

SỬ DỤNG HỌC MÁY ĐỂ PHÂN TÍCH TÍN HIỆU GIỌNG NÓI VÀ NHẬN BIẾT NGUY CƠ CHẬM NÓI Ở TRẺ NHỎ

Nguyễn Thị Thuý Hiền¹, Nguyễn Đức Nguyên¹, Nguyễn Quang Anh², Lại Thanh Vân³

1-Trường Đại học Giáo dục, Đại học Quốc gia Hà Nội

2-Trường Đại học Vancouver Island

3-Trường Quốc tế, Đại học Quốc gia Hà Nội

Tóm tắt: Chậm nói là một trong những dấu hiệu quan trọng của rối loạn phát triển ngôn ngữ ở trẻ nhỏ, có thể ảnh hưởng đến khả năng học tập và giao tiếp sau này. Việc phát hiện sớm nguy cơ chậm nói giúp can thiệp kịp thời, nâng cao hiệu quả điều trị. Nghiên cứu này áp dụng học máy để phân tích tín hiệu giọng nói của trẻ nhằm xây dựng mô hình dự đoán nguy cơ chậm nói. Dữ liệu giọng nói của trẻ từ 2-5 tuổi được thu thập và xử lý thông qua các phương pháp trích xuất đặc trưng âm thanh (MFCC, pitch, formant). Sau đó, các thuật toán học máy như Random Forest, SVM, và Mạng nơ-ron nhân tạo (ANN) được huấn luyện để phân loại trẻ có nguy cơ chậm nói. Kết quả nghiên cứu cho thấy mô hình có độ chính xác cao, hỗ trợ hiệu quả trong việc sàng lọc ban đầu.

Từ khóa: Chậm nói, học máy, phân tích giọng nói, phát hiện sớm, trẻ nhỏ, trích xuất đặc trưng âm thanh

USING MACHINE LEARNING FOR SPEECH SIGNAL ANALYSIS AND IDENTIFYING RISK OF SPEECH DELAY IN YOUNG CHILDREN

Nguyen Thi Thuy Hien¹, Nguyen Duc Nguyen¹, Nguyen Quang Anh², Lai Thanh Van³

1- University of Education, VNU

2-Vancouver Island University

3-International School, VNU

Abstract: Speech delay is one of the important signs of language development disorders in young children, which can affect their ability to learn and communicate later. Early detection of the risk of speech delay helps to intervene promptly and improve treatment effectiveness. This study applies machine learning to analyze children's speech signals to build a model to predict the risk of speech delay. Speech data of children aged 2-5 years old was collected and processed through audio feature extraction methods (MFCC, pitch, formant). Then, machine learning algorithms such as Random Forest, SVM, and Artificial Neural Network (ANN) were trained to classify children at risk of speech delay. The results of the study showed that the model had high accuracy, effectively supporting initial screening.

Keywords: Speech delay, machine learning, speech analysis, early detection, young children, acoustic feature extraction

Nhận bài: 12/03/2025

Phản biện: 26/03/2025

Duyệt đăng: 28/03/2025

I. ĐẶT VẤN ĐỀ

Chậm nói là một trong những dấu hiệu phổ biến của rối loạn phát triển ngôn ngữ ở trẻ nhỏ, có thể ảnh hưởng nghiêm trọng đến khả năng giao tiếp, học tập và hòa nhập xã hội của trẻ trong tương lai. Theo nghiên cứu của American Speech-Language-Hearing Association (ASHA, 2023), khoảng 10-15% trẻ từ 2-5 tuổi gặp vấn đề về phát triển ngôn ngữ, trong đó chậm nói là một triệu chứng điển hình. Ở Việt Nam, theo thống kê của Bệnh viện Nhi Trung ương, tỷ lệ trẻ gặp khó khăn trong phát triển ngôn ngữ chiếm khoảng 20% trong nhóm trẻ dưới 5 tuổi. Tuy nhiên, thực tế cho thấy việc phát hiện và can thiệp sớm ở trẻ có nguy cơ chậm nói vẫn còn nhiều hạn chế, do thiếu hệ thống sàng lọc khách quan và phụ thuộc quá nhiều vào đánh giá chủ quan của phụ huynh và chuyên gia.

Hiện nay, quá trình chẩn đoán chậm nói chủ yếu dựa vào quan sát lâm sàng, đánh giá hành vi ngôn ngữ của trẻ và thu thập thông tin từ gia đình.

Mặc dù cách tiếp cận này mang lại cái nhìn trực quan về sự phát triển ngôn ngữ của trẻ, nhưng nó có thể dẫn đến nhiều hạn chế như phát hiện muộn, chẩn đoán sai hoặc chưa đầy đủ. Trong khi đó, các nghiên cứu gần đây chỉ ra rằng học máy (Machine Learning - ML) có tiềm năng lớn trong việc tự động hóa quá trình phân tích giọng nói, giúp phát hiện sớm nguy cơ chậm nói với độ chính xác cao. Các phương pháp trích xuất đặc trưng âm thanh như MFCC (Mel-Frequency Cepstral Coefficients), pitch, formant đang được ứng dụng rộng rãi trong nghiên cứu về nhận dạng giọng nói và rối loạn ngôn ngữ. Những đặc trưng này có thể cung cấp thông tin quan trọng để xác định sự khác biệt giữa trẻ có nguy cơ chậm nói và trẻ phát triển bình thường.

II. NỘI DUNG NGHIÊN CỨU

2.1. Một số vấn đề lý luận về chậm nói

Khái niệm về chậm nói: Chậm nói ở trẻ nhỏ là tình trạng trẻ có sự chậm trễ đáng kể trong

việc phát triển kỹ năng ngôn ngữ so với các cột mốc phát triển bình thường. Theo Hiệp hội Ngôn ngữ - Lời nói - Thính giác Hoa Kỳ (American Speech-Language-Hearing Association - ASHA, 2023), trẻ được coi là chậm nói khi khả năng giao tiếp bằng lời nói của trẻ kém hơn đáng kể so với trẻ cùng độ tuổi, bao gồm việc hiểu ngôn ngữ (receptive language) và khả năng biểu đạt (expressive language). Bộ Giáo dục và Đào tạo Việt Nam đã đưa ra một số tiêu chí đánh giá sự phát triển ngôn ngữ của trẻ, bao gồm:

Khả năng hiểu ngôn ngữ: Trẻ có thể hiểu các mệnh lệnh đơn giản, nhận biết đồ vật và phản ứng với lời nói của người lớn.

Khả năng biểu đạt: Trẻ sử dụng từ ngữ để diễn đạt ý muốn, ghép từ thành câu có nghĩa, sử dụng câu phù hợp với độ tuổi.

Khả năng sử dụng ngôn ngữ phù hợp ngữ cảnh: Trẻ có thể giao tiếp với người khác, trả lời câu hỏi đơn giản, tham gia vào cuộc trò chuyện.

Thông thường, trẻ từ 18 tháng tuổi có thể nói ít nhất 10 từ đơn, trẻ 2 tuổi có thể nói khoảng 50 từ và bắt đầu ghép hai từ thành câu đơn giản. Đến 3 tuổi, trẻ có thể nói câu dài 4-5 từ và sử dụng ngôn ngữ để diễn đạt nhu cầu một cách rõ ràng hơn. Nếu trẻ không đạt được các mốc này, có thể nghi ngờ trẻ có dấu hiệu chậm nói. Tuy nhiên, đánh giá chậm nói không chỉ dựa vào số lượng từ vựng mà còn dựa trên sự phát triển toàn diện của các kỹ năng ngôn ngữ, bao gồm khả năng hiểu, biểu đạt, phát âm và giao tiếp xã hội. Việc chẩn đoán chậm nói cần được thực hiện bởi chuyên gia ngôn ngữ trị liệu hoặc bác sĩ nhi khoa thông qua các bài kiểm tra chuyên sâu như MacArthur-Bates Communicative Development Inventories (CDI) hoặc Preschool Language Scale (PLS-5).

Nguyên nhân của chậm nói

Nguyên nhân sinh học:

+ Rối loạn thính giác: Trẻ bị mất thính lực hoặc suy giảm thính lực có thể gặp khó khăn trong việc tiếp nhận và học ngôn ngữ.

+ Rối loạn thần kinh: Một số trẻ có bất thường về não bộ, chẳng hạn như bại não, động kinh hoặc tổn thương vùng ngôn ngữ của não.

+ Rối loạn phổ tự kỷ (ASD): Nhiều trẻ mắc tự kỷ có biểu hiện chậm nói hoặc mất khả năng ngôn ngữ sau một thời gian phát triển bình thường.

+ Chậm phát triển trí tuệ: Trẻ có chỉ số IQ thấp có thể gặp khó khăn trong việc tiếp thu và sử dụng ngôn ngữ.

Nguyên nhân môi trường:

+ Thiếu tương tác ngôn ngữ: Trẻ không được giao tiếp thường xuyên với cha mẹ hoặc người

chăm sóc sẽ có nguy cơ chậm nói cao hơn.

+ Lạm dụng thiết bị điện tử: Trẻ tiếp xúc quá nhiều với tivi, điện thoại, máy tính bảng có thể bị giảm khả năng giao tiếp trực tiếp.

+ Môi trường song ngữ không hỗ trợ đúng cách: Một số trẻ trong gia đình song ngữ có thể gặp khó khăn khi tiếp nhận cả hai ngôn ngữ cùng lúc.

Hậu quả của chậm nói: Nếu không được can thiệp sớm, chậm nói có thể dẫn đến nhiều hậu quả nghiêm trọng: ảnh hưởng đến khả năng giao tiếp và tương tác xã hội của trẻ; trẻ có thể gặp khó khăn trong việc học tập, đặc biệt là kỹ năng đọc, viết sau này; ảnh hưởng đến tâm lý, gây ra sự tự ti, lo âu, dẫn đến các vấn đề về hành vi như cáu gắt, bướng bỉnh hoặc thu mình; làm tăng nguy cơ rối loạn ngôn ngữ kéo dài đến giai đoạn trưởng thành.

2.2. Phương pháp chẩn đoán và can thiệp truyền thống

2.2.1. Phương pháp chẩn đoán

Chẩn đoán chậm nói thường dựa vào:

Quan sát lâm sàng: Chuyên gia đánh giá khả năng ngôn ngữ thông qua quan sát hành vi và phản ứng giao tiếp của trẻ.

Bảng kiểm đánh giá phát triển ngôn ngữ: Các công cụ như MacArthur-Bates CDI, Denver Developmental Screening Test II (DDST-II) hoặc Bayley Scales of Infant Development (BSID) thường được sử dụng để đo lường sự phát triển ngôn ngữ.

Đánh giá thính giác và thần kinh: Kiểm tra chức năng nghe để loại trừ các nguyên nhân do mất thính lực.

2.2.2. Phương pháp can thiệp truyền thống

Liệu pháp ngôn ngữ: Trẻ được chuyên gia ngôn ngữ trị liệu hướng dẫn các bài tập giúp tăng cường kỹ năng phát âm, xây dựng vốn từ vựng và cải thiện khả năng giao tiếp.

Tư vấn và hỗ trợ gia đình: Cha mẹ được hướng dẫn cách giao tiếp hiệu quả với trẻ, như khuyến khích trẻ đặt câu hỏi, mở rộng câu trả lời của trẻ, đọc sách cùng trẻ.

Ứng dụng công nghệ hỗ trợ: Một số trẻ có thể sử dụng các ứng dụng hỗ trợ giao tiếp hoặc các chương trình giáo dục ngôn ngữ trên máy tính bảng.

2.3 Công nghệ học máy trong xử lý giọng nói

Học máy (Machine Learning - ML) là một lĩnh vực quan trọng của trí tuệ nhân tạo (Artificial Intelligence - AI), trong đó các thuật toán được thiết kế để học từ dữ liệu và đưa ra dự đoán mà không cần lập trình cụ thể cho từng trường hợp. Trong những năm gần đây, học máy đã được ứng dụng rộng rãi trong xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP) và phân

tích giọng nói, mở ra nhiều hướng tiếp cận mới trong việc phát hiện và chẩn đoán các vấn đề liên quan đến ngôn ngữ và giao tiếp, bao gồm chậm nói ở trẻ nhỏ.

Giọng nói là một dạng tín hiệu thời gian liên tục chứa nhiều thông tin phức tạp về tần số, cường độ và sự biến đổi âm thanh theo thời gian. Việc phân tích tín hiệu giọng nói không chỉ giúp nhận diện lời nói mà còn cung cấp thông tin quan trọng về đặc điểm phát âm, cấu trúc âm vị và thậm chí là tình trạng sức khỏe của người nói. Các mô hình học máy có thể tự động học các mẫu giọng nói, phát hiện các bất thường trong cách phát âm, từ đó hỗ trợ phát hiện sớm các rối loạn ngôn ngữ như chậm nói, nói lắp hay rối loạn phổ tự kỷ (Autism Spectrum Disorder - ASD).

Ứng dụng của học máy trong phân tích giọng nói bao gồm:

Nhận diện giọng nói tự động (Automatic Speech Recognition - ASR): Hỗ trợ chuyển đổi giọng nói thành văn bản, ứng dụng rộng rãi trong các hệ thống trợ lý ảo như Siri, Google Assistant.

Phân tích cảm xúc trong giọng nói (Speech Emotion Recognition - SER): Nhận diện cảm xúc dựa trên đặc điểm âm thanh, hỗ trợ trong các lĩnh vực tâm lý học và chăm sóc sức khỏe.

Chẩn đoán rối loạn ngôn ngữ: Giúp phát hiện các dấu hiệu bất thường trong giọng nói, từ đó hỗ trợ quá trình chẩn đoán và can thiệp sớm đối với trẻ có nguy cơ chậm nói.

Trong phát hiện chậm nói, học máy cho phép tự động hóa quá trình phân tích giọng nói, giúp sàng lọc sớm nguy cơ ở trẻ em mà không phụ thuộc hoàn toàn vào đánh giá chủ quan của các chuyên gia. Điều này góp phần tăng độ chính xác trong chẩn đoán và tạo điều kiện thuận lợi cho việc triển khai các chương trình can thiệp kịp thời.

Các đặc trưng giọng nói quan trọng: MFCC, formant, pitch, độ dài nguyên âm

MFCC (Mel-Frequency Cepstral Coefficients): Đây là đặc trưng phổ biến nhất trong xử lý giọng nói, được sử dụng để trích xuất thông tin về cấu trúc phổ của âm thanh theo thang đo Mel, phản ánh cách tai người cảm nhận tần số âm thanh. MFCC giúp mô hình học máy nhận diện được sự khác biệt giữa các âm vị trong ngôn ngữ, từ đó hỗ trợ phân loại giọng nói và phát hiện các bất thường trong cách phát âm của trẻ.

Formant: Là các đỉnh phổ tần số của giọng nói, đặc biệt quan trọng trong việc xác định nguyên âm. Những trẻ có vấn đề về ngôn ngữ thường có sự thay đổi bất thường trong formant, điều này có thể giúp phân biệt trẻ bình thường và trẻ có nguy

cơ chậm nói.

Pitch (độ cao giọng nói): Phản ánh tần số cơ bản của giọng nói, liên quan đến cách một người phát âm và độ ổn định trong giọng nói. Trẻ chậm nói thường có sự biến đổi bất thường về pitch do chưa kiểm soát tốt các cơ quan phát âm.

Độ dài nguyên âm (Vowel Duration): Những trẻ gặp khó khăn trong phát triển ngôn ngữ thường có xu hướng kéo dài hoặc rút ngắn bất thường các nguyên âm trong lời nói. Đây là một đặc trưng quan trọng giúp nhận diện dấu hiệu chậm nói.

2.4. Các phương pháp học máy trong phân loại giọng nói

2.4.1. Đặc điểm

Support Vector Machine (SVM) là một thuật toán phân loại mạnh mẽ dựa trên việc tìm siêu phẳng tối ưu để tách biệt các lớp dữ liệu. SVM có thể áp dụng cho cả phân loại tuyến tính và phi tuyến tính bằng cách sử dụng các hàm kernel như RBF, polynomial hoặc sigmoid. Mô hình này hoạt động tốt với tập dữ liệu nhỏ và có độ nhiễu thấp, giúp tránh overfitting. Tuy nhiên, với dữ liệu lớn hoặc có số chiều cao, SVM thường chậm và khó mở rộng.

Random Forest là một mô hình học máy dựa trên tập hợp nhiều cây quyết định (Decision Trees) để tăng cường độ chính xác và giảm nguy cơ overfitting. Bằng cách kết hợp nhiều cây con với dữ liệu huấn luyện ngẫu nhiên, Random Forest có khả năng xử lý tốt dữ liệu nhiễu và cải thiện hiệu suất phân loại. Tuy nhiên, nhược điểm của mô hình này là khó giải thích khi số lượng cây quá lớn, khiến việc diễn giải quyết định của mô hình trở nên phức tạp.

Convolutional Neural Network (CNN) là một mô hình học sâu được thiết kế chủ yếu cho xử lý dữ liệu không gian, đặc biệt là hình ảnh. Trong phân tích giọng nói, CNN được sử dụng để phân loại tín hiệu âm thanh dưới dạng biểu đồ tần số (spectrogram). CNN có khả năng tự động trích xuất đặc trưng, giúp giảm thiểu công đoạn tiền xử lý dữ liệu. Tuy nhiên, mô hình này yêu cầu tập dữ liệu lớn và tài nguyên tính toán mạnh để huấn luyện hiệu quả.

Long Short-Term Memory (LSTM) là một biến thể của mạng nơ-ron hồi tiếp (Recurrent Neural Network - RNN), được thiết kế để xử lý dữ liệu chuỗi dài. Nhờ vào các cổng điều khiển, LSTM có thể ghi nhớ thông tin dài hạn, giúp mô hình hóa tốt các tín hiệu âm thanh có tính tuần tự cao. Tuy nhiên, nhược điểm của LSTM là thời gian huấn luyện lâu và yêu cầu tài nguyên tính toán lớn hơn so với các mô hình truyền thống.

2.4.2. So sánh các mô hình: SVM, Random Forest, CNN, LSTM

Hiện nay, nhiều phương pháp học máy khác nhau đã được áp dụng để phân tích tín hiệu giọng nói nhằm nhận diện các bất thường về phát triển ngôn ngữ, trong đó có chậm nói. Các mô hình phổ

biến bao gồm Support Vector Machine (SVM), Random Forest, Convolutional Neural Network (CNN) và Long Short-Term Memory (LSTM). Mỗi mô hình có ưu điểm và hạn chế riêng, phù hợp với từng loại dữ liệu và mục đích nghiên cứu khác nhau.

Mô hình	Ưu điểm	Nhược điểm	Ứng dụng
SVM (Support Vector Machine)	Hoạt động tốt với tập dữ liệu nhỏ, tránh hiện tượng quá khớp (overfitting).	Khó mở rộng với dữ liệu lớn, yêu cầu nhiều tài nguyên tính toán khi xử lý dữ liệu có số chiều cao.	Phù hợp khi tập dữ liệu nhỏ hoặc trung bình, phân loại giọng nói với số lượng đặc trưng ít.
Random Forest	Mạnh mẽ, ổn định, dễ triển khai và ít bị ảnh hưởng bởi nhiễu.	Khó giải thích logic bên trong mô hình, đôi khi chậm khi có quá nhiều cây quyết định.	Tốt cho các bài toán phân loại giọng nói dựa trên các đặc trưng trích xuất (MFCC, pitch, formant).
CNN (Convolutional Neural Network)	Xử lý tốt dữ liệu dạng hình ảnh như spectrogram của giọng nói, khả năng tự động trích xuất đặc trưng.	Yêu cầu GPU mạnh, cần tập dữ liệu lớn để đạt hiệu suất cao.	Phù hợp cho phân tích giọng nói bằng cách sử dụng spectrogram hoặc các bản đồ đặc trưng âm thanh.
LSTM (Long Short-Term Memory)	Tốt trong xử lý chuỗi thời gian, ghi nhớ thông tin dài hạn giúp phân tích tín hiệu giọng nói chính xác.	Chậm hơn do độ phức tạp cao, yêu cầu nhiều tài nguyên tính toán.	Phù hợp cho nhận diện giọng nói dựa trên dữ liệu thời gian liên tục, phân tích chuỗi tín hiệu âm thanh.

III. KẾT LUẬN

Tóm lại, trong nghiên cứu này việc ứng dụng học máy để phân tích tín hiệu giọng nói nhằm nhận diện nguy cơ chậm nói không chỉ giúp sàng lọc sớm mà còn hỗ trợ các nhà chuyên môn trong quá trình đánh giá, từ đó nâng cao hiệu quả của quá trình can thiệp. Một hệ thống nhận diện nguy cơ

chậm nói dựa trên học máy có thể giúp phát hiện những dấu hiệu bất thường trong giọng nói của trẻ, hỗ trợ cha mẹ và chuyên gia đưa ra quyết định kịp thời về các biện pháp can thiệp. Điều này đặc biệt quan trọng trong bối cảnh hiện nay khi tỷ lệ trẻ gặp vấn đề về phát triển ngôn ngữ ngày càng gia tăng, nhưng nguồn lực chuyên môn vẫn còn hạn chế.

TÀI LIỆU THAM KHẢO

American Speech-Language-Hearing Association (ASHA). (2023). *Late Language Emergence (LLE) in Toddlers*.
National Institute on Deafness and Other Communication Disorders (NIDCD). (2022). *Speech and Language Developmental Milestones*.
Bệnh viện Nhi Trung ương (2023). *Trẻ chậm phát triển ngôn ngữ: Phát hiện sớm, can thiệp hiệu quả*.
Snowling, M. J., & Hulme, C. (2021). Language Disorders in Children: Recent Advances in Research and Practice. *Journal of Child Psychology and Psychiatry*, 62(5), 523–540.
Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? *The case of developmental language disorder (DLD)*. *International Journal of Language & Communication Disorders*, 52(6), 671–680.
Bộ Giáo dục và Đào tạo Việt Nam. (2022). *Hướng dẫn đánh giá sự phát triển ngôn ngữ ở trẻ mầm non*.
Paul, R., & Norbury, C. (2012). *Language Disorders from Infancy through Adolescence*. Mosby Elsevier.
Bộ Giáo dục và Đào tạo Việt Nam. (2021). *Hướng dẫn hỗ trợ trẻ chậm nói tại trường mầm non*.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. Pearson.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*