

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG NHẬN DẠNG GIỌNG NÓI Ở BỆNH NHÂN PARKINSON

Phạm Thị Bích Đào^{1,2,✉}, Nguyễn Diệu My^{1,2}, Nguyễn Thị Hằng^{1,2}
Nguyễn Thị Anh Đào^{1,2}, Ngô Thị Ngọc³, Nguyễn Thị Thái Chung^{1,2}
Phan Xuân Nam^{1,2}, Lê Minh Đạt^{1,2}, Nguyễn Quang Quyền^{1,2}

¹Bệnh viện Đại học Y Hà Nội

²Trường Đại học Y Hà Nội

³Bệnh viện Đa khoa tỉnh Quảng Trị

Nghiên cứu ứng dụng trí tuệ nhân tạo (AI) trong nhận dạng giọng của bệnh nhân Parkinson (PD) so với người khỏe mạnh, dữ liệu tiếng Việt. Nghiên cứu cắt ngang, thực hiện tại Bệnh viện Đại học Y Hà Nội từ năm 2024 đến 2025, với 20 bệnh nhân được chẩn đoán PD và 60 người đối chứng khỏe mạnh. Giọng nói được ghi âm: kéo dài nguyên âm /a/, /i/, /u/ (≥ 3 giây, lặp lại 3 lần), đọc câu chuẩn 25 âm tiết và nói tự do trong 30 giây. Đánh giá các đặc trưng âm học gồm jitter, shimmer, độ hài thanh (HNR), tần số cơ bản (F0) trung bình/ độ lệch chuẩn và chỉ số âm học (MFCC). Huấn luyện hai mô hình AI: Máy vector hỗ trợ với hàm nhân cơ sở (SVM - RBF kernel) và Mạng nơ-ron tích chập trên biểu phổ Mel, đánh giá bằng hệ kiểm định chéo 5 phần. Kết quả: Mô hình CNN đạt độ chính xác với AUC 0,91, độ nhạy 88% và độ đặc hiệu 84%.

Từ khóa: Parkinson, giọng nói, trí tuệ nhân tạo, CNN, MFCC, tiếng Việt.

I. ĐẶT VẤN ĐỀ

Rối loạn giọng nói là biểu hiện không vận động phổ biến ở bệnh Parkinson (PD), ảnh hưởng đến khoảng 60 - 90% người bệnh trong tiến trình bệnh, bao gồm các triệu chứng như đơn điệu cao độ và cường độ, giảm âm lượng, run giọng, nói lắp và rối loạn nhịp điệu.¹⁻³ Những rối loạn này gây ảnh hưởng nghiêm trọng đến khả năng giao tiếp và chất lượng cuộc sống của bệnh nhân. Các phương pháp đánh giá truyền thống dựa trên cảm quan chuyên gia hoặc các chỉ số âm học đơn biến như jitter, shimmer và tỷ số tín hiệu trên nhiễu (HNR) thường phụ thuộc nhiều vào người đánh giá và có độ nhạy thấp trong việc phát hiện những thay đổi sớm ở giai đoạn đầu bệnh.^{4,5}

Sự phát triển của học máy và trí tuệ nhân tạo (AI) đã cho phép trích xuất và khai thác các đặc trưng phổ-thời gian phi tuyến từ giọng nói, giúp phát hiện sớm và theo dõi tiến triển PD thông qua các dấu ấn giọng nói (voice biomarkers).⁶⁻¹⁰ Các nghiên cứu ứng dụng mạng nơ-ron tích chập (CNN) và kiến trúc Transformer trên dữ liệu biểu phổ Mel (Mel-spectrogram) và hệ số Mel-frequency cepstral coefficients (MFCC) cho kết quả phân loại PD với độ chính xác cao, đồng thời hỗ trợ theo dõi đáp ứng điều trị levodopa và kích thích não sâu.¹¹⁻¹⁵ Tuy nhiên, hầu hết các nghiên cứu này tập trung vào dữ liệu tiếng Anh hoặc các ngôn ngữ phương Tây, trong khi dữ liệu tiếng Việt vẫn còn rất hạn chế.

Bên cạnh đó, tiếng Việt là ngôn ngữ đơn âm nhưng mang hệ thống thanh điệu phức tạp, bao gồm 6 thanh điệu với sự biến thiên lớn về cao độ, độ dốc, thời lượng và cường độ. Những đặc điểm này tạo ra dấu ấn âm học rất khác so với các ngôn ngữ Ấn - Âu, khiến quá

Tác giả liên hệ: Phạm Thị Bích Đào

Bệnh viện Đại học Y Hà Nội

Email: Daoptn0024@hmu.vn

Ngày nhận: 16/10/2025

Ngày được chấp nhận: 04/12/2025

trình nhận dạng giọng nói bệnh lý trở nên khó hơn. Ngoài ra, tiếng Việt còn có sự đa dạng vùng miền rõ rệt (Bắc - Trung - Nam), tạo nên khác biệt về khẩu hình, tốc độ nói và nhịp ngắt, ảnh hưởng trực tiếp đến các đặc trưng như F0, jitter, shimmer hay MFCC.^{6,9}

Do đó, khi áp dụng mô hình trí tuệ nhân tạo để nhận dạng giọng nói bệnh lý, việc xây dựng và huấn luyện mô hình trên dữ liệu tiếng Việt bản địa là đặc biệt quan trọng. Các mô hình được huấn luyện từ dữ liệu tiếng Anh hoặc các ngôn ngữ không thanh điệu thường không học được đầy đủ cấu trúc âm học đặc hữu của tiếng Việt, dẫn đến suy giảm độ chính xác. Đây chính là lý do việc phát triển mô hình AI phù hợp với ngôn ngữ tiếng Việt - có thanh điệu, biến thiên vùng miền và đặc điểm phát âm riêng - mang ý nghĩa khoa học và thực tiễn sâu sắc, góp phần nâng cao khả năng phát hiện sớm bất thường giọng nói ở bệnh nhân Parkinson tại Việt Nam.¹⁰

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Nghiên cứu được thực hiện trên 20 bệnh nhân Parkinson (PD) được chẩn đoán theo tiêu chuẩn MDS và 60 người đối chứng khỏe mạnh phù hợp về tuổi và giới. Người tham gia được thu âm giọng nói theo bộ nhiệm vụ chuẩn hóa gồm: kéo dài nguyên âm /a/, /i/, /u/ ≥ 3 giây (lặp lại 3 lần), đọc câu chuẩn 25 âm tiết chứa đủ 6 thanh điệu tiếng Việt và nói tự do 30 giây.

Đối với bệnh nhân PD, thời điểm thu âm được ghi nhận theo trạng thái dùng thuốc (ON/OFF levodopa). Những trường hợp giọng nói thay đổi rõ rệt ngay sau dùng thuốc hoặc đang trong giai đoạn dao động liều không ổn định được loại trừ để giảm nhiễu lên đặc trưng âm học.

Tiêu chuẩn loại trừ

Bệnh lý thanh quản, rối loạn phát âm bẩm sinh, viêm thanh quản cấp, trào ngược họng-

thanh quản nặng, tiền sử phẫu thuật vùng đầu-cổ, hoặc bất kỳ tình trạng nào gây ảnh hưởng độc lập đến giọng nói ngoài PD.

2. Phương pháp

Thiết kế nghiên cứu

Nghiên cứu cắt ngang mô tả, đánh giá khả năng phân loại PD bằng mô hình AI, tuân thủ hướng dẫn STARD.

Thời gian và địa điểm

Nghiên cứu được tiến hành tại Bệnh viện Đại học Y Hà Nội từ năm 2024 đến 2025.

Thu thập và ghi âm giọng nói

Ghi âm được thực hiện trong phòng yên tĩnh ($< 35\text{dB}$), dùng micro condenser đặt cách miệng 20cm, 44,1kHz - 16 bit - mono. Dữ liệu được chuẩn hóa RMS, loại bỏ khoảng lặng bằng voice activity detection và khử nhiễu bằng spectral gating:

+ Kéo dài nguyên âm /a/, /i/, /u/ trong ≥ 3 giây, mỗi nguyên âm lặp lại 3 lần.

+ Đọc câu chuẩn 25 âm tiết bao gồm đủ 6 thanh điệu tiếng Việt.

+ Nói tự do trong 30 giây.

Xử lý tín hiệu

Tín hiệu âm thanh được chuẩn hóa theo Root Mean Square (RMS), loại bỏ đoạn lặng và khử nhiễu bằng kỹ thuật *spectral gating*.

- Phân tích đặc trưng âm học: Các đặc trưng gồm jitter, shimmer, HNR, F0 trung bình/SD được trích xuất bằng Praat; đặc trưng MFCC (13 hệ số + delta & delta-delta) được tính từ toàn bộ nhiệm vụ phát âm.

Chia dữ liệu huấn luyện - kiểm định

Để tăng tính minh bạch và giảm sai lệch do cỡ mẫu nhỏ, nghiên cứu chia dữ liệu theo hai chiến lược:

- Kiểm định chéo phân tầng 5 phần (5-fold stratified cross-validation) cho quá trình huấn luyện - tối ưu siêu tham số.

- Giữ lại 20% dữ liệu độc lập (hold-out test) để đánh giá cuối cùng. Tỷ lệ phân bố mẫu PD/đối chứng được giữ nguyên trong từng tập con nhằm tránh lệch phân lớp.

Mô hình AI

- Hai mô hình được so sánh:

+ SVM (RBF kernel): mô hình đơn giản, ít tham số, được tối ưu bằng grid search.

+ CNN: gồm 4 khối convolution-batchnorm-ReLU-max pooling, dropout 0,3; tối ưu bằng Adam, áp dụng early stopping dựa trên tập validation của từng fold.

Do CNN có số tham số lớn so với cỡ mẫu (80 người), nguy cơ overfitting được lưu ý. Mặc dù đã dùng dropout, cross-validation và early stopping, mô hình vẫn có khả năng học thuộc dữ liệu khi số lượng mẫu ít. Sự vượt trội của CNN (AUC 0,91) so với SVM (AUC 0,82) cần được diễn giải thận trọng và kiểm chứng bằng tập dữ liệu độc lập quy mô lớn hơn.

- Đánh giá mô hình

+ Các chỉ số đánh giá gồm: AUC ROC, độ nhạy, độ đặc hiệu, F1-score. So sánh AUC giữa mô hình thực hiện bằng kiểm định DeLong. Đặc trưng quan trọng được giải thích bằng SHAP (SVM) và Grad-CAM (CNN).

Chiến lược chia dữ liệu Train - Validation

- Test

Phản biện yêu cầu làm rõ chiến lược chia dữ liệu; mục này đã được viết lại như sau:

Dữ liệu được chia theo hai tầng đánh giá độc lập:

(1) Giai đoạn huấn luyện và tối ưu hóa mô hình

Sử dụng 5-fold stratified cross-validation nhằm đảm bảo tỉ lệ PD/đối chứng được giữ nguyên trong từng fold. Mỗi vòng gồm:

- 4 phần cho huấn luyện (*training set*).

- 1 phần cho hiệu chỉnh siêu tham số (*validation set*).

Các kỹ thuật được áp dụng trong giai đoạn này gồm dropout, batch normalization, early stopping, nhằm giảm nguy cơ overfitting.

(2) Giai đoạn đánh giá cuối cùng

Sau khi tối ưu mô hình, 20% dữ liệu được giữ lại hoàn toàn độc lập (*independent hold-out test set*) để đánh giá hiệu năng cuối cùng của từng mô hình.

Chiến lược này giúp giảm thiểu sai lệch do cỡ mẫu nhỏ và tăng tính khách quan khi so sánh giữa các mô hình.

3. Đạo đức nghiên cứu

Tất cả người tham gia đều ký cam kết đồng ý tham gia; nghiên cứu không xâm lấn và được thực hiện theo đúng quy trình đạo đức y sinh.

III. KẾT QUẢ

Bảng 1. Đặc điểm mẫu nghiên cứu

Đối tượng	Số lượng	Tuổi trung bình (mean $\pm$ SD)	Giới tính (Nam/Nữ)	Giai đoạn Hoehn-Yahr (PD)
Bệnh nhân Parkinson	20	65,4 $\pm$ 7,2	12/8	I (30%), II (50%), III (20%)
Người đối chứng	60	64,1 $\pm$ 6,8	36/24	-

Mẫu nghiên cứu gồm 20 bệnh nhân PD giai đoạn từ I đến III và 60 người đối chứng khỏe mạnh, được cân bằng về tuổi và giới tính, đảm bảo tính đại diện cho nhóm nghiên cứu.

Bảng 2. Các nhiệm vụ ghi âm và đặc trưng âm học trích xuất

Giọng nói	Mô tả	Đặc trưng âm học
Kéo dài nguyên âm /a/, /i/, /u/	≥ 3 giây, lặp lại 3 lần	Jitter (local, RAP), Shimmer (local, APQ), HNR, F0 mean/SD
Đọc câu chuẩn 25 âm tiết	Bao gồm đủ 6 thanh điệu tiếng Việt	MFCC (13 hệ số + delta, delta-delta), các chỉ số trên
Nói tự do trong 30 giây	Tự do nói chuyện	Các đặc trưng âm học tổng hợp

Các nhiệm vụ được thiết kế đa dạng, bao phủ từ nguyên âm kéo dài đến câu chuẩn và nói tự do, nhằm khai thác tối đa các đặc trưng âm học liên quan đến rối loạn giọng nói của PD.

Bảng 3. Hiệu năng mô hình phân loại Parkinson vs. người khỏe mạnh

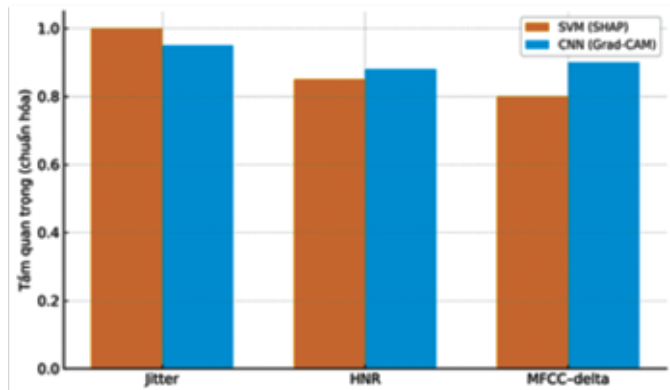
Mô hình	AUC (95% CI)	Độ nhạy (%)	Độ đặc hiệu (%)	Điểm F1	p-value (so sánh AUC với CNN)
CNN	0,91 (0,85 - 0,96)	88	84	0,86	-
SVM (RBF)	0,82 (0,75 - 0,89)	75	78	0,76	0,02

Mô hình CNN thể hiện hiệu suất vượt trội so với SVM trên tập dữ liệu tiếng Việt, với AUC cao, độ nhạy và độ đặc hiệu tốt, phù hợp cho nhận dạng giọng nói bất thường ở PD.

Bảng 4. Hiệu năng phân loại theo từng nhiệm vụ giọng nói (AUC)

Nhiệm vụ giọng nói	AUC
Kéo dài nguyên âm /a/	0,90
Kéo dài nguyên âm /i/	0,85
Kéo dài nguyên âm /u/	0,83
Đọc câu chuẩn 25 âm tiết	0,89
Nói tự do	0,80

Tác vụ kéo dài nguyên âm /a/ và đọc câu chuẩn có hiệu quả phân loại cao nhất, chứng tỏ đây là các nhiệm vụ phù hợp để phát hiện rối loạn giọng nói do PD trong tiếng Việt.

Bảng 5. Các đặc trưng âm học quan trọng được xác định qua giải thích mô hình**Đặc trưng âm học****Biểu đồ 1. Các đặc trưng âm học của nhóm đối tượng nghiên cứu**

Các biến Jitter, HNR và hệ số MFCC delta được xác định là những yếu tố quyết định trong việc phân loại, phản ánh rõ sự khác biệt trong đặc trưng giọng nói của bệnh nhân PD so với người khỏe mạnh.

IV. BÀN LUẬN

Nghiên cứu cho thấy mô hình mạng nơ-ron tích chập (CNN) đạt hiệu năng phân loại giọng nói bệnh nhân Parkinson cao, với giá trị AUC 0,91, vượt trội so với mô hình máy vector hỗ trợ (SVM) có AUC 0,82. Hiệu năng này tương đồng với các công trình quốc tế ứng dụng học sâu trong nhận dạng giọng nói của bệnh nhân Parkinson, với AUC thường dao động trong khoảng 0,85 - 0,93.^{1,3,6,12} Kết quả củng cố giả thuyết rằng các mô hình học sâu có khả năng tự động học các đặc trưng phổ - thời gian phức tạp từ tín hiệu giọng nói, phản ánh những dao động không đều của thanh môn và sự suy giảm phối hợp thần kinh - vận động trong quá trình phát âm của người bệnh. Ngược lại, SVM dựa trên các đặc trưng âm học thủ công như Jitter, shimmer, HNR hay MFCC chỉ mô tả được một phần sự biến thiên của tín hiệu, dẫn đến khả năng nhận dạng thấp hơn.

Ở các tác vụ ghi âm, việc phát âm kéo dài nguyên âm /a/ và đọc câu chuẩn cho hiệu quả phân loại cao nhất, trong khi nói tự do có độ biến thiên lớn hơn nên cho AUC thấp hơn. Điều này có thể lý giải bởi các tác vụ có cấu trúc ổn định giúp giảm nhiễu ngôn ngữ và ổn định cao độ, cường độ, thuận lợi cho việc trích xuất đặc trưng âm học. Kết quả phù hợp với nhiều nghiên cứu quốc tế chứng minh rằng các bài phát âm có cấu trúc rõ ràng, như kéo dài nguyên âm hoặc đọc câu cố định, là công cụ hiệu quả để phát hiện sớm rối loạn giọng nói trong Parkinson.^{4,7,10,11,13}

Từ góc nhìn lâm sàng, mô hình trí tuệ nhân tạo, đặc biệt CNN, có thể được ứng dụng trong sàng lọc và theo dõi bệnh nhân Parkinson từ xa thông qua các nền tảng di động, góp phần phát hiện sớm và giảm gánh nặng khám trực tiếp.^{9,14} Hệ thống còn có thể hỗ trợ đánh giá đáp ứng điều trị bằng cách theo dõi sự thay đổi các chỉ số âm học như jitter, HNR, hay MFCC-delta trong quá trình sử dụng levodopa hoặc kích thích não sâu.^{2,6,11} Ngoài ra, việc áp dụng mô hình nhận dạng giọng nói có thể hỗ trợ bác sĩ trong việc quyết định chỉ định các liệu pháp giọng nói chuyên sâu như LSVT LOUD, hoặc đánh giá

hiệu quả phục hồi giọng sau can thiệp.^{1,6}

Hiệu năng mô hình trong nghiên cứu này nằm trong khoảng giá trị đã công bố quốc tế, cho thấy tính khả thi khi áp dụng trên dữ liệu tiếng Việt - một ngôn ngữ có thanh điệu phức tạp và đặc thù vùng miền.^{3,7,10} Tuy vậy, nghiên cứu vẫn còn hạn chế về cỡ mẫu nhỏ và tiến hành tại một trung tâm, có thể ảnh hưởng đến tính đại diện của kết quả. Trạng thái ON/OFF thuốc của bệnh nhân chưa được chuẩn hóa tại thời điểm ghi âm, trong khi yếu tố này đã được chứng minh có thể tác động đến giọng nói. Ngoài ra, sự khác biệt vùng miền và tiếng ồn môi trường thực tế chưa được kiểm soát chặt chẽ, có thể làm giảm tính ổn định của mô hình.^{4,13,15}

Hướng nghiên cứu tiếp theo cần mở rộng quy mô đa trung tâm, thu thập dữ liệu dọc để đánh giá sự biến đổi giọng nói theo thời gian và chuẩn hóa bộ câu đọc nhằm bao phủ đầy đủ các thanh điệu tiếng Việt. Việc kết hợp các kỹ thuật *domain adaptation* và lọc nhiễu nâng cao sẽ giúp cải thiện tính ứng dụng của mô hình trong điều kiện thực tế ngoài bệnh viện.^{5,9,12}

Nghiên cứu này khẳng định tiềm năng của trí tuệ nhân tạo, đặc biệt mạng CNN, trong nhận dạng và theo dõi rối loạn giọng nói ở bệnh nhân Parkinson, mở ra hướng tiếp cận mới cho chẩn đoán sớm và quản lý bệnh bằng công nghệ ngôn ngữ tiếng Việt.

Mặc dù mô hình CNN cho kết quả khả quan, nghiên cứu còn hạn chế ở việc thiếu hình minh chứng SHAP/Grad-CAM, chưa phân tích sâu ảnh hưởng của thanh điệu tiếng Việt đến đặc trưng âm học và chưa mở rộng thảo luận về khả năng triển khai thực tế trên thiết bị di động. Việc bổ sung các nội dung này sẽ giúp mô hình có giá trị ứng dụng cao hơn trong sàng lọc và theo dõi bệnh Parkinson tại Việt Nam.

Nghiên cứu còn một số hạn chế như cỡ mẫu nhỏ dẫn đến nguy cơ overfitting của mô hình CNN, thiếu kiểm định trên dữ liệu tiếng ồn và

các tình huống ghi âm thực tế, và chưa có kiểm chứng đa vùng - đa giọng trong tiếng Việt. Các nghiên cứu tiếp theo cần mở rộng dữ liệu đa trung tâm, bổ sung tập kiểm định độc lập và đánh giá khả năng chống nhiễu của mô hình.

V. KẾT LUẬN

Mô hình CNN đạt độ chính xác cao với AUC 0,91, độ nhạy 88% và độ đặc hiệu 84%, vượt trội hơn so với SVM, cho thấy tiềm năng mạnh mẽ trong nhận dạng rối loạn giọng nói của bệnh nhân Parkinson trên dữ liệu tiếng Việt. Kết quả này mở ra triển vọng ứng dụng AI trong sàng lọc và theo dõi lâm sàng bệnh Parkinson hiệu quả và thuận tiện.

LỜI CẢM ƠN

Chúng tôi xin chân thành cảm ơn tất cả các bệnh nhân và người đối chứng đã tham gia nghiên cứu. Các tác giả tuyên bố không có xung đột lợi ích liên quan đến nội dung bài báo.

TÀI LIỆU THAM KHẢO

1. Little MA, McSharry PE, Hunter EJ, et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans Biomed Eng.* 2009; 56(4): 1015-22. doi:10.1109/TBME.2008.2002464.
2. Tsanas A, Little MA, McSharry PE, et al. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng.* 2012; 59(5): 1264-71. doi:10.1109/TBME.2011.2170335.
3. Orozco-Arroyave JR, Hönl F, Arias-Londoño JD, et al. Automatic detection of Parkinson's disease from continuous speech. *J Neurolinguistics.* 2016; 37: 141-60. doi:10.1016/j.jneuroling.2016.11.005.
4. Sakar CO, Serbes G, Gunduz A, et al. A comparative analysis of speech signal processing algorithms for Parkinson's

disease classification. *IEEE J Biomed Health Inform.* 2013; 17(4): 828-34. doi:10.1109/JBHI.2013.2245674.

5. Rusz J, Cmejla R, Ruzickova H et al. Quantitative acoustic measurements for characterization of speech and voice disorders in early PD. *J Acoust Soc Am.* 2011; 129(1): 350-67. doi:10.1121/1.3514381.

6. Al-Fatlawi A, et al. Deep CNNs for Parkinson's disease detection using spectrograms. *Comput Methods Programs Biomed.* 2021; 199: 105915. doi:10.1016/j.cmpb.2020.105915.

7. Vásquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, et al. Convolutional neural networks and transfer learning to classify PD from speech in three languages. *Lect Notes Comput Sci.* 2019; 11896: 697-708. doi:10.1007/978-3-030-26626-4_68.

8. Shen M, et al. Explainable AI to diagnose early Parkinson's using voice. *Sci Rep.* 2025; 15: 5487. doi:10.1038/s41598-025-96575-6.

9. Reddy GP, et al. Artificial intelligence-based effective detection of Parkinson's disease using voice. *Electronics.* 2024; 13(20): 4028. doi:10.3390/electronics13204028.

10. Orozco-Arroyave JR, Vásquez-Correa JC, Arias-Vergara T, et al. Apkinson: The smartphone application to monitor PD patients in the wild. *Sensors.* 2019; 19(21): 4802. doi:10.3390/s19214802.

11. Rusz J, Krack P, Tripoliti E. Digital speech biomarkers in PD: from prodromal stages to clinical trials. *Neurosci Biobehav Rev.* 2024; 156: 105922. doi:10.1016/j.neubiorev.2023.105922.

12. Tsanas A, et al. Novel speech signal processing algorithms for PD monitoring. *IEEE Trans Biomed Eng.* 2014; 61(8): 2180-94. doi:10.1109/TBME.2014.2314574.

13. Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S. Speech impairment in PD: a review. *Mov Disord.* 1998; 13(5): 676-88. doi:10.1002/mds.870130407.

14. Skodda S, Flasskamp A, Schlegel U. Instability of syllable repetition in PD measured by acoustic analysis. *J Voice.* 2011; 25(6): e265-75. doi:10.1016/j.jvoice.2010.06.009.

15. Dehak N, Kenny P, Dehak R et al. Front-end factor analysis for speaker verification. *IEEE Trans Audio Speech Lang Process.* 2011; 19(4): 788-98. doi:10.1109/TASL.2010.2064307.

Summary

APPLICATION OF ARTIFICIAL INTELLIGENCE IN VOICE RECOGNITION FOR PARKINSON'S DISEASE

This study applied artificial intelligence (AI) to recognize the voices of patients with Parkinson's disease (PD) compared with healthy controls using Vietnamese-language data. A cross-sectional study was conducted at Hanoi Medical University Hospital from 2024 to 2025, involving 20 diagnosed PD patients and 60 healthy controls. Voice samples were recorded, including sustained vowels /a/, /i/, /u/ (≥ 3 seconds, repeated three times), reading a standardized 25-syllable sentence, and 30 seconds of spontaneous speech. Acoustic features analyzed included jitter, shimmer, harmonic-to-noise ratio (HNR), mean and standard deviation of the fundamental frequency (F0), and Mel-frequency cepstral coefficients (MFCC). Two AI models were trained: a Support Vector Machine with a radial basis function kernel (SVM-RBF) and a Convolutional Neural Network (CNN) applied to Mel spectrograms. Performance was evaluated using 5-fold cross-validation. Results: The CNN model achieved an accuracy with an AUC of 0.91, sensitivity of 88%, and specificity of 84%.

Keywords: Parkinson's disease, voice, artificial intelligence, CNN, MFCC, Vietnamese.