Developing the New Proposal of Intelligence System in Heart and Cardiovascular Diagnosis

Hieu Le Ngoc
HCMC Open University
PhD. Student of Posts and Telecommunications Institute of Technology, HCMC Branch
Ho Chi Minh City, Vietnam
hieu.ln@ou.edu.vn

Hung Tran Cong
Posts and Telecommunications Institute of Technology, HCMC Branch
Ho Chi Minh City, Vietnam
conghung@ptithcm.edu.vn

Sang Do Phuoc
Department of Information Technology
Hanoi University of Home Affairs
Ho Chi Minh City, Vietnam
phuocsangpro@gmail.com

Khue Bui Dan
Department of Information Technology
HCMC Open University
Ho Chi Minh City, Vietnam
buikhue0959@gmail.com

***Abstract*—The World Health Organization estimates that annually, there are 17.5 million deaths due to cardiovascular-related diseases and the number of patients accumulates. This study aims to apply some AI technologies to the system of cardiovascular diagnosis, to provide a solution for building information channels so that people have an overview of cardiovascular pathology. With this approach, people can have self-diagnosis, classification of cardiovascular pathology and know how to plan the examination and treatment to achieve better outcomes. In this paper, the system is built by using reasoning algorithm combined with supervised learning to carry the diagnosis process through the collection of symptoms. The study was experimented on data collected directly from the medical records at General Hospital of Binh Duong Province, and the experimental results show that the accuracy of diagnosis is over 75% compared to the actual diagnostic results.**

Keywords— Cardiovascular diagnosis, Expert system, Supervised Learning, Diagnosis of the disease, Artificial intelligence.

I. INTRODUCTION

Cardiovascular pathology[1, 2] is the leading cause of death in the world, even in developed or developing countries. It is said that Cardiovascular pathology only occur in adults, but in fact the frequency of disease in young and middle age is higher. It can happen to anyone, at any age and the age of cardiovascular pathology is also increasingly rejuvenating. In Vietnam, 25% of the population has cardiovascular disease and hypertension. The rate of hypertension at the age of 25 and older in Vietnam was 47%. People with hypertension or most of cardiovascular complication are not aware of this. Deaths from heart failure, arrhythmia, coronary artery obstruction etc., very popular, this is an alarm. However, many Vietnamese have insufficient knowledge to understand the risks caused by cardiovascular disease. Therefore, it is necessary to provide the people with a channel of information including the basic knowledge about cardiovascular pathology.

In parallel, artificial intelligence (AI) [3, 4, 5] is increasingly evolving in all areas, including application in medical fields. The expert system is also one of the artificial intelligence applications. The most common form of expert system is a program consisting of a set of information analysis rules (often provided by system expert users) file about a specific class of issues, as well as giving analysis of those issues, and depending on the program design, gives advice on the sequence of action to be taken to solve the problem. This is a system that uses the reasoning abilities to reach the most appropriate conclusion. This proposed is built by forward reasoning algorithm combined with supervised learning. Forward reasoning (or forward chaining) is one of the two main methods of reasoning when using an inference engine and can be described logically as repeated application of modus ponens. Forward reasoning is a popular implementation strategy for expert system, business and production rule systems. Forward chaining starts with the available data and uses inference rules to extract more data (from an end user, for example) until a goal is reached. An inference engine using forward chaining searches the inference rules until it finds one where the antecedent (If clause) is known to be true. When such a rule is found, the engine can conclude, or infer, the consequent (Then clause), resulting in the addition of new information to its data. Inference engines will iterate through this process until a goal is reached. Another popular AI technique, Supervised learning [6] is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from *labeled training data* consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way (see inductive bias).

With the topic *"Supporting system for cardiovascular diagnosis"*, we study on the field of data mining to build a decision support system. The expert system is used to support diagnosis for cardiovascular pathologies. This study will help people to get an overview on cardiovascular pathologies, self-diagnosis and understand the classification of cardiovascular pathologies in order to plan examination, to treat early and to achieve better results.

978-1-7281-5163-2/19/$31.00 ©2019 IEEE

This article is organized as follows. Section 1: Introduce about Expert system, Supervised learning and forward reasoning; Section 2: Introduce some related works; Section 3: proposed work using Supporting system for cardiovascular pathology diagnosis; Section 4: Make some experiment; Section 5: General conclusions of the research topic.

II. RELATED WORKS

Recently, there have been many researches in the field of data mining, decision support systems and their applications in diagnosis and treatment support system. In 2000 - Expert: An Expert System for the Diagnosis of Sleep Disorders [10] is Fred, Filip, Partinen and Paiva studied based on cadastral data, system expansion motor relay is provided by CLIPS and support chain backwards. Naser and Mahdi in 2016, proposed the design of the proposed Expert System which was produced to help Podiatric physician in diagnosing many of the foot diseases [7] (Rheumatoid Arthritis, Gout, Heel Spur, Athlete's Foot, Morton's Neuroma, Cellulitis, Frostbite, Gangrene) are presented, an overview about the foot diseases are displayed, the cause of diseases are outlined and the treatment of disease whenever possible is given. SL5 Object Expert System language is used for designing and implementing the proposed expert system. In 2019, Mallika and UshaRani [8] had developed the fuzzy-based expert system to diagnose Alzheimer's disease. Bruce G. Buchanan applied research MYCIN system support diagnosis [9]. A detailed look at MYCIN, an expert system for diagnosing bacterial infections and prescribing treatment for them. Issues covered include detailed examinations of knowledge acquisition, reasoning, explanation, tutoring, performance evaluation and human interface.

Recent years in Vietnam, according to statistics at some universities in the fields of data mining, building a decision support system in the health can be mentioned as: Develop a support system for diagnosis decisions [11] written by Thanh Van The and Bao Tran Minh, the process of building this decision support system is based on a combination of fuzzy set theory, hedge algebra and fuzzy deductive methods to build a decision support system about Process of diagnosis through cardiovascular ultrasound data (2012); Decision support system in diagnosis: approach from fuzzy system [12] written by Lan Luong Thi Hong and Ngan Tran Thi, presents a model of building a decision support system in diagnosing the disease with a fuzzy inference approach, we also set up the proposed model on the liver disease data set.

III. INTELLIGENT SYSTEM OF CARDIOVASCULAR PATHOLOGY DIAGNOSIS

In this section, we present about building the diagnostic system. From the data gathered from the experts, specialized curriculum, references especially the data collected directly from Medical Records on a practical basic at Binh Duong General Hospital Province, we propose an Intelligence system for cardiovascular pathology diagnosis using forward reasoning algorithm combined with supervised learning.

A. *The intelligence system of cardiovascular pathology diagnosis*

The system was built for diagnostic purposes and related guidelines for the diagnosis and treatment of cardiovascular disease, so the system will ask the user for the symptoms then infer from those symptoms in the rules combined with supervised learning from sample data set to give conclusions about user's condition of illness.

The problem only applies to looking up some cardiovascular diseases including specific diseases collected from experts, references especially the data collected directly from Medical Records on a practical basic at Binh Duong General Hospital Province from June 2018 to December 2018.

Input data are disease symptoms that users import. The output data is disease conclusion and some advice.

1) *The proposed Algorithm:*

Diagnosis is based on the symptoms that the user declared conjunct with supervised learning, so forward reasoning algorithm and supervised learning algorithms were applied in research are as follows:

- Input:
    + Rule set = {r1, r2, …, rn}.
    + Set of declare user symptoms (GT ~ Theory).
    + Set of typical sample medical data (about 300 medical records) at provincial General Hospital.
- Output:
    + Announce that "you may be ill" + name of the illness + percentage probability (%) compared to sample data.
    + On the contrary, the message will be "We haven't detected your illness yet or you have declared insufficient symptoms!"

- Function **DiseaseDiagnoseByRule** (Reasoning)

{
TD:= GT;
T:= Fillter(Rule, TD);
While (Could not find the pathology $\not\subset$ TD) and (T<>∅) do
{
r := get(T);
TD:= TD + right(r);
Rule:=Rule \ {r};
T:= Fillter(Rule,TD);
}
If [Disease name] $\subset$ TD then
*Return **[Disease name]***
*Else Return **NULL***
}

- Function **DiseaseDiagnoseSL** (Supervised learning method)

{
TC:= GT;
T:= SupervisedLearning();
if [T] $\subset$ TD then
*Return **[Disease name]+ [probability accuracy in percent];***
*Else Return **NULL;***
}

- Function combined 2 functions above: **supervised learning and forward reasoning**.

{
T1 := DiseaseDiagnoseRule();

T2 := DiseaseDiagnoseSL();
If [T1] ∩[T2] ≠ Ø then
Return
"You may get sick" + [T1]
"You may get sick" + [T2] + probability accuracy in percent;
Else IF [T1] ≠ NULL then Return
"You may get sick" + [T1]
Else IF [T2] ≠ NULL then Return
"You may get sick" + [T2] + probability accuracy in percent;
Else
Return ***"The system hasn't found your disease";***
}

2) The modules of proposed algorithm:

Based on previous studies of Diagnostic support systems and supervised learning methods in machine learning, this thesis proposes the application of 5 main module groups:

(1) Module read and save data: symptoms, illness, advice and medical records

This module will perform the task of reading the necessary data and put into the database of the application: patient symptoms data, disease group data, symptom group.

Knowledge Base data: sample medical data, rules of diseases.

The sample data is sample taken out from the cardiovascular disease records treated at the Cardiology Department and the number of records is about 300 records.

(2) Diagnosis module by reasoning (forward - backward)

In this module will use forward reasoning and backward reasoning to infer disease from rules designed in the knowledge base, in addition also combine to the nature of the rule will have the following conclusion: 50% get sick; 100% get sick; or Unsure about the token of the disease;

(3) Diagnosis module by Supervised Learning

This module will use supervised learning algorithm (SVM or KNN, or other algorithms of supervised learning) to predict disease from input symptoms on the training dataset is sample of medical records stored in the database, in addition it is combined with the accuracy of the methods that can be concluded about diseases like: *p%* get sick; *100%* get sick;

(4) Module combining 2 diagnostic methods: forward reasoning and supervised learning

This module will incorporate the result of Module (2) and (3), using the method of extracting common characteristics from 02 results, or 01 of 02 results if either of the two results is empty. This Module will return corresponding disease result.

(5) Module giving conclusion and advice

This module will combine the result of Module (4) and query to disease data, symptoms, advice to draw conclusions and advice accordingly.

3) Algorithm description diagram:

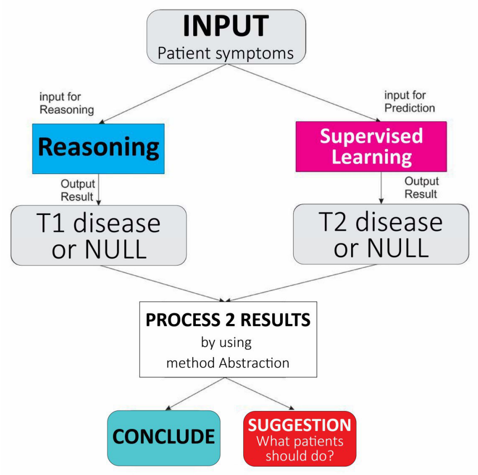

Fig. 1. Algorithm description diagram.

B. Design the knowledge-base and database

1) Data collection:

From expert: This is the beginning of the process of building an expert system, first collect the knowledge from the experts. Then, from the knowledge collected, analyzed, extracted the necessary information for the topic and taken the form can be stored on the database computer.

From medical records: Through the guidance of doctors, medical staff, this research has collected data directly from medical records.

Collected data is sample data from medical records, a rule book is a list of diseases with corresponding symptoms of disease and instructions attached to the disease (if available).

2) Database:

After analysis and classification knowledge from source data collected, build database to put data into the computer. Build database on the management system database MSSQL (SQL Server), ding the following tables:

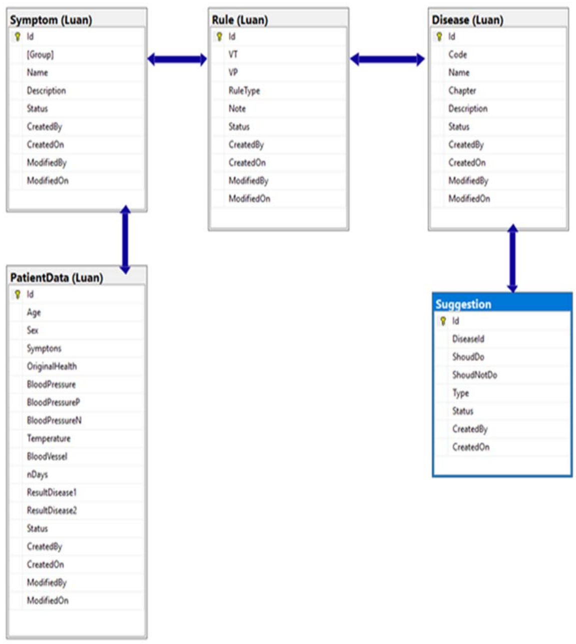

Fig. 2. The relationship between data tables.

3) Knowledge-base:

The knowledge base is described through the Rule, the rules are designed as below:

Column Name	Data Type	Allow Nulls
Id	int	☐
VT	varchar(MAX)	☑
VP	varchar(MAX)	☑
RuleType	smallint	☑
Note	nvarchar(MAX)	☑
Status	smallint	☑
CreatedBy	int	☑
CreatedOn	datetime	☑
ModifiedBy	int	☑
ModifiedOn	datetime	☑
		☐

Fig. 3. The rule database design.

	Id	VT	VP	RuleType	Note	Status	CreatedBy	CreatedOn	ModifiedBy	ModifiedOn
1	1	{118},{119},{120}	I05.0	1	Hẹp van hai lá	1	1	2019-05-22 09:42:34.670	NULL	NULL
2	2	{65},{118},{119}	I05.1	1	Hở van hai lá	1	1	2019-05-22 09:42:34.670	NULL	NULL
3	3	{47},{92},{83}	I10	1	Tăng huyết áp vô căn (nguyên phát)	1	1	2019-05-22 09:42:34.670	NULL	NULL
4	4	{94},{93},{97},{96},{95}	I20	1	Cơn đau thắt ngực ổn định	1	1	2019-05-22 09:42:34.670	NULL	NULL
5	5	{99},{93},{123}	I20.0	1	Cơn đau thắt ngực không ổn định	1	1	2019-05-22 09:42:34.670	NULL	NULL
6	6	{67},{93},{101},{102}	I21	1	Nhồi máu cơ tim cấp	1	1	2019-05-22 09:42:34.670	NULL	NULL
7	7	{93},{106},{100},{101},{102}	I21	1	Nhồi máu cơ tim cấp	1	1	2019-05-22 09:42:34.670	NULL	NULL
8	8	{94},{93},{97},{104},{95}	I25	1	Bệnh tim thiếu máu cục bộ mạn	1	1	2019-05-22 09:42:34.670	NULL	NULL
9	9	{67},{64},{93},{113},{117}	I47	1	Nhịp nhanh kịch phát	1	1	2019-05-22 09:42:34.670	NULL	NULL
10	10	{92},{104},{117},{6},{7},{91}	I50	1	Suy tim	1	1	2019-05-22 09:42:34.670	NULL	NULL
11	11	{117},{128},{129},{68}	J81	1	Phù phổi	1	1	2019-05-22 09:42:34.670	NULL	NULL
12	12	{99},{133},{83},{8},{63}	I64	1	Đột quị, không xác định do xuất huyết hay nhồi ...	1	1	2019-05-22 09:42:34.670	NULL	NULL
13	13	{10},{11}	I83	1	Dãn tĩnh mạch chi dưới	1	1	2019-05-22 09:42:34.670	NULL	NULL
14	14	{16}, {17}	I82.2	1	Thuyên tắc và huyết khối tĩnh mạch chủ	1	1	2019-05-22 09:42:34.670	NULL	NULL
15	15	{6},{7}, {17}, {21}	I87.2	1	Suy tĩnh mạch (mạn) (ngoại biên)	1	1	2019-05-22 09:42:34.670	NULL	NULL
16	16	{64}, {68}, {117}, {86}	I48	1	Rung nhĩ và cuồng nhĩ	1	1	2019-05-22 09:42:34.670	NULL	NULL
17	17	{92}, {68}, {117}, {74}	I49.0	1	Rung thất và cuồng thất	1	1	2019-05-22 09:42:34.670	NULL	NULL
18	18	{74}	I49.0	1	Rung thất và cuồng thất	1	1	2019-05-22 09:42:34.670	NULL	NULL
19	19	{67}, {93}	I47	1	Nhịp nhanh kịch phát	1	1	2019-05-22 09:42:34.670	NULL	NULL
20	20	{118},{120}	I05.0	1	Hẹp van hai lá	1	1	2019-05-22 09:42:34.670	NULL	NULL
21	21	{65},{118}	I05.1	1	Hở van hai lá	1	1	2019-05-22 09:42:34.670	NULL	NULL
22	22	{94},{93},{95}	I20	1	Cơn đau thắt ngực ổn định	1	1	2019-05-22 09:42:34.670	NULL	NULL
23	23	{99},{93}	I20.0	1	Cơn đau thắt ngực không ổn định	1	1	2019-05-22 09:42:34.670	NULL	NULL
24	24	{67},{93},{102}	I21	1	Nhồi máu cơ tim cấp	1	1	2019-05-22 09:42:34.670	NULL	NULL
25	52	{5},{6},{2},{89},{88}	I05.0	2	Hẹp van hai lá	1	1	2019-05-22 09:42:48.060	NULL	NULL
26	53	{108},{109},{122},{88},{78}	I05.1	2	Hở van hai lá	1	1	2019-05-22 09:42:48.060	NULL	NULL
27	54	{78},{108},{82},{56},{117},...	I10	2	Tăng huyết áp vô căn (nguyên phát)	1	1	2019-05-22 09:42:48.060	NULL	NULL
28	55	{117},{82}, {83},{117},{78}...	I20	2	Cơn đau thắt ngực ổn định	1	1	2019-05-22 09:42:48.060	NULL	NULL

Fig. 4. List of rules.

The list of rules was collected from experts who are doctors, specialized medical staff of the hospital, including 47 rules, in it are 24 of the rule has 100% accuracy and the rules remaining accurate of over 50%.

IV. EXPERIMENTS RESULTS

A. Experiment introduction

Simulated installation of module 2 and module 3 of the DOTNET-based diagnostic support system, in which module 3 uses ML-Accord library integrating in DOTNET to perform.

The data of the simulation test comes from the training data source which is typical case data from Binh Duong General Hospital. This data is taken from medical records of common cardiovascular diseases in the hospital through the guidance of doctors, health professionals. In this simulation, the study uses the BAYES NAÏVE classification algorithm, predicts the disease from the input symptoms, then extract combined with the expert system through the rule: conclusion of disease, percentage exactly, tips correspond to disease.

- *Disease and disease information data:* ICD10 International Classification of Diseases table is a list of diseases prescribed by the Ministry of Health under Decision 3970 / QD-BYT dated September 24, 2015, including 22.766 data lines, each line is the diseases and the corresponding disease code.
- *Sample medical data:* Is a typical medical record from General Hospital of Binh Duong Province, including 296 records each line is set of symptoms, conclusion of the disease (main disease, minor disease), medical history, vessels, blood pressure, sex, year of birth, treatment results, discharge status, the number of day of treatment.
- *Symptom data:* Table of symptoms are collected, analyzed, extracted from experts who are cardiologists, medical staff and specialized document. Including 131 data lines, each line is symptom type, symptom group, symptom name, description of symptom.
- *Rule data:* As the rule was collected, analyzed, extracted from experts who are cardiologists, medical staff, specialized documents, including 47 data lines, each line includes the code, disease name, symptom set always available, symptom set possible.

B. Results

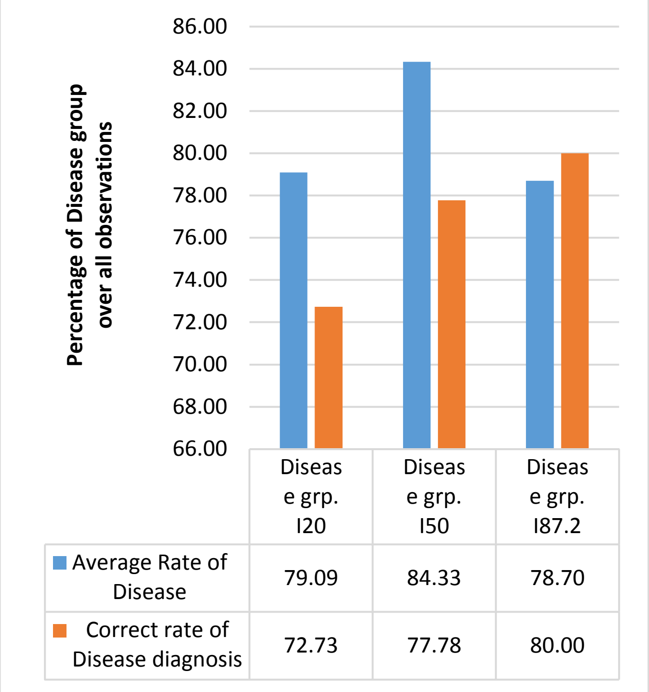

Fig. 5. Empirical results for 3 groups of diseases.

After running the program, retrieve the results that were just run against the actual sample data on the medical record, as follows:

- Empirical results with the group of disease I20:
  + Average incidence: 79.09%
  + The ratio of true compared to reality: 72.7%
- Empirical results with the group of disease I50:
  + Average incidence: 84.33%
  + The ratio of true compared to reality: 77.7%
- Empirical results with the group of disease I87.2:
  + Average incidence: 78.7%
  + The ratio of true compared to reality: 80%

From a review of 30 sample records of 3 groups of diseases (I20, I50 and I87.2) collected directly from the hospital record shows the incidence of each group of rule. In the group of I50 disease there is the highest average incidence, but the group I87.2 has the highest correct diagnostic rate.

V. CONCLUSION

In this paper, we propose a system that is built by using forward reasoning algorithm combined with supervised learning to carry the diagnosis process through the collection of symptoms. Of the 30 test records, the program runs out the diagnostic result coincides with the fact that 23 records correctly reached the rate of 77%. Although this is an experimental data, it has also been revealed that the program results in a relatively accurate and acceptable. With these initial diagnostic findings will be the basis for the patient to choose where to treat the disease appropriately. Beside we also found that this system has some disadvantages as the information was scanty and due to the constraint on time and resources there is no comparison of the proposed method results to related works. In the future, we continue to complete the functions, update more rules based system, use more diversity information to enhance the quantity of diagnosis. So, it is possible to study the development of this system in some diseases such as lungs, kidney failure, ...

ACKNOWLEDGMENT

Sincerely thank the leader of the Binh Duong General Hospital Province which facilitates the provision of figures for research papers and the international teacher and graduate school of Post & Telecoms Institute of Technology has communicated the knowledge and valuable experience throughout the research process.

REFERENCES

[1] E. Rene Rodriguez, Carmela D. Tan, “Cardiovascular Pathology”, in Essentials of anatomic pathology: Fourth edition, Springer, 2016, pp.1295-1351.

[2] B. Sampson, J. L. Hammers, “Forensic Aspects of Cardiovascular Pathology”, in Cardiovascular Pathology, Springer, 2016, pp.773-798.

[3] Vicenç Torra, H.Joe Steinhauer, “Artificial Intelligence”, in Data Science in Practice, Springer , 2019, pp.9-26.

[4] S. Russell and Peter Norvig, “A Modern Approach” in Artificial Intelligence, Prentice Hall, Englewood Cliffs, NJ, 2019.

[5] Ramjan Shaik, “Artificial Intelligence in Healthcare”, Indian Journal of Pharmacy Practice, 2019, pp.215-216.

[6] Paul D. McNicholas and Peter A. Tait, “Supervised Learning” in Data Science with Julia, Technics Publications, 2019, pp.93-128

[7] Samy S. Abu Naser, Ali Osama Mahdi, “A proposed expert system for foot diseases diagnosis”, the American Journal of Innovative Research and Applied Sciences. ISSN, 2016, 2429-5396.

[8] R. M. Mallika, K. UshaRani, Hemalatha Kulala, “A Fuzzy-Based Expert System to Diagnose Alzheimer’s Disease” in Internet of Things and Personalized Healthcare Systems, 2019, pp.65-74.

[9] Bruce G. Buchanan, “Rule- Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project”, Addiscion-Wesley, 2019, pp.717- 738.

[10] Ana L. N. Fred, Joaquim B Filipe, Markku Partinen and Teresa Paiva, "An Expert System for the Diagnosis of Sleep Disorders," in PSG-Expert, IOS Press 78, 2000, pp. 127-147.

[11] Văn Thế Thành, Trần Minh Bảo, “Xây dựng hệ hỗ trợ ra quyết định chẩn đoán bệnh”, Tạp chí khoa học, Đại học Huế, Việt Nam, 2012, pp.129-139.

[12] Lương Thị Hồng Lan, Trần Thị Ngân, “Hệ hỗ trợ ra quyết định trong chẩn đoán bệnh: tiếp cận từ hệ mờ phức”, Kỷ yếu Hội nghị Quốc gia lần thứ XII về nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Việt Nam, 2019.