & cs., 2020). Trong số các loại dữ liệu được thu thập từ các đàn ong, âm thanh do đàn ong phát ra (tiếng ong) đóng một vai trò quan trọng trong việc giám sát tổ ong tự động (Ferrari và cộng sự, 2008). Điều này là do tiếng vo ve của ong mang thông tin về hành vi của đàn ong. Mặc dù những người nuôi ong có kinh nghiệm có thể nhận biết được những thay đổi về âm thanh do các đàn ong bị căng thẳng tạo ra, nhưng không phải lúc nào họ cũng có thể xác định được nguyên nhân chính xác của những thay đổi đó mà không cần kiểm tra tổ ong. Do đó, các phương pháp dựa trên công nghệ hiện đại (AI - IoT) được khuyến khích với hy vọng giảm tác động xấu đến đàn ong,

tiết kiệm thời gian và đưa ra những cảnh báo kịp thời cho người nuôi ong mà không cần kiểm tra xâm lấn tổ ong, có khả năng phát hiện việc chia đàn ở ong mật. Ta có thể nhận biết việc chia đàn tự nhiên ở ong và sự hiện diện của ong chúa dựa vào nhận biết âm thanh (Cejrowski & cs., 2018).

Bài toán nhận dạng âm thanh: Gần đây, các phương pháp học máy (ML) nổi lên như một công cụ mạnh mẽ và có những đóng góp to lớn cho các hệ thống giám sát tổ ong tự động với chi phí thấp và hiệu suất tốt hơn. Các hệ thống này cung cấp các giải pháp hiệu quả về chi phí với hiệu suất cao. Trong các ứng dụng giám sát dựa vào dữ liệu âm thanh các mô hình ML được xây dựng dựa trên dữ liệu (data driven) do đó có khả năng phản ánh chính xác hơn các quan hệ giữa các biến dự báo (attributes/features) và các sự kiện (target variable). Điều này cho phép hệ thống đưa ra được những cảnh báo kịp thời đến người sử dụng. Từ dữ liệu âm thanh đầu vào là các file âm thanh (định dạng wav, …) sử dụng các phương pháp trích chọn đặc trưng để chuyển về dạng số (Nguyễn Thế Cường & cs., 2023). Sau đó sử dụng các phương pháp học máy để phân lớp các tập tin âm thanh này (Thái Thuận Thương, 2021). Bằng cách phân tích dữ liệu được thu thập từ tổ ong, các thuật toán ML có thể giải quyết một số vấn đề quan trọng trong việc giám sát các tổ ong như phát hiện sớm hiện tượng chia đàn, xác định sự có mặt của ong chúa trong đàn hay nhận biết các bệnh ở ong (Phan &cs., 2023).

Trong ngành nông nghiệp, ong mật được coi là một trong những loài côn trùng quan trọng nhất. Đây là loài côn trùng cực kỳ có giá trị và là nhân tố chính trong quá trình thụ phấn của cây (Nguyễn Thị Tuyết Nhung, 2014). Giám sát sức khỏe đàn ong đóng vai trò then chốt trong quá trình nuôi ong. Trong thực tế, công việc này thường được thực hiện bằng phương pháp thủ công. Tuy nhiên, phương pháp truyền thống dẫn đến nhiều yếu tố khiến việc kiểm tra không đạt hiệu quả cao vì phương pháp này tốn nhiều thời gian và phụ thuộc chủ yếu vào kinh nghiệm và kiến thức của người nuôi ong.

Trên thế giới, đã có nhiều nghiên cứu xây dựng các ứng dụng và các hệ thống giám sát theo dõi giám sát sức khỏe đàn ong sử dụng dữ liệu âm thanh ong, tuy nhiên những nghiên cứu này còn rất hạn chế ở Việt Nam. Đối với bài toán chia đàn tự nhiên, sau bước thu thập dữ liệu và gán nhãn chính xác, chúng ta cần chuyển đổi dữ liệu âm thanh thành định dạng phù hợp cho việc đầu vào của mô hình học máy. Bước tiếp theo cần lựa chọn được mô hình phù hợp cho với bài toán.

Câu hỏi nghiên cứu đặt ra là: Từ dữ liệu âm thanh thu được, cần rút trích đặc trưng âm thanh như thế nào để:

+ Phân lớp các đối tượng vào hai lớp đối tượng chia đàn và không chia đàn hiệu quả (thể hiện qua các chỉ số đánh giá khác nhau như độ chính xác, Precision hay Recall).

+ Mô hình phân lớp nào phù hợp cho bài toán chia đàn tự nhiên với bộ dữ liệu âm thanh ong thu thập ở Việt Nam.

+ Thời gian tính toán chấp nhận được.

Trả lời được các câu hỏi trên sẽ rất có ý nghĩa thực tế khi áp dụng các mô hình học máy cho bài toán nhận dạng đối tượng dựa trên âm thanh, cụ thể là phát hiện hiện tượng chia đàn ở loài ong mật.

Vì vậy, dưới sự hướng dẫn của TS. Phan Thị Thu Hồng và TS. Phạm Quang Dũng, tôi lựa chọn đề tài "**Nghiên cứu các kỹ thuật học máy áp dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh**" làm đề tài nghiên cứu cho luận văn thạc sĩ của mình.

1.2. CƠ SỞ KHOA HỌC

Cơ sở khoa học được đặt ra để trả lời câu hỏi trên là từ dữ liệu thô ban đầu, chúng ta cần biến đổi, trích xuất các thông tin thích hợp với đầu vào của mô hình học máy. Ở nghiên cứu này chúng tôi áp dụng thuật toán trích xuất đặc trưng phổ biến là hệ số tần số quang phổ Mel (MFCCs: Mel-frequency cepstral coefficients).

Ở đây có vấn đề được đặt ra là, trong việc trích xuất đặc trưng MFCCs của các tệp âm thanh thì các đặc trưng này thường được lấy là giá trị của khung hình (frame) đầu tiên của mỗi đặc trưng MFCC trượt trên mỗi đoạn âm thanh (ta gọi đây là trường hợp mặc định). Trong khi đó thì có nhiều khung hình trượt trên các đoạn âm thanh này (các khung hình này có thể chồng lấn lên nhau) vậy việc lấy giá trị của khung hình đầu tiên đó có đảm bảo cho phương pháp trích xuất đặc trưng MFCC có thể phân lớp với độ chính xác cao nhất không? Vì vậy chúng tôi xem xét các trường hợp lấy phần tử đại diện của MFCC dựa trên các giá trị của khung hình trượt trên mỗi đoạn âm thanh, cụ thể là: trường hợp mặc định như nói ở trên, trường hợp lấy giá trị nhỏ nhất trên tất cả các khung hình, trường hợp lấy giá trị lớn nhất trên tất cả các khung hình, và cuối cùng là trường hợp lấy giá trị trung bình trên tất cả các khung hình để thực hiện cài đặt trên các kịch bản lấy số

lượng đặc trưng khác nhau. Sau đó chúng tôi thực hiện chạy thử nghiệm phân lớp trên bộ dữ liệu âm thanh phát hiện nhận dạng ong chia đàn (swarmming) với các mô hình học máy kinh điển (như cây quyết định, rừng ngẫu nhiên, mô hình hồi quy logistic, k-láng giềng gần nhất, máy véc tơ hỗ trợ) để xem xét nghiên cứu trong đề án này.

1.3. MỤC TIÊU NGHIÊN CỨU

1.3.1. Mục tiêu tổng quát

Tìm hiểu một số thuật toán trích chọn đặc trưng âm thanh, các thuật toán học máy có giám sát và ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong.

1.3.2. Mục tiêu cụ thể

- Tìm hiểu về dữ liệu âm thanh, đặc trưng âm thanh

- Tìm hiểu thuật toán trích chọn đặc trưng âm thanh: MFCCS (Mel-frequency cepstral coefficients)

- Nghiên cứu một số thuật toán học máy có giám sát như: cây quyết định, máy véc tơ hỗ trợ (SVM), rừng ngẫu nhiên (RF), mô hình hồi quy logistic, k-láng giềng gần nhất (k-NN).

- Ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong: Tiền xử lý; Trích chọn đặc trưng; Phân lớp.

1.4. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

1.4.1. Đối tượng nghiên cứu

- Dữ liệu âm thanh.

- Các thuật toán học máy có giám sát.

- Các phương pháp trích chọn đặc trưng âm thanh.

1.4.2. Phạm vi nghiên cứu

- Nghiên cứu một số kỹ thuật học máy áp dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh

- Ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong: Tiền xử lý; Trích chọn đặc trưng; Phân lớp".

PHẦN 2. TỔNG QUAN TÀI LIỆU

2.1. TÌNH HÌNH NGHIÊN CỨU TRONG NƯỚC

Việc nghiên cứu nhận dạng âm thanh và ứng dụng vào trong các bài toán thực tiễn có nhiều ý nghĩa quan trọng và thu hút các nhà nghiên cứu quan tâm và tiến hành. Chúng ta có thể kể ra một số kết quả nghiên cứu của một số tác giả. Năm 2021, tác giả Thái Thuận Thương (2021) đã nghiên cứu sử dụng mạng nơ-ron tích chập để nghiên cứu nhận dạng tiếng nói điều khiển. Cũng năm 2021 này, Hoàng Thị Thanh Giang đã sử dụng phương pháp học máy Deep Boltzmann để nhận dạng giọng chữ cái tiếng Việt (Hoàng Thị Thanh Giang & cs., 2021).

Năm 2022, Nguyễn Chí Ngôn & cs. (2022) đã khảo sát kỹ thuật học sâu trên bài toán chẩn đoán hư hỏng động cơ điện dựa trên tiếng ồn vận hành. Các thuật toán học máy (ML) được huấn luyện dựa trên việc những dạng dữ liệu âm thanh thành một ma trận hoặc véc-tơ (trích xuất đặc trưng) để đưa vào các thuật toán học máy. Cũng trong năm 2022, Phan Thị Thu Hồng & cs. (2022) đã thực hiện so sánh hiệu suất của các thuật toán học máy khác nhau sử dụng năm giải thuật trích xuất đặc trưng âm thanh khác nhau. Đồng thời, nhóm tác giả cũng so sánh kết quả thí nghiệm của họ với kết quả từ nghiên cứu trước đó trong tài liệu. Kết quả thu được cho thấy bằng cách chọn đúng phương pháp trích xuất các đặc điểm quan trọng, hiệu suất của phương pháp ML trong phân loại âm thanh của ong có thể được cải thiện đáng kể (Phan &cs., 2022).

Năm 2023, các tác giả Nguyễn Thế Cường & cs. (2023) đã đề cập đến các cơ sở toán học và phương pháp MFCCs (Mel-Frequency Cepstral Coefficients) nhằm trích xuất các đặc trưng của dữ liệu dạng âm thanh (Nguyễn Thế Cường & cs., 2023). Năm 2023, Phan Thị Thu Hồng & cs. (2023) đã áp dụng một kỹ thuật tiên tiến để điều chỉnh các siêu tham số của các mô hình học máy và điều tra các tính năng hệ số cepstral tần số Mel (MFCC) mới. Mô hình này đã cải thiện đáng kể độ chính xác của các mô hình học máy trong việc nhận dạng và phân loại tiếng ong vo ve khỏi các tiếng ồn xung quanh khác, khiến chúng thậm chí còn tốt hơn một số thuật toán học sâu (Phan và cs., 2023).

2.2. TÌNH HÌNH NGHIÊN CỨU NGOÀI NƯỚC

Những năm gần đây, với sự phát triển mạnh mẽ của công nghệ 4.0 trong đó công nghệ IoT đã giúp thu thập nhiều dữ liệu giám sát tình trạng đàn ong. Do đó, nhu cầu phân tích nâng cao dữ liệu tổ ong và các dữ liệu khác liên quan đến

ong sử dụng các phương pháp học máy ngày càng được quan tâm (Dimitrijević & Zogović, 2022). Dimitrijević, S., & Zogović đã chỉ ra các trường hợp ứng dụng phổ biến nhất có liên quan đến đánh giá/xác thực chất lượng sản phẩm ong và nhận dạng/dự đoán các điều kiện của tổ ong. Để đạt được mục đích này, học có giám sát, cụ thể hơn là phân loại, đã được sử dụng chủ yếu. Các thuật toán được sử dụng thường xuyên nhất là Rừng ngẫu nhiên (RF), Máy vectơ hỗ trợ (SVM), hồi qui logistic (Thiele & cs, 2023) và Mạng nơ ron nhân tạo (ANN). Hơn nữa, nhiều loại dữ liệu được dùng làm đầu vào cho các mô hình bao gồm hình ảnh, âm thanh, dữ liệu từ cảm biến tổ ong, dữ liệu khí tượng, dữ liệu quang phổ trên các mẫu mật ong, v.v. Việc nghiên cứu các mô hình học máy ứng dụng cho nhận dạng dữ liệu âm thanh được nhiều nhà nghiên cứu quan tâm và phát triển (Amlathe, 2018; Das & cs., 2022)... Công nghệ trí tuệ nhân tạo (AI), đặc biệt là học máy, đã đưa ra những cách giải quyết các vấn đề về phân tích dữ liệu tổ ong lớn và phân tích nâng cao của nhiều dữ liệu khác liên quan đến ong như dữ liệu về độc tính của thuốc trừ sâu đối với ong mật hoặc dữ liệu quang phổ trên các mẫu mật ong để xác thực mật ong (Dimitrijević & Zogović, 2022).

Năm 2019, Cao & cs. (2019) đã nghiên cứu nhận dạng tiếng ồn đô thị với mạng nơ-ron tích chập. Năm 2020, Ashar & cs. (2020) đã đề xuất một kiến trúc mới được đề xuất sử dụng mạng thần kinh tích chập (CNN) và hệ số cepstral tần số Mel (MFCC) để xác định người nói trong môi trường ồn ào (Ashar & cs., 2020). Ở đây, Ashar & cs., 2020 sử dụng 39 đặc trưng MFCC và CNN để xác định giọng người nói trong môi trường ồn ào với độ chính xác đạt 87.5% (Ashar & cs., 2020). Các tác giả này không chỉ rõ cách xác định 39 đặc trưng này như thế nào. Cũng trong năm 2020, Ramsey & cs. (2020) chỉ ra một phương pháp không xâm lấn để theo dõi và dự đoán quá trình sinh trưởng của các đàn ong mật bằng cách sử dụng thông tin âm thanh quang phổ rung động. Nhóm tác giả đã áp dụng hai thuật toán phân tích thành phần chính (Principal Component Analysis) và phân tích hàm phân biệt (Discriminant Function Analysis) để dự đoán sự chia đàn, dựa trên dữ liệu rung động được ghi lại bằng gia tốc kế đặt trong tâm tổ ong mật để phân biệt thành công giữa các đàn chuẩn bị và không chia đàn với độ chính xác cao, trên 90% cho mỗi phương pháp, với khả năng dự đoán chia đàn thành công lên đến 30 ngày trước khi diễn ra sự kiện (Ramsey & cs., 2020). Trong nghiên cứu này các tác giả cũng xác định nghiên cứu của họ với hai chiến lược khác nhau là: phân biệt giữa quang phổ tức thời (Discrimination between

Instantaneous Spectra) và phân biệt giữa các quang phổ tiến hóa (Discrimination between Spectral Evolutions). Kết quả của họ chỉ ra chiến lược thứ nhất có độ chính xác và độ ổn định cao hơn với các dữ liệu lấy trong các điều kiện khác nhau (Ramsey & cs., 2020). Điều này gợi ý cho chúng tôi thấy cần nghiên cứu thuật toán với các kịch bản khác nhau để xác định được các kịch bản nào cho kết quả tốt hơn.

Năm 2021, Voudiotis & cs. (2021) đã trình bày một hệ thống giám sát tình trạng của ong được kết hợp với quy trình học sâu để phát hiện sự chia đàn của đàn ong. Hệ thống này bao gồm khả năng thu thập hình ảnh dễ sử dụng và các phương pháp tiếp cận nút cuối khác nhau cho cơ chế tại chỗ hoặc dựa trên đám mây. Hệ thống này cũng kết hợp công cụ CNN thông minh mới có tên Swarm-engine để phát hiện ong và đưa ra thông báo về các trường hợp có thể chia đàn ở ong cho người nuôi ong (Voudiotis & cs., 2021).

Năm 2022, các tác giả Dimitrios & cs. (2022) trình bày kịch bản thử nghiệm của họ về việc thu thập dữ liệu âm thanh của các sự chia đàn và không chia đàn. Họ sử dụng ảnh phổ mel (128x128 pixel) kết hợp với các phương pháp phân loại k-NN và SVM cũng như thuật toán CNN để đánh giá kết quả từ tập dữ liệu gồm có 2788 đối tượng có nhãn không chia đàn và 1435 đối tượng có nhãn chia đàn với tỉ lệ huấn luyện (80%) và kiểm tra (20%). Cuối cùng, các tác giả so sánh ba phương pháp này và trình bày kết quả so sánh chéo của phương pháp tối ưu để phát hiện sớm và muộn gần sự kiện của hiện tượng chia đàn. Trong nghiên cứu đó, các tác giả chỉ ra phương pháp SVM cho kết quả ổn định nhất với các kịch bản phát hiện trước khi chia đàn mười ngày (phát hiện sớm) đạt 90% và phát hiện trước khi chia đàn năm ngày (phát hiện muộn) là 97%, với k-NN tỉ lệ tương ứng là 85% và 98%, còn CNN đạt tỉ lệ 89% và 95% (Dimitrios & cs., 2022).

Các nghiên cứu như trên chưa được áp dụng cho dữ liệu về chia đàn của ong ở Việt Nam. Cho nên trong việc nghiên cứu các mô hình học máy áp dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh của ong là việc có ý nghĩa và cần thiết. Do đó, như mục 1.1 ở trên đã đề cập đến, luận văn này nghiên cứu các kỹ thuật học máy áp dụng cho bài toán nhận dạng đối tượng dựa trên âm thanh. Hơn nữa các mô hình học máy trên thế giới không có nói rõ việc lấy các đại diện đặc trưng MFCC như thế nào? Luận văn này trình bày rõ hơn cách lấy các đặc trưng với đại diện tưng ứng là giá trị trung bình (mean), giá trị mặc định (BT), giá trị nhỏ nhất (min), và giá trị lớn nhất (max) cho 3 kịch bản 13 đặc trưng

MFCCs, 26 đặc trưng MFCCs và 39 đặc trưng MFCCs. Những điều này sẽ được trình bày rõ trong phần kết quả của luận văn này. Hơn nữa kết quả của nhóm tác giả Dimitrios & cs. (2022) cho thấy các phương pháp học máy truyền thống như SVM và k-NN cho kết quả tốt hơn kết quả chạy CNN. Do đó trong luận văn này, chúng tôi nghiên cứu nhận dạng đối tượng dựa trên âm thanh (cụ thể là phát hiện chia đàn ở loài ong dựa trên âm thanh ong) bằng cách kết hợp phương pháp trích chọn đặc trưng MFCC và các mô hình học máy truyền thống như cây quyết định, rừng ngẫu nhiên, k-láng giềng, SVM và hồi quy logistic (LR).

Bên cạnh đó, các nghiên cứu trước đây đã chứng minh hiệu quả của MFCC (Mel-Frequency Cepstral Coefficients) trong việc phân loại và nhận dạng các hiện tượng sử dụng âm thanh thu thập trong tổ ong. Do đó, nghiên cứu này tập trung vào việc sử dụng MFCC cho bài toán nhận dang ong chia đàn tại Việt Nam. MFCC đã được chứng minh là có khả năng trích xuất các đặc trưng âm thanh hiệu quả, đặc biệt là đối với các tín hiệu âm thanh phức tạp như tiếng ong. MFCC có khả năng chống nhiễu tốt, giúp cải thiện độ chính xác của việc phân loại ong. MFCC có thể được tính toán hiệu quả, giúp giảm thiểu thời gian xử lý cho bài toán chia đàn.

PHẦN 3. PHƯƠNG PHÁP NGHIÊN CỨU VÀ XỬ LÝ SỐ LIỆU

3.1. ĐỊA ĐIỂM NGHIÊN CỨU

Khoa Công nghệ thông tin – Học viện Nông nghiệp Việt Nam.

3.2. THỜI GIAN NGHIÊN CỨU

Tháng 8/2023 đến 2/2024.

3.3. ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU

*** Đối tượng nghiên cứu:** Một số phương pháp trích chọn đặc trưng âm thanh và mô hình học máy ứng dụng cho bài toán phân lớp nhận dạng âm thanh.

*** Phạm vi nghiên cứu:** Ứng dụng các phương pháp trích chọn đặc trưng âm thanh và mô hình học máy cho bài toán phân lớp dữ liệu nhận dạng đối tượng dựa trên âm thanh áp dụng vào bài toàn chia đàn ở ong.

3.4. NỘI DUNG NGHIÊN CỨU

Nội dung 1: Tìm hiểu về dữ liệu âm thanh, đặc trưng âm thanh. Tìm hiểu một số thuật toán trích chọn đặc trưng âm thanh.

Nội dung 2: Nghiên cứu một số thuật toán học máy có giám sát.

Nội dung 3: Ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong: Tiền xử lý; Trích chọn đặc trưng; Phân lớp và cài đặt các thuật toán.

3.5. PHƯƠNG PHÁP NGHIÊN CỨU

- Phương pháp lý thuyết: Tham khảo các tài liệu đã được công bố về xử lý tín hiệu âm thanh, các thuật toán học máy cơ bản, các phương pháp lựa chọn đặc trưng.

- Phương pháp tham khảo ý kiến: Tham khảo ý kiến của giáo viên hướng dẫn.

- Phương pháp thực nghiệm: Cài đặt kiểm thử một số thuật toán học máy cho bài toán nhận dạng đối tượng sử dụng âm thanh

- Phương pháp đánh giá và đối sánh: Nhận xét, đánh giá, so sánh các kết quả thực nghiệm.

3.6. THU THẬP VÀ XỬ LÝ SỐ LIỆU

3.6.1. Thu thập dữ liệu

Dữ liệu ong được thu thập từ nhiều trung tâm nuôi ong ở Việt Nam như Trung tâm nghiên cứu ong và nuôi ong nhiệt đới (Học viện Nông nghiệp Việt Nam), trung tâm nuôi ong ở một số nơi khác như ở Bắc Giang, ở Đắc Lắc (Việt

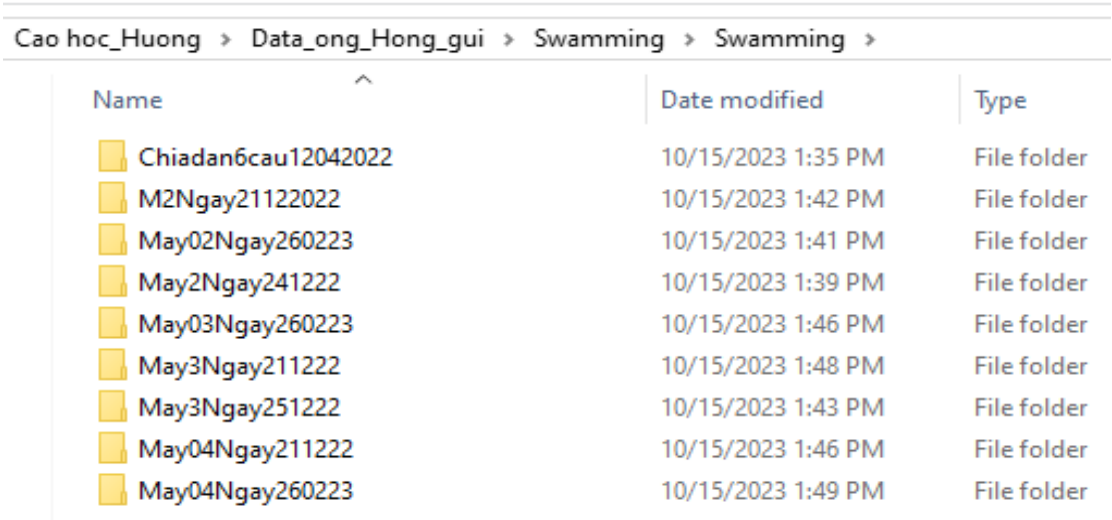
Nam). Riêng dữ liệu ong chia đàn phải nhờ các chuyên gia về ong thực hiện. Họ kiểm tra ong khi thấy dấu hiệu chia đàn thì họ dùng các thiết bị hỗ trợ như các máy thu chuyên dụng. Hiện tượng ong chia đàn là hiện tự nhiên và không biết trước nếu không thường xuyên kiểm tra tổ ong định kì. Các chuyên gia kiểm tra ong tập trung đông, có các hiện tượng bay ra bay vào nhiều thì họ kiểm tra tổ ong khi có dấu hiệu chia đàn thì thu thập dữ liệu chia đàn.

3.6.2. Xử lí dữ liệu âm thanh ong

Từ dữ liệu đã được thu thập, tiền xử lí trích chọn đặc trưng âm thanh và cài đặt một số thuật toán phân lớp ứng dụng vào bài toán phát hiện chia đàn tự nhiên ở ong theo từng công đoạn:

+ Lọc nhiễu: chuyên gia nhận định đó là tiếng ong hay nhiễu hoặc sử dụng các phương pháp nhận biết thông qua tần số thể hiện âm thanh ong. Từ đó ta thu được dữ liệu âm thanh ong.

+ Cắt các file âm thanh dài thành các đoạn âm thanh ngắn (các file.wav) để phục vụ cho trích chọn đặc trưng âm thanh. Ở đây các file âm thanh thu được từ các máy thu âm thanh theo các máy vào các thời điểm khác nhau (Hình 3.1).



Hình 3.1. Dữ liệu âm thanh ong thu được theo thời gian

+ Từ các tệp âm thanh file.wav thực hiện trích xuất đặc trưng âm thanh dựa trên phương pháp hệ số quang phổ tần số Mel (MFCC) để thu được file dữ liệu dạng tệp file.csv trong đó các cột là các đặc trưng âm thanh và dòng tương ứng là các file.wav ở trên. Tệp dữ liệu file.csv ở trên được đưa vào các mô hình học máy để phân lớp.

+ Lí do chọn MFCC là vì: đặc trưng MFCC tính toán các giá trị phổ của tín hiệu âm thanh giống với miền tần số mà tai người có thể cảm nhận được. Phương pháp này đã được sử dụng rộng rãi trên thế giới và có tính hiệu quả cao. MFCC đã được chứng minh là có khả năng trích xuất các đặc trưng âm thanh hiệu quả, đặc biệt là đối với các tín hiệu âm thanh phức tạp như tiếng ong. MFCC có khả năng chống nhiễu tốt, giúp cải thiện độ chính xác của việc phân loại ong. MFCC có thể được tính toán hiệu quả, giúp giảm thiểu thời gian xử lý cho bài toán chia đàn ong. Các nghiên cứu trước đây đã chứng minh hiệu quả của MFCC (Mel Frequency Cepstral Coefficients) trong việc phân loại và nhận dạng các hiện tượng sử dụng âm thanh thu thập trong tổ ong. Do đó, nghiên cứu này tập trung vào việc sử dụng MFCC cho bài toán nhận dạng ong chia đàn tại Việt Nam

PHẦN 4. KẾT QUẢ VÀ THẢO LUẬN

4.1. XỬ LÍ DỮ LIỆU ÂM THANH

4.1.1. Tổng quan về xử lý dữ liệu âm thanh

Tự động nhận dạng hoạt động của các đối tượng sử dụng dữ liệu âm thanh gần đây đã nhận được sự quan tâm đáng kể như một lĩnh vực nghiên cứu đầy hứa hẹn trong lĩnh vực máy học. Chẳn hạn trong lĩnh vực nuôi ong, việc phát hiện các hiện tượng của đàn ong như thiếu chúa (non-queen) hoặc chia đàn (swarming) của ong có ý nghĩa quan trọng đối với người nuôi ong.

Hiện nay các nghiên cứu liên quan đến nhận dạng âm thanh đã được thực hiện trên nhiều hướng phát triển, mục tiêu khác nhau và đạt hiệu quả cao. Có nhiều công cụ/kỹ thuật công nghệ đã được áp dụng nhiều trong thực tế, nhận dạng và xử lý âm thanh. Hệ thống nhận dạng âm thanh bao gồm hai bước chính: bước thứ nhất là rút trích và biểu diễn đặc trưng, bước thứ hai là huấn luyện mô hình máy học nhận dạng. Trong đó, rút trích và biểu diễn đặc trưng tín hiệu âm thanh thường được sử dụng (là hệ số phổ quang tần số thang Mel (MFCC: Mel-scale Frequency Cepstral Coefficient), hệ số dự đoán tuyến tính Linear Prediction coefficients), Biến đổi Fourier nhanh (FFT: Fast Fourier Transform). Các mô hình học máy có thể sử dụng là các phương pháp máy véc tơ hỗ trợ (SVM), rừng ngẫu nhiên, mạng nơ-ron nhân tạo (Mã Trường Thành & cs, 2015).

4.1.2. Âm thanh và biểu diễn âm thanh đối với học máy

a. Âm thanh

Âm thanh thường được định nghĩa là đề cập đến cảm giác thính giác hoặc sự xáo trộn trong môi trường gây ra cảm giác thính giác đó. Là một hiện tượng vật lý, ở đây, âm thanh được hiểu các sóng bắt nguồn từ đâu đó và sau đó truyền qua một môi trường nào đó đến một nơi khác, nơi chúng có thể được nghe hoặc đo được. Những sóng âm như vậy truyền trong chất rắn, chất lỏng và chất khí, và chúng có hai dạng là dọc và ngang.

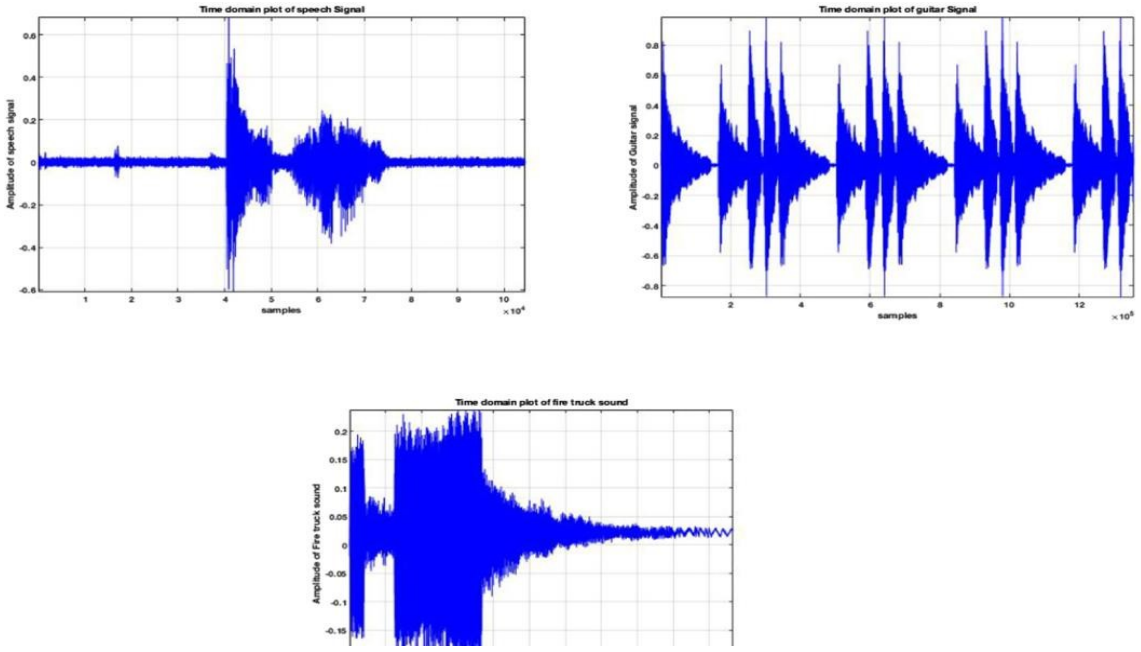
b. Các loại âm thanh

Các tín hiệu âm thanh nghe được được phân loại thành âm thanh giọng nói, âm nhạc và âm thanh môi trường.

- Lời nói: Lời nói được tạo ra bởi con người bằng cách sử dụng kết hợp các cơ quan khác nhau như phổi, miệng, mũi, bụng và não. Thanh quản và

dây thanh âm đóng một vai trò quan trọng trong việc tạo ra lời nói. Quá trình tạo giọng nói bắt đầu ở tần số 100 Hz và có thể lên đến tần số 17 kHZ (Nguyễn Thị Thu, 2022).

- Âm nhạc: Âm thanh do nhạc cụ hoặc con người tạo ra để tạo ra sự hài hòa và thể hiện cảm xúc. Âm nhạc có thể được mô tả theo nhiều khía cạnh khác nhau như thể loại, tâm trạng và đặc điểm âm thanh. Theo truyền thống, âm nhạc được phân thành các thể loại như rock, jazz, cổ điển hoặc pop. Dải tần số lý tưởng của âm nhạc thay đổi từ thấp đến 40 Hz và có thể lên đến 19,5 kHZ (Nguyễn Thị Thu, 2022).



Hình 4.1. Biểu diễn trong miền thời gian của giọng nói, tiếng đàn và âm thanh xe cứu hỏa

- Âm thanh môi trường: Trong cuộc sống hàng ngày, chúng ta bị bao quanh bởi vô số âm thanh môi trường như âm thanh của ô tô hoặc bất kỳ phương tiện nào khác, nước chảy, chuông cửa, chuông điện thoại, tiếng ồn của nhà máy, tiếng động vật, v.v. Những âm thanh này lan truyền trên toàn bộ phạm vi nghe được.

c. Biểu diễn âm thanh với học máy

Như trên ta có thể hiểu: âm thanh là các sóng lan truyền giao động cơ học

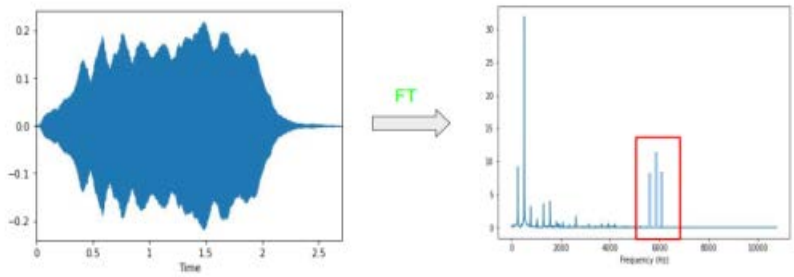
của các phần tử môi trường vật chất. Dạng sóng mang các yếu tố thông tin về tần số, cường độ và âm sắc, có thể tuần hoàn hoặc không tuần hoàn. Dạng sóng có biên độ lớn, ta nghe thấy âm thanh lớn, dạng sóng có tần số cao, ta nghe thấy âm thanh cao (Nguyễn Thế Cường & cs., 2023). Đối với lĩnh vực học máy (ML), một dạng sóng thường được biểu diễn bởi một hàm theo thời gian:

$$y(t) = A\ sin(2\pi ft + \varphi) \quad (1)$$

Trong đó A, f, t, φ tương ứng lần lượt là biên độ, tần số, thời gian và pha ban đầu của một dạng sóng âm thanh liên tục theo thời gian $y(t)$. Ta có thể sử dụng phương pháp lấy mẫu để chuyển đổi từ dạng sóng liên tục sang dạng tần số (dãy các giá trị rời rạc), tỉ lệ mẫu $S_r = 1\backslash T$ thường được chọn là 44100, với T là khoảng thời gian giữa 2 mẫu liên tiếp (Nguyễn Thế Cường & cs., 2023).

4.1.3. Hệ số quang phổ tần số Mel (MFCC)

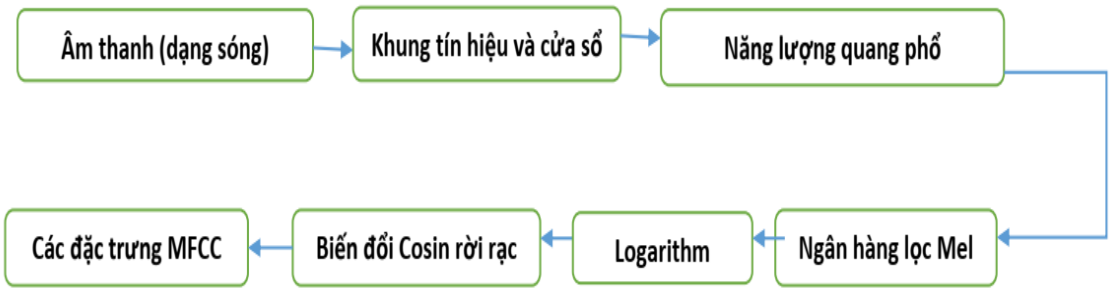
Trong lĩnh vực học máy, âm thanh có các dạng đặc trưng như: các đặc trưng miền thời gian (bao biên độ, căn bậc hai của trung bình của bình phương năng lượng, tỉ lệ băng qua trục hoành), các đặc trưng miền tần số (tỉ lệ dải năng lượng, tâm quang phổ, băng thông), quang phổ. Trong đó việc sử dụng biến đổi Fourier (FT) (Lyons, 2001) để chuyển từ miền thời gian về miền tần số nhằm trích xuất quang phổ. Biến đổi Fourier nhằm phân tích một âm thanh phức thành các thành phần tần số của nó (Hình 4.2)



Hình 4.2. Biến đổi Fourier đưa miền thời gian (t) về miền tần số (f)

Như vậy để trích xuất đặc trưng miền tần số của một tín hiệu âm thanh, chúng ta sử dụng phương pháp lấy mẫu và trên cơ sở của biến đổi Fourier phức để thu được các thông tin về tần số và độ lớn quang phổ. Từ đó tiến hành thêm các bước biến đổi tiếp sau biến đổi Fourier thời gian ngắn để chuyển sang hệ số quang phổ tần số Mel (MFCCs), một phương pháp trích xuất các đặc trưng một tín hiệu âm thanh phổ biến hơn và hiệu quả hơn quang phổ (Nguyễn Thế Cường & cs., 2023).

MFCC là một trong những đặc trưng được sử dụng phổ biến trong nhiều ứng dụng, đặc biệt là trong xử lý tín hiệu giọng nói như nhận dạng người nói, nhận dạng giọng nói và nhận dạng giới tính. MFCC có thể được tính bằng cách tiến hành năm quy trình liên tiếp, cụ thể là tạo khung tín hiệu, tính toán phổ công suất, áp dụng ngân hàng bộ lọc Mel cho phổ công suất thu được, tính toán giá trị logarit của tất cả các ngân hàng bộ lọc và cuối cùng áp dụng DCT (Abdul & Al-Talabani, 2022). Hình 4.3 minh họa quá trình tính toán MFCC.



Hình 4.3. Khung làm việc của MFCC

a. Nhấn mạnh trước

Nhấn mạnh trước (Pre-emphasis) là một trong những phương pháp tiền xử lý phổ biến trong lĩnh vực xử lý tín hiệu được sử dụng để bù tần số cao của tín hiệu đã bị triệt tiêu trong quá trình tạo tín hiệu. Nhấn mạnh trước là bước đầu tiên trong quá trình điều chỉnh MFCC, có thể được thực hiện bằng cách áp dụng bộ lọc thông cao có phương trình sai phân là:

$$y(n) = x(n) - \alpha.x(n-1) \quad (2)$$

trong đó $y(n)$ là mẫu tín hiệu ra sau khi nhấn mạnh trước; $x(n)$ là mẫu tín hiệu vào; $x(n-1)$ là mẫu vào trước $x(n)$; α là hằng số được chọn trong khoảng từ 0.9 đến 1.0 và thường sử dụng là 0.97.

b. Khung tín hiệu và cửa sổ

Tín hiệu được chia thành các khung chồng lên nhau (khoảng 50%-70% để tránh mất thông tin) để tính hệ số MFCC. Giả sử mỗi khung bao gồm N mẫu và để các khung liền kề được phân tách bằng M mẫu trong đó M < N. Mỗi khung được nhân với một cửa sổ Hamming trong đó phương trình cửa sổ Hamming được cho bởi:

$$W(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$

c. Năng lượng quang phổ

Phổ công suất có thể được mô tả là sự phân bố công suất của các thành phần tần số tạo nên tín hiệu (Abdul & Al-Talabani, 2022). Theo truyền thống, Biến đổi Fourier rời rạc (DFT) được sử dụng để tính toán phổ công suất. Phổ công suất của từng khung thu được phải được xác định dựa trên phương trình (4) dưới đây

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{i2\pi kn}{N}} \quad (4)$$

Với $k = 1,2,,\dots,N-1$ và $x(n)$ là tín hiệu rời rạc, N là chiều dài của tín hiệu.

Trong bước tiếp theo, tín hiệu miền tần số được chuyển đổi sang thang tần số Mel, phù hợp hơn với thính giác và nhận thức của con người. Điều này được thực hiện bằng một tập hợp các bộ lọc hình tam giác được sử dụng để tính toán tổng trọng số của các thành phần quang phổ sao cho đầu ra của quá trình xấp xỉ thang đo Mel. Đáp ứng tần số cường độ của mỗi bộ lọc có hình tam giác và bằng 1 ở tần số trung tâm và giảm tuyến tính về 0 ở tần số trung tâm của hai bộ lọc liền kề.

d. Bộ lọc Mel

Bộ lọc thông dải Mel là một tập hợp các bộ lọc được xây dựng dựa trên nhận thức cao độ (pitch perception). Bộ lọc thông dải Mel là một tập hợp các bộ lọc được thiết kế dựa trên cách thức con người nghe và nhận thức cao độ (pitch perception) của âm thanh. Cụ thể, bộ lọc Mel thường được sử dụng trong quá trình trích xuất đặc trưng từ tín hiệu âm thanh, như trong quá trình trích xuất MFCCs. Cách thức hoạt động của bộ lọc Mel bắt nguồn từ việc con người không nghe và cảm nhận tần số âm thanh theo cách đồng đều trên toàn bộ dải tần số, mà thay vào đó, họ cảm nhận tần số theo cách không phản ánh hoàn toàn khả năng của tai mình. Cụ thể, bộ lọc Mel chia dải tần số thành các dải tần số không đều, trong đó các dải tần số thấp có độ phân giải cao hơn so với các dải tần số cao. Điều này phản ánh cách mà tai con người cảm nhận và xử lý âm thanh, trong đó có sự nhạy cảm đặc biệt đối với các biến động ở dải tần số thấp, trong khi ít nhạy cảm hơn đối với các biến động ở dải tần số cao. Kết quả là, bộ lọc thông dải Mel giúp tăng cường tính nhất quán với cách thức con người nghe và nhận thức âm thanh, làm cho các đặc trưng trích xuất từ tín hiệu âm thanh (như MFCCs) trở nên phù hợp hơn với các nhu cầu của các ứng dụng như nhận dạng giọng nói. Bộ lọc Mel ban đầu được phát triển để phân tích giọng nói và giống như khả năng

nhận biết giọng nói của tai người, nó nhắm mục tiêu trích xuất biểu diễn phi tuyến tính của tín hiệu giọng nói. Ngân hàng bộ lọc Mel quy ước được xây dựng từ 40 bộ lọc hình tam giác. Hàm dịch chuyển của mỗi bộ lọc thứ m được tính qua hàm số

$$H_m(k) = \begin{cases} 0 & khi \quad k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & khi \quad f(m-1) \le k < f(m) \\ 1 & khi \quad k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & khi \quad f(m) < k \le f(m+1) \\ 0 & khi \quad k > f(m+1) \end{cases} \tag{5}$$

Trong đó, $f(m)$ là tần số trung tâm của bộ lọc tam giác và $\sum_m^{N-1} H_m(k) = 1$. Thang đo Mel theo tần số đáp ứng và ngược lại được tính theo phương trình (6) và (7)

$$m = 2595 \log_{10}(1 + \frac{f}{700}) \tag{6}$$

hay

$$f = 700(10^{\frac{m}{2595}} - 1) \tag{7}$$

Hàm dịch chuyển này có tác dụng xác định mức độ đóng góp của từng bộ lọc vào việc biểu diễn tín hiệu âm thanh theo các dải tần số Mel khác nhau. Điều này giúp tạo ra các đặc trưng biểu diễn tín hiệu phản ánh cách thức con người nghe và nhận thức âm thanh.

e. Biến đổi Cosin rời rạc (DCT)

Cuối cùng, để trích xuất cepstrum (tạm dịch là quang phổ) ta thực hiện biến đổi Cosin rời rạc (DCT- Discrete Cosine transform) để chuyển từ miền tần số về miền thời gian, ta thu được MFCCs.

Biến đổi Cosine rời rạc (DCT) biểu thị một chuỗi hữu hạn các điểm dữ liệu liên quan đến tổng các hàm cosine dao động ở các tần số khác nhau. DCT được Nasir Ahmed giới thiệu vào năm 1972. Trong quy trình MFCC, DCT được áp dụng trên ngân hàng bộ lọc Mel để chọn hầu hết các hệ số gia tốc hoặc để tách mối quan hệ trong cường độ phổ log khỏi ngân hàng bộ lọc. Giả sử $Y_t(m)$ là tín hiệu ngõ ra từ bộ lọc của ngân hàng học Mel, ta có, DCT được tính theo phương trình dưới đây.

$$\mathrm{cep_t}(k) = \sum_{m=1}^{M} \log(|Y_t(m)|^2) \cos\frac{k(m-0.5)\pi}{M} \tag{8}$$

Ở đây $1 \leq k \leq 12$ (lựa chọn 12 tín hiệu đầu tiên của thuộc tính phổ). M là số mẫu trong một khung hình.

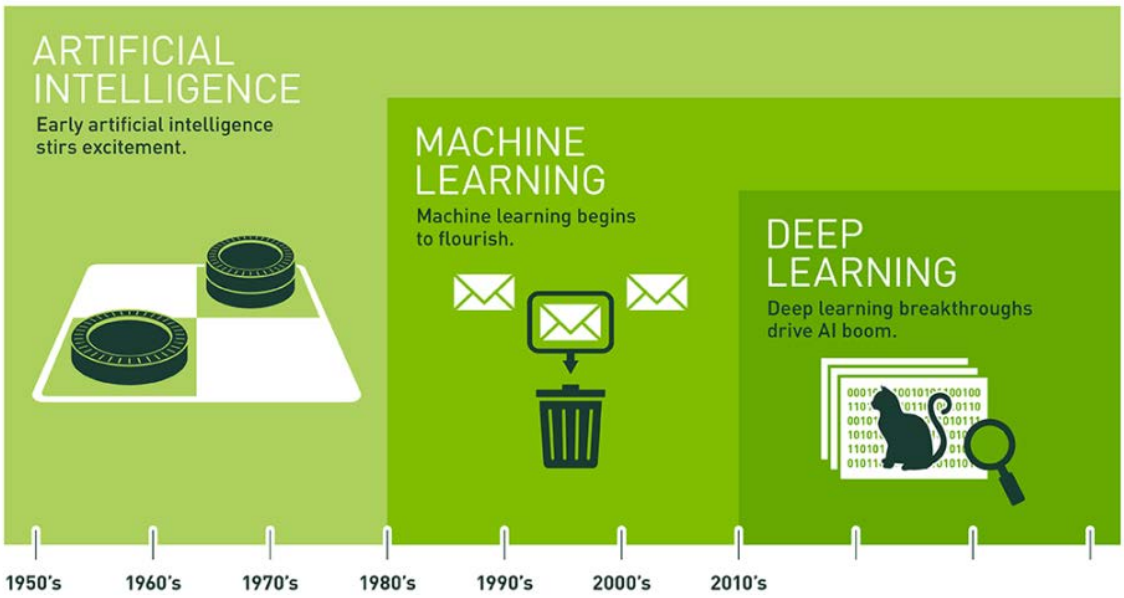
Từ đây ta có 12 hệ số rời rạc của hệ số quang phổ Mel kết hợp với năng lượng phổ của mỗi khung tín hiệu là ta có hệ số thứ 13 của các đặc trưng MFCCs (Kurzekar và cộng sự, 2014). Tiếp theo tính các sai phân đầu ra cấp 1 và cấp 2 của 13 hệ số rời rạc trên ta có tổng cộng 39 đặc trưng MFCC của mỗi khung tín hiệu.

Tổng kết: mục 4.1 đã trình bày về vấn đề xử lí âm thanh, các thu thập và biểu diễn âm thanh. Cũng như tìm hiểu các kĩ thuật trích xuất đặc trưng âm thanh như biến đổi Fourier và trích xuất đặc trưng MFCC để phục vụ cho nghiên cứu ở phần sau.

4.2. CÁC PHƯƠNG PHÁP HỌC MÁY

Trong phần này, đề án đề cập đến các phương pháp học máy được sử dụng cho bài toán phân lớp.

Ngày nay, trí tuệ nhân tạo (AI) mà học máy (ML-Machine Learning) là một nhánh của AI (hình 4.4) nổi lên như một minh chứng cho cuộc cách mạng công nghiệp 4.0. Trí tuệ nhân tạo đã và đang len lỏi vào mọi lĩnh vực trong cuộc sống của con người như: xe tự lái của Google và Tesla, hệ thống tự gắn thẻ của Facebook trong ảnh, trợ lý ảo Siri của Apple, hệ thống đề xuất sản phẩm của Amazon, hệ thống đề xuất phim của Netflix, máy nghe nhạc AlphaGo của Google, DeepMind, ... là một vài trong số rất nhiều ứng dụng của AI/ ML.



Hình 4.4. Mối liên hệ giữa AI, học máy và học sâu (deep learning)

4.2.1. Học máy

Học máy (Machine learning) là quá trình đào tạo máy tính bắt chước hành vi của con người. Trong khi con người học hỏi từ kinh nghiệm và thông qua các giác quan thì máy tính học hỏi từ dữ liệu. Trần Đăng Tú & cs. (2022) chỉ ra "Học máy là phương pháp phân tích dữ liệu tự động hóa thông qua mô hình phân tích. Bằng cách sử dụng các thuật toán hiện đại để học từ dữ liệu, học máy cho phép máy tính tìm thấy những thông tin, giá trị ẩn sâu mà không thể lập trình rõ ràng. Cách học của học máy rất quan trọng để khi các mô hình mạng này được tiếp xúc với dữ liệu mới, có thể thích ứng một cách độc lập nhờ học từ các tính toán trước đó để đưa ra quyết định cũng như kết quả lặp lại đáng tin cậy".

Học máy được xác định gồm các thành phần sau (Hồ Thị Ngọc, 2012): Cho trước một tập dữ liệu vũ trụ X. Cho S là một tập con của X, S được gọi là tập mẫu. Có một số hàm đích (quá trình ghi nhãn) $f: X \to Y$ với $y = f(x)$ trong đó Y là tập nhãn. Một tập huấn luyện D được xác định bởi $D = \{(x, y)|y = f(x), x \in S\}$. Tính toán một hàm $f': X \to Y$ với $y' = f'(x)$ bằng cách sử dụng D như là $f(x) \cong f'(x)$ với mọi $x \in X$.

Như vậy học máy gồm các thành phần (X, Y, f, S, D, f') trong đó X là tập dữ liệu vũ trụ (hay dữ liệu thuộc tính/đặc trưng), Y là tập nhãn, f là hàm gán nhãn, S là tập mẫu trên X, $D = \{(x, y)|y = f(x), x \in S\}$ và tính toán xác định $f': X \to Y$ với $y' = f'(x), \forall x \in X$ (trên tập huấn luyện D thì $f(x) \cong f'(x)$).

Các kĩ thuật học máy được chia ra thành:

+ Học không có giám sát (Unsupervised learning): Học với tập dữ liệu ban đầu hoàn toàn chưa được gán nhãn.

+ Học có giám sát (Supervised learning): Học với tập dữ liệu huấn luyện ban đầu hoàn toàn được gán nhãn.

Khi đó có hàm mất mát $L: Y \times Y \to \mathbb{R}^+$ với $L(y, y')$ là số thực không âm. Mất mát của hàm f' trên $S_1 \subset X$ được xem xét là $R(f') = \frac{1}{|S_1|} \sum_{i=1}^{|S_1|} L(y, y')$ với $|S_1|$ là số phần tử có trong S_1.

Các thuật toán học có giám sát còn được phân ra thành hai loại chính là phân lớp (Classification) và hồi quy (Regression).

+ Học bán giám sát (Semi-supervised learning): Học cả với dữ liệu có gán nhãn và chưa gán nhãn.

Ngoài ra còn có thêm cách học tăng cường được xác định như sau:

+ Học tăng cường (reinforcement learning): Học tăng tường hay học củng cố là bài toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất. Hiện tại, học tăng cường chủ yếu được áp dụng vào Lý Thuyết Trò Chơi (Game Theory), các thuật toán cần xác định nước đi tiếp theo để đạt được điểm số cao nhất. Chẳng hạn, AlphaGo - một phần mềm chơi cờ vây trên máy tính được xây dựng bởi Google DeepMind hay chương trình dạy máy tính chơi game Mario là những ứng dụng sử dụng học tăng cường. Trong lý thuyết trò chơi, học tăng cường thường được sử dụng để xây dựng các chiến lược tối ưu cho các tác nhân (agents) trong một môi trường tương tác. Cụ thể, học tăng cường giải quyết bài toán của các tác nhân đưa ra các hành động trong một môi trường nhất định để tối đa hóa một hàm phần thưởng (reward function) hoặc tối thiểu hóa một hàm chi phí (cost function).

4.2.2. Bài toán phân loại

Như trên đã nói, các thuật toán học có giám sát còn được phân ra thành hai loại chính là phân lớp (Classification) và hồi quy (Regression). Hay nói cách khác, phân loại là bài toán học có giám sát.

Bài toán phân loại

Một bài toán được gọi là phân loại (hay phân lớp) nếu các nhãn của dữ liệu đầu vào được chia thành một số hữu hạn lớp (miền giá trị là rời rạc). Chẳng hạn như tính năng xác định xem một email có phải là spam hay không của Gmail; xác định xem hình ảnh của con vật là chó hay mèo. Tương tự cho ví dụ nhận dạng khuôn mặt với hai lớp là phải và không phải khuôn mặt, …

4.2.3. Một số kĩ thuật học máy cho bài toán phân loại

Các kỹ thuật học máy có rất nhiều. Trong phần này chúng ta đề cập đến một số kỹ thuật học máy sẽ được sử dụng trong đề án này.

4.2.3.1. Mô hình cây quyết định

Một trong những mô hình học tập phổ biến nhất là Cây quyết định (DT-decision tree). Cây quyết định (DT) là mô hình dự đoán trong học tập có giám sát, được biết đến không chỉ vì tiện ích trong nhiều ứng dụng mà còn vì khả năng diễn giải và tính mạnh mẽ của chúng. Các mô hình này thường được biểu diễn theo cấu trúc giống như flowchart, trong đó mỗi nút bên trong là một phép thử

logic (được gọi là phần tách) và mỗi lá là một dự đoán. Trong quá trình suy luận, mỗi quan sát bắt đầu từ gốc và kết thúc ở một trong các lá, đi theo một đường dẫn duy nhất. Sự đơn giản của phương pháp này đi ngược lại tính chắc chắn của nó: DT được biết đến không chỉ như một mô hình có thể diễn giải, dễ nắm bắt mà còn là một phương pháp chính xác đã đứng vững trước thử thách của thời gian, kể từ đề xuất ban đầu của họ vào những năm 1960 (Morgan & Sonquist, 1963). Cây cũng có nhiều ưu điểm khác, chẳng hạn như chi phí tính toán thấp, có thể xử lý các giá trị còn thiếu và khả năng xử lý ngay lập tức các dữ liệu kết hợp hỗn hợp (Costa & Pedreira, 2023).

Với nhiệm vụ phân lớp, các bước thực hiện cây quyết định được khái quát hóa như sau (Costa & Pedreira, 2023):

(1) Cho X là tập dữ liệu với các thuộc tính $\{x_1, x_2, \dots, x_n\}$

(2) Nếu X được phân vùng đồng nhất với nhãn y thì tạo một lá dự đoán nhãn y và trở lại

(3) Ngược lại, mỗi thuộc tính x_i: Với mỗi nút hoặc giá trị c ta thực hiện các bước sau:

a) Lựa chọn ứng viên tốt nhất lựa chọn phân chia $(x_i \leq c)$ hoặc $(x > c)$ dựa theo tiêu chuẩn phân chia.

b) Tạo một nút bên trong với (x_i, c) là điểm tách.

c) Chia X thành hai tập dữ liệu $X\_trái$ và $X\_phải$, phù hợp, dựa theo điểm tách (x_i, c).

d) Tạo một nút con cho mỗi tập dữ liệu mới được tạo.

e) Đối với mỗi nút con, thực hiện bước a).

- Ưu điểm của phương pháp cây quyết định (Bansal & cs., 2022)

+ Có độ phức tạp rất thấp – Thuật toán này khá đơn giản để hiểu và không phụ thuộc vào bất kỳ chuyên ngành nào kiến thức liên quan đến thống kê để giải thích nó.

+ Hữu ích trong việc khám phá dữ liệu - Nó cũng có thể được sử dụng trong các giai đoạn khám phá dữ liệu như thuật toán cây quyết định chứng tỏ là một trong những các thuật toán nhanh nhất trong việc tạo hoặc xác định các tính năng mới.

+ Yêu cầu làm sạch dữ liệu ít hơn - Tương đối yêu cầu ít hơn bước làm sạch dữ liệu và không bị ảnh hưởng bởi các giá trị và dữ liệu bị thiếu.

+ Không hạn chế kiểu dữ liệu - Có thể xử lý số một cách linh hoạt như cũng như các biến có tính chất phân loại.

+ Phương pháp phi tham số - Cây quyết định sử dụng phương pháp phi tham số phương pháp, ngụ ý không đưa ra giả định nào về phân bố của dữ liệu.

- Nhược điểm của cây quyết định

Mặc dù cây quyết định là một phương pháp phân loại mạnh mẽ và dễ hiểu, nhưng nó cũng có một số nhược điểm:

+ Vấn đề quá khớp hay quá mức đào tạo (overfitting) – Vấn đề quá khớp là một trong những vấn đề thực tế chính ảnh hưởng đến mô hình cây quyết định. Cây quyết định có thể dễ bị quá mức đào tạo trên dữ liệu huấn luyện, đặc biệt là khi cây quá sâu hoặc khi có quá nhiều lá. Điều này dẫn đến việc mô hình hoạt động kém hiệu quả trên dữ liệu mới hoặc không được thấy trước.

+ Không ổn định với dữ liệu nhạy cảm với nhiễu: Cây quyết định dễ bị ảnh hưởng bởi nhiễu và các biến thừa trong dữ liệu, dẫn đến việc tạo ra các quy tắc quyết định không cần thiết hoặc không đáng tin cậy.

+ Khả năng xử lý dữ liệu liên tục và dữ liệu lớn: Trong một số trường hợp, cây quyết định không hiệu quả khi xử lý dữ liệu liên tục hoặc dữ liệu lớn, đặc biệt là khi số lượng biến và mẫu lớn. Cây quyết định mất một số thông tin có giá trị khi phân loại các biến theo nhiều loại khác nhau.

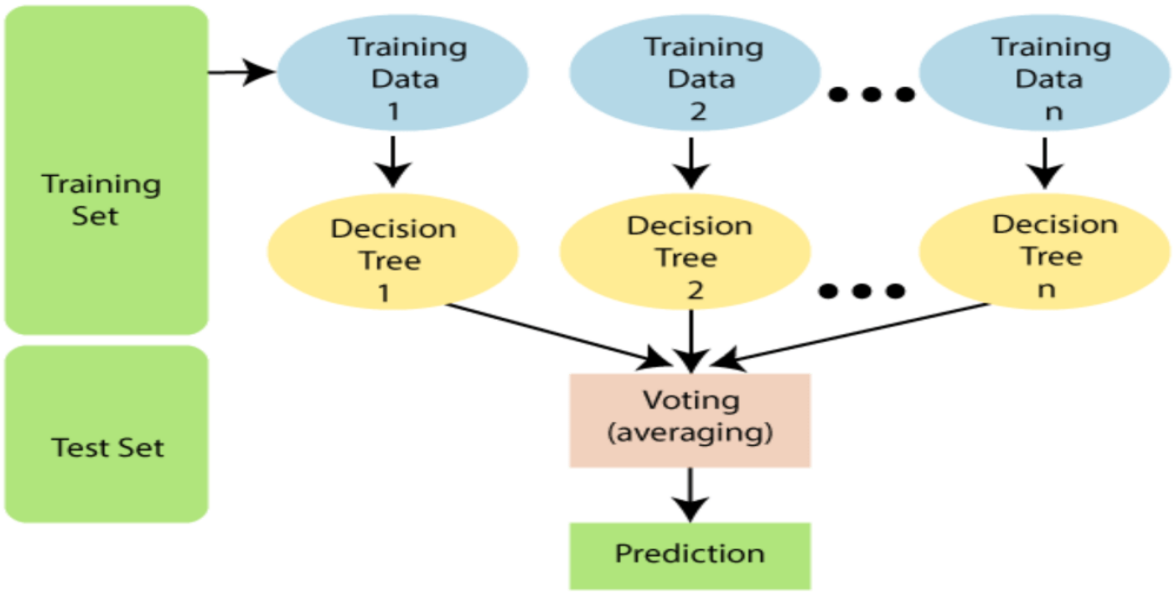
+ Tính không ổn định với các biến đầu vào quan trọng: Cây quyết định có thể không ổn định với các biến đầu vào quan trọng, khiến cho các quy tắc quyết định và cấu trúc của cây thay đổi mạnh mẽ khi dữ liệu thay đổi.

+ Tính chủ quan trong việc lựa chọn thuộc tính phân chia: Quyết định về thuộc tính phân chia có thể phụ thuộc nhiều vào phương pháp lựa chọn, dẫn đến sự không đồng nhất giữa các mô hình được tạo ra từ dữ liệu khác nhau.

4.2.3.2. Mô hình rừng ngẫu nhiên

Trong số nhiều thuật toán học tập hợp, thuật toán rừng ngẫu nhiên (RF), được giới thiệu ở Breiman (2001), có lẽ là một trong những phương pháp thành công nhất (Hồng & cs., 2023). Nó bao gồm nhiều cây đơn lẻ như trong hình 4.5, trong đó mỗi cây trong rừng được tạo trên một mẫu dữ liệu huấn luyện ngẫu

nhiên. Trong phương pháp này, những cây trưởng thành hoàn toàn được tạo ra mà không cần cắt tỉa để giữ độ lệch thấp. Mỗi cây được huấn luyện trên một tập mẫu bootstrap và để phân tách ở mỗi nút, một tập hợp con ngẫu nhiên của các thuộc tính sẽ được xem xét.



Hình 4.5. Cấu trúc của phân lớp rừng ngẫu nhiên

Bằng cách này, tính ngẫu nhiên tạo ra sự đa dạng giữa các cây và mối tương quan thấp giữa các cây được kiểm soát trong rừng. Kết quả là chúng chính xác và ổn định hơn cây cối được thêm vào.

Các thủ tục chính trong RF như sau (Hồng & cs., 2023):

- Từ tập dữ liệu huấn luyện, với m mẫu và n biến (đặc điểm), xây dựng cây quyết định T độc lập.
- Mô hình cây quyết định thứ t được xây dựng trên t-bootstrap bộ mẫu từ tập dữ liệu ban đầu (Learning set).
- Tại mỗi nút bên trong, chọn ngẫu nhiên n' biến ($n' << n$) và tính toán phân vùng tốt nhất dựa trên n' biến này.
- Cây chưa cắt tỉa được xây dựng với độ sâu tối đa.
- Kết quả cuối cùng thu được bằng cách tổng hợp các giá trị kết quả chẳng hạn như phương pháp voting theo kết quả được dự đoán bởi nhiều cây nhất hoặc phương pháp lấy giá trị trung bình.
- Ưu điểm của rừng ngẫu nhiên

+ Thuật toán này có thể áp dụng cho lớp bài toán phân loại và hồi quy.

+ Nó giải quyết vấn đề quá khớp với dữ liệu vì đầu ra dựa trên biểu quyết đa số hoặc lấy trung bình.

+ Nó hoạt động tốt ngay cả khi dữ liệu bị thiếu.

+ Nó có tính ổn định cao.

+ Nó duy trì sự đa dạng vì tất cả các thuộc tính không được xem xét trong khi tạo mỗi cây quyết định mặc dù nó không đúng trong mọi trường hợp.

+ Nó không bị ảnh hưởng khi số chiều dữ liệu tăng vì mỗi cây không xét tới tất cả các thuộc tính.

- Nhược điểm của phương pháp rừng ngẫu nhiên

+ Thuật toán Rừng ngẫu nhiên rất phức tạp khi so sánh với các cây quyết định nơi các quyết định có thể được thực hiện bằng cách đi theo con đường của cây.

+ Thời gian huấn luyện nhiều hơn so với các mô hình khác do tính phức tạp của nó. Bất cứ khi nào nó phải đưa ra dự đoán, mỗi cây quyết định phải tạo ra đầu ra cho dữ liệu đầu vào đã cho.

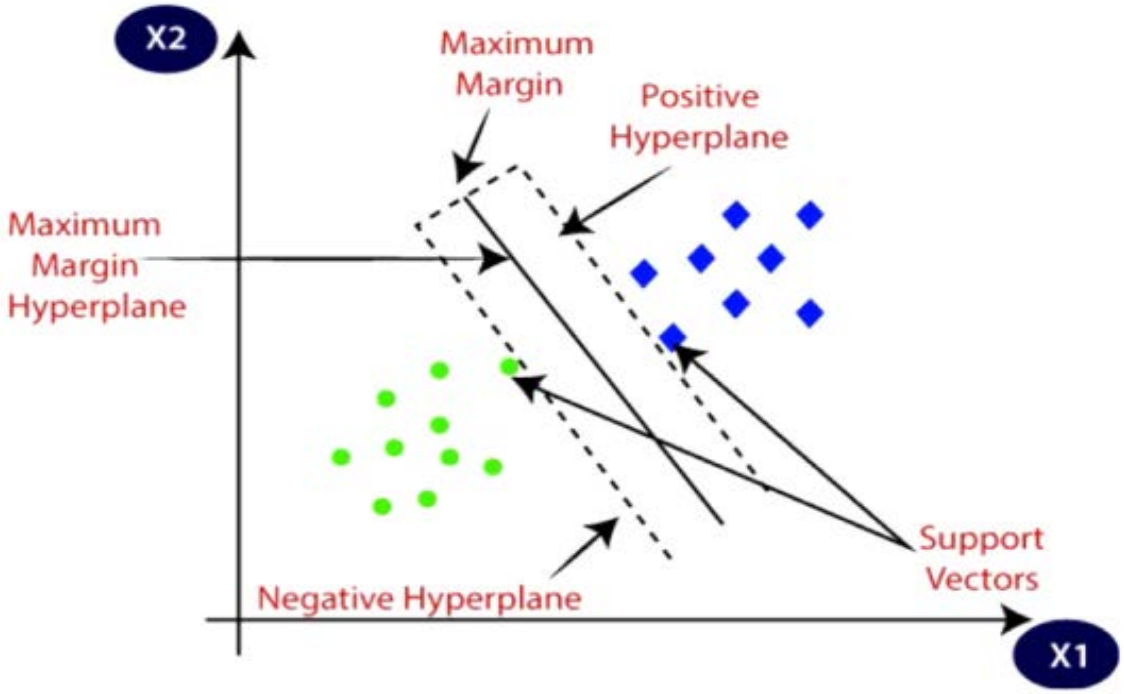
4.2.3.3. Mô hình máy véc tơ hỗ trợ (SVM)

Máy vectơ hỗ trợ (SVM-support vector machine) được tạo bởi Alexey ya. Chervonenkis và Vladimir N. Vapnik vào năm 1963 (Bansal & cs., 2022). Kể từ đó, kỹ thuật này đã được áp dụng rộng rãi để sử dụng trong các vấn đề phân tách và phân loại hình ảnh, siêu văn bản và văn bản. Các thuật toán này khá tiên tiến và có thể được sử dụng cho văn bản viết tay cũng như phân loại protein trong phòng thí nghiệm sinh học. Chúng được sử dụng trong nhiều lĩnh vực khác như ô tô tự lái, chatbot, nhận dạng khuôn mặt, v.v. (Bansal & cs., 2022). Trong số những thuật toán học kiểu có giám sát rất phổ biến, thuật toán Support Vector Machine dành cho các vấn đề hồi quy và phân loại. Thuật toán SVM nhằm mục đích hình thành giới hạn quyết định phù hợp nhất hoặc ranh giới, được gọi là siêu phẳng, ngăn cách không gian n chiều thành nhiều lớp khác nhau, giúp dễ dàng đặt một điểm khác trong thể loại thích hợp. Trong thuật toán SVM, các điểm vectơ cực trị được gọi là vectơ hỗ trợ được chọn để giúp tạo ra một siêu phẳng thích hợp. Kết quả là, nó có thể tối đa hóa khoảng cách giữa các điểm dữ liệu của cả hai lớp. Hình 4.6 cung cấp một hình dung trực quan cho ý tưởng cốt lõi của phương pháp này.

Siêu phẳng tách biệt có dạng hàm hồi quy tuyến tính, có thể được biểu diễn bằng công thức (Phan & cs., 2022):

$$f(x) = w^T x + b \quad (10)$$

trong đó, w là véc tơ trọng số, b là độ lệch, và x là véc tơ đầu vào.



Hình 4.6. Minh họa phân lớp tuyến tính với SVM

- Ưu điểm của phương pháp SVM (Bansal & cs., 2022):

+ Thuật toán SVM phù hợp nhất khi có sự phân chia rõ ràng giữa các lớp.

+ SVM cho thấy hiệu quả cao hơn khi nói đến không gian nhiều chiều hơn.

+ SVM tương đối hiệu quả về mặt bộ nhớ, đây là một tính năng khá đáng mong đợi.

+ Tính ổn định: Một thay đổi nhỏ đối với dữ liệu không ảnh hưởng lớn đến siêu phẳng và do đó không ảnh hưởng đến hiệu suất của mô hình SVM. Vì vậy, mô hình SVM ổn định.

- Nhược điểm của phương pháp SVM (Bansal & cs., 2022)

+ Thuật toán SVM không thể hoạt động phù hợp với các tập dữ liệu có kích thước khổng lồ.

+ SVM không hoạt động hiệu quả trong trường hợp tập dữ liệu chứa lượng nhiễu lớn, tức là các lớp mục tiêu chồng chéo, thực tế là hiện tượng này lại rất thường xuyên xảy ra.

+ SVM hoạt động kém trong trường hợp giá trị số của các đối tượng của mỗi điểm dữ liệu cao hơn mẫu dữ liệu huấn luyện.

+ Thuật toán SVM không có khả năng giải thích xác suất cho phân loại.

+ Thời gian huấn luyện dài: SVM mất nhiều thời gian huấn luyện trên các tập dữ liệu lớn.

4.2.3.4. Mô hình hồi quy logistic (LR)

Trong bài toán phân lớp nhị phân, mô hình hồi quy logistic cũng được sử dụng để phân lớp đối tượng (Gifford & Bayrak, 2023). Bài toán được xem xét trong nghiên cứu này là bài toán phân loại nhị phân. Xác suất của một đầu ra nhị phân, cho một tập hợp các biến, được biểu diễn như sau (Yaseliani & Khedmati, 2023):

$$p(y|x) = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}} \quad (11)$$

trong đó y là biến đầu ra, $x = (x_1, x_2, \dots, x_n)$ là tập hợp các biến được dự báo, $b = (b_0, b_1, b_2, \dots, b_n)$ là các hệ số, $p(y|x)$ là xác xuất biến y thuộc vào lớp 1 trong điều kiện các biến x được xác định trước.

Tỷ lệ chênh lệch (odds ratio) là một trong những phép đo thống kê được sử dụng trong việc ra quyết định. Tỷ lệ chênh lệch được định nghĩa là tỷ lệ giữa xác suất xảy ra và không xảy ra của một sự kiện:

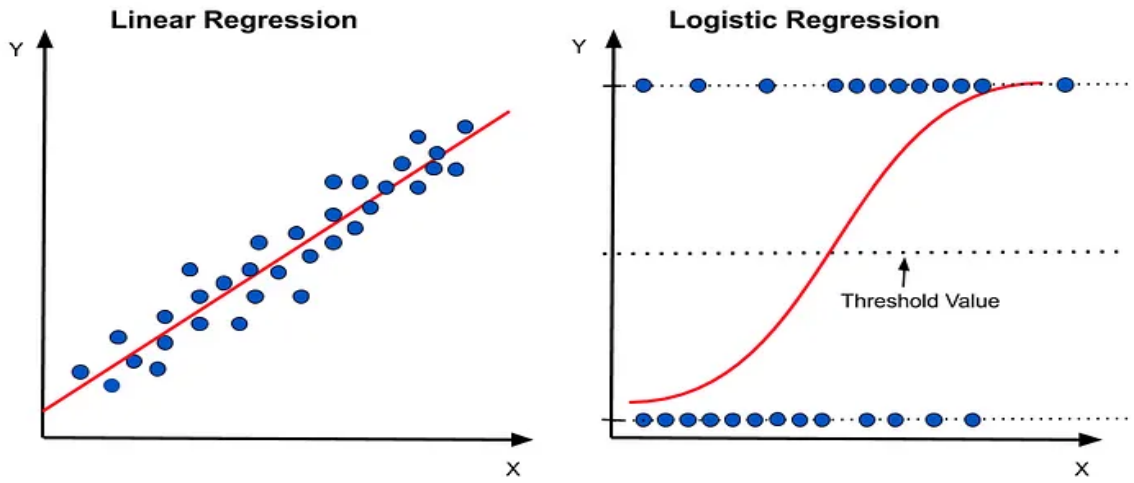
$$odds = \frac{p(y|x)}{1-p(y|x)} \quad (12)$$

Hàm logarit được xác định là logarit tự nhiên của tỷ lệ chênh lệch.

$$\ln(odds) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \quad (13)$$

Trong hồi quy logistic, thông thường kết quả thông qua một hàm đặc biệt được gọi là Hàm Sigmoid để dự đoán đầu ra y, tức là y được dự đoán thuộc về lớp 1 nếu $sigmoid(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n) \geq \theta$ với $\theta \in (0,1)$ là một ngưỡng (threathold value) và

$$sigmoid(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n) = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}} \quad (14)$$



Hình 4.7. Minh họa hồi quy logistic và hồi quy tuyến tính

- Ưu điểm của mô hình hồi quy logistic (Logistic-regression, 2024):

+ Hồi quy logistic là một trong những thuật toán học máy đơn giản nhất. Nó rất dễ dàng để thực hiện và giải thích.

+ Không có giả định cơ bản nào về sự phân bố của các đặc trưng.

+ Nó có thể được sử dụng cho cả trường hợp nhị thức và đa thức (hình 4.7)

+ Hồi quy logistic cung cấp cách tiếp cận dựa trên xác suất để dự đoán nhãn của biến mục tiêu.

+ Nó hoạt động tốt khi dữ liệu có thể phân tách tuyến tính.

+ Nó không bị ảnh hưởng bởi việc khớp quá mức đối với tập dữ liệu chiều thấp, tức là số lượng yếu tố dự đoán rất nhỏ so với cỡ mẫu.

+ Huấn luyện mô hình hồi quy logistic nhanh hơn nhiều so với các mô hình tương đối phức tạp.

+ Hồi quy logistic đưa ra các xác suất được hiệu chỉnh tốt cùng với kết quả phân loại. Điều này rất hữu ích để suy ra tính chính xác của dự đoán.

- Nhược điểm của mô hình hồi quy logistic (Logistic-regression, 2024)

+ Nếu số lượng quan sát nhỏ hơn số lượng thuộc tính thì không nên sử dụng Hồi quy logistic, nếu không có thể dẫn đến tình trạng quá mức đào tạo (overfitting).

+ Hạn chế chính của hồi quy logistic là giả định về tính tuyến tính giữa biến phụ thuộc và biến độc lập.

+ Đối với các dữ liệu có các mối quan hệ phức tạp thì mô hình hồi quy

logistic tỏ ra kém hiệu quả. Các thuật toán mạnh mẽ và nhỏ gọn hơn như Mạng thần kinh có thể dễ dàng vượt trội hơn thuật toán này.

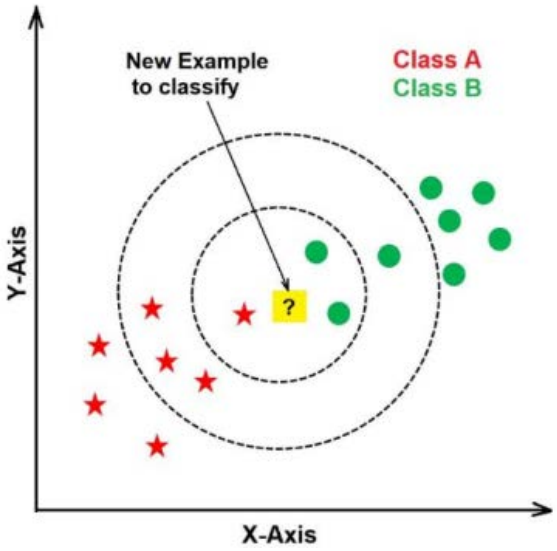
+ Trong hồi quy tuyến tính, các biến độc lập và phụ thuộc có liên quan tuyến tính. Nhưng hồi quy logistic cần các biến độc lập có liên quan tuyến tính với các tỷ lệ chênh lệch (odds).

4.2.3.5. Mô hình k-láng giềng gần nhất (k-NN)

K-Nearest Neighbor là một phương pháp học có giám sát không đưa ra giả định nào về phân phối xác suất của các vectơ đặc trưng. Đối với các vấn đề phân loại, các cá thể huấn luyện được lưu trữ trong không gian đặc trưng với nhãn lớp của chúng. Đối với một trường hợp thử nghiệm, khoảng cách từ nó đến tất cả các điểm huấn luyện trong tập dữ liệu được tính toán và lưu trữ theo thứ tự đã sắp xếp. Sau đó, nó được phân loại theo đa số phiếu bầu của các hàng xóm của nó, với trường hợp được gán cho lớp phổ biến nhất trong số các hàng xóm gần nhất của nó, trong đó k là số nguyên dương được truyền dưới dạng tham số. Thuật toán k-láng giềng gần nhất (k-NN: k-nearest neighbor) là một trong những thuật toán cần thiết và hiệu quả nhất để phân tách dữ liệu, có khả năng trở thành lựa chọn chính để triển khai, đặc biệt khi dữ liệu nhất định khá mơ hồ. Thuật toán này được phát minh vào năm 1951 bởi Evelyn Fix & Joseph Hodges để kiểm tra phân biệt khi việc quyết định mật độ xác suất bằng ước lượng tham số là tương đối khó khăn (Bansal & cs., 2022).

Trong phân lớp nhị phân, hai lớp có nhãn 0 hoặc nhãn 1, một điểm dữ liệu riêng biệt đã được chỉ định để xác định là thuộc về nhãn 0 hoặc nhãn 1. Trong trường hợp này, thuật toán k-NN có thể dễ dàng giúp người phân tích trong quy trình phân loại điểm mới khác với tập dữ liệu trên cơ sở chỉ số tương tự hoặc khoảng cách của điểm với cả hai trường hợp hiện có. Hình 4.8 cho thấy cách phân loại các đối tượng k-NN bằng cách sử dụng mô hình học tập có tính đến khoảng cách gần nhất với các đối tượng khác; nếu một đối tượng ở gần một đối tượng đang được phân loại thì nó được coi là thành viên của đối tượng gần nhất. Trong thuật toán k-NN, dữ liệu được nhập vào không gian đa chiều và mỗi chiều được liên kết với các thuộc tính dữ liệu khác nhau. Giá trị chính xác của thuật toán k-NN phụ thuộc nhiều vào nhiều yếu tố, bao gồm lựa chọn k, phương pháp đo khoảng cách, tiền xử lý dữ liệu và phân phối của dữ liệu. Thuật toán k-NN đưa ra giả định rằng những thứ có thể so sánh được sẽ xuất hiện ở gần hoặc ở các

vùng lân cận. Điều này ngụ ý rằng dữ liệu có thể so sánh được sẽ được đặt cạnh nhau. Thuật toán k-NN phân loại dữ liệu hoặc trường hợp mới bằng cách sử dụng tất cả dữ liệu có sẵn và các hàm tương tự hoặc khoảng cách. Lớp chứa phần lớn dữ liệu liền kề sau đó sẽ được cung cấp dữ liệu mới.



Hình 4.8. Mô phỏng phân lớp k_NN

- Ưu điểm của thuật toán k-NN

+ Không có thời gian huấn luyện: Nó không học được gì trong thời gian huấn luyện. Nó không tạo ra bất kỳ chức năng phân biệt nào từ dữ liệu huấn luyện. Nói cách khác, không có thời gian huấn luyện cho nó. Nó lưu trữ tập dữ liệu huấn luyện và chỉ học từ nó tại thời điểm đưa ra dự đoán thời gian thực. Điều này làm cho thuật toán k-NN nhanh hơn nhiều so với các thuật toán khác yêu cầu huấn luyện, ví dụ: SVM, Hồi quy tuyến tính, v.v. Vì thuật toán k-NN không yêu cầu huấn luyện trước khi đưa ra dự đoán, dữ liệu mới có thể được thêm liền mạch sẽ không ảnh hưởng đến độ chính xác của thuật toán

+ k-NN rất dễ thực hiện. Chỉ có hai tham số được yêu cầu để triển khai KNN, tức là giá trị của K và hàm khoảng cách (ví dụ: Euclidean hoặc Manhattan, v.v.)

- Nhược điểm của thuật toán k-NN

+ Không hoạt động tốt với tập dữ liệu lớn: Trong tập dữ liệu lớn, chi phí tính toán khoảng cách giữa điểm mới và mỗi điểm hiện có là rất lớn, điều này làm giảm hiệu suất của thuật toán.

+ Không hoạt động tốt với các dữ liệu có số chiều lớn: Thuật toán k-NN không hoạt động tốt với dữ liệu nhiều chiều vì với số lượng kích thước lớn, thuật toán sẽ trở nên khó khăn trong việc tính toán khoảng cách trong mỗi chiều.

+ Cần chuẩn hóa đặc trưng: Chúng ta cần thực hiện việc chuẩn hóa đặc trưng trước khi áp dụng thuật toán k-NN cho bất kỳ tập dữ liệu nào. Nếu chúng tôi không làm như vậy, k-NN có thể tạo ra các dự đoán sai.

+ Nhạy cảm với dữ liệu nhiễu, thiếu giá trị và giá trị ngoại lệ: k-NN nhạy cảm với nhiễu trong tập dữ liệu. Chúng ta cần đưa ra các giá trị bị thiếu theo cách thủ công và loại bỏ các giá trị ngoại lệ.

Tóm lại: Các thuật toán trên đều có những ưu điểm và nhược điểm riêng (xem Bảng 4.1). Do đặc điểm của các file âm thanh ong thu được sau khi thực hiện trích xuất đặc trưng nó có dạng số và có xu hướng tách thành các lớp thể hiện sự chia đàn và không chia đàn rất rõ (Xem hình phổ Hình 4.10, Hình 4.11 và Hình 4.13). Do đó việc thực hiện các phương pháp học cho mô hình này ở trong luận văn là cần thiết để thấy được ưu điểm và nhược điểm của các mô hình học máy ứng dụng cho bài toán này là cần thiết và từ đó đưa ra đề xuất mô hình phù hợp nhất cho bài toán.

Bảng 4.1. Ưu điểm và nhược điểm chính của các mô hình học máy

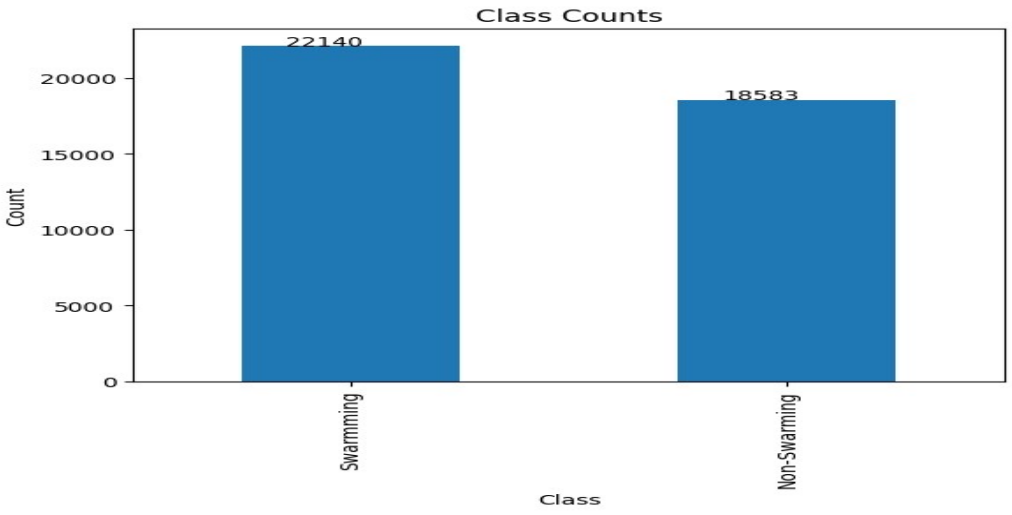
Mô hình học máy	Ưu điểm chính	Nhược điểm chính
Cây quyết định	+ Không hạn chế kiểu dữ liệu. + Không có yêu cầu giả định nào về phân bố của dữ liệu.	+ Tính không ổn định + Vấn đề quá khớp hay quá mức đào tạo (overfitting).
Rừng ngẫu nhiên	+ Giải quyết vấn đề quá khớp với dữ liệu vì đầu ra dựa trên biểu quyết đa số hoặc lấy trung bình. + Có tính ổn định cao.	Thời gian huấn luyện nhiều hơn so với các mô hình khác do tính phức tạp của nó.
Mô hình máy véc tơ hỗ trợ	+ Tính ổn định: Một thay đổi nhỏ đối với dữ liệu không ảnh hưởng lớn đến siêu phẳng. + Thuật toán SVM phù hợp nhất khi có sự phân chia rõ ràng giữa các lớp.	+ Thuật toán SVM không thể hoạt động phù hợp với các tập dữ liệu có kích thước khổng lồ + SVM không hoạt động hiệu quả trong trường hợp tập dữ liệu chứa lượng nhiễu lớn.
Mô hình hồi quy logistic	+ Dễ dàng để thực hiện và giải thích	+ Nếu số lượng quan sát nhỏ hơn số lượng thuộc tính

Mô hình học máy	Ưu điểm chính	Nhược điểm chính
	+ Hoạt động tốt khi dữ liệu có thể phân tách tuyến tính. + Huấn luyện mô hình hồi quy logistic nhanh hơn nhiều so với các mô hình tương đối phức tạp.	thì không nên sử dụng Hồi quy logistic vì có thể dẫn đến tình trạng trang bị quá mức (overfitting). + Đối với các dữ liệu có các mối quan hệ phức tạp thì mô hình hồi quy logistic tỏ ra kém hiệu quả.
Mô hình k-láng giềng gần nhất	+ Thuật toán k-NN nhanh hơn nhiều so với các thuật toán khác. + k-NN rất dễ thực hiện.	+ Không hoạt động tốt với tập dữ liệu lớn + Nhạy cảm với dữ liệu nhiễu.

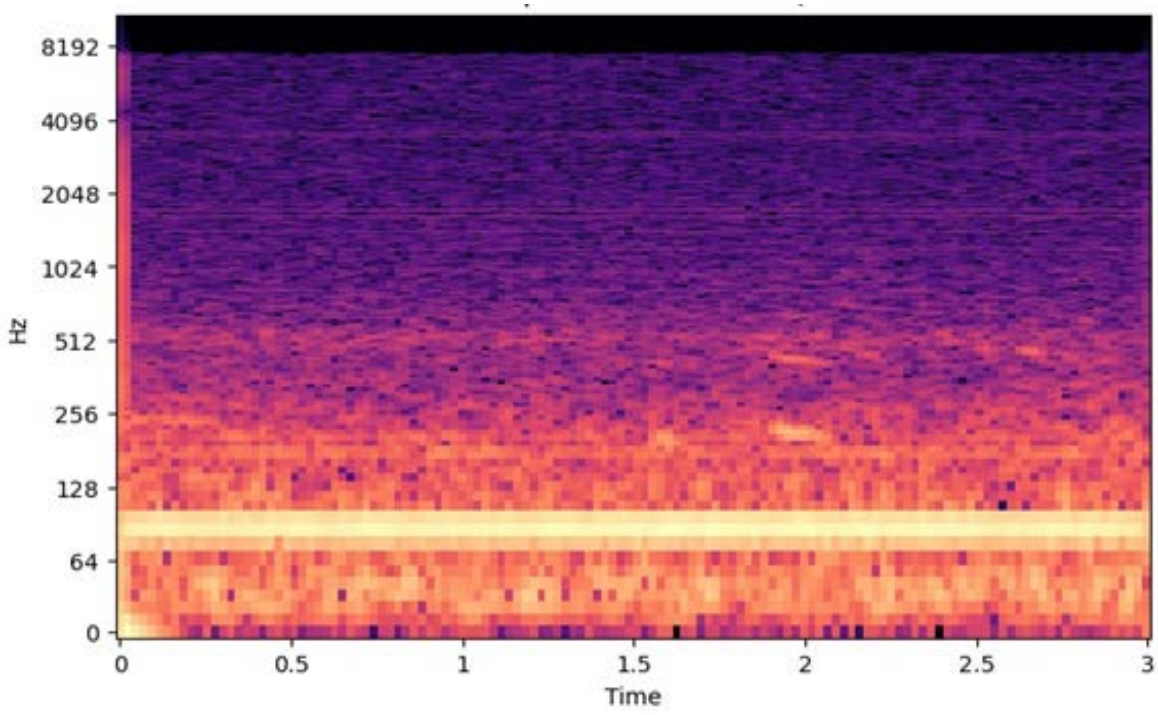
4.3. ỨNG DỤNG CÁC KỸ THUẬT HỌC MÁY CHO BÀI TOÁN NHẬN DẠNG ĐỐI TƯỢNG DỰA TRÊN ÂM THANH

4.3.1. Mô tả dữ liệu và tiền xử lí dữ liệu

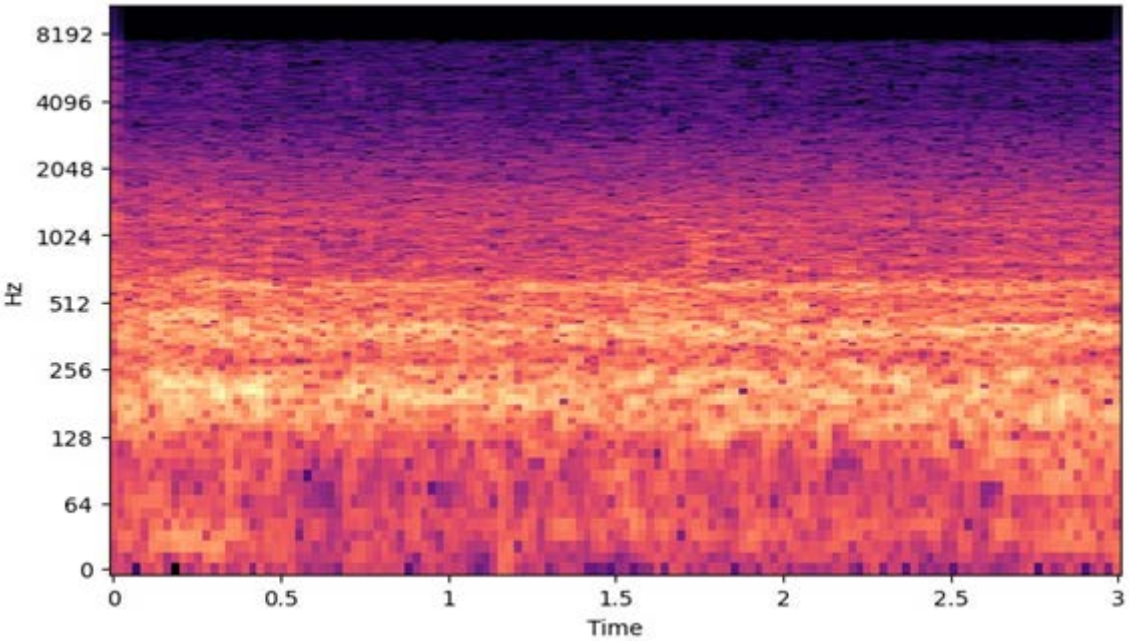
Tập dữ liệu âm thanh ong trong đề án này được lấy từ Trung tâm nghiên cứu ong và nuôi ong nhiệt đới, Học viện Nông nghiệp Việt Nam. Dữ liệu có hai bộ: bộ dữ liệu ong chia đàn (Swarmming) và dữ liệu ong ở trạng thái bình thường (Non-Swarmming). Trong đó, dữ liệu ong chia đàn có 22140 file.wav và bộ dữ liệu dữ liệu ong bình thường có 18583 file.wav (hình 4.9). Khi mã hóa, bộ dữ liệu ong chia đàn đánh nhãn là 1 và bộ dữ liệu ong ở trạng thái bình thường gán nhãn là 0. Hình ảnh 4.10 và hình ảnh 4.11, tương ứng là phổ của một mẫu âm thanh ong ở trạng thái bình thường và hình ảnh phổ của một mẫu âm thanh ong chia đàn. Đây là bài toán phân lớp nhị phân.



Hình 4.9. Số lượng file âm thanh ong được sử dụng để phân lớp

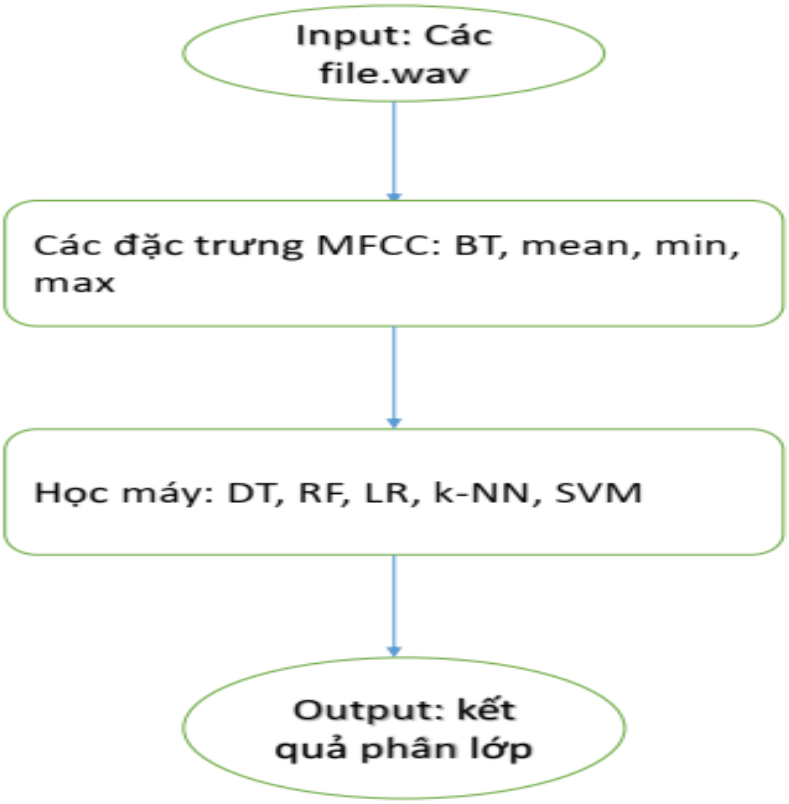


Hình 4.10. Ảnh phổ của một mẫu âm thanh ong ở trạng thái bình thường



Hình 4.11. Ảnh phổ của một mẫu âm thanh ong chia đàn

Việc thực hiện thực hiện nhận dạng đối tượng dựa trên âm thanh, ở đây, bao gồm hai giai đoạn chính như sau (Hình 4.12):



Hình 4.12. Sơ đồ thực hiện nhận dạng đối tượng dựa trên âm thanh

Giai đoạn 1: Thực hiện trích xuất đặc trưng: Ở đây với các dữ liệu đầu vào là các tệp âm thanh có nhãn (dạng file.wav). Dùng phương pháp trích chọn đặc trưng MFCC, ta thu được một file.csv đầu ra là một ma trận có các hàng là các đối tượng và các cột là các giá trị đặc trưng MFCC và cột cuối cùng là giá trị phân lớp.

Giai đoạn 2: Sử dụng các mô hình học máy để phân lớp. Đầu vào của giai đoạn 2 là file.csv (là kết quả của giai đoạn 1). Sử dụng năm mô hình học máy truyền thống (Cây quyết định, Rừng ngẫu nhiên, k-láng giềng, hồi quy logistic, máy véc tơ hỗ trợ) ta thu được đầu ra của giai đoạn 2 là kết quả phân lớp nhận dạng các file âm thanh.

4.3.2. Trích xuất đặc trưng MFCC

Theo phân tích trong phần 4.1. về trích chọn đặc trưng MFCC của âm thanh ta thu được 39 đặc trưng của âm thanh. Hình 4.13 là phân bố dữ liệu của một mẫu 39 MFCC.

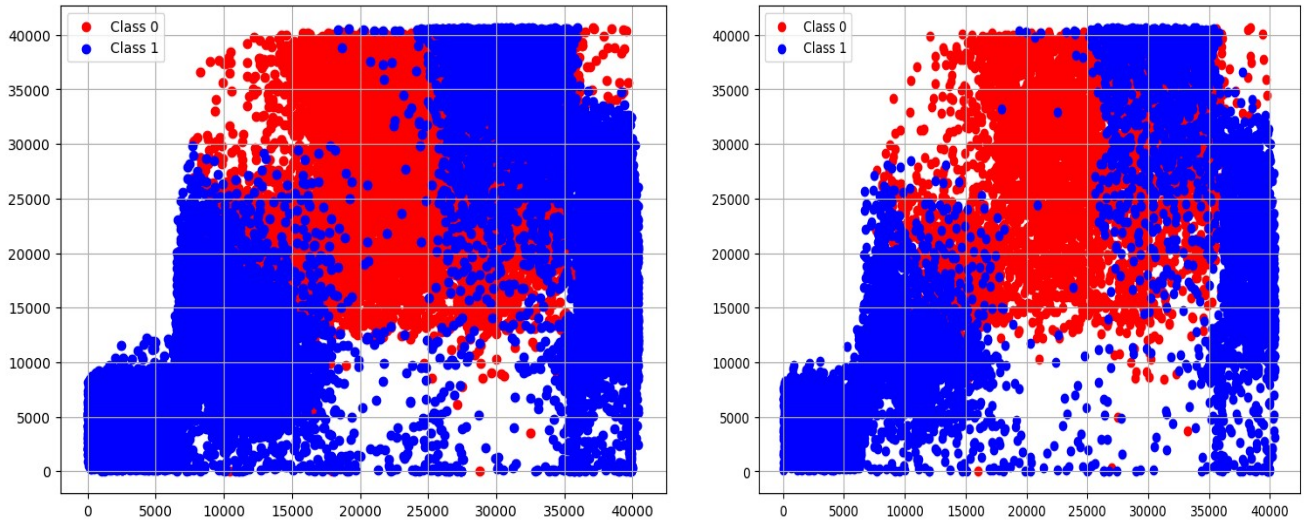
Ngoài ra thì lựa chọn trích chọn đặc trưng của mỗi thuộc tính MFCC ta có thể chọn theo định hướng chẳng hạn thường là các giá trị đặc trưng: giá trị trung bình, độ lệch chuẩn, phương sai, giá trị lớn nhất, giá trị nhỏ nhất của mỗi véc tơ giá trị số của mỗi thuộc tính MFCC.

4.3.3. Cài đặt thử nghiệm

4.3.3.1. Các mô hình học máy có giám sát

Trong nghiên cứu này chúng tôi cài đặt thử nghiệm năm mô hình phân lớp (học có giám sát) là cây quyết định, rừng ngẫu nhiên, máy véc tơ hỗ trợ, và mô hình mạng nơ-ron tích chập. Tham số của các mô hình như sau:

- *Mô hình cây quyết định*: DecisionTreeClassifier(random_state=41)
- *Mô hình rừng ngẫu nhiên*: RandomForestClassifier(random_state=41)
- Mô hình k-NN: KNeighborsClassifier(n_eighnors = k) với hai trường hợp k = 5 và k =10.
- *Mô hình hồi quy logistic*: LogisticRegression()
- *Mô hình máy véc tơ hỗ trợ*: SVC(kernel='linear')
- *Tỉ lệ dữ liệu chia huấn luyện và kiểm tra*: 70% huấn luyện, 30% thử nghiệm.



Tập huấn luyện Tập kiểm tra

Hình 4.13 Phân bố dữ liệu của một mẫu 39 MFCC

4.3.3.2. Các chỉ số phân lớp

Đánh giá mô hình là một nhiệm vụ quan trọng trong phân loại và một số tham số đã được thể hiện ở khía cạnh này. Tiêu chí đánh giá được sử dụng thường xuyên nhất trong dạng nghiên cứu này là độ chính xác phân lớp (accuracy), khả năng thu hồi (recall), độ chính xác (pricision), và điểm F (F-socre). Các chỉ số này được xác định qua các chỉ số của ma trận nhầm lẫn như bảng sau (Bảng 4.2):

Bảng 4.2. Ma trận nhầm lẫn đối với phân lớp nhị phân

	Giá trị thực (1)	**Giá trị thực (0)**
Dự đoán (1)	TP	FP
Dự đoán (0)	FN	TN

Trong đó:

TP (dương tính đúng, true positive): Dự đoán là đối tượng thuộc lớp 1 và thực sự thì đối tượng dự đoán thuộc lớp 1.

TN (âm tính đúng, true negative): Dự đoán đối tượng thuộc lớp 0 và thực sự thì đối tượng dự đoán thuộc lớp 0.

FP (dương tính giả, false positive): Dự đoán đối tượng thuộc lớp 1 nhưng thực sự thì đối tượng thuộc lớp 0.

FN (âm tính giả, false negative): Dự đoán đối tượng thuộc lớp 0, nhưng thực sự thì đối tượng thuộc lớp 1.

Các chỉ số đánh giá hiệu quả của bài toán phân lớp được sử dụng dựa trên ma trận nhầm lẫn thường là các chỉ số sau:

+ Độ chính xác phân lớp (accuracy): xác định mức độ chính xác của kết quả phân lớp cho cả phân lớp 0 và phân lớp 1.

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \quad (14)$$

+ Độ chính xác của phép đo (precision): xác định tỉ số phân lớp đúng trên tổng số dự đoán là đúng.

$$Pre = \frac{TP}{TP+FP} \quad (15)$$

+ Tỉ lệ thu hồi: cho thấy tỷ lệ các đối tượng thực sự được phân loại đúng trên tập các đối tượng có giá trị đúng cần phân lớp.

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

Nếu Recall gần bằng 1, điều đó có nghĩa là mô hình dự đoán có khả năng phát hiện tất cả hoặc gần như tất cả các trường hợp đúng. Tuy nhiên, một Recall thấp có thể chỉ ra rằng mô hình dự đoán bỏ sót nhiều trường hợp đúng và có thể không phát hiện được một số trường hợp quan trọng.

+ Điểm F (F-score): là trung bình điều hòa giữa độ chính xấc của phép đo và tỉ lệ thu hồi theo công thức

$$F - score = 2 * \frac{Pre*Recall}{Pre+Recall} \quad (17)$$

Một mô hình có F-score cao cho thấy mô hình có Precision và Recall tốt, tức là nó có khả năng dự đoán chính xác các điểm dữ liệu thuộc cả hai lớp.

4.3.3.3. Các kịch bản và kết thử nghiệm rút trích đặc trưng MFCC

Trong nghiên cứu này, chúng tôi thực hiện rút trích đặc trưng MFCC của các tệp âm thanh dạng file.wav bằng thư viện librosa trên python với độ dài của một frame áp dụng cho biến đổi Fourier nhanh n_ftt =1024 và bước nhảy giữa các frame là hop_length = 512. Vì độ dài của các tệp âm thanh (đã được cắt thành đoạn âm thanh nhỏ lưu dưới dạng file.wav) ở đây là bằng nhau, nên số lượng khung trên mỗi tệp âm thanh được tính ra bằng 130 khung (bằng chiều dài mỗi file.wav chia cho bước nhảy). Các kết quả này được hiện bằng ngôn ngữ Python qua môi trường Jupyter Notebook trên máy PC có cấu hình: Core i5, ram 8G.

Trong kết quả chạy với 39 đặc trưng MFCC (như đã nói đến trong phần 4.1, ta kí hiệu là MFCC_BT), chúng tôi có kết quả phân lớp như Bảng 4.3. Trong các bảng kết quả, giá trị in đậm là giá trị tốt nhất.

Từ bảng 4.3 ta thấy, với 39 đặc trưng MFCC_BT (lấy mặc định theo thư viện của Librosa là phần tử đầu tiên của khung hình thứ nhất trên mỗi file âm thanh, có thời gian thực hiện khoảng 1135s) mô hình rừng ngẫu nhiên cho kết quả phân lớp cao nhất trong năm mô hình được xem xét đến ở đây với độ chính xác phân lớp đạt 97.23% nhưng chiếm thời gian thực hiện khá cao 394 giây (s); tiếp theo là các mô hình cây quyết định có độ chính xác khoảng 94.64% với thời gian chạy khoảng

4.19s; mô hình máy véc tơ hỗ trợ có độ chính xác 94.25% có thời gian thực hiên 434.97s; mô hình hồi quy logistis cho độ chính xác 94.22% có thời gian thực hiện 0.44s. Hai trường hợp chạy mô hình k-NN (k=5 và k=10) cho độ chính xác phân lớp lần lượt là 92.31% và 91.35% với thời gian thực hiện lần lượt là 1.52s và 1.47s.

Bảng 4.3. Kết quả phân lớp của các mô hình với 39 đặc trưng MFCC_BT

	Acc	**Thời gian chạy (giây-s)**
DT	0.9464	4.19
k-NN (10)	0.9135	1.47
k-NN (5)	0.9231	1.53
LR	0.9422	**0.44**
RF	**0.9723**	394
SVM	0.9425	434.97

Trong trường hợp trên ta thấy rằng về độ chính xác thì mô hình RF cho kết quả vượt trội so với các mô hình còn lại. Nhưng về mặt thời gian tính toán thì mô hình RF chiếm thời gian chỉ ngắn hơn mô hình SVM, trong khi đó thời gian thực hiện của RF là lâu hơn so với các mô hình còn lại (xem bảng 4.2).

Chúng ta có thể cải thiện độ chính xác này với cách thực hiện như sau: Tại mỗi khung mỗi đặc trưng MFCC sẽ nhận giá trị là một số thực. Vì vậy với 130 khung trên một tệp âm thanh thì ta nhận được 130 giá trị số thực (tương ứng với từng khung) của một đặc trưng MFCC. Ở đây ta có thể coi mỗi MFCC là một véc tơ M có số chiều bằng số khung này (ở đây là véc tơ 130 chiều, tương ứng với 130 khung hình). Từ đây ta có thể lấy giá trị nhỏ nhất (min), giá trị lớn nhất (max) và giá trị trung bình (mean) dựa trên véc tơ M này của mỗi MFCC làm đại diện cho MFCC tại từng tệp âm thanh (file.wav). Do đó với 39 đặc trưng MFCCs chúng ta thực hiện bốn trường hợp lấy đại diện cho mỗi đặc trưng MFCC là: lấy phần tử đầu tiên của M (đây chính là trường hợp mặc định khi xác định MFCC trong thư viện librosa cho tập các tệp âm thanh), lấy phần tử nhỏ nhất (min) trên véc tơ M, lấy phần tử lớn nhất (max) của M, lấy giá trị trung bình (mean) của véc tơ M. Điều này được chỉ rõ hơn trong phần phụ lục.

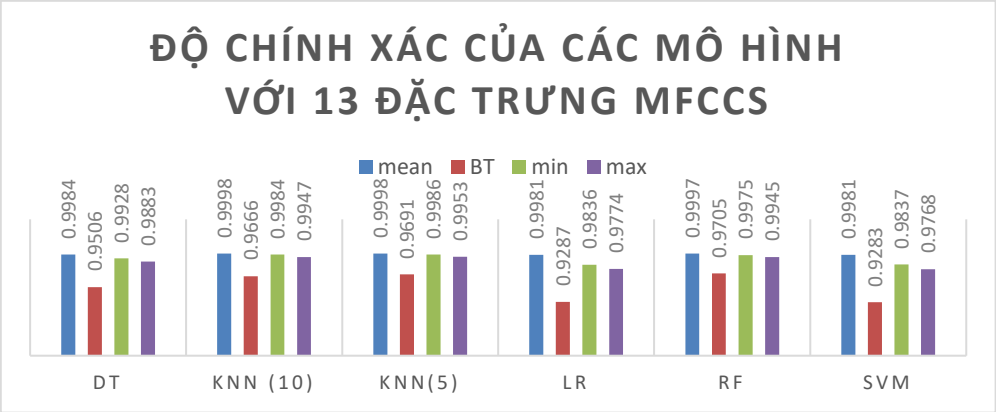
Ngoài ra nghiên cứu này cũng xét thêm các kịch bản là chỉ lấy 13 đặc trưng MFCC đầu tiên của các file âm thanh; kịch bản nữa là lấy 13 đặc trưng đầu tiên và 13 đạo hàm cấp một của chúng là 26 đặc trưng MFCC của các tệp âm thanh.

Do đó đề án tiến hành thử nghiệm ba kịch bản khác nhau:

Kịch bản 1: Thực hiện chương trình với 13 đặc trưng âm thanh đầu tiên. Thời gian thực hiện để trích xuất đặc trưng theo cách thức lấy phần tử đại diện trên mỗi đặc trưng MFCC là 801.76s, 1002s, 616.45s và 643s cho các đại diện tương ứng là giá trị trung bình (mean), giá trị mặc định (BT), giá trị nhỏ nhất (min), và giá trị lớn nhất (max).

Bảng 4.4. Độ chính xác của các mô hình khi chạy với kịch bản 13 đặc trưng MFCC

	mean	**BT**	**min**	**max**
DT	0.9984	0.9506	0.9928	0.9883
k-NN (10)	**0.9998**	0.9666	0.9984	0.9947
k-NN (5)	**0.9998**	0.9691	0.9986	0.9953
LR	0.9981	0.9287	0.9836	0.9774
RF	0.9997	0.9705	0.9975	0.9945
SVM	0.9981	0.9283	0.9837	0.9768



Hình 4.14. Độ chính xác của các mô hình với 13 đặc trưng MFCCs

Bảng 4.5. Thời gian thực hiện của các mô hình với 13 đặc trưng MFCC (tính theo giây)

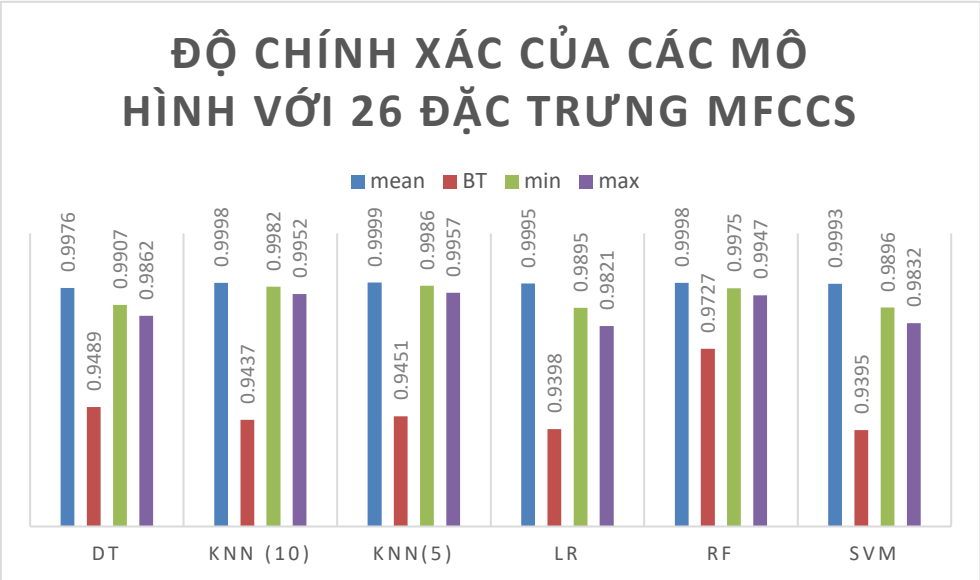
	mean	**BT**	**min**	**max**
DT	2.62	5.95	3.55	3.45
k-NN (10)	4.16	7	5.77	5.79
k-NN (5)	5.45	7	5.34	3.14
LR	1.08	1.28	0.98	**0.59**
RF	387.6	633	412.46	261.46
SVM	13.54	1127	142	102.25

Trong kịch bản 1 này thì mô hình k-láng giềng gần nhất (với k =5) cho kết quả độ chính xác cao nhất trên cả sau mô hình thực hiện và thời gian thực hiện cũng gần như tốt nhất trên tất cả các trường hợp (chỉ trừ trường hợp lấy đại diện trung bình (mean)) (Bảng 4.4 và Hình 4.14). Mô hình hồi quy Logistic có thời gian chạy nhanh nhất, nhưng cũng là mô hình có độ chính xác phân lớp chênh lệch nhiều nhất trong các trường hợp lấy phần tử đại diện ở đây (thấp nhất ở lấy đại diện mặc định và cao nhất với lấy đại diện trung bình) (Bảng 4.5).

Kịch bản 2: Thực hiện chương trình với 26 đặc trưng âm thanh đầu tiên. Thời gian thực hiện để trích xuất đặc trưng theo cách thức lấy phần tử đại diện trên mỗi đặc trưng MFCC là 303.56s, 450.36s, 397.86s và 365.07s cho các đại diện tương ứng là giá trị trung bình (mean), giá trị mặc định (BT), giá trị nhỏ nhất (min), và giá trị lớn nhất (max).

Bảng 4.6. Độ chính xác của các mô hình khi chạy với kịch bản 26 đặc trưng MFCC

	mean	**BT**	**min**	**max**
DT	0.9976	0.9489	0.9907	0.9862
k-NN (10)	0.9998	0.9437	0.9982	0.9952
k-NN (5)	**0.9999**	0.9451	0.9986	0.9957
LR	0.9995	0.9398	0.9895	0.9821
RF	0.9998	0.9727	0.9975	0.9947
SVM	0.9993	0.9395	0.9896	0.9832



Hình 4.15. Độ chính xác của các mô hình với 26 đặc trưng MFCCs

Trong kịch bản 2 này ta nhận thấy về độ chính xác không có mô hình nào chiếm ưu thế hoàn toàn (cao hơn) trong các trường hợp lấy đại diện trên mỗi đặc trưng MFCCs (Bảng 4.6 và Hình 4.15). Trong các trường hợp lấy đại diện thì lấy đại diện mặc định vẫn có độ chính xác kém nhất trên tất cả mô hình và thời gian chạy trường hợp này cũng là cao nhất trên tất cả các mô hình so với các trường hợp khác, đặc biệt trường hợp này mô hình học máy véc tơ hỗ trợ còn có độ chính xác thấp hơn nhiều so với các trường hợp khác (Bảng 4.7).

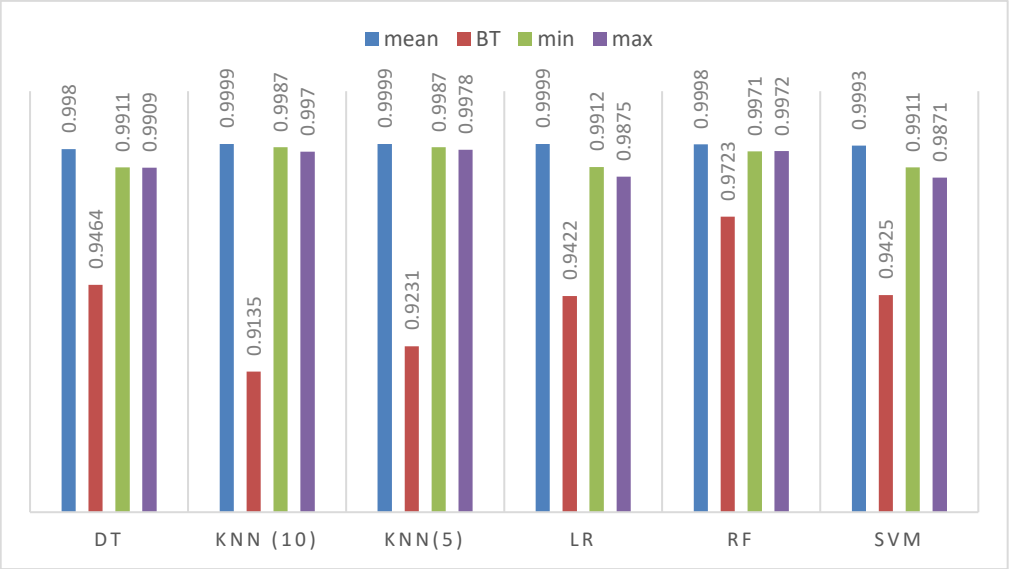
Bảng 4.7. Thời gian thực hiện của các mô hình với 26 đặc trưng MFCC (tính theo giây)

	mean	**BT**	**min**	**max**
DT	1.77	3	2.38	2.76
k-NN (10)	1.37	1.44	1.34	2.23
k-NN (5)	1.33	1.39	1.36	1.34
LR	0.41	0.41	**0.37**	**0.37**
RF	179.47	310.32	252.9	262.9
SVM	1.47	372.28	27.79	78.2

Kịch bản 3: Thực hiện chương trình với 39 đặc trưng âm thanh đầu tiên. Thời gian thực hiện để trích xuất đặc trưng theo cách thức lấy phần tử đại diện trên mỗi đặc trưng MFCC là 930s, 1135s, 872.4s và 641s cho các đại diện tưng ứng là giá trị trung bình (mean), giá trị mặc định (BT), giá trị nhỏ nhất (min), và giá trị lớn nhất (max).

Bảng 4.8. Độ chính xác của các mô hình khi chạy với kịch bản 39 đặc trưng MFCC

	mean	**BT**	**min**	**max**
DT	0.9980	0.9464	0.9911	0.9909
k-NN (10)	**0.9999**	0.9135	0.9987	0.997
k-NN (5)	**0.9999**	0.9231	0.9987	0.9978
LR	**0.9999**	0.9422	0.9912	0.9875
RF	0.9998	0.9723	0.9971	0.9972
SVM	0.9993	0.9425	0.9911	0.9871



Hình 4.16. Độ chính xác của các mô hình với 39 đặc trưng MFCCs

Bảng 4.9. Thời gian thực hiện của các mô hình với 39 đặc trưng MFCC (tính theo giây)

	mean	BT	min	max
DT	2.75	4.19	4.16	5.86
k-NN (10)	1.45	1.47	1.36	2.23
k-NN (5)	1.5	1.53	1.26	2.52
LR	**0.39**	0.44	0.43	0.77
RF	223.33	394	325.41	351.55
SVM	0.74	434.97	18.41	39.72

Trong kịch bản 3 này ta nhận thấy tình hình cũng gần giống với kịch bản 2: về độ chính xác không có mô hình nào chiếm ưu thế hoàn toàn (cao hơn) trong các trường hợp lấy đại diện trên mỗi đặc trưng MFCCs (Bảng 4.8 và Hình 4.16). Trong các trường hợp lấy đại diện thì lấy đại diện mặc định vẫn có độ chính xác kém nhất trên tất cả mô hình và thời gian chạy trường hợp này cũng là cao nhất trên tất cả các mô hình so với các trường hợp khác, đặc biệt trường hợp này mô hình học máy véc tơ hỗ trợ còn có độ chính xác thấp hơn nhiều so với các trường hợp khác (Bảng 4.9).

4.4. THẢO LUẬN

Trong số các kịch bản chạy với số đặc trưng lần lượt là 13MFCCs, 26MFCCs và 39 MFCC với bốn trường hợp lấy đại diện trong các kịch bản lần lượt là: lấy phần tử trung bình (mean), lấy mặc định (phần tử đầu tiên), phần tử

có giá trị nhỏ nhất (min) và phần tử lớn nhất (max) của véc tơ *M* có các tọa độ được tạo thành từ các giá trị tương ứng của mỗi đặc trưng MFCC trên các khung hình (frame) chạy trên từng tệp âm thanh ta nhận thấy trường hợp lấy giá trị trung bình (mean) có độ chính xác cao nhất trong các kịch bản (đạt gần như tuyệt đối trên 99.7%). Trường hợp lấy giá trị mặc định có độ chính xác không ổn định nhất trên các mô hình và trong các kịch bản (từ 91% đến 97%). Mô hình rừng ngẫu nhiên cho độ chính xác ổn định nhất trong tất cả các kịch bản và trong tất cả các trường hợp (từ trên 97% đến trên 99%), trong khi đó các mô hình khác có độ chính xác thay đổi rất nhiều từ 94% đến trên 99% (mô hình cây quyết định và hồi quy logistic), và từ khoảng 92% đến trên 99% (với các mô hình còn lại). Về thời gian tính toán thì mô hình hồi quy logistic thời gian chạy nhanh nhất. Mô hình rừng ngẫu nhiên có thời gian chạy lâu nhất tiếp đến là mô hình máy véc tơ hỗ trợ (SVM). Xét trên toàn diện cả về thời gian và độ chính xác trong các kịch bản và các trường hợp thì không có mô hình nào chiếm ưu thế hoàn toàn theo nghĩa độ chính xác cao nhất, thời gian thực hiện ít nhất. Cũng từ những kết quả trên, chúng tôi đưa ra gợi ý dựa trên cả độ chính xác và thời gian thực hiện từ việc trích xuất các đặc trưng đến thực hiện trên các mô hình thì chúng tôi đề nghị lựa chọn mô hình hồi quy logistic với 26 đặc trưng MFCCs và lấy đại diện là phần tử trung bình.

- ***Ưu điểm của phương pháp đề xuất:***

+ Xây dựng được các kịch bản trích xuất đặc trưng MFCCs và các trường hợp lấy phần tử đại diện cho mỗi đặc trưng MFCC và thử nghiệm với bộ dữ liệu ở Trung tâm nghiên cứu ong và nuôi ong nhiệt đới, Học viện Nông nghiệp Việt Nam. Trong đó dữ liệu gồm hai nhãn: Nhãn dấu hiệu ong chia đàn (Swarming) và nhãn ong ở trạng thái bình thường (Non-Swarming).

+ Cài đặt thử nghiệm với một số mô hình học máy như: cây quyết định, rừng ngẫu nhiên, máy véc tơ hỗ trợ, k-láng giềng gần nhất (k=5, k=10) và mô hình hồi quy logistic.

+ Độ chính xác phân lớp khá cao với các trường hợp lấy đại diện cho mỗi đặc trưng MFCCs là phần tử mặc định (khoảng từ 92% đến 97%), trong khi đó các trường hợp còn lại là trên 98% đặc biệt là lấy đại diện là phần tử trung bình thì độ chính xác phân lớp đạt gần như 100% như đã phân tích ở phần 4.4.

+ Thời gian chạy nhanh: Với 40722 tệp âm thanh (file.wav), mỗi file có độ dài 3 giây. Mỗi lần thực hiện trích xuất đặc trưng MFCC như đã trình bày ở trên và chạy cả năm mô hình phân lớp chỉ chiếm 18-25 phút.

+ Chỉ ra được trường hợp tốt nhất là lấy giá trị trung bình làm đại diện cho mỗi đặc trưng MFCC và mô hình phân lớp phù hợp nhất (trên tất cả các kịch bản và các trường hợp) cho dữ liệu này là mô hình k-láng giềng gần nhất, bỏ qua thời gian chạy thì mô hình rừng nhẫu nhiên (Random Forest) cho độ chính xác ổn định nhất. Nếu ưu tiên cả về thời gian và độ chính xác thì chúng tôi đề nghị mô hình hồi quy logistic với 26 đặc trưng MFCCs và lấy đại diện là phần tử trung bình để phân lớp cho dữ liệu này.

- **Hạn chế của phương pháp đề xuất:**

+ Chưa thử nghiệm được với các đặc trưng cụ thể của các file âm thanh như: trọng tâm của phổ, độ rộng của phổ, năng lượng tín hiệu trung bình của file âm thanh,...

+ Chưa chạy được nhiều mô hình phân lớp khác như các mô hình học sâu: CNN, RNN, ANN.

+ Chưa thực hiện được chuyển âm thanh sang phổ hình ảnh để chạy các mô hình nơ-ron. Trong khi chạy các file ảnh mới là thế mạnh của các mạng nơ-ron.

PHẦN 5. KẾT LUẬN VÀ KIẾN NGHỊ

5.1. KẾT LUẬN

Đề án này đã tìm hiểu tổng quan về dữ liệu âm thanh, đặc trưng âm thanh, tìm hiểu về bài toán phân lớp. Đồng thời đề án cũng nghiên cứu các thuật toán học máy áp dụng cho bài toán phân lớp điển hình như cây quyết định, rừng ngẫu nhiên, máy học vector hỗ trợ hồi quy logistic, và k- láng giềng gần nhất.

Bên cạnh đó, đề án cũng tìm hiểu, nghiên cứu giải thuật lựa chọn đặc trưng MFCC. Sau đó tôi áp dụng vào bài toán nhận dạng đối tượng dựa trên âm thanh. Cụ thể đối tượng trong nghiên cứu này là dữ liệu âm thanh về chia đàn tự nhiên ở ong và tình trạng thiếu chúa của ong. Bộ dữ liệu âm thanh này được thu tại Trung tâm nghiên cứu ong và nuôi ong nhiệt đới, Học viện Nông nghiệp Việt Nam.

Ngoài ra, nghiên cứu này cũng xét các kịch bản là chỉ lấy 13 đặc trưng MFCC đầu tiên của các file âm thanh; kịch bản nữa là lấy 13 đặc trưng đầu tiên và 13 đạo hàm cấp một của chúng là 26 đặc trưng MFCC của các tệp âm thanh và kịch bản thứ 3 là lấy 39 đặc trưng âm thanh. Các kịch bản được thực hiện bốn trường hợp lấy đại diện cho mỗi đặc trưng MFCC là: lấy phần tử đầu tiên của M (đây chính là trường hợp mặc định khi xác định MFCC trong thư viện librosa cho tập các tệp âm thanh), lấy phần tử nhỏ nhất (min) trên véc tơ M, lấy phần tử lớn nhất (max) của M, lấy giá trị trung bình (mean) của véc tơ M.

Kết quả chỉ ra trường hợp lấy giá trị trung bình làm phần tử đại diện cho mỗi đặc trưng MFCC trên các file âm thanh cho kết quả tốt nhất. Độ chính xác đạt trên 99.7%. Chúng tôi đề nghị mô hình hồi quy logistic với 26 đặc trưng MFCCs và lấy đại diện là phần tử trung bình để phân lớp cho dữ liệu này. Kết quả này cũng là một sự gợi ý cho việc lấy đại diện cho mỗi đặc trưng MFCC cho trích chọn đặc trưng trong việc phân loại âm thanh, cụ thể ở đây là âm thanh ong.

5.2. KIẾN NGHỊ

- Như đã chỉ ra ở phần hạn chế của nghiên cứu. Trong tương lai, chúng tôi sẽ tìm hiểu thêm các phương pháp trích chọn đặc trưng khác như mô hình phân lớp khác như các mô hình học sâu: CNN, RNN, ANN và thực hiện thử nghiệm với các đặc trưng cụ thể của các file âm thanh như: trọng tâm của phổ, độ rộng của phổ, năng lượng tín hiệu trung bình của file âm thanh,...

- Tìm hiểu thêm việc thực hiện được chuyển âm thanh sang phổ hình ảnh để chạy các mô hình nơ-ron. Trong khi chạy các file ảnh mới là thế mạnh của các mạng nơ-ron.

TÀI LIỆU THAM KHẢO

Abdul, Z. K. & Al-Talabani, A. K. (2022). Mel Frequency Cepstral Coefficient and its applications: A Review. *IEEE Access*. V.10, pp. 122136-122158.

Amlathe, P. (2018). *Standard machine learning techniques in audio beehive monitoring: Classification of audio samples with logistic regression, K-nearest neighbor, random forest and support vector machine* (Doctoral dissertation, Utah State University).

Ashar, A., Bhatti, M. S. & Mushtaq, U. (2020). Speaker identification using a hybrid cnn-mfcc approach. In 2020 International Conference on Emerging Trends in Smart Technologies (ICETST) (pp. 1-4). IEEE.

Bansal, M., Goyal, A. & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long shortterm memory algorithms in machine learning. *Decision Analytics Journal. 3*. 100071.

Cao, J., Cao, M., Wang, J., Yin, C., Wang, D. & Vidal, P. P. (2019). Urban noise recognition with convolutional neural network. Multimedia Tools and Applications. 78: 29021-29041.

Cejrowski, T., Szymański, J., Mora, H. & Gil, D. (2018). Detection of the bee queen presence using sound analysis. In Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings. Part II 10: 297-306. Springer International Publishing.

Costa, V. G. & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. Artificial Intelligence Review. 56(5): 4765-4800.

Das, J. K., Chakrabarty, A. & Piran, M. J. (2022). Environmental sound classification using convolution neural networks with different integrated loss functions. Expert Systems. 39(5): e12804.

Dimitrijević, S. & Zogović, N. (2022). Machine Learning Advances in Beekeeping.

Dimitrios, K.I., Bellos, C.V., Stefanou, K.A., Stergios, G.S., Andrikos, I., Katsantas & T. and Kontogiannis, S., 2022. Performance Evaluation of Classification Algorithms to Detect Bee Swarming Events Using Sound. Signals. 3(4): 807-822.

Ferrari, S., Silva, M., Guarino, M. & Berckmans, D. (2008). Monitoring of swarming sounds in bee hives for early detection of the swarming period. Computers and electronics in agriculture 64(1): 72-77.

Gifford, M. & Bayrak, T. (2023). A predictive analytics model for forecasting outcomes

in the National Football League games using decision tree and logistic regression. Decision Analytics Journal, 8, 100296.

Hồ Thị Ngọc (2012). Nghiên cứu ứng dụng học bán giám sát (Luận văn thạc sĩ, Đại học Đà Nẵng).

Hoàng Thị Thanh Giang, Nguyễn Thị Thúy Hạnh & Nguyễn Trọng Kương(2021). Nhận dạng giọng chữ cái tiếng Việt sử dụng Deep BOLTZMANN machines. Tạp chí Khoa học Nông nghiệp Việt Nam 19(4): 435-442

Kanakala, R. & Reddy, K. (2023). Modelling a deep network using CNN and RNN for accident classification. Measurement: Sensors, 100794.

Kurzekar, P. K., Deshmukh, R. R., Waghmare, V. B. & Shrishrimal, P. P. (2014). A comparative study of feature extraction techniques for speech recognition system. International Journal of Innovative Research in Science, Engineering and Technology. 3(12): 18006-18016.

Lyons, R. (2001). *Understanding digital signal processing's frequency domain*. RF DESIGN. 24(11): 36-49.

Mã Trường Thành, Đỗ Thanh Nghị, Phạm Nguyên Khang & Châu Ngân Khánh (2015) Điều khiển robot Pioneer P3-DX bằng tiếng nói với đặc trưng MFCC và giải thuật Naïve Bayes Nearest Neighbor. Kỷ yếu Hội nghị Quốc gia lần thứ VIII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR); Hà Nội, ngày 9_10/7/2015.

Morgan JN & Sonquist JA (1963). Problems in the analysis of survey data, and a proposal. J Am Stat Assoc. 58(302): 415–434.

Nassif, A. B., Shahin, I., Hamsa, S., Nemmour, N. & Hirose, K. (2021). CASA-based speaker identification using cascaded GMM-CNN classifier in noisy and emotional talking conditions. Applied Soft Computing 103: 107141.

Nguyễn Chí Ngôn, Lê Thanh Tú, Lương Hoàng Vĩnh Thuận & Nguyễn Chánh Nghiệm (2022). Khảo sát kỹ thuật học sâu trên bài toán chẩn đoán hư hỏng động cơ điện dựa trên tiếng ồn vận hành. Tạp chí Khoa học Đại học cần Thơ. 58(1): 27-40.

Nguyễn Thế Cường, Nguyễn Thanh Vi & Trương Ngọc Hải (2023). Cơ sở toán và trích xuất đặc trưng âm thanh. Tạp chí Khoa học. 20(7): 1155.

Nguyễn Thị Thu (2022). Nghiên cứu và áp dụng các kĩ thuật học máy cho bài toán phát hiện đối tượng dựa trên dữ liệu âm thanh (Lunaj văn thạc sĩ, Học viện kĩ thuật quân sự).

Nguyễn Thị Tuyết Nhung. (2014). Ảnh hưởng của sự thay đổi vi khí hậu trong và ngoài tổ đến sức khỏe trứng của ong chúa và hàm lượng nước có trong mật ong tại

huyện Chợ Lách-tỉnh Bến Tre và huyện Kế Sách- tỉnh Sóc Trăng. Tạp chí Khoa học Đại học cần Thơ. (35): 54-64.

Nguyen, H. D., Nguyen, D. D., Vu, T.L., Van Hoang Nguyen, Hong Thai Pham, Thanh Ngoc Phan, Viet Long Nguyen & Thi Thu Hong Phan. (2020). Audio beehive monitoring based on IoT-AI techniques: a survey and perspective." Tap chi Khoa hoc Nong nghiep Viet Nam/Vietnam Journal of Agricultural Sciences 3, no. 1(2020): 530-540.

Phan, T. T. H., Nguyen, H. D. & Nguyen, D. D. (2022, August). Evaluation of Feature Extraction Methods for Bee Audio Classification. In International Conference on Intelligence of Things. 194-203. Cham: Springer International Publishing.

Phan, T. T. H., Nguyen-Doan, D., Nguyen-Huu, D., Nguyen-Van, H. & Pham-Hong, T. (2023). Investigation on new Mel frequency cepstral coefficients features and hyper-parameters tuning technique for bee sound recognition. *Soft Computing*. *27*(9): 5873-5892.

Ramsey, M.T., Bencsik, M., Newton, M.I., Reyes, M., Pioz, M., Crauser, D., Delso, N.S. & Le Conte, Y. (2020). The prediction of swarming in honeybee colonies using vibrational spectra. Scientific reports. 10(1): 9798.

Thái Thuận Thương (2021). Nhận dạng tiếng nói điều khiển với convolutional neural network (CNN). Tạp chí Khoa học Đại học cần Thơ. (57): 30-39.

Thiele, F., Windebank, A. J. & Siddiqui, A. M. (2023). Motivation for using data-driven algorithms in research: A review of machine learning solutions for image analysis of micrographs in neuroscience. Journal of Neuropathology & Experimental Neurology. 82(7): 595-610.

Trần Đăng Tú, Lê Thế Hùng, Trần Xuân Quý, Đoàn Huy Hiên, Phạm Trường Giang & Lưu Đinh Tùng (2022). Ứng dụng thuật toán học máy để dự báo khai thác cho đối tượng móng nứt mẻ, vòm trung tâm, mỏ Bạch Hổ. Tạp chí Dầu khí (9): 16-23.

Voudiotis, G., Kontogiannis, S. & Pikridas, C. (2021). Proposed smart monitoring system for the detection of bee swarming. Inventions. 6(4): 87.

Yaseliani, M. & Khedmati, M. (2023). Prediction of heart diseases using logistic regression and likelihood ratios. International Journal of Industrial Engineering & Production Research. 34(1): 1-15.

PHỤ LỤC

Rút trích MFCC của file 1:

```
Kích thước của ma trận MFCCs X số Frame trong file.wav thứ nhất
(0_Qeen3105_0.wav): (39, 130)
mfccs: [[-3.0162225e+02 -5.6770343e+02 -5.6078253e+02 ... -
5.6915576e+02
  -5.7236359e+02 -5.2062042e+02]
 [ 1.3371736e+02  1.2822083e+02  1.3943568e+02 ...  1.3112314e+02
   1.2978787e+02  1.2673549e+02]
 [-4.5993225e+01  1.4805899e+01  2.2534040e+01 ...  2.2432705e+01
   1.6646236e+01  1.3394249e+01]
 ...
 [-4.4086128e-01  1.6489916e+00  2.4335680e+00 ...  7.0713758e+00
  -1.2935636e+00 -2.6950016e+00]
 [-1.9544930e+00 -5.1152272e+00 -5.3093415e-01 ... -1.5076618e+00
  -5.8417177e+00 -7.5416050e+00]
 [ 5.2074140e-01  3.1297982e-01  2.6073744e+00 ...  1.4884791e+00
  -1.1487808e+00 -1.7133909e+00]]
-----------------
giá trị đầu tiên trên từng frame của file.wav thứ nhất
(0_Qeen3105_0.wav):
 [-301.62225     133.71736     -45.993225     33.25535      -
5.2263546
   33.25508     -22.286139     37.39849     -18.023365
9.641236
    6.962144     -5.6509867    16.229042     -2.1064825
6.6884823
    3.4454799     8.505168     -1.2364424     8.728474     -
0.7439035
    5.8531413     3.1563604    -0.53541195    7.5417595    -
2.5964096
    4.7266955     0.6031545     1.863605     -1.2594565
2.7496696
   -2.3011847    -0.32245374    1.1005534    -3.4589458
1.5190207
   -2.476883     -0.44086128   -1.954493      0.5207414 ]
```

MFCC_1	MFCC_2	MFCC_3	MFCC_4	MFCC_5	MFCC_6	MFCC_7	MFCC_8	MFCC_9	MFCC_10	MFCC_11	MFCC_12	MFCC_13	MFCC_14	MFCC_15
-301.622	133.7174	-45.9932	33.25535	-5.22635	33.25508	-22.2861	37.39849	-18.0234	9.641236	6.962144	-5.65099	16.22904	-2.10648	6.688482

---

giá trị trung bình trên từng frame:

```
 [-5.6668976e+02  1.3103934e+02  1.9087854e+01  7.1327225e+01
  8.6524992e+00  4.8331326e+01 -4.7313533e+00  2.0391275e+01
  5.5753031e+00  2.2229488e+01  7.7750063e+00  1.4274504e+00
  1.7888672e+01  5.7513223e+00  1.3910352e+01 -5.9649639e+00
  6.3385620e+00 -2.1911745e+00  6.3188438e+00 -9.5271111e+00
 -5.3659544e+00  3.5226038e+00 -3.5702689e+00  1.3489943e+00
 -7.8041172e+00  3.9783268e+00 -1.3520060e+00  5.3057799e+00
 -4.3275099e+00 -1.3648185e+00  1.7792124e+00 -4.4988022e+00
 -3.3014898e+00 -5.5654898e+00  5.9882790e-01 -3.8787186e+00
 -1.6456193e-01 -7.5357490e+00 -2.3694208e+00]
```

MFCC_1	MFCC_2	MFCC_3	MFCC_4	MFCC_5	MFCC_6	MFCC_7	MFCC_8	MFCC_9	MFCC_10	MFCC_11	MFCC_12	MFCC_13	MFCC_14	MFCC_15
-566.69	131.0393	19.08785	71.32723	8.652499	48.33132	-4.73135	20.39128	5.575303	22.22949	7.775006	1.42745	17.88867	5.751322	13.91035

---

giá trị nhỏ nhất trên từng frame

```
min_values_per_frame:
 [-579.9491     119.21466    -45.993225    33.25535     -5.2263546
   33.25508    -22.286139     9.191466   -18.023365     9.641236
  -1.8681679   -8.880687      7.033729    -6.6023993    4.3959284
  -14.582114   -7.429231     -11.965993   -2.4002194   -17.859234
  -15.833658   -6.3011475    -14.683865   -10.493055   -17.822098
  -5.837042    -10.64846     -7.1446314   -14.078825   -10.842155
  -7.415408    -15.193752    -13.274603   -13.697229    -6.000227
  -11.743406   -10.208092    -14.986547   -10.825048]
```

MFCC_1	MFCC_2	MFCC_3	MFCC_4	MFCC_5	MFCC_6	MFCC_7	MFCC_8	MFCC_9	MFCC_10	MFCC_11	MFCC_12	MFCC_13	MFCC_14	MFCC_15
-579.949	119.2147	-45.9932	33.25535	-5.22635	33.25508	-22.2861	9.191466	-18.0234	9.641236	-1.86817	-8.88069	7.033729	-6.6024	4.395928

---

giá trị lớn nhất trên từng frame

```
min_values_per_frame: [-301.62225     147.95848      29.485634
82.81587      17.538605
   64.15127       8.157434     37.39849      14.846053
34.676514
   21.21693      11.874265     29.338589     17.39907      24.66743
   3.8207228     18.077168      8.355494     16.376938
0.77879566
   5.8760414     14.942226      8.752022     12.497249
2.2104561
   15.276368     10.782095     14.378485      7.251279
10.7585945
   12.939228      9.664794     12.412324      7.344386      8.06173
    5.007368      9.149814      2.1474037     3.3692346 ]
```

MFCC_1	MFCC_2	MFCC_3	MFCC_4	MFCC_5	MFCC_6	MFCC_7	MFCC_8	MFCC_9	MFCC_10	MFCC_11	MFCC_12	MFCC_13	MFCC_14	MFCC_15
-301.622	147.9585	29.48563	82.81587	17.53861	64.15127	8.157434	37.39849	14.84605	34.67651	21.21693	11.87427	29.33859	17.39907	24.66743