A Framework for Analyzing Uncertainty in Data Using Computational Intelligence Techniques

M. Sujatha, G. Lavanya Devi and N. Naresh

Abstract In this study, use of soft computing techniques for analyzing medical data is presented. A medical dataset usually contains objects/records of patients that include a set of symptoms that a patient experiences. Analysis of such medical data could reveal new insights that would definitely help in efficient diagnosis and also in drug discovery. A novel fuzzy-rough based classification approach is described and its performance is evaluated using a medical dataset having multiclass values for response variable. Novel approaches for data preprocessing using fuzzy-rough concepts are introduced for attaining complete and consistent data. Thus, automated medical diagnosis can be done efficiently by using computational intelligence (CI) techniques for the benefit of mankind to live a healthy long life.

Keywords Rough set theory · Fuzzy set theory · Fuzzy-rough concepts · Computational intelligence

1 Introduction

Artificial Intelligence (AI) is an emulating human intelligence on machines to act and think similar to human beings. It contains reasoning, planning, intelligent search, machine learning and perception building. Traditional AI problem resolving methods are focused on problem states and rule set design to draw transitions in problem states. AI techniques are better for inductive and analogy-based learning compared to supervised learning [1]. These techniques are less feasible for opti-

M. Sujatha · G. Lavanya Devi (✉) · N. Naresh
Computer Science and System Engineering,
Andhra University College of Engineering, Visakhapatnam, India
e-mail: lavanyadevig@yahoo.co.in

M. Sujatha
e-mail: sujathamadugulacse@gmail.com

N. Naresh
e-mail: naresh855@gmail.com

mization and machine learning with regard to uncertain data. These failures of traditional AI opened up with solutions from CI techniques for real world problems. CI adopts techniques motivated by nature, that possess the ability to learn and deal with new situations with high computational speed. Also, they are less error-prone to noisy information sources [2].

1.1 Motivation

There are several data mining techniques to analyze data. Still there is always a scope for new approaches to analyze data for better decision making. The current trend is to develop techniques for data analysis based on domain, volume and type of the data. This approach has every chance of better decision making than generic approaches. Hence, techniques tailored based on the domain (banking, finance, social media, medical etc.) would perform better. Every domain has some or other form of uncertainty that has to be interpreted properly. This research is focused on medical domain.

In general, medical data has uncertainty due to the reason that a patient suffering from a specific illness cannot be completely determined by one or more symptoms; a certain set of symptoms can only indicate that there is a probability of a particular illness.

Broadly medical misdiagnosis can be classified into three classes.

False positive: misdiagnosis of a disease that is not actually present.
False negative: failure to diagnose a disease that is present.
*Equivocal result*s: inconclusive interpretation without a definite diagnosis [60].

General data mining methods are inefficient in dealing with cognitive uncertainties such as vagueness and ambiguity.

Proper analysis of medical datasets is one of the best ways to discover potentially useful information for diagnosis and also drug discovery. This research work aims to address the problem of uncertainty in medical datasets. Fuzzy-rough set based data preprocessing methodologies have been proposed for handling medical datasets. Finally an ensemble of rule based fuzzy-rough classifier is built for analyzing medical datasets.

The proposed work has been carried out to achieve the following objectives that address the above mentioned problem.

1.2 Objectives of the Proposed Work

To study the role of uncertainty in real world problems and to understand the scope of RST for analyzing uncertain data.

To design an efficient classifier that handles ambiguity and vagueness in medical datasets for better diagnosis of illness.

To design and built a rule based fuzzy-rough classifier for analyzing uncertainty in medical dataset.

To perform experimental analysis to evaluate the performance of the proposed classification model.

To test the performance of the proposed approach with existing state-of-the-art approaches.

1.3 Computational Models for Prediction

In general, systems built based on CI techniques are used for decision making. Decision making is essential in all fields of human activities and affect every sphere of our life. Today's real world problems are mostly data driven. Data accumulation in every domain of life is a current challenge for data engineers. Often, data is imprecise, ambiguous, vague and uncertain. Decision making is a difficult process due to the uncertainty in data. Firstly, data has to be set free from ambiguity and uncertainty. Then, it can be used for decision making. These situations can be better handled with CI based solutions. CI techniques constitute of neural networks (NN), genetic algorithms (GA), fuzzy systems, rough set and hybrid systems (combinations of NN, GA, fuzzy system and rough sets).

Neural networks learn by instances. Thus, NN architectures identifies known instances of a problem before they are tested for their 'inference' capability on known instances of the problem [3]. Therefore, it identifies new instances before untrained. It has ability to generalize and predict new outcomes.

Genetic algorithms are computerized search and optimization techniques based on procedure of selecting natural genetics. GA and evolutionary strategies mimic the principle of natural genetics and natural selection to construct search and optimization procedures [4].

Fuzzy set theory is proven to solve uncertainty in data. Rough set is a new approximation of a crisp set. It deals with uncertainty in data by using approximation [5]. Uncertain data in a given model can be handled by indiscernibility for rough sets and vagueness for fuzzy sets.

2 Uncertainty in Data

Uncertainty may be due to lack of knowledge or insufficient information [6–9]. Vagueness and ambiguity are the two major forms of uncertainty. Vagueness is associated with the difficulty of making sharp or precise distinctions in the real world [10, 11].
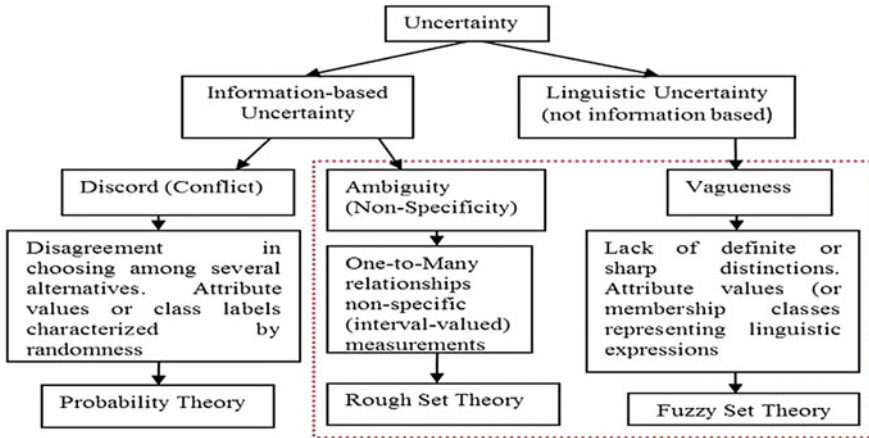
Fig. 1 Classification of uncertainty

Ambiguity is associated with two or more alternatives such that the choice between them is left unspecified. Broad classification of uncertainty is depicted in Fig. 1.

Information based uncertainty arises from lack of information. It is classified into discord and ambiguity. Discord is associated with the conflict in choosing among several alternatives of the attribute. It is handled by the probability theory [12]. Ambiguity can be addressed by rough set theory, to discover hidden patterns in data. It finds partial or total dependencies in databases, eliminates redundant and missing data.

Linguistic uncertainty that arises in natural language is vague and also the precise meaning of the words can change over time. Linguistic variables are words or sentences in a natural or artificial language. The meaning of linguistic variable is given as vagueness. Vagueness is existence of objects which cannot be uniquely classified relative to a set or its complement [1].

2.1 Classical Set Theory Versus Fuzzy Set Theory Versus Rough Set Theory

This section provides an overview of classical set, fuzzy set and rough set theories which includes the basic concepts, notations, applications and limitations. A summary of the same is given in Table 1.

Classical Set Theory

A set is defined as collections of objects which share certain characteristics. Set theory is the branch of mathematical logic that studies sets. It was initiated by

Table 1 Classical set theory versus fuzzy set theory versus rough set theory

	Classical set theory	Fuzzy set theory	Rough set theory
Definitions	Elements can belong to a set or not at all. It either belongs to the set definitely (denoted as 1) or does not belong to the set definitely (denoted as 0) Also known as crisp sets	Fuzzy sets are those whose elements have degrees of membership. It permits the gradual assessment of the membership of elements valued in the real unit interval [0, 1] Crisp sets are special cases of fuzzy sets	Rough set theory is an extension to set theory that enables approximation in decision making by using a pair of sets {Lower Approximation, Upper Approximation} of original set Rough set is a formal approximation of a crisp set
Notations	Let U be a space of objects, referred to as the universe of discourse, and x an element of U A classical (crisp) set A, $A \subseteq U$ is defined as a collection of elements $x \in U$, such that each element x can belong to the set or not belong	Let U be a finite and non-empty set called universe A fuzzy set A of U is defined by a membership function: $\mu_A: U \to [0, 1]$	The decision system $(U, A \cup \{d\})$, where U is the finite universe of discourse, A the set of conditional attributes and d the decision attribute, with $d \notin A$ $P \subseteq A$ there is associated an equivalence relation $IND_A(P)$: $\{(x, y) \in U^2 \mid \forall_a \in P, a(x) = a(y))\}$
Operations	Union: $A \cup B = \{x \mid (x \in A) \, or \, (x \in B)\}$ Intersection: $A \cap B = \{x \mid (x \in A) \, and \, (x \in B)\}$ Difference: $B - A = \{x \mid (x \in B) \, and \, not \, (x \in A)\}$ Complement: $A^c = \{x \mid (x \notin A), \, (x \in U)\}$ where A and B are crisp sets	Union: $\mu_{A \cup B}(x) = max(\mu_A(x), B(x))$ Intersection: $\mu_{A \cap B}(x) = min(\mu_A(x), B(x))$ Difference: $A - B = (A \cap \bar{B})$ Complement: $\mu_A^{-1}(x) = 1 - \mu_A(x)$ where A and B are fuzzy sets	Lower approximation: $\underline{P}X = \bigcup_{x \in U} \{P(x): P(x) \subseteq X\}$ Upper approximation: $\overline{P}X = \bigcup_{x \in U} \{P(x): P(x) \cap X \neq \emptyset\}$ Positive region: $Pos_P(Q) = \bigcup_{X \in U/Q} \underline{P}X$ Boundary region: $BND_P(Q) = \bigcup_{X \in U/Q} \overline{P}X - \bigcup_{X \in U/Q} \underline{P}X$
Applications	Used in different mathematical approaches like vector spaces, graphs, rings and manifolds described by axiomatic properties	Used for knowledge discovery in databases, to design controllers for electronic instruments, automobiles and traffic monitoring systems	Used in computational intelligence, pattern recognition, acoustic analysis and metrological pattern classification
Limitations	No partial membership is allowed	Membership value of an element does not depend on other elements	Membership value of an element depends on other elements

Georg Cantor and Richard Dedekind in 1870s. After the discovery of paradoxes in naive set theory, numerous axiom systems were proposed in the early twentieth century, of which the Zermelo–Fraenkel axioms, with the axiom of choice, are the best-known [47]. Table 1 gives a summary on the basics, operations and applications of set theory.

Fuzzy Set Theory

Fuzzy Set Theory permits gradual assessment of membership of elements in a set with real value interval [0, 1]. It represents classical bivalent sets as *crisp* sets [13]. It is used when information is vague, incomplete or imprecise. In real life, human imprecision causes an information available to be vague or fuzzy. Vagueness is handled by making use of "soft" boundaries of fuzzy sets i.e., graded membership which gives subjective knowledge to define these attributes. Fuzzifying attributes is tedious when precise information is available. Brief on basics, operations and applications of fuzzy theory is depicted in Table 1.

Rough Set Theory

Rough Set Theory is an extension of conventional set theory that supports approximations in analyzing decisions [14]. It is represented as a pair of sets, *Lower* and *Upper* approximation of crisp set. Lower approximation identifies objects that certainly belong to subset of interest, whereas upper approximation identifies objects that possibly belong to subset [15–18].

Rough set theory based models efficiently handles incomplete or imperfect knowledge [19–21]. Table 1 gives the outline of basics, operations and applications of classical set, fuzzy set and rough set theories.

2.2 Combining Fuzzy and Rough Set Theories

Fuzzy and rough set theories evolved as successful approaches to represent and compute imperfect data in real world applications [22–24]. Both of the theories are not competitive rather they are complementary to each other [25–27]. Along with the complementary nature, the similarities between them have motivated to develop a hybrid theory that covers the strengths of each. Lynn Deer et al., have proposed the process of fuzzifying the lower and upper approximations [28–30].

In this theory, given incomplete information (A) representing a subset of given universe (U) containing examples of a concept (C), the lower and upper approximations, along with an equivalence relation (R) that models indiscernibility can be expanded in the following two ways:

i. Objects in set A can belong to a concept with varying degree, i.e., making the set into a fuzzy set.

ii. Objects are classified into classes with soft boundaries, using similarities among the objects which are represented by a fuzzy relation *R*. Here, the fuzzy relation is used instead of indiscernibility relation.

This article does not intend to cover fuzzy-RST in its entirety. Rather, it confines to a brief introduction of rough-fuzzy and fuzzy-rough concepts.

Rough-Fuzzy Sets

A rough-fuzzy set is simplified from approximation of a fuzzy set in a crisp approximation space, where decision attribute values are fuzzy and conditional values are crisp [42]. In rough-fuzzy set, lower and upper approximations of objects belonging to these sets are defined as

$$\mu_{\underline{P}X}\big([x]_P\big) = inf\big\{\mu_X(x)|x \in [x]_P\big\} \tag{1}$$

$$\mu_{\overline{P}X}\big([x]_P\big) = sup\big\{\mu_X(x)|x \in [x]_P\big\} \tag{2}$$

where $[x]_P$ is crisp, $\mu_X(x)$ is degree of *x* fit into fuzzy equivalence class *X*, tuple $<\underline{P}X, \overline{P}X>$ is represented as rough-fuzzy set. Rough-fuzzy sets simplified to fuzzy-rough sets, when equivalence classes are fuzzy.

Fuzzy-Rough Sets

Competing with rough-fuzzy set theory, researchers focused on fuzzy-rough set theory [44, 45]. Earlier, such type of hybridization makes data analysis on information systems. Dubois suggested fuzzy-rough set theory which approximate fuzzy sets [43]. Rough sets are defined in terms of fuzzy membership function. It represents boundary regions, positive and negative region. In boundary region objects belong to membership value is 0.5, positive region objects belong to membership value is 1and negative region objects belong to membership value is 0.

2.3 Neural Networks

Neural networks are simplified models of the biological nervous systems. They can be defined as data processing systems, consisting of a large number of simple, highly interconnected processing elements (artificial neurons), in an architecture inspired by the structure of the cerebral cortex of the brain.

Learning methods in NN are broadly classified into three types: *supervised, unsupervised and reinforced* [31].

In supervised learning, every input pattern that is used to train the network is associated with an output pattern, which is the target or the desired pattern.

In unsupervised learning, the target output is not presented to the network.
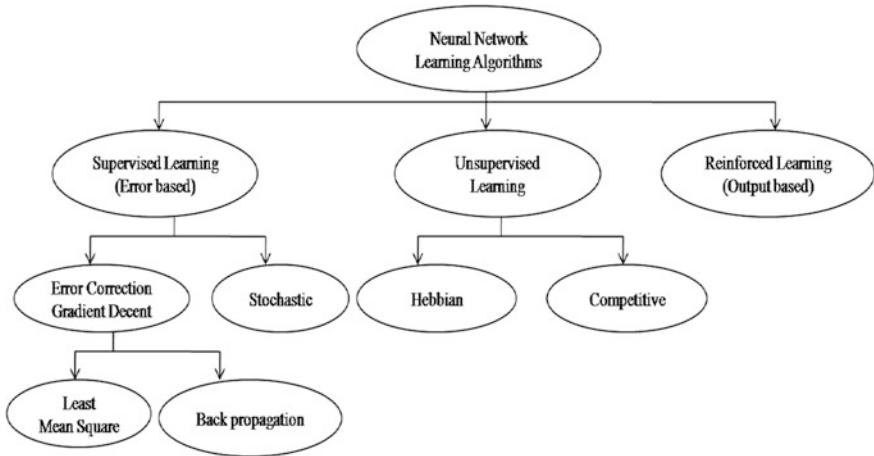
Fig. 2 Classification of learning algorithms

Table 2 The classification of NN system with respect to type of architecture

Type of architecture			
	Single layer feedforward network	Multilayer feedforward network	Recurrent network
Diagram			
Description	Comprises of two layers: input layer and output layer	Comprises of multiple layers: input layer, hidden layer (one or more) and output layer	Comprises of multiple layers as in feedforward networks. In addition, they have at least one feedback loop.
Popular methods	ADALINE, Hopfield, Perceptron	Cascade correlation, radial basis function	Adaptive resonance theory, Boltzman machine, Cauchy machine
Remarks	Can only deal linear problems	Can deal non-linear problems	It is often difficult to predict the optimal neural network size for a particular application
		With the increase of number of hidden layers computational complexity increases	

In reinforced learning, indication whether the computed output is correct or incorrect is given. A reward is given for the correct output and incorrect output is penalized. Classification of learning algorithms as depicted in Fig. 2.

Table 2 shows the classification of the NN systems listed above, according to their learning methods and architectural types.

2.4   Genetic Algorithms

Genetic algorithms are adaptive heuristic search methods based on principles of evolutionary ideas of natural selection and genetics [40]. GAs represents decision attributes of a search problem into fixed-length strings of certain cardinality. Strings results as candidate solutions to search problem for specified chromosomes, alphabets are referred to as genes and values of genes are named as alleles. When problem is encoded in a chromosomal manner and a fitness measure for discriminating good solutions from bad ones have been chosen, solutions to search problem start to evolve using *Initialization, Evaluation, Selection, Recombination and Mutation* [41].

Initial population of candidate solutions is generated randomly across the search space. When population is initialized, fitness values of the candidate solutions are evaluated. Selection allocates more copies of those solutions with higher fitness values and thus imposes the survival-of-the-fittest procedure on candidate solutions. Recombination combines parts of two or more parental solutions to create new, possibly better solutions. The offspring under recombination will not be identical to any particular parent. Table 3 presents most commonly used selection methods, recombination (crossover) operators, mutation operators and replacement of GAs [39].

While recombination operates on two or more parental chromosomes, mutations are changes in genetic sequence of a chromosome. These changes occur at many different levels of individual chromosomes. The offspring population created by selection, recombination, and mutation replaces the original parental population.

2.5   Hybrid Systems

Hybrid systems is combination of two or more techniques to overcome limitations of individual techniques. It is impossible to deal with either as purely continuous or discrete-event system without ignoring important phenomena that result from the combination of continuous and discrete movements of this system [35].

Mohammed Hamed Ahmed Elhebir described minimum support requirement dictates the efficiency of association rule mining [48]. If support threshold is low, then not truly interesting rules are generated. On the other hand, if the support threshold is high, then interesting rules are missed from the rule set. Eliminating redundant rules and clustering decreased the size of the generated rule set for obtain interestingness rules.

Abraham et al., presented biological motivation on particle swarm optimization and ant colony optimization algorithms. The basic data mining terminologies are explained by using swarm intelligence techniques [49].

Table 3 Genetic algorithm operators (GAO)

GAO	Methods	Description
Selection methods	Fitness proportionate selection	Selection is performed based on the fitness function derived
		Methods mostly used are Roulette-wheel Selection, Boltzmann Selection, Rank Selection, Steady State Selection and Stochastic Universal Selection
	Ordinal selection	Selection procedure is independent of fitness function
		Methods mostly used are Tournament Selection and Truncation Selection
Recombination (crossover) operators	Crossover operators	Individuals from the mating pool are recombined to create new, hopefully better, offspring
	k-point crossover	Crossover site is selected at random over string length and alleles on one side of site are exchanged between individuals
	Uniform crossover	Every allele is exchanged between a pair of randomly selected chromosomes with a certain probability
	Uniform order-based crossover	k-point and uniform crossover methods described above are not well suited for search problems with permutation codes such as the ones used in the traveling salesman problem
	Order-based crossover	It is a variation of uniform order-based crossover in which two parents are randomly selected and two random crossover sites are generated
	Partially matched crossover	Two parents are randomly selected and two random crossover sites are generated
	Cycle crossover	Two randomly selected parents P_1 and P_2 that are solutions to a traveling salesman problem. Offspring C_1 receives first variable from P_1. Then choose variable that maps onto same position in P_2
Mutation operators	It is designed if two parents has same allele at a given gene then one-point crossover will not change that i.e., gene have same allele forever	
Replacement	Delete-all	Deletes all members of current population and replaces them with same number of chromosomes that have just been created
	Steady-state	Deletes n old members and replaces them with n new members. The number to delete and replace, n, at any one time is a parameter to this deletion technique
	Steady-state-no-duplicates	It checks no duplicate chromosomes are added to population

Table 4 Hybrid systems and their applications

Hybrid system	Applications
KMLP-ANN (K-means Multilayer Perceptron Artificial Neural Networks)	It investigates *k*-means algorithm for data filtering and MLP-ANN for predicting customer behavior is very important for real life market and competition and it is essential to manage it. In the industry of information and communication technology [36]
NN-GA (Neural Networks-Genetic Algorithm)	NN are becoming of significant importance in automatic diagnosis and prognoses of different diseases. GA used to find the optimum network structure with high classification accuracy [37, 61–63]
HIDSS (Hybrid Intelligent Decision Support Systems)	It is implemented in an agent-based architecture, a case study system European Monetary Union-HIDSS is built for risk analysis and prediction of evolving economic clusters in Europe [46]. Neural network and rule-based agents are evolved from incoming data and expert knowledge if a decision making process requires this
gHFNN (Genetically Optimized Hybrid Fuzzy Neural Networks)	It is used for modeling vehicle for nonlinear and complex systems. It is constructed by combining Fuzzy NNs with polynomial NNs. Rule-based FNN is optimized by genetic algorithms and back-propagation (BP) learning algorithm: GAs leads to the auto-tuning of vertexes of membership function, while the BP algorithm helps produce optimal parameters of the consequent polynomial of fuzzy rules through learning. Polynomial NN that is the consequent structure of the gHFNN [38]

Benxian Yue et al., investigated optimal reducts using a particle swarm optimization approach. This approach observed change of positive region as particles proceed throughout search space is best attribute [50].

Dong-Hwa Kim et al., proposed hybrid approach genetic algorithms and bacterial foraging algorithms for function optimization problems [51]. Performance of hybrid approach is studied on mutation, chemotactic steps, crossover, variation of step sizes and lifetime of bacteria.

Various applications of the hybrid systems is given in Table 4.

3 Classification Modeling Using CI Techniques

Fuzzy and rough sets are well known for analyzing different aspects of uncertainty. Combination of these techniques will enable to build a robust mathematical approach for combating problems of uncertainty. This work suggests fuzzy-rough theory based solution for classifying medical datasets. Medical datasets are considered as special domain as most of the times, medical data is incomplete, vague and noisy. Fuzzy-rough rule induction (FRRI) is developed to generate rules for classification. Further, proposed FRRI is tested to understand the performance on

Table 5 Performance measures

Metrics	Formula	Evaluation process
Accuracy	$\dfrac{\sum_{i=1}^{l}\frac{tp_i+tn_i}{tp_i+fn_i+fp_i+tn_i}}{l}$	Average per-class effectiveness of a classifier
Error rate	$\dfrac{\sum_{i=1}^{l}\frac{fp_i+fn_i}{tp_i+fn_i+fp_i+tn_i}}{l}$	Average per-class classification error
Precision	$\dfrac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fp_i)}$	Agreement of the data class labels with those of a classifiers
Recall	$\dfrac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fn_i)}$	Effectiveness of a classifier to identify class labels

par with other CI techniques such as Multi-Layer Perceptron (MLP) and Fuzzy-Genetic Modeling (Table 6). The metrics are used for perforamce analysis of classifiers is shown in Table 5.

3.1 Proposed Algorithm for Rule Generation

This section gives the proposed algorithm for rule generation that uses Fuzzy-Rough Rule Induction. In the proposed hybrid classifier lower and upper approximation of RST is fuzzified by adopting the approach given below:

- A fuzzy set in X, is simplified to set A, allowing these objects can fit to class label to different degrees.
- Alternatively, estimate objects indiscernibility, their approximate equality, represented by a fuzzy relation R may be measured. Subsequently, objects are classified into classes with ''soft'' boundaries based on their equality to each other.

By definition, abrupt transitions between classes are adjusted by gradual ones, allowing that an element can fit (to varying degrees) to more than one class.

Fuzzy Indiscernibility

An information system I is considered. The fuzzy indiscernibility relation R_a is used for any fuzzy relation that determines degree to which two objects are indiscernible. The following equations are based on a quantitative attribute in tolerance relations R_a:

$$R_a(x, y) = \max\left(\min\left(\frac{a(y) - a(x) + \sigma_a}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a}{\sigma_a} \right), 0 \right) \qquad (3)$$

where $\forall x, y \in U$ with σ_a denoting the standard deviation of a. Fuzzy-rough set theory lower and upper approximations are defined by an implicator I and t-norm τ.

The following are the fuzzy B-lower and B-upper approximations of a fuzzy set A in U.

$$(R_B{\downarrow}A)(y) = \inf_{x \in U} \tau(R_B(x, y)A(x)) \tag{4}$$

$$(R_B{\uparrow}A)(y) = \sup_{x \in U} \tau(R_B(x, y)A(x)) \tag{5}$$

$R_B \downarrow A$ is set of elements necessarily satisfying concept (strong membership), while $R_B \uparrow A$ is set of elements possibly belonging to concept (weak membership). Mainly, these were designed to deal with uncertainty in data. Fuzzy B-positive region is defined based on fuzzy B-indiscernibility relations, for $y \in X$.

$$POS_B(y) = \left(\bigcup_{x \in U} R_B \downarrow R_d x\right)(y) \tag{6}$$

Degree of dependency of d on B, γ_B by

$$\gamma_B = \frac{|POS_B|}{|U|} = \frac{\sum_{x \in X} POS_B(x)}{|U|} \tag{7}$$

An algorithm for classification of medical dataset with a hybrid approach using fuzzy-rough set theory is represented as Fuzzy-Rough Rule Induction (FRRI).

Algorithm: Fuzzy-Rough Rule Induction (FRRI)

Given medical dataset M, select randomly a subset *B* of conditional attribute A.

(a) Initially subsets B of conditional attribute, ruleset *R* and cover set Cov are empty.

(b) For each attribute $a \in M$, repeat the following steps.

(c) For each object $o_1 \in M$, repeat the following steps.

(d) Compute γ degree D_1 of belongingness of o_1 to positive region of attribute *a*.

(e) Compute γ degree D_2 of belongingness of o_1 to positive region for given dataset M.

(f) If degree D_1 equals to D_2.

(g) Then construct the rule r, for the object o_1 and attribute subset $B \cup a$.

(h) Add the rule to ruleset R if r does not have same or more coverage than existing rules in ruleset R. Update the coverage set.

This algorithm generates fuzzy-rough if-then rules for classifying a given medical dataset. However, it is observed that in such rules, antecedents are covering almost all attributes of decision system which eventually increase computational time in classification. It is favorable to combine rule induction and attribute

selection process. Thereby, fuzzy rules are generated from best attributes that maximally cover the decision system.

3.2 Multi-layer Perceptron for Classification

Multilayer perceptron (MLP) classifier is based on the feed forward artificial neural network [32]. Here, information moves in only forward direction from input nodes and passed through next hidden nodes to output nodes. There are no cycles or loops in the network. It consists of multiple layers of nodes. Each layer is fully connected to the next layer in the network. In input layer, nodes represent input data. All other nodes map inputs to outputs by a linear combination of inputs with the node's weights w and bias b by applying an activation function. This can be written in matrix form for MLP with $K + 1$ layers as follows:

$$y(X) = f_K\left(\cdots f_2\left(w_2^T f_1\left(w_1^T X + b_1\right) + b_2\right)\cdots + b_K\right) \tag{8}$$

A multi-layer neural network can compute a continuous output instead of real numbers called step function. A common choice is the so-called logistic function. Sigmoid function refers to the special case of the logistic function. Nodes in intermediate layers use sigmoid (logistic) function $f(Z_i)$.

These nodes of networks consists of multiple layers are interconnected in a feed-forward direction. Every neuron in one layer shown feed-forward directed network to its subsequent layer. This process is implemented by sigmoid function as an activation function in this network.

$$f(Z_i) = \frac{1}{1 + e^{-Z_i}} \tag{9}$$

Nodes in the output layer use softmax function:

$$f(Z_i) = \frac{e^{Z_i}}{\sum_{K=1}^{N} e^{Z_K}} \tag{10}$$

Number of nodes in the output layer corresponds to the number of classes.

3.3 Genetic-Fuzzy Modeling (GFM)

This section gives a hybrid CI approach that have attracted a lot of attention in past decade, which successfully offered solutions to many real world problems. In this work, hybridization of fuzzy logical and genetic algorithm (GA) is used to construct the classification model. GA are stochastic search based on natural selection and

genetics. The fitness function simply defined as candidate solution to the problem as input and produces as output how "fit" our how "good" the solution is with respect to the problem in consideration. Calculation of fitness value is done repeatedly in a GA and therefore it should be sufficiently fast. The significant attributes selected from dataset helps in diagnosing system to built a classification fuzzy inference model. The rules for the fuzzy system are generated from dataset. These rule sets are most significant and optimal subset of rules are selected using genetic algorithm. The benefits of GA and fuzzy inference system for effective prediction of heart disease in patients. Genetic-Fuzzy Logic (GFL) model for effective heart disease prediction [33].

Genetic-Fuzzy Logic

GFL finds fitness of a chromosome in a population by genetic algorithm, as it decides termination criterion. The attributes in the dataset are selected using GA and the fuzzy inference system for classification. The fitness function value is the measurement that helps to check the nearness of the optimal solution. Genetic operators used in genetic algorithms are analogous to those in the real life: survival of the fittest, or selection; reproduction (crossover, also called recombination); and mutation.

Selection is a process of selecting parents among the population by using roulette wheel selection. Intermediate crossover is used to select parents and interchanges position of values based on crossover point fixed. The values before fixed point from one chromosome is replaced with first part of new chromosome and the values that are in the second chromosome are replaced with new ones, thus inheriting the features of both the parents. Gaussian mutation is a process of changing gene values based on its given probability 0.05 and 1. The stochastic search of the genetic algorithm stops based on the convergence criteria. GA combine high performance notions to achieve better performance for getting optimal solution.

Membership Functions

Membership function represents the fuzzy set and measures degree of similarity. It is defined as fuzzy set A on the universe of discourse X is defined as $\mu_A : X \to [0, 1]$, where each element of X is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of membership of the element in X to the fuzzy set A.

$$\mu_A(x, c, s, m) = exp\left[-\frac{1}{2}\left|\frac{x-c}{s}\right|^m \right] \tag{11}$$

where c is center, s is width and m is fuzzification factor.

Fuzzy Inference System

A fuzzy inference system maps inputs to output using predefined fuzzy rules available in knowledge base. The knowledge base consists of if-then fuzzy rules that specify the relationship between the input and output fuzzy sets. If

$a1, a2, \ldots, an$ are the attributes and $c1, c2, \ldots, cm$ are class labels then a fuzzy rules are based on the linguistic values. As it requires the input in fuzzy values, the input is fuzzified and output from the inference system is defuzzified.

Fuzzy Classifier

Fuzzy classifier learns data in the form of rules and predicts target value for set of test data. It uses cross validation tenfold technique for estimating the test error. This process identifies subset for testing with all the other subsets as training subsets. Using trained model testing subset is classified. It continues for ten times with different training and testing data. The actual labels are provided based on the data in the dataset whereas the modified labels are framed based on the fuzzy classifier. Here, defuzzification methods replaces fuzzy values to their corresponding crisp values by using centroid method.

In GFL, genetic algorithm implements stochastic search on dataset to reduce number of features. Fuzzy inference system predicts test data by fuzzy Gaussian membership function and centroid defuzzification method.

3.4 Rule Classifier Using Fuzzy Ant Colony Optimization

Ant Colony Optimization (ACO), an inspired algorithm from nature, has been successfully applied to classification tasks of data mining [54–56]. A rule-based system for medical data mining by using a combination of ACO and fuzzy set theory, called Fuzzy ACO-Miner was proposed by Mostafa Fathi Ganji and Mohammad Saniee Abadeh [52].

Fuzzy ACO-Miner operates in rule generations and its optimization phases. Initially, an ACO algorithm is applied to learn fuzzy rules. This algorithm applies the artificial ants to explore among the training samples and gradually deriving fuzzy rules. The ants learns the rules related to each class separately corresponding fuzzy rules.

Ant constructs rule randomly by adding one term at a time and in the next iterations the ants modify rule. Each ant chooses term to modify (or add to current rule in the first iteration) with following probability:

$$P_{i,j} = \frac{\tau_{i,j}(t) \cdot \eta_{i,j}}{\sum_{i}^{a} \sum_{j}^{b_i} \tau_{i,j}(t) \cdot \eta_{i,j}} \qquad (12)$$

$\eta_{i,j}$ is a problem-dependent heuristic value for term. The function that defines the problem-dependent heuristic value. $\tau_{i,j}$ is the amount of pheromone currently

available (at time t) between attributes. *I* is the set of attributes that are not yet used by the ant.

It is necessary to decrease the pheromone of terms that have not participated in the construction of rules. For this purpose, pheromone evaporation is simulated. To simulate the pheromone evaporation in real ant colony, the amount of pheromone associated with each term that does not occur in the constructed rule must be decreased. The pheromone of unused terms is decreased by dividing the amount of the value of each $\tau_{i,j}$ by the summation of all $\tau_{i,j}$ [53].

Fuzzy ACO-Miner has also have additional features that make it different from existing classifiers based on ACO meta-heuristic. Unknown data is classified using fuzzy ACO-Miner based on averaging both the number of rules and the covering value to classify the input data.

3.5 Fuzzy Discrete Particle Swarm Optimization Classifier for Rule Classification

In contrast to traditional mining approaches, biologically inspired algorithms such as evolutionary algorithms and swarm intelligence approaches can also be used for the purpose of classification task. Hence, fuzzy discrete particle swarm optimization classifier by local search (fuzzy DPSO-LS) classifier is proposed to handle imprecise and uncertain data is explained by Min Chen and Simone A. Ludwig [57].

FDPSO-LS classifier, which uses a rule base to represent a 'particle' that evolves rule base over time. It is implemented as a matrix of rules, representing fuzzy IF-THEN classification rules, that have conjunctive antecedents and one consequent. It is applied to both discrete and continuous data sets [58, 59]. In addition, a local mutation search strategy was incorporated in order to take care of the premature convergence of PSO. As the number of rules rises, an efficient algorithm that can automatically find the fuzzy rules is important and necessary. Normally, several rules of the rule base are fired in the fuzzy rule classification system. The predicted class for a given instance is determined by the membership degree of the input variables. Specifically, for each class *k*,

$$\beta_{Classk} = \arg \max_k \sum_{1 \le i \le n} \prod_{1 \le j \le m} \mu_{ij} \tag{13}$$

where μ_{ij} is the input membership degree of the *i*th rule of the *j*th antecedent. The class that has the largest β value is selected as the predicted class. Limitation of this approach are, it may not efficiently normalize discrete datasets using linguistic terms, leads to overfitting of the model and may result in decrease of classification of accuracy.

4 Experimental Evaluation of the Classification Model

The degree of belief on classifier's capability to classify the unknown instances can be acknowledged by looking at its generalization ability on trained instances. The classifier learned on training dataset has to be tested experimentally for its performance. Experimental procedures adopted for evaluation of the proposed classification model and analysis of results is shown in this section. The experimental process is carried out by using various classification evaluation metrics for assessing the classifier's performance on a dataset.

4.1 Evaluation Metrics

Most of the study on medical data analysis is done using binary classification models. This study focuses on medical data analysis with multiclass classification problem, as most in most of the cases the automated medical diagnosis cannot be formulated as binary classification problems. In multiclass classification, the instances are to be classified into only one class from set of non-overlapping classes. Subsequently, a fuzzy-rough classifier for classification of medical datasets having uncertain data is developed. The experimental evaluation process involves in comparison of proposed classification algorithms performance with existing computational intelligence techniques.

The choice of appropriate metrics for evaluation of the classifiers is vital to measure the actual performance. Here, the focus should be on performance but not the perception of the trained classifier. The most common metric to evaluate classifiers are accuracy and error-rate. Accuracy tells the ratio of instances the classifier classified correctly. Error-rate specifies the ratio of incorrect classification done by classifier. Although, specificity and sensitivity are the popular metrics to indicate success in medical diagnosis classification models along with medical test, they are not preferred here as the experimental evaluation is done for multiclass classification problems. Precision and Recall are known to be quality metrics for multiclass classification problems. Precision specifies classifiers ability in precisely (exactness) identifying the relevant class labels by the classifier for each class. Recall specifies completeness of the classifier.

Also performance of the proposed approach is given by comparing the results obtained with that of existing well known methods.

4.2 Description of the Dataset Used

The main focus is to analyze uncertainty in medical dataset, so that efficient diagnosis of diseases can be made by medical practitioners. This study made use of

Cleveland heart disease dataset retrieved from UCI machine learning repository [34]. The dataset has 303 instances with 14 attribute. The attribute indicating level of heart disease is considered as response attribute which is distributed into five classes.

4.3 Experimental Analysis

In the experimental evaluation process, Cleveland heart disease dataset is randomly partitioned into training and testing sets. Holdout method approach is used in splitting training and test sets with 2/3, 1/3 of total instances respectively. More instances of each class can make the classifier to attain good generalization capabilities. Training dataset is given as input to the classifiers to build classification model. For evaluation of the classifier the test set is used. The results obtained are showcased subsequently. Also, a performance of the proposed classifier is evaluated by comparing results obtained with that of the other classifier' built using computational intelligence techniques on the same dataset.

Fuzzy-Rough Rule Induction

Rule based classifiers are known for their better representation and interpretation of classification model. A rule based classifier (FRRI) is built on Cleveland heart

Table 6 Classification performance

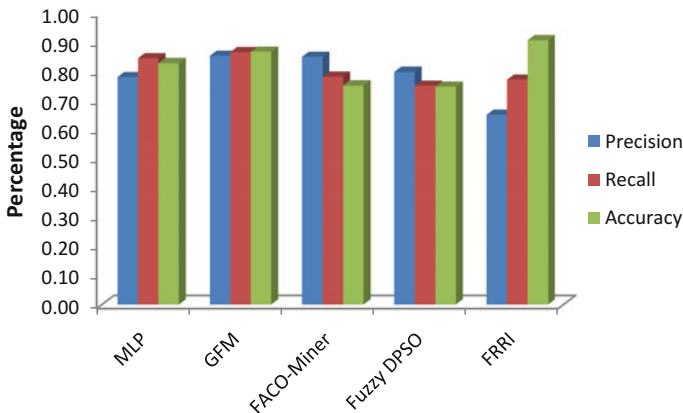| Classification models	No. of rules	Precision	Recall	Accuracy	Error rate
MLP	8	0.781	0.846	0.829	0.171
GFM	10	0.854	0.867	0.868	0.132
FACO-Miner	5	0.851	0.782	0.752	0.248
Fuzzy DPSO	7	0.799	0.751	0.748	0.252
Proposed FRRI	4	0.651	0.772	0.908	0.092

Fig. 3 Classification models

disease dataset. The proposed FRRI has generated 4 rules. Classification performed by these rules has given an accuracy of 90.8%. The performance of FRRI is tested with other CI techniques. The techniques used in comparative analysis are, Multi-layer perceptron and also, fuzzy based hybrid classification techniques: genetic-fuzzy modeling (GFM), fuzzy ant colony optimization (FACO-Miner), fuzzy particle swarm optimization (Fuzzy DPSO). It is observed that results are inferior to that of FRRI.

Table 6 depicts the classification performances of aforementioned computational intelligence techniques.

The graphical representation of classification models are depicted in Fig. 3.

5 Conclusions and Future Research Directions

The problem of analyzing medical data for better diagnosis by using computational intelligence techniques has been studied. This work emphasized on analysing uncertainty in medical data which is inevitable while collecting information from patients. A predictive hybrid model is built using fuzzy-rough classifiers that classifies a new instance/record of a patient and labels/conveys whether the patient might suffer from the disease or not (this research is carried on heart dataset). To achieve this, the model is trained to generate rules for classification. Fuzzy-rough concepts are used for the entire process as it is proven to handle uncertainty efficiently. A rule based fuzzy-rough classifier FRRI, which classifies uncertain medical datasets has been built.

Similarly, the framework can be extended to different domains other than medical field. The behavior of a netizen can be analyzed by such hybrid models. Also, the impact of social media can also be studied. Computational intelligence techniques can be used to build robust decision making systems. Further, ensemble of random forests can be used for classification. Parallel processing can be adopted to reduce the computational complexity of random forests.

References

1. G. Selvachandran, A vague soft set theoretic approach to multi attribute decision making problems. Appl. Math. Sci. **8**(134), 6937–6949 (2014)
2. J. Fulcher, *Advances in Applied Artificial Intelligence*. Computational Intelligence and its Applications Series (2006)
3. W. Sibanda, P. Pretorius, Novel application of multi-layer perceptrons (MLP) neural networks to model HIV in South Africa using seroprevalence data from antenatal clinics. Int. J. Comput. Appl. **35**(5) (2011)
4. A.J. Umbarkar, M.S. Joshi, P.D. Sheth, Dual population genetic algorithm for solving constrained optimization problems. Int. J. Intell. Syst. Appl. **2**, 34–40 (2015)

5. M. Kumar, N. Yadav, Fuzzy rough sets and its application in data mining field. Adv. Comput. Sci. Inf. Technol. **2**(3), 237–240 (2015)
6. A. Umut, S. Ayberk, Failure mode and effects analysis under uncertainty: a literature review and tutorial. Intell. Decis. Making Qual. Manage. **97**, 265–325 (2016)
7. C.M.D. Cornelis, E. Cock, E. Kerre, Intuitionistic fuzzy rough sets: at the crossroads of imperfect Knowledge. Expert Syst. **20**(5), 260–270 (2003)
8. J. David Hunter, Uncertainty in the Era of Precision Medicine. The New England Journal of Medicine. pp. 711–713, (2016)
9. J. Kacprzyk, S. Zadrozny, G. De Tre, Fuzziness in database management systems. Fuzzy Sets Syst. **281**, 300–307 (2015)
10. B. Zhang, A new measure of similarity between vague sets, in *International Conference on Oriental Thinking and Fuzzy Logic*, vol. 443 (2016), pp. 601–610
11. G. De Tre, S. Zadrożny, *Soft Computing in Database and Information Management*. Springer Handbook of Computational Intelligence (2015)
12. G.J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory* (Wiley, Hoboken, 2006), p. 499
13. L. Zadeh, *Fuzzy Sets, Information and Control* (1965), pp. 338–353
14. Z. Pawlak, Rough sets. Int. J. Comput. Inf. Sci. **11**(5) (1982)
15. I. Duntsch, G. Gediga, Rough set data analysis. Encycl. Comput. Sci. Technol. **43**(28), 281–301 (2000)
16. E.H. Shortliffe, J.J. Cimino, Biomedical informatics. Comput. Appl. Health Care and Biomed. (2014)
17. A. Skowron, S.K. Pal, Rough sets. Pattern recognition and data mining. Pattern Recogn. Lett **24**(6), 829–933 (2003)
18. S. Udhayakumar, H. Hannah, Inbarani: a novel neighborhood rough set based classification approach for medical diagnosis. Proc. Comput. Sci. **47**, 351–359 (2015)
19. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data* (Dordrecht Kluwer Academic, 1991)
20. A. Skowron, Z. Pawlak, J. Komorowski, L. Polkowski, A rough set perspective on data and knowledge, in *Handbook of Data Mining and Knowledge Discovery* (2002), pp. 134–149
21. D. Jianhua, Rough set approach to incomplete numerical data. Inf. Sci. **241**, 43–57 (2013)
22. D.S. Yeung, D. Chen, E.C.C. Tsang, J.W.T. Lee, W. Xizhao, On the generalization of fuzzy rough sets. IEEE Trans. Fuzzy Syst. **13**(3), 43–361 (2005)
23. I. Masahiro, W.-Z. Wu, C. Cornelis, *Fuzzy-Rough Hybridization*. Springer Handbook of Computational Intelligence (2015), pp. 425–451
24. W.Z. Wu, J.S. Mi, W.X. Zhang, Generalized fuzzy rough sets. Inf. Sci. **151**, 263–282 (2003)
25. D. Boixader, J. Jacas, J. Recasens, Upper and lower approximations of fuzzy sets. Int. J. Gen. Sys. **29**(4), 555–568 (2000)
26. E. Saleh, A. Valls, A. Moreno, P. Romero-Aroca, S. dela Riva-Fernandez, R. Sagarra Alamo, Diabetic retinopathy risk estimation using fuzzy rules on electronic health record data. Model. Decis. Artif. Intell. **9880**, 263–274 (2016)
27. S. Chimphlee, N. Salim, M.S.B. Ngadiman, W. Chimphlee, S. Srinoy, Independent component analysis and rough fuzzy based approach to web usage mining, in *Proceedings of the Artificial Intelligence and Applications* (2006), pp. 422–427
28. L. D'eer, N. Verbiest, C. Cornelis, L. Godo, A comprehensive study of implicator–conjunctor-based and noise-tolerant fuzzy rough sets: definitions, properties and robustness analysis. Fuzzy Sets Syst. (2014)
29. S.P. Tiwari, S. Sharan, V.K. Yadav, Fuzzy closure spaces vs. fuzzy rough sets. Fuzzy Inf. Eng. **6**(1), 93–100 (2014)
30. J. Zhan, K. Zhu, A novel soft rough fuzzy set: z-soft rough fuzzy ideals of hemirings and corresponding decision making. Soft Comput. 1–14 (2016)

31. S. Rajasekaran, G.A. Vijayalakshmi Pai: Neural networks, fuzzy logic and genetic algorithms synthesis and applications. PHI Learn. (2011)
32. A. Rajendra, S. Priti Srinivas, Artificial neural network. Intell. Tech. Data Sci. 125–155 (2016)
33. T. Santhanam, E.P. Ephzibah, Heart disease prediction using hybrid genetic fuzzy model. Indian J. Sci. Technol. **8**(9), 797–803 (2015)
34. UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml
35. A.P. Markopoulos, W. Habrat, N.I. Galanis, Modelling and optimization of machining with the use of statistical methods and soft computing. Des. Exp. Prod. Eng. 39–88 (2016)
36. A. Hudaib, R. Dannoun, O. Harfoushi, R. Obiedat, H. Faris, Hybrid data mining models for predicting customer churn. Int. J. Commun. Network Syst. Sci. **8**, 91–96 (2015)
37. Md. Mijanur Rahman, T. Akter Setu, An implementation for combining neural networks and genetic algorithms. Int. J. Comput. Sci. Technol. **6**(3) (2015)
38. S.-K. Oh, W. Pedrycz, Genetically optimized hybrid fuzzy neural networks: analysis and design of rule-based multi-layer perceptron architectures, in *Engineering Evolutionary Intelligent Systems* (Springer, 2008)
39. K.-L. Du, M.N.S. Swamy, *Genetic Algorithms. Search and Optimization by Metaheuristics* (2016), pp. 37–69
40. J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan, 1975)
41. D.E. Goldberg, K. Sastry, A practical schema theorem for genetic algorithm design and tuning, in *Proceedings of the Genetic and Evolutionary Computation Conference* (2001), pp. 328–335
42. D.H. Kraft, E. Colvin, G. Bordogna, Fuzzy information retrieval systems: a historical perspective, in *Fifty Years of Fuzzy Logic and its Applications*, vol. 326, pp. 267–296 (2015)
43. D. Dubois, H. Prade, *Putting Rough Sets and Fuzzy Sets Together* (1992), pp. 203–232
44. T. Beaubouef, F. Petry, Information systems uncertainty design and implementation combining: rough, fuzzy and intuitionistic approaches. Flexible approaches in data. Inf. Knowl. Manage. **497**, 143–164 (2014)
45. M. Cai, Q. Li, G. Lang, Shadowed sets of dynamic fuzzy sets. Granular Comput. 1–10 (2016)
46. Kaklauskas Arturas, Intelligent decision support systems. Biometric Intell. Decis. Making Support **81**, 31–85 (2015)
47. G.M. Bergman, Ordered sets, induction, and the axiom of choice, in *An Invitation to General Algebra and Universal Constructions* (2015), pp. 119–171
48. M.H.A. Elhebir, *Machine Learning Methods for Mining Web Access Patterns* (Sudan University for Science & Technology, 2016)
49. A. Abraham, C. Grosan, V. Ramos, *Swarm Intelligence and Data Mining*. Studies in Computational Intelligence (Springer, 2006), p. 270
50. B. Yue, W. Yao, A. Abraham, H. Liu, A new rough set reduct algorithm based on particle swarm optimization, in *International Work-Conference on the Interplay Between Natural and Artificial Computation* (2007), pp. 397–406
51. D.H. Kim, A. Abraham, J.H. Cho, Hybrid genetic algorithm and bacterial foraging approach for global optimization. Inf. Sci. **177**(18), 3918–3937 (2007)
52. M.F. Ganji, M.S. Abadeh, *Parallel Fuzzy Rule Learning Using an ACO-Based Algorithm for Medical Data Mining* (IEEE, 2010)
53. R.S. Parpinelli, H.S. Lopes, A.A. Freitas, Data mining with an ant colony optimization algorithm. IEEE Trans. Evol. Comput. **6**, 321–332 (2002)
54. K.M. Salama, A.M. Abdelbar, Learning neural network structures with ant colony algorithms. Swarm Intell. **9**, 229–265 (2015)
55. P. Mrutyunjaya, A. Ajith, Hybrid evolutionary algorithms for classification data mining. Neural Comput. Appl. **26**, 507–523 (2015)

56. A. Thannob, S. Siriporn, L. Chidchanok, Optimizing the modified fuzzy ant-miner for efficient medical diagnosis. Appl. Intell. **37**, 357–376 (2012)
57. M. Chen, S.A. Ludwig, A fuzzy discrete particle swarm optimization classifier for rule classification. Int. J. Hybrid Intell. Syst. **11**(3) (2014)
58. H. Ishibuchi, T. Nakashima, T. Murata, Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Trans. Syst. Man Cybern. **29**(5), 601–618 (1999)
59. L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples. IEEE Trans. Syst. Man Cybern. **22**(6) (1992)
60. Medical Misdiagnosis in America 2008, http://www.premerus.com/news/Misdiagnosis_in_America.pdf
61. O.W. Samuel, G.M. Asogbon, A.K. Sangaiah, P. Fang, G. Li, An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Syst. Appl. **68**, 163–172 (2017)
62. A.K. Sangaiah, A.K. Thangavelu, X.Z. Gao, N. Anbazhagan, M.S. Durai, An ANFIS approach for evaluation of team-level service climate in GSD projects using Taguchi-genetic learning algorithm. Appl. Soft Comput. **30**, 628–635 (2015)
63. A.K. Sangaiah, A.K. Thangavelu. An adaptive neuro-fuzzy approach to evaluation of team-level service climate in GSD projects. Neural Comput. Appl. **23**(8) (2013). doi:10.1007/s00521-013-1521-9