

RESEARCH ARTICLE

Open Access

A genetic fuzzy system for unstable angina risk assessment

Wei Dong¹, Zhengxing Huang^{2*}, Lei Ji³ and Huilong Duan²

Abstract

Background: Unstable Angina (UA) is widely accepted as a critical phase of coronary heart disease with patients exhibiting widely varying risks. Early risk assessment of UA is at the center of the management program, which allows physicians to categorize patients according to the clinical characteristics and stratification of risk and different prognosis. Although many prognostic models have been widely used for UA risk assessment in clinical practice, a number of studies have highlighted possible shortcomings. One serious drawback is that existing models lack the ability to deal with the intrinsic uncertainty about the variables utilized.

Methods: In order to help physicians refine knowledge for the stratification of UA risk with respect to vagueness in information, this paper develops an intelligent system combining genetic algorithm and fuzzy association rule mining. In detail, it models the input information's vagueness through fuzzy sets, and then applies a genetic fuzzy system on the acquired fuzzy sets to extract the fuzzy rule set for the problem of UA risk assessment.

Results: The proposed system is evaluated using a real data-set collected from the cardiology department of a Chinese hospital, which consists of 54 patient cases. 9 numerical patient features and 17 categorical patient features that appear in the data-set are selected in the experiments. The proposed system made the same decisions as the physician in 46 (out of a total of 54) tested cases (85.2%).

Conclusions: By comparing the results that are obtained through the proposed system with those resulting from the physician's decision, it has been found that the developed model is highly reflective of reality. The proposed system could be used for educational purposes, and with further improvements, could assist and guide young physicians in their daily work.

Keywords: Unstable angina risk assessment, Fuzzy association rule mining, Genetic algorithm

Background

Unstable Angina (UA) is a kind of chest discomfort or pain that occurs in a continuous and unpredictable way [1,2]. The unstable pain can result from the disruption of an atherosclerotic plaque in narrowed coronary vessels with lessened flexibility, embolization and vasospasm. As a major type of Cardiovascular Disease (CVD), UA lays its symptoms between stable angina and acute myocardium infarction and a further sudden death [2]. While the risk of UA is high, the population of UA is huge, especially for aged people and those with associated disease such as

hypertension and diabetes [3]. To this end, reliable assessment of risk levels for individual UA patients will be of significant value and interest.

A number of models for UA risk assessment have been proposed in literature. Most of these models are derived from databases of clinical trials, e.g., the Thrombolysis in Myocardial Infarction (TIMI) [4], platelet glycoprotein IIb/IIIa in unstable angina: Receptor Suppression Using Integrilin (PURSUIT) [5], and the Global Registry of Acute Coronary Events (GRACE) [6], etc. They use standard patient features that are part of the routine medical evaluation of UA patients, and lead to a score to define prognostic groups [2]. Although there are many benefits related to the design and use of these prognostic models, a number of studies have highlighted possible shortcomings [2,4]. One serious drawback is that existing models

*Correspondence: zhengxing.h@gmail.com

²College of Biomedical Engineering and Instrument Science, Zhejiang University, 310008, Zhou Yiqing Building 510, Zheda road 38#, Hangzhou, Zhejiang, China

Full list of author information is available at the end of the article

lack the ability to deal with the intrinsic uncertainty about patient features utilized in UA risk assessment. Note that vagueness is fundamental and indispensable aspects of knowledge, so as in many practical problems, the experts face vagueness in feature vectors. According to Bellman and Zadeh “much of the decision making in the real world takes place in an environment in which the goals, the constraints, and consequences of possible actions are not known precisely” [7]. Regarding UA risk assessment, many patient features are vague, and not easy to be handled by existing models. It is, therefore, necessary to develop a new UA risk assessment model to deal with vague information.

In this paper, a novel UA risk assessment model has been developed using fuzzy set theories. The proposed model represents patient features with fuzzy sets and then extracts useful information with a descriptive rule induction approach based on fuzzy systems. To derive fuzzy rules from data, the proposed model employs genetic algorithms (GAs) to learn rule base from the collected data-set. GAs are search algorithms based on natural genetics that provide robust search capabilities in complex spaces [8]. The hybridization between fuzzy systems and GAs, called genetic fuzzy system (GFS), has attracted considerable attention in the computational intelligence community [9-12]. Our main goal is to develop a novel GFS such that we derive from clinical data-set a set of assessment rules, which has good interpretability before determining an efficient assessment model in order to get high accuracy of UA risk stratification. Since the accuracy of an assessment model can be largely affected by processing vague patient features, this paper also discusses a clustering-based method for patient feature partitioning.

This paper is organized as follows. Section ‘Preliminary’ presents preliminary knowledge used in this paper. Section ‘Method’ describes the development of the genetic-fuzzy system for UA risk assessment. Experimental studies of the performance of the proposed approach are presented in Section ‘Results and discussion’. Section ‘Conclusion’ concludes the paper.

Preliminary

Let $D = \{\sigma_1, \dots, \sigma_n\}$ be a patient data-set consisting of a finite set of UA patient cases. Let $A = \{a_1, \dots, a_n\}$ represent all patient features that appear in D and $Class = \{low-risk, medium-risk, high-risk\}$ be a set of UA risk levels. Each feature a may have a categorical or numerical underlying domain, denoted $dom(a)$. Each patient case σ ($\sigma \in D$) contains values of some patient features from A . Let $\sigma(a)$ ($\sigma(a) \in dom(a)$) be the target feature value for the patient case σ for feature a .

For example, Table 1 shows an example patient data-set, which consists of five patient cases. Each case contains

Table 1 An example patient data-set

	Age (year)	Sex	Smoking	Heart events recently	Physician assessment
σ_1	74	Male	No	Yes	Medium-risk
σ_2	81	Female	No	Yes	Medium-risk
σ_3	74	Male	Yes	Yes	High-risk
σ_4	71	Male	No	Yes	High-risk
σ_5	76	Female	No	No	Low-risk
σ_6	67	Male	Yes	No	Medium-risk

The cases are simplified information extraction from patient records of Chinese PLA general hospital.

1 numerical patient features (i.e., *age*) and 3 categorical patient features (i.e., *sex*, *smoking*, and *has event recently*).

For numerical patient feature a ($a \in A$), let $\{l_a^1, l_a^2, \dots, l_a^m\}$ be a set of linguistic terms defined over a . Let $\mu_{a,j}(\sigma(a))$ be the membership degree on the value of a feature a of the patient case σ to the fuzzy set corresponding to the linguistic label l_a^j for this feature a . Note that the degree of membership of each value of a in any of the fuzzy sets specified for a is directly based on the evaluation of the membership function of the particular fuzzy set with the value of a as input. The fuzzy partition of $dom(a)$ is composed of $\{l_a^1, \dots, l_a^m\}$ that satisfies $\sum_{j=1}^m \mu_{a,j}(x) = 1, \forall x \in dom(a)$.

In this study, we employ fuzzy rules of the following form [9,12]:

$$r : Cond \rightarrow Class \quad (1)$$

where *Cond* is the antecedent part of the rule, and *Class* is the consequent part of the rule. For example, a fuzzy rule can be expressed as:

$$R : \text{IF } a_1 \text{ is } (l_a^1 \text{ or } l_a^3) \text{ and } a_5 \text{ is } l_a^4 \text{ THEN } low-risk. \quad (2)$$

It must be noted that any subset of the complete set of patient features, with any combination of linguistic labels related to the operators *and* and *or*, can take part in the rule antecedent. For this kind of fuzzy rule, we say that a patient case σ supports the antecedent part of a rule r if

$$APC(\sigma, r) = \frac{1}{n} \sum_{i=1}^n \max\{\mu_{a_i,1}(\sigma(a_i)), \dots, \mu_{a_i,m}(\sigma(a_i))\} > 0 \quad (3)$$

where $\mu_{a_i,j}(\sigma(a_i))$ is the membership degree of patient feature a_i for σ to the fuzzy set corresponding to the linguistic label $l_{a_i}^j$ for a_i ; and *APC* is the antecedent part compatibility between a patient case and the antecedent part of a fuzzy rule. For the categorical features, the degrees of membership are zero or one.

For a patient case σ , the support degree of σ by a specific rule r is calculated as follows:

$$Supp(\sigma, r) = \begin{cases} APC(\sigma, r) & \text{the actual risk level of } \sigma \text{ is equal to the class of } r \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In general, a fuzzy rule can be considered to be a classification rule if the antecedent contains fuzzy item sets, and the consequent part contains only one class label such as *low-risk*, *medium-risk*, or *high-risk* in this study. A fuzzy rule $r : Cond \rightarrow Class$ could be measured directly in terms of support and confidence as follows:

$$Support(r) = \frac{\sum_{\sigma \in D} Supp(\sigma, r)}{|D|} \quad (5)$$

$$Confidence(r) = \frac{\sum_{\sigma \in D} Supp(\sigma, r)}{\sum_{\sigma \in D} APC(\sigma, r)} \quad (6)$$

Method

In this section, we describe the process of utilizing GFS to develop an intelligent system for the problem of UA risk assessment. As shown in Figure 1, the proposed method consists of three steps. At first, all the numerical patient features of the data-set are given as input for the fuzzy clustering module for calculating the membership functions. Then, calculated function values are given to the rule generation module for obtaining UA risk assessment rules. Based on the derived rules, a classification model for UA risk assessment is generated.

The case study was performed in the Cardiology Department at the Chinese PLA General hospital. Prior approval was obtained from the data protection committee of the hospital to conduct the study. We state that the patient data was anonymized in this study and in the Method section of this paper.

Fuzzy clustering for numerical feature discretization

One of the most important steps in UA risk assessment is to deal with the intrinsic uncertainty about the variables

utilized. As described in [13], fuzzy set is a common tool for facilitating the interpretation of rules in linguistic terms, and avoiding unnatural boundaries in the partitioning of the variable domains. It is especially useful in clinical settings where the boundaries of a piece of information used may not be clearly defined. Regarding our task of UA risk assessment, the quality of the results produced relies quite crucially on the appropriateness of fuzzy sets to the given patient features. So, fuzzy sets must be consistent with the values of the corresponding feature.

Fuzzy sets can be provided by physicians. However, the provided fuzzy sets by physicians may not be suitable for mining fuzzy association rules from data-set. Also, it is extremely difficult for physicians to estimate the most appropriate fuzzy sets. In order to cope with these problems, we first concentrate on how fuzzy sets of the given features are determined automatically from the collected data-set. Clustering techniques are usually employed as a preprocessing step to partition numerical features [14]. In this study, we employed a hierarchical agglomerative clustering [15-17] algorithm to partition numerical features.

As shown in algorithm 1, hierarchical agglomerative clustering begins with each value as a separate cluster and merges them into successively larger clusters. The process is repeated until the similarity between any pair of clusters is less than a threshold value ϵ . Consequently, the algorithm builds a structure called dendrogram, i.e., a tree illustrating the merging process and intermediate clusters. Similarity between two clusters c_1 and c_2 can be measured as follows:

$$sim(c_1, c_2) = \frac{1}{|c_1||c_2|} \sum_{x \in c_1} \sum_{y \in c_2} |x - y| \quad (7)$$

where $|c_1|$ and $|c_2|$ are the number of clusters c_1 and c_2 , respectively.

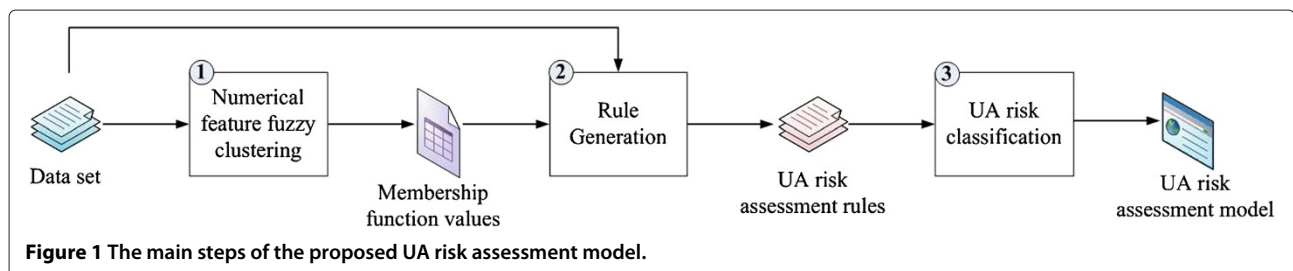


Figure 1 The main steps of the proposed UA risk assessment model.

Algorithm 1 Hierarchical Agglomerative Clustering

- 1: **Input:**
 2: D_a is a value set of a specific patient feature a
 3: δ is a cluster similarity measure
 4: ε is a merging threshold value
 5: **Output:**
 6: C is the set of clusters
 7: **Steps:**
 8: Let $C = \{\text{initial clusters}\}$ where each value in D_a forms an initial cluster
 9: Repeat
 10: Let (c_1, c_2) be a pair of clusters which are most similar in C
 11: If $\text{sim}(c_1, c_2) \geq \varepsilon$ then
 12: Let $c_3 \leftarrow c_1 \cup c_2$
 13: Add c_3 into C
 14: Remove c_1 and c_2 from C
 15: End If
 16: Until cannot find c_1 and c_2 with $\text{sim}(c_1, c_2) \geq \varepsilon$
 17: **End Procedure**

This way, the values of each patient feature in the data-set are distributed over a set of derived clusters using Algorithm 1. For each patient feature, the centroids of the clusters are the set of midpoints of the fuzzy sets. To illustrate the process, suppose we want to find fuzzy sets for a specific patient feature a , which is quantitative with a range from $\min(\text{dom}(a))$ to $\max(\text{dom}(a))$. Let $\{v_a^1, \dots, v_a^m\}$ be the set of mid-points of the fuzzy sets for a . As a result, the derived fuzzy sets will have the following ranges: $[v_a^0, v_a^1], [v_a^1, v_a^2], \dots, [v_a^{m-1}, v_a^m]$ and $[v_a^m, v_a^{m+1}]$, where $v_a^0 = \min(\text{dom}(a))$, and $v_a^{m+1} = \max(\text{dom}(a))$.

After the fuzzy sets of each numerical feature are obtained, the corresponding membership function can be generated for each fuzzy set. In this study, we used membership functions of both semi-trapezoidal shape and triangular shape because they are in general the most appropriate shapes and the most widely used in fuzzy systems. For example, for the fuzzy set with a range from v_a^0 to v_a^1 , the membership function is given by

$$\mu_{a,0}(x) = \begin{cases} 1 & \text{if } x < v_a^0 \\ \frac{v_a^1 - x}{v_a^1 - v_a^0} & \text{if } v_a^0 \leq x \leq v_a^1 \\ 0 & \text{if } x > v_a^1 \end{cases} \quad (8)$$

For each fuzzy set with midpoint v_a^j , where $1 \leq j \leq m$, the membership function is given by

$$\mu_{a,j}(x) = \begin{cases} \frac{x - v_a^{j-1}}{v_a^j - v_a^{j-1}} & \text{if } v_a^{j-1} \leq x \leq v_a^j \\ \frac{v_a^{j+1} - x}{v_a^{j+1} - v_a^j} & \text{if } v_a^j \leq x \leq v_a^{j+1} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

And for the fuzzy set with a range from v_a^m to v_a^{m+1} , the membership function is given by

$$\mu_{a,m+1}(x) = \begin{cases} 1 & \text{if } x > v_a^{m+1} \\ \frac{x - v_a^m}{v_a^{m+1} - v_a^m} & \text{if } v_a^m < x \leq v_a^{m+1} \\ 0 & \text{if } x \leq v_a^m \end{cases} \quad (10)$$

For example, given a numerical feature, say *age* with three different ranges, i.e., [30, 56], [56, 74], and [74, 87]. The values of *age* range from 30 to 87, and can be classified into four fuzzy sets, as shown in Figure 2.

Fuzzy association rule mining

This subsection presents a GFS for mining fuzzy association rules from a data-set. The proposed system uses fuzzy rule format defined in Equation (2), which offers a flexible structure to the rules, allowing each patient feature to take more than one value and facilitating the extraction of general UA risk assessment rules.

Chromosome representation

As mentioned above, a US risk assessment rule r consists of an antecedent part *Cond* and a consequent part *Class*. In this study, we code the antecedent part *Cond* of r as one chromosome consisting of a set of segments. Each segment corresponds to a specific patient feature. The set of possible values for the categorical features is that indicated by the problem, and for numerical features, it is the set of linguistic terms determined by the clustering method presented above. The consequent part *Class* of r is prefixed to one of the possible values of risk levels, i.e., *high-risk*, *medium-risk*, and *low-risk*.

Table 2 describes a representation for a rule with numerical and categorical features for the values of a specific risk level *high-risk*. Note that a bit for each one of the possible values of each feature is stored. In this way, if the value of the corresponding element is 0, it indicates that the value is not used in the rule. Otherwise, if the value is 1, it indicates that the corresponding value is included. If a rule contains all the elements corresponding to a feature of the value 1, or all of them contain the value 0, this indicates

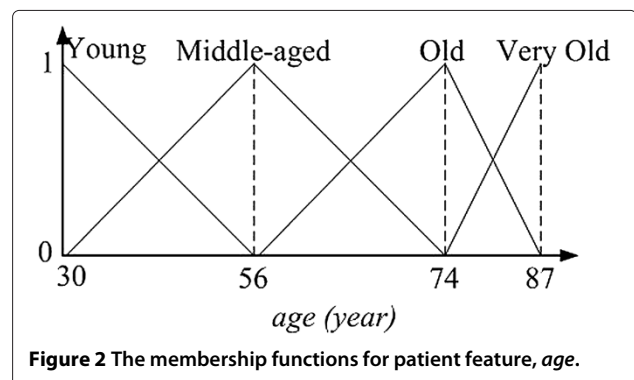


Figure 2 The membership functions for patient feature, *age*.

Table 2 Representation of a fuzzy rule with numerical and categorical features in UA risk assessment

Patient feature	Age				Sex		Smoking		Has heart events recently		
Linguistic label	Young	Middle-aged	Old	Very old	Male	Female	Yes	No	Yes	No	
Rule	0	0	1	1	0	0	1	0	1	0	High-risk

that this feature has no relevance for the information contributed in the rule, and so this feature is ignored. In these cases, the feature does not take part in the rule. For example, as shown in Table 2, the rule is represented by a binary string $\langle(0011)(00)(10)(10) : high-risk\rangle$, where parentheses are to separate segments, and “:” is to separate the IF part and the THEN part of the rule. In this example, a_1 has four possible values and $a_2, a_3,$ and a_4 have two possible values. Note that a_2 does not take part in the rule as a_2 takes none of its values, and thus a_2 is irrelevant for the rule. This binary string can be interpreted as the following rule: “IF age is (old or very old) and smoking is true and has heart events recently is true THEN UA risk is high”.

Fitness function and selection process

The objective of this step is to find the accurate and general rules for UA risk assessment. Thus, given a specific rule r , the GA method uses the composite fitness function consisting of support and confidence in the following way:

$$Fitness(r) = \frac{Support(r) + Confidence(r)}{2} \tag{11}$$

The objective of the fitness function is defined as the composite measure of support and confidence. This composite measurement provides an effective selection environment which balances the accuracy and generality of the rules.

Three operators, i.e., selection, crossover, and mutation, are applied in the proposed GA method to generate the offspring population, which are illustrated as follows:

- The selection procedure is used for evolution where two individuals are selected randomly from the current population and used for crossover and mutation operators. During each generation, individuals with higher fitness values survive while those with lower fitness values are destroyed.
- Two parents are selected and recombined according to the predefined crossover probability during crossover. In this work, one point crossover is applied due to its simplicity, which can randomly select different cutoff points for each parent to generate offspring rule sets.
- Each element in the chromosome is applied to mutation with a predefined mutation probability. The value of a randomly selected element is converted to 0 if its value is 1, and vice versa. Elimination of existing rules and addition of new rules can also be used as

mutation operations. As a result, the number of rules in the rule sets string can be changed accordingly.

During GA operations, redundant rules might be produced. For example, we say a rule “If Age is (Young) and ST is (Low) then Risk is (Low)” is redundant w.r.t the other rule “If Age is (Young OR Middle-ages) and ST is (Low) then Risk is (Low)”, if both rules have the same support degree on a data-set. Thus, the proposed algorithm must check the rule sets and maintains single among all the rules, to guarantee the consistency of fuzzy systems. The stopping criterion of the proposed algorithm is the number of generations. The scheme of the proposed algorithm is shown in Algorithm 2.

Algorithm 2 The GA-based rule mining algorithm for UA risk assessment.

- 1: **Input:**
- 2: UA data-set, algorithm parameters
- 3: **Output:**
- 4: \mathfrak{R} is the set of derived UA risk assessment rules
- 5: **Steps:**
- 6: For each target class c of UA risk levels, apply the following steps
- 7: **Step 1:** Initialize P_c as a set of UA risk assessment rules
- 8: **Step 2:** Fitness evaluation for each individual $r \in P_c$
- 9: **Step 3:** While not (termination condition) do
- 10: **Step 3.1:** Select r_{p_1} and r_{p_2} from P_c with high fitness value
- 11: **Step 3.2:** Apply crossover to r_{p_1} and r_{p_2} for obtaining r_{c_1} and r_{c_2}
- 12: **Step 3.3:** Apply mutation to r_{c_1} and r_{c_2}
- 13: **Step 3.4:** Check if r_{c_1} and r_{c_2} have been stored in P_c , if not
- 14: **Step 3.4.1:** Update P_c by deleting the worst solution and adding r_{c_1} and r_{c_2}
- 15: **Step 4:** Add P_c to rule pool \mathfrak{R}
- 16: Output selected rule set from \mathfrak{R}
- 17: **End Procedure**

Taking the data set shown in Table 1 as an example, and assuming the population size as 50, the number of generation as 1000, the crossover rate as 0.5, and the mutation

Table 3 Rules derived from the example patient data-set shown in Table 1

# Rule	Rule	Support	Confidence	Fitness
1	IF Age (Young) AND Heart Events Recently is (False) THEN Risk is Low	0.083	0.5	0.291
2	IF Age (Young) AND Sex is (Female) AND Smoke is (False) THEN Risk is Low	0.11	0.33	0.22
3	IF Age is (Young OR Meddle-aged) AND Smoke is (False) AND Heart Events Recently is (True) THEN Risk is Medium	0.244	0.514	0.378
4	IF Age is (Old) AND Smoke is (False) THEN Risk is Medium	0.339	0.465	0.402
5	IF Age is (Old OR Very-Old) AND Smoke is (True) THEN Risk is High	0.269	0.377	0.323
6	IF Age is (Meddle-aged OR Old OR Very-Old) AND Sex is (Male) AND Heart Events Recently is (True) THEN Risk is High	0.324	0.455	0.390

rate as 0.2 in the proposed genetic algorithm, we can obtain an example rule-set, as shown in Table 3.

UA risk assessment model

Based on the derived rule set, we can generate a classification model for UA risk assessment. Formally, let \mathfrak{R} be the set of derived UA risk assessment rules. For each $r : Cond \rightarrow Class (r \in \mathfrak{R})$, the score value of the target class ($class \in \{low-risk, medium-risk, high-risk\}$) of r with respect to a given patient case σ can be assessed, by using the following equation:

$$v_{class} = \sum_{r \in \mathfrak{R}} Confidence(r)\beta(r, \sigma) \tag{12}$$

where $Confidence(r)$ is the confidence value of rule r , and $\beta(r, \sigma)$ is the firing strength of the input patient case σ on the antecedent part of rule r .

The firing strength $\beta(r, \sigma)$ is defined as

$$\beta(r, \sigma) = \sum_{\langle (a, l_a^1), \dots, (a, l_a^m) \rangle \in Cond} \max\{\mu_{a,1}(\sigma(a)), \dots, \mu_{a,m}(\sigma(a_i))\} \tag{13}$$

where $\mu_{a,j}$ represents the fuzzy membership function for the pair of a and the linguistic term l_a^j in the $\langle (a, l_a^1), \dots, (a, l_a^m) \rangle (1 \leq j \leq m)$ in the antecedent part $Cond$ of rule r .

With respect to the target values of risk levels, i.e., *low-risk*, *medium-risk*, *high-risk*, the corresponding scores v_l, v_m, v_h can be generated based on Equation (12). And the risk level with the top score in the scoring vector v will be the predicted risk level for the patient case σ .

$$\widehat{Class}_\sigma = argmax(\{v_l, v_m, v_h\}) \tag{14}$$

Taking σ_1 shown in Table 1 as an example, and using the derived rule-set shown in Table 3, the score values for σ_1 are calculated as $v_l = 0.4$, $v_m = 2.088$, and $v_h = 1.745$, by Equation (12). Thus, the predicted risk level for σ_1 is *medium-risk*.

Table 4 Patient features utilized in UA risk assessment

Numerical feature	
Name	Fuzzy sets
Age	{Young, middle-aged, old, very old}
Systolic Blood Pressure (SBP)	{Low, medium, high}
Diastolic Blood Pressure (DBP)	{Low, low-medium, high-medium, high}
Creatinine	{Low, medium, high}
AST	{Low, medium, high}
LDH	{Low, medium, high}
CK	{Low, medium, high}
CK MB	{Very low, low, medium, high}
CnT	{Low, high}
Categorical feature	
Name	Crisp sets
Sex	{Male, female}
Smoke	{Yes, no}
Atrial premature beat	{Yes, no}
Significant change in ST in ECG	{Yes, no}
Has heart events recently	{Yes, no}
Anamnesis of coronary heart disease	{Yes, no}
Anamnesis of renal insufficiency	{Yes, no}
Anamnesis of bleeding	{Yes, no}
Anamnesis of diabetes	{None, type 1, type 2}
Anamnesis of hyperlipaemia	{Yes, no}
Anamnesis of hypertension	{None, level I, level II, level III}
Has aspirin in recent 7 days	{Yes, no}
Has percutaneous coronary intervention	{Yes, no}
Lungs event as expectoration with blood	{Yes, no}
Change in T	{Yes, No}
Heart event by drinking	{Yes, no}
Irregular pulse	{Yes, no}

Results and discussion

To evaluate the feasibility of the presented methods, a clinical case study is conducted through the cooperation with the Cardiology Department of the Chinese PLA General Hospital. The data-set collected from the hospital consists of 54 patient cases. The target classes of UA risk levels include: *low-risk*, *medium-risk*, and *high-risk*. Physicians that evaluated these cases are experienced clinicians working for the hospital with 10 years of working experience on average. As a result, 16 cases are classified into the *low-risk* group, 33 cases are classified into the *medium-risk* group, and 5 cases are classified into the *high-risk* group, respectively. Patient features (9 numerical features and 17 categorical features) that appear in the data-set are shown in Table 4. These features are regularly recorded in UA treatment practice.

All experiments were performed on a Lenevo Compatible PC with an Intel Pentium IV CPU 2.8 GHz, 4G byte main memory running on Microsoft Windows 7. The algorithms were implemented using Microsoft C#. A 10-fold cross-validation was performed to evaluate the proposed method, by using a 90% of the data-set as the training set, and the remaining 10% as the validation set.

To reduce variability, 10 rounds of this validation process were performed by using different partitions.

For the first step of patient feature discretization, we applied Hierarchical Agglomerative Clustering method to each numerical feature to generate a set of fuzzy sets. The derived fuzzy sets of input numerical patient features are shown in Figure 3.

To mine fuzzy association rules, we have taken the population size as 100, the number of generation as 1000, the crossover rate as 1.0, and the mutation rate as 0.2 in the proposed genetic algorithm. By using these parameters, we run our genetic algorithm for each target class of UA risk level, one by one, to obtain a set of fuzzy rules for a given class.

Table 5 shows the rules obtained, which have the best fitness values for the target classes of UA risk level (i.e., *low-risk*, *medium-risk*, and *high-risk*). In this table, the number of patient features involved in each rule (# of Feature.), and the Support and Confidence of each rule are shown. The values of Support and Confidence are between zero and one. High values in support means that the rule covers most of patient cases which are categorized into the class, and high values in confidence means that

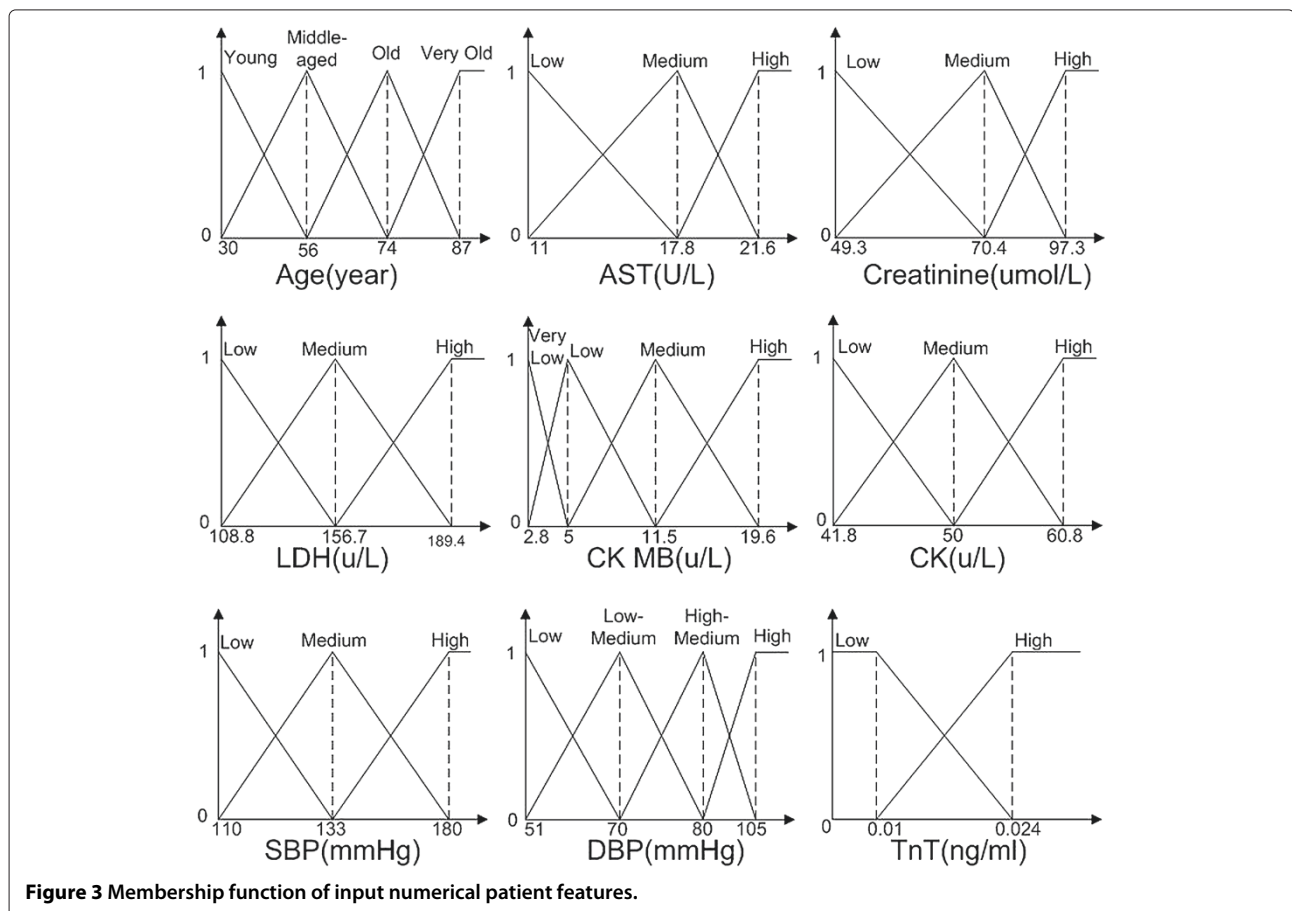


Figure 3 Membership function of input numerical patient features.

Table 5 Results for low-risk, medium-risk, and high-risk

Risk level	# of feature	Support	Confidence
Low	9	0.204	0.375
	8	0.195	0.380
Medium	10	0.448	0.633
	10	0.449	0.631
High	10	0.067	0.114
	9	0.064	0.115

the rule has few negative patient cases [12]. Note that the knowledge discovered for each target value of risk level is understandable by physicians due to the use of fuzzy logic and the low number of rules and conditions in the rule antecedents (below 40% of 26 patient features). Tables 6, 7 and 8 show the rules obtained which have the best fitness values corresponding to the target classes of risk level.

Now, acquiring fuzzy rule base, it is possible to complete UA risk assessment through the proposed classification model. As we mentioned above, the ensemble of fuzzy rules and the proposed classification model perform the role of a mathematical function to obtain the system output. This output is the stratification of UA risk. This way, for each patient whose data are informed as system inputs, the most likely risk level for that patient is generated.

The comparison between the proposed model and physicians' decisions is done for each partition of dataset. The comparison is given in Figure 4. The proposed system made the same decisions as the physician in 46 (out of a total of 54) tested cases (85.2%). From Figure 4, the proposed model did not predict the risk well in the case of the fifth (the physician assessment is medium-risk while the proposed model assessment is low-risk), the twenty third (the physician assessment is low-risk while the proposed model assessment is medium-risk), the twenty ninth (the physician assessment is medium-risk while the proposed model assessment is low-risk), the thirty third (the physician assessment is low-risk while the proposed model assessment is medium-risk), the thirty fourth (the physician assessment is medium-risk while the

Table 6 Rules for low-risk

# Rule	Rule
1	IF Age is (Young OR Meddle-aged) AND AST is (Low OR Medium) AND Creatinine is (Low OR Medium) AND CK is (Low OR Medium) AND LDH is (Low OR Medium) AND CK MB is (Medium) AND has PCI is (No) AND Hypertension is (No OR Level I OR Level II) AND Arrhythmia is (No) THEN Risk is Low
2	IF Age is (Young OR Meddle-aged) AND AST is (Low OR Medium) AND Creatinine is (Low OR Medium) AND CK is (Low OR Medium) AND LDH is (Low OR Medium) AND CK MB is (Medium) AND has PCI is (No) AND Hypertension is (No OR Level I OR Level II) THEN Risk is Low

Table 7 Rules for medium-risk

# Rule	Rule
1	IF Age is (Old) AND Creatinine is (Low OR Medium) AND CK is (Low OR Medium) AND LDH is (Medium) AND SBP is (Medium OR High) AND Heart Events Recently is (Yes) AND Aspirin in Recently 7 days is (No) AND has PCI is (Yes) AND Hypertension is (Level I OR Level II OR Level III) AND Hyperlipaemia is (No) THEN Risk is Medium
2	IF Age is (Middle-aged OR Old) AND Creatinine is (Low OR High) AND CK is (Low OR High) AND LDH is (Medium) AND SBP is (Medium OR High) AND Heart Events Recently is (Yes) AND Aspirin in Recently 7 days is (No) AND has PCI is (Yes) AND Hypertension is (Level I OR Level II OR Level III) AND Hyperlipaemia is (No) THEN Risk is Medium

proposed model assessment is low-risk), the forty (the physician assessment is medium-risk while the proposed model assessment is low-risk), the forty second (the physician assessment is medium-risk while the proposed model assessment is low-risk), and the forty seventh (the physician assessment is medium-risk while the proposed model assessment is low-risk).

Furthermore, we measure the accuracy of the proposed approach using the sum of two performance measures: sensitivity (probability that the test correctly classifies a case with a specific risk level) and specificity (probability of correctly classifying a case without a specific risk level).

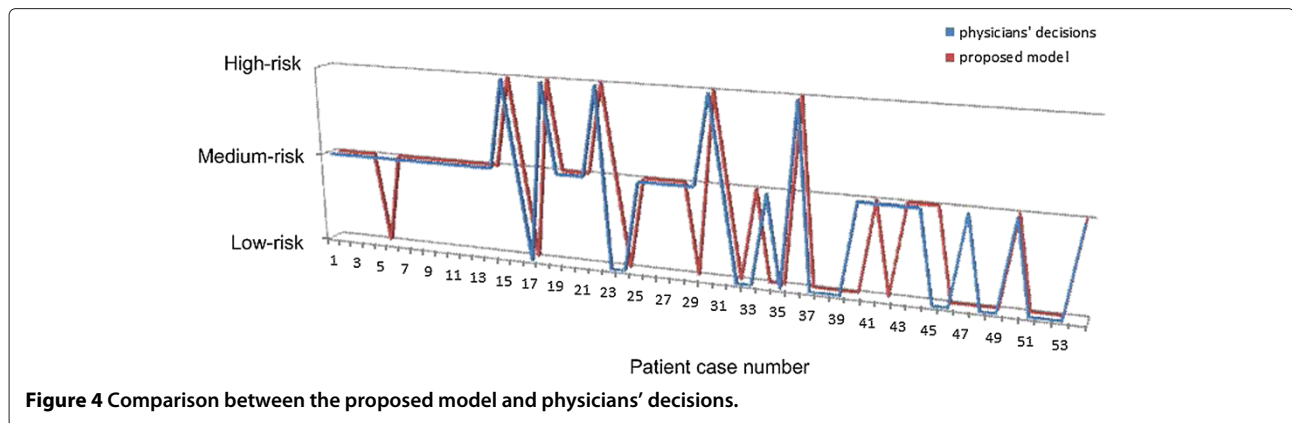
$$\text{Sensitivity}(v_{class}) = \frac{\sum_{i=1}^n TP(\sigma_i)}{\sum_{i=1}^n TP(\sigma_i) + \sum_{i=1}^n FN(\sigma_i)} \quad (15)$$

$$\text{Specificity}(v_{class}) = \frac{\sum_{i=1}^n TN(\sigma_i)}{\sum_{i=1}^n TN(\sigma_i) + \sum_{i=1}^n FP(\sigma_i)} \quad (16)$$

where $v_{class} \in \{low-risk, medium-risk, high-risk\}$; TP is the set True Positive, patient cases with the specific risk level v_{class} classified correctly; FN is the set False Negative, patient cases with the specific risk level v_{class} classified as other risk levels; TN is the set True Negative, patient cases without the specific risk level v_{class} classified; FP is the

Table 8 Rules for high-risk

# Rule	Rule
1	IF Age is (Old OR Very-Old) AND AST is (High) AND Creatinine is (High) AND CK is (Medium) AND LDH is (Medium OR High) AND CK MB is (High) AND SBP is (Low OR Medium) AND Smoke is (Yes) AND Hypertension is (Level II OR Level III) AND Diabetes is (Type 1 OR Type 2) THEN Risk is High
2	IF Age is (Old OR Very-Old) AND AST is (High) AND Creatinine is (High) AND CK is (Medium) AND LDH is (Medium OR High) AND CK MB is (High) AND SBP is (Low OR Medium) AND Hypertension is (Level II OR Level III) AND Diabetes is (Type 1 OR Type 2) THEN Risk is High



set False Positive, patient cases without the specific risk level v_{class} classified as v_{class} . In Table 9 the sensitivity and specificity obtained for each risk level are presented. The experimental results indicate that the proposed method is feasible for predicting risk levels of unstable angina patients.

Conclusion

In this paper, we have presented an intelligent system for UA risk assessment by combining genetic algorithm and fuzzy association rule mining. The developed approach has been tested on a data-set consisting of 54 UA patient cases from the Cardiology department of Chinese PLA General hospital. The experimental results show that considerable agreement is achieved between the proposed approach and physicians' problem solving knowledge.

The main novelty of the developed model is that it represents a valuable objective tool for UA risk assessment. In medical literature, physicians are in discrepancies about the risk factors highlighted. This research has focused on the application of computational intelligence. In particular, a genetic-fuzzy system, to identify the key factors behind UA, is proposed, which could be used for educational purposes, and with further improvements, could assist and guide young physicians in their daily work.

For future studies, there may be a comparison of effectiveness in terms of the proposed system with traditional UA risk assessment models, such as TIMI, GRACE, etc. The application of the proposed system to other kinds of CVD, such as heart failure, will also be investigated.

Table 9 Sensitivity and specificity with different risk levels

	Sensitivity	Specificity
Low-risk	0.8125	0.8421
Medium-risk	0.8182	0.85
High-risk	1.0	1.0

Furthermore, other computational intelligence techniques can be associated with the developed system.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WD: conception, design, analysis of data, development, interpretation and evaluation of results, and drafting of the manuscript. ZH: conception, design, and development of methods, drafting of the manuscript, and critical revision of manuscript. LJ: conception, design, extraction of data, interpretation and evaluation of results. HD: design of methods, analysis of data, and critical revision of manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Nature Science Foundation of China under Grant No 81101126, and the National Hi-Tech R&D Plan of China under Grant No 2012AA02A601. The author would like to give special thanks to all experts who cooperated in the evaluation of the proposed method.

Author details

¹Cardiology Department of Chinese PLA General Hospital, Beijing, China. ²College of Biomedical Engineering and Instrument Science, Zhejiang University, 310008, Zhou Yiqing Building 510, Zheda road 38#, Hangzhou, Zhejiang, China. ³IT Department of Chinese PLA General Hospital, Beijing, China.

Received: 10 June 2013 Accepted: 13 February 2014
 Published: 18 February 2014

References

1. Yeghiazarians Y, Braunstein JB, Askari A, Stone PH: **Unstable angina pectoris**. *N Engl J Hum Serv* 2000, **342**(2):101-114.
2. Graham CA, Tsay SX, Rotheray KR, Rainer TH: **Validation of the TIMI risk score in Chinese patients presenting to the emergency department with chest pain**. *Int J Cardiol* 2013, **168**:597-598.
3. 2012 Writing Committee Members, Jneid H, Anderson JL, Wright RS, Adams CD, Bridges CR, Casey DE Jr, Ettinger SM, Fesmire FM, Ganiats TG, Lincoff AM, Peterson ED, Philippides GJ, Theroux P, Wenger NK, Zidar JP, Anderson JL, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines: **2012 ACCF/AHA focused update of the guideline for the management of patients with Unstable Angina/Non-ST-Elevation myocardial infarction (updating the 2007 guideline and replacing the 2011 focused update)**. *Circulation* 2012, **126**(7):875-910.
4. Kozieradzka A, Kamiński KA, Maciorkowska D, Olszewska M, Dobrzycki S, Nowak K, Kralisz P, Prokopczuk P, Musiał WJ: **GRACE, TIMI, Zwolle and CADILLAC risk scores - do they predict 5-year outcomes after ST-elevation myocardial infarction treated invasively?** *Int J Cardiol* 2011, **148**:70-75.

5. Boersma E, Pieper KS, Steyerberg EW, Wilcox RG, Chang WC, Lee KL, Akkerhuis KM, Harrington RA, Deckers JW, Armstrong PW, Lincoff AM, Califf RM, Topol EJ, Simoons ML: **For the PURSUIT Investigators. Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation. Results from an international trial of 9461 patients.** *Circulation* 2000, **101**:2557–2567.
6. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, Van De Werf F, Avezum A, Goodman SG, Flather MD, Fox KA, Global Registry of Acute Coronary Events Investigators: **Predictors of hospital mortality in the global registry of acute coronary events.** *Arch Intern Med* 2003, **163**:2345–2353.
7. Bellman R, Zadeh L: **Decision-making in a fuzzy environment.** *Manag Sci* 1970, **17**:141–164.
8. Vidal F, Villard PF, Lutton E: **Tuning of patient-specific deformable models using an adaptive evolutionary optimization strategy.** *Biomed Eng IEEE Trans* 2012, **59**(10):2942–2949.
9. Buczak A, Koshute P, Babin S, Feighner B, Lewis S: **A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data.** *BMC Med Inform Decis Mak* 2012, **12**:124. [http://dx.doi.org/10.1186/1472-6947-12-124]
10. Mago V, Mehta R, Woolrych R, Papageorgiou E: **Supporting meningitis diagnosis amongst infants and children through the use of fuzzy cognitive mapping.** *BMC Med Inform Decis Making* 2012, **12**:98. [http://dx.doi.org/10.1186/1472-6947-12-98]
11. Carmona C, Gonzalez P, del Jesus M, Herrera F: **NMEEF-SD: non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery.** *Fuzzy Syst IEEE Trans* 2010, **18**(5):958–970.
12. del Jesus M, Gonzalez P, Herrera F, Mesonero M: **Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing.** *Fuzzy Syst IEEE Trans* 2007, **15**(4):578–592.
13. Casillas J, Carse B, Bull L: **Fuzzy-XCS: a michigan genetic fuzzy system.** *Fuzzy Syst IEEE Trans* 2007, **15**(4):536–550.
14. Dubois D, Hülermeier E, Prade H: **A systematic approach to the assessment of fuzzy association rules.** *Data Min Knowl Discov* 2006, **13**:167–192.
15. Bang YK, Lee CH: **Fuzzy time series prediction using hierarchical clustering algorithms.** *Expert Syst Appl* 2011, **38**(4):4312–4325.
16. Horng SC, Yang FY, Lin SS: **Hierarchical fuzzy clustering decision tree for classifying recipes of ion implanter.** *Expert Syst Appl* 2011, **38**:933–940.
17. West G, McArthur S, Towle D: **Industrial implementation of intelligent system techniques for nuclear power plant condition monitoring.** *Expert Syst Appl* 2012, **39**(8):7432–7440.

doi:10.1186/1472-6947-14-12

Cite this article as: Dong et al.: A genetic fuzzy system for unstable angina risk assessment. *BMC Medical Informatics and Decision Making* 2014 **14**:12.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

