DUJASE Vol. 6 (1) 16-21, 2021 (January)

Fuzzy Hypertension Risk Prediction System

Iffat Arefa*, Abu Bakar Siddique and M. S. Alam

Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka, Bangladesh
Euro-Bangla Heart Hospital Ltd. Dhaka, Bangladesh

**E-mail: iffat_arefa@yahoo.com*

Received on 15.05.20, Accepted for publication on 14.6.21

ABSTRACT

This paper presents a fuzzy decision support system for risk prediction of hypertension. Risk factors relating to physiological, pathological and information about patients' lifestyle are considered to assess risk based on medical knowledge. To capture/utilize imprecise risk factors in a precise way, a fuzzy inference system is proposed where the rule-base is developed based on experts' (cardiologists and consultants) knowledge in the problem domain. For tuning and testing the system, primary patient data are collected from a heart hospital in Bangladesh. The output risk scores of the decision support system are compared against decisions made by cardiologists as diagnosed from patients' record. The results demonstrate that the proposed fuzzy decision support system is capable of predicting the risk of hypertension with sufficient accuracy.

Keywords: *Fuzzy Logic, Decision Support System, Hypertension*

1. Introduction

Hypertensive heart disease refers to heart conditions caused by sustained high blood pressure. The heart working under high pressure increases risk of coronary heart disease (CHD), left ventricular hypertrophy (LVH), heart failure, arrhythmia, heart attack, sudden cardiac arrest, stroke and sudden death [1]. According to WHO data published in 2017, cardiovascular disease is the leading cause of death globally, taking an estimated 17. 9 million lives each year [2]. In Bangladesh, CHD caused 14.31% of total deaths [3]. According to the Framingham Heart Study, hypertension has a 2-fold increase in the development of heart failure in men and a 3-fold increase for women [4]. The 2015 SPRINT trial suggested that proper management of hypertension correlates with a 64% reduction in the development of heart failure [5]. Therefore, to reduce mortality, it is necessary to predict the disease as early as possible. But it is challenging for doctors to predict hypertension early since various symptoms, such as fatigue, pounding in chest, neck, or ears and factors, such as, smoking, high blood pressure, diabetes, physical inactivity, overweight, high blood cholesterol, age, gender, race-ethnicity are needed to be considered with due importance to make a decision [6-7]. More importantly, some symptoms and risk factors are required to be carefully and disproportionately emphasized subjective to patients depending on prevalent symptoms and risk factors. Even experienced medical professionals and consultants, although numbers are extremely inadequate proportionate to the huge population in Bangladesh, find it difficult to provide decision in many cases. Moreover, for many patients, symptoms appear to be imprecise, confusing and in many cases misleading that clinical people cannot diagnose/decide with certainty or confidence. Furthermore, many patients, especially elderly and having disabilities of different forms, are not precise in expressing their situation, rather they use ambiguous terms and language that doctors, without having much experience, may interpret differently. Among the artificial intelligence paradigms, fuzzy logic is an excellent choice to capture/describe fuzziness, vagueness, and imprecise data/facts precisely. Unlike other methods, fuzzy logic allows the use of imprecise /ambiguous values as inputs and develop an inference mechanism among linguistic terms, the numerical values of the disease parameters and expert knowledge. Fuzzy logic reflects how people think. It attempts to model our sense of words and our decision making. To deal with imprecise data, indistinct symptoms and physiological states of patients, and severe scarcity/ unavailability of experienced cardiologists/ consultants the authors intend to develop a hypertension risk prediction system using fuzzy logic considering its success in a wide range of diseases including cardiovascular diseases as reported in [8-13].

2. Methods and Materials

2.1 Data Collection and Processing

In order to develop fuzzy decision support system, a number of symptoms, modifiable risk factors and non-modifiable risk factors are considered in consultation with experienced cardiologists and consultants.

A total of 16 variables are selected as predictor inputs. Three categorical output (i. e. responses) are chosen such as No Risk, Medium Risk and High Risk. 129 patients' data are collected from local hospital over a period of three months. The selected predictor inputs and output are presented in Table 1 and Table 2 respectively along with ranges of fuzzy sets. Some input data are measured in numerical values whereas some data are represented categorically. For example, in case of input 'Diabetes Disease', if a patient has diabetes then input is used as 1 and for non-diabetic as 0.

Table 1. Predictor Inputs and Ranges of fuzzy sets

Input Field	Range	Fuzzy Set
Systolic Pressure (mm of Hg) (X_1)	<120	Normal (X_{1N})
	120-139	Prehypertension (X_{1P})
	140-159	Stage 1 (X_{1S1})
	>=160	Stage 2 (X_{1S2})
Diastolic Pressure (mm of Hg) (X_2)	< 80	Normal (X_{2N})
	80-89	Prehypertension (X_{2P})
	90-99	Stage 1 (X_{2S1})
	>=100	Stage 2 (X_{2S2})
Heart Rate (X_3)	Up to <=100	Normal (X_{3N})
	> 100	High (X_{3H})
BMI (X_4)	10-25	Normal Weight (X_{4N})
	24-30	Overweight (X_{4Ov})
	29-65	Obesity (X_{4Ov})
Diabetes Disease (X_5)	0	Non-diabetic (X_{5N})
	1	Diabetic X_{5D}
Other Disease [Kidney Disease, Sleep Apnea etc.] (X_6)	0	NEG (X_{6N})
	1	POS (X_{6Y})
Smoking Condition (X_7)	0	No Consumer (X_{7N})
	0. 5	Low Consumer (X_{7L})
	1	High Consumer (X_{7H})
Palpitation Related (X_8)	0	NEG (X_{8N})
	1	POS (X_{8P})
Genetic Factor/Family History (X_9)	0	NEG (X_{9N})
	1	POS (X_{9P})
Physical Activity/ Sedentary Lifestyle (X_{10})	0	Active (X_{10A})
	1	Not Active (X_{10N})
Unhealthy Food Consumption (X_{11})	0	NEG (X_{11N})
	0. 5	Sometimes (X_{11S})
	1	Regular (X_{11R})
Stress Level (X_{12})	0	No Stressful Life (X_{12N})
	0. 5	Moderate Stressful Life (X_{12M})
	1	Stressful Life (X_{12S})
Environment (X_{13})	0	Comfortable (X_{13C})
	1	Uncomfortable (X_{13U})
Salt Intake (X_{14})	0	Normal Consumer (X_{14N})
	1	High Consumer (X_{14H})
Special Symptoms [Severe Headache/Drowsiness/ Fatigue/Confusion/Visio n Problems] (X_{15})	0	NEG (X_{15N})
	1	POS (X_{15Y})
Blood in the urine (X_{16})	0	NEG (X_{16N})
	1	POS (X_{16Y})

The risk for hypertension is classified into three levels within the range of [0 – 100] represented by three fuzzy sets shown in Table 2.

Table 2. Classification of Output Risk level

Output Field	Range	Fuzzy Set
Risk Level	0-30	No Risk
	25-60	Medium Risk
	55-100	High Risk

No Risk condition is considered as normotensive condition in medical term as the patient have no risk of hypertension in present condition. Prevention steps should be taken to maintain healthy condition [1]. Medium Risk is considered as risk of Hypertension as pre-stage of Hypertension or mild hypertension. Prevention steps have to be taken in this condition for preventing future risk of hypertension. High Risk is considered when the patient is diagnosed as hypertension positive or having high probability in future risk of hypertension. The patient needs to consult with doctor urgently [1].

2.2 The Proposed Risk Prediction System for Hypertension

The risk prediction system is designed based on Mamdani fuzzy inference mechanism for ease of incorporating expert knowledge in the form of IF-THEN. Contrary to other fuzzy inference systems, such as Sugeno and Tsukamoto, the IF-THEN rules of Mamdani fuzzy system do not require any derivation of functions from a large number of patient data. Patient information, collected as 16 inputs and converted into fuzzy sets shown in Table 1, are fed to the system to infer the risk factor based on the rule-base and defuzzification technique. Here centroid method is used for defuzzification. The predicted risk score will be a rounded to numerical value. The Mamdani fuzzy inference system comprising four processes such as fuzzification, inference mechanism, rule-base, and defuzzification is shown in Figure 1.

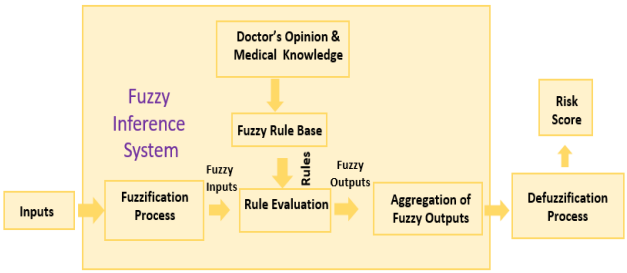


Fig. 1. Proposed hypertension risk prediction system based on Fuzzy logic

2.3 Knowledge Base

The output risk level is computed based on the 16 input predictors collected from patient record, the inference mechanism, rule-base and the defuzzification method used in the prediction system. The rule-base is the knowledge-base of the fuzzy system and comprises a set of rules where multiple inputs are combined using operators to infer an output. The rule-base requires the knowledge of the domain experts. People having all healthy data inputs are considered as no risk state and people having at least one unhealthy data input are considered as risk state. That is the key consideration to design rules. It is important to note that many symptoms and variables are needed to be considered in different proportion to decide risk prediction and in such case, experienced cardiologists'/consultants' knowledge is explored. The knowledge of domain expert (cardiologists and consultants) is incorporated or encoded as rules in the rule-base of fuzzy expert system. The designed rules are listed below:

Rule-1: IF Systolic Pressure is **NORMAL**, Diastolic Pressure is **NORMAL**, Heart Rate is **NORMAL**, BMI is **NORMAL**, Diabetes Disease is **NON DIABETIC**, Other

Disease is **NEG**, Smoking Condition is **NO CONSUMER**, Palpitation Related is **NEG,** Family History is **NEG**, Lifestyle is **ACTIVE**, Unhealthy Food Consumption is **NEG,** Stress Level is **NO STRESSFUL LIFE**, Environment is **COMFORTABLE**, Salt Intake is **NORMAL CONSUMER**, Special Symptoms is **NEG**, Blood in the Urine is **NEG** (=AND operator)

THEN the person is in "**NO RISK**" state.

Rule-2: IF Systolic Pressure is **PREHYPERTENSION** or **STAGE 1**, Diastolic Pressure is **PREHYPERTENSION** or **STAGE 1**, BMI is **OVERWEIGHT**, Smoking Condition is **LOW CONSUMER**, Lifestyle is **NOT ACTIVE**, Unhealthy Food Consumption is **SOMETIMES**, Stress Level is **MODERATE STRESSFUL LIFE** Special Symptoms is **POS**, Blood in the Urine is **POS** (,=OR operator)

THEN the person is in "**MEDIUM RISK**" state.

Rule-3: IF Systolic Pressure is **STAGE 2**, Diastolic Pressure is **STAGE 2**, Heart Rate is **HIGH**, BMI is **OBESITY**, Diabetes Disease is **DIABETIC**, Other Disease is **POS**, Smoking Condition is **HIGH CONSUMER**, Palpitation Related is **POS**, Family History is **POS**, Unhealthy Food Consumption is **REGULAR**, Stress Level is **STRESSFUL LIFE**, Environment is **UNCOMFORTABLE**, Salt Intake is **HIGH CONSUMER** (,=OR operator)

THEN the person is in "**HIGH RISK**".

2.4 Implementation and Performance Analysis

In this paper, 129 patients information was collected from local hospital, out of that 96 were used for tuning the fuzzy system and remaining 33 were used for testing the system. Here confusion matrix has been used for performance analysis. To form the confusion matrix, four notations are used;

TP=True Positive values
TN=True Negative Values
FP=False Positive Values
FN=False Negative Values

For example, if the system predicts a person is in high risk who is actually is in high risk, then it is the TP but if it mistakenly predicts as No Risk, then it will be FN. Similarly, if a person is predicted as No Risk who is actually is in no risk state, then it is TN but if it is predicted as High Risk, it is considered as FP. Thus the confusion matrix is formed and is used to evaluate performance of any prediction system.

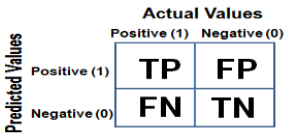


Fig. 2. Confusion Matrix

From confusion matrix we can also calculate Sensitivity, Specificity and Accuracy. Sensitivity is formulated as

$$Sensitivity = \frac{TP}{TP + FN}$$

It measures the proportion of actual positives that are correctly identified. Specificity is calculated as

$$Specificity = \frac{TN}{TN + FP}$$

This measures the proportion of actual negatives that are correctly identified. Finally Accuracy is denoted by

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

First, sensitivity, specificity and accuracy of the designed fuzzy decision system are calculated. For clarity, some notations are used in following discussion and those are as follows:

H = Number of high risk patients
M = Number of medium risk patients
N = Number of no risk patients

A total of 96 patient information is used that includes all three types of patients; numbers of high risk, medium risk and no risk patients are 71, 19 and 6 respectively. It means,

H = 71; N = 6; M = 19 and

Total = H + M + N = 96. The distribution is shown below (Fig. 3).

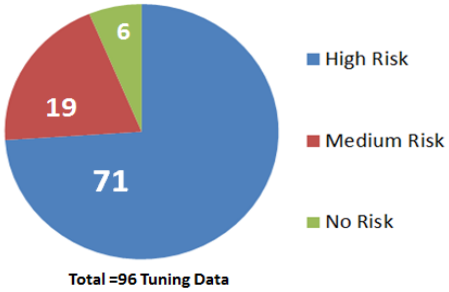


Fig. 3. Data set for Tuning

At the initial run, out of 96 cases, the system could correctly predicted risk of 75 patients with 16 errors, that is, the system wrongly predicted 15 medium risk patients as high risk and one no risk patient as high risk.

So, considering "No Risk" as Negative class and "Medium Risk" and "High Risk" as Positive class. It can be summarized as TN=75, TP=5, FP=1, FN=15. Using these values, sensitivity, specificity and accuracy are calculated as below:

$$Sensitivity = \frac{5}{5 + 15} = 0.25 = 25\%$$

$$Specificity = \frac{75}{75 + 1} = 0.9868 = 98.68\%$$

$$Accuracy = \frac{5 + 75}{5 + 15 + 1 + 75} = 83.33\%$$

Although results seem satisfactory but for a risk prediction system, especially, related to deadly diseases, it is required to be absolutely perfect/accurate. To improve performance

measures, as calculated above, the designed system was tuned, described in following section.

2.5 Tuning Process

For designing the system perfectly, tuning is necessary for fuzzy system. The tuning process was designed like an automated supervised learning algorithm. The overview is given in Fig. 4. In this paper, 96 out of 129 dataset were used to define MF (membership functions) of different fuzzy sets of input variables having deciding effects on hypertension risk assessment based on expert knowledge. To begin the tuning process, the inputs were given to fuzzy system and system outputs were compared against the expected/desired outputs as diagnosed from patients' record. If risk level found is different from doctor's diagnosis (as recorded in patient's record) then error is detected and the MFs of inputs are adjusted. This process is repeated until the system is tuned. Moreover, this process was repeated for all training data set in an automated manner. Finally, the system shows higher performance indicating in the confusion matrix.

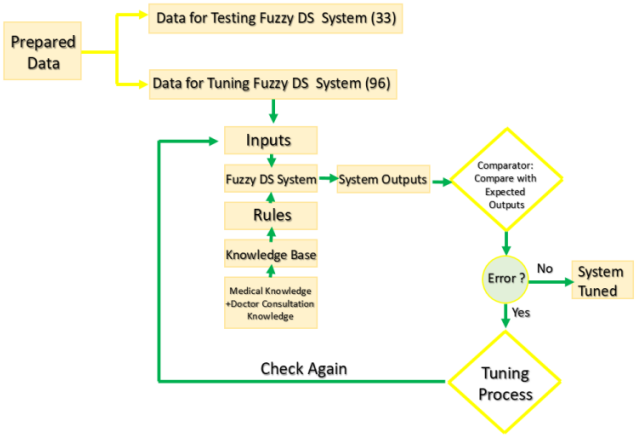
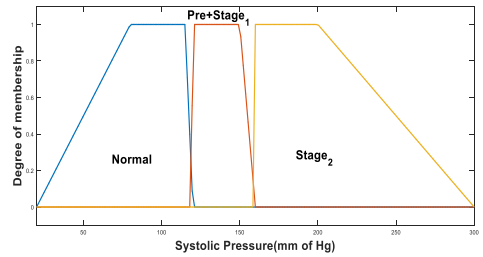
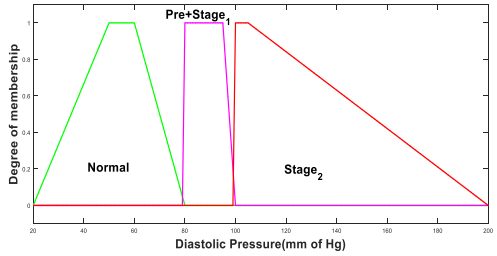


Fig. 4. Tuning Process



(a)



(b)

Fig. 5. Some Input Membership Functions (a) Systolic Pressure (mm of Hg); (b) Diastolic Pressure (mm of Hg)

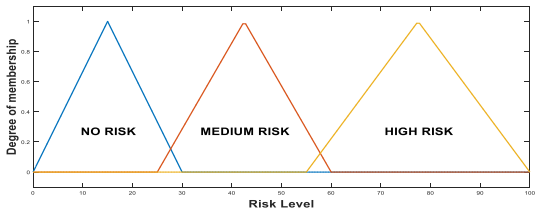


Fig. 6. Membership Functions of Risk Level

3. Results and Discussion

In this section, confusion matrix parameters: sensitivity, specificity and accuracy are calculated after tuning the as explained in preceding section. Moreover, it is noted that same 96 patient information is used in this section. Here, we consider Medium risk and High Risk as Positive in calculating confusion matrix, i. e., P = 19+71 = 90 and N = 6 in 96 patient information. So the desired/ expected parameters of confusion matrix: TP, TN, FP and FN should be 90, 6, 0 and 0 respectively to achieve maximum accuracy.

It is very important to note that the designed system, after tuning, yielded the exact desired output as diagnosed by doctors. Sensitivity, specificity and accuracy are calculated as follows:

Sensitivity = 90/ (90+0) = 100%;
Specificity=6/ (6+0) =100%;
Accuracy = (90+6)/ (90+6+0+0) =100%

The result signifies the proportion of actual positives is correctly identified and the proportion of actual negatives is correctly identified in the developed system. So, the above results show that the system has been tuned successfully. The following tables show initial results vs results after tuning.

Table 3.

Before Tuning (96 Data) Results from Confusion Matrix		
Sensitivity	Specificity	Accuracy
25%	98. 68%	83. 33%

Table 4.

After Tuning (96 Data) Results from Confusion Matrix		
Sensitivity	Specificity	Accuracy
100%	100%	100%

Confusion Matrix Calculation in Testing

In the testing process, 33 patient information is used that also includes all three types of patients. Numbers of high risk, medium risk and no risk patients are 20, 6 and 7 respectively. It means, H = 20; N = 7 and M = 6. The distribution is shown in Fig. 7.

Case Study-1(Patient ID-2018/00113)

According to patient's record, inputs were, X_1=110, X_2=60, X_3=60, X_4=20, X_5=0, X_6=0, X_7=0, X_8=0, X_9=0, X_{10}=0, X_{11}=0, X_{12}=0, X_{13}=0, X_{14}=0, X_{15}=0, X_{16}=0; and according to doctor's diagnosis, it was labeled as No Risk

(Expected output for this patient). When these values were fed to the designed decision support system, the output was recorded as 15, which was evaluated as NO RISK state as well. So the decision support system predicted risk level correctly.

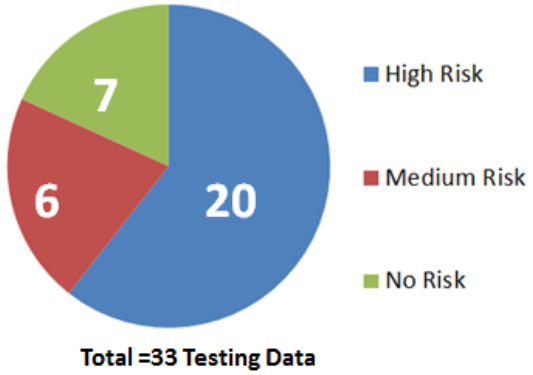


Fig. 7. Test Dataset

Case Study-2(Patient ID-P-2018/0087)

Inputswere,X_1=180, X_2=100, X_3=72, X_4=20, X_5=1, X_6=0, X_7=1, X_8=0, X_9=0, X_{10}=1, X_{11}=0. 5,X_{12}=0. 5, X_{13}=0, X_{14}=0, X_{15}=0, X_{16}=0; and Expected Output was High Risk. When these values were applied to the designed system it gave a risk score of 77 which was evaluated as "HIGH RISK". Thus, the system calculated risk level correctly.

Let's consider No risk as Negative Data in Calculating Confusion Matrix, N=7, so, TN=7. Also, we consider Medium risk and High Risk as Positive Data in Calculating Confusion Matrix, P=20+6=26, so, TP=26.

Therefore, P=26; N=7; TP=26; TN=7; FP=0; FN=0

Sensitivity=26/ (26+0) =100%;
Specificity=7/ (7+0) =100%;
Accuracy = (26+7)/ (26+7+0+0) =100%

So, the decision support system has been tested successfully.

Table V.

Test Data Set Results from Confusion Matrix		
Sensitivity	Specificity	Accuracy
100%	100%	100%

4. Conclusion

A hypertension risk prediction system has been designed using fuzzy logic. Like clinical practice, a number of symptoms and risk factors have been considered in consultation with experienced cardiologists and consultants to assess the risk. Mamdani fuzzy inference system is used in order to incorporate expert knowledge into the system in the form of IF-THEN rules. The rules are designed in similar way as cardiologists and consultants usually make decision and treat patient. Unlike Sugeno and other fuzzy inference systems, the rule-base of Mamdani system, comprising of a number of IF-THEN rules, is clearly understandable to doctors and clinical people. An automated 'supervised learning' type procedure is developed to tune the fuzzy risk prediction system. First hand data, collected from a local hospital are used to tune the system. Moreover, a separate set of first hand patient data, collected from same hospital, are used to test and validate the designed system. The performance of the system is assessed in terms of standard parameters as used in similar research. After the tuning, results/ decisions made by the designed system seem to match with the decisions made by cardiologists and consultants although number of patient considered in the research is not large. However, such fuzzy logic based decision support system can help doctors as well as general people for early assessment of this disease which will make the diagnosis process faster and reduce the risk of wrong diagnosis. Following this work, the authors intend to investigate more comprehensive decision support system for a wide range of diseases including heart diseases.

Acknowledgement

The research was supported by Fab Lab DU, University of Dhaka, Dhaka, Bangladesh. The authors would like to thank "Euro-Bangla Heart Hospital Limited" for prividing data.

References

1. S. S. Davidson, D. E. Newby, N. R. Grubb, "Cardiology", in S. H. Ralston, Ian D. Penman, W. Mark, J. Strachan, and R. P. Hobson, Davidson's Principles and Practice of Medicine, Elsevier, 441-544, 2018.
2. https://www. who. int/health-topics/cardiovascular-diseases/#tab=tab_1 [accessed: February 26, 2020, 6:50pm].
3. https://www.worldlifeexpectancy.com/bangladesh-coronary-heart-disease [accessed: February 26, 2020, 6:50pm].
4. S. Kenchaiah, and R. S. Vasan, "Heart failure in women–insights from the Framingham Heart study", *Cardiovascular Drugs and Therapy*, vol. 29 no. 4, pp. 377-390, 2015.
5. SPRINT Research Group. "A randomized trial of intensive versus standard blood-pressure control", *New England Journal of Medicine*, vol. 373 no. 22, pp. 2103-2116, 2015.
6. F. Sparrenberger, F. T. Cichelero, A. M. Ascoli, F. P. Fonseca, G. Weiss, O.,Berwanger, S. C. Fuchs, L. B. Moreira and F. D. Fuchs, "Does psychosocial stress cause hypertension? A systematic review of observational studies", *Journal of Human Hypertension*, vol. 23 no. 1, pp. 12, 2009.
7. M. J. O'Donnell, S. L. Chin, S. Rangarajan, D. Xavier, L. Liu, H. Zhang, P. Rao-Melacini, X. Zhang, P. Pais, S. Agapay, and P. Lopez-Jaramillo, "Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study", *The Lancet*, vol. 388 no. 10046, pp-761-775, 2016.
8. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University-Computer and Information Sciences*, vol. 24, no. 1, pp. 27-40, 2012.
9. P. Melin, I. Miramontes, and G. Prado-Arechiga, "A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk

diagnosis", *Expert Systems with Applications*, vol. 107, pp. 146-164, 2018.

10. M. Pulido, P. Melin, and G. Prado-Arechiga, "A new model based on a fuzzy system for arterial hypertension classification". in *Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications* (pp. 319-327). Springer, Cham, 2018.
11. G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis", *Evolutionary Intelligence*, vol. 13, no. 2, pp. 185-196, 2020.
12. S. Thukral and V. Rana, "Versatility of fuzzy logic in chronic diseases: A review", *Medical Hypotheses*, vol. 122, pp. 150-156, 2019.
13. G. Gürsel, "Healthcare, uncertainty, and fuzzy logic", *Digital Medicine*, vol. 2, no.3, p. 101, 2016.