A HYBRID INTELLIGENT SYSTEM FOR THE DISEASE RISK EVALUATION

Sheng Jen Jian
Fu Jen Catholic University, Taiwan

Kang-Hong Liou
Fu Jen Catholic University, Taiwan

Ruey Kei Chiu
Fu Jen Catholic University, Taiwan
rkchiu@mail.fju.edu.tw

Abstract:
This paper attempts to develop a hybrid intelligent system for the risk evaluation of coronary heart disease by leveraging the artificial intelligent techniques of combining artificial neural network and fuzzy expert system. The statistical analytical method and the multilayer perceptron of neural network are taken respectively to identify the significant causing factors for the coronary heart disease. The factors are then applied for developing the fuzzy expert system of risk evaluation. Furthermore, the system is deployed to the cloud for being used by the publics. The result of cloud deployment is also evaluated in terms of advantages, and disadvantages. The accuracy rate of this system from case evaluation may reach to 80 percent. Although this figure is acceptable from the viewpoints of experts but it may be improved in order to be used with more reliable in practice. Additionally, the cost, performance, and data security are also concerned and need to be further evaluated in the subsequent study of this paper.

Keywords: knowledge, machine learning, hybrid intelligent system, fuzzy expert system, coronary heart disease, risk evaluatiion

1. INTRODUCTION

According to the statistic figures and information released on (DOH, 2012) and (Framingham Heart Study, 2011), the coronary heart disease (CHD) is one of the most popular heart diseases. The coronary heart disease, also called coronary artery heart disease, is one of the most common types of heart diseases which are detected both in Taiwan and in the world. The major risk factors include men over 45 years and women over 55 years, smoking, high blood pressure, diabetes, obesity, and so on. How to have an early detecting the risk of probably suffering CHD becomes one of the most important issues to a human or the public health authority of a nation

The study of this paper includes four major activities of system implementation. Firstly, the preliminary risk factors to cause CHD are identified through the literature review and the preliminary data analysis by employing a neural network model, Secondly, the statistical method of Factorial ANOVA is applied to filtering out the significant risk factors to cause CHD from the risk factors identified in the first activity. Thirdly, a fuzzy expert system (Zahan, 2001, pp. 271-275; Negnevitsky, 2011, pp. 87-127) is developed to assess the severity of the risk of coronary heart disease. The final part of this paper is to describe the deployment of the system developed in this study to the cloud by leveraging Google cloud platform. Consequently, the publics may take a self-evaluation to his/her risk severity of contracting CHD. Meanwhile, the physician of CHD may also take the system generated to assess the patient's heart health so that a proper treatment to each individual patient may be provided.

2. CORONARY HEART DISEASE

Coronary heart disease refers to that the supply of coronary blood could not meet the required blood of metabolism resulted in the occurrence of myocardial ischemia or infarction. Coronary artery disease also refers to sclerosis of coronary disease. When the coronary artery wall cholesterol accumulation, it will stimulate the destruction of the vascular wall resulted in impaired blood vessel wall to form a clot. Generally speaking, coronary heart disease is the disease caused by the coronary blood supply to the heart when it is partly or completely blocked. The risk factors for coronary heart disease can be divided into many facets. Some are caused by natural, but some are paid by more attention and prevention. Some risk factors such as gender can't be other than force control. However, some risk factors are different with the habits of life, a different diet or the effect of environment. Epidemiological studies have shown that the related factors of causing cardiovascular disease (CVD) include obesity, smoking, drinking, low-density lipoprotein cholesterol (LDL), high density lipoprotein cholesterol (HDL) and total cholesterol (TC). Foreign research institutions and the Framingham heart study (Framingham heart study, 2011) have identified the major risk factors including high cholesterol, hypertension, smoking, obesity, diabetes, and so on.

3. CLOUD COMPUTING

The definition of Cloud Computing presented by the National Institute of Standards and Technology (NIST) indicates that cloud computing is a new computing paradigm for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell&Grance, 2012). The services of cloud computing consist of three service models including (1) Infrastructure as a Service (IaaS). The well-known IaaS providers are such as Google, Amazon, and IBM, and so on; (2) Platform as a Service (PaaS). The typical PaaS providers are such as Microsoft Windows Azure and Google App Engine (GAE) (Google, 2012); and (3) Software as a Service (SaaS). ypical examples of well-known SaaS providers and its products are such as Salesforce''s CRM, Google's Google Docs, Gmail, Google Map and Microsoft's Live Mail and health vault. The most well-known social network web sites such as Facebook, Twitter, Line, YouTube are thought as another kinds of service Software over the cloud.

The system developed is also deployed on Google cloud platforms for the use by publics. Therefore, in this paper the advantages and disadvantages of the popular cloud service platform is firstly analyzed followed by the selection of an appropriate cloud platform for subsequent system development and deployment. The main facets of the consideration on the choice of cloud platform for this paper are the stability, cost, and subsequent updates of system development. As a result, the Google cloud platform is chosen for this system development due to its advantages over other peer rivalry in terms of its unique system development engine, data management such as Google file

system (GFS) and distributed and parallel computing facilities and mechanism MapReduce (Dean&Ghemawat, 2004).

4. RESEARCH MATERIALS AND METHODS

In this paper, the domain knowledge of coronary heart disease is acquired from the literature review and the survey from professional experts. The Factorial ANOVA analysis is firstly applied to identify the preliminary set of risk factors for causing coronary heart disease. The fuzzy rules for building the fuzzy expert system is then implemented based on the knowledge acquired from the domain experts and risk factors identified.

4.1. The Determination of Risk Factors

We firstly identify 14 risk factors from the original data taken from the data set of UCI database Cleveland heart disease which consists of 303 case records. This is the original set of risk factors which are determined. This data file again inputs to ANOVA analysis of statistics method by using R statistical software (Software R, 2012) for multi-factor analysis of variance to obtain most significant risk factors in order to provide follow-up study. There are ten significant factors remain from ANOVA analysis. These ten factors are age, sex, chest, pain type, trestbps (the resting blood pressure), resting ECG, max heart rate, oldpeak, exercise induced angina, number of vessels colored and Thal. The set of 10 risk factors is called as the screening set of risk factors.

4.2. Simulation and Modelling

In order to obtain the most significant set of risk factors to be used as the input variables for developing fuzzy expert system, the neural network model is applied for the further verification. The original set of 14 preliminary risk factors initially identified from the data set and the set of 10 risk factors screened from the ANOVA analysis are respectively taken as the input for the verification of neural network model.

The NeuroSolutions 5.0 (NeuroSolutions, 2012) which is a well-known Neural Network software tool is employed to construct a multi-layer neural network model (Negnevitsky, 2011, pp. 165-216) for this study. The learning process of network model development uses K-fold cross validation network training and testing. Confusion Matrix presented in Table 1 shows accuracy, sensitivity and specificity in order to assess the differences in the model experiment and simulation by taking two different sets of risk factors as the input for the model training and testing in developing neural networks. The results showed that the set of risk factors after the screening may produce a higher accuracy rate of 0.9824 than that of 0.9794 which is produced by the original set of risk factors. Hence, the set of 10 factors from the result of ANOVA analysis shows a more proper set. In order to simplify the model building for the system, the 10 risk factors are further screened to 7 risk factors by consulting from professional experts of CHD. These 7 risk factors shown in Table 2 are then determined as the input variables (also called as fuzzy variables) for the development of fuzzy expert system.

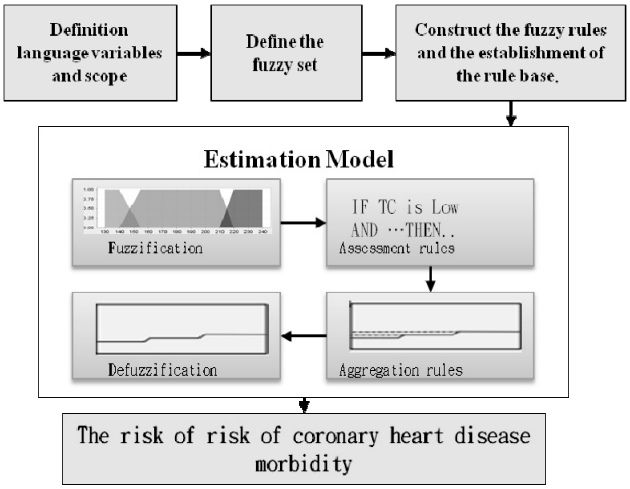
Table 1: Confusion matrix

		Accuracy	Sensitivity	Specificity
The set of risk factors After the screening	Mean	0.9824	0.9922	0.9726
	St. Dev	0.0039	0.0063	0.0084
The original set of risk factors	Mean	0.9794	0.9898	0.9689
	St. Dev	0.0067	0.0036	0.0124

4.3. System Implementation

We use jFuzzyLogic (jFuzzyLogic, 2014) and JESS (JESS, 2014) fuzzy expert software tools respectively to develop this system. The system development process is shown in Picture 1.

Picture 1: Fuzzy expert system development process



The fuzzy sets Zadeh, 1965, pp. 338-353) with respect to their respective fuzzy range for each of 7 risk factors (i.e., fuzzy variables) are shown in Table 2.

Table 2: The fuzzy set of CHD risk factors

Fuzzy Variables	Fuzzy sets	Fuzzy range
Age	Young	<30years
	Summer	25-45 years
	Middle	40-65 years
	Old	>60years
Diabetes	True	1
	False	0
Smoking	True	1
	False	0
Blood Pressure (Systolic)	Low	<129 mm/hg
	Middle	125-139 mm/hg
	Rather High	135-159 mm/hg
	High	>155 mm/hg
Total Cholesterol	Middle	<239 mg/dl
	Rather High	235-279 mg/dl
	High	>275 mg/dl
LDL Cholesterol	Middle	<159 mm/hg
	Rather High	155-190 mm/hg
	High	>185 mm/hg
Risk ratio(Output)	Very Low	<20%
	Low	15-30%
	Rather Low	25-40%
	Middle	35-50%
	Rather High	45-60%
	High	55-70%
	Very High	>65%

The fuzzy rules for the risk assessment established by the system are shown in Table 3.

Table 3: Fuzzy Rules for ChD

Rules	Inputs				Output
	Age	Diabetes	…	LDL	Risk
Rule 1	Young	False	…	Middle	Very Low
Rule 2	Young	False	…	Rather High	Very Low
Rule 3	Young	False	…	High	Low
…	…	…	…	…	…
Rule 432	Old	True	…	High	Very High

The knowledge about these rules are solicited from professional experts of local collaborative hospital, the documents and reports regarding to the risk assessment of CHD. Through proper Fuzzy inference, the risk level of getting the CHD for each testing case may thus be measured. The user interface of using this system for conducting the risk assessment of CHD is illustrated in Picture 2. After the user assessing the risk of CHD function, the system will provide relevant health education and suggestions of overcoming condition worsens according to their degree of risk to help users to refer. The outcome, action Plans and suggestions for a Risk Assessment is shown in Picture 3. The summarized action plans for all risk levels of CHD are shown in Table 4.

Picture 2: The user interface

Risk Assessment Tool for Estimating Your Coronary Heart Risk

The risk assessment tool below uses information from the Framingham Heart Study to predict a person chance of having a Coronary Heart Disease.

Age :

Diabetes : yes no

Smoking : yes no

Systolic Blood Pressure : **mm/Hg**

Total Cholesterol : **mg/dL**

LDLl Cholesterol : **mg/dL**

submit

Picture 3: Action plans and suggestions for a risk assessment

Risk Assessment Tool for Estimating Your Coronary Heart Risk

The risk assessment tool below uses information from the Framingham Heart Study to predict a person chance of having a Coronary Heart Disease.

back to Risk Assessment
Age : 73.0
Diabetes : 1.0
Smoking : 1.0
Systolic Blood Pressure : 130.0 mm/Hg
Total Cholesterol : 240.0 mg/dL
LDL Cholesterol : 190.0 mg/dL
Your CHD Risk is : 52%
The morbidity risk the possibility of: (High)
1. Quit smoking, control high cholesterol, high-fat foods, eat more Fruit and vegetables
2. Maintaining a good mood., more exercise and consult a doctor.
3. Note that blood pressure and blood sugar.
Systolic blood pressure normal range 90-140 mm / Hg, and before meals and blood sugar normal range 70 - 100 mg/dl

Table 4: Risks and action plan

Risk		Action plans
The morbidity risk the possibility of: (Low)	25-40%	1. Less smoking, weight control and more exercise. (smoking)
		2. Weight control and more exercise. (no-smoking)
		3. Eating habits of more attention and adequate sleep.
The morbidity risk the possibility of: (Medium)	41-50%	1. Quit smoking, control high cholesterol, high-fat foods, eat more Fruit and vegetables. (smoking)
		2. Control high cholesterol, high-fat foods, eat more Fruit and vegetables (no-smoking)
		3. Maintaining a good mood, more exercise and regular health checks.
The morbidity risk the possibility of: (High)	>51%	1. Quit smoking, control high cholesterol, high-fat foods, eat more Fruit and vegetables (smoking)
		2. Control high cholesterol, high-fat foods, eat more Fruit and vegetables (no-smoking)
		3. Maintaining a good mood ,more exercise and consult a doctor. 4. Note that blood pressure and blood sugar. Systolic blood pressure normal range 90-140 mm / Hg, and before meals and blood sugar normal range 70 - 100 mg/dl

4.4. System Performance Evaluation

The system performance is measured through the evaluation against forty has-been diagnosed patient cases of CKD. The risk levels are categorized into low, middle, and high risks. As the outcome shown in Table 5, there are thirty-two cases are accurately identified by the system out of forty testing cases. The accuracy rate is 80 percent. Although this figure is acceptable but it still has a space which ought to be improved.

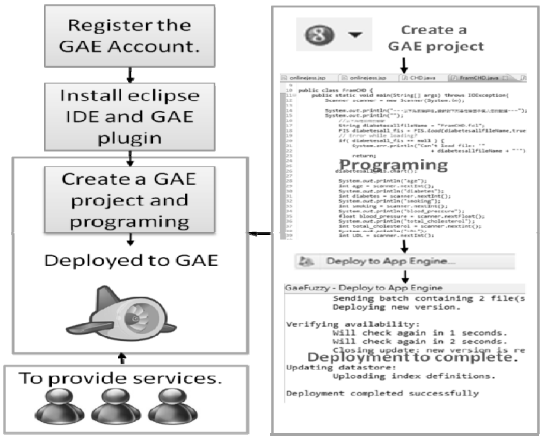
Table 5: The confusion table for system evaluation for has-been diagnosed cases

System / Has-been diagnosed	Low	Middle	High
Low	9	0	0
Middle	1	11	3
High	0	4	12

4.5. System Deployment to the Cloud

At this stage, we take Google cloud platform for deploying our system developed because it provides an easier, more flexible and scalable infrastructure for system development. The system deployment process with GAE to Google cloud platform is illustrated in Picture 4. It needs to register an account number of Google first followed by the installation of eclipse IDE and GAE plugin before creating a GAE project and programming on GAE cloud platform. Once the system project is created, it may be deployed to GAE. In additions, GAE also provides other services such as versions control, discharge observation, data access services, statistics and multi-media data access.

Picture 4: The system deployment process under GAE



Therefore, we can take the advantage of Google App Engine (GAE) provided by this platform to develop and run web applications on Google's cloud platform. Meanwhile, Google App Engine applications are easier to build, maintain, and scale as the demand of system computation and data storage grow.

Lastly, the objective of deploying the system developed on the cloud is to provide a high availability and scalability of system for the publics to use this system for assessing the risk of contracting the coronary heart disease in an early stage Hopefully, our deployment experience may be taken by other researchers for the implementation of same or analogous system development.

5. CONCLUSION

This paper develops a fuzzy expert system for the public users to self-assess his/her risk severity of coronary heart disease over Google cloud-service platform. Hopefully, this system may provide a convenient and efficient way to assess the risk of coronary artery heart disease. Meanwhile, it may also provide the medical professionals an alternative to understand the patient's health condition so that a better treatment can be given to an individual patient.. Although the accuracy rate of this system from case evaluation may reach to 80 percent which is acceptable from the viewpoints of experts but it may be improved in order to be used with more reliable in practice. Additionally, the cost, performance, and data security are also concerned and need to be further evaluated in the subsequent study of this paper.

REFERENCE LIST

1. Department of Health (DOH) (2012), Executive Yuan, Taiwan, "The statistics on the causes of death, health statistics in Taiwan, 2010," Retrieved on December 2012, from http://www.doh.gov.tw/EN2006/DM/DM2.aspx?now_fod_list_no=12537&class_no=390&level_no=2.
2. Zahan, S.(2001), "Fuzzy approach to computer-assisted myocardial ischemia diagnosis," Artificial Intelligent in Medicine (21), pp. 271-275.
3. Negnevitsky M. (2011), Artificial Intelligence, A Guide to Intelligent Systems, 3rd Ed., Addison Wesley, pp. 87-127, pp. 165-256.
4. Framingham Heart Study (2011), "Coronary heart disease (10-year risk) ," retrieved on October 2010 from http://www.framinghamheartstudy.org/.
5. Mell P. and Grance T. (2012), "The NIST definition of cloud computing," retrieved on December 2012, from http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf/.
6. Google (2012), "Google App Engine," Retrieved on August 20, from http://code.google.com/intl/zh-TW/appengine.
7. Dean J. & Ghemawat S. (2004), "MapReduce: simplified Data processing on large clusters," 6th Symposium on Operating Systems Design and Implementation USENIX Association.
8. Software R (2012), "What is R?" retrieved on June 2012, from http://www.r-project.org/.

9. NeuroSolutions (2012), “NeuroSolutions, the premier neural network development platform,” Retrieved on June, 2012 from http://www.jessrules.com/jess/index.shtml.
10. jFuzzyLogic (2014), “What is jFuzzyLogic?, dateless,” retrieved on October, 2014 from http://jfuzzylogic.sourceforge.net/html/index.html.
11. Jess (2014), “Jess, the rule engine for Java Platform,” retrieved on June 2014, from http://www.jessrules.com/jess/index.shtml.
12. Zadeh L.A.(1965), “Fuzzy sets,” Information and Control, vol. 8, issue 3, 1965, pp. 338-353.