

ĐẠI HỌC BÁCH KHOA HÀ NỘI
TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



SOICT

PROJECT II

BÁO CÁO GIỮA KỲ

CHUẨN ĐOÁN BỆNH TIM MẠCH BẰNG SỬ DỤNG MÔ HÌNH SƠ ĐỒ TRI THỨC MỜ DẠNG CẶP FKG-PAIRS

Giảng viên hướng dẫn: PGS.TS Phạm Văn Hải

Sinh viên thực hiện: Đào Thành Mạnh 20211014

Hà Nội, Ngày 20 tháng 4 năm 2024

MỤC LỤC

MỤC LỤC.....	2
1. Bài toán.....	4
2. Mục tiêu nghiên cứu	4
2.1. Mục tiêu chung	4
2.2. Mục tiêu cụ thể	4
3. Dữ liệu đầu vào	5
3.1. Tập dữ liệu.....	5
3.2. Xử lý dữ liệu.....	6
4. Kiến thức liên quan	7
4.1. Tập mờ và logic mờ.....	7
4.2. Hệ suy diễn mờ.....	8
4.3. Đồ thị tri thức mờ	8
4.4. Suy luận xấp xỉ.....	9
5. Mô hình và thuật toán.....	9
5.1. Phát biểu bài toán	9
5.2. Mô hình đề xuất	10
5.2.1. Mô hình bài toán.....	10
5.2.2. Chức năng hệ thống.....	11
5.2.3. Phương pháp thực hiện	12
5.3. Thuật toán cài đặt FKG-Pairs	13
7. Ví dụ thực tế về FKG-Pairs-1 và FKG-Pairs-2 chuẩn đoán bệnh tim	14
7.1. Chuẩn đoán bệnh tim bệnh nhân dựa trên mô hình M-CFIS-FKG sử dụng FKG-Pairs-1	14
7.2. Chuẩn đoán bệnh tim bệnh nhân dựa trên mô hình M-CFIS-FKG sử dụng FKG-Pairs-2.....	19
8. Kết quả đánh giá 2 mô hình FKG-Pairs-1 và FKG-Pairs-2 bằng phương pháp toán học.....	27
9. Dự kiến tiến độ công việc tiếp theo	28

10.	Tài liệu tham khảo	28
------------	---------------------------------	-----------

1. Bài toán

Trong lĩnh vực chăm sóc sức khỏe, đặc biệt là trong việc đưa ra quyết định và chuẩn đoán về bệnh tim mạch, việc có sự hiểu biết sâu sắc về thông tin y tế và khả năng áp dụng các phương pháp và kiến thức y tế hiện đại là vô cùng quan trọng. Tuy nhiên, mặc dù đã có sự tiến bộ trong việc thu thập và tổ chức dữ liệu y tế, những vấn đề vẫn tồn tại do tính phức tạp và đa dạng của thông tin này, đặc biệt là khi đối mặt với các bệnh lý như bệnh tim mạch.

Sơ đồ tri thức (KG) truyền thống không thể hoàn toàn giải quyết các thách thức này. Mặc dù KG có thể tổ chức dữ liệu y tế và xây dựng các mối quan hệ giữa các yếu tố y tế khác nhau, nhưng nó thường không thể xử lý được tính mờ và không chắc chắn của thông tin y tế, cũng như không đủ linh hoạt để áp dụng các quy tắc và quan hệ phức tạp trong việc đưa ra quyết định và chuẩn đoán về bệnh tim mạch.

Trong khi đó, Sơ đồ Tri thức Mờ dạng cặp (FKG) mở ra một tiềm năng mới để cải thiện quá trình ra quyết định và chuẩn đoán về bệnh tim mạch. FKG không chỉ giúp biểu diễn thông tin y tế một cách linh hoạt hơn và xử lý tính mờ và không chắc chắn của dữ liệu, mà còn giải quyết vấn đề hiệu suất thấp của KG thông thường bằng cách sử dụng các cặp thông tin thay vì chỉ dùng một cặp đơn lẻ. Điều này tăng cường khả năng suy luận và đưa ra quyết định chính xác hơn trong việc điều trị và quản lý bệnh tim mạch, giúp nâng cao chất lượng chăm sóc sức khỏe và kết quả điều trị cho bệnh nhân.

2. Mục tiêu nghiên cứu

2.1. Mục tiêu chung

Phát triển một hệ thống/ứng dụng hỗ trợ ra quyết định và chuẩn đoán bệnh tim mạch dựa trên mô hình Sơ Đồ Tri Thức Mờ Dạng Cặp FKG-Pairs. Dự án tập trung vào xây dựng một cơ sở dữ liệu y tế đa dạng và chất lượng cao, sử dụng FKG-Pairs để biểu diễn thông tin y tế một cách linh hoạt và hiệu quả. Mục tiêu là phát triển một mô hình dự đoán bệnh tim mạch đáng tin cậy và chính xác, kết hợp cả khía cạnh dự đoán và phân tích giải thích. Cuối cùng, dự án sẽ đánh giá và tối ưu hóa hiệu suất của hệ thống trước khi triển khai vào thực tế y tế, nhằm cung cấp một công cụ hữu ích giúp nâng cao chất lượng chuẩn đoán và điều trị bệnh tim mạch, đồng thời tăng cường sự hiểu biết và tin cậy từ phía người sử dụng.

2.2. Mục tiêu cụ thể

Xây dựng các mô hình Sơ Đồ Tri Thức Mờ Dạng Cặp FKG-Pairs với các cặp FKG pairs k ($k = 1, 2, 3$), phát triển thuật toán và phương pháp để xây dựng mô hình FKG-Pairs với độ chính xác và linh hoạt cao, tiến hành đánh giá hiệu năng

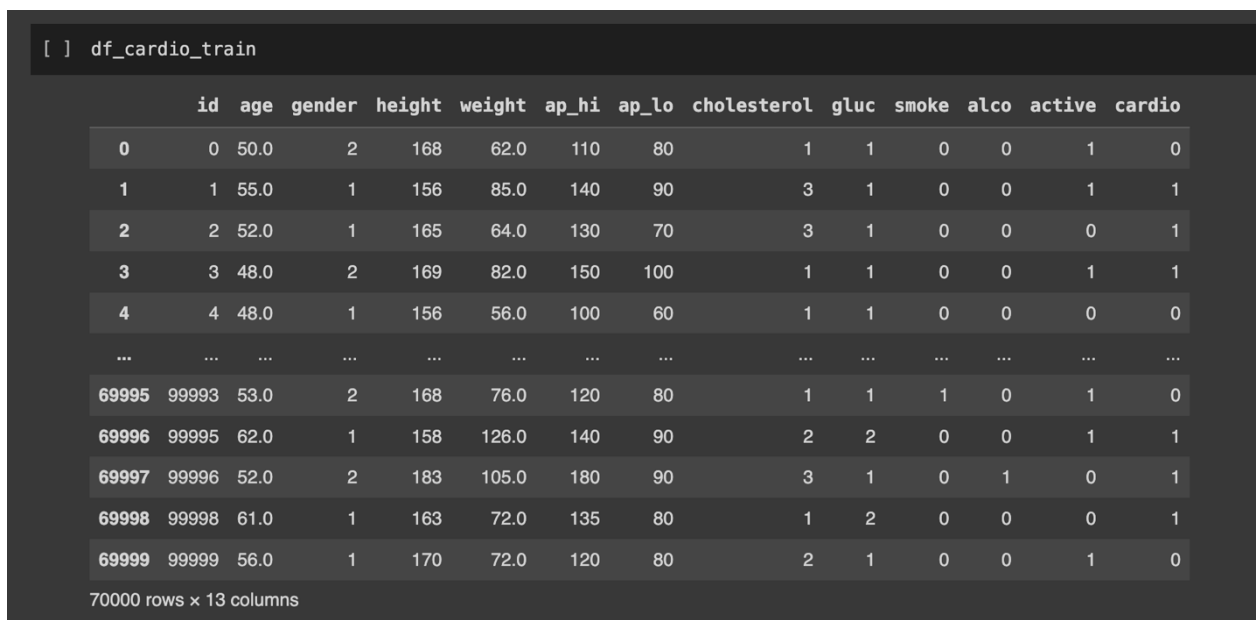
của các mô hình FKG-Pairs dựa trên các tiêu chí như độ chính xác, độ phân loại và độ tin cậy, so sánh và phân tích kết quả của các mô hình FKG-Pairs để xác định mô hình tối ưu nhất cho việc chuẩn đoán bệnh tim mạch, tối ưu hóa và cải thiện mô hình tối ưu nhất để đảm bảo hiệu suất và độ chính xác cao nhất trước khi triển khai vào thực tế y tế, cung cấp hướng dẫn và báo cáo chi tiết về quy trình và kết quả của dự án để chia sẻ kinh nghiệm và kiến thức cho cộng đồng y tế và nghiên cứu.

3. Dữ liệu đầu vào

3.1. Tập dữ liệu

Dữ liệu thực nghiệm được thu thập trực tiếp từ trang chính của Kaggle: <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data>

Bộ dữ liệu bao gồm 70 000 dữ liệu bệnh nhân, 11 thuộc tính được lưu dưới dạng file **cardio_train.csv**:



	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	50.0	2	168	62.0	110	80	1	1	0	0	1	0
1	1	55.0	1	156	85.0	140	90	3	1	0	0	1	1
2	2	52.0	1	165	64.0	130	70	3	1	0	0	0	1
3	3	48.0	2	169	82.0	150	100	1	1	0	0	1	1
4	4	48.0	1	156	56.0	100	60	1	1	0	0	0	0
...
69995	99993	53.0	2	168	76.0	120	80	1	1	1	0	1	0
69996	99995	62.0	1	158	126.0	140	90	2	2	0	0	1	1
69997	99996	52.0	2	183	105.0	180	90	3	1	0	1	0	1
69998	99998	61.0	1	163	72.0	135	80	1	2	0	0	0	1
69999	99999	56.0	1	170	72.0	120	80	2	1	0	0	1	0

70000 rows x 13 columns

Mô tả dữ liệu

Có 3 loại tính năng đầu vào:

- Mục tiêu: thông tin thực tế;
- Khám: kết quả khám sức khỏe;
- Chủ quan: thông tin do bệnh nhân cung cấp.

Thuộc tính:

Tuổi | Tính năng khách quan | tuổi | int (ngày)

Chiều cao | Tính năng khách quan | chiều cao | int (cm) |
Cân nặng | Tính năng khách quan | cân nặng | phao (kg) |
Giới tính | Tính năng khách quan | giới tính | mã phân loại |
Huyết áp tâm thu | Tính năng kiểm tra | ap_hi | int |
Huyết áp tâm trương | Tính năng kiểm tra | ap_lo | int |
Cholesterol | Tính năng kiểm tra | cholesterol | 1: bình thường, 2: trên mức bình thường, 3: trên mức bình thường |
Đường | Tính năng kiểm tra | gluc | 1: bình thường, 2: trên mức bình thường, 3: trên mức bình thường |
Hút thuốc | Tính năng chủ quan | khói | nhị phân |
Uống rượu | Tính năng chủ quan | rượu | nhị phân |
Hoạt động thể chất | Tính năng chủ quan | hoạt động | nhị phân |
Sự hiện diện hay vắng mặt của bệnh tim mạch | Biến mục tiêu | tim mạch | nhị phân |
Tất cả các giá trị tập dữ liệu được thu thập tại thời điểm kiểm tra y tế.

3.2. Xử lý dữ liệu

Xử lý tập dữ liệu lấy từ kaggle bằng Google Colab sau đó lưu dữ liệu đã xử lý vào Drive:

Link:

https://colab.research.google.com/drive/1m6IaNr2wEebYCINRZ1peqyE_DYsZrUT-?usp=sharing

Tập dữ liệu sau khi xử lý xong đã tách thành hai tập dữ liệu bệnh nhân mắc bệnh tim và bệnh nhân không mắc bệnh tim chia theo tỉ lệ 1:1 để tiện sử dụng trong quá trình viết các luật tri thức mờ và huấn luyện các mô hình FKG-Pairs-k.

Dữ liệu sau khi xử lý xong được trả về kết quả như hình dưới đây:

Tập **positive_data_3000.csv**:

positive_data_3000

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	1	55.0	1	156	85.0	140	90	3	1	0	0	1	1
1	2	52.0	1	165	64.0	130	70	3	1	0	0	0	1
2	3	48.0	2	169	82.0	150	100	1	1	0	0	1	1
3	4	62.0	2	178	95.0	130	90	3	3	0	0	1	1
4	5	46.0	2	172	112.0	120	80	1	1	0	0	0	1
...
2995	2996	52.0	1	161	76.0	140	80	1	1	0	0	1	1
2996	2997	53.0	1	160	70.0	110	70	2	1	0	0	1	1
2997	2998	51.0	2	170	69.0	140	90	1	3	0	0	0	1
2998	2999	60.0	2	164	77.0	150	120	1	1	0	0	1	1
2999	3000	59.0	1	151	103.0	140	90	1	1	0	0	1	1

3000 rows x 13 columns

Tập negative_data_3000.csv:

negative_data_3000

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	50.0	2	168	62.0	110	80	1	1	0	0	1	0
4	4	48.0	1	156	56.0	100	60	1	1	0	0	0	0
5	8	60.0	1	151	67.0	120	80	2	2	0	0	0	0
6	9	61.0	1	157	93.0	130	80	3	1	0	0	1	0
8	13	48.0	1	158	71.0	110	70	1	1	0	0	1	0
...
6014	8545	49.0	1	155	87.0	120	80	1	1	0	0	1	0
6018	8552	48.0	1	158	88.0	120	80	1	1	0	0	0	0
6020	8555	54.0	1	147	51.0	110	70	1	1	0	0	1	0
6021	8556	40.0	1	168	67.0	120	80	1	1	0	0	1	0
6023	8559	62.0	2	175	85.0	120	90	1	1	0	0	1	0

3000 rows x 13 columns

4. Kiến thức liên quan

4.1. Tập mờ và logic mờ.

Tập mờ lần đầu được giới thiệu bởi Zadeh vào năm 1965 , được giới thiệu như một công cụ toán học mới để giải quyết các vấn đề với thông tin mơ hồ, không chắc chắn. Khác với các tập hợp thông thường, đánh giá quan hệ thành viên của tập hợp theo logic nhị phân “một phần tử là thuộc hoặc không thuộc tập hợp”, logic mờ đánh giá quan hệ thành viên

của một phần tử thông qua một hàm liên thuộc $\mu \rightarrow [0,1]$, thể hiện độ thuộc của một phần tử đối với một tập hợp

4.2. Hệ suy diễn mờ.

Suy diễn mờ là quá trình tìm ra kết luận cho tập giá trị đầu vào, dựa trên hệ luật mờ đã được tổng hợp. Các phương pháp suy diễn mờ thường được nhắc đến như suy diễn mờ Mamdani, suy diễn mờ Takagi-Sugeno, ... các hệ suy diễn kể trên còn được biết đến là các phương pháp suy diễn cổ điển, đã được sử dụng rộng rãi trong các hệ điều khiển tự động. Đồ thị tri thức mờ được biết đến như một phương pháp suy diễn mới, hiệu quả, và có độ chính xác cao hơn so với các phương pháp suy diễn trước đó. Quy tắc chung cho quá trình áp dụng hệ suy diễn mờ được thể hiện qua ba bước:

- Mờ hóa (fuzzification) : Trong bước này, chúng ta cần xác định được thang đo giá trị và thuật ngữ mức độ tương ứng của từng thuộc tính đầu vào của bộ dữ liệu, sau đó là quá trình chuyển đổi từ các giá trị rõ của bộ dữ liệu đầu vào thành các giá trị mờ, dựa trên thang đo giá trị kết hợp với hàm thuộc đã được xây dựng trước đó, cuối cùng là kết hợp các giá trị mờ của từng mẫu dữ liệu đầu vào bằng các toán tử mờ (AND, OR, NOT) để đưa ra các luật biểu diễn dưới dạng các mệnh đề IF-THEN và đưa vào hệ cơ sở luật mờ
- Suy diễn mờ (fuzzy inference) : Sử dụng phương pháp suy diễn mờ để tìm ra kết quả đầu ra dựa trên hệ cơ sở luật mờ đã xây dựng ở bước 1
- Giải mờ (deuzification) : Chuyển đổi các giá trị đầu ra mờ được tìm ra ở bước 2 thành các giá trị rõ, đưa ra kết quả của bài toán.

4.3. Đồ thị tri thức mờ

Thuật ngữ đồ thị tri thức mờ được giới thiệu lần đầu được tích hợp trong mô hình M-CFIS-FKG với mục đích ban đầu là mở rộng mô hình M-CFIS-R làm cho quá trình suy luận của mô hình này trong pha testing trở nên nhanh hơn. Kế thừa đặc điểm của đồ thị tri thức, nên về mặt hình thức, một đồ thị tri thức mờ bao gồm các đỉnh đại diện cho các nhân ngôn ngữ của thuộc tính và nhãn đầu ra của các luật, các cạnh tương ứng là cung nối giữa các đỉnh.

Cách tính giá trị trọng số của các cạnh của đồ thị tri thức mờ đã được trình bày chi tiết trong [7], và được tóm tắt gọn lại như sau:

- Đối với các cạnh nối giữa hai đỉnh thuộc tính, với mỗi cặp giá trị (X_i, X_j) , $1 \leq i \leq j \leq m$, trong luật thứ t R_t , $t = \overline{1, k}$, trọng số A_{ij}^t của cạnh này được tính theo công thức:

$$A_{ij}^t = \frac{|X_i \text{ quan hệ với } X_j \text{ trong luật thứ } t|}{|R|}$$

- Đối với các cạnh nối giữa đỉnh thuộc tính và đỉnh nhãn đầu ra, với mỗi cặp (X_i, l) , $1 \leq i \leq j \leq m$, $l = \overline{1, C}$, trong luật thứ t R_t , $t = \overline{1, k}$, trọng số B_{ij}^t của cạnh này được tính theo công thức:

$$B_{il}^t = \left(\sum A_{ij}^t \right) \times \frac{|X_i \text{ quan hệ với nhãn } l \text{ trong luật thứ } t|}{|R|}$$

Kết quả của hai bộ trọng số được lưu trữ trong một ma trận kề, đại diện cho đồ thị tri thức mờ đã được xây dựng.

4.4. Suy luận xấp xỉ.

Suy luận xấp xỉ được định nghĩa như một công cụ để suy luận từ các mệnh đề có ý nghĩa không cụ thể rõ ràng thông qua logic mờ. Thông thường phương pháp suy luận xấp xỉ có độ chính xác của kết quả không cao bằng các kỹ thuật suy luận thông thường đối với các dữ liệu rõ, tuy nhiên ưu thế của suy luận xấp xỉ là nó có thể thực hiện lập luận đối với các biến ngôn ngữ, hay ngôn ngữ tự nhiên là dữ liệu có nghĩa nhập nhằng, không rõ ràng

5. Mô hình và thuật toán

5.1. Phát biểu bài toán

Đầu vào: Để xây dựng đồ thị tri thức mờ sử dụng trong bài toán này, ta cần tập các mẫu bệnh nhân đã được chẩn đoán bởi các bác sĩ và chuyên gia dựa trên các thuộc tính được đặt ra. Tập dữ liệu mẫu này thông qua quá trình tiền xử lý (mờ hóa) và lưu vào hệ cơ sở luật mờ như bảng 1, hệ cơ sở luật mờ này bao gồm n luật R_1, R_2, \dots, R_n đại diện cho các mẫu bệnh nhân, m thuộc tính S_1, S_2, \dots, S_m đại diện cho các triệu chứng của bệnh, và C nhãn đầu ra $1, 2, 3, \dots, C$ đại diện cho kết luận chẩn đoán của bác sĩ

Ngoài ra, có thêm một bệnh nhân mới nằm ngoài hệ luật trên, được biểu diễn như sau

IF S_1 là “Low” và S_2 là “Low” và S_3 là “High” và S_4 là “Very high” và ... và S_{m-1} là “High” và S_m là “Low” THEN Kết luận = ?

	S_1	S_2	...	S_{m-1}	S_m	Kết luận
R_1	High	High	...	Very high	High	1
R_2	Medium	Medium	...	Medium	Low	2
...

R_{n-1}	Medium	Medium	...	Medium	Medium	2
R_n	Low	Medium	...	Low	Low	3
Triệu chứng của bệnh nhân	High Medium Low	High Medium	...	Very high High Medium Low	High Medium High	0,1,..., C

6. Bảng 1: Hệ cơ sở luật mờ

Đầu ra: Kết quả của hệ thống chẩn đoán đầu ra của bệnh nhân mới dựa trên cơ sở luật mờ đã có.

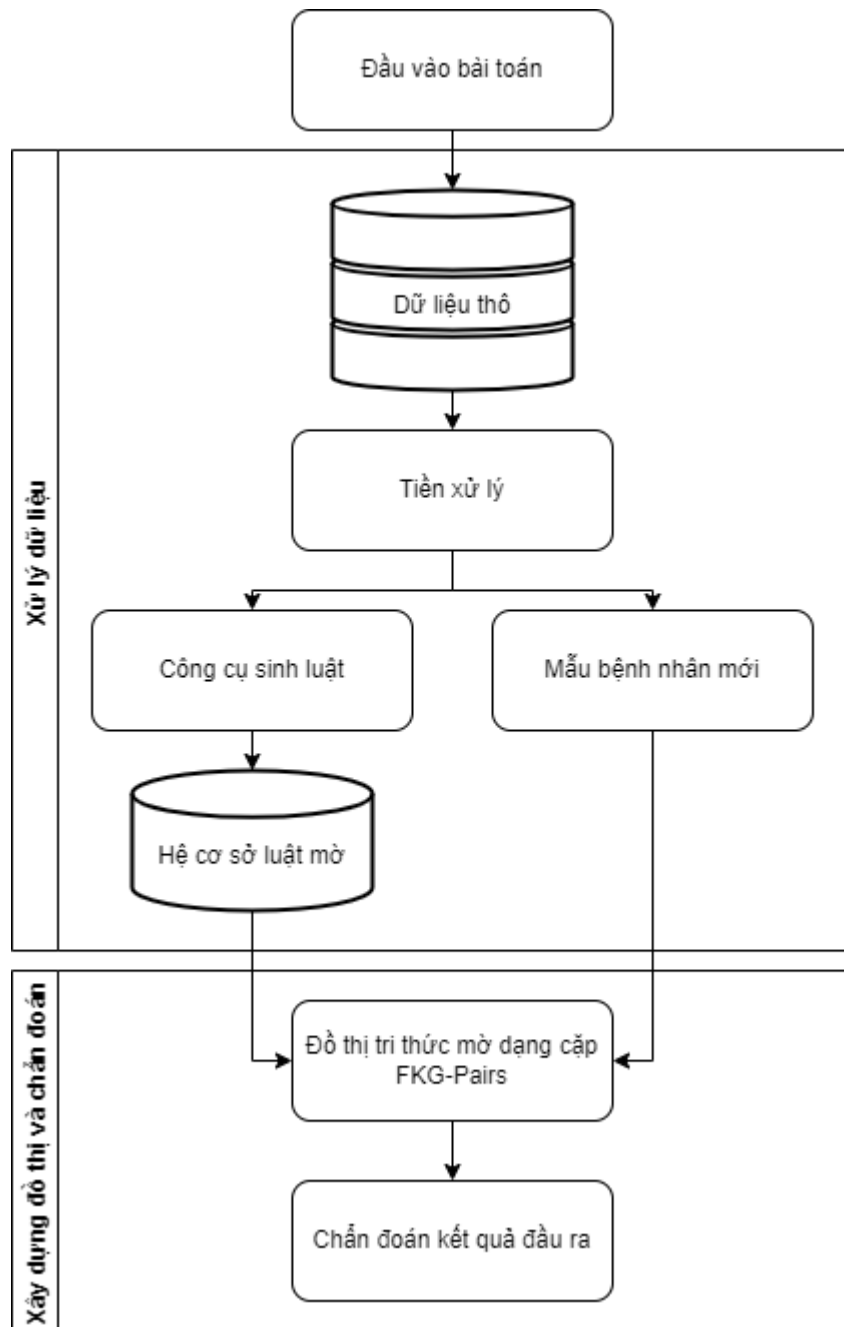
5.2. Mô hình đề xuất

5.2.1. Mô hình bài toán.

Hệ thống gồm 2 giai đoạn cần xử lý:

- Giai đoạn xử lý dữ liệu: Dữ liệu thu thập được từ các bác sĩ và chuyên gia là dữ liệu thô, dữ liệu này đi vào pha xử lý dữ liệu, bước tiền xử lý bao gồm công đoạn phân chia thang đo, thiết kế thuật ngữ mức độ tương ứng, sau đó kết hợp với công cụ sinh luật để đưa ra hệ cơ sở luật mờ của bài toán, làm cơ sở cho giai đoạn xây dựng đồ thị tri thức mờ và chẩn đoán
- Giai đoạn xây dựng đồ thị và chẩn đoán : dựa trên hệ cơ sở luật mờ đã được hình thành ở bước trên, tiến hành xây dựng đồ thị tri thức mờ FKG bằng cách tính toán các bộ trọng số cạnh và lưu thành ma trận kề, từ đó sử dụng dữ liệu của đồ thị tri thức mờ FKG đó để chẩn đoán ca bệnh mới

Mô hình cho hai giai đoạn trên của bài toán đã được mô tả trong [8], được vẽ lại dưới đây.



Hình 1: Mô hình bài toán chẩn đoán bệnh tim bằng FKG-Pairs

5.2.2. Chức năng hệ thống

Hệ thống cần có các chức năng thực hiện được 2 giai đoạn kể trên

- Đối với giai đoạn chuẩn bị dữ liệu : dữ liệu thô được thu thập bởi người thiết kế hệ thống làm đầu vào của bài toán cần được phân tích, loại bỏ nhiễu, lỗi, và đưa ra bộ dữ liệu đầu vào hoàn chỉnh. Bộ dữ liệu này cần được làm mờ dựa trên thang đo và thuật ngữ mức độ được thống nhất

với các chuyên gia y tế trong lĩnh vực trên, thông qua công cụ sinh luật để tổng hợp thành hệ cơ sở luật mờ sử dụng cho các chức năng sau đó của hệ thống

- Đối với giai đoạn biểu diễn và chẩn đoán : Sau khi có được hệ cơ sở luật mờ, hệ thống sẽ phải xây dựng được đồ thị tri thức mờ FKG và lưu trữ đồ thị này sử dụng cho bước chẩn đoán. Tại bước chẩn đoán, hệ thống sẽ nhận đầu vào, kết hợp với đồ thị tri thức mờ FKG để tiến hành chẩn đoán và đưa ra kết luận.

5.2.3. Phương pháp thực hiện

Theo như mô tả trong phần chức năng hệ thống, bài toán sẽ được cài đặt theo các bước sau:

- Bước 1: Xác thực và xử lý dữ liệu
Sau khi đã xây dựng được thang đo, tiến hành xây dựng các hàm và thủ tục để điều khiển xử lý tự động dữ liệu với các điều kiện cho từng thuộc tính dựa trên thang đo, kết hợp với công cụ sinh luật, cho đầu ra là hệ cơ sở luật mờ như trong bảng 1.
- Bước 2: Tính các bộ trọng số \tilde{A}, \tilde{B}
Bộ trọng số \tilde{A} của đồ thị tri thức mờ dạng cặp là trọng số của cạnh nối giữa các nhãn ngôn ngữ của các thuộc tính trong các luật $t (R_t)$. Các trọng số này được tính bởi công thức sau:

$$\tilde{A}_{ij...k}^t = \frac{|S_i \rightarrow S_j \rightarrow \dots \rightarrow S_{k+1} \text{ trong luật thứ } t|}{|R|} \quad (1)$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1$.

Bộ trọng số \tilde{B} của đồ thị tri thức mờ dạng cặp là các trọng số của các cạnh nối giữa các nhãn của các cặp thuộc tính với nhãn đầu ra trong luật $t (R_t)$. Các trọng số này được tính bởi công thức sau :

$$\tilde{B}_{ij...kl}^t = \left(\sum \tilde{A}_{ij...k+1}^t \right) \times \min \left(\frac{|S_i \rightarrow l \text{ trong luật } t|}{|R|}, \frac{|S_j \rightarrow l \text{ trong luật } t|}{|R|}, \dots, \frac{|S_k \rightarrow l \text{ trong luật } t|}{|R|} \right) \quad (2)$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$.

- Bước 3: Lưu trữ đồ thị tri thức mờ dạng cặp
Các bộ trọng số sau khi tính toán cần được lưu trữ dưới dạng các ma trận kề để tiện cho việc sử dụng cho tính toán các bước tiếp theo, trong

thiết kế này, ma trận kề của các bộ trọng số được lưu trong 2 sheet của một file excel thuộc tệp hệ thống.

- Bước 4: Áp dụng quá trình suy luận xấp xỉ và đưa ra kết quả chẩn đoán bệnh.

Trong [4], tác giả Lương Thị Hồng Lan đã sử dụng thuật toán FISA để suy luận xấp xỉ đầu ra của bài toán xác định đầu ra của luật mới dựa trên đồ thị tri thức mờ, sau đó, tác giả Cù Kim Long đã cải tiến thuật toán trên trong bài báo [8] của mình cho phù hợp với việc xác định đầu ra của luật mới dựa trên đồ thị tri thức mờ dạng cặp, thuật toán được mô tả qua các bước dưới đây:

Trước tiên, cần tính toán tổng trọng số của các cạnh (\tilde{C}) từ các siêu đỉnh đến nhãn đầu ra bằng cách sử dụng công thức sau:

$$\tilde{C}_{ij\dots kl} = \sum_t \tilde{B}_{ij\dots kl}^t \quad (3)$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$.

Sau đó, áp dụng các phép toán Max-Min để tính các giá trị (\tilde{D}), các giá trị này giúp ước tính gần đúng các cặp thuộc tính mới trong tập Testing với các cặp thuộc tính tương ứng trong đồ thị tri thức mờ dạng cặp, nhằm tìm ra mức độ ảnh hưởng đến các nhãn đầu ra. Giá trị của (\tilde{D}) được tính dựa trên các toán tử Max-Min theo công thức sau:

$$\tilde{D}_l = \text{Max}_{1 \leq i \leq j < \dots < k} (\tilde{C}_{ij\dots kl}) + \text{Min}_{1 \leq i \leq j < \dots < k} (\tilde{C}_{ij\dots kl}) \quad (4)$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$

Cuối cùng, kết quả chẩn đoán nhãn đầu ra của luật mới được kết luận bằng cách sử dụng phép toán Max như sau:

$$\text{Label} = p \text{ If } \tilde{D}_p = \text{Max}_{l=\overline{1, C}} (\tilde{D}_l) \quad (5)$$

5.3. Thuật toán cài đặt FKG-Pairs

Thuật toán cài đặt FKG dạng cặp

- 1 **Dữ liệu vào:** Tập dữ liệu kiểm thử, m: Số thuộc tính của mỗi luật, n: Số mẫu của tập dữ liệu, C: Số nhãn của mỗi thuộc tính.
- 2 **Dữ liệu ra:** Nhãn của mẫu mới.
- 3 **Begin**
- 4 Nhập các giá trị ;
- 5 Nhận tập dữ liệu kiểm thử;
- 6 Tiến hành mờ hóa tập dữ liệu kiểm thử;
- 7 **for** $i = 1$ **to** m **do**

```

8      for  $t = 1$  to  $n$  do
9          for  $l = 1$  to  $C$  do
10             while  $1 \leq i \leq j < \dots \leq k$  do
11                 Tính  $\tilde{C}_{ij\dots kl}^t = \sum_t \tilde{B}_{ij\dots kl}^t$ 
12                 Tính  $\tilde{D}_l = \text{Max}_{1 \leq i \leq j < \dots \leq k}(\tilde{C}_{ij\dots kl}) + \text{Min}_{1 \leq i \leq j < \dots \leq k}(\tilde{C}_{ij\dots kl})$ 
13             end
14             Xác định nhãn của mẫu t:  $\text{Label} = p$  If  $\tilde{D}_p = \text{Max}_{l=1, \dots, C}(\tilde{D}_l)$ 
15         end
16         Nhận nhãn của mẫu t và lặp lại các bước từ 5 đến 12 để tìm nhãn của
        các mẫu khác cho đến khi kết thúc
17     end
18 end
19 end

```

Thuật toán cài đặt bài toán FKG dạng cặp

7. Ví dụ thực tế về FKG-Pairs-1 và FKG-Pairs-2 chuẩn đoán bệnh tim

7.1. Chuẩn đoán bệnh tim bệnh nhân dựa trên mô hình M-CFIS-FKG sử dụng FKG-Pairs-1

Input: Giả sử đầu vào của bài toán là một danh sách gồm 6 bệnh nhân $\{R_1, R_2, R_3, R_4, R_5, R_6\}$, mỗi bệnh nhân có các kết quả xét nghiệm được thể hiện qua các thuộc tính $\{S_1, S_2, S_3, S_4, S_5, S_6\}$. Những trường hợp bệnh nhân nói trên đã được kiểm tra và chẩn đoán dựa trên các kết quả xét nghiệm bởi bác sĩ, các kết luận chẩn đoán “Bình thường”, “Tim ” và “Tim nặng” được thể hiện tương ứng với các nhãn 0, 1, 2. Sau khi qua giai đoạn “Xử lý dữ liệu” thu được một hệ cơ sở luật mờ như trong Bảng 1.

	S_1	S_2	S_3	S_4	S_5	S_6	Kết luận
R_1	High	Medium	High	Medium	High	High	2
R_2	High	Medium	Medium	Medium	High	High	2
R_3	Medium	Medium	Medium	Medium	Medium	Medium	0
R_4	Medium	High	Medium	High	Medium	High	1
R_5	High	High	Medium	High	Low	High	1
R_6	High	High	High	Medium	High	Medium	2

Bảng 1: Hệ cơ sở luật mờ giả định kết quả khám bệnh của sáu bệnh nhân đã được kết luận chẩn đoán bởi bác sĩ

Bên cạnh đó, Input có thêm một trường hợp bệnh nhân mới được biểu diễn như sau:

IF $S_1 = \text{"High"}$, $S_2 = \text{"Medium"}$, $S_3 = \text{"Medium"}$, $S_4 = \text{"Medium"}$, $S_5 = \text{"High"}$, $S_6 = \text{"Medium"}$ **THEN** Kết luận = ?

Output: Đưa ra kết luận chẩn đoán cho bệnh nhân trên, dựa trên hệ cơ sở luật mờ đã cho bởi Input.

Các bước thực hiện cho bài toán trên được thực hiện lần lượt như sau:

- **Bước 1:** Tính toán các bộ trọng số \tilde{A} , \tilde{B}

Bộ trọng số \tilde{A} gồm cạnh nối giữa các nhãn ngôn ngữ của các thuộc tính của bệnh nhân, được tính toán theo công thức (1)

Ví dụ, trong trường hợp $\{1\}$, trọng số \tilde{A} sẽ được tính toán như sau:

$$\begin{aligned}\tilde{A}_{12}^1 &= \frac{|\text{High} \rightarrow \text{Medium}|}{R} = \frac{1}{3} \\ \tilde{A}_{13}^1 &= \frac{|\text{High} \rightarrow \text{High}|}{R} = \frac{1}{3} \\ \tilde{A}_{14}^1 &= \frac{|\text{High} \rightarrow \text{Medium}|}{R} = \frac{1}{2}\end{aligned}$$

Bộ trọng số \tilde{B} của cạnh nối giữa nhãn của các cặp thuộc tính với các nhãn kết quả đầu ra được tính theo công thức:

$$\tilde{B}_{il}^t = \left(\sum \tilde{A}_{i+1}^t \right) \times \min \left(\frac{|S_i \rightarrow l \text{ trong luật } t|}{|R|}, \frac{|S_j \rightarrow l \text{ trong luật } t|}{|R|}, \dots, \frac{|S_k \rightarrow l \text{ trong luật } t|}{|R|} \right)$$

$$\begin{aligned}\tilde{B}_{1l}^1 &= (\tilde{A}_{12}^1 + \tilde{A}_{13}^1 + \tilde{A}_{14}^1 + \tilde{A}_{15}^1 + \tilde{A}_{16}^1 + \tilde{A}_{23}^1 + \tilde{A}_{24}^1 + \tilde{A}_{25}^1 + \tilde{A}_{26}^1 + \\ &\tilde{A}_{34}^1 + \tilde{A}_{35}^1 + \tilde{A}_{36}^1 + \tilde{A}_{45}^1 + \tilde{A}_{46}^1 + \tilde{A}_{56}^1) \times \frac{|\text{High} \rightarrow 2|}{R} = \frac{11}{4}\end{aligned}$$

Kết quả tính toán của toàn bộ ma trận trọng số được thể hiện qua Bảng 2 và Bảng 3.

	R_1	R_2	R_3	R_4	R_5	R_6
\tilde{A}_{12}^t	1/3	1/3	1/6	1/6	1/3	1/3
\tilde{A}_{13}^t	1/3	1/3	1/3	1/3	1/3	1/3

\tilde{A}_{14}^t	1/2	1/2	1/6	1/6	1/6	1/2
\tilde{A}_{15}^t	1/2	1/2	1/3	1/3	1/6	1/2
\tilde{A}_{16}^t	1/2	1/2	1/6	1/6	1/2	1/6
\tilde{A}_{23}^t	1/6	1/3	1/3	1/3	1/3	1/6
\tilde{A}_{24}^t	1/2	1/2	1/2	1/3	1/3	1/6
\tilde{A}_{25}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{26}^t	1/3	1/3	1/6	1/3	1/3	1/6
\tilde{A}_{34}^t	1/3	1/3	1/3	1/3	1/3	1/3
\tilde{A}_{35}^t	1/3	1/6	1/3	1/3	1/6	1/3
\tilde{A}_{36}^t	1/6	1/2	1/6	1/2	1/2	1/6
\tilde{A}_{45}^t	1/2	1/2	1/6	1/6	1/6	1/2
\tilde{A}_{46}^t	1/3	1/3	1/3	1/3	1/3	1/3
\tilde{A}_{56}^t	1/3	1/3	1/6	1/6	1/6	1/6

Bảng 2: Kết quả tính ma trận trọng số \tilde{A}

	R_1	R_2	R_3	R_4	R_5	R_6
\tilde{B}_{1l}^t	11/4	35/12	23/36	25/36	13/18	13/6
\tilde{B}_{2l}^t	11/6	35/18	23/36	25/18	13/9	13/18
\tilde{B}_{3l}^t	11/6	35/36	23/36	25/18	13/9	13/9
\tilde{B}_{4l}^t	11/4	35/12	23/36	25/18	13/9	13/6
\tilde{B}_{5l}^t	11/4	35/12	23/36	25/36	13/18	13/6
\tilde{B}_{6l}^t	11/6	35/18	23/36	25/18	13/9	13/18

Bảng 3: Kết quả tính ma trận trọng số \tilde{B}

- ⇒ Các bộ trọng số sẽ được kết hợp với hệ cơ sở luật mờ để biểu diễn đồ thị tri thức mờ.
- **Bước 2:** Áp dụng phương pháp suy luận xấp xỉ để đưa ra kết quả chẩn đoán bệnh.

Sau khi có được đồ thị tri thức mờ (được biểu diễn dựa trên các bộ trọng số và hệ cơ sở luật mờ ở Bước 1), tiếp tục tiến hành chẩn đoán kết quả của bệnh nhân mới bằng phương pháp suy luận xấp xỉ. Đầu tiên, cần tính toán tổng trọng số của các cạnh (\tilde{C}) từ các siêu đỉnh đến nhãn đầu ra dựa theo công thức:

$$\tilde{C}_{ij\dots kl} = \sum_t \tilde{B}_{ij\dots kl}^t$$

$$\tilde{C}_{11} = \sum_t \tilde{B}_{11}^t = \tilde{B}_{11}(\text{Rule 4}) + \tilde{B}_{11}(\text{Rule 5}) = \frac{25}{36} + \frac{13}{18} = \frac{17}{12}$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$.

Kết quả tính toán cụ thể được tổng hợp lại trong Bảng 4.

	<i>Label 0</i>	<i>Label 1</i>	<i>Label 2</i>
\tilde{C}_{1l}	23/36	17/12	47/2
\tilde{C}_{2l}	23/36	17/3	149/18
\tilde{C}_{3l}	23/36	17/3	271/36
\tilde{C}_{4l}	23/36	17/3	47/2
\tilde{C}_{5l}	23/36	17/12	47/2
\tilde{C}_{6l}	23/36	17/3	149/18

Bảng 4: Kết quả tính ma trận trọng số \tilde{C}

Dựa vào bảng trên, tiếp tục tính các giá trị (\tilde{D}) theo công thức:

$$\tilde{D}_l = \text{Max}_{1 \leq i \leq j < \dots < k} (\tilde{C}_{ij\dots kl}) + \text{Min}_{1 \leq i \leq j < \dots < k} (\tilde{C}_{ij\dots kl})$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$

Input có thêm một trường hợp bệnh nhân mới được biểu diễn như sau:

IF $S_1 = \text{"High"}$, $S_2 = \text{"Medium"}$, $S_3 = \text{"Medium"}$, $S_4 = \text{"Medium"}$, $S_5 = \text{"High"}$, $S_6 = \text{"Medium"}$ **THEN** Kết luận = ?

Label 0:

$$\tilde{C}_{\text{High}1,0} = 0, \tilde{C}_{\text{Medium}2,0} = 23/36, \tilde{C}_{\text{Medium}3,0} = 23/36,$$

$$\tilde{C}_{\text{Medium}4,0} = 23/36, \tilde{C}_{\text{High}5,0} = 0, \tilde{C}_{\text{Medium}6,0} = 23/36$$

$$\begin{aligned}
\tilde{D}_0 &= \max(\tilde{C}_{High1,0}, \tilde{C}_{Medium2,0}, \tilde{C}_{Medium3,0}, \tilde{C}_{Medium4,0}, \tilde{C}_{High5,0}, \tilde{C}_{Medium6,0}) \\
&+ \min(\tilde{C}_{High1,0}, \tilde{C}_{Medium2,0}, \tilde{C}_{Medium3,0}, \tilde{C}_{Medium4,0}, \tilde{C}_{High5,0}, \tilde{C}_{Medium6,0}) \\
&= 23/36
\end{aligned}$$

Label 1:

$$\begin{aligned}
\tilde{C}_{High1,1} &= 13/18, \tilde{C}_{Medium2,1} = 0, \tilde{C}_{Medium3,1} = 17/6, \\
\tilde{C}_{Medium4,1} &= 0, \tilde{C}_{High5,1} = 0, \tilde{C}_{Medium6,1} = 0
\end{aligned}$$

$$\begin{aligned}
\tilde{D}_1 &= \max(\tilde{C}_{High1,1}, \tilde{C}_{Medium2,1}, \tilde{C}_{Medium3,1}, \tilde{C}_{Medium4,1}, \tilde{C}_{High5,1}, \tilde{C}_{Medium6,1}) \\
&+ \min(\tilde{C}_{High1,1}, \tilde{C}_{Medium2,1}, \tilde{C}_{Medium3,1}, \tilde{C}_{Medium4,1}, \tilde{C}_{High5,1}, \tilde{C}_{Medium6,1}) \\
&= 17/6
\end{aligned}$$

Label 2:

$$\begin{aligned}
\tilde{C}_{High1,2} &= 47/6, \tilde{C}_{Medium2,2} = 34/9, \tilde{C}_{Medium3,2} = 35/36, \\
\tilde{C}_{Medium4,2} &= 47/6, \tilde{C}_{High5,2} = 47/6, \tilde{C}_{Medium6,2} = 13/18
\end{aligned}$$

$$\begin{aligned}
\tilde{D}_2 &= \max(\tilde{C}_{High1,2}, \tilde{C}_{Medium2,2}, \tilde{C}_{Medium3,2}, \tilde{C}_{Medium4,2}, \tilde{C}_{High5,2}, \tilde{C}_{Medium6,2}) \\
&+ \min(\tilde{C}_{High1,2}, \tilde{C}_{Medium2,2}, \tilde{C}_{Medium3,2}, \tilde{C}_{Medium4,2}, \tilde{C}_{High5,2}, \tilde{C}_{Medium6,2}) \\
&= 47/6 + 13/18 \\
&= 77/9
\end{aligned}$$

Từ đó ta có:

$$D_0 = \frac{23}{36}$$

$$D_1 = \frac{17}{6}$$

$$D_2 = \frac{77}{9}$$

Theo công thức:

$$Label = p \text{ If } \tilde{D}_p = \text{Max}_{l=\overline{1,C}}(\tilde{D}_l)$$

$$\text{Ta có: } \text{Max}_{l=\overline{0,2}}(\tilde{D}_l) = D_2 = \frac{77}{9}$$

$$\text{Do đó } Label = 2$$

Sử dụng phép toán Max ta thu được nhãn đầu ra của bệnh nhân mới là $2 \left(D_2 = \frac{77}{9} \right)$, từ đó có thể kết luận, bệnh nhân mới có dấu hiệu mắc bệnh Tim nặng.

7.2. Chuẩn đoán bệnh tim bệnh nhân dựa trên mô hình M-CFIS-FKG sử dụng FKG-Pairs-2

Input: Giả sử đầu vào của bài toán là một danh sách gồm 6 bệnh nhân $\{R_1, R_2, R_3, R_4, R_5, R_6\}$, mỗi bệnh nhân có các kết quả xét nghiệm được thể hiện qua các thuộc tính $\{S_1, S_2, S_3, S_4, S_5, S_6\}$. Những trường hợp bệnh nhân nói trên đã được kiểm tra và chẩn đoán dựa trên các kết quả xét nghiệm bởi bác sĩ, các kết luận chẩn đoán “Bình thường”, “Tim” và “Tim nặng” được thể hiện tương ứng với các nhãn 0, 1, 2. Sau khi qua giai đoạn “Xử lý dữ liệu” thu được một hệ cơ sở luật mờ như trong Bảng 1.

	S_1	S_2	S_3	S_4	S_5	S_6	Kết luận
R_1	High	Medium	High	Medium	High	High	2
R_2	High	Medium	Medium	Medium	High	High	2
R_3	Medium	Medium	Medium	Medium	Medium	Medium	0
R_4	Medium	High	Medium	High	Medium	High	1
R_5	High	High	Medium	High	Low	High	1
R_6	High	High	High	Medium	High	Medium	2

Bảng 1: Hệ cơ sở luật mờ giả định kết quả khám bệnh của sáu bệnh nhân đã được kết luận chẩn đoán bởi bác sĩ

Bên cạnh đó, Input có thêm một trường hợp bệnh nhân mới được biểu diễn như sau:

IF $S_1 = \text{“High”}$, $S_2 = \text{“Medium”}$, $S_3 = \text{“Medium”}$, $S_4 = \text{“Medium”}$, $S_5 = \text{“High”}$, $S_6 = \text{“Medium”}$ **THEN** Kết luận = ?

Output: Đưa ra kết luận chẩn đoán cho bệnh nhân trên, dựa trên hệ cơ sở luật mờ đã cho bởi Input.

Các bước thực hiện cho bài toán trên được thực hiện lần lượt như sau:

- **Bước 1:** Tính toán các bộ trọng số \tilde{A} , \tilde{B}

Bộ trọng số \tilde{A} gồm cạnh nối giữa các nhãn ngôn ngữ của các thuộc tính của bệnh nhân, được tính toán theo công thức (1)

Ví dụ, trong trường hợp $\{1\}$, trọng số \tilde{A} sẽ được tính toán như sau:

$$\begin{aligned}\tilde{A}_{123}^1 &= \frac{|\text{High} \rightarrow \text{Medium} \rightarrow \text{High}|}{R} = \frac{1}{6} \\ \tilde{A}_{124}^1 &= \frac{|\text{High} \rightarrow \text{High} \rightarrow \text{Medium}|}{R} = \frac{1}{3} \\ \tilde{A}_{125}^1 &= \frac{|\text{High} \rightarrow \text{Medium} \rightarrow \text{High}|}{R} = \frac{1}{2}\end{aligned}$$

Bộ trọng số \tilde{B} của cạnh nối giữa nhãn của các cặp thuộc tính với các nhãn kết quả đầu ra được tính theo công thức:

$$\tilde{B}_{il}^t = \left(\sum \tilde{A}_{ij}^t \right) \times \min \left(\frac{|S_i \rightarrow l \text{ trong luật } t|}{|R|}, \frac{|S_j \rightarrow l \text{ trong luật } t|}{|R|}, \dots, \frac{|S_k \rightarrow l \text{ trong luật } t|}{|R|} \right)$$

$$\begin{aligned}\tilde{B}_{12l}^1 &= (\tilde{A}_{123}^1 + \tilde{A}_{124}^1 + \tilde{A}_{125}^1 + \tilde{A}_{126}^1 + \tilde{A}_{134}^1 + \tilde{A}_{135}^1 + \tilde{A}_{136}^1 + \tilde{A}_{145}^1 + \\ &\tilde{A}_{146}^1 + \tilde{A}_{156}^1 + \tilde{A}_{234}^1 + \tilde{A}_{235}^1 + \tilde{A}_{236}^1 + \tilde{A}_{245}^1 + \tilde{A}_{246}^1 + \tilde{A}_{256}^1 + \\ &\tilde{A}_{345}^1 + \tilde{A}_{346}^1 + \tilde{A}_{356}^1 + \tilde{A}_{456}^1) \times \frac{|\text{High} \rightarrow \text{Medium} \rightarrow 2|}{R}\end{aligned}$$

$$= \frac{17}{3} \times \frac{1}{3}$$

$$= \frac{17}{9}$$

Kết quả tính toán của toàn bộ ma trận trọng số được thể hiện qua Bảng 2 và Bảng 3.

	R_1	R_2	R_3	R_4	R_5	R_6
\tilde{A}_{123}^t	1/6	1/6	1/6	1/6	1/6	1/6
\tilde{A}_{124}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{125}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{126}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{134}^t	1/3	1/6	1/6	1/6	1/6	1/3
\tilde{A}_{135}^t	1/3	1/6	1/3	1/3	1/6	1/3
\tilde{A}_{136}^t	1/6	1/3	1/6	1/6	1/3	1/6
\tilde{A}_{145}^t	1/2	1/2	1/6	1/6	1/6	1/2
\tilde{A}_{146}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{156}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{234}^t	1/6	1/3	1/3	1/3	1/3	1/6
\tilde{A}_{235}^t	1/6	1/6	1/6	1/6	1/6	1/6
\tilde{A}_{236}^t	1/6	1/6	1/6	1/3	1/3	1/6
\tilde{A}_{245}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{246}^t	1/3	1/3	1/6	1/3	1/3	1/6
\tilde{A}_{256}^t	1/3	1/3	1/6	1/6	1/6	1/6
\tilde{A}_{345}^t	1/3	1/6	1/6	1/6	1/6	1/3
\tilde{A}_{346}^t	1/6	1/6	1/6	1/3	1/3	1/6
\tilde{A}_{356}^t	1/6	1/6	1/6	1/6	1/6	1/6
\tilde{A}_{456}^t	1/3	1/3	1/6	1/6	1/6	1/6

Bảng 2: Kết quả tính ma trận trọng số \tilde{A}

	R_1	R_2	R_3	R_4	R_5	R_6
\tilde{B}_{12l}^t	17/9	11/6	11/18	25/36	25/36	25/36
\tilde{B}_{13l}^t	17/9	11/12	11/18	25/36	25/36	25/18

\tilde{B}_{14l}^t	17/6	11/4	11/18	25/36	25/36	25/12
\tilde{B}_{15l}^t	17/6	11/4	11/18	25/36	25/36	25/12
\tilde{B}_{16l}^t	17/9	11/6	11/18	25/36	25/36	25/36
\tilde{B}_{23l}^t	17/18	11/12	11/18	25/18	25/18	25/36
\tilde{B}_{24l}^t	17/9	11/6	11/18	25/18	25/18	25/36
\tilde{B}_{25l}^t	17/9	11/6	11/18	25/18	25/18	25/36
\tilde{B}_{26l}^t	17/9	11/6	11/18	25/18	25/18	25/36
\tilde{B}_{34l}^t	17/9	11/12	11/18	25/18	25/18	25/18
\tilde{B}_{35l}^t	17/9	11/6	11/18	25/36	25/36	25/18
\tilde{B}_{36l}^t	17/18	11/12	11/18	25/18	25/18	25/36
\tilde{B}_{45l}^t	17/6	11/4	11/18	25/36	25/36	25/12
\tilde{B}_{46l}^t	17/9	11/6	11/18	25/18	25/18	25/36
\tilde{B}_{56l}^t	17/9	11/6	11/18	25/36	25/36	25/36

Bảng 3: Kết quả tính ma trận trọng số \tilde{B}

⇒ Các bộ trọng số sẽ được kết hợp với hệ cơ sở luật mờ để biểu diễn đồ thị tri thức mờ.

- **Bước 2:** Áp dụng phương pháp suy luận xấp xỉ để đưa ra kết quả chẩn đoán bệnh.

Sau khi có được đồ thị tri thức mờ (được biểu diễn dựa trên các bộ trọng số và hệ cơ sở luật mờ ở Bước 1), tiếp tục tiến hành chẩn đoán kết quả của bệnh nhân mới bằng phương pháp suy luận xấp xỉ. Đầu tiên, cần tính toán tổng trọng số của các cạnh (\tilde{C}) từ các siêu đỉnh đến nhân đầu ra dựa theo công thức:

$$\tilde{C}_{ij...kl} = \sum_t \tilde{B}_{ij...kl}^t$$

$$\tilde{C}_{121} = \sum_t \tilde{B}_{121}^t = \tilde{B}_{121}^t(\text{Rule 4}) + \tilde{B}_{121}^t(\text{Rule 5}) = \frac{25}{36} + \frac{25}{36} = \frac{25}{18}$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$.

Kết quả tính toán cụ thể được tổng hợp lại trong Bảng 4.

	<i>Label 0</i>	<i>Label 1</i>	<i>Label 2</i>
\tilde{C}_{12l}	11/18	25/18	53/12
\tilde{C}_{13l}	11/18	25/18	79/9
\tilde{C}_{14l}	11/18	25/18	23/3
\tilde{C}_{15l}	11/18	25/18	23/3
\tilde{C}_{16l}	11/18	25/18	53/2
\tilde{C}_{23l}	11/18	25/9	23/9
\tilde{C}_{24l}	11/18	25/9	53/12
\tilde{C}_{25l}	11/18	25/9	53/12
\tilde{C}_{26l}	11/18	25/9	53/12
\tilde{C}_{34l}	11/18	25/9	151/36
\tilde{C}_{35l}	11/18	25/18	46/9
\tilde{C}_{36l}	11/18	25/9	23/9
\tilde{C}_{45l}	11/18	25/18	23/3
\tilde{C}_{46l}	11/18	25/9	53/12
\tilde{C}_{56l}	11/18	25/18	53/12

Bảng 4: Kết quả tính ma trận trọng số \tilde{C}

Dựa vào bảng trên, tiếp tục tính các giá trị (\tilde{D}) theo công thức:

$$\tilde{D}_l = \text{Max}_{1 \leq i \leq j < \dots \leq k} (\tilde{C}_{ij \dots kl}) + \text{Min}_{1 \leq i \leq j < \dots \leq k} (\tilde{C}_{ij \dots kl})$$

Trong đó $t = \overline{1, n}, 1 \leq i \leq j < \dots < k < m - 1, l = \overline{1, C}$

Input có thêm một trường hợp bệnh nhân mới được biểu diễn như sau:

IF $S_1 = \text{"High"}$, $S_2 = \text{"Medium"}$, $S_3 = \text{"Medium"}$, $S_4 = \text{"Medium"}$, $S_5 = \text{"High"}$, $S_6 = \text{"Medium"}$ **THEN** Kết luận = ?

Label 0:

$$\tilde{C}_{\text{High1} \rightarrow \text{Medium2},0} = 0,$$

$$\tilde{C}_{\text{High1} \rightarrow \text{Medium3},0} = 0,$$

$$\tilde{C}_{\text{High1} \rightarrow \text{Medium4},0} = 0,$$

$$\tilde{C}_{\text{High1} \rightarrow \text{High5},0} = 0,$$

$$\begin{aligned}
\tilde{C}_{High1 \rightarrow Medium6,0} &= 0, \\
\tilde{C}_{Medium2 \rightarrow Medium3,0} &= 11/18, \\
\tilde{C}_{Medium2 \rightarrow Medium4,0} &= 11/18, \\
\tilde{C}_{Medium2 \rightarrow High5,0} &= 0, \\
\tilde{C}_{Medium2 \rightarrow Medium6,0} &= 11/18, \\
\tilde{C}_{Medium3 \rightarrow Medium4,0} &= 11/18, \\
\tilde{C}_{Medium3 \rightarrow High5,0} &= 0, \\
\tilde{C}_{Medium3 \rightarrow Medium6,0} &= 11/18, \\
\tilde{C}_{Medium4 \rightarrow High5,0} &= 0, \\
\tilde{C}_{Medium4 \rightarrow Medium6,0} &= 11/18, \\
\tilde{C}_{High5 \rightarrow Medium6,0} &= 0
\end{aligned}$$

$$\begin{aligned}
&\tilde{D}_0 = \\
&\max(\\
&\quad \tilde{C}_{High1 \rightarrow Medium2,0}, \tilde{C}_{High1 \rightarrow Medium3,0}, \tilde{C}_{High1 \rightarrow Medium4,0}, \tilde{C}_{High1 \rightarrow High5,0}, \tilde{C}_{High1 \rightarrow Medium6,0}, \\
&\quad \tilde{C}_{Medium2 \rightarrow Medium3,0}, \tilde{C}_{Medium2 \rightarrow Medium4,0}, \tilde{C}_{Medium2 \rightarrow High5,0}, \\
&\quad \tilde{C}_{Medium2 \rightarrow Medium6,0}, \tilde{C}_{Medium3 \rightarrow Medium4,0}, \tilde{C}_{Medium3 \rightarrow High5,0}, \tilde{C}_{Medium3 \rightarrow Medium6,0}, \\
&\quad \tilde{C}_{Medium4 \rightarrow High5,0}, \tilde{C}_{Medium4 \rightarrow Medium6,0}, \tilde{C}_{High5 \rightarrow Medium6,0}) + \\
&\min(\\
&\quad \tilde{C}_{High1 \rightarrow Medium2,0}, \tilde{C}_{High1 \rightarrow Medium3,0}, \tilde{C}_{High1 \rightarrow Medium4,0}, \tilde{C}_{High1 \rightarrow High5,0}, \tilde{C}_{High1 \rightarrow Medium6,0}, \\
&\quad \tilde{C}_{Medium2 \rightarrow Medium3,0}, \tilde{C}_{Medium2 \rightarrow Medium4,0}, \tilde{C}_{Medium2 \rightarrow High5,0}, \\
&\quad \tilde{C}_{Medium2 \rightarrow Medium6,0}, \tilde{C}_{Medium3 \rightarrow Medium4,0}, \tilde{C}_{Medium3 \rightarrow High5,0}, \tilde{C}_{Medium3 \rightarrow Medium6,0}, \\
&\quad \tilde{C}_{Medium4 \rightarrow High5,0}, \tilde{C}_{Medium4 \rightarrow Medium6,0}, \tilde{C}_{High5 \rightarrow Medium6,0}) \\
&= 11/18
\end{aligned}$$

Label 1:

$$\begin{aligned}
\tilde{C}_{High1 \rightarrow Medium2,1} &= 0, \\
\tilde{C}_{High1 \rightarrow Medium3,1} &= 25/36, \\
\tilde{C}_{High1 \rightarrow Medium4,1} &= 0, \\
\tilde{C}_{High1 \rightarrow High5,1} &= 0,
\end{aligned}$$

$$\begin{aligned}
\tilde{C}_{High1 \rightarrow Medium6,1} &= 0, \\
\tilde{C}_{Medium2 \rightarrow Medium3,1} &= 0, \\
\tilde{C}_{Medium2 \rightarrow Medium4,1} &= 0, \\
\tilde{C}_{Medium2 \rightarrow High5,1} &= 0, \\
\tilde{C}_{Medium2 \rightarrow Medium6,1} &= 0, \\
\tilde{C}_{Medium3 \rightarrow Medium4,1} &= 0, \\
\tilde{C}_{Medium3 \rightarrow High5,1} &= 0, \\
\tilde{C}_{Medium3 \rightarrow Medium6,1} &= 0, \\
\tilde{C}_{Medium4 \rightarrow High5,1} &= 0, \\
\tilde{C}_{Medium4 \rightarrow Medium6,1} &= 0, \\
\tilde{C}_{High5 \rightarrow Medium6,1} &= 0
\end{aligned}$$

$$\begin{aligned}
\tilde{D}_1 &= \\
&\max(\\
&\tilde{C}_{High1 \rightarrow Medium2,1}, \tilde{C}_{High1 \rightarrow Medium3,1}, \tilde{C}_{High1 \rightarrow Medium4,1}, \tilde{C}_{High1 \rightarrow High5,1}, \tilde{C}_{High1 \rightarrow Medium6,1}, \\
&\tilde{C}_{Medium2 \rightarrow Medium3,1}, \tilde{C}_{Medium2 \rightarrow Medium4,1}, \tilde{C}_{Medium2 \rightarrow High5,1}, \\
&\tilde{C}_{Medium2 \rightarrow Medium6,1}, \tilde{C}_{Medium3 \rightarrow Medium4,1}, \tilde{C}_{Medium3 \rightarrow High5,1}, \tilde{C}_{Medium3 \rightarrow Medium6,1}, \\
&\tilde{C}_{Medium4 \rightarrow High5,1}, \tilde{C}_{Medium4 \rightarrow Medium6,1}, \tilde{C}_{High5 \rightarrow Medium6,1}) + \\
&\min(\\
&\tilde{C}_{High1 \rightarrow Medium2,1}, \tilde{C}_{High1 \rightarrow Medium3,1}, \tilde{C}_{High1 \rightarrow Medium4,1}, \tilde{C}_{High1 \rightarrow High5,1}, \tilde{C}_{High1 \rightarrow Medium6,1}, \\
&\tilde{C}_{Medium2 \rightarrow Medium3,1}, \tilde{C}_{Medium2 \rightarrow Medium4,1}, \tilde{C}_{Medium2 \rightarrow High5,1}, \\
&\tilde{C}_{Medium2 \rightarrow Medium6,1}, \tilde{C}_{Medium3 \rightarrow Medium4,1}, \tilde{C}_{Medium3 \rightarrow High5,1}, \tilde{C}_{Medium3 \rightarrow Medium6,1}, \\
&\tilde{C}_{Medium4 \rightarrow High5,1}, \tilde{C}_{Medium4 \rightarrow Medium6,1}, \tilde{C}_{High5 \rightarrow Medium6,1}) \\
&= 25/36
\end{aligned}$$

Label 2:

$$\begin{aligned}
\tilde{C}_{High1 \rightarrow Medium2,2} &= 67/18, \\
\tilde{C}_{High1 \rightarrow Medium3,2} &= 11/12, \\
\tilde{C}_{High1 \rightarrow Medium4,2} &= 23/3, \\
\tilde{C}_{High1 \rightarrow High5,2} &= 23/3, \\
\tilde{C}_{High1 \rightarrow Medium6,2} &= 25/36, \\
\tilde{C}_{Medium2 \rightarrow Medium3,2} &= 11/12,
\end{aligned}$$

$$\tilde{C}_{Medium2 \rightarrow Medium4,2} = 67/18,$$

$$\tilde{C}_{Medium2 \rightarrow High5,2} = 67/18,$$

$$\tilde{C}_{Medium2 \rightarrow Medium6,2} = 0,$$

$$\tilde{C}_{Medium3 \rightarrow Medium4,2} = 11/12,$$

$$\tilde{C}_{Medium3 \rightarrow High5,2} = 11/6,$$

$$\tilde{C}_{Medium3 \rightarrow Medium6,2} = 0,$$

$$\tilde{C}_{Medium4 \rightarrow High5,2} = 23/3,$$

$$\tilde{C}_{Medium4 \rightarrow Medium6,2} = 25/36,$$

$$\tilde{C}_{High5 \rightarrow Medium6,2} = 25/36$$

$$\tilde{D}_2 =$$

$$\max($$

$$\tilde{C}_{High1 \rightarrow Medium2,2}, \tilde{C}_{High1 \rightarrow Medium3,2}, \tilde{C}_{High1 \rightarrow Medium4,2}, \tilde{C}_{High1 \rightarrow High5,2}, \tilde{C}_{High1 \rightarrow Medium6,2},$$

$$\tilde{C}_{Medium2 \rightarrow Medium3,2}, \tilde{C}_{Medium2 \rightarrow Medium4,2}, \tilde{C}_{Medium2 \rightarrow High5,2},$$

$$\tilde{C}_{Medium2 \rightarrow Medium6,2}, \tilde{C}_{Medium3 \rightarrow Medium4,2}, \tilde{C}_{Medium3 \rightarrow High5,2}, \tilde{C}_{Medium3 \rightarrow Medium6,2},$$

$$\tilde{C}_{Medium4 \rightarrow High5,2}, \tilde{C}_{Medium4 \rightarrow Medium6,2}, \tilde{C}_{High5 \rightarrow Medium6,2}) +$$

$$\min($$

$$\tilde{C}_{High1 \rightarrow Medium2,2}, \tilde{C}_{High1 \rightarrow Medium3,2}, \tilde{C}_{High1 \rightarrow Medium4,2}, \tilde{C}_{High1 \rightarrow High5,2}, \tilde{C}_{High1 \rightarrow Medium6,2},$$

$$\tilde{C}_{Medium2 \rightarrow Medium3,2}, \tilde{C}_{Medium2 \rightarrow Medium4,2}, \tilde{C}_{Medium2 \rightarrow High5,2},$$

$$\tilde{C}_{Medium2 \rightarrow Medium6,2}, \tilde{C}_{Medium3 \rightarrow Medium4,2}, \tilde{C}_{Medium3 \rightarrow High5,2}, \tilde{C}_{Medium3 \rightarrow Medium6,2},$$

$$\tilde{C}_{Medium4 \rightarrow High5,2}, \tilde{C}_{Medium4 \rightarrow Medium6,2}, \tilde{C}_{High5 \rightarrow Medium6,2})$$

$$= 23/3$$

Từ đó ta có:

$$D_0 = \frac{11}{18}$$

$$D_1 = \frac{25}{36}$$

$$D_2 = \frac{23}{3}$$

Theo công thức:

$$Label = p \text{ If } \tilde{D}_p = \text{Max}_{l=\overline{1,C}}(\tilde{D}_l)$$

$$\text{Ta có: } \text{Max}_{l=\overline{0,2}}(\tilde{D}_l) = D_2 = \frac{23}{3}$$

$$\text{Do đó } Label = 2$$

Sử dụng phép toán Max ta thu được nhãn đầu ra của bệnh nhân mới là 2 ($D_2 = \frac{23}{3}$), từ đó có thể kết luận, bệnh nhân mới có dấu hiệu mắc bệnh Tim nặng.

8. Kết quả đánh giá 2 mô hình FKG-Pairs-1 và FKG-Pairs-2 bằng phương pháp toán học

Model	FKG-Pairs-1	FKG-Pairs-2
D_0	23/36	11/18
D_1	17/6	25/26
D_2	77/9	23/3

Dựa trên kết quả của hai cặp Pairs-1 và Pairs-2 và đánh giá chênh lệch giữa các giá trị D_0, D_1, D_2 , ta có thể đưa ra một số nhận xét như sau:

1. Hiệu suất của mô hình Pair 2: Pair 2 có độ chênh lệch giữa các giá trị D_0, D_1, D_2 nhỏ hơn so với Pair 1. Điều này gợi ý rằng mô hình Pair 2 có khả năng phân loại và dự đoán nguy cơ một cách rõ ràng và chính xác hơn so với Pair 1. Sự nhất quán trong việc dự đoán giữa các nhãn cũng tăng cường tính đáng tin cậy của mô hình.
2. Độ chính xác của mô hình Pair 1: Pair 1 có độ chênh lệch lớn hơn giữa các giá trị D_0, D_1, D_2 , cho thấy mô hình có thể gặp khó khăn trong việc phân loại rõ ràng giữa các nhóm nguy cơ. Sự chênh lệch lớn có thể gợi ý rằng mô hình cần được điều chỉnh hoặc cải thiện để tăng tính chính xác và đáng tin cậy của dự đoán.
3. Tính linh hoạt và tính ứng dụng: Cả hai mô hình đều sử dụng mô hình M-CFIS-FKG với FKG-Pairs-k để dự đoán nguy cơ mắc bệnh tim nặng hay nhẹ. Tuy nhiên, hiệu suất của chúng có thể phụ thuộc vào cách thức xây dựng mô hình và cách xử lý dữ liệu. Đối với ứng dụng cụ thể này, Pair 2 có vẻ linh hoạt hơn và hiệu quả hơn trong việc dự đoán nguy cơ.
 - ⇒ Cần tiếp tục nghiên cứu và cải thiện: Đánh giá mô hình chỉ là một phần của quá trình. Cần tiếp tục nghiên cứu và phát triển các phương pháp mới để cải thiện hiệu suất của mô hình trong việc dự đoán nguy cơ mắc bệnh tim. Sự tiếp tục theo dõi và đánh giá các mô

hình là quan trọng để đảm bảo tính chính xác và đáng tin cậy của dự đoán trong ứng dụng thực tế.

9. Dự kiến tiến độ công việc tiếp theo

- Tham khảo ý kiến các bác sĩ và tra cứu các ngưỡng giới hạn để xác định ngưỡng của các thuộc tính đặc trưng như chiều cao, cân nặng, huyết áp tâm thu, huyết áp tâm trương,... từ đó viết các luật tri thức mờ
⇒ Xây dựng bộ luật cơ sở tri thức mờ cho tập dữ liệu để xây bộ dữ liệu hoàn chỉnh cho huấn luyện mô hình chuẩn đoán bệnh tim.
- Tìm thêm các bộ dữ liệu chuẩn đoán bệnh tim khác nhau để phục vụ cho quá trình huấn luyện và kiểm thử mô hình chuẩn đoán bệnh tim.
- Tiếp tục phát triển ra hệ thống/ứng dụng chuẩn đoán bệnh hoàn chỉnh sử dụng Python.

10. Tài liệu tham khảo

- [1] Abdel-Basset M, Gamal A, Manogaran G, Son LH, Long HV (2019) A novel group decision making model based on neutrosophic sets for heart disease diagnosis. *Multimed Tools Appl* 79:9977–10002. <https://doi.org/10.1007/s11042-019-07742-7>
- [2] Hai V.Phạm, Cu Kim Long, Phan Hung Khanh and Ha Quoc Trung (2023) A Fuzzy Knowledge Graph Pairs-Based Application for Classification in Decision Making: Case Study of Preeclampsia Signs. <https://www.mdpi.com/2078-2489/14/2/104>
- [3] Alves MA et al (2021) Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Comput Biol Med* 132. <https://doi.org/10.1016/j.compbiomed.2021.104335>
- [4] Bai W, Ding J, Zhang C (2020) Dual hesitant fuzzy graphs with applications to multi-attribute decision making. *Int J Cogn Comput Eng* 1:18–26. <https://doi.org/10.1016/j.ijcce.2020.09.002>
- [5] Bakhshipour A et al (2020) Application of decision trees and fuzzy inference system for quality classification and modeling of black and green tea based on visual features. In: *Proc. Food Meas. Characterization*, pp 1–15
- [6] FKG-Group (2021). Datasets and source codes of this paper are available at the following: <https://github.com/CodePaper/FKG-Group>
- [7] Luong Thi Hong Lan, Tran Manh Tuan and others (2020) A New Complex Fuzzy Inference System With Fuzzy Knowledge Graph and Extensions in Decision Making. <https://ieeexplore.ieee.org/abstract/document/9184876>

[8] *Phạm Minh Chuẩn, Trần Mạnh Tuấn, Cù Kim Long, Nguyễn Hồng Tân (2023) ỨNG DỤNG ĐỒ THỊ TRI THỨC MỜ TRONG HỖ TRỢ CHẨN ĐOÁN CHO BỆNH NHÂN BỊ ĐÁI THÁO ĐƯỜNG.* <https://jst.tnu.edu.vn/jst/article/view/9132>

[9] Xiaonan Li, Kai Zhang, Guanyu Li & Bin Zhu (2021) A Chinese Knowledge Graph for Cardiovascular Disease. https://link.springer.com/chapter/10.1007/978-981-15-8411-4_239

[10] Boya Cheng, Yuan Zhang, Cai Dejun, Wan Qiu & 1 other (2018) Construction of traditional Chinese medicine Knowledge Graph using Data Mining and Expert Knowledge. <https://typeset.io/papers/construction-of-traditional-chinese-medicine-knowledge-graph-crk2w3ri1f>