**UỶ BAN NHÂN DÂN**

**THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

****

**Đề cương nghiên cứu**

**(Nhóm 14)**

**Thuộc nhóm ngành khoa học: PP NCKH trong CNTT**

**Chủ nhiệm đề tài: Đỗ Như Tài**

**Thành viên tham gia:**

**1. Lê Nguyễn Anh Đức - 3122410087**

**2. Cao Tấn Đạt - 3122410069**

**3. Nguyễn Trọng Đạt - 3122410075**

**Giáo viên hướng dẫn: TS.Đỗ Như Tài**

**Thành phố Hồ Chí Minh, 2/2025**

[**1. Lý do chọn đề tài 4**](#_uy0l2fu5s07s)

[**2. Tổng quan vấn đề nghiên cứu 5**](#_sxrm6pew3dor)

[2.1. Tình hình nghiên cứu hiện tại 5](#_ckz2oqomytxr)

[2.2. Hướng tiếp cận của đề tài 6](#_z3opd6z3mq4m)

[**3. Mục đích và nhiệm vụ nghiên cứu 6**](#_4dr3seg4g8wk)

[3.1 Mục đích nghiên cứu: 6](#_sc7mw158bjxc)

[3.2 Nhiệm vụ nghiên cứu: 7](#_kqf5or80rrm)

[**4. Đối tượng và phạm vi nghiên cứu 7**](#_dguanhje499u)

[4.1 Đối tượng nghiên cứu: 7](#_il2k58sxqwox)

[4.2 Phạm vi nghiên cứu: 8](#_g16l7z4i4xss)

[4.2.1 Phạm vi dữ liệu: 8](#_w6x1zkbycgwh)

[4.2.2 Phạm vi kỹ thuật: 8](#_em7x3m53k40g)

[4.2.3 Phạm vi ứng dụng: 9](#_cnwhphso5dyu)

[**5. Phương pháp nghiên cứu 9**](#_csrsaj51xroi)

[5.1 Phương pháp lý thuyết: 9](#_f7uo47k24ejr)

[5.2 Phương pháp thực nghiệm: 10](#_pg77d7jlp254)

[5.3 Phương pháp chuyên gia: 10](#_rml18ah4wb61)

[**6. Giả thuyết khoa học 11**](#_psdtd57zsjh)

[**7. Những đóng góp mới của đề tài 11**](#_mex6nwoo9so0)

[**8. Dự kiến kế hoạch nghiên cứu 12**](#_56tsm1r5zpkb)

[**9. Dự kiến nội dung của luận văn 12**](#_szfpbe6l8c9i)

[Chương 1: Giới thiệu 12](#_o7m9viue9ljy)

[Chương 2: Cơ sở lý thuyết 13](#_t7hp5zrx6sp3)

[Chương 3: Phương pháp nghiên cứu và mô hình đề xuất 13](#_town173tq02d)

[Chương 4: Thực nghiệm và kết quả 13](#_bdvx2eyolv8z)

[Chương 5: Kết luận và hướng phát triển 14](#_317w5d4gqh0f)

[**10. Danh mục tài liệu tham khảo 14**](#_socb41dt5fe1)

# 1. Lý do chọn đề tài

Trong bối cảnh phát triển nhanh chóng của trí tuệ nhân tạo (AI), giao tiếp tự nhiên giữa con người và máy móc đang trở thành một trong những hướng đi trọng điểm. Một trong những nền tảng cốt lõi của giao tiếp này chính là khả năng nhận diện và xử lý tiếng nói của con người, nhằm giúp máy có thể hiểu, phản hồi và tương tác hiệu quả hơn với người sử dụng. Tuy nhiên, để thực hiện điều đó, một bước quan trọng và cần thiết chính là xác định chính xác các đoạn âm thanh có chứa tiếng người, phân biệt chúng với các loại âm thanh khác trong môi trường thực tế.

Các mô hình học sâu hiện đại như CNN và LSTM đã chứng minh hiệu quả vượt trội trong nhận dạng âm thanh so với các phương pháp truyền thống. CNN có khả năng học đặc trưng âm học phức tạp tốt hơn nhiều so với các mô hình dựa trên đặc trưng thủ công [2]. Trong khi đó, việc kết hợp CNN với LSTM giúp mô hình không chỉ học được không gian đặc trưng từ phổ tần, mà còn nắm bắt được sự thay đổi theo thời gian – đặc biệt hữu ích trong xử lý chuỗi âm thanh như giọng nói [3].

Trong thực tiễn, điều khiển bằng giọng nói là một chức năng quan trọng trên nhiều thiết bị di động và hệ thống nhà thông minh. Đặc biệt, đây còn là giải pháp hữu ích giúp người hạn chế vận động có thể điều khiển các thiết bị trong cuộc sống hằng ngày một cách dễ dàng hơn [1]. Từ đó kéo theo nhu cầu phát hiện nhanh và chính xác các đoạn tiếng người trong dòng âm thanh liên tục. Đây không chỉ là bài toán kỹ thuật mà còn là yếu tố sống còn để đảm bảo hiệu quả giao tiếp giữa người và máy.

Vì vậy, đề tài "Phân loại tiếng người – Tiền đề giao tiếp với AI" được nhóm lựa chọn với mong muốn:

* Góp phần giải quyết một trong những vấn đề cơ bản của tương tác người – máy.
* Nghiên cứu, so sánh và đề xuất mô hình hiệu quả trong nhận diện tiếng người giữa môi trường âm thanh phức tạp.
* Tạo tiền đề cho các nghiên cứu tiếp theo trong lĩnh vực xử lý tiếng nói và ứng dụng AI vào đời sống.

# 2. Tổng quan vấn đề nghiên cứu

## 2.1. Tình hình nghiên cứu hiện tại

Việc phát hiện và phân loại tiếng người trong tín hiệu âm thanh là một trong những bài toán nền tảng trong lĩnh vực xử lý tín hiệu và trí tuệ nhân tạo. Bài toán này đóng vai trò thiết yếu trong nhiều ứng dụng như nhận dạng giọng nói, trợ lý ảo, lọc âm thanh trong hệ thống giám sát, và đặc biệt là trong giao tiếp người – máy.

Trên thế giới, nhiều nghiên cứu đã ứng dụng các kỹ thuật học máy và học sâu để giải quyết bài toán phân loại giọng nói. Các phương pháp truyền thống như Random Forest và XGBoost tỏ ra hiệu quả khi xử lý đầu vào có đặc trưng đã được trích xuất như MFCC [5, 6], tuy nhiên các phương pháp này vẫn phụ thuộc vào chất lượng đặc trưng và khó thích nghi với dữ liệu có tính biến động cao.

Gần đây, các mô hình học sâu như Convolutional Neural Networks (CNN) và Recurrent Neural Networks (RNN), đặc biệt là sự kết hợp giữa CNN và LSTM (Long Short-Term Memory), đã cho thấy hiệu quả vượt trội trong việc nhận diện và phân loại âm thanh [5]. CNN có khả năng học các đặc trưng không gian từ phổ tần, trong khi LSTM lại nổi bật trong việc khai thác các mối quan hệ theo chuỗi thời gian – vốn rất quan trọng trong ngữ điệu và ngữ cảnh giọng nói.

Bên cạnh đó, các tập dữ liệu lớn và đa dạng như AudioSet, ESC-50 và FSD50K đã tạo điều kiện thuận lợi để huấn luyện và đánh giá hiệu quả các mô hình. Đặc biệt, FSD50K là một trong những tập dữ liệu mở có gán nhãn thủ công, chứa nhiều lớp âm thanh khác nhau bao gồm cả tiếng người (speech) với độ dài, chất lượng và độ phức tạp gần gũi với môi trường thực tế [4].

Tuy vậy, hầu hết các nghiên cứu hiện nay chủ yếu tập trung vào phân loại đa lớp các loại âm thanh tổng quát. Việc tập trung chuyên biệt vào phân biệt speech và non-speech vẫn còn tương đối ít và chưa khai thác đầy đủ tiềm năng của các mô hình lai như CNN + LSTM.

## 2.2. Hướng tiếp cận của đề tài

Đề tài “Phân loại tiếng người – Tiền đề giao tiếp với AI” lựa chọn hướng tiếp cận chuyên sâu vào bài toán nhị phân: phân biệt tiếng người (speech) và phi tiếng người (non-speech) từ tín hiệu âm thanh ngắn.

Cách tiếp cận bao gồm các giai đoạn chính:

* Tiền xử lý và chọn lọc dữ liệu từ tập FSD50K, chỉ tập trung vào các lớp âm thanh liên quan đến tiếng người (speech, conversation, male/female speech/singing, child speech, v.v.) [4].
* Trích xuất đặc trưng âm học từ tín hiệu, sử dụng phổ log-mel hoặc MFCC để phục vụ cho cả mô hình học máy và học sâu.
* So sánh hiệu quả của các mô hình học máy (Random Forest, XGBoost) và học sâu (CNN, CNN + LSTM) nhằm xác định mô hình phù hợp nhất trong bối cảnh thực tế.
* Phân tích chi tiết ưu – nhược điểm của từng mô hình, không chỉ qua độ chính xác mà còn thông qua khả năng tổng quát hóa, tốc độ xử lý và độ ổn định với dữ liệu nhiễu.

Thông qua đó, đề tài không chỉ đưa ra được một mô hình khả thi để nhận diện tiếng người trong âm thanh mà còn xây dựng được một hướng nghiên cứu có thể mở rộng cho các ứng dụng cao cấp hơn như nhận dạng người nói, tách lời thoại, hoặc điều khiển thông minh bằng giọng nói.

# 3. Mục đích và nhiệm vụ nghiên cứu

## 3.1 Mục đích nghiên cứu:

Xây dựng một hệ thống phân loại tiếng người hiệu quả, có khả năng phát hiện chính xác các đoạn âm thanh có chứa tiếng nói con người từ môi trường âm thanh tổng hợp.

So sánh hiệu năng giữa các mô hình học máy truyền thống (Random Forest, XGBoost), mô hình học sâu CNN, và mô hình kết hợp CNN + LSTM để xác định phương pháp tối ưu cho bài toán nhận diện tiếng người.

Góp phần tạo tiền đề cho các ứng dụng giao tiếp người – máy, đặc biệt trong các hệ thống trợ lý ảo, điều khiển bằng giọng nói, và các giải pháp tương tác thông minh khác.

## 3.2 Nhiệm vụ nghiên cứu:

Thu thập và xử lý tập dữ liệu âm thanh có gán nhãn – sử dụng tập FSD50K để xây dựng bài toán phân loại nhị phân speech / non-speech.

Trích xuất đặc trưng từ tín hiệu âm thanh (MFCC, log-mel spectrogram, v.v.) để làm đầu vào cho các mô hình học máy.

Huấn luyện và đánh giá ba mô hình nền tảng:

* Random Forest (RF)
* XGBoost
* CNN
* Xây dựng mô hình học sâu kết hợp CNN + LSTM để khai thác cả đặc trưng không gian và đặc trưng thời gian từ tín hiệu âm thanh.

So sánh hiệu suất giữa các mô hình theo các chỉ số như Accuracy, Precision, Recall, F1-score.

Phân tích ưu – nhược điểm của từng phương pháp và đề xuất mô hình phù hợp nhất cho ứng dụng thực tế.

# 4. Đối tượng và phạm vi nghiên cứu

## 4.1 Đối tượng nghiên cứu:

Đối tượng nghiên cứu của đề tài này là các mô hình học máy và học sâu được áp dụng trong bài toán phân loại tiếng nói của con người từ các đoạn âm thanh tổng hợp có chứa nhiều loại âm thanh khác nhau. Cụ thể:

Các mô hình học máy truyền thống như Random Forest (RF) và XGBoost, được huấn luyện trên đặc trưng âm học rút trích từ dữ liệu âm thanh để thực hiện phân loại nhị phân có tiếng người / không có tiếng người.

Mô hình học sâu Convolutional Neural Network (CNN), khai thác khả năng học đặc trưng từ phổ tần (spectrogram) hoặc log-mel của tín hiệu âm thanh.

Mô hình kết hợp CNN + LSTM, tận dụng điểm mạnh của cả hai: CNN xử lý đặc trưng không gian từ ảnh phổ tần, trong khi LSTM xử lý thông tin thời gian để nắm bắt ngữ cảnh và chuỗi âm thanh liên tục – rất đặc trưng cho dữ liệu tiếng nói.

Đặc trưng âm học MFCC (Mel-Frequency Cepstral Coefficients), đóng vai trò là đầu vào tiêu chuẩn cho các mô hình này.

Bên cạnh đó, dữ liệu âm thanh được sử dụng là từ tập FSD50K, một tập dữ liệu lớn, đa dạng, có nhãn rõ ràng và phù hợp với việc xây dựng bài toán nhận diện tiếng người trong môi trường thực tế.

## 4.2 Phạm vi nghiên cứu:

Phạm vi nghiên cứu của đề tài được xác định rõ ràng nhằm tập trung vào giải quyết một bài toán nền tảng nhưng quan trọng trong lĩnh vực xử lý âm thanh – đó là phân loại nhị phân tiếng người, với các giới hạn cụ thể sau:

### 4.2.1 Phạm vi dữ liệu:

Nghiên cứu sử dụng duy nhất tập dữ liệu FSD50K, bao gồm hai phần: dev set và eval set. Trong đó, dev set sẽ được chia thành train và validation để huấn luyện và tinh chỉnh mô hình; eval set được sử dụng để đánh giá cuối cùng.

Dữ liệu được xử lý trước để lọc ra các đoạn âm thanh thuộc hai nhóm: speech (theo danh sách các nhãn tiếng người đã xác định) và non-speech (tất cả các âm thanh còn lại).

### 4.2.2 Phạm vi kỹ thuật:

Nghiên cứu tập trung vào bài toán phân loại nhị phân, không đi sâu vào phân tích nội dung tiếng nói, nhận dạng người nói, ngôn ngữ, giới tính hay cảm xúc.

Việc trích xuất đặc trưng âm học và huấn luyện mô hình được giới hạn trong các kỹ thuật phổ biến và đã được xác minh hiệu quả: MFCC, log-mel spectrogram, CNN, LSTM.

Mô hình sẽ được đánh giá theo các chỉ số kỹ thuật như Accuracy, Precision, Recall, F1-score để đảm bảo tính khách quan và dễ so sánh.

### 4.2.3 Phạm vi ứng dụng:

Đề tài dừng lại ở mức nghiên cứu mô hình và đánh giá hiệu quả. Chưa triển khai thành ứng dụng thực tế hoặc tích hợp với hệ thống nhận diện tiếng nói hoàn chỉnh.

Tuy nhiên, kết quả nghiên cứu có thể đóng vai trò làm nền tảng cho các ứng dụng tương lai như trợ lý ảo, hệ thống phản hồi bằng giọng nói, nhận dạng hoạt động trong môi trường âm thanh phức hợp, v.v.

# 5. Phương pháp nghiên cứu

## 5.1 Phương pháp lý thuyết:

Tổng quan tài liệu khoa học: Thu thập, phân tích và tổng hợp các nghiên cứu liên quan đến phân loại âm thanh, nhận diện tiếng người, cũng như các phương pháp học máy và học sâu phổ biến như Random Forest, XGBoost, CNN và CNN+LSTM.

Nghiên cứu đặc trưng âm học: Tìm hiểu các đặc trưng âm thanh thường được sử dụng trong phân tích tín hiệu giọng nói, như MFCC, log-mel spectrogram, chroma features, từ đó lựa chọn phương pháp trích xuất phù hợp với mục tiêu nghiên cứu.

Phân tích so sánh mô hình: Nghiên cứu các tiêu chí đánh giá hiệu quả mô hình phân loại (accuracy, precision, recall, F1-score) và lý thuyết đằng sau cơ chế hoạt động của từng mô hình để từ đó thiết kế khung so sánh khách quan.

Thiết lập giả thuyết nghiên cứu: Đưa ra giả thuyết rằng mô hình CNN kết hợp LSTM sẽ cho hiệu quả vượt trội so với các mô hình còn lại trong bài toán phân loại tiếng người từ tín hiệu âm thanh thực tế.

## 5.2 Phương pháp thực nghiệm:

Tiền xử lý dữ liệu: Áp dụng quy trình lọc nhãn từ tập dữ liệu FSD50K để tách hai lớp speech và non-speech một cách chính xác, đồng thời chia dữ liệu thành tập train/validation/test hợp lý.

Trích xuất đặc trưng: Thực hiện trích xuất các đặc trưng âm học như MFCC và log-mel spectrogram từ các đoạn âm thanh để làm đầu vào cho mô hình học máy và học sâu.

Huấn luyện mô hình: Xây dựng và huấn luyện lần lượt các mô hình:

* Mô hình học máy: Random Forest, XGBoost (với dữ liệu đặc trưng dạng vector)
* Mô hình học sâu: CNN và CNN + LSTM (với đầu vào dạng ảnh hoặc chuỗi thời gian)

Tối ưu hóa tham số: Sử dụng phương pháp thử nghiệm (grid search, validation loss tracking) để chọn siêu tham số phù hợp nhất cho từng mô hình.

Đánh giá và so sánh kết quả: Đo lường hiệu quả mô hình thông qua tập test, ghi nhận các chỉ số đánh giá và tiến hành so sánh chi tiết giữa các mô hình.

## 5.3 Phương pháp chuyên gia:

• Trao đổi thường xuyên với giảng viên hướng dẫn để làm rõ các vấn đề khoa

học liên quan, nhận phản hồi và định hướng cho từng giai đoạn nghiên cứu.

• Dựa trên ý kiến của giảng viên, điều chỉnh cách tiếp cận, tối ưu hóa quá trình

triển khai thực nghiệm, và đảm bảo tiến độ thực hiện luận văn.

Phương pháp nghiên cứu này kết hợp chặt chẽ giữa lý thuyết và thực nghiệm,

cùng với sự hỗ trợ từ chuyên gia, nhằm đảm bảo kết quả nghiên cứu có tính khoa

học, khả thi và giá trị ứng dụng cao.

# 6. Giả thuyết khoa học

Đề tài đặt ra giả thuyết rằng các mô hình học sâu, đặc biệt là mô hình kết hợp giữa CNN và LSTM, có thể mang lại hiệu quả vượt trội so với các phương pháp học máy truyền thống như Random Forest hay XGBoost. CNN có khả năng học các đặc trưng không gian từ biểu diễn phổ tần của tín hiệu âm thanh, trong khi LSTM lại giúp mô hình nắm bắt được các mối quan hệ theo chuỗi thời gian – một yếu tố then chốt trong đặc điểm ngữ điệu và nhịp điệu của tiếng người.

Đồng thời, đề tài cũng giả định rằng việc tuyển chọn cẩn thận các nhãn liên quan đến tiếng người từ tập dữ liệu FSD50K như "speech", "conversation", "singing", v.v., sẽ giúp mô hình tập trung học sâu vào đặc điểm ngôn ngữ và giọng nói của con người, từ đó nâng cao độ chính xác và khả năng phân biệt tiếng người trong môi trường thực tế. Ngoài hiệu suất cao về độ chính xác, mô hình CNN+LSTM còn được kỳ vọng có tính ổn định hơn khi xử lý dữ liệu âm thanh có nhiễu hoặc biến động, nhờ khả năng phối hợp giữa học đặc trưng không gian và chuỗi thời gian.

# 7. Những đóng góp mới của đề tài

Đề tài đóng góp một hướng tiếp cận chuyên biệt cho bài toán phân loại tiếng người từ tín hiệu âm thanh, bằng cách xây dựng mô hình CNN kết hợp LSTM nhằm khai thác đồng thời đặc trưng không gian và chuỗi thời gian. Bên cạnh đó, đề tài xác định rõ tập nhãn liên quan đến tiếng người từ FSD50K, đồng thời so sánh hiệu quả mô hình học sâu với các mô hình học máy truyền thống (RF, XGBoost) trong cùng điều kiện thực nghiệm, từ đó làm rõ ưu – nhược điểm của từng phương pháp và cung cấp quy trình xử lý có thể tái sử dụng cho các nghiên cứu tương tự.

# 8. Dự kiến kế hoạch nghiên cứu

|  |  |  |
| --- | --- | --- |
| STT | Nội dung | Dự kiến thời gian thực hiện |
| 1 | Nghiên cứu, chọn đề tài, xây dựng đề cương luận văn | 2 tháng |
| 2 | Sửa chữa hoàn thiện đề cương, nộp đề cương | 0.5 tháng |
| 3 | Nghiên cứu, viết, hoàn thiện luận văn |  |
| Chương 1: Giới thiệu | 0.5 tháng |
| Chương 2: Các nghiên cứu liên quan | 0.5 tháng |
| Chương 3: Thực nghiệm | 1 tháng |
| Chương 4: Kết luận và hướng phát triển | 0.5 tháng |
| 5 | Chỉnh sửa, hoàn thiện luận văn và hồ sơ bảo vệ luận văn | 1 tháng |

# 9. Dự kiến nội dung của luận văn

## Chương 1: Giới thiệu

Lý do chọn đề tài.

Mục tiêu và nhiệm vụ nghiên cứu.

Đối tượng và phạm vi nghiên cứu.

Giả thuyết khoa học.

Phương pháp nghiên cứu.

Những đóng góp mới của đề tài.

Cấu trúc luận văn.

## Chương 2: Cơ sở lý thuyết

Tổng quan về tín hiệu âm thanh và đặc trưng MFCC.

Giới thiệu các mô hình học máy và học sâu:

* Random Forest và XGBoost: cơ chế hoạt động, ưu – nhược điểm.
* Mạng nơ-ron tích chập (CNN) và mạng hồi tiếp (LSTM).
* Kiến trúc CNN+LSTM cho dữ liệu chuỗi thời gian.

Các nghiên cứu liên quan đến phân loại tiếng người.

Các chỉ số đánh giá mô hình: Accuracy, Precision, Recall, F1-score, ROC-AUC.

## Chương 3: Phương pháp nghiên cứu và mô hình đề xuất

Quy trình nghiên cứu tổng quát.

Xây dựng tập nhãn tiếng người từ dữ liệu FSD50K.

Tiền xử lý dữ liệu âm thanh:

* Trích xuất MFCC, chia tập train/val/test từ Dev và Eval.

Thiết kế mô hình huấn luyện:

* Cài đặt và huấn luyện các mô hình RF, XGBoost, CNN và CNN+LSTM.
* Điều chỉnh tham số, kỹ thuật tránh overfitting.

Mô tả cách đánh giá và so sánh mô hình.

## Chương 4: Thực nghiệm và kết quả

Mô tả tập dữ liệu và thông số huấn luyện.

Đánh giá kết quả huấn luyện các mô hình:

* So sánh hiệu suất giữa RF, XGBoost, CNN và CNN+LSTM.
* Phân tích ưu – nhược điểm từng mô hình trong bài toán phân loại tiếng người.

Trình bày kết quả qua bảng biểu và hình ảnh.

## Chương 5: Kết luận và hướng phát triển

Tóm tắt kết quả đạt được.

Đánh giá khả năng áp dụng mô hình CNN+LSTM trong thực tế.

Hạn chế của nghiên cứu.

Hướng phát triển:

* Áp dụng mô hình vào phân loại nâng cao như: phân biệt người nói, nhận diện ngữ điệu.
* Kết hợp thêm các kỹ thuật Attention hoặc sử dụng Transformer cho dữ liệu âm thanh.

# 10. Danh mục tài liệu tham khảo

[1] Thái, Thuận Thương. “Nhận dạng tiếng nói điều khiển với convolutional neural network (CNN)”. Can Tho University Journal of Science, vol 57, số p.h 4, Tháng Tám 2021, tr 30–39. DOI.org (Crossref), https://doi.org/10.22144/ctu.jvn.2021.111.

[2] Zhang, Ying, et al. Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks. arXiv:1701.02720, arXiv, 10 Jan. 2017. arXiv.org, <https://doi.org/10.48550/arXiv.1701.02720>.

[3] Donahue, Jeff, et al. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. arXiv:1411.4389, arXiv, 31 May 2016. arXiv.org, <https://doi.org/10.48550/arXiv.1411.4389>.

[4] Fonseca, Eduardo, et al. FSD50K: An Open Dataset of Human-Labeled Sound Events. arXiv:2010.00475, arXiv, 23 Apr. 2022. arXiv.org, https://doi.org/10.48550/arXiv.2010.00475.

[5] Mohan, Meera, và c.s. “Speech Emotion Classification Using Ensemble Models with MFCC”. Procedia Computer Science, vol 218, 2023, tr 1857–68. DOI.org (Crossref), <https://doi.org/10.1016/j.procs.2023.01.163>.

[6] Khadar Nawas, K., et al. “Speaker Recognition Using Random Forest.” ITM Web of Conferences, edited by J. Kannan R. et al., vol. 37, 2021, p. 01022. DOI.org (Crossref), https://doi.org/10.1051/itmconf/20213701022.