HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

-----ꙮ 🕮 ꙮ-----

PROJECT II

FINAL REPORT

DIAGNOSING CARDIOVASCULAR DISEASE USING FKG-PAIRS Fuzzy Knowledge Graph Model

Instructor : Assoc. Prof. PhD. Pham Van Hai

Student : Dao Thanh Manh 20211014

Hanoi , June 30th, 2024

TABLE OF CONTENTS

1. Introduce an overview of the problem

In the healthcare field, especially in cardiovascular disease decision-making and diagnosis, it is important to have a deep understanding of medical information and the ability to apply medical methods and knowledge. Modernity is extremely important. However, although there has been progress in collecting and organizing medical data, problems still exist due to the complexity and diversity of this information, especially when dealing with medical conditions. like cardiovascular disease.

Traditional knowledge graphs (KGs) cannot completely address these challenges. Although KG can organize medical data and build relationships between different medical elements, it often cannot handle the fuzziness and uncertainty of medical information, nor does it Flexible enough to apply complex rules and relationships in cardiovascular disease decision-making and diagnosis.

Meanwhile, the pairwise Fuzzy Knowledge Graph (FKG) opens up a new potential to improve the decision-making and diagnosis process of cardiovascular disease. FKG not only helps to represent medical information more flexibly and handle data fuzziness and uncertainty, but also solves the low performance problem of conventional KG by using information pairs instead because only a single pair is used. This enhances the ability to reason and make more accurate decisions in the treatment and management of cardiovascular disease, helping to improve the quality of healthcare and patient outcomes.

2. Objectives of the study

2.1. General objective

Develop a system/application to support decision making and cardiovascular disease diagnosis based on the FKG-Pairs Fuzzy Knowledge Graph model. The project focuses on building a diverse and high-quality medical database, using FKG-Pairs to represent medical information flexibly and efficiently. The goal is to develop a reliable and accurate cardiovascular disease prediction model that incorporates both predictive and explanatory analytical aspects. Finally, the project will evaluate and optimize the performance of the system before deploying into medical practice, to provide a useful tool to help improve the quality of diagnosis and treatment of cardiovascular diseases, and at the same time. while increasing understanding and trust from users.

2.2. Detail goal

Build FKG-Pairs Fuzzy Knowledge Graph models with k FKG pairs (k = 1, 2, 3), develop algorithms and methods to build FKG-Pairs models with accuracy and highly flexible, conduct performance evaluation of FKG-Pairs models based on criteria such as accuracy, classification and reliability, compare and analyze the results of FKG-Pairs models to determine the most optimal model for cardiovascular disease diagnosis, optimize and improve the most optimal model to ensure the highest performance and accuracy before deploying into medical practice, providing Provide detailed guidance and reports on project processes and results to share experiences and knowledge with the medical and research community.

3. Input data

3.1. Data set

Empirical data collected directly from Kaggle main page: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data

The dataset includes 1025 patient data, 13 attributes saved as **cardio_train.csv file** :

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholestoral in mg/dl
6. fasting blood sugar > 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina
10. oldpeak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by flourosopy
13. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

Figure 3.1: Data of 13 attributes of the heart disease data set

Data description:

STT	Thuộc tính	Thang đo	Biến ngôn ngữ	Ký hiệu biến	Ghi chú
1	age	<40	Low	L1	Tuổi của bệnh nhân (tính theo năm)
		40-60	Medium	M1	
		61-80	High	H1	
2	sex	0	Female	0	Giới tính của bệnh nhân (0: nữ, 1: nam)
		1	Male	1	
3	cp	0	Typical angina	0	Loại đau ngực (0: Đau thắt ngực điển hình, 1: Đau thắt ngực không điển hình, 2: Đau không do thắt ngực, 3: Không triệu chứng)
		1	Atypical angina	1	
		2	Non-anginal pain	2	
		3	Asymptomatic	3	
4	trestbps	<90	Low	L4	Huyết áp khi nghỉ ngơi (tính bằng mm Hg) của bệnh nhân khi nhập viện
		90-120	Medium	M4	
		>120	High	H4	
5	chol	<200	Medium	M5	Mức cholesterol trong huyết thanh (tính bằng mg/dl)
		200-240	High	H5	
		>240	Very High	VH5	
6	fbs	0	Medium	M6	Mức đường huyết lúc đói (trên 120 mg/dl được xem là 1 và <= 120 mg/dl được xem là 0)
		1	High	H6	
7	restecg	0	Medium	M7	Kết quả điện tâm đồ khi nghỉ ngơi. Được phân loại thành ba loại: 0: Bình thường 1: Có bất thường sóng ST-T (đảo ngược sóng T và/hoặc ST tăng hoặc giảm > 0.05 mV) 2: Cho thấy có khả năng hoặc chắc chắn phì đại thất trái theo tiêu chí của Estes
		1	High	H7	
		2	Very High	VH7	
8	thalach	<100	Medium	M8	Nhịp tim tối đa đạt được
		>100	High	H8	
9	exang	0	No	0	Đau thắt ngực do gắng sức (1 = có, 0 = không)
		1	Yes	1	
10	oldpeak	<2	Low	L9	Suy giảm ST do gắng sức so với khi nghỉ ngơi (đơn vị -> suy giảm)
		>=2	High	H9	
11	slope	0	Upsloping	0	Độ dốc của đoạn ST khi gắng sức. Được phân loại thành ba loại: 0: Độ dốc tăng 1: Bằng phẳng 2: Độ dốc giảm
		1	Flat	1	
		2	Downsloping	2	
12	ca	0	Normal	0	Số lượng mạch máu chính (0-3) được nhuộm màu bằng cách chụp cắt lớp thallium: Kết quả kiểm tra căng thẳng thallium (0: Bình thường, 1: Khuyết tật cố định, 2: Khuyết tật có thể đảo ngược, 3: Không được mô tả)
		1	Fixed defect	1	
		2	Reversible defect	2	
		3	Not described	3	
		4		4	
13	thal	0	Normal	0	Bệnh máu gọi là thalassemia. Được phân loại thành ba loại: 1: Bình thường 2: Khuyết tật cố định 3: Khuyết tật có thể đảo ngược
		1	Fixed	1	
		2	Reversible	2	
		3		3	

Figure 3.2: Data description of 13 heart disease attributes

All dataset values were collected at the time of medical examination.

The data file after downloading is compiled into the internal **heart sheet** excel file **Data.xlsx:**

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3,1	0	0	3	0
70	1	0	145	174	0	1	125	1	2,6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1,9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4,4	0	3	1	0
55	1	0	160	289	0	0	145	1	0,8	1	1	3	0
46	1	0	120	249	0	0	144	0	0,8	2	0	3	0
54	1	0	122	286	0	0	116	1	3,2	1	2	2	0
71	0	0	112	149	0	1	125	0	1,6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0,7	2	0	2	1
51	1	0	140	298	0	1	122	1	4,2	1	3	3	0
52	1	0	128	204	1	1	156	1	1	1	0	0	0
34	0	1	118	210	0	1	192	0	0,7	2	0	2	1
51	0	2	140	308	0	0	142	0	1,5	2	1	2	1
54	1	0	124	266	0	0	109	1	2,2	1	1	3	0
50	0	1	120	244	0	1	162	0	1,1	2	0	2	1
58	1	2	140	211	1	0	165	0	0	2	0	2	1
60	1	2	140	185	0	0	155	0	3	1	0	2	0
67	0	0	106	223	0	1	142	0	0,3	2	2	2	1
45	1	0	104	208	0	0	148	1	3	1	0	2	1
63	0	2	135	252	0	0	172	0	0	2	0	2	1
42	0	2	120	209	0	1	173	0	0	1	0	2	1
61	0	0	145	307	0	0	146	1	1	1	0	3	0
44	1	2	130	233	0	1	179	1	0,4	2	0	2	1
58	0	1	136	319	1	0	152	0	0	2	2	2	0
56	1	2	130	256	1	0	142	1	0,6	1	1	1	0
55	0	0	180	327	0	2	117	1	3,4	1	0	2	0
44	1	0	120	169	0	1	144	1	2,8	0	0	1	0
50	0	1	120	244	0	1	162	0	1,1	2	0	2	1
57	1	0	130	131	0	1	115	1	1,2	1	1	3	0
70	1	2	160	269	0	1	112	1	2,9	1	1	3	0
50	1	2	129	196	0	1	163	0	0	2	0	2	1
46	1	2	150	231	0	1	147	0	3,6	1	0	2	0

heart | data | Properties | Data1 | FuzzyData_heart | fuzzyData | Test | fuzzyTest

Figure 3.3: Heart disease data in excel file before processing

3.2. Data processing

Process the data set taken from kaggle using Visual Studio Code then save the processed data into sheets of the original excel file **Data.xlsx**

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Medium	1	0	High	High	Medium	High	High	0	Low	2	2	3	0
Medium	1	0	High	High	High	Medium	High	1	High	0	0	3	0
High	1	0	High	Medium	Medium	High	High	1	High	0	0	3	0
High	1	0	High	High	Medium	High	High	0	Low	2	1	3	0
High	0	0	High	Very High	High	High	High	0	Low	1	3	2	0
Medium	0	0	Medium	Very High	Medium	Medium	High	0	Low	1	0	2	1
Medium	1	0	Medium	Very High	Medium	Very High	High	0	High	0	3	1	0
Medium	1	0	High	Very High	Medium	Medium	High	1	Low	1	1	3	0
Medium	1	0	Medium	Very High	Medium	Medium	High	0	Low	2	0	3	0
Medium	1	0	High	Very High	Medium	Medium	High	1	High	1	2	2	0
High	0	0	Medium	Medium	Medium	High	High	0	Low	1	0	2	1
Medium	0	0	High	Very High	High	Medium	High	1	High	1	0	3	0
Low	0	1	Medium	High	Medium	High	High	0	Low	2	0	2	1
Medium	1	0	High	Very High	Medium	High	High	1	High	1	3	3	0
Medium	1	0	High	High	High	High	High	1	Low	1	0	0	0
Low	0	1	Medium	High	Medium	High	High	0	Low	2	0	2	1
Medium	0	2	High	Very High	Medium	Medium	High	0	Low	2	1	2	1
Medium	1	0	High	Very High	Medium	Medium	High	1	High	1	1	3	0
Medium	0	1	Medium	Very High	Medium	High	High	0	Low	2	0	2	1
Medium	1	2	High	High	High	Medium	High	0	Low	2	0	2	1
Medium	1	2	High	Medium	Medium	Medium	High	0	High	1	0	2	0
High	0	0	Medium	High	Medium	High	High	0	Low	2	2	2	1
Medium	1	0	Medium	High	Medium	Medium	High	1	High	1	0	2	1
High	0	2	High	Very High	Medium	Medium	High	0	Low	2	0	2	1
Medium	0	2	Medium	High	Medium	High	High	0	Low	1	0	2	1
High	0	0	High	Very High	Medium	Medium	High	1	Low	1	0	3	0
Medium	1	2	High	High	Medium	High	High	1	Low	2	0	2	1
Medium	0	1	High	Very High	High	Medium	High	0	Low	2	2	2	0
Medium	1	2	High	Very High	High	Medium	High	1	Low	1	1	1	0
Medium	0	0	High	Very High	Medium	Very High	High	1	High	1	0	2	0
Medium	1	0	Medium	Medium	Medium	High	High	1	High	0	0	1	0
Medium	0	1	Medium	Very High	Medium	High	High	0	Low	2	0	2	1
Medium	1	0	High	Medium	Medium	High	High	1	Low	1	1	3	0
High	1	2	High	Very High	Medium	High	High	1	High	1	1	3	0
Medium	1	2	High	Medium	Medium	High	High	0	Low	2	0	2	1
Medium	1	2	High	High	Medium	High	High	0	High	1	0	2	0

heart | data | Properties | Data1 | FuzzyData_heart | fuzzyData | Test | fuzzyTest

Figure 3.4: Heart disease data in excel file after threshold processing

4. Related knowledge

4.1. Fuzzy sets and fuzzy logic.

Fuzzy sets were first introduced by Zadeh in 1965, introduced as a new mathematical tool for solving problems with ambiguous, uncertain information. Unlike normal sets, which evaluate the membership relationship of a set according to binary logic "an element belongs or does not belong to the set", fuzzy logic evaluates the membership relationship of an element through a function. membership $\mu \rightarrow [0,1]$, represents the membership of an element to a set

4.2. Fuzzy inference system.

Fuzzy inference is the process of finding conclusions for a set of input values, based on a synthesized fuzzy rule system. Fuzzy inference methods are often referred to as Mamdani fuzzy inference, Takagi-Sugeno fuzzy inference, etc. The above inference systems are also known as classical inference methods, which have been widely used in automatic control systems. Fuzzy knowledge graph is known as a new, effective, and more accurate inference method than previous inference methods. The general rule for applying the fuzzy inference system is shown in three steps:

- Fuzzification: In this step, we need to determine the value scale and corresponding level terms of each input attribute of the data set, followed by the conversion process from explicit values. of the input data set into fuzzy values, based on the value scale combined with the previously built membership function, finally combining the fuzzy values of each input data sample using operators. fuzzy (AND, OR, NOT) to provide representation rules in the form of IF-THEN clauses and put them into the fuzzy rule base system
- Fuzzy inference: Use fuzzy inference method to find output results based on the fuzzy rule system built in step 1
- Defuzzification: Converts the fuzzy output values found in step 2 into clear values, giving the results of the problem.

4.3. Fuzzy knowledge graph

The term fuzzy knowledge graph was first introduced and integrated in the M-CFIS-FKG model with the initial purpose of expanding the M-CFIS-R model to make the inference process of this model in the testing becomes faster. Inheriting the characteristics of knowledge graphs, formally, a fuzzy knowledge graph includes vertices representing the linguistic labels of attributes and output labels of rules, the corresponding edges are arcs. connection between vertices.

The way to calculate the weight values of edges of fuzzy knowledge graphs has been presented in detail in [7], and is briefly summarized as follows:

- For edges connecting two attribute vertices, for each pair of values $(X_i, X_j), 1 \leq i \leq j \leq m$, in rule t $R_t, t = \overline{1, k}$, the weight $A_{ij}^t$of this edge is calculated according to the formula:
$$A_{ij}^t = \frac{|X_i \text{ quan hệ với } X_j \text{ trong luật thứ } t|}{|R|}$$
- For edges connecting attribute vertices and output label vertices, for each pair $(X_i, l), 1 \leq i \leq j \leq m,\ l = \overline{1, C}$, in rule t $R_t, t = \overline{1, k}$, the weight $B_{ij}^t$of this edge is calculated according to the formula:
$$B_{il}^t = \left(\sum A_{ij}^t\right) \times \frac{|X_i \text{ quan hệ với nhãn } l \text{ trong luật thứ } t|}{|R|}$$

The results of the two sets of weights are stored in an adjacency matrix, representing the constructed fuzzy knowledge graph.

4.4. Approximate inference.

Approximate reasoning is defined as a tool for reasoning from propositions whose meaning is not clearly defined through fuzzy logic. Normally, the approximate reasoning method's accuracy of results is not as high as conventional reasoning techniques for clear data, however, the advantage of approximate reasoning is that it can perform argumentative reasoning. With linguistic variables, or natural language is data with ambiguous and unclear meanings

5. Proposed model

5.1. State the problem

Input: To build the fuzzy knowledge graph used in this problem, we need to set patient samples that have been diagnosed by doctors and experts based on the given attributes. This sample data set goes through preprocessing (fuzzy) and is saved into a fuzzy rule base system as shown in table 1. This fuzzy rule base system includes n rules $R_1, R_2, \dots, R_n$representing patient samples, m $S_1, S_2, \dots, S_m$representative attributes. represents the symptoms of the disease, and C output labels 1, 2, 3, …, C represents the doctor's diagnostic conclusion

In addition, there is one new patient outside the above rule system, shown as follows

IF $S_1$is “Low” and $S_2$is “Low” and $S_3$is “High” and $S_4$is “Very high” and … and $S_{m-1}$is “High” and $S_m$is “Low” THEN Conclusion = ?

	S_1	S_2	…	S_{m-1}	S_m	Conclude
R_1	High	High	…	Very high	High	first
R_2	Medium	Medium	…	Medium	Low	2
…	…	…	…	…	…	…
R_{n-1}	Medium	Medium	…	Medium	Medium	2
R_n	Low	Medium	…	Low	Low	3
Patient's symptoms	High Medium Low	High Medium	…	Very high High Medium Low	High Medium High	0,1,…,C

Table 5.1: Fuzzy rule base system

Output: Results of the new patient output diagnostic system based on existing fuzzy rules.

5.2. Proposed model

5.2.1. Problem model.

The system includes 2 stages that need to be processed:

- Data processing phase: Data collected from doctors and experts is raw data, this data goes into the data processing phase, the pre-processing step includes scale division, design corresponding level terms, then combined with the rule generation tool to provide a fuzzy rule base system of the problem, as a basis for the phase of building a fuzzy knowledge graph and diagnosis.
- Graph construction and diagnosis phase: based on the fuzzy rule base system formed in the above step, proceed to build the fuzzy knowledge graph FKG by calculating edge weight sets and saving it as an adjacency matrix. , thereby using the data of that FKG fuzzy knowledge graph to diagnose new cases

The model for the above two stages of the problem has been redrawn below:

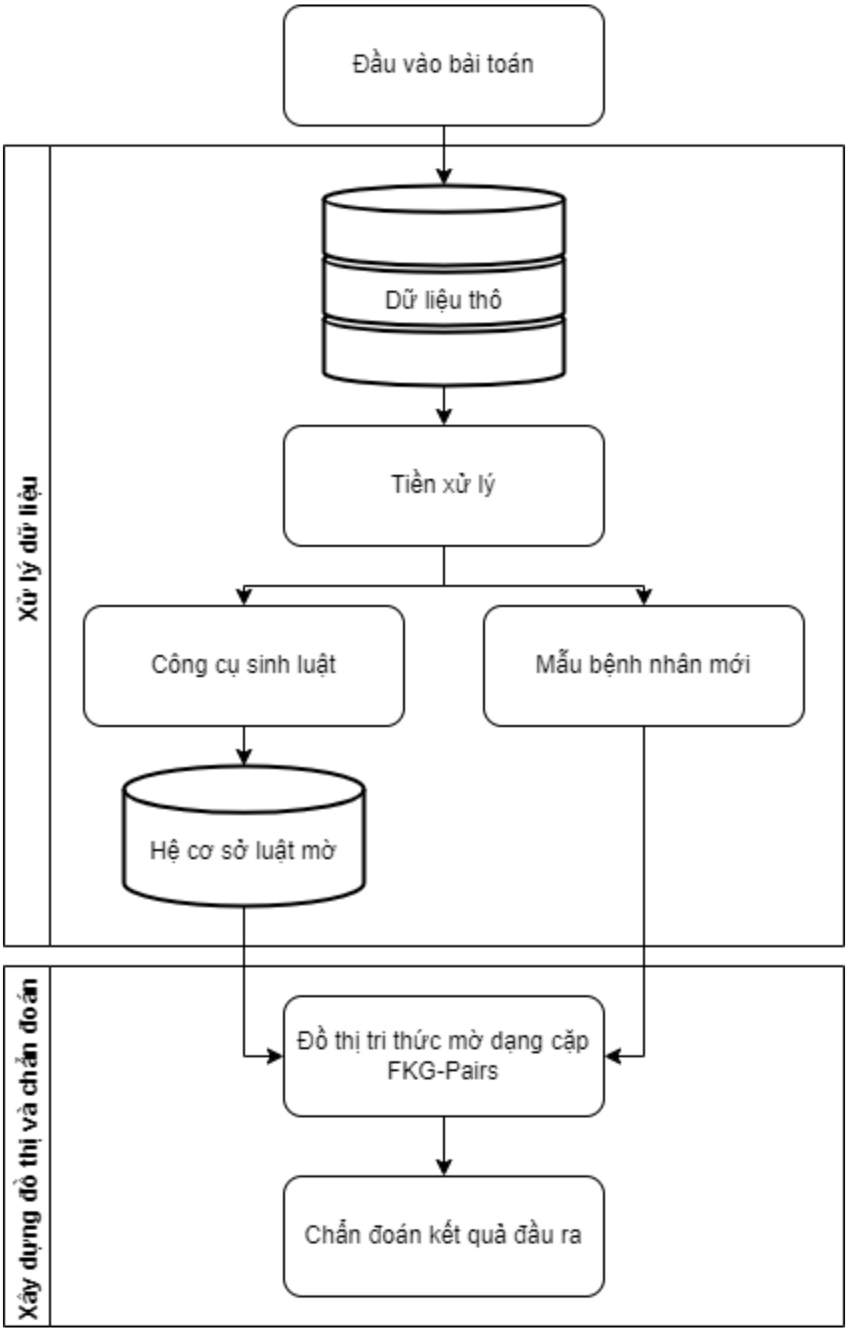

Figure 5.1: Problem model for diagnosing heart disease using FKG-Pairs

5.2.2. System functions

The system needs to have functions that can perform the two stages mentioned above

- For the data preparation phase: the raw data collected by the system designer as input to the problem needs to be analyzed, noise and errors removed, and a complete input data set provided. This data set needs to be blurred based on the scale and level terminology agreed upon with medical experts in the above field, through the rule generation tool to synthesize a fuzzy rule base system for use in medical research. subsequent functionality of the system

- For the performance and diagnosis phase: After obtaining the fuzzy rule base system, the system will have to build a fuzzy knowledge graph FKG and store this graph for use in the diagnosis step. At the diagnosis step, the system will receive input, combined with the fuzzy knowledge graph FKG to conduct diagnosis and draw conclusions.

5.2.3. Methods of implementation

As described in the system functions section, the problem will be installed according to the following steps:

- Step 1: Validate and process data

  After building the scale, build functions and procedures to control automatic data processing with conditions for each attribute based on the scale, combined with a rule generation tool, to give input. The output is the fuzzy rule base system as shown in table 1.

- Step 2: Calculate the weight sets$\tilde{A}, \tilde{B}$

  The set of weights $\tilde{A}$of the pairwise fuzzy knowledge graph is the weight of the edge connecting the linguistic labels of the attributes in rules t (R_t). These weights are calculated by the following formula:

$$\tilde{A}_{ij\ldots k}^{t} = \frac{\left|S_i \to S_j \to \cdots \to S_{k+1}\ trong\ luật\ thứ\ t\ \right|}{|R|} \quad (1)$$

  In there$t = \overline{1, n}, 1 \le i \le j < \cdots < k < m - 1.$

  The set of weights $\tilde{B}$of the pairwise fuzzy knowledge graph are the weights of the edges connecting the labels of the attribute pairs with the output label in the t-rule. (R_t).These weights are calculated by the following formula:

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

  In there$t = \overline{1, n}, 1 \le i \le j < \cdots < k < m - 1, l = \overline{1, C}.$

- Step 3: Store the pairwise fuzzy knowledge graph

The weight sets after calculation need to be stored as adjacency matrices for convenience in calculating the next steps. In this design, the adjacency matrices of the weight sets are stored in 2 sheets. of an excel file belonging to the file system.

- Step 4: Apply the approximate inference process and provide disease diagnosis results.

In [4], author Luong Thi Hong Lan used the FISA algorithm to approximate the output of the problem of determining the output of a new law based on the fuzzy knowledge graph, then author Cu Kim Long improved the above algorithm in his article [8] to be suitable for determining the output of new rules based on pairwise fuzzy knowledge graphs. The algorithm is described through the steps below:

First, it is necessary to calculate the weighted sum of the edges $(\tilde{C})$from the super vertices to the output label using the following formula:

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

In there$t = \overline{1,n}, 1 \le i \le j < \cdots < k < m-1, l = \overline{1,C}$.

Then, apply Max-Min operations to calculate the values $(\widetilde{D})$, these values help approximate the new attribute pairs in the Testing set with the corresponding attribute pairs in the pairwise fuzzy knowledge graph, to find out the level of influence on the output labels. The value of $(\widetilde{D})$is calculated based on the Max-Min operators according to the following formula:

$$\widetilde{D}_l = Max_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\dots kl}\right) + Min_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\dots kl}\right) \quad (4)$$

In there$t = \overline{1,n}, 1 \le i \le j < \cdots < k < m-1, l = \overline{1,C}$

Finally, the output label diagnosis result of the new rule is concluded using the Max operation as follows:

$$Label = p \; If \; \widetilde{D}_p = Max_{l=\overline{1,C}}\left(\widetilde{D}_l\right) \quad (5)$$

5.3. FKG-Pairs installation algorithm

Paired FKG installation algorithm

1 **Input data** : Test data set, m: Number of attributes of each rule, n: Number of samples of the data set, C: Number of labels of each attribute.

2 **Output data** : Label of new sample.

3 **Begin**
4 Enter values ;
5 Get test data set;
6 Conduct fuzzification of the test data set;
7 $\boldsymbol{for\ i = 1\ to\ m\ do}$
8 $\boldsymbol{for\ t = 1\ to\ n\ do}$
9 $\boldsymbol{for\ l = 1\ to\ C\ do}$
10 $\boldsymbol{while\ 1 \le i \le j < \cdots \le k\ do}$
11 $\boldsymbol{Tính\ \widetilde{C}^{t}_{ij\ldots kl} = \sum_{t} \widetilde{B}^{t}_{ij\ldots kl}}$
12 $\boldsymbol{Tính\ \widetilde{D}_{l} = Max_{1\le i\le j<\cdots\le k}\left(\widetilde{C}_{ij\ldots kl}\right) + Min_{1\le i\le j<\cdots\le k}\left(\widetilde{C}_{ij\ldots kl}\right)}$
13 $\boldsymbol{end}$
14 Determine the label of sample t:$\boldsymbol{Label = p\ If\ \widetilde{D}_{p} = Max_{l=\overline{1,C}}\left(\widetilde{D}_{l}\right)}$
15 $\boldsymbol{end}$
16 Get the label of sample t and repeat steps 5 to 12 to find the labels of other samples until the end
17 $\boldsymbol{end}$
18 $\boldsymbol{end}$
19 $\boldsymbol{end}$

Algorithm for setting up pairwise FKG problem

5.4. Technology selection

After analyzing the problem, as well as the calculation and programming steps, to be suitable for building the system in the most effective way, the technologies selected for the specific system design are as follows:

a. Choose programming language.

The main programming language chosen for the above system design is Python, specifically Python version 3.10 because of the following advantages:

- Python is an easy-to-learn, easy-to-use programming language with simple command structures that are easy to understand and execute. At the same time, Python also demonstrates flexibility, efficiency, reliability and high speed, and can be used in many different environments.
- Strong support community, because Python is an increasingly popular and powerful language, the user community is also getting larger. When problems arise during the design process, it will be easier to find and debug them.
- Supported by hundreds of Python libraries and frameworks, due to receiving a lot of attention from users as well as famous businesses, Python is built with many powerful support libraries, in addition to many other cloud media service that provides cross-platform support through library-like tools.
- Python is used as a core programming language in academia due to its countless applications in Artificial Intelligence, Deep Learning, Data Science

- Automation, due to the support of many available tools and modules, Python is also the best performance enhancing tool in the process of automation and software testing.
- Pycharm IDE supports fast Python code editing, with many extensions to support faster code editing and most features are free with an edu account.

b. Choose an application interface design tool.

The tool that comes with Python to design the interface chosen is Kivy, this is a powerful tool, used for many platforms, supported by a large number of users in the programming community.

6. Mathematical method

6.1. Diagnosis of patient's heart disease based on the M-CFIS-FKG model using FKG-Pairs-1

Input: Suppose the input of the problem is a list of 6 patients $\{R_1, R_2, R_3, R_4, R_5, R_6\}$, each patient has test results expressed through attributes $\{S_1, S_2, S_3, S_4, S_5, S_6\}$. The above patient cases have been examined and diagnosed based on test results by doctors, the diagnostic conclusions "Normal", "Heart" and "Severe heart" are shown respectively. labels 0, 1, 2. After going through the "Data Processing" stage, a fuzzy rule base system is obtained as shown in Table 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

Table 1: The fuzzy rule base system assumes the medical examination results of six patients who have been diagnosed by a doctor.

Besides, Input has an additional new patient case shown as follows:

IF *S_1= "High", S_2= "Medium", S_3= "Medium", S_4= "Medium", S_5= "High", S_6= "Medium"* ***THEN*** *Conclusion = ?*

Output: Provide diagnostic conclusions for the above patient, based on the fuzzy rule base system given by Input.

The steps for the above problem are performed sequentially as follows:

- **Step 1:** Calculate the weight sets $\tilde{A}$, $\tilde{B}$

  The weight set $\tilde{A}$includes edges connecting the linguistic labels of the patient's attributes, calculated according to formula (1)

  For example, in case {1}, the weights $\tilde{A}$will be calculated as follows:

$$\tilde{A}_{12}^{1} = \frac{|High \rightarrow Medium|}{R} = \frac{1}{3}$$

$$\tilde{A}_{13}^{1} = \frac{|High \rightarrow High|}{R} = \frac{1}{3}$$

$$\tilde{A}_{14}^{1} = \frac{|High \rightarrow Medium|}{R} = \frac{1}{2}$$

  The weight set $\tilde{B}$of the edge connecting the labels of the attribute pairs with the output labels is calculated according to the formula:

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

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

  The calculation results of the entire weight matrix are shown in Tables 2 and 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

$\tilde{A}^t_{45}$	1/2	1/2	1/6	1/6	1/6	1/2
$\tilde{A}^t_{46}$	1/3	1/3	1/3	1/3	1/3	1/3
$\tilde{A}^t_{56}$	1/3	1/3	1/6	1/6	1/6	1/6

Table 2: Results of calculating the weight matrix$\widetilde{A}$

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

Table 3: Results of calculating the weight matrix$\widetilde{B}$

⇨ The weight sets will be combined with the fuzzy rule base system to represent the fuzzy knowledge graph.

- **Step 2:** Apply approximate inference method to provide disease diagnosis results.
  After obtaining the fuzzy knowledge graph (represented based on the weight sets and fuzzy rule base system in Step 1), continue to diagnose the new patient's results using the approximate inference method. First, it is necessary to calculate the total weight of the edges $(\tilde{C})$from the super vertices to the output label based on the formula:

$$\tilde{C}_{ij\ldots kl} = \sum_t \tilde{B}^t_{ij\ldots kl}$$

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

In there$t = \overline{1, n}, 1 \le i \le j < \cdots < k < m - 1, l = \overline{1, C}$.

Specific calculation results are summarized in Table 4.

	$Label$ 0	$Label$ 1	$Label$ 2
$\tilde{C}_{1l}$	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

Table 4: Results of calculating the weight matrix $\widetilde{\boldsymbol{C}}$

Based on the table above, continue to calculate the values $(\widetilde{D})$according to the formula:

$$\widetilde{D}_l = Max_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\ldots kl}\right) + Min_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\ldots kl}\right)$$

In there$t = \overline{1,n}, 1 \le i \le j < \cdots < k < m-1, l = \overline{1,C}$

Input has an additional new patient case represented as follows:

IF *S_1= "High", S_2= "Medium", S_3= "Medium", S_4= "Medium", S_5= "High", S_6= "Medium"* ***THEN*** *Conclusion = ?*

Label 0:

$\tilde{C}_{High1,0}$= *0,* $\tilde{C}_{Medium2,0}$= *23/36,* $\tilde{C}_{Medium3,0}$= *23/36,*
$\tilde{C}_{Medium4,0}$= *23/36,* $\tilde{C}_{High5,0}$= *0,* $\tilde{C}_{Medium6,0}$= *23/36*

$\widetilde{D}_0$= *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

Label 1:

$\tilde{C}_{High1,1}$= *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*

$\widetilde{D}_1$= *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($\tilde{C}_{High1,1}, \tilde{C}_{Medium2,1}, \tilde{C}_{Medium3,1}, \tilde{C}_{Medium4,1}, \tilde{C}_{High5,1}, \tilde{C}_{Medium6,1}$*)*

= June 17

Label 2:

$\tilde{C}_{High1,2}$= *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*

$\widetilde{D}_2$= *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

From there we have:

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

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

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

According to the formula:

$$Label = p \; If \; \widetilde{D}_p = Max_{l=\overline{1,C}}(\widetilde{D}_l)$$

We have: $Max_{l=\overline{0,2}}(\widetilde{D}_l) = D_2 = \frac{77}{9}$

Therefore Label = 2

Using the Max operation, we obtain the output label of the new patient as 2 $\left(D_2 = \frac{77}{9}\right)$, from which we can conclude that the new patient shows signs of severe heart disease.

6.2. Diagnosis of patient's heart disease based on the M-CFIS-FKG model using FKG-Pairs-2

Input: Suppose the input of the problem is a list of 6 patients { $R_1, R_2, R_3, R_4, R_5, R_6$}, each patient has test results expressed through attributes { $S_1, S_2, S_3, S_4, S_5, S_6$}. The above patient cases have been examined and diagnosed based on test results by doctors, the diagnostic conclusions "Normal", "Heart" and "Severe heart" are shown respectively. labels 0, 1, 2. After going through the "Data Processing" stage, a fuzzy rule base system is obtained as shown in Table 1.

	S_1	S_2	S_3	S_4	S_5	S_6	Result
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

Table 1: The fuzzy rule base system assumes the medical examination results of six patients who have been diagnosed by a doctor.

Besides, Input has an additional new patient case shown as follows:

IF *S_1= "High", S_2= "Medium", S_3= "Medium", S_4= "Medium", S_5= "High", S_6= "Medium"* ***THEN*** *Conclusion = ?*

Output: Provide diagnostic conclusions for the above patient, based on the fuzzy rule base system given by Input.

The steps for the above problem are performed sequentially as follows:

- **Step 1:** Calculate the weight sets $\tilde{A}$, $\tilde{B}$

  The weight set $\tilde{A}$includes edges connecting the linguistic labels of the patient's attributes, calculated according to formula (1)

  For example, in case {1}, the weights $\tilde{A}$will be calculated as follows:

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

  The weight set $\tilde{B}$of the edge connecting the labels of the attribute pairs with the output labels is calculated according to the formula:

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

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

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

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

The calculation results of the entire weight matrix are shown in Tables 2 and 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

Table 2: Results of calculating the weight matrix$\tilde{A}$

	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

Table 3: Results of calculating the weight matrix $\widetilde{\mathbf{B}}$

⇨ The weight sets will be combined with the fuzzy rule base system to represent the fuzzy knowledge graph.

- **Step 2:** Apply approximate inference method to provide disease diagnosis results.
After obtaining the fuzzy knowledge graph (represented based on the weight sets and fuzzy rule base system in Step 1), continue to diagnose the new patient's results using the approximate inference method. First, it is necessary to calculate the total weight of the edges $(\tilde{C})$from the super vertices to the output label based on the formula:

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

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

In there$t = \overline{1,n}, 1 \le i \le j < \dots < k < m-1, l = \overline{1,C}$.

Specific calculation results are summarized in Table 4.

	$Label\ 0$	$Label\ 1$	$Label\ 2$
$\tilde{C}_{12l}$	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

Table 4: Results of calculating the weight matrix$\widetilde{C}$

Based on the table above, continue to calculate the values $(\widetilde{D})$according to the formula:

$$\widetilde{D}_l = Max_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\ldots kl}\right) + Min_{1\le i\le j<\cdots\le k}\left(\tilde{C}_{ij\ldots kl}\right)$$

In there$t = \overline{1,n}, 1 \le i \le j < \cdots < k < m-1, l = \overline{1,C}$

Input has an additional new patient case represented as follows:

IF *S_1= "High", S_2= "Medium", S_3= "Medium", S_4= "Medium", S_5= "High", S_6= "Medium"* ***THEN*** *Conclusion = ?*

Label 0:

$\tilde{C}_{High1\to Medium2,0}$= *0,*

$\tilde{C}_{High1\to Medium3,0}$= *0,*

$\tilde{C}_{High1\to Medium4,0}$= *0,*

$\tilde{C}_{High1\to High5,0}$= *0,*

$\tilde{C}_{High1\to Medium6,0}$= *0,*

$\tilde{C}_{Medium2\to Medium3,0}$= *11/18,*

$\tilde{C}_{Medium2\to Medium4,0}$= *11/18,*

$\tilde{C}_{Medium2\to High5,0}$= *0,*

$\tilde{C}_{Medium2\to Medium6,0}$= *11/18,*

$\tilde{C}_{Medium3\to Medium4,0}$= *11/18,*

$\tilde{C}_{Medium3\to High5,0}$= *0,*

$\tilde{C}_{Medium3\to Medium6,0}$= *11/18,*

$\tilde{C}_{Medium4\to High5,0}$= *0,*

$\tilde{C}_{Medium4\to Medium6,0}$= *11/18,*

$\tilde{C}_{High5\to Medium6,0}$= *0*

$\widetilde{D}_0$=
max(
$\tilde{C}_{High1\to Medium2,0}, \tilde{C}_{High1\to Medium3,0}, \tilde{C}_{High1\to Medium4,0}, \tilde{C}_{High1\to High5,0}, \tilde{C}_{High1\to Medium6,0},$
$\tilde{C}_{Medium2\to Medium3,0}, \tilde{C}_{Medium2\to Medium4,0}, \tilde{C}_{Medium2\to High5,0},$

$\tilde{C}_{Medium2\to Medium6,0}$, $\tilde{C}_{Medium3\to Medium4,0}$, $\tilde{C}_{Medium3\to High5,0}$, $\tilde{C}_{Medium3\to Medium6,0}$,
$\tilde{C}_{Medium4\to High5,0}$, $\tilde{C}_{Medium4\to Medium6,0}$, $\tilde{C}_{High5\to Medium6,0}$) +
min(
$\tilde{C}_{High1\to Medium2,0}$, $\tilde{C}_{High1\to Medium3,0}$, $\tilde{C}_{High1\to Medium4,0}$, $\tilde{C}_{High1\to High5,0}$, $\tilde{C}_{High1\to Medium6,0}$,
$\tilde{C}_{Medium2\to Medium3,0}$, $\tilde{C}_{Medium2\to Medium4,0}$, $\tilde{C}_{Medium2\to High5,0}$,
$\tilde{C}_{Medium2\to Medium6,0}$, $\tilde{C}_{Medium3\to Medium4,0}$, $\tilde{C}_{Medium3\to High5,0}$, $\tilde{C}_{Medium3\to Medium6,0}$,
$\tilde{C}_{Medium4\to High5,0}$, $\tilde{C}_{Medium4\to Medium6,0}$, $\tilde{C}_{High5\to Medium6,0}$)

= 11/18

Label 1:

$\tilde{C}_{High1\to Medium2,1} = 0$,
$\tilde{C}_{High1\to Medium3,1} = 25/36$,
$\tilde{C}_{High1\to Medium4,1} = 0$,
$\tilde{C}_{High1\to High5,1} = 0$,
$\tilde{C}_{High1\to Medium6,1} = 0$,
$\tilde{C}_{Medium2\to Medium3,1} = 0$,
$\tilde{C}_{Medium2\to Medium4,1} = 0$,
$\tilde{C}_{Medium2\to High5,1} = 0$,
$\tilde{C}_{Medium2\to Medium6,1} = 0$,
$\tilde{C}_{Medium3\to Medium4,1} = 0$,
$\tilde{C}_{Medium3\to High5,1} = 0$,
$\tilde{C}_{Medium3\to Medium6,1} = 0$,
$\tilde{C}_{Medium4\to High5,1} = 0$,
$\tilde{C}_{Medium4\to Medium6,1} = 0$,
$\tilde{C}_{High5\to Medium6,1} = 0$

$\widetilde{D}_1 =$
max(
$\tilde{C}_{High1\to Medium2,1}$, $\tilde{C}_{High1\to Medium3,1}$, $\tilde{C}_{High1\to Medium4,1}$, $\tilde{C}_{High1\to High5,1}$, $\tilde{C}_{High1\to Medium6,1}$,
$\tilde{C}_{Medium2\to Medium3,1}$, $\tilde{C}_{Medium2\to Medium4,1}$, $\tilde{C}_{Medium2\to High5,1}$,
$\tilde{C}_{Medium2\to Medium6,1}$, $\tilde{C}_{Medium3\to Medium4,1}$, $\tilde{C}_{Medium3\to High5,1}$, $\tilde{C}_{Medium3\to Medium6,1}$,
$\tilde{C}_{Medium4\to High5,1}$, $\tilde{C}_{Medium4\to Medium6,1}$, $\tilde{C}_{High5\to Medium6,1}$) +
min(

$\tilde{C}_{High1\to Medium2,1}, \tilde{C}_{High1\to Medium3,1}, \tilde{C}_{High1\to Medium4,1}, \tilde{C}_{High1\to High5,1}, \tilde{C}_{High1\to Medium6,1},$
$\tilde{C}_{Medium2\to Medium3,1}, \tilde{C}_{Medium2\to Medium4,1}, \tilde{C}_{Medium2\to High5,1},$
$\tilde{C}_{Medium2\to Medium6,1}, \tilde{C}_{Medium3\to Medium4,1}, \tilde{C}_{Medium3\to High5,1}, \tilde{C}_{Medium3\to Medium6,1},$
$\tilde{C}_{Medium4\to High5,1}, \tilde{C}_{Medium4\to Medium6,1}, \tilde{C}_{High5\to Medium6,1})$
= 25/36

Label 2:

$\tilde{C}_{High1\to Medium2,2}$ = 67/18,
$\tilde{C}_{High1\to Medium3,2}$ = 11/12,
$\tilde{C}_{High1\to Medium4,2}$ = March 23,
$\tilde{C}_{High1\to High5,2}$ = March 23,
$\tilde{C}_{High1\to Medium6,2}$ = 25/36,
$\tilde{C}_{Medium2\to Medium3,2}$ = 11/12,
$\tilde{C}_{Medium2\to Medium4,2}$ = 67/18,
$\tilde{C}_{Medium2\to High5,2}$ = 67/18,
$\tilde{C}_{Medium2\to Medium6,2}$ = 0,

$\tilde{C}_{Medium3\to Medium4,2}$ = 11/12,
$\tilde{C}_{Medium3\to High5,2}$ = June 11,
$\tilde{C}_{Medium3\to Medium6,2}$ = 0,
$\tilde{C}_{Medium4\to High5,2}$ = March 23,
$\tilde{C}_{Medium4\to Medium6,2}$ = 25/36,
$\tilde{C}_{High5\to Medium6,2}$ = 25/36

$\tilde{D}_2 =$
max(
$\tilde{C}_{High1\to Medium2,2}, \tilde{C}_{High1\to Medium3,2}, \tilde{C}_{High1\to Medium4,2}, \tilde{C}_{High1\to High5,2}, \tilde{C}_{High1\to Medium6,2},$
$\tilde{C}_{Medium2\to Medium3,2}, \tilde{C}_{Medium2\to Medium4,2}, \tilde{C}_{Medium2\to High5,2},$
$\tilde{C}_{Medium2\to Medium6,2}, \tilde{C}_{Medium3\to Medium4,2}, \tilde{C}_{Medium3\to High5,2}, \tilde{C}_{Medium3\to Medium6,2},$
$\tilde{C}_{Medium4\to High5,2}, \tilde{C}_{Medium4\to Medium6,2}, \tilde{C}_{High5\to Medium6,2})$ +
min(
$\tilde{C}_{High1\to Medium2,2}, \tilde{C}_{High1\to Medium3,2}, \tilde{C}_{High1\to Medium4,2}, \tilde{C}_{High1\to High5,2}, \tilde{C}_{High1\to Medium6,2},$
$\tilde{C}_{Medium2\to Medium3,2}, \tilde{C}_{Medium2\to Medium4,2}, \tilde{C}_{Medium2\to High5,2},$
$\tilde{C}_{Medium2\to Medium6,2}, \tilde{C}_{Medium3\to Medium4,2}, \tilde{C}_{Medium3\to High5,2}, \tilde{C}_{Medium3\to Medium6,2},$
$\tilde{C}_{Medium4\to High5,2}, \tilde{C}_{Medium4\to Medium6,2}, \tilde{C}_{High5\to Medium6,2})$
= March 23

From there we have:

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

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

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

According to the formula:

$$Label = p \; If \; \widetilde{D}_p = Max_{l=\overline{1,C}}(\widetilde{D}_l)$$

We have: $Max_{l=\overline{0,2}}(\widetilde{D}_l) = D_2 = \frac{23}{3}$

Therefore Label = 2

Using the Max operation, we obtain the output label of the new patient as 2 $\left(D_2 = \frac{23}{3}\right)$, from which we can conclude that the new patient shows signs of severe heart disease.

7. Results evaluated using mathematical methods

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

Table 7.1: Results of evaluating two models FKG-Pairs-1 and FKG-Pairs-2 using mathematical methods

Based on the results of Pairs-1 and Pairs-2 and evaluating the difference between the values D_0, D_1, D_2, we can make some comments as follows:

1. Performance of the Pairs 2 model: Pairs 2 has a smaller difference between values than Pair 1. This suggests that the Pair 2 model has the ability to clearly and clearly classify and predict risks. D_0, D_1, D_2 more accurate than Pair 1. Consistency in prediction across labels also enhances model reliability.
2. Pairs 1 model accuracy: Pairs 1 has a larger difference between values D_0, D_1, D_2, indicating that the model may have difficulty clearly classifying risk groups. Large discrepancies may suggest that the model needs to be adjusted or improved to increase the accuracy and reliability of predictions.
3. Flexibility and applicability: Both models use the M-CFIS-FKG model with FKG-Pairs-k to predict the risk of severe or mild heart disease. However, their performance can depend on how the model is built and how the data is processed.

For this particular application, Pair 2 appears to be more flexible and more effective in predicting risk.

8. Installation results

8.1. Evaluation methods

8.1.1. Accuracy

Accuracy is the ratio of correct predictions to the total number of predictions. The calculation formula is as follows:

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

- TP (True Positive): The number of predictions that are correct and are actually also correct.
- TN (True Negative): The number of false predictions and actual errors.
- FP (False Positive): Number of predictions that are correct but actually wrong.
- FN (False Negative): Number of false predictions but actually correct.

8.1.2. Time Calculating

Computation time is the time it takes a model to complete the training or prediction process. This time is measured in seconds, minutes or hours depending on the complexity of the model and the size of the data.

8.1.3. Precision (Predictive accuracy)

Precision is the ratio of correct predictions (True Positive) to the total number of predictions determined to be correct (True Positive + False Positive):

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

High Precision means that of the cases predicted to be positive, the majority are correct.

8.1.4. Recall

Recall is the ratio of correct predictions (True Positive) to the total number of actual cases that are correct (True Positive + False Negative):

$$\text{Recall} = \frac{TP}{TP+FN}$$

High recall means the model can detect most of the positive cases.

8.1.5. F1-Score

F1-Score is the harmonic average of Precision and Recall, providing a balanced measure of Precision and Recall:

$$\text{Precision} = 2 \text{ x} \frac{\text{Precision x Recall}}{Precision + Recall}$$

A high F1-Score shows that the model has both high Precision and Recall.

8.2. Experimental results

8.2.1. Accuracy and Caculating time

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
ACCURACY	FKG-Pairs-2	67,51	85,71	79,7	68,53	89,34	75,51	83,76	95,92	80,71	75,63	95,43	69,9	70,92	92,89	83,76	96,45	70,05	86,22	72,59	83,25	**81,189**
	FKG-Pairs-3	67,51	80,1	93,4	94,92	92,39	90,82	97,97	95,41	95,43	99,49	98,98	97,96	97,45	98,48	97,97	96,95	98,48	100	98,48	97,91	**94,505**
CACULATING TIME	FKG-Pairs-2	0,57	0,57	0,56	0,59	0,56	0,57	0,74	0,8	0,84	0,65	0,76	0,68	0,59	0,61	0,57	0,61	0,58	0,6	0,58	0,6	**0,6315**
	FKG-Pairs-3	0,74	0,92	1,04	1,04	1,34	1,65	1,39	1,48	1,59	1,8	1,97	2,6	2,43	2,82	3,02	3,58	3,56	4,04	4,2	4,75	**2,298**

Figure 8.1. Results of training FKG-Pairs 2 and FKG-Pairs 3

	FKG-Pairs 2	FKG-Pairs 3
Accuracy (%)	81,189	94,505
Time Calculating (s)	0,6315	2,298

Table 8.1. Accuracy and Time caculating are obtained from training results

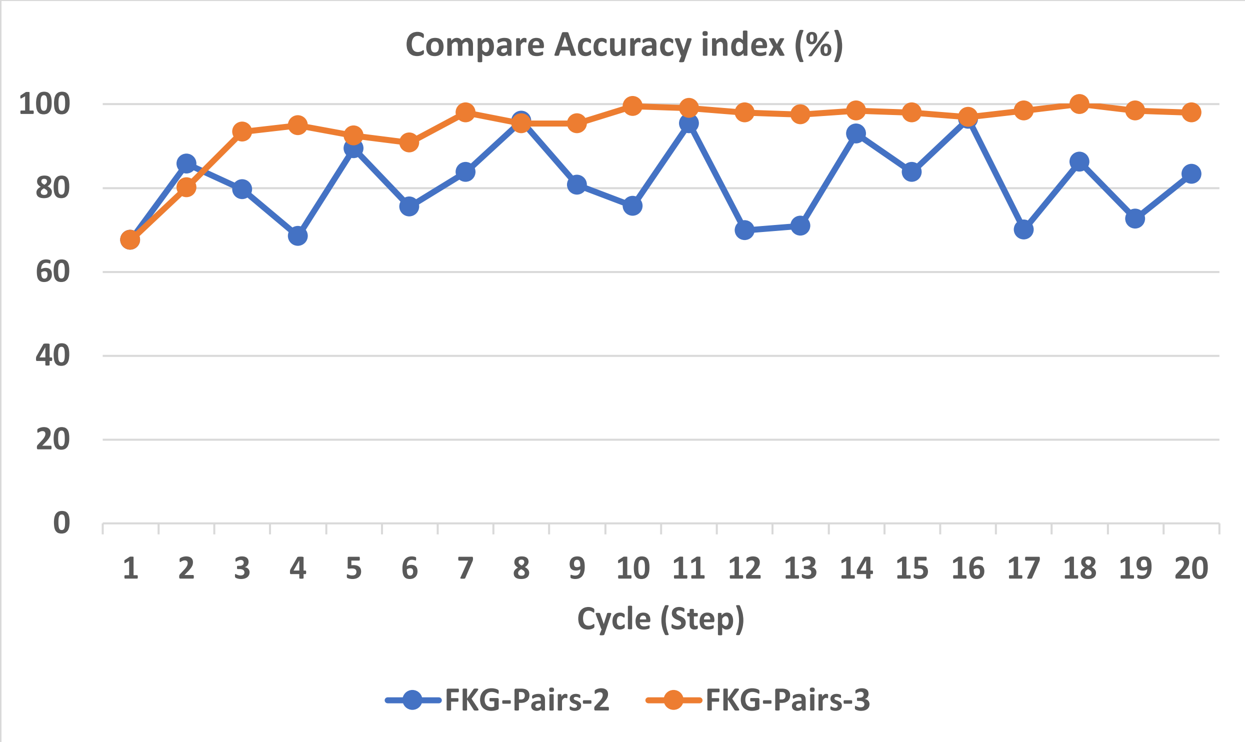

Figure 8.2. Compare the Accuracy index of FKG-Pairs 2 and FKG-Pairs 3

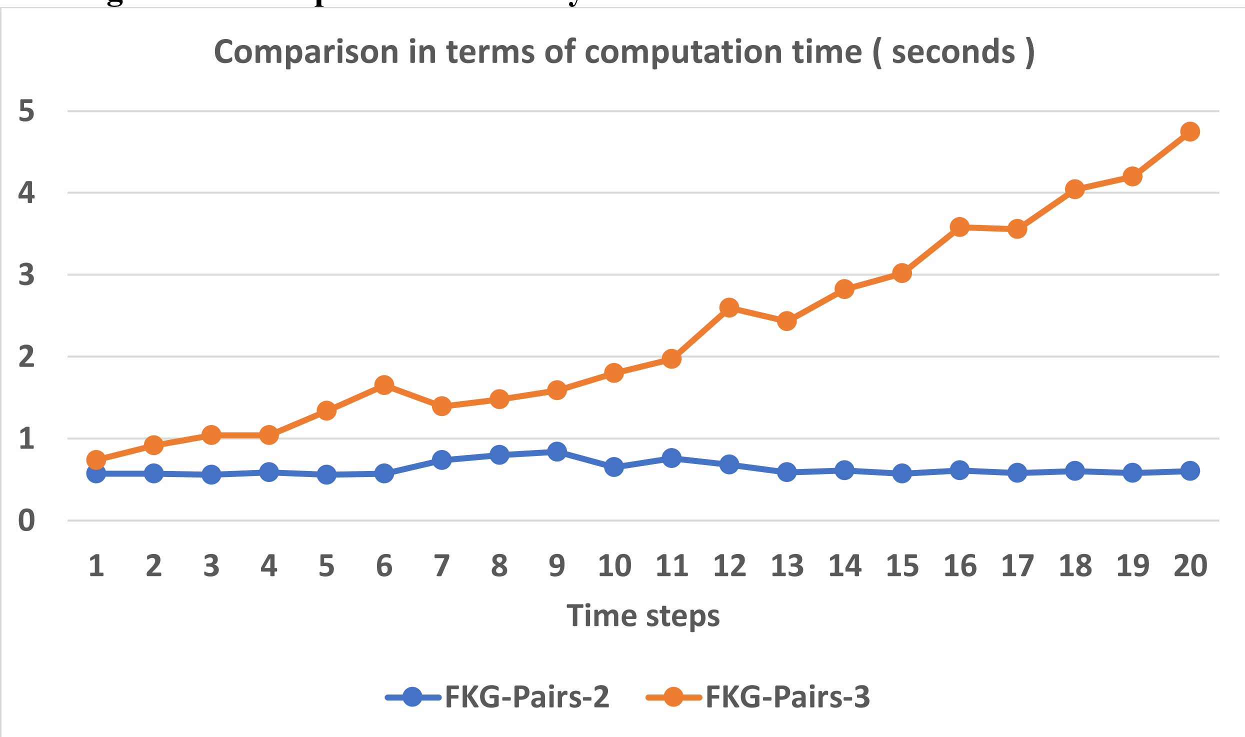

Figure 8.3: Comparison of calculation time of FKG-Pairs 2 and FKG-Pairs 3

8.2.2. Precision, Recall and F1-Score

Steps	Precision		Recall		F1-Score	
	FKG-Pairs 3	FKG-Pairs 2	FKG-Pairs 3	FKG-Pairs 2	FKG-Pairs 3	FKG-Pairs 2
1	42,86	42,86	100	100	60	60
2	62,86	70,21	100	100	77,2	82,5
3	82,67	86,67	100	41,94	90,51	56,53
4	86,11	0	100	0	92,54	0
5	77,27	70,83	100	100	87,18	82,92
6	74,29	52	100	100	85,25	68,42
7	94,67	97,56	100	56,34	97,26	71,43
8	85,25	86,67	100	100	92,04	92,86
9	87,32	96,15	100	40,32	93,23	56,81
10	98	100	100	2,04	98,99	4
11	96,49	85,94	100	100	98,21	92,44
12	93,75	100	100	1,67	96,77	3,29
13	93,65	71,43	98,33	8,33	95,93	14,92
14	94,64	93,33	100	79,25	97,25	85,72
15	93,94	65,96	100	100	96,88	79,49
16	90,48	89,06	100	100	95	94,21
17	95,16	0	100	0	97,52	0
18	100	100	100	52,63	100	68,96
19	94,64	0	100	0	97,25	0
20	92,31	60	100	100	96	75
Avengers	**86,818**	**68,4335**	**99,9165**	**59,126**	**92,2505**	**54,475**

Figure 8.4: Training results for FKG-Pairs 2 and FKG-Pairs 3

	FKG-Pairs 2	FKG-Pairs 3
Precision (%)	68,4335	86,818
Recall (%)	59,126	99,9165
F1-Score (%)	54,475	92,2505

Table 8.2: Precision, Recall and F1-Score extracted from training results

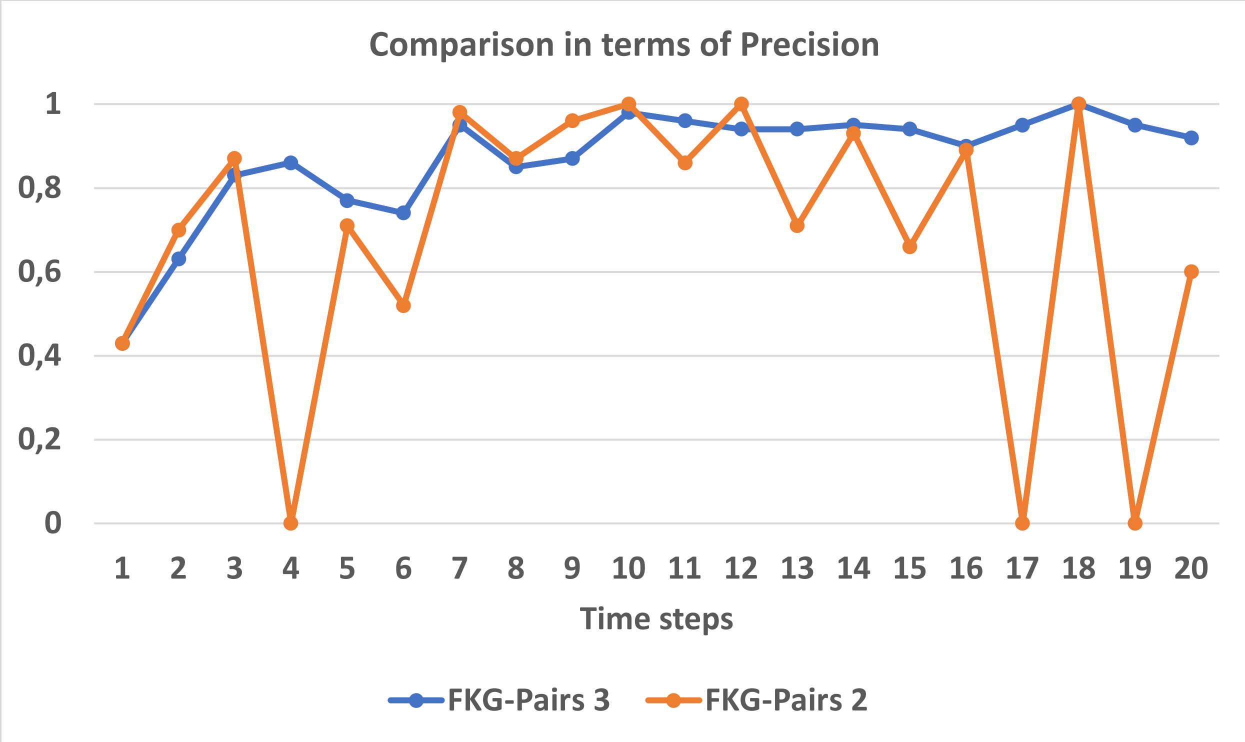

Figure 8.5: Comparison of Precision of FKG-Pairs 2 and FKG-Pairs 3

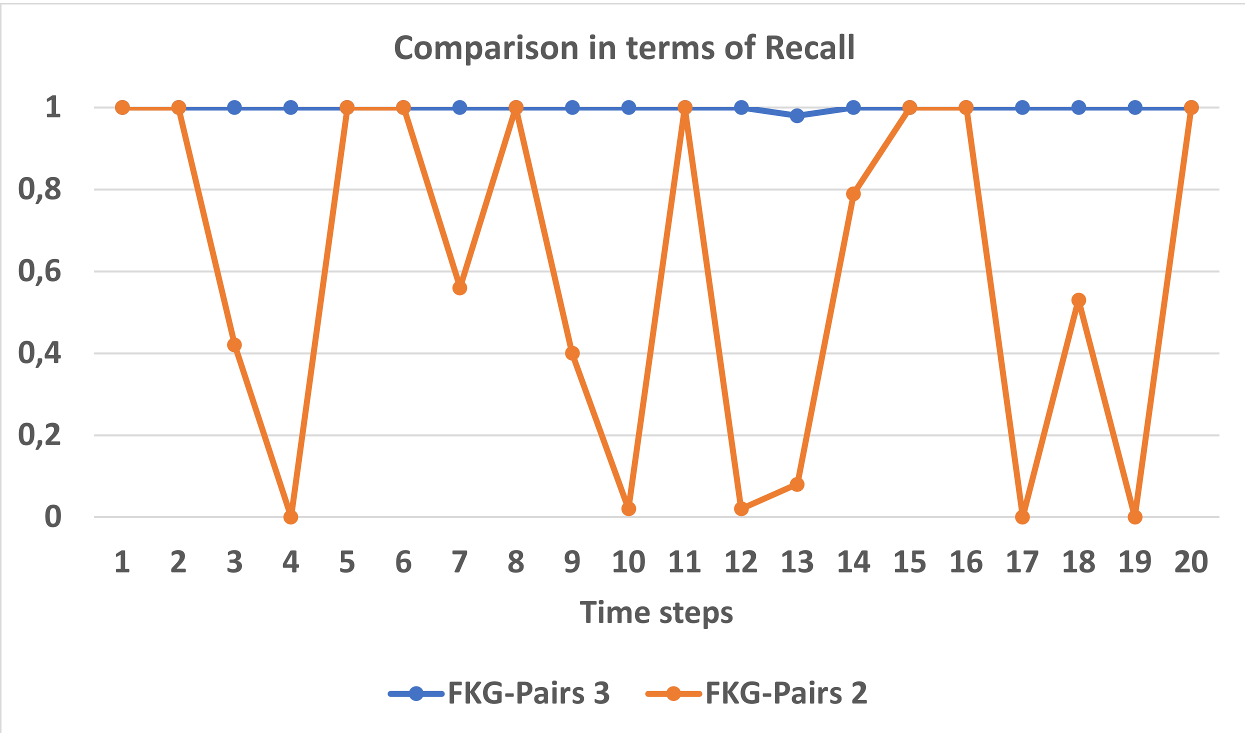

Figure 8.6: Comparison of Recall of FKG-Pairs 2 and FKG-Pairs 3

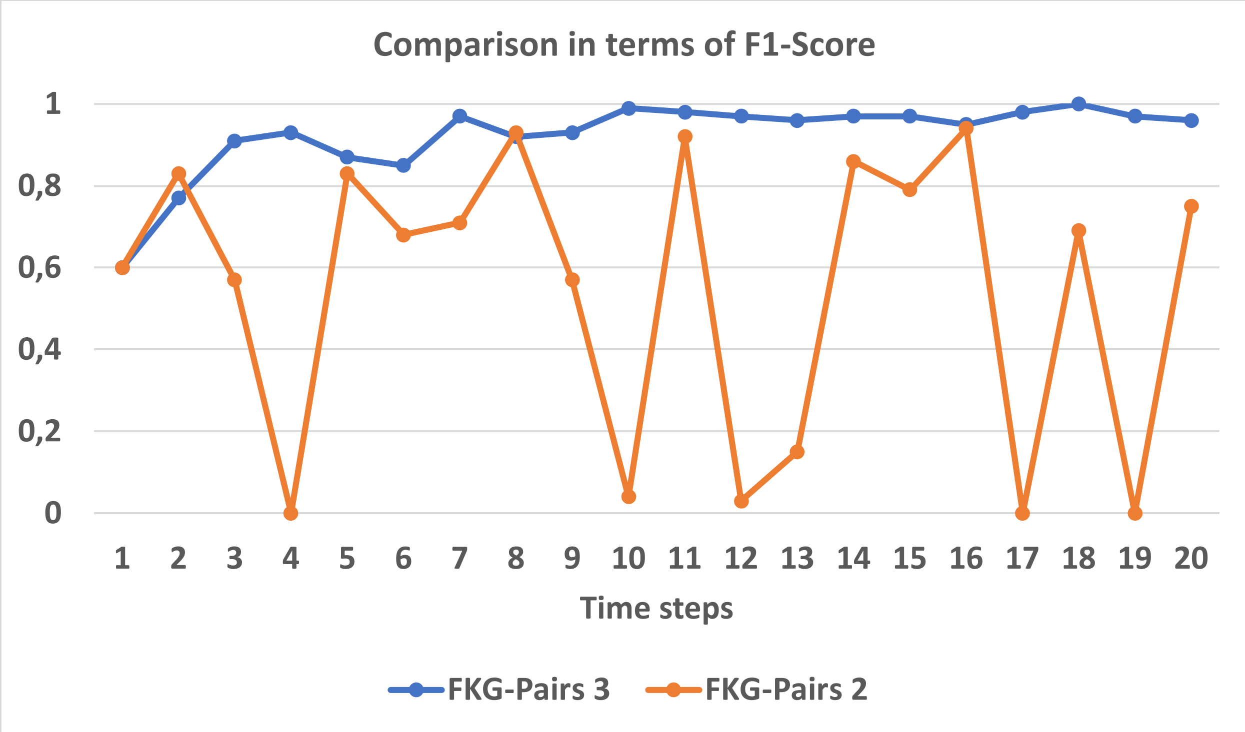

Figure 8.7: Comparison of F1-Score of FKG-Pairs 2 and FKG-Pairs 3

8.3. Evaluate experimental results

8.3.1. FKG-Pairs 2 model

Advantages : Calculation time is very fast.

Disadvantages : Accuracy, Precision, Recall and F1-Score are all average or low.

Suitable for : Applications that require high computational speed and accept average prediction performance.

8.3.2. FKG-Pairs model 3

Pros : Accuracy, Precision, Recall and F1-Score are all very high, indicating good overall performance.

Disadvantage : Longer calculation time compared to FKG-Pairs 2.

Suitable for : Applications requiring high precision and performance, longer calculation times can be tolerated.

8.4. Interface of heart disease diagnosis application using FKG-Pairs

8.4.1. Main screen

The main screen of the application is simple, including only the main function of the application which is disease diagnosis, and detailed instructions for using the application are implemented in the form of navigation buttons as shown in Figure 1 .

8.4.2. Disease diagnosis function

After clicking the diagnosis button, the application will redirect to the patient's preliminary information input screen (Figure 3), after entering the information in the corresponding box, click to redirect to the patient input screen. The patient's obtained test index (Figure 4), here, continue to enter the test indexes to make the diagnosis (note that indexes marked with (*) cannot be blank). , the diagnosis results will be displayed as shown in Figure 5 , including the diagnosis results, along with the doctor's recommendations for the patient, here, the application user can have the option to save the patient case new to the application database (Figure 6), the fuzzy knowledge graph will be periodically updated to increase the accuracy of the diagnostic model.

8.4.3. Instruction screen

In this screen (Figure 2) , there will be instructions for using the application similar to the one presented above.

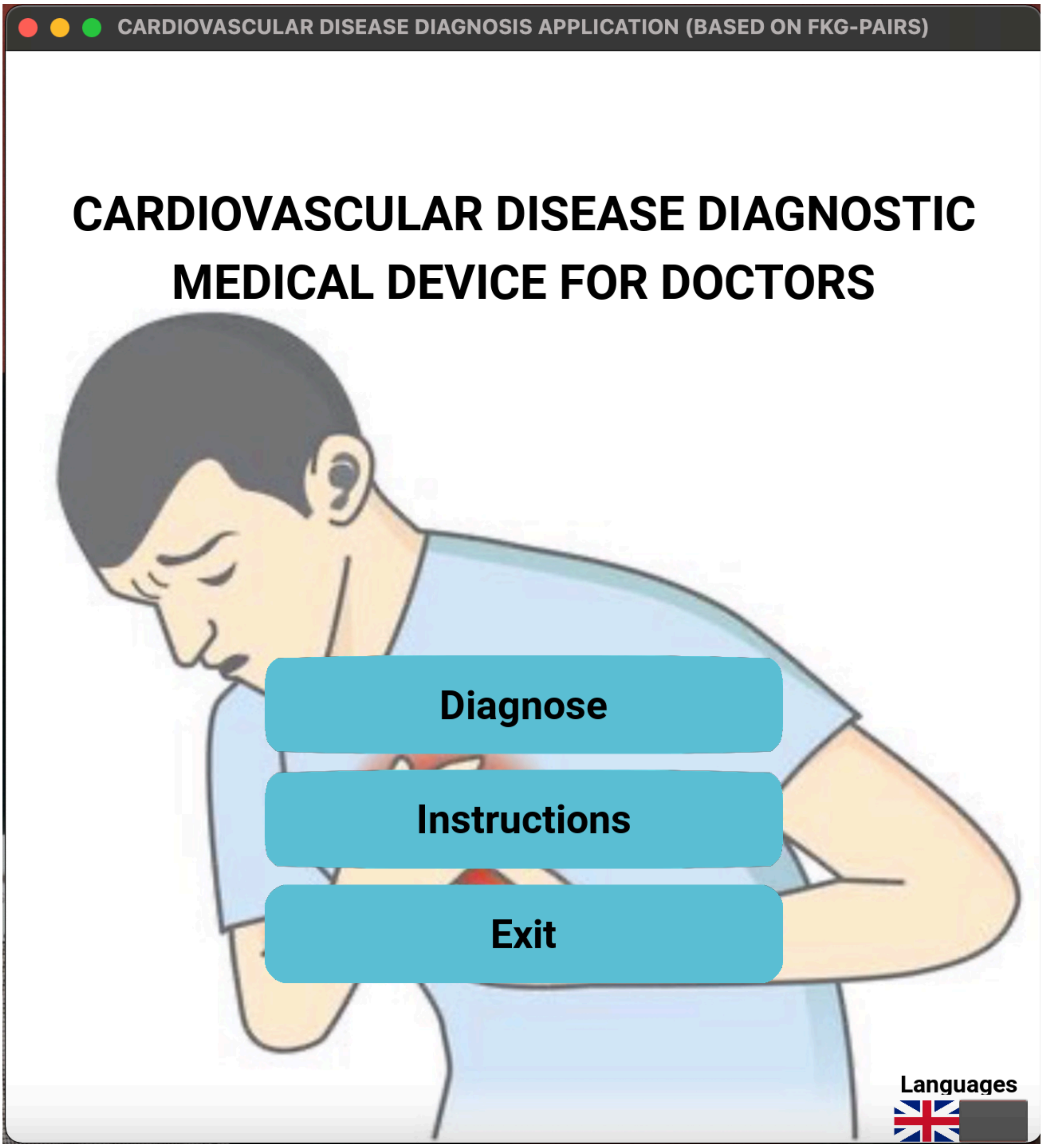

Figure 1: Main screen

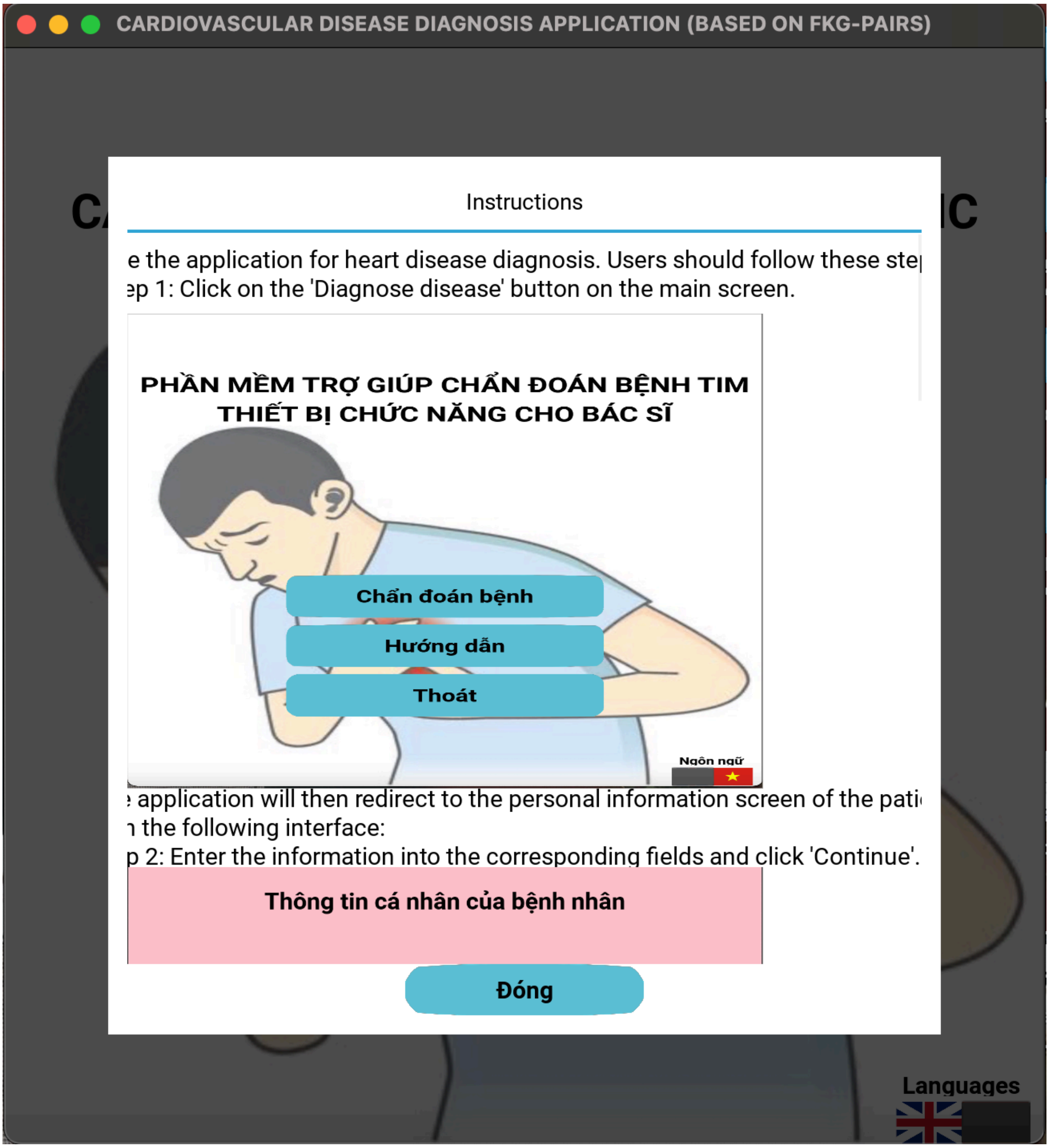

Figure 2: User manual screen

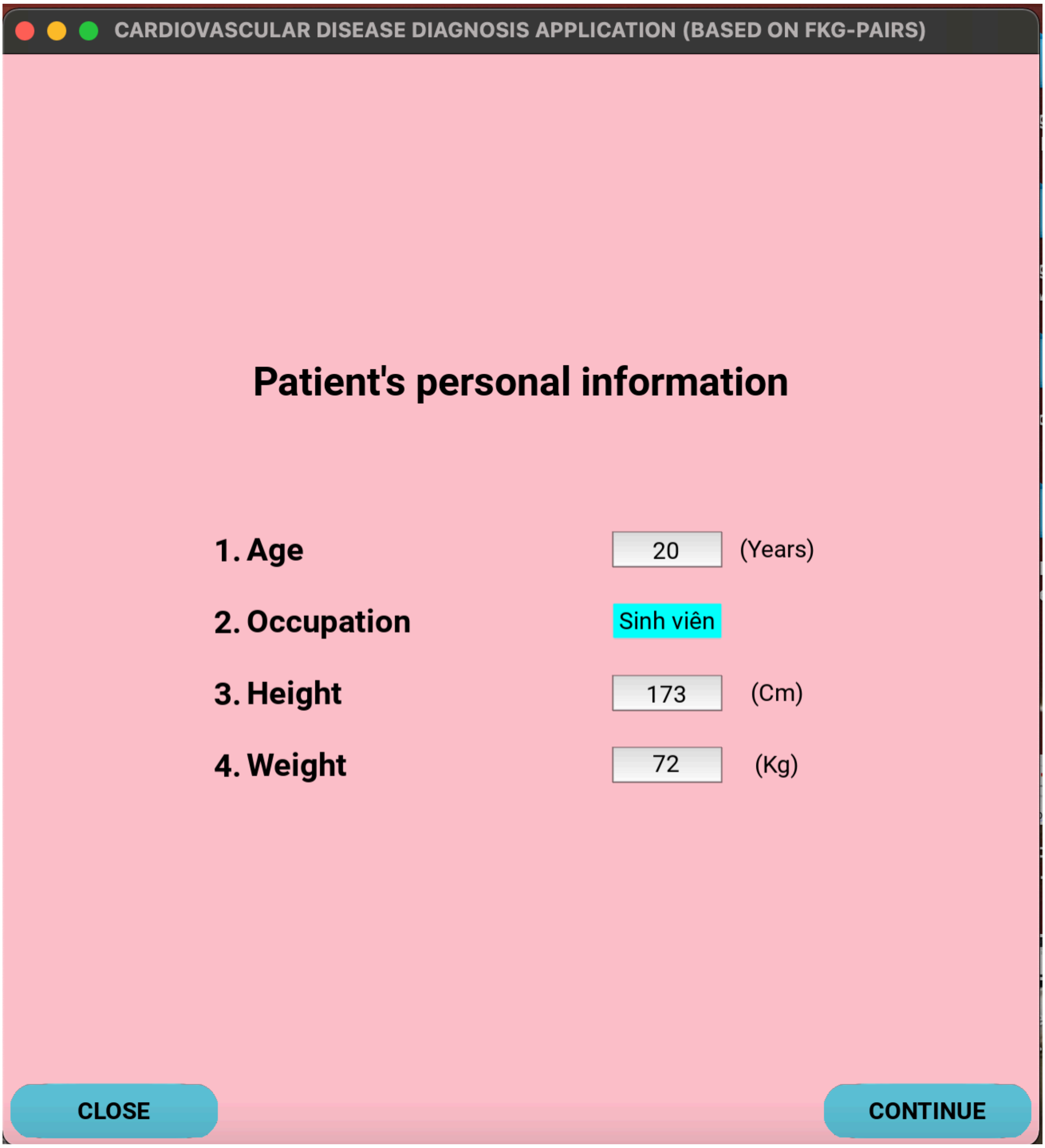

Figure 3: Screen for entering personal information

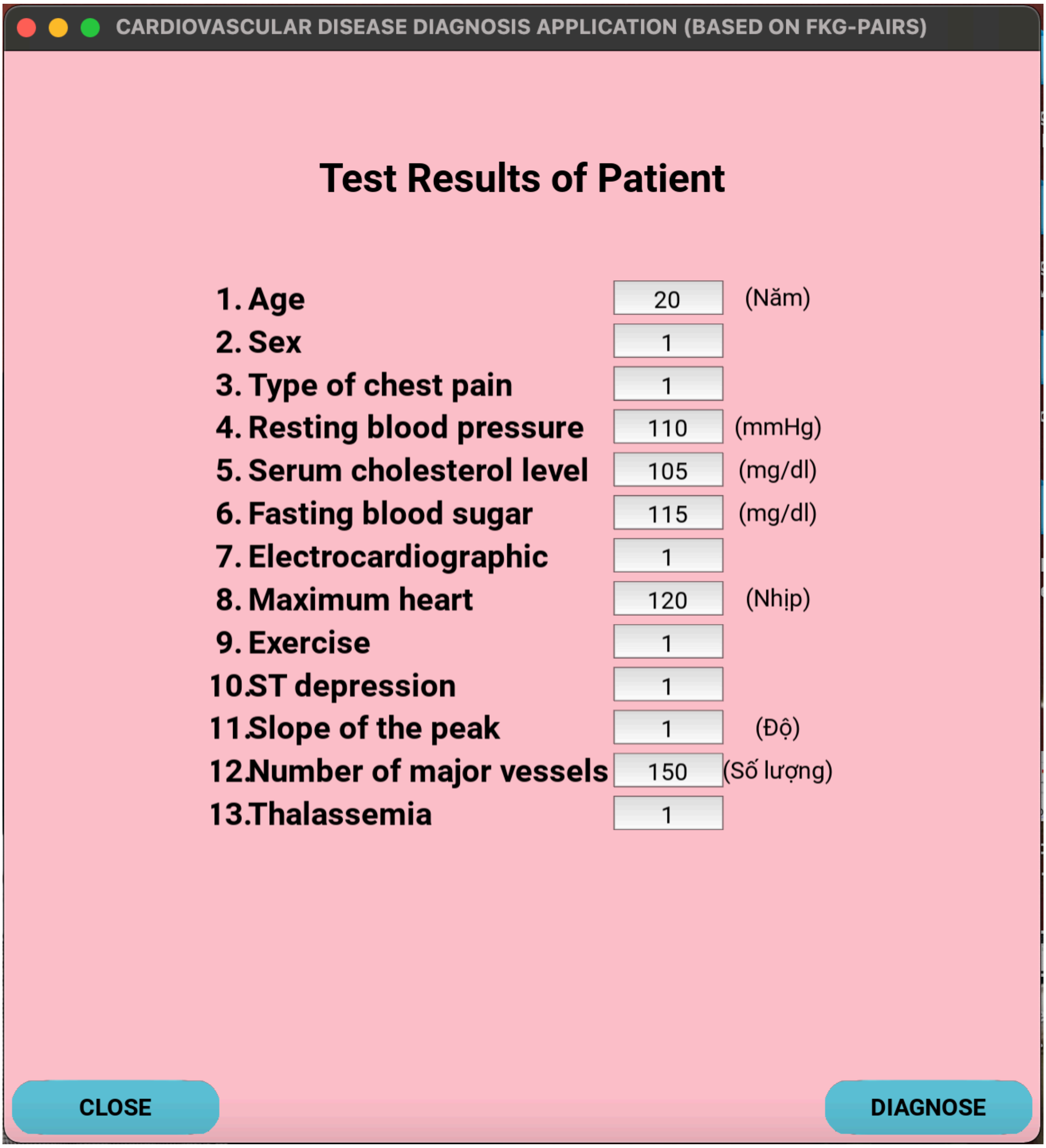

Figure 4: Patient's test result input screen

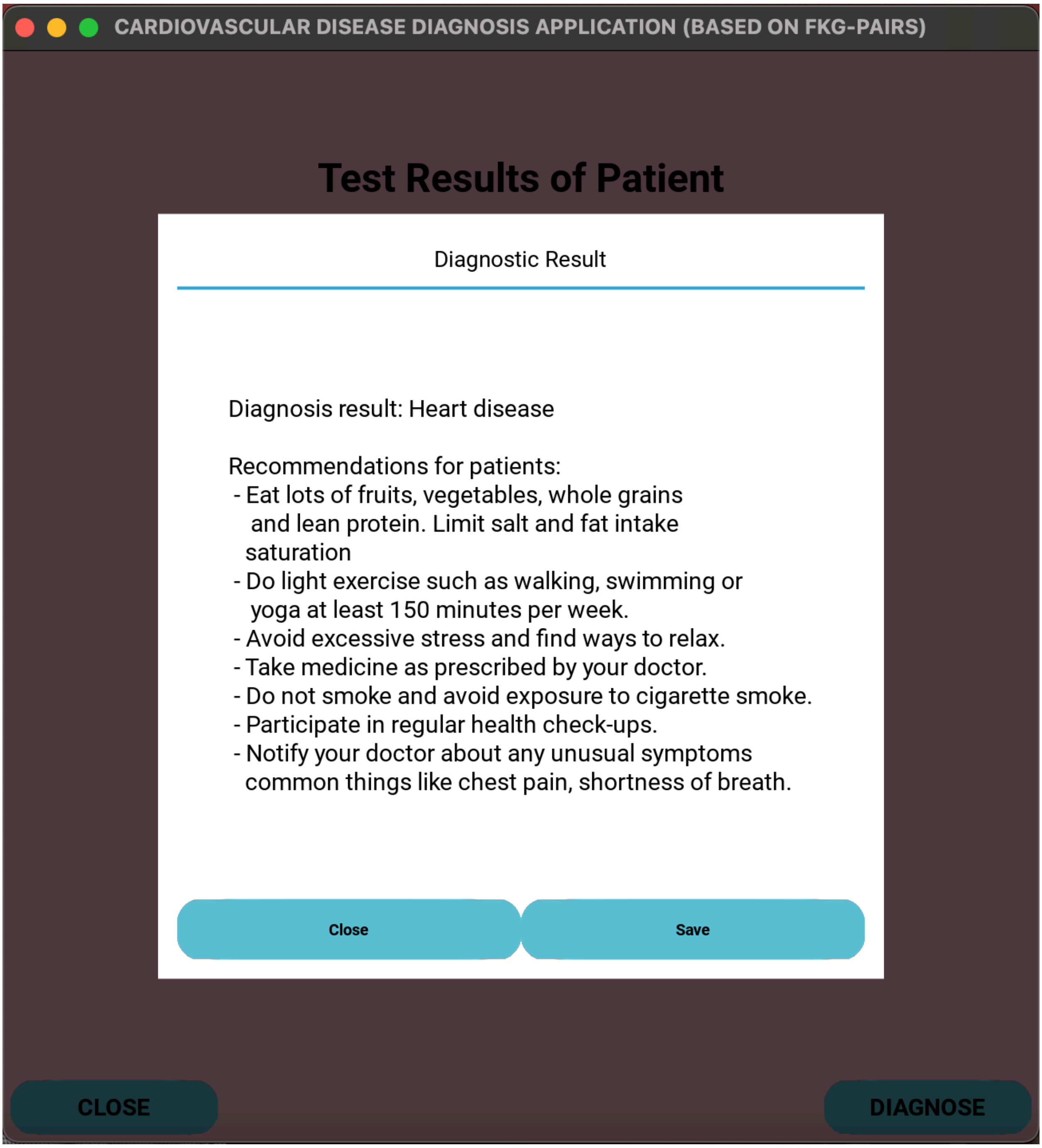

Figure 5: Screen prints out diagnostic results and recommendations

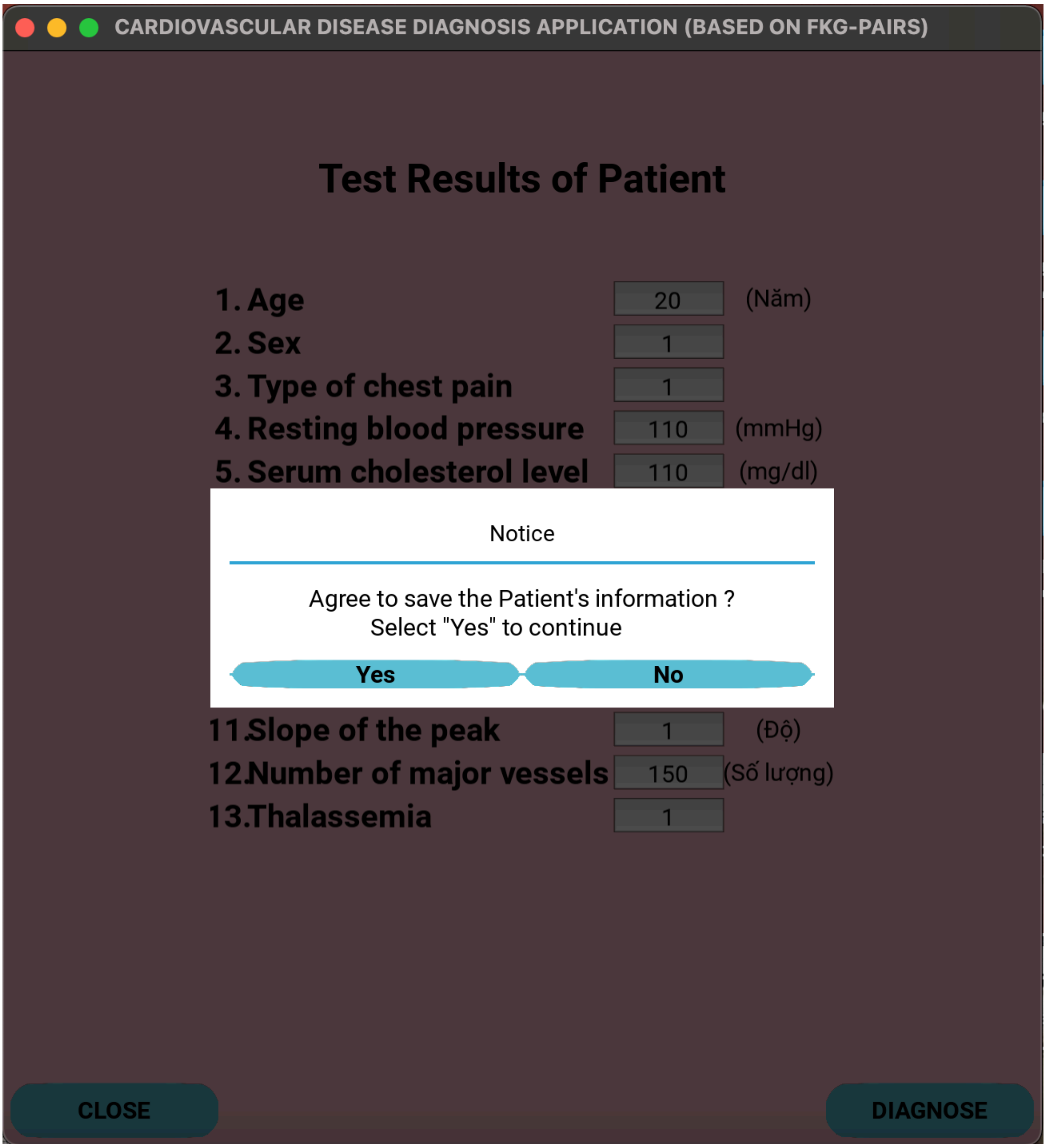

Figure 6: Information saving function

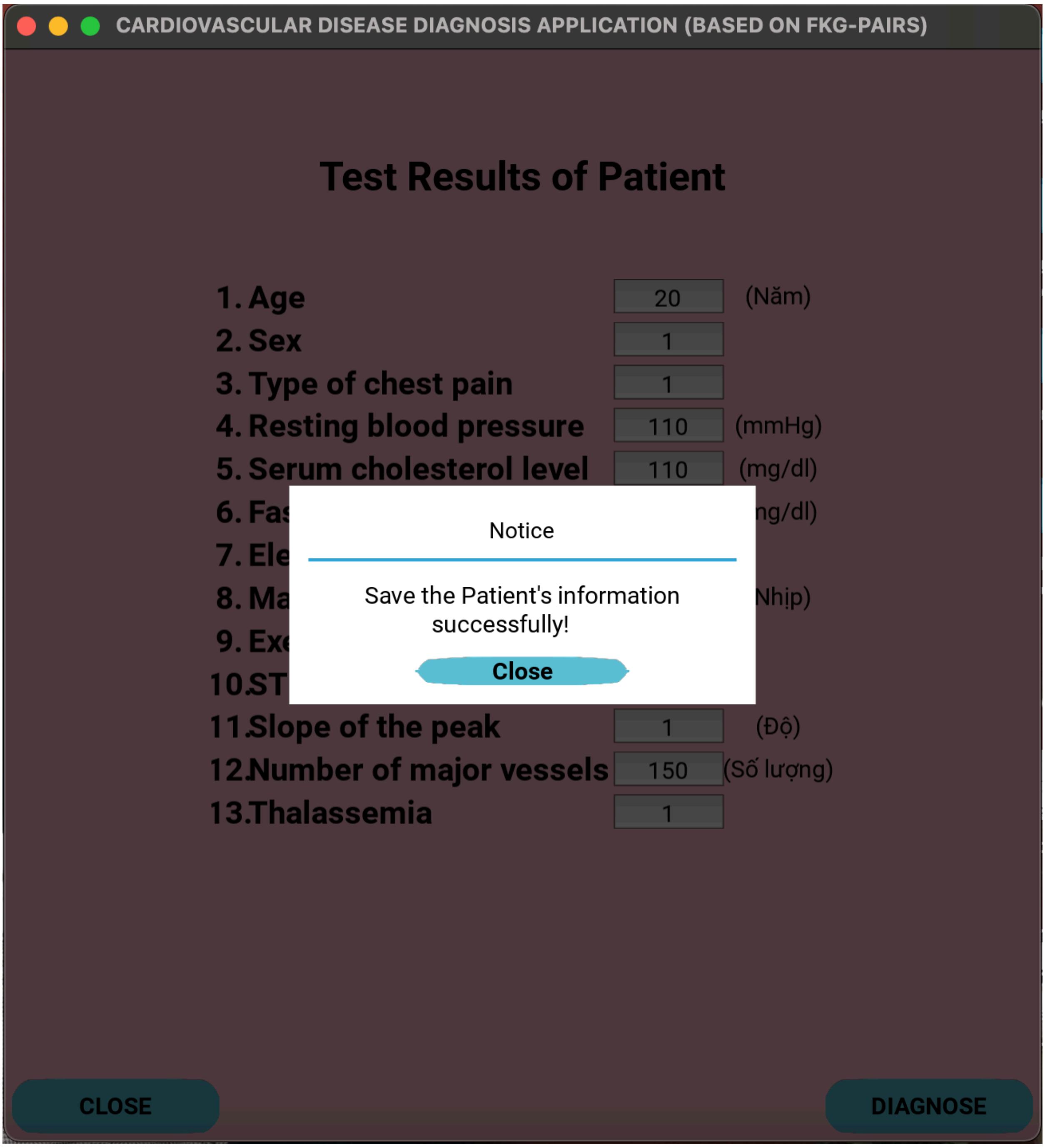

Figure 7: The screen displays information saved successfully

9. Conclude

In this report, I presented the process of researching, understanding, and designing an application to support heart disease diagnosis based on pairwise fuzzy knowledge graphs. Through the implementation process, I used Mathematical and experimental methods to evaluate 3 pairwise fuzzy knowledge models. At the same time, I also learned about new knowledge related to knowledge value graphs, fuzzy inference systems, as well as improving knowledge. Improve programming and application design skills.

10. References

[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.Pham , 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] *Pham Minh Chuan , Tran Manh Tuan , Cu Kim Long , Nguyen Hong Tan (2023) APPLICATION OF FuzzY KNOWLEDGE GRAPH IN SUPPORTING DIAGNOSIS FOR PATIENTS WITH DIABETES.* 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 D ejun , 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