

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO (AI) VÀ NGHIỆM PHÁP CHỨC NĂNG THĂNG BẰNG TRONG CHẨN ĐOÁN BỆNH TIỀN ĐÌNH

Đỗ Trâm Anh^{1,✉}, Hiromasa Takakura², Masatsugu Asai²
Naoko Ueda², Hideo Shojaku²

¹Bệnh viện Quân Y 103

²Đại học Toyama, Nhật Bản

Bệnh tiền đình ngoại biên (peripheral vestibular disease - PV) là nguyên nhân phổ biến gây chóng mặt. Tuy nhiên cần nhiều thời gian thăm khám lâm sàng, đánh giá toàn diện và phân biệt với bệnh không phải tiền đình ngoại biên (non-peripheral vestibular disease: non-PV) để đưa ra chẩn đoán chính xác. Non-PV bao gồm nhiều loại bệnh khác nhau, trong đó chóng mặt do hạ huyết áp tư thế (hemodynamic orthostatic dizziness/vertigo - HO) chiếm số lượng lớn bệnh nhân trong tập dữ liệu. Mục tiêu nghiên cứu này nhằm xác định khả năng ứng dụng máy học, trí tuệ nhân tạo (AI) trong việc chẩn đoán và phân loại bệnh HO với nhóm PV và non-PV. Các mô hình phân loại đa lớp được áp dụng trên 1009 bệnh nhân (497 PV, 157 HO và 355 non-PV) cho thấy tính đúng (accuracy) cao nhất đạt được là 72% và điểm F1 của nhóm PV, non-PV và HO lần lượt là 0,78; 0,64; 0,71. Kết quả cho thấy AI hoàn toàn có thể trở thành công cụ hữu ích trong thực hành lâm sàng giúp tiết kiệm thời gian và nâng cao độ chính xác, hiệu quả chẩn đoán bệnh.

Từ khóa: Máy học, nghiệm pháp chức năng thăng bằng, tiền đình ngoại biên, chóng mặt do hạ huyết áp tư thế.

I. ĐẶT VẤN ĐỀ

Chóng mặt là một trong những triệu chứng khiến bệnh nhân thường xuyên phải đến gặp bác sĩ. Tuy nhiên, việc xác định nguyên nhân chính xác của chóng mặt rất phức tạp vì có rất nhiều loại bệnh liên quan đến chúng. Rối loạn chức năng hệ thống tiền đình ngoại biên (PV) là một trong những nguyên nhân phổ biến nhất gây chóng mặt, như chóng mặt tư thế kịch phát lành tính, viêm dây thần kinh tiền đình và bệnh Meniere.¹ Tiêu chuẩn chẩn đoán bệnh PV chủ yếu dựa vào tiền sử bệnh nhân và các xét nghiệm khám lâm sàng hệ thống tiền đình, kiểm tra vận nhãn và hệ thống tiểu não theo các tiêu chuẩn chẩn đoán định hướng lâm sàng của

Hiệp hội Bárány.²

Trong nghiên cứu này, chúng tôi tiến hành các bài kiểm tra chức năng thăng bằng khác nhau, như bài kiểm tra nhiệt lượng và rung giật nhãn cầu tự phát trên những bệnh nhân chưa có chẩn đoán xác định, điều trị không hiệu quả hoặc triệu chứng chóng mặt, mất thăng bằng tồn tại kéo dài. Ngoài ra, các xét nghiệm chẩn đoán hình ảnh như cộng hưởng từ não (MRI), chụp cắt lớp vi tính não (CT scan), CT xương thái dương và siêu âm mạch máu cổ cũng được thực hiện ở hầu hết các bệnh nhân. Do vậy cần nhiều thời gian để đánh giá toàn diện và phân biệt với bệnh không phải tiền đình ngoại biên (non-PV) để đưa ra chẩn đoán chính xác.

Trí tuệ nhân tạo (AI) đang là xu hướng nổi bật trong mọi lĩnh vực toàn cầu, bao gồm cả lĩnh vực y học. Nhiều ứng dụng AI trong y tế đã ra đời để giảm tải gánh nặng cho các bệnh viện, cải thiện trình độ chuyên môn và hiệu quả công

Tác giả liên hệ: Đỗ Trâm Anh

Bệnh viện Quân Y 103

Email: tramanhdo3110@gmail.com

Ngày nhận: 07/08/2024

Ngày được chấp nhận: 28/08/2024

việc lâm sàng. AI trong y học là thuật ngữ tổng quát mô tả việc áp dụng thuật toán như máy học (Machine Learning) để phân tích và hành động dựa trên dữ liệu y học, thường nhằm mục đích dự đoán một kết quả cụ thể. Ngày nay, việc áp dụng AI đang phát triển nhanh chóng, đặc biệt là chẩn đoán hình ảnh như X-quang, hệ thống phẫu thuật robot và chẩn đoán bệnh trên lâm sàng.^{3,4} Gần đây, máy học đã được nghiên cứu như một phương pháp phần mềm hữu ích để hỗ trợ đưa ra quyết định y tế đối với bệnh rối loạn chức năng tiền đình.⁵ Những nghiên cứu này cho thấy máy học đang trở thành một công cụ tiềm năng giúp các bác sĩ xử lý và sử dụng hiệu quả nhất lượng thông tin lớn từ bệnh nhân, giúp tiết kiệm thời gian để đưa ra quyết định chẩn đoán chính xác.

Nghiên cứu trước đây của chúng tôi dựa trên kết quả khám chức năng thăng bằng của 1009 bệnh nhân cho thấy máy học có khả năng chẩn đoán bệnh tiền đình ngoại biên (PV) và bệnh tiền đình không ngoại biên (non-PV) với tính đúng là 79% bởi một mô hình và khi kết hợp đồng thời nhiều mô hình khác nhau, khả năng dự đoán đúng bệnh PV tăng lên 83% và bệnh non-PV là 85%. Nghiên cứu lần này được

thực hiện nhằm đánh giá khả năng hỗ trợ của AI khi áp dụng mô hình phân loại đa lớp nhằm phân loại một mặt bệnh cụ thể như chóng mặt do hạ huyết áp tư thế (hemodynamic orthostatic dizziness/vertigo - HO) với hai nhóm bệnh PV và non-PV.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Nghiên cứu được thực hiện trên 1009 bệnh nhân (bao gồm 497 PV, 157 HO và 355 non-PV). Tất cả các bệnh nhân đều được chẩn đoán bệnh theo tiêu chuẩn phân loại quốc tế về rối loạn tiền đình của Hiệp hội Bárány và hướng dẫn của Hiệp hội nghiên cứu thăng bằng Nhật Bản (Bảng 1).^{6,7}

Các bài kiểm tra cân bằng của chúng tôi được thực hiện đối với các bệnh nhân ngoại trú đến khám tại khoa với triệu chứng chóng mặt, không bao gồm bệnh nhân cấp cứu. Ngoài ra, chúng tôi còn khám một số bệnh nhân sau nhập viện từ nhiều khoa khác trong bệnh viện, chẳng hạn như khoa cấp cứu, nội thần kinh và phẫu thuật thần kinh, hay bệnh nhân từ các bệnh viện khác.

Bảng 1. Đặc điểm và chẩn đoán lâm sàng của bệnh nhân

Bệnh tiền đình ngoại biên (PV)	n = 497
Tuổi, trung bình năm (khoảng):	55,9 (7 - 93)
Giới tính, n (%):	
Nam	220 (44,3)
Nữ	277 (55,7)
Chẩn đoán, n (%):	
U dây VIII (Koos I/II)	18 (3,6)
Chóng mặt kịch phát lành tính	44 (8,8)
Bệnh lý tiền đình 2 bên	2 (0,4)

Bệnh tiền đình ngoại biên (PV)	n = 497
Cholesteatoma/ Viêm tai giữa mạn tính	29 (5,8)
Phù nội dịch chậm (DEH)	19 (3,8)
Liệt dây VII / Hội chứng Hunt	13 (2,6)
Rối loạn tai trong	156 (31,4)
Bệnh Meniere	120 (24,1)
Rò ngoại dịch	4 (0,8)
Điếc đột ngột kèm chóng mặt	44 (8,8)
Viêm dây thần kinh tiền đình	48 (9,6)
Bệnh không phải tiền đình ngoại biên (non-PV)	n = 512
Tuổi, trung bình năm (khoảng):	55,3 (7 - 91)
Giới tính, n (%):	
Nam	391 (76,4)
Nữ	121 (23,6)
Chẩn đoán, n (%):	
Nhồi máu/ chảy máu não	20 (3,9)
U não	20 (4)
Rối loạn thăng bằng trung ương	103 (20,1)
Rung giật nhãn cầu bẩm sinh	8 (1,5)
Hội chứng Disembarkment	3 (0,6)
Chóng mặt không rõ nguyên nhân	32 (6,2)
Chấn thương vùng đầu	9 (1,7)
Chóng mặt do hạ huyết áp tư thế (HO)	157 (30)
Đau nửa đầu	6 (1,2)
Bệnh thần kinh trung ương khác (n < 5)	23 (4,49)
Parkinson	5 (1)
Chóng mặt nhận thức-tư thế dai dẳng	23 (4,5)
Chóng mặt do bệnh lý tâm thần kinh	33 (6,4)
Thoái hoá tuỷ sống	22 (4,3)
Thiếu năng động mạch đốt sống thân nền	48 (9,3)

2. Phương pháp

Thiết kế nghiên cứu

Nghiên cứu hồi cứu, mô tả.

Thời gian nghiên cứu

Từ năm 2009 đến năm 2019.

Địa điểm nghiên cứu

Khoa Tai mũi họng và Phẫu thuật đầu cổ, Đại học Toyama, Nhật Bản.

Các bước tiến hành nghiên cứu

Kiểm tra chức năng thính bằng:

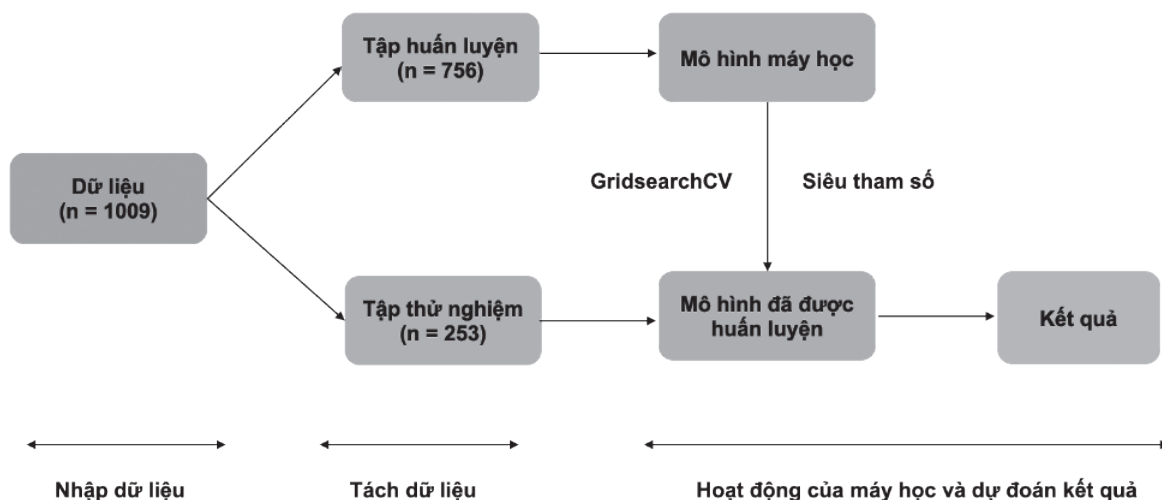
Tất cả các bệnh nhân được trải qua 16 bài kiểm tra chức năng thính bằng tiêu chuẩn.⁸ Các bài kiểm tra này bao gồm 44 đặc tính, chia thành hai loại: biến liên tục và phân loại. Các đặc tính liên tục với các giá trị dưới dạng số và các tính năng phân loại được mã hóa dưới dạng số nguyên từ 0 đến 3. Ngoài ra, bảng câu hỏi kiểm kê điểm chóng mặt “Dizziness Handicap Inventory” (DHI) cũng được tiến hành trên mỗi

bệnh nhân.

Ứng dụng máy học phân loại bệnh:

Trong nghiên cứu hiện tại, chúng tôi đã sử dụng các mô hình máy học có giám sát để thực hiện phân loại, nhằm mục đích dự đoán các loại quan sát mới dựa trên dữ liệu tập huấn luyện đã biết. Chương trình được tạo trên “Google Colaboratory” sử dụng phiên bản Python (v) 3.7.12, scikit-learn v1.0.2, NumPy v1.21.5, SciPy v1.4.1, Pandas v1.3.5 và Matplotlib v3.2.2.⁹ Bảy thuật toán phổ biến nhất bao gồm Decision Tree (DT), AdaBoost (AB), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Gradient Boosting (GB) và KNeighbors (KN) đã được áp dụng.

Tổng quan quá trình hoạt động của máy học bao gồm các bước: nhập dữ liệu, tách dữ liệu, đào tạo mô hình máy học tốt nhất để dự đoán kết quả được thực hiện như Sơ đồ 1.



Sơ đồ 1. Tổng quan quá trình hoạt động của máy học

Các phương pháp đánh giá kết quả thử nghiệm

Trong phân loại nhị phân, một trong hai nhóm được dự đoán được gọi là nhóm âm (negative: N) và nhóm còn lại được gọi là nhóm dương (positive: P). Đối với mô hình đa phân

loại, khi nhóm bệnh cần đánh giá được gọi là P, hai nhóm còn lại được coi là N.

- Ma trận nhầm lẫn (confusion matrix) thường được sử dụng để đánh giá khả năng chẩn đoán của bộ phân loại.¹⁰ Bảng 2 hiển thị kết quả dự đoán theo từng mô hình trong mỗi bốn loại:

True Positive: TP (dương tính thật), False Positive: FP (dương tính giả), False Negative: FN (âm tính giả) và True Negative: TN (âm tính thật).

- Phương pháp đánh giá kết quả kiểm tra được sử dụng để đánh giá hiệu suất dự đoán của máy học như sau: tính đúng (accuracy),

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

tính chính xác (precision), độ nhạy (recall), hệ số điểm F1 (F1-score) là trung bình điều hoà giữa tính chính xác và độ nhạy. Các phương pháp trên được hiển thị dưới dạng các giá trị số nằm trong khoảng từ 0 đến 1. Giá trị càng lớn thì hiệu suất dự đoán càng cao.

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Xử lý số liệu

Thử nghiệm Mann-Whitney U được sử dụng để đánh giá thống kê về độ tuổi, tính chính xác, độ nhạy, và điểm F1 giữa hai nhóm bệnh PV và không PV. Kiểm định Chi bình phương được sử dụng để đánh giá thống kê tỷ lệ giới tính. BellCurve for Excel v3.21 (Công ty Thông tin Nghiên cứu Khảo sát Xã hội, Nhật Bản) đã được sử dụng để phân tích với $p < 0,05$ được coi là có ý nghĩa thống kê.

3. Đạo đức nghiên cứu

Nghiên cứu này đã được phê duyệt bởi Ủy ban đạo đức của Bệnh viện Đại học Toyama, Toyama, Nhật Bản (Số: R2019003). Tất cả các phương pháp được thực hiện theo các hướng dẫn và quy định có liên quan. Thông tin của bệnh nhân được mã hóa và bảo mật, phục vụ nghiên cứu khoa học, nhằm nâng cao chất

lượng điều trị, không có bất kỳ mục đích khác.

III. KẾT QUẢ

Đánh giá khả năng dự đoán bệnh được tính toán bởi tính đúng (accuracy) và hệ số điểm F1 (F1-score). Bảy mô hình máy học được tiến hành trên bộ dữ liệu huấn luyện ($n = 756$) và áp dụng chúng cho một bộ dữ liệu thử nghiệm ($n = 253$).

Kết quả dự đoán của bảy mô hình máy học khác nhau được tổng hợp trong Bảng 2 cho thấy tính đúng dao động từ 62% đến 72%. Điểm F1 ở nhóm bệnh PV dao động từ 0,71 đến 0,78 ở tất cả các mô hình máy học, và thấp hơn ở nhóm bệnh non-PV và HO. Mô hình SVM cho thấy kết quả tốt nhất trong tất cả các mô hình với tính đúng 72% và điểm F1 của 3 nhóm bệnh PV, non-PV và HO lần lượt là 0,78; 0,64; 0,71.

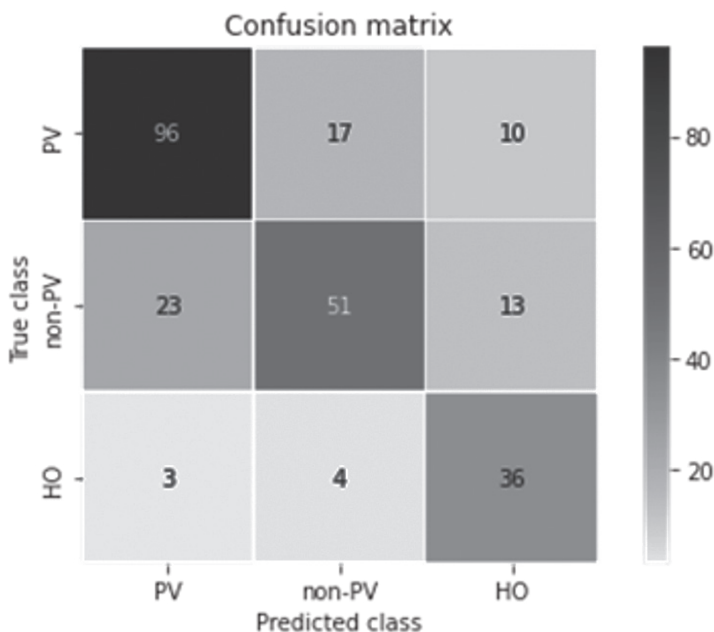
Bảng 2. Kết quả dự đoán bệnh của các mô hình máy học khác nhau

Mô hình máy học	Tính đúng (Accuracy)	Tính chính xác (Precision)			Độ nhạy (Recall)			Điểm hệ số F1 (F1-score)		
		PV	Non-PV	HO	PV	Non-PV	HO	PV	Non-PV	HO
DT	0,66	0,74	0,57	0,63	0,69	0,59	0,72	0,71	0,58	0,67
RF	0,66	0,67	0,65	0,64	0,85	0,55	0,33	0,75	0,6	0,43
AB	0,64	0,63	0,65	0,75	0,85	0,59	0,14	0,73	0,61	0,24
GB	0,68	0,70	0,69	0,63	0,88	0,44	0,63	0,78	0,54	0,63

Mô hình máy học	Tính đúng (Accuracy)	Tính chính xác (Precision)			Độ nhạy (Recall)			Điểm hệ số F1 (F1-score)		
		PV	Non-PV	HO	PV	Non-PV	HO	PV	Non-PV	HO
SVM	0,72	0,79	0,71	0,61	0,78	0,59	0,84	0,78	0,64	0,71
LR	0,70	0,72	0,68	0,67	0,82	0,52	0,72	0,77	0,59	0,70
KN	0,62	0,65	0,61	0,56	0,81	0,36	0,63	0,72	0,45	0,59
Average	0,67	0,70	0,65	0,64	0,81	0,52	0,57	0,75	0,57	0,57
SD	0,03	0,06	0,05	0,06	0,06	0,09	0,25	0,03	0,06	0,17

Khung cơ bản của ma trận nhầm lẫn hiển thị kết quả được dự đoán bởi từng mô hình theo bốn loại: TP, FP, FN, TN.

Số lượng dự đoán đúng của nhóm bệnh PV, non-PV và HO lần lượt là 96, 51 và 36 bệnh nhân trong nhóm dữ liệu thử nghiệm 25% (n = 253).



Hình 1. Khung cơ bản của ma trận

Chỉ số “feature_importances” trong scikit-learn cho thấy tầm quan trọng của mỗi tính năng đóng góp vào việc dự đoán một mô hình. Chỉ số này cho biết những đặc điểm nào có thể phù hợp nhất hoặc ít phù hợp nhất với mục tiêu nghiên cứu.¹¹ Trong nghiên cứu này, 10 tính năng đóng vai trò quan trọng nhất được đánh giá theo thứ tự bởi mô hình RF, AB, và GB (Biểu đồ 1).

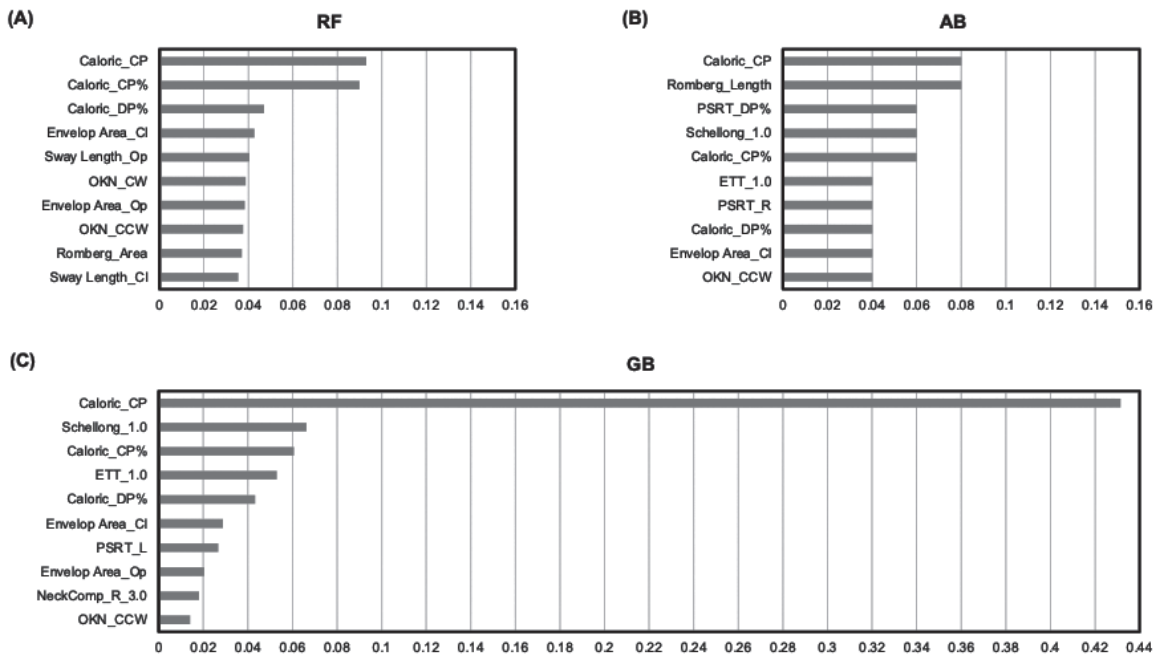
Mỗi tính năng được tính với một giá trị nằm trong khoảng từ 0 đến 1, trong đó 0 có nghĩa là “hoàn toàn không đóng vai trò trong việc dự đoán” và 1 có nghĩa là “dự đoán hoàn hảo với giá trị mục tiêu”, giá trị số càng cao thì tính năng đó càng quan trọng.

Giữa các tính năng đánh giá chức năng tiền đình, các tính năng của bài kiểm tra nhiệt lượng (Caloric_CP, Caloric_CP%) là xếp hạng

cao nhất trong cả ba mô hình. Điều này khẳng định rằng CP trong bài kiểm tra nhiệt lượng là một thông số đóng vai trò quan trọng trong việc phân loại bệnh PV và non-PV.

Đối với các tính năng liên quan đến thử nghiệm đo thăng bằng tư thế (Stabilometry test), tỉ lệ chiều dài khi lắc lư cơ thể (Romberg_Length) cũng nằm trong 10 tính năng hàng đầu của mô hình AB, tương tự Caloric_CP. Các tính năng khác, như Envelop Area_Op, Envelop Area_CI, Sway length_Op,

Sway length_CI, và Romberg_Area, cũng có mặt trong bảng xếp loại của 3 mô hình này. Trong số các tính năng để đánh giá chức năng tiểu não và thân não, hai tính năng của thử nghiệm rung giật nhãn cầu tự phát (OKN_CW, OKN_CCW) được đánh giá cao bởi mô hình RF. Các tính năng của bài kiểm tra theo dõi mắt (ETT), bài kiểm tra Schellong (Schellong) và thử nghiệm xoay theo hình sin kiểu con lắc (PSRT_R, PSRT_L) cũng được xếp hạng bởi mô hình AB và GB.



Biểu đồ 1. Top 10 tính năng quan trọng nhất, được xếp hạng từ cao xuống thấp theo mô hình phân loại

Random forest (RF); (b) Adaboost (AB); (c) Gradient boosting (GB)

- CP (canal paresis):* liệt ống bán khuyên trong nghiệm pháp nhiệt,
- DP (directional preponderance)* ưu thế định hướng trong nghiệm pháp nhiệt,
- PSRT (pendular sinusoidal rotation test):* thử nghiệm xoay theo hình sin kiểu con lắc,
- R (right):* phải, *L (left):* trái, *Op (open):* mở mắt, *CI (close):* nhắm mắt,
- OKN (optokinetic), CW (clockwise):* cùng chiều kim đồng hồ,
- CCW (counterclockwise):* ngược chiều kim đồng hồ,
- FFS (failure of fixation suppression test):* thử nghiệm thất bại trong việc cố định ánh nhìn,
- ETT (eye tracking test):* thử nghiệm theo dõi chuyển động của mắt

IV. BÀN LUẬN

Đối với bệnh nhân có triệu chứng chóng mặt, điều quan trọng cần phải phát hiện, chẩn đoán đúng bệnh PV và phân biệt với các nhóm bệnh khác. Tuy nhiên, nhóm bệnh không phải PV bao gồm rất nhiều bệnh khác nhau, ví dụ bệnh có căn nguyên não như u não, nhồi máu não, thoái hóa tủy sống tiểu não, thiếu năng động mạch đốt sống thân nền và những bệnh khác mà chẩn đoán không kịp thời có thể dẫn đến hậu quả đe dọa tính mạng. Vì vậy, ứng dụng công nghệ trí tuệ nhân tạo hay cụ thể là các mô hình máy học là cần thiết để hỗ trợ bác sĩ đưa ra quyết định chẩn đoán không chỉ đối với các bệnh PV mà còn đối với các bệnh non-PV. Hiệu suất dự đoán này có thể được đánh giá bằng cách sử dụng tính đúng, độ nhạy và hệ số điểm F1. Trong đó, điểm F1 là thước đo có thể đánh giá toàn diện tính chính xác và độ nhạy. Theo như bảng kết quả, điểm F1 của SVM là tốt nhất với 0,78 đối với nhóm PV, 0,71 ở nhóm HO và 0,64 đối với nhóm non-PV.

Trong nghiên cứu trước, chúng tôi đã sử dụng năm mô hình máy học phổ biến là RF, AB, GB, SVM và LR để phân biệt chẩn đoán giữa 2 nhóm bệnh PV và non-PV. Kết quả cho thấy khả năng dự đoán đúng lên tới 79% bởi mô hình SVM. Đáng chú ý hơn khi kết hợp 5 mô hình với nhau, khả năng dự đoán đúng bệnh PV tăng lên 83% và bệnh non-PV là 85%. Bệnh chóng mặt do hạ huyết áp tư thế (HO) thuộc nhóm non-PV chiếm số lượng lớn trong tập dữ liệu. Do vậy, chúng tôi hi vọng có thể ứng dụng công nghệ AI và máy học để chẩn đoán bệnh HO và phân biệt với 2 nhóm PV và non-PV còn lại. Trong bảy mô hình máy học khác nhau, SVM cho thấy kết quả tốt hơn các mô hình khác với tính đúng là 72% và điểm F1 là 0,78 (PV), 0,64 (non-PV) và 0,71 (HO). Kết quả trong nghiên cứu này thấp hơn so với nghiên cứu trước của chúng tôi trong việc phân loại

PV và non-PV với tính đúng là 79%. Kết quả từ hai nghiên cứu của chúng tôi cho thấy mô hình SVM dường như vượt trội so với các mô hình khác và là công cụ phân loại hữu ích phân biệt giữa bệnh PV, non-PV và HO.

Masankaran và cộng sự đã sử dụng bốn mô hình phân loại (RF, SVM, k-nearest Neighbor, và naïve Bayes) với bảng câu hỏi DHI để phân biệt chóng mặt tư thế kịch phát lành tính với tính đúng cao nhất là 73,91%.¹² Priesol và cộng sự áp dụng năm mô hình phân loại (DT, RF, LR, AB và SVM) với độ chính xác là 76%.¹³ So với những báo cáo này, hiệu suất của mô hình phân loại tốt nhất của chúng tôi trong nghiên cứu trước có độ chính xác cao hơn là 79%, tuy nhiên ở nghiên cứu đa phân loại này giảm hơn là 72%. Nguyên nhân có thể do số lượng đối tượng nghiên cứu tương đối khiêm tốn, sự chênh lệch về số lượng dữ liệu giữa 3 nhóm bệnh, đặc biệt là ở nhóm bệnh HO (497 PV, 157 HO và 355 non-PV). Số lượng dữ liệu mất cân bằng giữa 3 nhóm có thể ảnh hưởng đến kết quả điểm F1 giảm ở nhóm non-PV và HO.

Trong Biểu đồ 1, các tính năng liên quan đến bài kiểm tra nhiệt lượng là những đặc tính quan trọng nhất. Ngoài ra, những đặc tính khác như kiểm tra rung giật nhãn cầu tự phát, kiểm tra theo dõi chuyển động mắt, kiểm tra Schellong, kiểm tra xoay theo hình sin kiểu con lắc và nghiệm pháp thăng bằng tư thế cũng được xếp hạng trong top 10. Do đó, việc kết hợp nhiều loại bài kiểm tra chức năng thăng bằng có thể giúp nâng cao chất lượng của tập dữ liệu huấn luyện cho máy học.

Ngày nay, sự gia tăng không ngừng của thông tin y khoa đặt ra một thách thức không nhỏ đối với các chuyên gia y tế trong việc cập nhật kiến thức mới. Trước thách thức này, việc sử dụng AI được xem như là giải pháp hiệu quả nhất để đương đầu với khối lượng thông tin

khổng lồ trong lĩnh vực y tế. Ở nước ta hiện nay việc ứng dụng AI cũng như các thuật toán, mô hình máy học trong thực hành y học đang phát triển nhưng còn nhiều hạn chế, đặc biệt về việc chuẩn hoá lượng dữ liệu lớn đầu vào. Trong khi, hoạt động của máy học cho thấy nguồn dữ liệu học máy càng lớn, chẩn đoán đưa ra càng chính xác, càng ít bỏ sót các bệnh hiếm gặp hay khó chẩn đoán. Chính vì vậy, trong y học mặc dù máy học có thể hỗ trợ đưa ra dự đoán tốt nhưng hiện nay nó chưa thể thay thế hoàn toàn con người, do vậy bác sĩ cần đưa ra chẩn đoán cuối cùng.

Nghiên cứu của chúng tôi còn một số hạn chế, bao gồm các đặc điểm và số lượng đối tượng nghiên cứu tương đối khiêm tốn cũng như việc tối ưu hóa các mô hình máy học. Tỷ lệ nam và nữ chưa được đưa vào tập dữ liệu huấn luyện và kiểm tra. Ngoài ra, các bài kiểm tra chức năng thăng bằng và các xét nghiệm như được trình bày ở đây sẽ không phù hợp với việc ra quyết định lâm sàng trong trường hợp bệnh nhân chóng mặt cấp tính đang trong tình trạng cấp cứu hoặc ở trung tâm y tế không có đầy đủ các máy móc, thiết bị thăm dò chức năng tiền đình.

V. KẾT LUẬN

Nghiên cứu hiện tại cho thấy tính hiệu quả của việc sử dụng mô hình máy học dựa trên kết quả các bài kiểm tra chức năng thăng bằng, đặc biệt là mô hình SVM như một công cụ hỗ trợ để chẩn đoán bệnh HO và phân biệt với các bệnh PV, non-PV với tính đúng là 72%. Kết quả đã chứng minh rằng AI hoàn toàn có khả năng không chỉ dự đoán nhóm bệnh PV, non-PV mà còn có thể đưa ra chẩn đoán chính xác một bệnh cụ thể, ví dụ HO như trong nghiên cứu này. Trong tương lai, cần có những nghiên cứu sâu hơn sử dụng các mô hình tổng hợp được tối ưu hóa với số lượng lớn bệnh nhân để cải thiện khả năng chẩn đoán của máy học

với các mặt bệnh cụ thể khác. Các nghiên cứu bước đầu của chúng tôi cho thấy AI và máy học có thể trở thành công cụ hữu ích trong thực hành lâm sàng, giúp bác sĩ giải phóng thời gian và nâng cao tính chính xác, hiệu quả của việc chẩn đoán bệnh.

TÀI LIỆU THAM KHẢO

1. Labuguen RH. Initial evaluation of vertigo [published correction appears in *Am Fam Physician*. 2006 May 15;73(10):1704]. *Am Fam Physician*. 2006; 73(2): 244-251.
2. Strupp M, Feil K, Zwergal A. Diagnose und Differenzialdiagnose von peripheren und zentralen Schwindelsyndromen [Diagnosis and Differential Diagnosis of Peripheral and Central Vestibular Disorders]. *Laryngorhinotologie*. 2021; 100(3): 176-183. doi:10.1055/a-1057-3239.
3. Mayo RC, Leung J. Artificial intelligence and deep learning - Radiology's next frontier?. *Clin Imaging*. 2018; 49: 87-88. doi:10.1016/j.clinimag.2017.11.007.
4. Egert M, Steward JE, Sundaram CP. Machine Learning and Artificial Intelligence in Surgical Fields. *Indian J Surg Oncol*. 2020; 11(4): 573-577. doi:10.1007/s13193-020-01166-8.
5. Ahmadi SA, Vivar G, Navab N, et al. Modern machine-learning can support diagnostic differentiation of central and peripheral acute vestibular disorders. *J Neurol*. 2020; 267(Suppl 1): 143-152. doi:10.1007/s00415-020-09931-z.
6. Bisdorff A, Von Brevern M, Lempert T, Newman-Toker DE. Classification of vestibular symptoms: towards an international classification of vestibular disorders. *J Vestib Res*. 2009; 19(1-2): 1-13. doi:10.3233/VES-2009-0343.
7. Japan Society for Equilibrium Research. [https:// www. memai. jp/ guide/](https://www.memai.jp/guide/).

8. Anh DT, Takakura H, Asai M, Ueda N, Shojaku H. Application of machine learning in the diagnosis of vestibular disease. *Sci Rep*. 2022; 12(1): 20805. Published 2022 Dec 2. doi:10.1038/s41598-022-24979-9.
9. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018; 284(6): 603-619. doi:10.1111/joim.12822.
10. F. Pedregosa, G. Varoquaux, A. Gramfort, et al. *Journal of Machine Learning Research*. 2011; 12: 2825–2830.
11. Genuer, R., Poggi, J. M. & Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett*. 2010; 31: 2225–2236.
12. Masankaran, L., Viyanon, W. & Mahasittiwat, V. Classification of benign paroxysmal positioning vertigo types from Dizziness Handicap Inventory using machine learning techniques. *International Conference on Intelligent Informatics and Biomedical Sciences, ICIBMS*. 2018; 209–214.
13. Priesol AJ, Cao M, Brodley CE, Lewis RF. Clinical vestibular testing assessed with machine-learning algorithms. *JAMA Otolaryngol Head Neck Surg*. 2015; 141(4): 364-372. doi:10.1001/jamaoto.2014.3519.

Summary

RESEARCH ON ARTIFICIAL INTELLIGENCE (AI) AND EQUILIBRIUM FUNCTION TESTS IN THE DIAGNOSIS OF VESTIBULAR DISEASE

Peripheral vestibular (PV) disease is a common cause of dizziness. However, an accurate diagnosis of this condition requires significant time for clinical examination, comprehensive assessment, and differentiation from non-peripheral vestibular diseases (non-PV). Non-PV diseases encompass a wide range of conditions, with hemodynamic orthostatic dizziness/vertigo (HO) being the most prevalent in our data. This study aims to evaluate the effectiveness of applying machine learning which is a branch of artificial intelligence (AI) to diagnose HO and classify peripheral vestibular (PV) and non-peripheral vestibular (non-PV) categories. Multi-class classification models were tested on a dataset of 1,009 patients (497 PV, 157 HO, and 355 non-PV) and achieved an overall accuracy of 72%, with F1 scores of 0.78 for PV, 0.64 for non-PV, and 0.71 for HO. Our results show that AI can become a useful tool in clinical practice, permitting time saving and improve the accuracy and efficiency of disease diagnosis.

Keywords: Machine learning, neuro-otological examinations, peripheral vestibular disease, hemodynamic orthostatic dizziness/vertigo.