COMPARISON OF THE PERFORMANCE OF TWO ARTIFICIAL INTELLIGENCE MODELS USING THE DECISION TREE ALGORITHM FOR PREDICTING HBA1C ≥ 6.5% BASED ON BIOCHEMICAL TEST RESULTS AT BAC KAN GENERAL HOSPITAL IN 2026

Nong Văn Diep[1], Hoàng Xuan Son[1], Trieu Thi Bien[1], Hoang Thi Thuan[1], Nguyen Duc Quan[2,3]

[1]Bac Kan General Hospital - Residential Group No. 2, Duc Xuan Ward, Thai Nguyen Province, Vietnam
[2]Department of Life Science, University of Science and Technology of Hanoi, VAST - Building A21, 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi City, Vietnam
[3]Institute of Biology, Vietnam Academy of Science and Technology - No. 18, Hoang Quoc Viet Street, Nghia Do Ward, Hanoi City, Vietnam

Received: 14/07/2026
Reviced: 19/08/2026; Accepted: 10/09/2026

ABSTRACT

Objective: Comparison of the performance of two artificial intelligence models using the Decision Tree algorithm for predicting HbA1c ≥ 6.5% based on biochemical test results at Bac Kan General Hospital in 2026.

Methods: A cross-sectional study was conducted at Bac Kan General Hospital in 2026, involving 880 patients who underwent HbA1c and biochemical testing. Two Decision Tree models were developed using SPSS software based on the CART method, with a 70/30 training-to-testing data split. Model A utilized the variables of age, glucose, total cholesterol, triglycerides, AST, and ALT; Model B utilized age, total cholesterol, triglycerides, AST, and ALT (excluding glucose). The performance of both models was evaluated based on accuracy, sensitivity, specificity, and risk estimate.

Results: Among the 880 patients studied, the group with HbA1c ≥ 6.5% exhibited statistically significantly higher glucose and triglyceride levels compared to the group with HbA1c < 6.5% ($p < 0.05$). Model A demonstrated superior performance to Model B on the test set, achieving an accuracy of 79.8% versus 57.5%; sensitivities of 63.0% and 12.7%, respectively; specificities of 89.7% and 90.6%; and error estimates of 0.202 and 0.425. In Model A, glucose was the most important variable and served as the initial splitting node of the decision tree; when glucose was removed, triglycerides became the model's most important variable in Model B.

Conclusions: The Decision Tree algorithm can assist in predicting HbA1c levels ≥ 6.5% based on biochemical test results. Incorporating the glucose variable significantly improves the model's predictive performance, underscoring the crucial role of glucose in developing artificial intelligence models to screen for patients at risk of elevated HbA1c.

Keywords: Artificial intelligence, Decision Tree, HbA1c, glucose, biochemical test.

*Corresponding author
Email: nongvandiep1093@gmail.com **Phone**: (+84) 891125822 **DOI:** 10.52163/yhc.v67iCD14.6477

SO SÁNH HIỆU NĂNG HAI MÔ HÌNH TRÍ TUỆ NHÂN TẠO SỬ DỤNG THUẬT TOÁN DECISION TREE TRONG DỰ BÁO HBA1C ≥ 6,5% TỪ CÁC XÉT NGHIỆM HÓA SINH TẠI BỆNH VIỆN ĐA KHOA BẮC KẠN NĂM 2026

Nông Văn Diệp[1], Hoàng Xuân Sơn[1], Triệu Thị Biển[1], Hoàng Thị Thuần[1], Nguyễn Đức Quân[2,3]

[1]Bệnh viện đa khoa Bắc Kạn - Tổ dân phố số 2,phường Đức Xuân, tỉnh Thái Nguyên, Việt Nam
[2]Khoa Khoa học Sự sống, Đại học Khoa học và Công nghệ Hà Nội, VAST - Tòa nhà A21, 18 đường Hoàng Quốc Việt, phường Nghĩa Đô, thành phố Hà Nội, Việt Nam
[3]Viện Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam - Số 18, đường Hoàng Quốc Việt, phường Nghĩa Đô, Thành phố Hà Nội, Việt Nam

Ngày nhận bài: 14/07/2026
Ngày chỉnh sửa: 19/08/2026; Ngày duyệt đăng: 10/09/2026

TÓM TẮT

Mục tiêu: So sánh hiệu năng của hai mô hình trí tuệ nhân tạo sử dụng thuật toán Decision Tree trong dự báo HbA1c ≥ 6,5% từ các xét nghiệm hóa sinh tại Bệnh viện Đa khoa Bắc Kạn năm 2026.

Phương pháp nghiên cứu: Nghiên cứu cắt ngang trên 880 người bệnh được thực hiện xét nghiệm HbA1c và các xét nghiệm hóa sinh tại Bệnh viện Đa khoa Bắc Kạn năm 2026. Hai mô hình Decision Tree được xây dựng bằng phần mềm SPSS theo phương pháp CART với tỷ lệ chia dữ liệu huấn luyện/kiểm tra là 70/30. Mô hình A sử dụng các biến tuổi, glucose, cholesterol toàn phần, triglycerid, AST và ALT; mô hình B sử dụng các biến tuổi, cholesterol toàn phần, triglycerid, AST và ALT (không bao gồm glucose). Hiệu năng của hai mô hình được đánh giá thông qua độ chính xác, độ nhạy, độ đặc hiệu và sai số phân loại.

Kết quả: Trong 880 người bệnh nghiên cứu, nhóm HbA1c ≥ 6,5% có nồng độ glucose và triglycerid cao hơn có ý nghĩa thống kê so với nhóm HbA1c < 6,5% (p<0,05). Mô hình A cho hiệu năng vượt trội so với mô hình B trên tập kiểm tra, với độ chính xác đạt 79,8% so với 57,5%; độ nhạy lần lượt là 63,0% và 12,7%; độ đặc hiệu là 89,7% và 90,6%; sai số phân loại là 0,202 và 0,425. Trong mô hình A, glucose là biến có mức độ quan trọng cao nhất và là nút phân tách đầu tiên của cây quyết định. Khi loại bỏ biến glucose, triglycerid trở thành biến quan trọng nhất của mô hình B.

Kết luận: Thuật toán Decision Tree có khả năng hỗ trợ dự báo tình trạng HbA1c ≥ 6,5% từ các xét nghiệm hóa sinh. Việc bổ sung biến glucose giúp cải thiện rõ rệt hiệu năng dự báo của mô hình, góp phần khẳng định vai trò quan trọng của glucose trong xây dựng mô hình trí tuệ nhân tạo hỗ trợ sàng lọc người bệnh có nguy cơ tăng HbA1c.

Từ khóa: Trí tuệ nhân tạo, Decision Tree, HbA1c, glucose, xét nghiệm hóa sinh.

1. ĐẶT VẤN ĐỀ

Đái tháo đường (ĐTĐ) là bệnh không lây nhiễm có tỷ lệ mắc ngày càng gia tăng và gây gánh nặng lớn đối với hệ thống y tế trên toàn cầu [1]. HbA1c phản ánh nồng độ glucose máu trung bình trong khoảng 2–3 tháng và được sử dụng rộng rãi trong chẩn đoán, theo dõi kiểm soát đường huyết và đánh giá hiệu quả điều trị ĐTĐ [2]. Tuy nhiên, do chi phí và yêu cầu về trang thiết bị, xét nghiệm HbA1c chưa được chỉ định thường quy tại nhiều cơ sở y tế. Trong khi đó, các xét nghiệm hóa sinh như glucose, cholesterol toàn phần, triglycerid, AST và ALT được thực hiện hằng ngày với chi phí thấp và sẵn có tại hầu hết các phòng xét nghiệm. Việc khai thác dữ liệu từ các xét nghiệm này để hỗ trợ dự báo HbA1c có ý nghĩa trong sàng lọc người bệnh và tối ưu hóa chỉ định xét nghiệm.

Những năm gần đây, trí tuệ nhân tạo (Artificial Intelligence - AI), đặc biệt là các thuật toán học máy (Machine Learning - ML), được ứng dụng ngày càng nhiều trong y học nhằm hỗ trợ chẩn đoán và dự báo bệnh [3]. Trong đó, Decision Tree (cây quyết định) là thuật toán có ưu điểm dễ diễn giải, dễ triển khai và phù hợp với các hệ thống hỗ trợ quyết định lâm sàng [4]. Mặc dù đã có nhiều nghiên cứu ứng dụng học máy trong dự báo đái tháo đường từ nhiều yếu tố lâm sàng [5], nhưng các nghiên cứu sử dụng dữ liệu xét nghiệm hóa sinh thường quy để dự báo HbA1c ≥ 6,5% đặc biệt là so sánh trực tiếp hiệu năng giữa mô hình có và không có biến glucose vẫn còn hạn chế. Xuất phát từ thực tiễn đó, nghiên cứu được thực hiện nhằm so sánh hiệu năng của hai mô hình Decision Tree dự báo HbA1c ≥ 6,5% từ các xét nghiệm hóa sinh thường quy và đánh giá vai trò của glucose trong việc nâng cao hiệu năng dự báo của mô hình.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thiết kế nghiên cứu

Nghiên cứu cắt ngang trên 880 người bệnh kết hợp xây dựng và đánh giá mô hình dự báo bằng thuật toán học máy Decision Tree(Classification and Regression Tree - CART).

2.2. Địa điểm, thời gian nghiên cứu và đối tượng nghiên cứu

Nghiên cứu được tiến hành tại Bệnh viện Đa khoa Bắc Kạn trong thời gian từ tháng 01/2026 đến tháng 07/2026

*Tác giả liên hệ
Email: nongvandiep1093@gmail.com **Điện thoại**: (+84) 891125822 **DOI:** 10.52163/yhc.v67iCD14.6477

trên đối tượng trên 880 người, chọn mẫu thuận tiện. Tất cả các trường hợp đáp ứng tiêu chuẩn lựa chọn trong thời gian nghiên cứu đều được đưa vào phân tích.

Tiêu chuẩn lựa chọn:

- Người bệnh từ 18 tuổi trở lên.
- Có kết quả xét nghiệm HbA1c đồng thời với các xét nghiệm hóa sinh gồm glucose, cholesterol toàn phần, triglycerid, AST và ALT.
- Hồ sơ xét nghiệm đầy đủ và hợp lệ trên hệ thống thông tin xét nghiệm(LIS).
- Người bệnh đồng ý tham gia nghiên cứu.

Tiêu chuẩn loại trừ:

- Thiếu một hoặc nhiều biến nghiên cứu.
- Người bệnh không đồng ý tham gia nghiên cứu.

2.3. Biến số nghiên cứu

Biến phụ thuộc: HbA1c được phân loại thành hai nhóm: HbA1c < 6,5% và HbA1c ≥ 6,5%.

Biến độc lập: Tuổi(năm), glucose(mmol/L), cholesterol toàn phần(mmol/L), triglycerid(mmol/L), AST(U/L), ALT(U/L).

Biến đầu ra của mô hình: Độ chính xác(accuracy), độ nhạy(sensitivity), độ đặc hiệu(specificity), sai số phân loại(risk estimate), mức độ quan trọng của biến(variable importance)

2.4. Kỹ thuật, công cụ và quy trình thu thập số liệu

Dữ liệu được trích xuất từ hệ thống quản lý thông tin xét nghiệm (LIS) của Bệnh viện Đa khoa Bắc Kạn.

Lấy 4 ml máu tĩnh mạch và các chỉ số glucose, cholesterol toàn phần, triglycerid, AST và ALT được thực hiện trên hệ thống xét nghiệm hóa sinh tự động AU 680 Beckman Coulter, HbA1c được định lượng trên máy HA-1500 Dagnostic GmbH.

Sau khi làm sạch dữ liệu, bộ dữ liệu được chia ngẫu nhiên theo tỷ lệ 70% cho tập huấn luyện (Training) và 30% cho tập kiểm tra (Testing).

Hai mô hình Decision Tree được xây dựng và phân tích:

- Mô hình A: sử dụng các biến đầu vào gồm: tuổi, glucose, cholesterol toàn phần, triglycerid, AST và ALT.
- Mô hình B: sử dụng các biến đầu vào gồm: tuổi, cholesterol toàn phần, triglycerid, AST và ALT (không bao gồm glucose).

2.5. Xử lý và phân tích số liệu

Số liệu được xử lý bằng phần mềm IBM SPSS Statistics 26.0 sau khi làm sạch và mã hóa. Mô hình Decision Tree được sử dụng nhằm phân tầng nguy cơ HbA1c ≥ 6,5% dựa trên các biến đầu vào bao gồm tuổi và các chỉ số hóa sinh. Dữ liệu được phân chia thành tập huấn luyện (70%) và tập kiểm định (30%) để xây dựng và đánh giá mô hình. Hiệu năng mô hình được đánh giá thông qua các chỉ số bao gồm độ chính xác (accuracy), độ nhạy (sensitivity), độ đặc hiệu (specificity), sai số phân loại (risk estimate), các quy tắc phân nhánh của mô hình được phân tích nhằm phục vụ diễn giải lâm sàng.

2.6. Đạo đức nghiên cứu

Nghiên cứu được sự chấp thuận của Ban Giám đốc Bệnh viện đa khoa Bắc Kạn và Hội đồng khoa học Sở y tế tỉnh Thái Nguyên theo quyết định số 1339/QĐ-SYT ngày 15 tháng 06 năm 2026. Nghiên cứu sử dụng bộ số liệu xét nghiệm, không can thiệp trực tiếp vào người bệnh. Thông tin cá nhân được mã hóa và bảo mật tuyệt đối chỉ phục vụ cho nghiên cứu.

3. KẾT QUẢ NGHIÊN CỨU

Bảng 1. Đặc điểm của đối tượng nghiên cứu theo phân loại HbA1c (n=880)

Đặc điểm	HbA1c <6,5% (n=544)	HbA1c ≥6,5% (n=336)	p
Tuổi (năm)	59,76 ± 12,71	60,86 ± 13,17	0,221
Glucose (mmol/L)	5,29 ± 1,22	10,11 ± 6,24	< 0,001
Cholesterol (mmol/L)	5,21 ± 1,20	5,28 ± 1,53	0,454
Triglycerid (mmol/L)	2,18 ± 2,13	2,69 ± 3,74	0,024
AST (U/L)	30,11 ± 14,81	32,66 ± 31,14	0,160
ALT (U/L)	27,17 ± 20,02	28,30 ± 19,44	0,412

Bảng 1 cho thấy nhóm HbA1c ≥6,5% có nồng độ glucose và triglycerid cao hơn có ý nghĩa thống kê so với nhóm HbA1c <6,5% (p<0,05), trong khi tuổi, cholesterol, AST và ALT không khác biệt có ý nghĩa thống kê.

Bảng 2. So sánh hiệu năng dự báo của hai mô hình Decision Tree

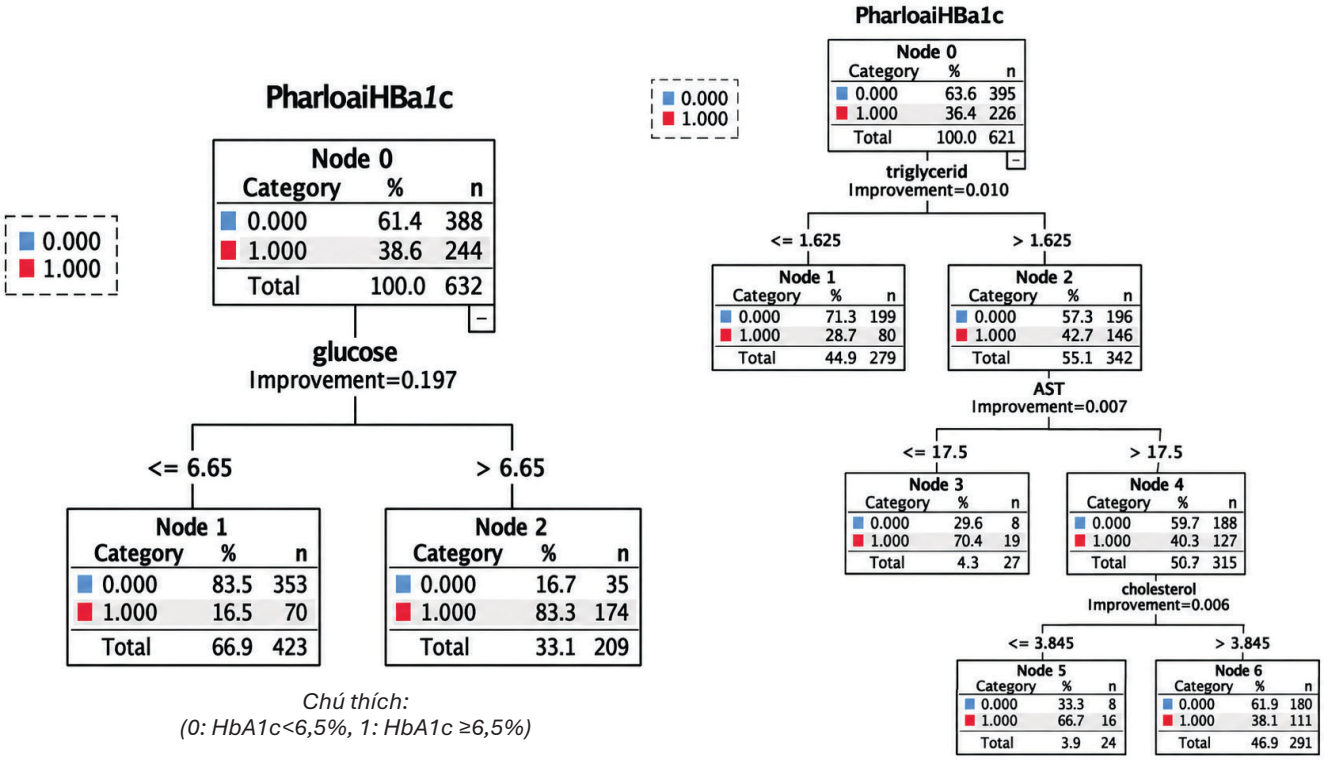
Chỉ số	Mô hình A (Có glucose)	Mô hình B (Không glucose)
Số biến đầu vào	6	5
Biến phân tách đầu tiên	Glucose	Triglycerid
Độ chính xác tập huấn luyện (%)	83,4	66,7
Độ chính xác tập kiểm tra (%)	79,8	57,5
Độ nhạy tập kiểm tra (%)	63,0	12,7
Độ đặc hiệu tập kiểm tra (%)	89,7	90,6
Sai số phân loại tập kiểm tra	0,202	0,425
Số mức phân chia của cây	1	3

Bảng 2 cho thấy mô hình A đạt độ chính xác trên tập kiểm tra 79,8%, cao hơn rõ rệt so với mô hình B 57,5%, đồng thời có sai số phân loại thấp hơn, cho thấy khả năng dự báo tốt hơn.

Bảng 3. Thứ tự mức độ quan trọng của các biến trong hai mô hình

TT	Mô hình A	Giá trị (Mức độ quan trọng)	Mô hình B	Giá trị (Mức độ quan trọng)
1	Glucose	0,197 (100,0%)	Triglycerid	0,010 (100,0%)
2	Triglycerid	0,007 (3,5%)	ALT	0,008 (72,9%)
3	AST	0,005 (2,4%)	AST	0,007 (68,9%)
4	ALT	0,001 (0,5%)	Cholesterol	0,006 (55,7%)
5	Cholesterol	0,001 (0,4%)	Tuổi	0,001 (10,0%)

Bảng 3 cho thấy trong mô hình A, glucose là biến quan trọng nhất với mức độ quan trọng 100%. Khi loại bỏ glucose khỏi mô hình, triglycerid trở thành biến có vai trò dự báo lớn nhất(100%), tiếp theo là kết hợp với ALT, AST và cholesterol, tuổi.



Chú thích:
(0: HbA1c<6,5%, 1: HbA1c ≥6,5%)

Hình 1. Cấu trúc cây quyết định của hai mô hình Decision Tree A (trái) và B (phải)

Hình 1 cho thấy mô hình A sử dụng glucose là biến phân tách đầu tiên với ngưỡng 6,65 mmol/L. Trong nhóm người bệnh có glucose ≤ 6,65 mmol/L, 83,5% thuộc nhóm HbA1c < 6,5%; ngược lại, ở nhóm có glucose > 6,65 mmol/L, 83,3% thuộc nhóm HbA1c ≥ 6,5%. Mô hình B sử dụng triglycerid là biến phân tách đầu tiên với điểm cắt 1,625 mmol/L. Các nhánh tiếp theo được phân tách bởi AST (17,5 U/L) và cholesterol (3,845 mmol/L), tạo thành 7 nút. Nhóm người bệnh có triglycerid ≤ 1,625 mmol/L chủ yếu thuộc nhóm HbA1c < 6,5% (71,3%), trong khi ở nhánh triglycerid > 1,625 mmol/L, mô hình tiếp tục phân loại bằng AST và cholesterol.

4. BÀN LUẬN

Trong nghiên cứu này, nhóm HbA1c ≥ 6,5% có nồng độ glucose trung bình cao hơn rõ rệt so với nhóm HbA1c < 6,5% (10,11 ± 6,24 mmol/L so với 5,29 ± 1,22 mmol/L; p < 0,001). Kết quả này phù hợp với cơ sở sinh học của HbA1c, vì HbA1c phản ánh nồng độ glucose máu trung bình trong khoảng 2-3 tháng trước đó [2]. Theo Hiệp hội Đái tháo đường Hoa Kỳ (ADA), cả HbA1c và glucose huyết tương đều được sử dụng trong chẩn đoán đái tháo đường [6]. Do đó, sự khác biệt rõ rệt về glucose giữa hai nhóm trong nghiên cứu này là phù hợp và bước đầu khẳng định glucose là biến có giá trị quan trọng trong dự báo HbA1c ≥ 6,5%. Nồng độ triglycerid ở nhóm HbA1c ≥ 6,5% cũng cao hơn có ý nghĩa thống kê so với nhóm HbA1c < 6,5% (p = 0,024). Kết quả này phù hợp với các nghiên cứu cho thấy tình trạng kháng insulin và rối loạn chuyển hóa lipid thường đi kèm tăng triglycerid ở người bệnh đái tháo đường [7]. Tuy nhiên, do triglycerid chỉ phản ánh gián tiếp rối loạn chuyển hóa glucose và chịu ảnh hưởng của nhiều yếu tố, giá trị dự báo của biến này thấp hơn glucose. Cholesterol toàn phần, AST, ALT và tuổi không khác biệt có ý nghĩa thống kê giữa hai nhóm, cho thấy các biến này có khả năng phân biệt hạn chế khi sử dụng riêng lẻ.

Nghiên cứu cho thấy mô hình A (có glucose) đạt hiệu năng dự báo vượt trội so với mô hình B (không có glucose), với độ chính xác trên tập kiểm tra đạt 79,8% so với 57,5% và sai số phân loại giảm từ 0,425 xuống còn 0,202. Kết quả này phù hợp với các tổng quan gần đây của Fahed và cộng sự (2023), về ứng dụng học máy trong dự báo đái tháo đường, trong đó glucose (đặc biệt là glucose huyết tương lúc đói) luôn được ghi nhận là một trong những biến dự báo quan trọng nhất và được sử dụng phổ biến trong các mô hình trí tuệ nhân tạo [8]. Điều này khẳng định vai trò trung tâm của glucose trong nâng cao hiệu năng mô hình Decision Tree.

Điểm mới của nghiên cứu không chỉ là so sánh hiệu năng giữa hai mô hình mà còn lượng hóa ảnh hưởng của biến glucose đối với cấu trúc và hiệu năng của thuật toán Decision Tree. Trong mô hình A, glucose được lựa chọn làm nút phân tách đầu tiên và có mức độ quan trọng cao nhất. Khi loại bỏ glucose, thuật toán phải sử dụng triglycerid kết hợp với AST và cholesterol để xây dựng cây quyết định nhưng hiệu năng vẫn giảm đáng kể. Kết quả này cho thấy không có sự kết hợp nào giữa các chỉ số hóa sinh còn lại có thể thay thế vai trò của glucose trong dự báo HbA1c ≥ 6,5% [4].

Ngoài độ chính xác, mô hình A còn đạt độ nhạy cao hơn nhiều so với mô hình B (63,0% so với 12,7%), trong khi độ đặc hiệu của hai mô hình tương đương (89,7% và 90,6%). Đối với mục tiêu sàng lọc, mô hình có độ nhạy cao giúp giảm nguy cơ bỏ sót người bệnh cần được chẩn đoán và theo dõi. Đồng thời, sai số phân loại của mô hình A thấp hơn rõ rệt (0,202 so với 0,425), cho thấy khả năng phân loại chính xác tốt hơn. Mặc dù hiệu năng chưa cao bằng một số thuật toán học máy phức tạp hơn, Decision Tree có ưu điểm là cấu trúc trực quan, dễ diễn giải và thuận lợi triển khai trong thực hành lâm sàng. Với độ chính xác gần 80% cùng khả năng giải thích rõ ràng các quy tắc phân

loại, mô hình có tiềm năng hỗ trợ sàng lọc người bệnh có nguy cơ HbA1c ≥ 6,5%, đặc biệt tại các cơ sở y tế có nguồn lực còn hạn chế.

Kết quả nghiên cứu cho thấy trong mô hình A, glucose có mức độ quan trọng chuẩn hóa đạt 100% và là biến duy nhất được lựa chọn để phân tách cây quyết định với ngưỡng 6,65 mmol/L, trong khi các biến còn lại chỉ đóng góp rất nhỏ. Phù hợp nguyên lý hoạt động của thuật toán Decision Tree (CART), trong đó thuật toán sẽ ưu tiên lựa chọn biến có khả năng làm giảm độ hỗn tạp lớn nhất tại mỗi bước phân nhánh. Việc glucose được chọn làm nút gốc hoàn toàn phù hợp với cơ chế sinh lý bệnh của ĐTĐ. Chính vì vậy, chỉ với một lần phân tách theo ngưỡng glucose 6,65 mmol/L, mô hình đã phân loại chính xác phần lớn người bệnh. Kết quả này cũng phù hợp với nhiều nghiên cứu ứng dụng học máy trong dự báo ĐTĐ, trong đó glucose luôn được xếp là biến dự báo quan trọng nhất, bất kể sử dụng mô hình nào [9].

Khi loại biến glucose khỏi mô hình, thuật toán lựa chọn triglycerid làm nút gốc do đây là biến còn lại có khả năng phân biệt hai nhóm HbA1c tốt nhất. Điều này có thể giải thích bởi tăng triglycerid là hậu quả thường gặp của tình trạng kháng insulin và rối loạn chuyển hóa lipid ở người bệnh ĐTĐ. Tuy nhiên, do triglycerid chỉ phản ánh gián tiếp tình trạng rối loạn chuyển hóa glucose, mô hình phải tiếp tục sử dụng thêm AST và cholesterol để hoàn thiện quá trình phân loại nhưng hiệu năng vẫn thấp hơn rất nhiều so với mô hình có glucose. Kết quả này cho thấy không có sự kết hợp nào giữa các chỉ số hóa sinh còn lại có thể vượt trội hơn vai trò của glucose trong dự báo HbA1c. Một ưu điểm nổi bật của Decision Tree là khả năng diễn giải trực quan, khác với nhiều thuật toán học máy phức tạp hơn, cây quyết định có cấu trúc minh bạch và biểu diễn rõ ràng các quy tắc phân loại dưới dạng các ngưỡng xét nghiệm cụ thể, giúp bác sĩ dễ dàng hiểu cơ sở của từng quyết định dự báo và tạo thuận lợi cho việc ứng dụng trong hệ thống hỗ trợ quyết định lâm sàng [10]. Điều này làm tăng khả năng ứng dụng trong thực hành lâm sàng, đặc biệt tại các bệnh viện tuyến tỉnh hoặc cơ sở y tế có nguồn lực hạn chế, nơi các mô hình có khả năng giải thích cao thường thuận lợi hơn trong hỗ trợ ra quyết định. Đây là một trong số ít nghiên cứu tại Việt Nam đánh giá trực tiếp sự thay đổi cấu trúc và hiệu năng của mô hình Decision Tree khi loại bỏ biến glucose, qua đó lượng hóa vai trò của glucose trong dự báo HbA1c ≥ 6,5%.

Hạn chế của nghiên cứu

Nghiên cứu được thực hiện trên dữ liệu của một bệnh viện và mới đánh giá mô hình bằng phương pháp chia tập huấn luyện/kiểm tra nội bộ (70/30), do đó cần được nghiên cứu trên các quần thể độc lập và đa trung tâm với cỡ mẫu lớn hơn để khẳng định khả năng khái quát mô hình. Mặc dù vậy, kết quả đã cung cấp những bằng chứng ban đầu về khả năng ứng dụng Decision Tree trong dự báo HbA1c từ các xét nghiệm hóa sinh thường quy, tạo cơ sở cho các nghiên cứu phát triển và hoàn thiện các mô hình đa biến và phức tạp hơn.

5. KẾT LUẬN

Nghiên cứu đã xây dựng và so sánh hai mô hình Decision Tree dự báo HbA1c ≥ 6,5% từ các xét nghiệm hóa sinh thường quy trên 880 người bệnh. Mô hình có sử dụng biến glucose cho hiệu năng vượt trội so với mô hình không sử dụng glucose, đồng thời glucose là biến quan trọng nhất và được lựa chọn làm nút phân tách đầu tiên của cây quyết định. Khi loại bỏ biến glucose, triglycerid trở thành biến dự báo quan trọng nhất nhưng không thể thay thế vai trò của glucose. Kết quả cho thấy Decision Tree có tiềm năng hỗ trợ sàng lọc người bệnh có nguy cơ tăng HbA1c từ các xét nghiệm hóa sinh thường quy và là cơ sở để phát triển các công cụ hỗ trợ quyết định trong thực hành lâm sàng.

6. TÀI LIỆU THAM KHẢO

[1] IDF Diabetes Atlas. 11th ed. Brussels: International Diabetes Federation; 2025.
[2] Diagnosis and Classification of Diabetes: Standards of Care in Diabetes-2024. *Diabetes care*. 2024;47(Suppl 1). doi:10.2337/dc24-S002.
[3] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.
[4] Kavakiotis I, Tsave O, Salifoglou A, et al. Machine Learning and Data Mining Methods in Diabetes Research. *Comput Struct Biotechnol J*. 2017;15:104-116. doi:10.1016/j.csbj.2016.12.005.
[5] Tuppad A, Patil SD. Machine learning for diabetes clinical decision support: a review. *Adv Comput Intell*. 2022;2(2):22. doi:10.1007/s43674-022-00034-y.
[6] American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes—2026. Diabetes Care. 2026;49(Suppl 1):S27-S46.
[7] Ginsberg HN, Zhang YL, Hernandez-Ono A. Regulation of plasma triglycerides in insulin resistance and diabetes. *Arch Med Res*. 2005;36(3):232-240. doi:10.1016/j.arcmed.2005.01.005. PMID: 15925013.
[8] Mohsen F, Al-Absi HRH, Yousri NA, et al. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digit Med*. 2023;6(1):197. doi:10.1038/s41746-023-00933-5. PMID: 37880301.
[9] Lugner M, Rawshani A, Helleryd E, Eliasson B. Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. Sci Rep. 2024. doi: 10.1038/s41598-024-52023-5.
[10] Xu Q, Xie W, Liao B, et al. Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. Comput Math Methods Med. 2023;2023:9919269. doi:10.1155/2023/9919269.