

APPLICATION OF FUZZY KNOWLEDGE GRAPH IN DIAGNOSIS SUPPORT FOR DIABETES PATIENTS

Pham Minh Chuan¹, Tran Manh Tuan², Cu Kim Long³, Nguyen Hong Tan^{4*}

¹Hung Yen University of Technology and Education, ²Thuyloi University

³Information Technology Center - Ministry of Science and Technology

⁴TNU - University of Information and Communication Technology

ARTICLE INFO	ABSTRACT
Received: 02/11/2023	Recently, the needs of human in health care are extremely necessary. Among common diseases, diabetes is one of the dangerous diseases and many people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, lifestyle, unreasonable diet, high blood pressure, etc. are all causes of diabetes. People, who have diabetes, are at high risk of developing diseases like heart disease, kidney disease, stroke, eye problems, nerve damage, etc. Information technology tools used in diagnostic support help doctors detect a patient's disease condition quickly and accurately. Based on the diagnosis, the appropriate treatment regimens for patients can be quickly indicated. In this article, we focus on researching fuzzy knowledge graph models in supporting diabetes diagnosis. The knowledge graph model, mainly based on graph theory combined with fuzzy inference, is a new research direction in recent years. This model is applied to detect potentially data patterns that are prone to disease and assist doctors in diagnosis. To evaluate the performance of the proposed model, the implementations are performed on a data set collected from the doctors at Hung Yen General Hospital. Experimental results show that our new model gives better results than the compared models.
Revised: 29/11/2023	
Published: 29/11/2023	
KEYWORDS	
Diabetes	
Fuzzy Knowledge Graph	
Diagnosis support	
Performance	
Accuracy	

ỨNG DỤNG ĐỒ THỊ TRI THỨC MỜ TRONG HỖ TRỢ CHẨN ĐOÁN CHO BỆNH NHÂN BỊ ĐÁI THÁO ĐƯỜNG

Phạm Minh Chuẩn¹, Trần Mạnh Tuấn², Cù Kim Long³, Nguyễn Hồng Tân^{4*}

¹Trường Đại học Sư phạm Kỹ thuật Hưng Yên, ²Trường Đại học Thủy lợi

³Trung tâm Công nghệ thông tin - Bộ Khoa học và Công nghệ

⁴Trường Đại học Công nghệ Thông tin và Truyền thông - ĐHTH Thái Nguyên

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 02/11/2023	Nhu cầu chăm sóc sức khỏe con người là hết sức cần thiết trong thời gian gần đây. Trong số các bệnh thường gặp, Đái tháo đường là một trong những căn bệnh nguy hiểm và rất nhiều người đang mắc phải căn bệnh này. Tuổi tác, béo phì, lười vận động, tiểu đường di truyền, lối sống, chế độ ăn uống không hợp lý, cao huyết áp,... đều là nguyên nhân gây ra bệnh Đái tháo đường. Những người mắc bệnh tiểu đường có nguy cơ cao mắc các bệnh như bệnh tim, bệnh thận, đột quỵ, các vấn đề về mắt, tổn thương thần kinh,... Các công cụ công nghệ thông tin được sử dụng trong hỗ trợ chẩn đoán giúp cho các bác sĩ phát hiện tình trạng bệnh của bệnh nhân một cách nhanh chóng chính xác để từ đó hỗ trợ bác sĩ sớm đưa ra các phác đồ điều trị phù hợp cho người bệnh. Trong bài này, chúng tôi tập trung nghiên cứu mô hình đồ thị tri thức mờ trong việc hỗ trợ chẩn đoán bệnh đái tháo đường. Mô hình đồ thị tri thức dựa trên lý thuyết đồ thị kết hợp với suy diễn mờ là hướng nghiên cứu mới trong thời gian gần đây. Mô hình này được sử dụng để phát hiện các mẫu dữ liệu có khả năng mắc bệnh và trợ giúp bác sĩ trong chẩn đoán. Để đánh giá hiệu năng của mô hình, các thực nghiệm mô hình đồ thị tri thức được thực hiện trên bộ dữ liệu thu thập từ các bác sĩ tại Bệnh viện đa khoa Hưng Yên. Kết quả thực nghiệm cho thấy mô hình đồ thị tri thức mờ cho kết quả tốt hơn các mô hình so sánh.
Ngày hoàn thiện: 29/11/2023	
Ngày đăng: 29/11/2023	
TỪ KHÓA	
Đái tháo đường	
Đồ thị tri thức mờ	
Hỗ trợ chẩn đoán	
Hiệu năng	
Độ chính xác	

DOI: <https://doi.org/10.34238/tnu-jst.9132>

*Corresponding author. Email: nhtan@ictu.edu.vn

1. Giới thiệu

Đái tháo đường (hay còn gọi là tiểu đường) là một trong những bệnh không lây phổ biến ở nhiều nước trên thế giới. Theo thống kê năm 2021 của Liên đoàn đái tháo đường thế giới, hiện nay có khoảng 537 triệu người trưởng thành (20-79 tuổi) đang sống chung với bệnh đái tháo đường, và con số này được dự đoán sẽ tăng lên khoảng 643 triệu vào năm 2030, chủ yếu là tại các quốc gia có thu nhập thấp và trung bình [1]. Tại Mỹ, tổng chi phí y tế ước tính vào năm 2017 liên quan đến điều trị bệnh đái tháo đường là 327 tỷ USD, trong đó chi phí trực tiếp y tế là 237 tỷ USD, với chi phí thuốc là 71,2 tỷ USD. Trung bình, những người mắc bệnh đái tháo đường có mức chi trả cho y tế cao hơn khoảng 3,3 lần so với mức chi trả của người không mắc bệnh [2].

Ở Việt Nam, bệnh đái tháo đường được dự báo là một trong bảy căn bệnh gây tử vong và tàn tật hàng đầu vào năm 2030. Theo dữ liệu cập nhật của Liên đoàn đái tháo đường thế giới, năm 2021 Việt Nam có gần 4 triệu người mắc bệnh, với tỷ lệ được điều chỉnh theo tuổi là 6,1%. Trong đó, tỷ lệ mắc bệnh chưa được chẩn đoán là 51,5%. Chi phí trung bình cho điều trị của mỗi bệnh nhân đái tháo đường là 418,1 USD/năm [1], [3]. Chi phí cho điều trị bệnh đái tháo đường là rất lớn, với thành phần chi phí rất đa dạng, ngoài các chi phí trực tiếp y tế, còn có các chi phí trực tiếp ngoài y tế và các chi phí gián tiếp khác. Đái tháo đường là bệnh mạn tính. Người bệnh phải dùng thuốc liên tục và suốt đời, tạo ra một gánh nặng kinh tế không nhỏ cho bản thân họ và gia đình, cũng như cho hệ thống y tế.

Để phát hiện bệnh tiểu đường, rất nhiều công nghệ và thuật toán đã được các nhà nghiên cứu sử dụng trong vài năm qua. Học máy (Machine Learning - ML) là một trong số các công nghệ đó. Trong cuộc cách mạng công nghiệp lần thứ tư, học máy đã được chứng minh là một công cụ có giá trị trong các lĩnh vực khác nhau, bao gồm cả chăm sóc sức khỏe [4]–[6]. Trí tuệ nhân tạo, khai phá dữ liệu, mạng nơ ron nhân tạo và nhiều mạng khác được coi là các nhánh thiết yếu của ML. Việc áp dụng các công nghệ tiên tiến vào lĩnh vực chăm sóc sức khỏe là rất quan trọng, đặc biệt trong việc phát hiện bệnh đái tháo đường [7]. Quá trình phát hiện bệnh và suy luận bệnh gặp một số khó khăn khi sử dụng công nghệ học máy [8]. Trong khi hầu hết các công nghệ học máy, trí tuệ nhân tạo có thể được sử dụng để dự đoán bệnh một cách chính xác, thì quá trình thiết kế và tiến trình suy diễn thường không thể diễn giải được, khiến chúng trở lên khó hiểu. Vì vậy, việc sử dụng các công nghệ có thể giải thích được và dễ hiểu đối với con người cũng đóng vai trò quan trọng. Hơn nữa, một nhược điểm khác của công nghệ học máy, học sâu là chúng không thể giải quyết một cách hiệu quả đối với các bài toán có dữ liệu có tính không chắc chắn.

Logic mờ được phát triển để giải quyết những vấn đề xảy ra khi cần biểu diễn dữ liệu có tính mơ hồ, không chắc chắn. Logic mờ được giới thiệu lần đầu tiên bởi Zadeh [9] và được xem như là mở rộng của logic đại số. Điểm khác biệt của logic mờ là các giá trị nằm trong khoảng từ 0 đến 1, được gọi là mức độ thành viên (hoặc độ thuộc). Logic mờ được trang bị các lập luận giống như con người ở cách tư duy hệ thống nên nó phù hợp khi áp dụng cho bài toán mà dữ liệu có tính chất không rõ ràng, không chắc chắn. Bằng khả năng biểu diễn và tính toán mạnh mẽ với dữ liệu mơ hồ và không chắc chắn, logic mờ là một công cụ có giá trị cho các bài toán phân lớp. Hơn nữa, việc sử dụng các hệ thống dựa trên quy tắc mờ (Fuzzy Rule Based System - FRBS), các luật dạng IF - THEN chứng minh khả năng biểu diễn, phù hợp với bài toán phân lớp và làm tăng tính diễn giải của hệ thống [10]. So với các công nghệ học máy hiện đại, các hệ thống FRBS có tính diễn giải cao khi các luật IF - THEN được biểu diễn dạng biến ngôn ngữ [11], [12]. Với những đặc điểm như vậy, logic mờ được khẳng định là một kỹ thuật hữu ích để dự đoán chính xác và sớm bệnh đái tháo đường.

Một hướng nghiên cứu gần đây là biểu diễn và suy luận tri thức dựa vào đồ thị tri thức và đồ thị tri thức mờ được cộng đồng nghiên cứu quan tâm rộng rãi. Đồ thị tri thức (*Knowledge Graph*) [13] là một kỹ thuật được sử dụng để giải quyết các bài toán phân loại, hỗ trợ ra quyết định rất hiệu quả. Tuy nhiên, đồ thị tri thức đã gặp khó khăn trong các quá trình biểu diễn đồ thị và suy luận xấp xỉ trên các tập dữ liệu đầu vào có chứa các thông tin nhập nhằng, mơ hồ, không rõ ràng. Gần đây, đồ thị tri thức mờ (*Fuzzy Knowledge Graph - FKG*) lần đầu được giới thiệu vào năm 2020 và được tích hợp trong mô hình M-CFIS-FKG [14]. Nó đã được đưa ra để giải quyết những

hạn chế của đồ thị tri thức bằng cách sử dụng các nhãn ngôn ngữ cho các thuộc tính trong tập dữ liệu huấn luyện. Đầu tiên, một đồ thị tri thức mờ được xây dựng sau giai đoạn chuẩn bị dữ liệu (hay còn gọi là biểu diễn đồ thị). Sau đó, quá trình suy luận xấp xỉ được sử dụng để tìm nhãn của các bản ghi mới trong tập dữ liệu kiểm thử. Đồ thị tri thức mờ suy luận dựa trên luật IF-THEN, trong đó tác động của nhãn ngôn ngữ có khả năng tạo ra nhãn đầu ra tương ứng.

Trong nghiên cứu này, một hướng tiếp cận mới dựa trên mô hình FKG được đề xuất và áp dụng cho bài toán chẩn đoán bệnh đái tháo đường. Bộ dữ liệu huấn luyện và kiểm tra được cung cấp từ bệnh viện Đa khoa tỉnh Hưng Yên. Những đóng góp chính về mặt khoa học của bài báo này bao gồm:

- Thu thập và tiền xử lý dữ liệu về bệnh đái tháo đường tại bệnh viện Đa khoa Hưng Yên;
- Xây dựng mô hình đồ thị tri thức mờ cho hỗ trợ chẩn đoán bệnh đái tháo đường;
- Cài đặt, thực nghiệm mô hình trên bộ dữ liệu đã thu thập.

Nội dung tiếp theo của bài báo được cấu trúc như sau: Phần 2 hệ thống dựa trên luật mờ, đồ thị tri thức mờ, cách xây dựng đồ thị tri thức mờ trong hỗ trợ chẩn đoán bệnh đái tháo đường. Môi trường cài đặt, các kết quả thực nghiệm để kiểm chứng mô hình đồ thị tri thức mờ với bộ dữ liệu bệnh đái tháo đường được trình bày ở phần 3. Phần cuối cùng là kết luận và một số hướng phát triển của các nghiên cứu tiếp theo.

2. Mô hình đồ thị tri thức mờ trong hỗ trợ chẩn đoán đái tháo đường

2.1. Hệ thống dựa trên luật mờ (FRBS)

FRBS là sự mở rộng của các hệ thống dựa trên quy tắc cổ điển. Hệ thống này còn được gọi là hệ thống sản xuất hoặc các hệ thống chuyên gia. Về cơ bản, chúng được thể hiện dưới dạng luật "IF A THEN B" trong đó A và B là các tập mờ. A được gọi là phần tiền đề và B được gọi là phần hệ quả của quy tắc. Dựa vào cấu trúc thiết lập các luật, có hai mô hình FRBS cơ bản: mô hình Mamdani và mô hình Takagi Sugeno Kang [15].

Mô hình Mamdani [15]

Mô hình này được giới thiệu bởi Mamdani năm 1974 và được phát triển bởi Mamdani và Assilian năm 1975. Mô hình Mamdani bao gồm các biến ngôn ngữ trong cả phần tiền đề và phần hệ quả của các luật. Mô hình này được xem là hệ thống nhiều đầu vào và một đầu ra (Multi Input Single Output - MISO), các luật IF-THEN mờ được mô tả ở dạng như sau:

$$\text{IF } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_2 \text{ and } \dots \text{ and } X_n \text{ is } A_n \text{ THEN } Y \text{ is } B \quad (1)$$

trong đó, X_i và Y lần lượt là các biến ngôn ngữ đầu vào và đầu ra, còn A_i và B là các giá trị ngôn ngữ.

Kiến trúc tiêu chuẩn cho mô hình Mamdani, bao gồm bốn thành phần: Giao diện mờ hóa, Cơ sở tri thức, Bộ suy diễn và Bộ giải mờ. Trong đó:

- Giao diện mờ hóa chuyển đổi các đầu vào rõ thành các giá trị ngôn ngữ.
- Cơ sở tri thức gồm Cơ sở dữ liệu và Cơ sở luật mờ. Cơ sở dữ liệu bao gồm các định nghĩa tập mờ và tham số của các hàm thành viên còn cơ sở luật mờ chứa tập hợp các luật mờ IF THEN.
- Bộ suy luận thực hiện các phép toán suy luận trên luật mờ và dữ liệu đầu vào.
- Bộ giải mờ tạo ra các giá trị sắc nét từ các giá trị ngôn ngữ như kết quả cuối cùng.

Mô hình Takagi Sugeno Kang (TSK) [15]

Mô hình TSK tương tự như mô hình Mamdani, chỉ khác ở phần hệ quả của các luật IF - THEN và được mô tả như công thức (2) dưới đây:

$$\text{IF } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_2 \text{ and } \dots \text{ and } X_n \text{ is } A_n \text{ THEN } Y = f(X_1, X_2, \dots, X_n) \quad (2)$$

Trong đó, $Y = f(X_1, X_2, \dots, X_n)$ là tổ hợp tuyến tính của các biến đầu vào. Vì có một hàm trên phần hệ quả của luật IF THEN, đầu ra là giá trị thực, do đó trên mô hình TSK không có bộ giải mờ.

2.2. Đồ thị tri thức mờ

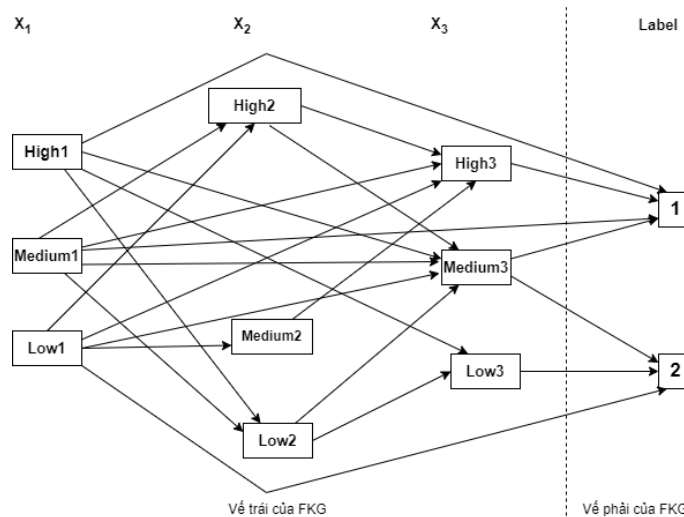
Đồ thị tri thức mờ [14] là bộ gồm năm thành phần được ký hiệu là $FKG = (R, V, L, A, B)$, trong đó:

- $R = \{R_j | j = \underline{1}, \underline{n}\}$ là tập các luật trong cơ sở luật mờ.
- $V = \{X_i | i = \underline{1}, \underline{m}\}$ là tập các thuộc tính đầu vào, mỗi thuộc tính được biểu diễn thông qua giá trị biến ngôn ngữ (sau đây gọi là vé trái của đồ thị FKG).
- $L = \{1, 2, \dots, C\}$ là tập đỉnh nhân đầu ra (sau đây gọi là vé phải của đồ thị FKG).
- A : là tập các cạnh biểu diễn mối quan hệ giữa các đỉnh nhân thuộc tính đầu vào.
- B : là tập các cạnh biểu diễn mối quan hệ giữa các đỉnh thuộc tính đầu vào và đỉnh nhân đầu ra.

Ví dụ 2.1. Ta có tập các luật mờ như Bảng 1. Khi đó hình ảnh của đồ thị tri thức mờ được xây dựng như trên Hình 1.

Bảng 1. Cơ sở luật mờ với sáu luật

Luật	X_1	X_2	X_3	Nhân đầu ra
R_1	Medium1	High2	High3	1
R_2	High1	Low2	Low3	2
R_3	Low1	Medium2	High3	1
R_4	Low1	High2	Medium3	1
R_5	High1	Low2	Medium3	2
R_6	Medium1	Low2	Low3	2



Hình 1. Đồ thị FKG dựa trên cơ sở luật mờ

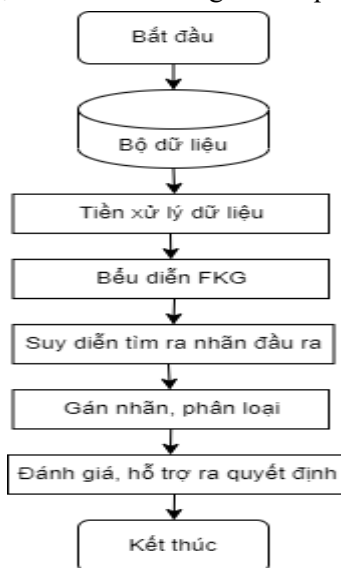
2.3. Mô hình đồ thị tri thức mờ trong hỗ trợ chẩn đoán đái tháo đường

Phần này trình bày mô hình hỗ trợ ra quyết định dựa trên FKG và ứng dụng FKG trong hỗ trợ chẩn đoán bệnh đái tháo đường.

Mô hình hỗ trợ ra quyết định dựa trên FKG gồm một số bước chính được biểu diễn như trong Hình 2.

Cụ thể, các bước thực hiện của mô hình được chi tiết hóa như sau:

Bước 1. Thu thập bộ dữ liệu bệnh đái tháo đường từ kết quả khám chữa bệnh.



Hình 2. Sơ đồ tổng quan của mô hình hỗ trợ ra quyết định dựa trên FKG

Bước 2. Tiền xử lý dữ liệu.

- ✓ Sử dụng ý kiến chuyên gia để xây dựng một thang đo với từng thuộc tính đầu vào.
- ✓ Thống nhất cách sử dụng các giá trị biến ngôn ngữ đối với từng thuộc tính đầu vào.
- ✓ Tham khảo ý kiến chuyên gia để loại bỏ các nhiễu và làm sạch bộ dữ liệu.
- ✓ Thực hiện một số phép toán thống kê cơ bản trên bộ dữ liệu để hiểu biết về phân bố của dữ liệu và kiểm chứng độ tin cậy của bộ dữ liệu.
- ✓ Nhận cơ sở luật mờ từ quá trình suy diễn mờ để tiến hành bước tiếp theo.

Bước 3. Biểu diễn FKG [16].

- Tính toán bộ trọng số cạnh $A_{ij\dots k}^t$ bằng cách áp dụng công thức (3)

$$A^t(X_i X_j \dots X_k) = \frac{|\sigma(X_i) \rightarrow \sigma(X_j) \rightarrow \dots \rightarrow \sigma(X_{k+1}) \text{ in rule } k|}{n} \quad (3)$$

trong đó, $1 \leq i < j < \dots < k \leq m - 1, t = 1, n$;

n : số mẫu; $\sigma(X_i), \sigma(X_j), \dots, \sigma(X_k)$: giá trị ngôn ngữ của thuộc tính X_i, X_j, \dots, X_k .

- Tính toán bộ trọng số cạnh $B_{ij\dots kl}^t$ bằng cách áp dụng công thức (4)

$$B_{ij\dots kl}^t = \left(\sum_{j=i+1}^m A_{ij\dots k+1}^t \right) \quad (4)$$

$$\times \text{MIN} \left(\frac{|\sigma(X_i) \rightarrow l|}{n}, \frac{|\sigma(X_j) \rightarrow l|}{n}, \dots, \frac{|\sigma(X_k) \rightarrow l|}{n} \right)$$

với $1 \leq i < j < \dots < k < m, t = 1, n, l = \{1, 2, \dots, C\}$.

- Biểu diễn FKG sau khi tính các bộ trọng số $A_{ij\dots k}^t$ và $B_{ij\dots kl}^t$.

Bước 4. Suy diễn tìm nhãn đầu ra.

- Tính bộ trọng số $C_{ij\dots kl}$ được xác định dựa trên tổng trọng số các quan hệ đến nhãn ($B_{ij\dots kl}^t$) theo công thức (5):

$$C_{ij\dots kl} = \sum_{t=1}^n B_{ij\dots kl}^t \quad 1 \leq i < j < \dots < k' \leq k, l = \{1, 2, \dots, C\} \quad (5)$$

$C_{ij\dots kl}$: được sử dụng để xác định mức độ ảnh hưởng của từng nhãn đến giá trị ngôn ngữ của thuộc tính trên dữ liệu cần gán nhãn.

- Tính giá trị trọng số D_l bằng cách áp dụng toán tử **max - min** theo công thức (6):

$$D_l = \max_{1 \leq i < j < \dots < k' < k} (C_{ij \dots k'l}) + \min_{1 \leq i < j < \dots < k' < k} (C_{ij \dots k'l}) \quad (6)$$

D_l : xác định mức độ khả năng nhận có thể nhận được trên bộ dữ liệu cần tìm nhãn

- Tìm nhãn đầu ra sau khi áp dụng toán tử **max** theo công thức (7):

$$\text{Nhãn đầu ra} = p \text{ nếu } D_p = \max_{l=1, \dots, C} (D_l) \quad (7)$$

Bước 5. Gán nhãn đầu ra, phân loại các mẫu mới.

Bước 6. Đánh giá, hỗ trợ ra quyết định (kết luận cuối cùng có thể kết hợp ý kiến chuyên gia nếu cần thiết).

3. Kết quả thực nghiệm

Nhóm nghiên cứu đã thực hiện thực nghiệm trên bộ dữ liệu về các bệnh nhân bị đái tháo đường ở thời kỳ mang thai thu thập tại bệnh viện Đa khoa Hưng Yên. Tập dữ liệu (BN-DaiThaoDuong_ThaiNghen) được sử dụng để thực nghiệm gồm 759 bệnh nhân với 22 thuộc tính liên quan và kết luận của bác sĩ tại bệnh viện.

Môi trường thực nghiệm tiến hành trên máy tính xách tay (DELL Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz 2.70 GHz), ngôn ngữ Matlab 2014 đã được cài đặt, dữ liệu được chia theo tỉ lệ 70% để training (530 mẫu để Training) và 30% dùng để testing (229 mẫu để Testing). Độ đo để đánh giá hiệu năng của mô hình, chúng tôi sử dụng độ đo độ chính xác (accuracy) và thời gian tính toán (computational time). Chúng tôi thực hiện đánh giá hiệu năng của các phương pháp FKG-Pairs khác nhau (bao gồm cặp đôi: là cặp giá trị ngôn ngữ của 2 thuộc tính, cặp ba: là cặp giá trị ngôn ngữ của 3 thuộc tính, và cặp tư: là cặp giá trị ngôn ngữ của 4 thuộc tính).

Bảng 2. Kết quả thực nghiệm

	Thời gian (Giây)	Độ chính xác
FKG1	1,593	0,783
FKG2	1,942	0,823
FKG3	2,583	0,832
FKG4	3,152	0,854

Trong kết quả thực nghiệm tại Bảng 2 chúng tôi chứng tỏ rằng phương pháp FKG-Pairs4 tốt nhất về mặt độ chính xác (0,854), nhưng thời gian cao gấp khoảng hai lần so FKG-Pairs2. Điều này cho việc suy luận và chẩn đoán bệnh dựa trên đồ thị tri thức mờ dạng cặp (FKG-Pairs) về mặt tham số độ chính xác đã được tăng lên khi tăng tổ hợp kết hợp giá trị ngôn ngữ của các thuộc tính (theo mức độ cặp đôi, cặp ba, cặp tư), trong khi tham số thời gian tính toán lại tăng nhanh. Việc tăng sự kết hợp các cặp giá trị ngôn ngữ của các thuộc tính thì độ chính xác của kết quả hỗ trợ chẩn đoán tăng nhưng thời gian thực thi thuật toán cũng tăng theo. Phù hợp với thực tế khi có nhiều giá trị các thuộc tính (các thông số xét nghiệm) được kết hợp với nhau thì kết quả hỗ trợ chẩn đoán được chính xác hơn.

4. Kết luận

Trong bài báo này, chúng tôi đã trình bày một phương pháp hỗ trợ chẩn đoán bệnh trên đồ thị tri thức mờ dạng cặp (FKG-Pairs). Bài báo có một số đóng góp chính như sau: (i) biểu diễn đồ thị và hỗ trợ chẩn đoán bệnh. Trong đó, nhóm tác giả đã cải tiến xây dựng mô FKG-Pairs vào trong các bài toán hỗ trợ chẩn đoán bệnh; (ii) đã cài đặt, thu thập dữ liệu thực tế tại bệnh viện Đa khoa Hưng Yên; (iii) kết quả thực nghiệm dựa trên độ đo Accuracy, Time cũng đã cho thấy việc tăng sự kết hợp giữa các cặp thuộc tính thì mức độ chính xác được tăng lên nhưng thời gian thực thi mô hình tăng. Nghiên cứu này, làm tiền đề cho việc phát triển nghiên cứu sau này cho việc giải một số bài toán hỗ trợ chẩn đoán y tế.

Lời cảm ơn

Nghiên cứu này được thực hiện dưới sự tài trợ bởi đề tài mã số: B2022-SKH-01 từ Bộ Giáo dục và Đào tạo.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] International Diabetes Federation (IDF), “The IDF Diabetes Atlas 10th Edition,” 2021.
- [2] American Diabetes Association, “Economic Costs of Diabetes in the U.S. in 2017,” 2018, pp. 917-928.
- [3] B. N. Nguyen, L. Z. Lu, and A. Waqas, “Diabetes: What Challenges Lie Ahead for Vietnam?” *Ann. Glob. Health*, vol. 86, no. 1, pp. 1-9, Jan 2, 2020.
- [4] M. F. Ijaz, M. Attique, and Y. Son, “Data-Driven Cervical Cancer Prediction Model with Outlier Detection and Over-Sampling Methods,” *Sensors*, vol. 20, no. 10, 2020, Art. no. 2809, doi: 10.3390/s20102809.
- [5] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, “Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM,” *Sensors*, vol. 21, 2021, Art. no. 2852, doi: 10.3390/s21082852.
- [6] M. Mandal, P.K. Singh, M.F. Ijaz, J. Shafi, and R. Sarkar, “A Tri-Stage Wrapper-Filter Feature Selection Framework for Disease Classification,” *Sensors*, vol. 21, no. 16, 2021, Art. no. 5571, doi: 10.3390/s21165571.
- [7] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, “Hybrid Prediction Model for Type 2 Diabetes and Hypertension Using DBSCAN Based Outlier Detection, Synthetic Minority Over Sampling Technique (SMOTE), and Random Forest,” *Appl. Sci.*, vol. 8, 2018, Art. no. 1325, doi: 10.3390/app8081325.
- [8] O. Biran and C. Cotton, “Explanation and Justification in Machine Learning: A Survey,” *IJCAI-17*, 2017, pp. 8–13.
- [9] L.A. Zadeh, *Advances in Fuzzy Systems—Applications and Theory*, Word Scientific Publisher: Singapore, 1996, pp. 394–432.
- [10] F. Mehmanpazir and S. Asadi, “Development of an evolutionary fuzzy expert system for estimating future behavior of stock price,” *J. Ind. Eng. Int.*, vol. 13, pp. 29–46, 2017.
- [11] F. Mansourypoor and S. Asadi, “Development of a Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System for diabetes diagnosis,” *Comput. Biol. Med.*, vol. 91, pp. 337–352, 2017.
- [12] S. Guillaume and B. Charnomordic, “Learning interpretable fuzzy inference systems with FisPro,” *Inf. Sci.*, vol. 181, pp. 4409–4427, 2011.
- [13] P. A. Bonatti, S. Decker, A. Polleres, and V. Presutti, “Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371),” *Dagstuhl Reports*, vol. 8, no. 9, pp. 29-111, Dagstuhl-Leibniz Castle Center for Computer Science, 2019.
- [14] T. H. L. Luong, M. T. Tran, T. N. Tran, H. S. Le, L. G. Nguyen, T. N. N. Vo, and V. H. Pham, “A new complex fuzzy inference system with fuzzy knowledge graph and extensions in decision making,” *IEEE Access*, vol. 8, pp. 164 899– 164 921, 2020.
- [15] L. S. Riza, C. Bergmeir, F. Herrera, and J. M. Benitez, “frbs: Fuzzy Rule-Based Systems for Classification and Regression in R,” *Journal of Statistical Software*, vol. 65, no. 6, pp. 1–30, 2015.
- [16] K. L. Cu, V. H. Pham, M. T. Tran, T. H. L. Luong, M. C. Pham, and H. S. Le, “A novel fuzzy knowledge graph pairs approach in decision making,” *Multimedia Tools and Applications*, vol. 81, no. 18, pp. 26505-26534, 2022.