OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

# Risk Assessment of Cardiovascular Disease Using Fuzzy Expert System for a Medical Examination Database

Jenn-Long Liu[1] and Yun-Yun Chang[2]

[1]*Professor,Dept. of Information Management, I-Shou University, Taiwan.*

[2]*Graduate Student, Dept. of Information Management, I-Shou University, Taiwan.*

*Corresponding Author: Jenn-Long Liu*

***ABSTRACT:****In recent years, the aging population is growing much faster than the total population in Taiwan. This change in the population structure on age has shifted Taiwan gradually to an aged society. As a result, elderly people increasingly concern about their health such that they take physical examinations regularly. Accordingly, this work focuses on the study of analyzing a medical examination database by using data mining and fuzzy logic techniques. The medical examination database, conducted by Nutrition and Health Survey in Taiwan from 2004 to 2008, is used by adopting 1,843 from 3,671 samples, excluding 1,827 ones with missing data, for the data mining and attribute analysis. Among the available samples, the confirmed cases with cardiovascular disease (CVD) are used and analyzed using classifiers with C4.5, NB tree, Naïve Bayes, Bayes Net, and Multi-Layer Perceptron, respectively. Experiments indicated that the result of a patient with hypertension, hyperlipidemia, stroke, and heart diseases is a reasonable combination to diagnose him or her as suffering from CVD or not. The accuracy of classification for the training set is 81% obtained using C4.5. Furthermore, this study adopts a fuzzy inference system with giving some significant attributes selected from the attribute analysis and healthcare knowledge. The significant attributes include five physiological indexes obtained from physical examination items and also three lifestyle factors. The eight crisp inputs of the significant attributes are selected from the medical examination database and then input into the fuzzy inference system to assess the risk of a patient suffering from CVD. The analyzed results using proposed data mining and fuzzy logic techniques for the medical examination database are shown to be good references for patients' self-health management and physician's disease diagnosis.*

***KEY WARDS:****Medical examination database,Data mining,Fuzzy logic,Cardiovascular Disease.*

---

Date of Submission: 28-05-2020 Date of acceptance: 14-06-2020

---

## I. INTRODUCTION

Population aging is common in many nations due to the fast growth of aging population and low birthrate of neonates. Especially, this problem in Japan is the most serious compared with other countries, and Japan also is experiencing a "super-aged" society [1]. Like Japan and most Western countries, the aging population in Taiwan has become larger in recent years. With 15.28% of its elder population aged over 65 and 119.82% of its population aging index bythe end of 2019, as listed in Table I [2], Taiwan become an "aged" society since the end of 2017. Furthermore, the elderly people will reach 20% of its total population and Taiwan will become a super-aged society at the end of 2025 [3, 4], according to the estimation from Council for Economic Planning and Development in Taiwan. Therefore, Taiwan government is actively promoting health care industry and conducting a 10-year long-term care plan now. As a result, the demands on medical treatment and health care will become more and more ardent in the future. Besides, an official report proposed by Ministry of Health and Welfare in Taiwan indicated that 81.1% of elderly people over the age of 65 complained of chronic diseases, were mainly hypertension (54%), osteoporosis (33%), diabetes (25%) and heart disease (21%) [5]. Accordingly, the expenses related to long-term medical treatment and health care for elder people will be a heavy burden on individuals and families in the future as the aging population increases year by year and also mostly elderly people have the tendency to suffer from chronic diseases.

Under the concepts of disease prevention and health promotion, how to maintain body health has become an important issue for modern people. Therefore, health care and health-related industries have emerged rapidly in recent years. People pay more attention to nutrition and mental fitness in conjunction with physical training. They also take periodical health examinations by precision instruments to screen out unknown diseases in the body and to make sense about themselves states of health. According to the statistics on adult preventive health services at medical service agencies counted by the Directorate General of Budget, Accounting and Statistics of Executive Yuan in Taiwan, the number of people declarations for participating health examination increased from 1,715,419 in 2008 to 1,853,961 in 2014, an increase of 138,543 in the six years. This shows that there has been a steady increase on people's participation of preventive health care in recent years. In view of the vigorous development of economy, the lifestyles and dietary habits of Taiwanese people are gradually changing. Some factors, such as air pollution, working pressure, fine diet, bad lifestyle, and personal aging, etc., do harm to personal health and increase the risk for people on affecting cardiovascular disease (CVD). The higher the risk factors in life, the greater the risk of suffering from CVD. In fact, CVD is a complex and fatal illness and it involves the deterioration of functions on heart or blood vessels. Cardiovascular diseases include coronary artery diseases (CAD), stroke, heart failure, hypertensive heart disease, rheumatic heart disease, cardiomyopathy, heart arrhythmia, and so on. According to an analytical report released by the Ministry of Health and Welfare, three types of CVD--heart disease, cerebrovascular disease and hypertensive disease--were ranked in the top 10 causes of death in the past few decades. The total number of deaths for the three types of chronic diseases was second only to the first malignant tumor. The ratios for deaths due to CVD were 22.4% in 2014, 21.0% in 2015, and 22.4% in 2016. This shows that the ratios of people dying from CVD are high. Clearly, CVD has become the major killer of chronic diseases. The top 10 causes of death for Taiwanese people in 2016-2018 were shown in Table II.

**Table I. Population structure of Taiwan for the last five years**

| Year | Population | Ratio of age≥65 | Ratio of age≤14 | Population aging index[1] |
|---|---|---|---|---|
| 2015 | 23,492,074 | 12.51 | 13.57 | 92.18% |
| 2016 | 23,539,816 | 13.20 | 13.35 | 98.86% |
| 2017 | 23,571,227 | 13.86 | 13.12 | 105.70% |
| 2018 | 23,588,932 | 14.56 | 12.92 | 112.64% |
| 2019 | 23,603,121 | 15.28 | 12.75 | 119.82% |

[1]: the ratio of (population of age≥65)÷(population of age≤14)

**Table II. The top 10 causes of death for Taiwanese people in the past three years (2016-2018)**

| **Disease**[2] | **Deaths per 100,000 population** | | |
|---|---|---|---|
| | 2016 | 2017 | 2018 |
| malignant tumor (i.e. cancer) | 203.1 | 203.9 | 206.9 |
| heart disease | 88.5 | 87.6 | 91.5 |
| pneumonia | 51.9 | 53.0 | 56.9 |
| cerebrovascular disease | 50.4 | 49.9 | 48.9 |
| diabetes | 42.4 | 41.8 | 39.8 |
| accident injury | 30.6 | 29.6 | 29.0 |
| chronic lower respiratory tract disease | 28.9 | 26.6 | 26.1 |
| hypertensive disease | 25.0 | 25.8 | 25.4 |
| nephritis, nephrotic | 22.2 | 22.8 | 23.4 |
| chronic liver disease and cirrhosis | 20.1 | 19.3 | 18.3 |

[2]: The rank of diseases is based on the top 10 causesof death in 2018 [5].

With the advancement of medical science and technology, the concept of preventive medicine has been accepted by most people. Therefore, the medical and healthcare industries rise up abundantly, especially in this aging society. In [6], authors presented the results of intelligent data analysis to analyze trends in lifestyle related chronic diseases for a periodical medical examination data with 300 records. As mentioned above, CVD, which includes heart disease, cerebrovascular disease and hypertensive disease, are ranked second only to malignant tumor (i.e. cancer) among the top 10 causes of death in Taiwan. Therefore, the advocacy on periodical health examination is needed for people to establish the concepts of early detection and early treatment to effectively reduce the incidence of CVD to death and also the impact of lethality. This study uses a medical examination database, conducted by Nutrition and Health Survey in Taiwan from 2004 to 2008 [7] to analyze the attributes of affecting CVD from physical examination records of patients. The cardiovascular disease-related cases are also evaluated to assess the performance of classification and to determine the important attributes related to CVD. Furthermore, this work adopts a fuzzy inference system based on fuzzy logic theory to systematically evaluate the risk of suffering from CVD for a patient by inputting the analyzed significant attributes and additional lifestyle factors.

## II. RESEARCH METHODS

### *2.1 Data Mining Analysis Description and Climate*

This study uses Weka software, developed by University of Waikato in New Zealand [8], as an analytical tool for data mining analysis. The medical examination database based on the subjects' physical examination reports was applied to implement the classification analysis. Five classifiers, including C4.5, NB tree, Bayes Net, Naïve Bayes and Multi-Layer Perceptron, were applied for classification using Weka software. The five classifiers were described as follows.

1- C4.5: The algorithm based on information gain ratio is proposed by Ross Quinlan for addressing the defects of ID3 algorithm [9]. As we known, ID3 chooses split attribute with the highest information gain and favors attributes with large number of divisions. Further, C4.5 is able to deal with noise, missing data and evading over-fitting by a pruning technique.

2- NB Tree: It is a simple Bayesian tree classification rule, which is suitable for the category prediction. Its leaf node of decision tree is based on a simple Bayesian classifier, which uses cross-validation to identify nodes and then splits them into internal nodes.

3- Bayes Net: Bayesian Net is a probabilistic network based on probabilistic reasoning, whereas Bayesian formulas are the basis for probabilistic network.

4- Naïve Bayes: In machine learning, Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. First, it classifies the training set and memorizes the relationship between the attributes. Then according to the relationship between the attributes used, it uses the concept of learning for non-classified data objects, type of prediction to get the objective goal of test data.

5- Multi-Layer Perceptron (MLP): The "Back Propagation Algorithm" belongs to the type of error correction learning to train the key value of the network, and then classify the instances.

### *2.2Significant Attributes Analysis*

The physical examination database used in this study is based on a medical examination report of Nutrition and Health Survey in Taiwan from 2004 to 2008, conducted by the Research Center for Humanities and Social Sciences of Academia Sinica, Taiwan [10]. The research objects include adults aged 19 or above for the survey design and planning of physical examination. The records of questionnaires and physical examination reports are 4,666 and 3,671, respectively. The reports were released for academic research since 2013. In this work, the Cardiovascular Disease Screening Table reported by Health Checkup Center of Taiwan Adventist Hospital (TAH) was also used to screen the properties of CVD with 20 numerical attributes and a diagnostic result related to the test items. The 20 numerical attributes and one diagnostic result were: height, weight, bmi (Body Mass Index), rbc (Red Blood Cells), hgb (Hemoglobin), hct (Hematocrit), mcv (Mean Corpuscular Volume), mch (Mean Corpuscular Hemoglobin), mchc (Mean Corpuscula-hemoglobin Concentration), wbc (White Blood Cell), plt (Platelets), fbs (Fasting Blood Sugar), crp (C-reactive Protein), t_cho (Total Cholesterol), tg (Triglyceride), ldl (Low-density Lipoprotein), hdl (High-density Lipoprotein), sbp (Systolic Blood Pressure), dbp (Diastolic Blood Pressure), pulse, and result (Diagnostic Result). Then, the questionnaires of the patients with CVD and physical examination reports were connected and got a total of 1,843 available dataset and 1,828 invalid dataset with missing data. This work just uses the 1,843 available dataset only. In addition, the cardiovascular-related diseases studied in this work were hypertension disease, hyperlipidemia disease, stroke, and heart disease only.

### *2.3Fuzzy Inference System*

In this study, a fuzzy inference system was used to implement the assessment of CVD risk. Since fuzzy logic theory can handle linguistic fuzzy variables, the fuzzy rules of the physical examination database can be expressed with the form of "If-Then" rule. The process of fuzzy logic inference involves all of the pieces that are characterized in membership functions, logical operations, and If-Then rules [11]. In this work, the fuzzy variables were selected with five physiological indexes and three lifestyle factors. The five variables for physiological indexes were sbp (Systolic Blood Pressure), bmi (Body Mass Index), fbs (Fasting Blood Sugar), tg (Triglyceride), crp (C-Reactive Protein), and the three variables for lifestyle factors were smoking, drinking, and exercise. Table III lists the linguistic variables and their ranges for the physiological indexes and lifestyle factors.

**Table III.Linguistic variables and their ranges**

| LinguisticVariables | Linguistic Variables | Range |
|---|---|---|
| sbp (Systolic Blood Pressure) | low | <120 |
| | middle | 120～160 |
| | high | ≧140 |
| bmi (Body Mass Index) | low | 18.5～24 |
| | middle | 22.5～27 |
| | high | ≧26.5 |
| fbs (Fasting Blood Sugar) | low | 80～100 |
| | middle | 90～140 |
| | high | ≧130 |
| tg (Triglyceride) | low | 50～151 |
| | middle | 141～200 |
| | high | ≧191 |
| crp (C-Reactive Protein) | low | <0.5 |
| | middle | 0.45～0.99 |
| | high | ≧0.76 |
| drinking | yes | 1 |
| | no | 0 |
| smoking | yes | 1 |
| | no | 0 |
| exercise | yes | 0 |
| | no | 1 |
| risk | very low | <20% |
| | low | 15～30% |
| | rather low | 25～40% |
| | middle | 35～50% |
| | rather high | 45～60% |
| | high | 55～70% |
| | very high | >65% |

Moreover, this study uses Mamdani inference system [12, 13] to find out the output result. There are six steps in the processes of Mamdani inference system: (1) determining a set of fuzzy rules, (2) fuzzifying the inputs using the input membership functions, (3) combining the fuzzified inputs according to the fuzzy rules to establish a rule strength, (4) finding the consequence of the rule by combining the rule strength and the output membership function, (5) combining the consequences to get an output distribution, and (6) defuzzifying the output distribution (this step is only if a crisp output (class) is needed). There are several defuzzification methods, but probably the most popular one is the centroid technique. After defuzzification, the system outputs a crisp value, which is the probability of the risk of a patient suffering from CVD.

## III. RESULTS AND DISCUSSION

### *3.1Classification Analysis*

In the classification analysis, the training data is generally used to evaluate whether the model is good or not. Therefore, the diagnostic results were first divided into two groups by negative and positive cases for CVD. The result of classification was expressed by a confusion matrix, as shown in Table IV. In Table IV, the “True Negative” (TN) means the case that the test makes a negative prediction for CVD, and the subject has a negative result for CVD, and the “false negative” (FN) means the case that the test makes a negative prediction for CVD, and the subject has a positive result for CVD. Also, the “True Positive” (TP) means the case that the test makes a positive prediction for CVD, and the subject has a positive result for CVD, and the “False Positive” (FP) means the case that the test makes a positive prediction for CVD, and the subject has a negative result for CVD. Furthermore, the cost matrix was adopted to show the cost to bear when an error in the predicted result occurs. If one actually shows negative for CVD, but the predicted result is positive for CVD, the cost is set to be 1; if patients actually has positive for CVD, but the predicted result is negative for CVD, the cost is increased by 5 times due to the severity of misdiagnosis. The cost matrix was also shown in Table 4, which the values of cost were listed inside parentheses. Accordingly, if the cost is higher, it means that the predicted performance is worse.

**Table IV.Confusion matrix and Cost matrix**

| Result \ Prediction | Negative | Positive |
|---|---|---|
| Negative | TN (0) | FP (1) |
| Positive | FN (5) | TP (0) |

Thereafter, five classifiers, C4.5, NB Tree, Bayes Net, Naïve Bayes and Multi-Layer Perceptron, were applied to evaluate the performance of classification for the medical examination database. The predicted performance of classifiers was measured by following the five parameters.

1- Accuracy: Accuracy is the ratio of the total number of true prediction divided by the overall number of the dataset. The higher of the accuracy value means the higher value of TN+TP and the lower value of FN+FP.

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+FN+FP+TP)} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

2- Sensitivity: The number of cases of suffering from CVD by prediction divided by the number of actual cases of suffering from CVD, also known as True Positive Rate (TPR) , which is

$$\text{Sensitivity (P)} = \frac{TP}{(FN+TP)} \quad \ldots\ldots\ldots\ldots\ldots (2)$$

3- Specificity: The number of cases of not suffering from CVD by prediction divided by the number of actual cases of not suffering from CVD, also known as True Negative Rate, (TNR), which is

$$\text{Specificity (R)} = \frac{TN}{(TN+FP)} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

4- F-measure: The parameter of F-measure is to measure the effect of Sensitivity (P) and Specificity (R) by a harmonic average. If both values, P and R, are high, the value of F-measure is high, indicating that prediction accuracy for TP and TN is also high. The definition of F-measure is

$$\text{F-measure} = \frac{2 \times P \times R}{(P+R)} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

5- Cost: The cost is designed to pay when a fault prediction occurs in order to measure the severity of fault prediction for a classifier. In this work, the values of cost to be paid for FP and FN were 1 and 5, respectively.

$$\text{Cost} = FP \times 1 + FN \times 5 \ldots\ldots\ldots\ldots\ldots (5)$$

**Table V.Confusion matrixes obtained using the five classifiers**

| | Prediction | | |
|---|---|---|---|
| Actuality | Classifiers | Negative | Positive |
| Negative | J48 | 1062 | 213 |
| | NB Tree | 1115 | 160 |
| | Bayes Net | 995 | 280 |
| | Naïve Bayes | 997 | 278 |
| | Multi-Layer Perceptron | 1066 | 209 |
| Positive | J48 | 137 | 431 |
| | NB Tree | 268 | 300 |
| | Bayes Net | 265 | 303 |
| | Naïve Bayes | 293 | 275 |
| | Multi-Layer Perceptron | 165 | 403 |

**Table VI. Comparisons of classified performance obtained using the five classifiers**

| Parameters | J48 | NB Tree | Bayes Net | Naïve Bayes | MLP |
|---|---|---|---|---|---|
| Accuracy | 81% | 77% | 70% | 69% | 80% |
| Sensitivity | 76% | 53% | 53% | 48% | 71% |
| Specificity | 83% | 87% | 78% | 78% | 84% |
| F-Measure | 79% | 66% | 63% | 60% | 77% |
| Cost | 898 | 1500 | 1605 | 1743 | 1034 |

From Tables V and VI, the accuracy of the C4.5 classification algorithm outperforms other classifiers with the best accuracy, sensitivity, F-measure, and cost. It is found that the accuracy for classification obtained using C4.5 was the best and the value of 81% was acceptable, whereas, the Naïve Bayes classification has the lowest accuracy. Accordingly, the combination of hypertension, hyperlipidemia, stroke, and heart diseases for CVD indeed was a good index to judge whether a patient suffers from CVD or not.

### *3.2 Attribute Analysis*

To measure the degree of importance of attributes, the attribute analysis was conducted to determine the significance. First, the dataset were divided into two groups by negative and positive cases for CVD. Then we used information gain theory to rank the attributes. The outcome of the first ranking attribute obtained using the attribute analysis was sbp. Moreover, the top 5 significant attributes in this study were sbp, bmi, fbs, tg, and crp, according to the ranking of importance of attribute. The detailed results for the attribute analysis were summarized in Table VII.

**Table VII. The ranking of attributes related with suffering from CVD**

| Ranking | Attribute | Importance of attribute |
|---|---|---|
| 1 | sbp | 0.09138 |
| 2 | bmi | 0.03200 |
| 3 | fbs | 0.03193 |
| 4 | tg | 0.02726 |
| 5 | crp | 0.02154 |
| 6 | dbp | 0.02048 |
| 7 | height | 0.01797 |
| 8 | hdl | 0.01392 |
| 9 | t_cho | 0.00927 |
| 10 | ldl | 0.00849 |

### *3.3 Fuzzy Inference for Risk Assessment*

The fuzzy inference was based on the JFuzzyLogic package which was designed by Cingolani and Alcalá-Fdez [14] using Java Language. The inference engine is the core of fuzzy expert system and can be used as a tool for access the risk of suffering from CVD. After selecting significant attributes using information gain theory as mentioned in Sec. 3.2, this work decides to use five physiological index variables and three lifestyle factors as the inputs of JFuzzyLogic to implement the risk assessment of suffering from CDV. The definitions of the membership functions and their levels for the eight fuzzy variables (i.e. inputs) were shown in Figs. 1(a)-(h).

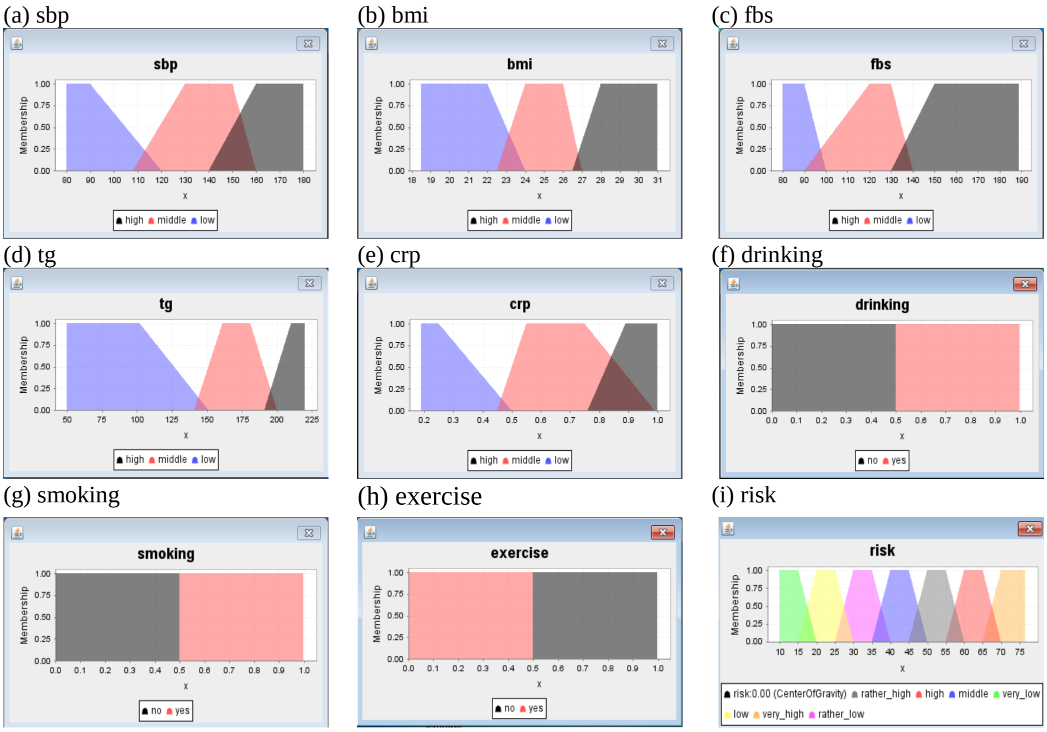
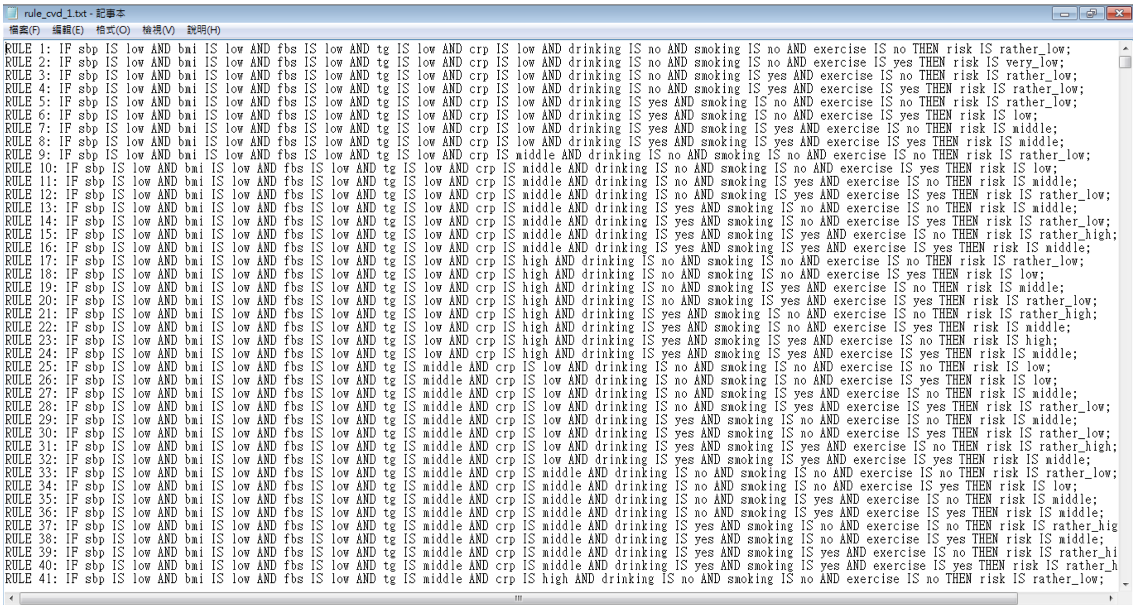



**Figure 1:Membership functions for selected input variables**

In addition, the membership function of the risk (i.e. output) was designed to seven levels of linguistic variable: very low, rather low, low, middle, high, rather high, and very high, as shown in Fig. 1(i). After the fuzzy rules were generated based on the levels of fuzzy variables for the eight inputs and one output, the knowledge rules can be established. As displayed in Fig. 2, a total of 1944 knowledge rules were generated and the rules were stored in the knowledge base for the use of the fuzzy inference. This study applied Microsoft Visual Basic 2010 software to build an interface window for entering the inputs of designated fuzzy variables and connecting the fuzzy inference system. The interface window as displayed in Fig 3(a). Users can enter the crisp inputs relating to the physiological indexes and lifestyle factors from the records of health examination database. The inputs of physiological indexes include Systolic Blood Pressure (sbp), Body Mass Index (bmi), Fasting Blood Sugar (fbs), Triglyceride (tg), and C-reactive Protein (crp). In addition, three lifestyle factors, drinking, smoking and exercise, were used. The evaluated result assessment was a crisp numeric and displayed graphically using the JFuzzyLogic. An example of output result with risk of 70.52% was shown in Fig. 3(b) by using a record of physical examination database with inputs: sbp=143, bmi=28.8, fbs=107, tg=129, crp=0.76, drinking=no, smoking=yes, and exercise=no.



```
rule_cvd_1.txt - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)
RULE 1: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS rather_low;
RULE 2: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS no AND smoking IS no AND exercise IS yes THEN risk IS very_low;
RULE 3: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS no AND smoking IS yes AND exercise IS no THEN risk IS rather_low;
RULE 4: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS no AND smoking IS yes AND exercise IS yes THEN risk IS rather_low;
RULE 5: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS yes AND smoking IS no AND exercise IS no THEN risk IS rather_low;
RULE 6: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS yes AND smoking IS no AND exercise IS yes THEN risk IS low;
RULE 7: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS yes AND smoking IS yes AND exercise IS no THEN risk IS middle;
RULE 8: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS low AND drinking IS yes AND smoking IS yes AND exercise IS yes THEN risk IS middle;
RULE 9: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS rather_low;
RULE 10: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS no AND smoking IS no AND exercise IS yes THEN risk IS low;
RULE 11: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS no AND smoking IS yes AND exercise IS no THEN risk IS middle;
RULE 12: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS no AND smoking IS yes AND exercise IS yes THEN risk IS rather_low;
RULE 13: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS yes AND smoking IS no AND exercise IS no THEN risk IS middle;
RULE 14: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS yes AND smoking IS no AND exercise IS yes THEN risk IS rather_low;
RULE 15: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS yes AND smoking IS yes AND exercise IS no THEN risk IS rather_high;
RULE 16: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS middle AND drinking IS yes AND smoking IS yes AND exercise IS yes THEN risk IS middle;
RULE 17: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS rather_low;
RULE 18: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS no AND smoking IS no AND exercise IS yes THEN risk IS low;
RULE 19: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS no AND smoking IS yes AND exercise IS no THEN risk IS middle;
RULE 20: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS no AND smoking IS yes AND exercise IS yes THEN risk IS rather_low;
RULE 21: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS yes AND smoking IS no AND exercise IS no THEN risk IS rather_high;
RULE 22: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS yes AND smoking IS no AND exercise IS yes THEN risk IS middle;
RULE 23: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS yes AND smoking IS yes AND exercise IS no THEN risk IS high;
RULE 24: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS low AND crp IS high AND drinking IS yes AND smoking IS yes AND exercise IS yes THEN risk IS middle;
RULE 25: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS low;
RULE 26: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS no AND smoking IS no AND exercise IS yes THEN risk IS low;
RULE 27: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS no AND smoking IS yes AND exercise IS no THEN risk IS middle;
RULE 28: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS no AND smoking IS yes AND exercise IS yes THEN risk IS rather_low;
RULE 29: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS yes AND smoking IS no AND exercise IS no THEN risk IS middle;
RULE 30: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS yes AND smoking IS no AND exercise IS yes THEN risk IS rather_low;
RULE 31: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS yes AND smoking IS yes AND exercise IS no THEN risk IS rather_high;
RULE 32: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS low AND drinking IS yes AND smoking IS yes AND exercise IS yes THEN risk IS middle;
RULE 33: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS rather_low;
RULE 34: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS no AND smoking IS no AND exercise IS yes THEN risk IS low;
RULE 35: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS no AND smoking IS yes AND exercise IS no THEN risk IS middle;
RULE 36: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS no AND smoking IS yes AND exercise IS yes THEN risk IS middle;
RULE 37: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS yes AND smoking IS no AND exercise IS no THEN risk IS rather_hig
RULE 38: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS yes AND smoking IS no AND exercise IS yes THEN risk IS middle;
RULE 39: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS yes AND smoking IS yes AND exercise IS no THEN risk IS rather_hi
RULE 40: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS middle AND drinking IS yes AND smoking IS yes AND exercise IS yes THEN risk IS rather_h
RULE 41: IF sbp IS low AND bmi IS low AND fbs IS low AND tg IS middle AND crp IS high AND drinking IS no AND smoking IS no AND exercise IS no THEN risk IS rather_low;
```

**Figure 2:Knowledge rules with total of 1944 rules**

**(a) (b)**

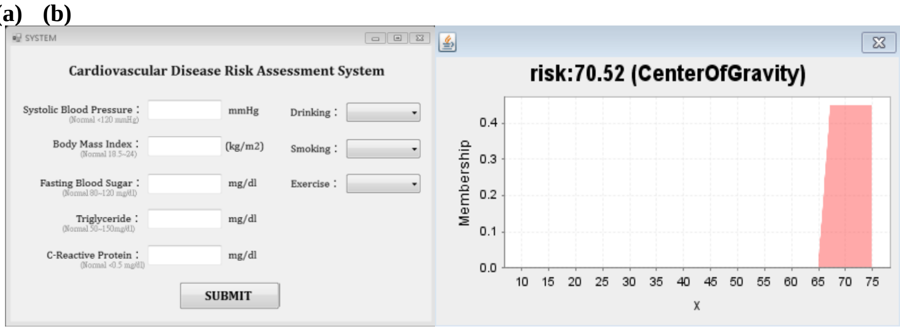



**Figure 3:(a) Interface window of the risk assessment system and (b) Crisp output of the risk assessment**

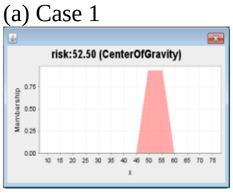
Furthermore some data are randomly selected from physical examination database to test the performance of the present fuzzy interference engine system. As listed in Table VIII, ten cases with five physiological indexes and three lifestyle factors were tested. The present risk assessment system can automatically read the inputs and find out the corresponding output rule based on the rules of knowledge base. The output results of risk assessment for the ten tests were displayed in Fig. 4. Moreover, as shown in Table VIII, the values of the risk for CVD in Case 1, Case 3, Case 5, Case6 , Case 7 and Case 10 were found to be over 50%, with high possibility of affecting cardiovascular-related diseases, as a result the six cases were predicted to have suffered from CVD. Meanwhile, the values of the risk for CVD in Case 2, Case 4, Case 8, and Case 9 were all less than 50% with a low possibility of affecting cardiovascular-related diseases. As a result, the four cases were predicted without suffering from CVD. Clearly, the systematic calculations of the ten cases using present fuzzy inference system for predicting CVD coincided with the diagnostic result of CVD in the physical examination data. It was also found that when Systolic Blood Pressure (sbp) and Body Mass Index (bmi) were at high values, the risk of suffering from CVD was high.

**Table VIII. Comparisons of ten test cases between the diagnosed results and risk assessments using fuzzy inference system**

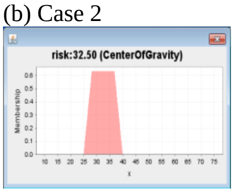
| Case | ID | Records in the physical examination database | | | | | | | | | Prediction by fuzzy inference system | | Comparison |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | sbp | bmi | fbs | tg | crp | drinking | smoking | exercise | diagnosed positive? (*) | risk assessment | level of risk | consistency ? |
| 1 | A753 | 131 | 20.8 | 73 | 105 | 0.25 | yes | yes | no | 1 | 52.50% | rather high | matched |
| 2 | A771 | 101 | 20.6 | 88 | 65 | 0.07 | no | no | no | 0 | 32.20% | rather low | matched |
| 3 | B386 | 143 | 28.8 | 107 | 129 | 0.76 | no | yes | no | 1 | 70.52% | very high | matched |
| 4 | A746 | 109 | 18.4 | 94 | 53 | 0.07 | yes | no | no | 0 | 38.32% | rather low | matched |
| 5 | B355 | 135 | 32 | 109 | 123 | 1.21 | no | no | no | 1 | 62.50% | high | matched |
| 6 | D437 | 162 | 26 | 173 | 79 | 1.25 | no | yes | no | 1 | 71.09% | very high | matched |
| 7 | D595 | 210 | 30.6 | 143 | 110 | 0.2 | no | yes | yes | 1 | 70.74% | very high | matched |
| 8 | H757 | 129 | 22.6 | 86 | 62 | 0.07 | no | no | yes | 0 | 23.44% | low | matched |
| 9 | B096 | 148 | 20.2 | 103 | 58 | 0.07 | no | no | no | 0 | 47.31% | middle | matched |
| 10 | A005 | 115 | 28.9 | 104 | 128 | 0.3 | no | no | no | 1 | 59.11% | rather high | matched |

Remarks: (*)1: positive on affecting CVD; 0: negative on affecting CVD

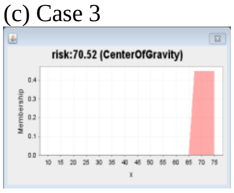
(a) Case 1



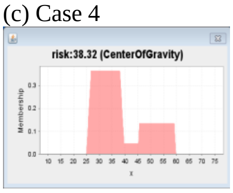

(b) Case 2



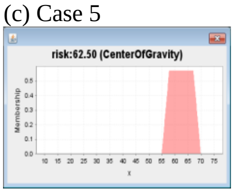

(c) Case 3



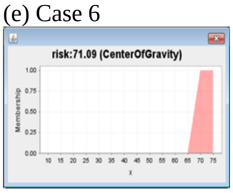

(c) Case 4



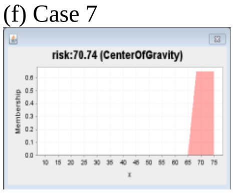

(c) Case 5



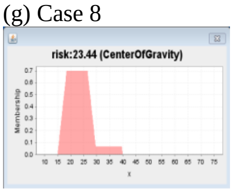

(e) Case 6



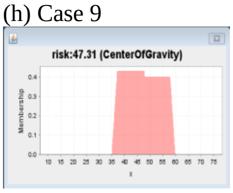

(f) Case 7



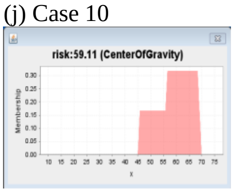

(g) Case 8




(h) Case 9




(j) Case 10




To further evaluate the changes in the risk to CVD related to regular exercise habits, this work changed the input values of variable "exercise" for the aforementioned ten cases to observe the presence or absence of exercise-related factor to the risk of suffering from CVD. If the original case in the physical examination database has no regular exercise habits, then the testing case has regular exercise habits. On the contrary, if the original case in the physical examination database has regular exercise habits, then the testing case has no regular exercise habits. Therefore, we can check whether there is an important degree of influence for the lifestyle factor "exercise" to CVD. In Table IX, it was found that if there was presence of exercise, the risk of suffering from CVD diseased was high. However, if the physiological indexes have been at critical high levels, the predicted values of risk of suffering from CVD did not make any changes. That is, from the results of Cases 5-7, we found that when the body-related physiological indexes of patients were higher than the normal standards, they are still at high risk of suffering from CVD whether they have regular exercise habits or not.

**Table IX.The effect of regular exercise habits to CVD**

| Case | ID | Records in the physical examination database | | | | | | | | | Exercise effect using fuzzy inference system | | Verification |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | sbp | bmi | fbs | tg | crp | drinking | smoking | exercise | diagnosed positive? (*) | exercise: yes | exercise: no | Has an effect? |
| 1 | A753 | 131 | 20.8 | 73 | 105 | 0.25 | yes | yes | no | 1 | 42.50% | 52.50% | yes |
| 2 | A771 | 101 | 20.6 | 88 | 65 | 0.07 | no | no | no | 0 | 14.24% | 32.20% | yes |
| 3 | B386 | 143 | 28.8 | 107 | 129 | 0.76 | no | yes | no | 1 | 63.70% | 70.52% | yes |
| 4 | A746 | 109 | 18.4 | 94 | 53 | 0.07 | yes | no | no | 0 | 26.49% | 38.32% | yes |
| 5 | B355 | 135 | 32 | 109 | 123 | 1.21 | no | no | no | 1 | 62.50% | 62.50% | no |
| 6 | D437 | 162 | 26 | 173 | 79 | 1.25 | no | yes | no | 1 | 71.09% | 71.09% | no |
| 7 | D595 | 210 | 30.6 | 143 | 110 | 0.2 | no | yes | yes | 1 | 70.74% | 70.74% | no |
| 8 | H757 | 129 | 22.6 | 86 | 62 | 0.07 | no | no | yes | 0 | 23.44% | 33.44% | yes |
| 9 | B096 | 148 | 20.2 | 103 | 58 | 0.07 | no | no | no | 0 | 37.31% | 47.31% | yes |
| 10 | A005 | 115 | 28.9 | 104 | 128 | 0.3 | no | no | no | 1 | 42.50% | 52.50% | yes |

Remarks: (*)1: positive on affecting CVD; 0: negative on affecting CVD

## IV. CONCLUSIONS

Due to population structure on age changes in Taiwan, health care industries have been flourishing in recent years. This study conducted the medical examination data from the research project of Nutrition and Health Survey in Taiwan from 2004 to 2008. There are 1844 available data after excluding 1827 records with missing data. The available data were divided into two groups: with/without suffering from hypertension, hyperlipidemia, stroke, and heart diseases. Using the classification algorithms by C4.5, NB Tree, Bayes Net, Naïve Bayes and Multi-Layer Perceptron in Weka software, we got the confusion matrixes of classifications to study the performance of the classifiers. From the classification results, the parameters of accuracy, sensitivity, and F-measure are obtained using C4.5 outperformed other classifiers, with an accuracy of 81%, sensitivity of 76%, F-measure of 79%, and cost of 898. From the attribute analysis using information gain theory, this work selected five physiological indexes, which were sbp (Systolic Blood Pressure), bmi (Body Mass Index), fbs (Fast Blood Sugar), tg (Triglyceride), and crp (C-reactive Protein), as the significant attributes according to the ranking of importance of attributes. In addition, the five attributes of physiological indexes and additional three lifestyle factors of drinking, smoking, and exercise were selected as the inputs of fuzzy inference system. by using JFuzzyLogic tool. Then, this work selected ten records randomly from the physical examination database for risk assessment to evaluate the performance of fuzzy expert system. From the comparisons, the results of diagnosed outcome and present predicted values of risk for CVD agreed well. It was also found that when the levels of systolic blood pressure and body mass index were high, the risks of suffering from CVD were high. To further examine the effect of regular exercise habits to the risk of suffering from CVD, the work changed the input values of attribute of exercise for the selected ten data and evaluated them by the fuzzy expert system. Results showed that there was indeed a significant decrease in the risk of CVD with regular exercise habits when the relevant physiologic indexes were below high levels. However, there were not any changes for the risk values, when the body-related physiological indexes were above high levels. From the research results, the present fuzzy expert system with a simple interface window was effective and available for accessing the risk of suffering from CVD. The system can be a useful tool to achieve the goals of self-health management, health promotion, and ill-health prevention so as to reduce the cost of medical resources.

## REFERENCES

[1]. Muramatsu N. and Akiyama, H.,(2011). “Japan: Super-Aging Society Preparing for the Future,” The Gerontologist, 51, 4, 425–432, DOI=https://doi.org/10.1093/geront/gnr067.

[2]. Interior National Indicators, https://www.moi.gov.tw/files/news_file/w10904%E4%BA%BA%E5%8F%A3%E5%B9%B4%E9%BD%A1.pdf. published in Chinese Language.

[3]. Aging in Taiwan - EASP | East Asian Social Policy research network, http://www.welfareasia.org/5thconference/papers/Hsueh%20J_living%20arrangement%20and%20well-being.pdf.

[4]. Hsueh, J. C. and Yun-Tung Wang, Y. T., “Living Arrangement and the Well-being of the Elderly in Taiwan,” http://www.welfareasia.org/5thconference/papers/Hsueh%20J_living%20arrangement%20and%20well-being.pdf.

[5]. Ministry of Health and Welfare, https://www.mohw.gov.tw/cp-16-48057-1.html, presented in Chinese Language.

[6]. Waghmare, K. and Pai, A. R.,(2013). “Analytical study using data mining for periodical medical examination of employees,” Aswatha Kumar M. et al. (Eds.), Proceedings of ICAdC, AISC 174, pp. 221–227, Springer India.

[7]. Tu, S. H., Chen, C., Hsieh, Y. T., Chang, H. Y., Yeh, C. J., Lin, Y. C., and Pan, W. H.,(2011). “Design and sample characteristics of the 2005-2008: Nutrition and health survey in Taiwan,” Asia Pacific Journal of Clinical Nutrition, 20, 2, 225-237.

[8]. Witten, I. H., Frank, E., Hall, M., and Pal, C.,(2016). Data Mining: Practical Machine Learning Tools and Techniques, fourth edition, Morgan Kaufmann. https://www.cs.waikato.ac.nz/ml/weka/.

[9]. Quinlan, J. R.,(1979). Discovering rules form large collections of examples: A case study. Expert Systems in the Micro Electronic Age. Edinburgh Press: Edingburgh.

[10]. Health Checkup Center of Taiwan Adventist Hospital, https://www.tahsda.org.tw/en/main.php.

[11]. Yager, R.R.,(1996). “On the interpretation of fuzzy if then rules,” Applied Intelligence, 6, 2, 141–151.

[12]. Mamdani, E. H. and Assilian, S.,(1975). “An experiment in linguistic synthesis with a fuzzy logic controller,” International Journal of Man-Machine Studies, 7, 1, 1-13.

[13]. Mamdani, E. H.,(1977). “Application of fuzzy logic to approximate reasoning using linguistic synthesis,” IEEE Transactions on Computers, 100, 12, 1182-1191.

[14]. Cingolani, P. and Alcalá-Fdez, J.,(2013). “jFuzzyLogic: A java library to design fuzzy logic controllers according to the standard for fuzzy control programming,” International Journal of Computational Intelligence Systems, 6, sup1, 61-75.

Jenn-Long Liu, et. al." Risk Assessment of Cardiovascular Disease Using Fuzzy Expert System for a Medical Examination Database." *International Journal of Modern Engineering Research (IJMER)*, vol. 10(05), 2020, pp 01-09.