



RS-HeRR: a rough set-based Hebbian rule reduction neuro-fuzzy system

Feng Liu¹ · Arif Ahmed Sekh² · Chai Quek¹ · Geok See Ng¹ · Dilip K. Prasad²

Received: 23 January 2020 / Accepted: 2 May 2020 / Published online: 1 June 2020
© The Author(s) 2020

Abstract

Interpretability is one of the desired characteristics in various classification task. Rule-based system and fuzzy logic can be used for interpretation in classification. The main drawback of rule-based system is that it may contain large complex rules for classification and sometimes it becomes very difficult in interpretation. Rule reduction is also difficult for various reasons. Removing important rules may effect in classification accuracy. This paper proposes a hybrid fuzzy-rough set approach named RS-HeRR for the generation of effective, interpretable and compact rule set. It combines a powerful rule generation and reduction fuzzy system, called Hebbian-based rule reduction algorithm (HeRR) and a novel rough-set-based attribute selection algorithm for rule reduction. The proposed hybridization leverages upon rule reduction through reduction in partial dependency as well as improvement in system performance to significantly reduce the problem of redundancy in HeRR, even while providing similar or better accuracy. RS-HeRR demonstrates these characteristics repeatedly over four diverse practical classification problems, such as diabetes identification, urban water treatment monitoring, sonar target classification, and detection of ovarian cancer. It also demonstrates excellent performance for highly biased datasets. In addition, it competes very well with established non-fuzzy classifiers and outperforms state-of-the-art methods that use rough sets for rule reduction in fuzzy systems.

Keywords Pattern classification · Neuro-fuzzy system · Hebbian-based rule reduction · Rough set · Rule reduction

1 Introduction

Recent demands of interpretable machine learning [1–4] opened up new challenges to the research community. Majority of the human experts depend on rules in many classification tasks [5–8]. Such rule-based classifications are interpretable and highly accurate. Fuzzy systems are usually employed to model a practical problem that is too complex to be described using mathematical models [9, 10]. Acquired knowledge can be understood and readily explained using the **IF-THEN** fuzzy rules. The fuzzy logic [11] provides a mathematical framework to deal with uncertainty through employing fuzzy rules to represent the knowledge of the model. Fuzzy logic centric

methods widely applied in various application such as modeling of heat transfer mechanism [12], energy storage system [13], bio-medical engineering [14], and intelligent system design [15]. Olivas et al. [16] proposed ant colony optimization using type-2 fuzzy, Xiao et al. [17] used fuzzy theory in multi-criteria decision making, Nilashi et al. [18] proposed a fuzzy-based classifier applied in medical images. Fuzzy-based systems also used in many emerging applications such as neural network design [19] and quantum intelligence [20]. There are two important issues in fuzzy system domain; namely: the accuracy (the ability to classify patterns accurately), and the interpretability (the capability to describe the system in an understandable way). These two requirements are usually contradictory to each other [21]. Tuning the fuzzy membership functions of the rules can diminish the modeling error. But the interpretability may also be degraded as the fuzzy sets may drift closer to each other and end up overlapping [22]. Thus, a balance between accuracy and interpretability is desirable.

✉ Dilip K. Prasad
dilip.prasad@uit.no

¹ Nanyang Technological University, Nanyang Avenue, Singapore, Singapore

² UiT The Arctic University of Norway, Tromsø, Norway

Rule reduction is one of the research objectives in various rule-based systems. In some analyses, too many useless rules may be a concern in a rule-based system. In this area of researches, data-driven methods [23, 24] are popular but it is data dependent. A Hebbian-based Rule Reduction (HeRR) method has been proposed for this purpose [25]. In the HeRR method, consistent and interpretable rule set is generated through an iterative tuning and reduction process [26].

Although the iterative tuning and reduction process aims to balance the accuracy and interpretability, a good compromise is not always achievable. Furthermore, HeRR may result into a rule set with some amount of redundancy. Further reduction of the rule set is necessary to reduce the model complexity while maintaining high accuracy [24]. It is desirable to modify the previous HeRR system to derive an effective (high accuracy), interpretable (easy to understand) and compact (less redundancy) rule set.

This paper presents a Rough-Set-based Hebbian Rule Reduction (RS-HeRR) system to automatically generate effective, interpretable and compact fuzzy rules for classification problem. In order to maintain the interpretability of HeRR method, the reduction part of the HeRR is performed without any tuning process. To improve the accuracy, a post-processing step is used to polish the generated rule set. This step is implemented using the theory of rough sets. Rough set is a mathematical tool to deal with vague concepts and is domain independent. Hence, it is highly suitable for this problem [27, 28].

Here, we discuss the previously proposed fuzzy systems that exploit the rough set theory [29, 30] and form the basis of RS-HeRR. One of them is the rough-fuzzy multi-layer perceptron (MLP) [31–33]. The method successfully applied in robot modeling [34], controlling nonlinear systems [35], and demand prediction [36]. In the rough-fuzzy MLP, the domain knowledge is encoded in the connection weights and rules are generated from the decision table by computing the relative reducts. This model is further pursued using a modular approach in [37, 38]. A multi-class problem is divided into multiple 2-class sub-problems. The rough-fuzzy MLP modules, generated for each class, are subsequently combined to formulate fuzzy rules automatically using genetic algorithm.

Recently developed granulation approach computes granules of fuzzy rules are through rough set approach [39]. While rough sets are used directly in the definition and formation of the fuzzy neural networks, rough sets are not used for direct reduction of rules and attributes. Another rough-fuzzy approach for the identification of fuzzy rules is proposed in [40], which corresponds to tandem approach of hybridization of fuzzy neural

networks and rough sets as mentioned in [29]. This types of architectures are also used with support vector machine [41]. The proposed approach is also a tandem approach. The reduction approach of [40] uses the rough set-based attribute reduction algorithms (QuickReduct and QuickReduct II) to reduce the attributes of fuzzy rules derived from the rule induction algorithm (RIA) [42]. This approach attempts to compute a minimal reduct without exhaustively generating all possible subsets of attributes. However, it does not guarantee to always generate a minimal reduct. Further improvement for this approach is proposed in [40]. It uses the fuzzy-rough set [43] to provide better guidance in the selection of attributes. A hybrid of fuzzy neural network and the rough set, called rough sets-based outer-product (RSPOP), has been proposed in [44, 45] and extended in [46]. The RSPOP algorithm utilizes the rough set theory to perform the reduction of the attributes of the fuzzy rules generated by the pseudo outer-product (POP) rule identification algorithm [47, 48]. It uses the accuracy which is derived using subsets of attributes, to guide the computation of the relative reduct. We note that attribute reduction of RS-HeRR can be incorporated into RF-MLP and granulation approaches for applications such as medical image segmentation problems [49, 50].

The main contributions of the paper are as follows:

1. We have proposed a generalized rule generation and reduction method, namely RS-HeRR. The method produce state-of-the-art results in four variety of classification tasks without compromising accuracy.
2. Motivated by the existing methods, RS-HeRR performs optimization of accuracy, interpretability, and compactness of rule set using a rough-set-based attribute set selection approach using partial dependency.

The main novelty of our method is that we pay attention to the interpretability of learning methods. We have introduced an automatic rule generation and reduction module to build a compact learning framework. The method utilizes a combination of fuzzy logic and rough-set to achieve the goal. We made a theoretical foundation of the proposed method and also applied on four public classification dataset. The results conclude that the method is generic and can be applied to different classification tasks.

This paper is organized as follows. Section 2 describes a summary of the Hebbian-based Rule Reduction (HeRR) algorithm and Sect. 3 present the proposed Rough-Set-based Hebbian Rule Reduction (RS-HeRR) system. Four sets of experimental results are presented in Sect. 4 to benchmark and demonstrate the effectiveness of the proposed method; Sect. 5 presents the conclusion.

2 Hebbian-based rule reduction system

Fuzzy models usually consist of four components, a fuzzifier, a rule base, an inference engine and a defuzzifier. Crisp inputs are fuzzified to fuzzy inputs through a fuzzifier. Fuzzy inference is performed in the inference engine based on a set of fuzzy rules, as shown in Fig. 1.

It transforms the fuzzy inputs to the fuzzy outputs. These fuzzy outputs are defuzzified to the crisp outputs through a defuzzifier. The crisp input and output vectors are represented as $\mathbf{X}^T = [x_1, x_2, \dots, x_i, \dots, x_{n1}]$, $\mathbf{Y}^T = [y_1, y_2, \dots, y_i, \dots, y_{n5}]$, where $n1$ and $n5$ are the input and output dimensions, respectively [3]. Gaussian membership functions (MF) are used for fuzzification and defuzzification in this paper. Although other types of membership functions such as triangular, trapezoidal, etc. can also be used in our case. We have used Gaussian MF because of its advantage of being smooth and nonzero at all points. The centroids and widths of the MFs are variables in the framework and denoted as $(c_{i,j}^{in}, \delta_{i,j}^{in})$, the i -th input dimension and the j -th MF, and $(c_{l,m}^{out}, \delta_{l,m}^{out})$, the l -th output dimension and the m -th MF, for the input and output counterparts of the rules, respectively. Denote $iL_k(i)$ as the input label in the i -th input dimension of the k -th rule, and as the output label in the m -th output dimension of the k -th rule. Although the framework of Fig. 1 is valid for multi-input-multi-output (MIMO) systems, we consider only classification problems which are multi-input-single-output (MISO). The centroids of the single output are denoted as c_l^{out} (l -th MF), and the widths of the output MF are not involved in the computation. The output label of the k -th rule is represented as oL_k . The minimum and the maximum functions are used for fuzzy union and intersection operators, respectively. Given the input vector $\mathbf{X}^T = [x_1, x_2, \dots, x_i, \dots, x_{n1}]$, the output of the fuzzy system can be expressed as Eq. (1), where f_k is called the firing strength of the input point \mathbf{X}^T by the k -th rule and define in Eq. (2).

$$o = c_{k'}^{out}, k' = \arg \max_k (f_k) \tag{1}$$

$$f_k = \min_i \left(\exp \left(\frac{(x_i - c_{i,iL_k(i)}^{in})^2}{(\delta_{i,iL_k(i)}^{in})^2} \right) \right) \tag{2}$$

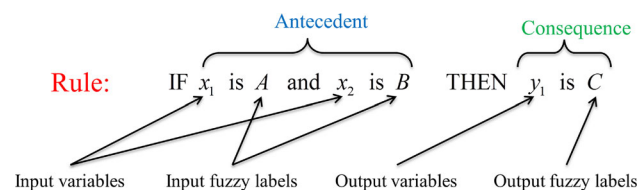


Fig. 1 Fuzzy rule (Mamdani type)

Hebbian-based rule reduction neuro-fuzzy system [25] consists of 4 steps shown in Fig. 2, namely initial rule set generation, rule ranking, membership functions merging, and redundancy removal. We discuss each step in detail as this framework is used directly by RS-HeRR.

The initial rule set is generated from the training data samples. If there are no rules in the rule base or no rules with sufficient firing strength (threshold θ), a new rule node is created. When a new rule is created, the centroids of the newly generated MFs for the input dimensions are assigned the value of the current data sample, while the widths of the MFs are set to a predefined value in proportion to the scale of each dimension. The centroid of the output dimension is assigned a value equal to the class label. The flowchart of the procedure is described in Fig. 3.

In the next step, all the initial rules are ranked and sorted based on their Hebbian importance. In fuzzy modeling, each fuzzy rule corresponds to a sub-region of the decision space. Some rules lying in an appropriate region may represent many samples and have much influence on the final result, while some other rules may occasionally be generated by noise and become redundant in the rule base. As the membership functions of a rule are determined, the training sample (\mathbf{X}^T, y_i) is fed into both the condition and consequence parts of the rule simultaneously. The input \mathbf{X}^T is used to determine the firing strength of the rule. If the class label of the rule (the centroid of the consequence part) is the same as y_i , it is given a membership of 1. Otherwise, the membership is assigned 0. If the input-output samples repeatedly fire a rule by the product of their firing strength and the membership values of consequence part, such that the accumulated strength surpasses that of other rules, it indicates the existence of such a rule. The larger the product is, the more important the rule is deemed. The Hebbian importance of the k -th rule is defined in Eq. (3).

$$I_k = \sum_{i=1}^n f_{k,i} \times \mu_{oL_k}(y_i) \tag{3}$$

where $f_{k,i}$ is the firing strength of the k -th rule on the i -th data point and

$$\mu_{oL_k}(y_i) = \begin{cases} 0 & \text{if } c_k^{out} \neq y_i \\ 1 & \text{if } c_k^{out} = y_i. \end{cases}$$

The fuzzy rules are sorted in a decreasing order of I_k . This is called the Hebbian ordering of the rules.

The next step is merging of membership functions and rules reduction by redundancy and inconsistency removal, which are performed iteratively. Their flowchart is shown in Fig. 4. In the beginning of this flowchart, Hebbian ordered rules are contained in the original rule set R and the initial reduced rule set R' is null. If there is no rule in the reduced rule set, the rule with highest Hebbian importance

Fig. 2 The four procedures of the Hebbian-based rule reduction neuro-fuzzy system

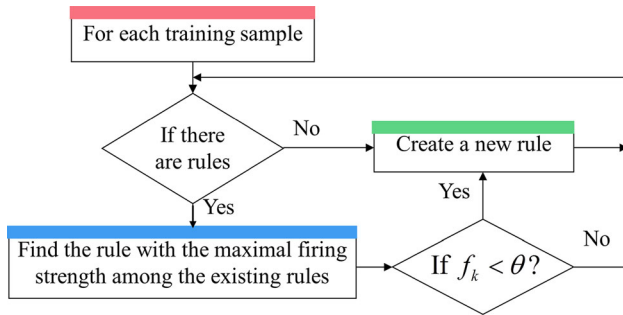
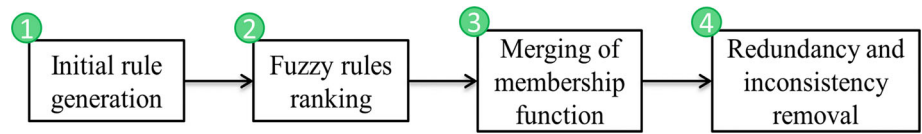


Fig. 3 Flowchart of rule initialization algorithm

in R is added into R' , otherwise, the fuzzy set in each dimension of the rule is added or merged into the fuzzy sets in R' according to a set of criteria defined below. All the newly added or merged fuzzy sets are linked together to formulate a new rule in the reduced rule set R' .

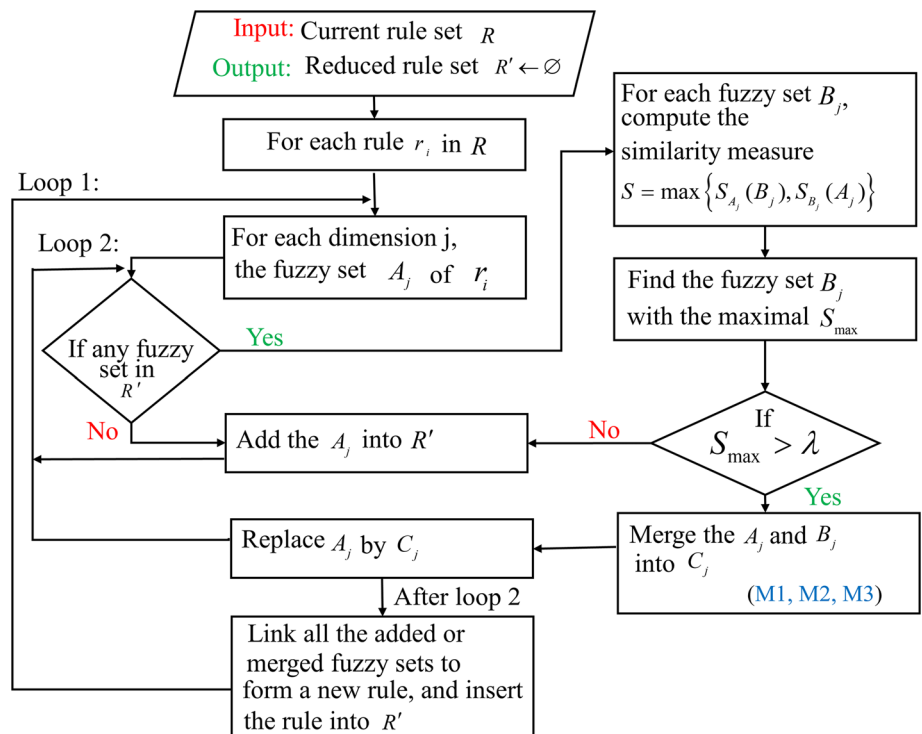
Since some fuzzy sets may be shared by several rules, changing a fuzzy set is equivalent to modifying these rules simultaneously. Thus, to merge the MFs, the fuzzy set associated with each MF is assigned an importance value which is equal to the sum of importance of the rules which share the fuzzy set. Let us denote the importance value of a

fuzzy set F as \hat{I}_F . The value of \hat{I}_F changes during the merging process. Each time the k -th rule (i.e., the k -th rule is “If $x_i = A_i$ then $y_1 = C_k$ ”) in the original rule set R is presented, its fuzzy sets $A_i(i = 1..n_1)$ in the condition part, would have the same importance with the Hebbian importance of the associated rule and is described in Eq. (4).

$$\hat{I}_{A_i} = I_k(i = 1..n_1) \tag{4}$$

As the consequence part of a rule is the class label, C_1 does not participate in the merging process. In each input dimension i , among the fuzzy sets in the reduced rule set R' of this dimension, the fuzzy set A'_i with the maximum degree of overlap over A_i (the degree of overlap is defined in Eq. (8)) is selected. If the maximum degree of overlap satisfies a specified criterion, they are merged into another fuzzy set A''_i , otherwise, the fuzzy set A_i is directly added into the reduced rule set of this dimension. The centroid and the width of A_i and A'_i ; namely: (c_{A_i}, δ_{A_i}) and $(c_{A'_i}, \delta_{A'_i})$, respectively, are merged into $(c_{A''_i}, \delta_{A''_i})$ using Eqs. (5) and (6).

Fig. 4 Flowchart of merging of membership functions and rule reduction algorithm



$$c_{A''_i} = \frac{\hat{I}_{A_i} c_{A_i} + \hat{I}_{A'_i} c_{A'_i}}{\hat{I}_{A_i} + \hat{I}_{A'_i}} \tag{5}$$

M2:

$$\delta_{A''_i} = \frac{\hat{I}_{A_i} \delta_{A_i} + \hat{I}_{A'_i} \delta_{A'_i}}{\hat{I}_{A_i} + \hat{I}_{A'_i}} \tag{6}$$

The importance of the new fuzzy set A''_i is determined using Eq. (7).

M3:

$$\hat{I}_{A''_i} = \hat{I}_{A_i} + \hat{I}_{A'_i} \tag{7}$$

In other words, the centroid and variance of the resultant fuzzy set is the weighted average of the two fuzzy sets in accordance with their importance, and the degree of importance of this final fuzzy set is the sum of the importance of the two. Subsequently the newly generated fuzzy set A''_i replaces the previous fuzzy set A'_i . Finally, the newly added or merged fuzzy sets in all the dimensions are linked together to form a fuzzy rule in the reduced rule set R' . Given fuzzy sets A and B , the degree of overlap of A by B is defined as in Eq. (8).

$$S_A(B) = \left| \frac{A \cap B}{A} \right| \tag{8}$$

As the Gaussian membership function is used in the proposed method, the overlap measure can be derived using the centroids and widths of the MFs. For fuzzy sets A and B , with membership functions given in Eqs. (9), (10), respectively, where c_A and c_B are their centroids, δ_A and δ_B are their widths; and assuming that $c_A \geq c_B$, then $|A|$ and $|A \cap B|$ can be expressed in Eqs. (11) and (12).

$$\mu_A(x) = \exp \left\{ \frac{-(x - c_A)^2}{\delta_A^2} \right\} \tag{9}$$

$$\mu_B(x) = \exp \left\{ \frac{-(x - c_B)^2}{\delta_B^2} \right\} \tag{10}$$

$$|A| = \sqrt{\pi} \delta_A \tag{11}$$

$$|A \cap B| = \frac{1}{2} \left(\frac{h^2 [c_B - c_A + \sqrt{\pi}(\delta_A + \delta_B)]}{\sqrt{\pi}(\delta_A + \delta_B)} + \frac{h^2 [c_B - c_A + \sqrt{\pi}(\delta_A - \delta_B)]}{\sqrt{\pi}(\delta_B - \delta_A)} + \frac{h^2 [c_B - c_A - \sqrt{\pi}(\delta_A - \delta_B)]}{\sqrt{\pi}(\delta_A - \delta_B)} \right) \tag{12}$$

where $h(x) = \max\{0, x\}$. The criterion of merging is as follows, if $S_A(B) > \lambda$ or $S_B(A) > \lambda$, the two fuzzy sets A and B are merged, where λ is a specified threshold that determines the maximum degree of overlap between fuzzy sets that the system can tolerate. Higher λ may increase the

accuracy but degrade the interpretability, while lower λ may force a larger number of rules to be reduced but at the risk of the number of rules being less than necessary to maintain the required modeling accuracy. In the algorithm, the current rule set R consists of the fuzzy rules derived from the rule initialization algorithm and the reduced rule set R' is the resultant rule set and is set to null at the beginning. There are two loops in the flowchart, where the outer loop is for each fuzzy rule r_i and the inner loop is for each input dimension j . If some fuzzy set of the dimension j in the R' is very similar to the fuzzy set of dimension j of the rule r_i (i.e., the similarity measure S is larger than a predefined value), these two fuzzy sets are merged into a new fuzzy set according to Eqs. (5) and (6). Otherwise, the fuzzy set of dimension j of the rule r_i are added into the reduced rule set R' . All the newly merged or added fuzzy sets are linked together and inserted into R' . The last step is the redundancy and inconsistency removal. After the merging process, the reduced rule set R' consists of a set of modified rules. The following three steps are used to remove redundancy in R' :

1. If there is only one membership function within one dimension, this dimension (feature) is removed;
2. If there is any rule that has the same conditions and consequences with the others, it is removed;
3. If there are any conflicting rules that have equivalent conditions but different consequences, the one with the higher degree of importance is preserved and the others are deleted.

3 RS-HeRR: the hybrid of Hebbian-based rule reduction system and rough set

Although the Hebbian rule reduction method is proposed to remove redundant rules and attributes through the merging process, redundancy may still remain. For example, if a conditional attribute, which is unrelated with the decision attribute, cannot be merged into a membership function, this conditional attribute will become redundant for the rule set. Thus, further reduction of the generated rule set is desirable. This is the goal of RS-HeRR. Rule set is initially generated through the Hebbian rule reduction algorithm and the rough set theory is used as a post-processing step to further reduce the generated rule set without affecting the classification accuracy.

As discussed in the introduction, the proposed RS-HeRR uses both the partial dependency (used in QuickReduct algorithms [40]) and the system’s performance (used in RSPOP [44, 45]) to select attributes and reduce fuzzy rules. QuickReduct algorithms aim to find the ‘reduct’ of the rough set through rule selection based on partial

dependency. We describe the concept of dependency, and partial dependency below.

In the theory of rough sets, a knowledge base can be represented as a relational system $K = (\mathbf{U}, \mathbf{R})$, where $\mathbf{U} \neq \emptyset$, called the universe, is a set of objects, and \mathbf{R} is a set of relations. In the knowledge base, \mathbf{R} particularly consists of a set of attributes, each of which forms an equivalence relation. For any $\mathbf{P} \subseteq \mathbf{R}$, an associated equivalence relation, called indiscernibility relation over P , denoted by $IND(\mathbf{P})$, is formed in Eq. (13):

$$IND(\mathbf{P}) = \left\{ (x, y) \in U^2 \mid \forall a \in \mathbf{P}, a(x) = a(y) \right\} \tag{13}$$

The family of all equivalence classes of the equivalence relation $IND(\mathbf{P})$ is denoted by $\mathbf{U}/IND(\mathbf{P})$. With each subset $X \subseteq U$ and an equivalence relation R , the rough set theory uses the R -lower ($\underline{R}X$) and R -upper ($\overline{R}X$) approximations to approximate X , as described in Eqs. (14) and (15).

$$\underline{R}X = \cup \{Y \in \mathbf{U}/R \mid Y \subseteq X\} \tag{14}$$

$$\overline{R}X = \cup \{Y \in \mathbf{U}/R \mid Y \cap X \neq \emptyset\} \tag{15}$$

By employing the knowledge R , $\underline{R}X$ is the set of all objects of \mathbf{U} which can be classified with certainty as the elements of X , and $\overline{R}X$ is the set of objects of \mathbf{U} which can be possibly classified as elements of X . Let P and Q be two equivalence relations in \mathbf{U} , the positive, negative and boundary regions are defined by Eqs. (16)–(18):

$$POS_P(Q) = \cup_{X \in \mathbf{U}/Q} \underline{P}X \tag{16}$$

$$NEG_P(Q) = \mathbf{U} - \cup_{X \in \mathbf{U}/Q} \overline{P}X \tag{17}$$

$$BND_P(Q) = \cup_{X \in \mathbf{U}/Q} \overline{P}X - \cup_{X \in \mathbf{U}/Q} \underline{P}X \tag{18}$$

For a knowledge base, dependency between attributes may exist. Let $\mathbf{P}, \mathbf{Q} \subseteq \mathbf{R}$. \mathbf{Q} depends on \mathbf{P} if $IND(\mathbf{P}) \subseteq IND(\mathbf{Q})$. The dependency can also be partial which means that only part of \mathbf{Q} depends on \mathbf{P} . For a partial dependency, \mathbf{Q} depends on \mathbf{P} in a degree $\gamma_{\mathbf{P}}(\mathbf{Q})$ ($0 \leq \gamma_{\mathbf{P}}(\mathbf{Q}) \leq 1$), described in Eq. (19).

$$\gamma_{\mathbf{P}}(\mathbf{Q}) = \frac{card\ POS_{\mathbf{P}}(\mathbf{Q})}{card\ \mathbf{U}} \tag{19}$$

Where $card()$ is the cardinality of the input set, i.e., $card\ POS_{\mathbf{P}}(\mathbf{Q})$ and $card\ \mathbf{U}$ denote the number of elements in $POS_{\mathbf{P}}(\mathbf{Q})$ and \mathbf{U} , respectively.

A simple depiction of RS-HeRR is given in Fig. 5. Initially, when training samples are available, a rule set is generated by the Hebbian rule reduction method. Then, using this set of rules (denoted as $RuleSet(A, C)$), the accuracy due to each attribute is also computed. Similar to QuickReduct, an additive approach is used for building the

new rule set, however, guided primarily by the classification accuracy. This implies that the new attribute set R is initially empty. The attribute which single handedly gives the best classification accuracy is the attribute first transferred to R . It is investigated if the accuracy of R is poorer than the accuracy of $RuleSet(A, C)$. For each remaining attribute $\{a\}$, the classification accuracy of $R \cup \{a\}$ is computed. The attribute, whose inclusion gives the best accuracy is included in R . In the case of a tie for best accuracy, partial dependency $\gamma_{R \cup \{a\}}(C)$ and $\gamma_R(C)$ is computed for the attributes resulting into the tie. The attribute with maximum partial dependency is added to R . If there is a situation of tie in the partial dependency as well, one of them is randomly chosen. The algorithm terminates when the partial dependency becomes one and the classification rate using the subset of attributes is not worse than that using the whole set of the attributes. Conceptually, it implies that RS-HeRR tackles the conditional attribute for redundancy handling which is uncorrelated with the decision attributes (i.e., contributors to accuracy). The pseudo-code of the attribute selection algorithm is presented in Algorithm 1. We also present an example in Sect. 4.1 to explain the procedure.

Algorithm 1 Rough-set based attribute selection

Require: $RuleSet(A, C)$, the rule set formed by the condition attribute set A and class attribute C . $ACCU(RuleSet(A, C))$, the classification accuracy of the fuzzy system using the rule set formed by attribute set A and C .

```

1:  $O_A \leftarrow ACCU(RuleSet(A, C))$ 
2:  $R \leftarrow \phi$ 
3: do
4:    $T \leftarrow R$ 
5:   for each:  $a \in (A - R)$ 
6:      $O_{R \cup \{a\}} = ACCU(RuleSet(R \cup \{a\}, C))$ 
7:     if  $O_{R \cup \{a\}} > O_T$  OR  $(O_{R \cup \{a\}} = O_T$  AND  $O_{R \cup \{a\}}(C) > \gamma_T(C))$  then
8:        $T \leftarrow R \cup \{a\}$ 
9:     end if
10:   $R \leftarrow T$ 
11: while  $\#1$ 
12: Return  $R$ 

```

4 Experimental results

In this section, we use four binary classification problems and demonstrate the effectiveness of RS-HeRR in meeting the three requirements, namely classification performance, interpretability, and compactness. The datasets of the classification problems are detailed in individual sub-sections but a brief summary is provided in Table 1. The amount of cross-folding validation (CV) used and the ratio of number of training data samples (#Train) to the number of test data samples (#Test) is given in Table 1. Notably, we use smaller subset for training as compared to the test subset. The purpose of using the smaller subset of data as the training set is to test and compare the generalization ability of the proposed system and other benchmarked models.

Fig. 5 The hybrid of the Hebbian rule reduction fuzzy system and the rough set

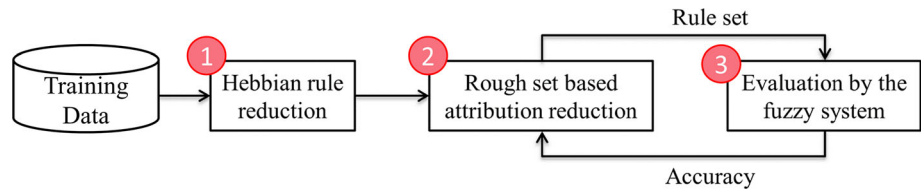


Table 1 The details of the datasets and the associated binary classification problem

	Pima Indian diabetes dataset	Urban water treatment plant monitoring dataset	Sonar dataset	Ovarian cancer dataset
Description	Medical data of female patients of at least 21 years old Pima Indian heritage	Data measured from an urban waste water treatment plant over a period of 527 days	Processed data on each antenna in the sonar array	Gene patterns from DNA microarray gene expression
Positive class	Patients having diabetes	Normal function	Target is metallic	Gene is from an ovarian tumor
Negative class	Patients not having diabetes	Abnormal event	Target is rocky	Gene is from a normal tissue
#of features	8	38	60	1536
#of data samples	768	527	208	54
#of positive sample	500	513	111	30
#of negative samples	268	14	97	24
Cross-validation	10	10	10	3
#Train: #Test	1:9	1:9	1:9	1:2

Classification performance is quantitatively assessed using overall classification accuracy over both classes, specificity for the positive class, and sensitivity for the negative class. For binary classification problem, the sensitivity is the accuracy for the positive class, and the specificity is the accuracy for the negative class. They are computed using Eqs. (20) and (21), respectively.

$$\text{Sensitivity} = \frac{\text{Number of positive samples correctly predicted}}{\text{Total number of positive samples}} \tag{20}$$

$$\text{Specificity} = \frac{\text{Number of negative samples correctly predicted}}{\text{Total number of negative samples}} \tag{21}$$

We compare the classification performance of the RS-HeRR with Support Vector Machine (SVM) [51], the C4.5 decision tree algorithm [52], the Naïve Bayesian classifier (NB) [53], and the Rough-Set-based Pseudo Outer-Product fuzzy rule identification algorithm (RSPOP) [45]. Our selection of these methods for comparison is explained as

follows. For comparison of accuracy, we consider the classification methods that concern with different formats of knowledge representation and concern with achieving accuracy rather than factors such as interpretability or compactness. Our selection was also guided by the availability of the codes or executables of the methods being compared.

Interpretability is measured by the number of fuzzy rules used for representing the knowledge. For other factors concerning interpretability remaining the same, such as the type of membership functions, the architecture, etc., the interpretability due to the attribute selection strategy is better if the number of rules after selection is less than the number of rules obtained using other attribute selection strategies. Similarly, compactness of the fuzzy system is characterized by the number of attributes as well as the number of rules. Here also, a fair comparison requires that other factors of the fuzzy system are kept the same and the compactness due to the attribute selection strategies is compared. Thus, for analyzing the advantage of RS-HeRR

regarding interpretability and compactness, we use same initial fuzzy systems as obtained using the basic HeRR strategy explained in Sect. 2 and compare RS-HeRR with the cases where no attribute is used (original HeRR) or QuickReduct and QuickReduct II are used for attribute selection. We refer to this comparison as attribute selection comparison.

4.1 Pima indian diabetes dataset

The Pima-Indian Diabetes dataset [54] consists of a total of 768 data instances, all of which are for the patients who are female of at least 21 years old Pima Indian heritage. There are 8 input features for each instance, listed in Table 2. Figure 6 shows the benchmark results of the RS-HeRR against the SVM, C4.5, NB, and RSPOP. The RS-HeRR system achieve the highest overall testing accuracy, though its sensitivity is lower than that of SVM by only 4.36 and its specificity is worse than that of C4.5 and NB by less than 4.37%. This poor performance is attributed to the use of Mamdani model of fuzzification, which favors interpretability over accuracy.

The results for attribute selection comparison are shown in Table 3. The overall accuracy, the number of generated rules and the number of selected features are given in the table for comparison. The classification accuracy of the HeRR without rough set is the worst among these models, and the number of rules and features are much higher than others. It indicates that the redundancy in the rule set could have degraded the testing accuracy of the HeRR system and the removal of redundant rules and features is necessary to maintain and boost the performance. The RS-HeRR system achieves a decrease of 58.37% for the number of rules and 60.76% for the number of features, compared with the HeRR system without any feature selection process. The decrease is larger than the HeRR system plus QuickReduct and QuickReduct II. It indicates the rule set derived by RS-HeRR is more compact than that derived by HeRR plus QuickReduct & QuickReduct II. The experimental results thus show the superior performance of the proposed rough-set-based attribute selection.

After the Hebbian rule reduction algorithm shown in Fig. 2, a total of 17 fuzzy rules are derived. The classification accuracy on the training and testing sets are 80.52% and 71.35%, respectively. At the beginning of RS-HeRR, the attribute set A consists of all the attributes $x_1 - x_8$, and the output attribute set R is empty. For each attribute a in $A - R$, the training accuracy and the partial dependency using the rule set with the attribute set $R \cup \{a\}$ are shown in Table 4. It is seen that when the attribute x_2 is added to R , the training classification achieved the maximum among all the 8 attributes. Thus, R becomes $\{x_2\}$. As the training accuracy is worse than that using the whole attribute set,

the selection process continues. In the second round of the loop, the training accuracy and partial dependency of the remaining 7 attributes are shown in Table 5. In the table, both x_6 and x_8 achieves the best in training accuracy. However, the partial dependency of x_6 is larger than that of x_8 . Thus, the attribute x_6 is selected. As the training accuracy is still lower than 80.52%, which is achieved by the whole attribute set, the algorithm comes into the third round. The training accuracy and partial dependency are shown in Table 6. The best accuracy is achieved by adding x_7 to R . Hence, x_7 is selected. As the accuracy is not lower than 80.52% and the partial dependency becomes 1, the algorithm terminates. At the end of the algorithm, the attribute set $\{x_2, x_6, x_7\}$ is produced. Using the 3 attributes, the rule set is simplified into 6 rules by deleting repeated rules. The training accuracy is still maintained at 80.52%. However, the testing accuracy is boosted to 75.40%, which is better than before, due to the selection of the attributes.

The interpretability of the proposed RS-HeRR system is evaluated in Figs. 7, 8, 9 and Table 7. Figures show the membership functions of the attributes: x_2, x_6 and x_7 . x_2 is divided into 2 MFs (Low and High), and x_6 and x_7 are divided into 3 MFs (Low, Medium and High). It is shown that the divided MFs are all clearly separated and have distinguished semantic meanings. The resultant rule set which consists of a total of 6 fuzzy rules is shown in Table 8. From the rule set, it can be seen that, when the x_2 (Plasma glucose concentration) is low and the x_6 (Body mass index) and x_7 (Diabetes pedigree function) are not high, the patient will be tested as positive for diabetes, and when x_2 (Plasma glucose concentration) is high and the x_6 (Body mass index) is not low, the patient will be tested as negative for diabetes. The derived rule set can be easily understood by humans and used to assist the decision making of the clinicians.

4.1.1 Urban water treatment plant monitoring dataset

The urban water treatment dataset¹ consists of a set of historical data measured from an urban waste water treatment plant over a period of 527 days. The objective is to classify the operational state of the plant in order to predict faults through the state variables of the plant at each of the stages in the treatment process. In total, 38 input features for each data point, organized into 5 aspects, are used (see Table 8).

The status of the plant is represented by one of 13 different categories. Some represent the normal operation of the plant and others point out various faults of the plant. They are listed in Table 9. As all the faults appear in a short period and are dealt with immediately, there are not

¹ The UCI Machine Learning Database Repository.

Table 2 The description of the eight features of the Pima Indian diabetes dataset

Feature	Description
1	Number of times pregnant
2	Plasma glucose concentration a 2 h in an oral glucose tolerance test
3	Diastolic blood pressure (mm Hg)
4	Triceps skin fold thickness (mm)
5	2-h serum insulin (mu U/ml)
6	Body mass index (weight in kg/(height in m) ²)
7	Diabetes pedigree function
8	Age (years)

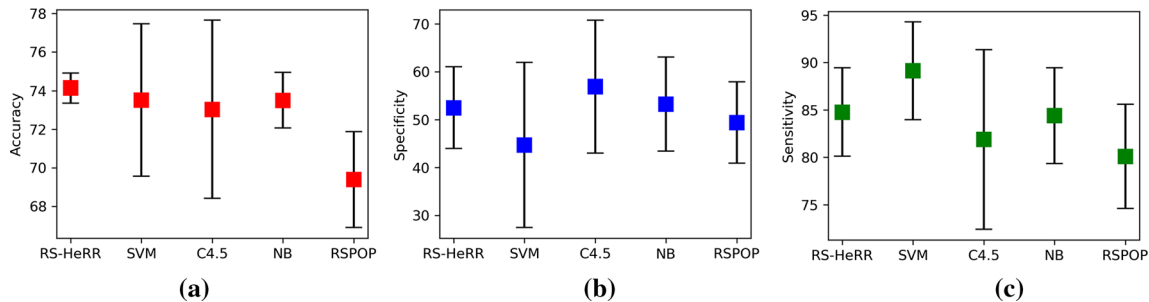


Fig. 6 Classification results on the Pima Indian diabetes dataset (statistics over 10 cross-fold validation)

Table 3 Results on the Pima Indian diabetes dataset, compared with the QuickReduct, QuickReduct II and without rough set reduction (statistics over 10 cross-fold validation)

Models	Pima Indian diabetes dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Dec %	Mean	Std.	Dec %
HeRR without RS	71.86	1.72	43.00	21.29	N/A	7.90	0.31	N/A
RS-HeRR	74.13	0.78	17.90	19.79	58.37	3.10	1.91	60.76
HeRR + QuickReduct	73.35	1.49	32.90	26.42	23.49	3.80	2.04	51.90
HeRR + QuickReductII	72.82	1.50	34.00	20.48	20.93	5.10	1.59	35.44

Table 4 The training accuracy and partial dependency using x_1 to x_8

Attributes	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Accuracy %	63.64	77.92	61.04	38.96	61.04	40.26	42.86	62.34
Dependency	0.00	0.67	0.33	0.33	0.00	0.50	0.20	0.20

Table 5 The training accuracy and partial dependency using $x_1, x_3, x_4, x_5, x_6, x_7, x_8$

Attributes	x_1	x_3	x_4	x_5	x_6	x_7	x_8
Accuracy %	77.92	77.92	76.62	76.62	79.22	77.92	79.22
Dependency	0.96	1.00	1.00	1.00	1.00	1.00	0.96

Table 6 The training accuracy and partial dependency using $x_1, x_3, x_4, x_5, x_7, x_8$

Attributes	x_1	x_3	x_4	x_5	x_7	x_8
Accuracy %	77.92	76.62	76.62	79.22	80.52	77.92
Dependency	1.00	1.00	1.00	1.00	1.00	1.00

enough training data points for these categories. However, for the statuses normal function (green shaded row in Table 9) and abnormal event (blue shaded row in Table 9), there are 513 and 14 data samples, respectively. Even after

clustering statuses corresponding to normal and abnormal performance, the dataset is a highly skewed dataset and thus it is difficult to obtain good specificity (accuracy for abnormal samples). Further, there are missing values for

Fig. 7 Derived fuzzy membership functions of attribute x_2

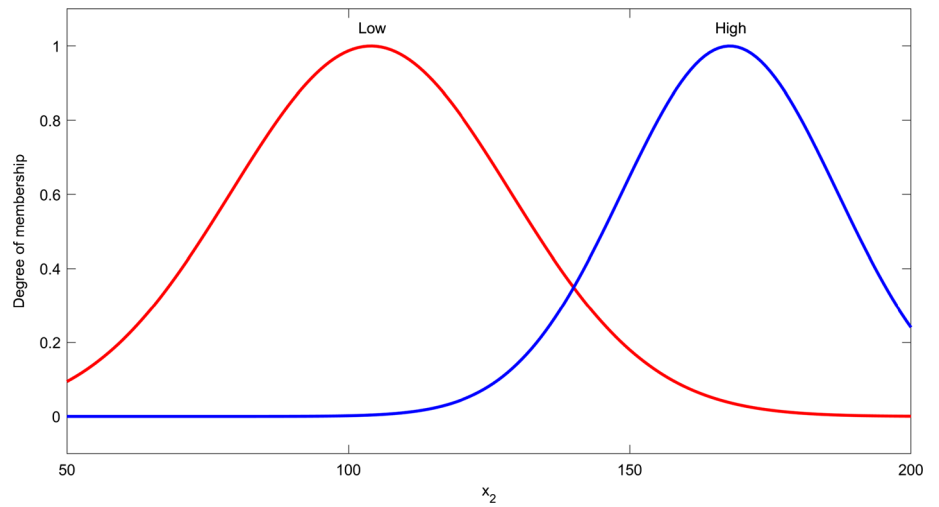


Fig. 8 Derived fuzzy membership functions of attribute x_6

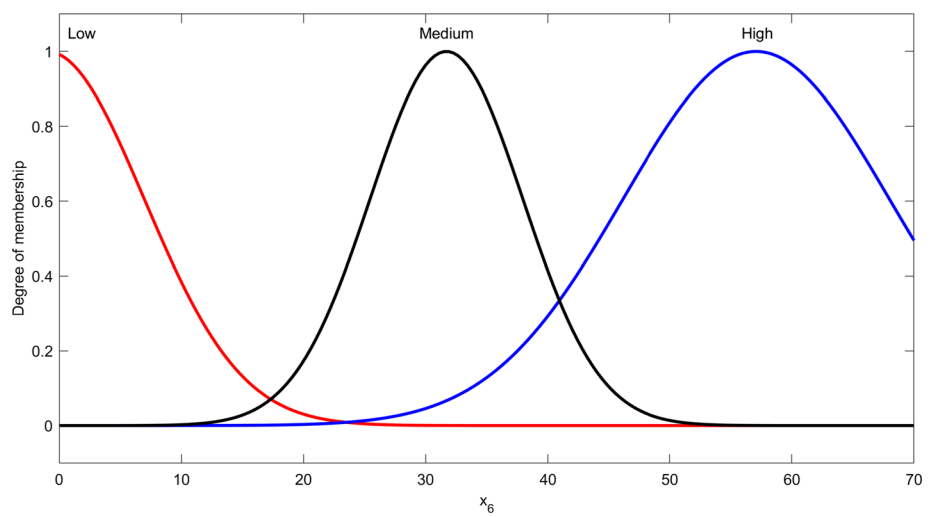
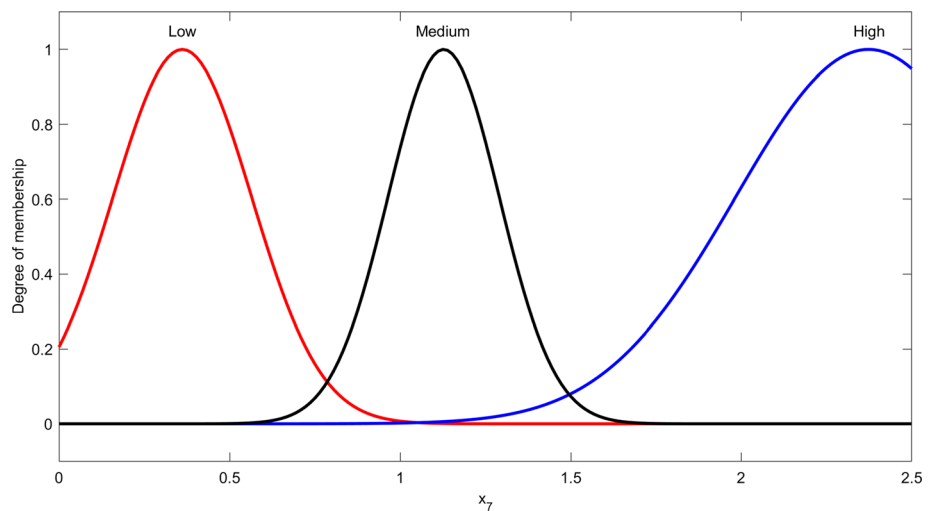


Fig. 9 Derived fuzzy membership functions of attribute x_7



some features in this dataset. The missing values are replaced by the average value of the corresponding

attribute in this experiment. This substitution may be considered as noise in the data samples.

Table 7 The six derived fuzzy rules

Rule #1:	IF x_2 is low and x_6 is medium and x_7 is medium THEN y is positive class
Rule #2:	IF x_2 is low and is x_6 medium and x_7 is low THEN is positive class
Rule #3:	IF x_2 is low and x_6 is low and x_7 is low THEN y is positive class
Rule #4:	IF x_2 is high and x_6 is medium and x_7 is low THEN y is negative class
Rule #5:	IF x_2 is high and x_6 is medium and x_7 is medium THEN y is negative class
Rule #6:	IF x_2 is high and x_6 is high and x_7 is high THEN y is negative class

Table 8 The five aspects of the input features

Aspect	Number of features
Input to plant	9
Input to primary settler	6
Input to secondary settler	7
Output from plant	7
Plant performance	9
Total	38

The classification results are shown in Fig. 10. RSPOP achieves a sensitivity of 100%, classifying all the positive samples correctly across all the CV groups. However its performance in classifying negative samples is bad, as its specificity is lower than others. In this experiment, RS-HeRR achieves the best specificity. It shows that the RS-HeRR system outperforms other models in overall testing accuracy, although its sensitivity is slightly lower than that of SVM, C4.5 and RSPOP. This is attributed to the property of RS-HeRR in favoring accuracy even while identifying the critically important attributes through partial dependency. The experimental results show the classification capability of the RS-HeRR in the dataset where the classes are unbalanced distributed.

Table 10 shows the attribute selection comparison results. The testing accuracy achieved by RS-HeRR is the same as that of HeRR without RS, while the HeRR plus QuickReduct and QuickReduct II achieve lower accuracy. Under the percentage decrease in the number of rules and features, compared with the HeRR without RS, the proposed RS-HeRR outperforms the HeRR with QuickReduct & QuickReduct II. We note that in some cross-validation sets, RS-HeRR, HeRR + QuickReduct, and HeRR + QuickReductII, all select only one feature. But, the selected feature in RS-HeRR is the one that provides the best accuracy rather than maximum partial dependency. This is critical since such set of CV may have only one or two negative samples. It indicates that the RS-HeRR derives a more compact rule set while still maintaining the accuracy.

4.2 Sonar dataset

The Sonar dataset [55] consists of a total 208 pattern samples with 60 input features and two classes (metal target and rock target). Figure 11 shows the classification results of RS-HeRR, compared with SVM, C4.5, NB, and RSPOP. The proposed RS-HeRR system yields an accuracy of $89.41 \pm 1.75\%$. The standard deviation of both the accuracy and sensitivity is lower than that of other models. In shows that classification capability of the proposed RS-HeRR is better than the others in this dataset.

Table 9 The 13 statuses of the water treatment plant

Categories	Number of samples	Description
1	275	Normal situation
2	1	Secondary settler problem, type 1
3	1	Secondary settler problem, type 2
4	4	Secondary settler problem, type 3
5	116	Normal situation with performance over the mean
6	3	Solids overload, type 1
7	1	Secondary settler problem, type 4
8	1	Storm, type 1
9	69	Normal situation with low influent
10	1	Storm, type 2
11	53	Normal situation
12	1	Storm, type 3
13	1	Solids overload, type 2

Blue shaded rows indicate abnormal events

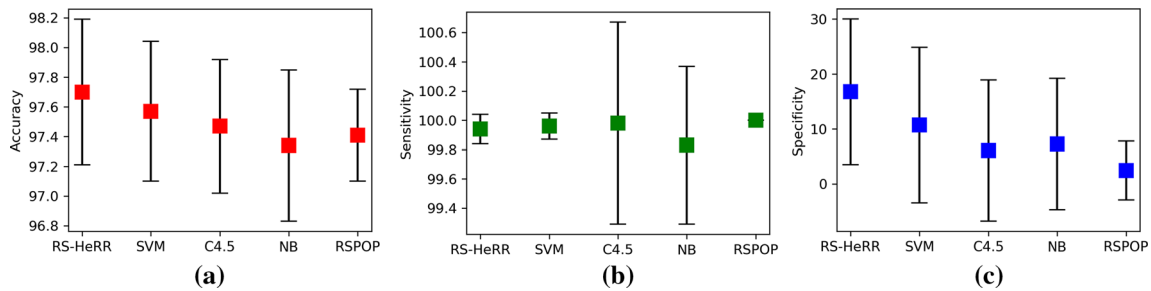


Fig. 10 Classification results on the urban water treatment dataset (statistics over 10 cross-fold validation)

Table 10 Classification results on the urban water treatment dataset, compared with the QuickReduct, QuickReduct II and without rough set reduction (statistics over 10 cross-fold validation)

Models	Urban water treatment dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Dec %	Mean	Std.	Dec %
HeRR without RS	97.70	0.52	27.50	13.67	N/A	23.60	8.62	N/A
RS-HeRR	97.70	0.49	3.50	3.87	87.27	1.10	0.32	95.34
HeRR + QuickReduct	97.68	0.49	4.80	7.27	82.45	1.20	0.42	94.92
HeRR + QuickReductII	97.53	0.63	8.40	13.03	69.45	1.20	0.42	94.92

Bold indicates this value is better in terms of criteria of performance metric

The attribute selection comparison results are given in Table 10. The HeRR without the RS attribute selection yields the highest testing accuracy, while the performance of the proposed RS-HeRR is at the second place. Notably, Table 11 shows that the HeRR system will involve all the 60 input features in the classification if no attribute selection process is incorporated. Such system is difficult to interpret. RS-HeRR only involves $11 \pm 4.58\%$ input features, at a cost of about 1% degradation in accuracy. The results show that, a balance between accuracy and interpretability is achieved by the proposed RS-HeRR. The HeRR with QuickReduct competes well with RS-HeRR. Although its accuracy is slightly lower than that of RS-HeRR, the number of both derived rules and selected features is less than that of RS-HeRR. This is because RS-HeRR prioritizes accuracy over partial dependency and the accuracy may inherently be poor in noisy measurements,

such as in the case of sonars which have poor signal to noise ratio in the presence of cluttered background. The HeRR with QuickReduct II derives the least number of rules, but its accuracy is much lower than others.

4.3 Ovarian cancer dataset

The ovarian cancer DNA microarray gene expression dataset [56] consists of a total of 54 patterns where 30 examples are derived from ovarian tumors and 24 examples are normal. Each of these examples comprises 1536 features. This is a good example of a highly complex and uninterpretable problem where attribute reduction and compact rule base is highly desirable.

In Fig. 12, the RS-HeRR yields an accuracy of $96.30 \pm 3.21\%$. The SVM, C4.5 and NB yield the same average accuracy, 90.74%, where the standard deviation of

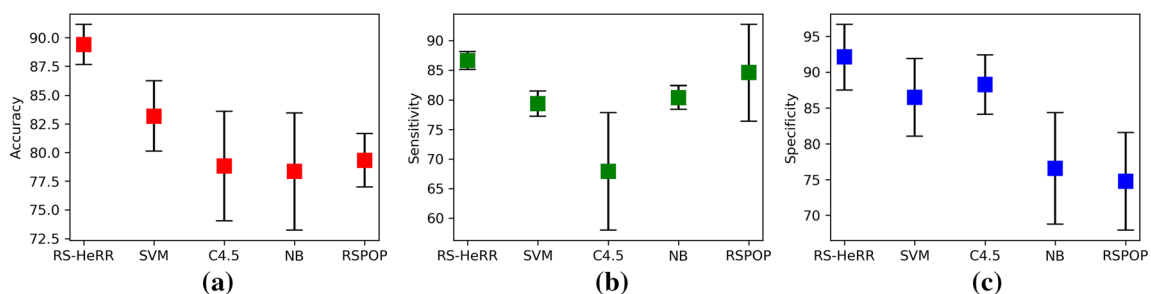


Fig. 11 Classification results on the sonar dataset (statistics over 10 cross-fold validation)

Table 11 Classification results on the sonar dataset, compared with the QuickReduct, QuickReduct II and without rough set reduction (statistics over 10 cross-fold validation)

Models	Sonar dataset								
	Accuracy %		No. of rules			No. of features			
	Mean	Std.	Mean	Std.	Dec %	Mean	Std.	Dec %	
HeRR without RS	90.86	2.23	135.33	1.15	N/A	60.00	0.00	N/A	
RS-HeRR	89.41	1.75	127.67	5.51	5.66	11.00	4.58	81.67	
HeRR + QuickReduct	88.94	0.79	118.00	5.57	12.80	5.33	0.58	91.11	
HeRR + QuickReductII	77.89	1.59	112.67	4.16	16.74	10.33	0.58	82.78	

Bold indicates this value is better in terms of criteria of performance metric

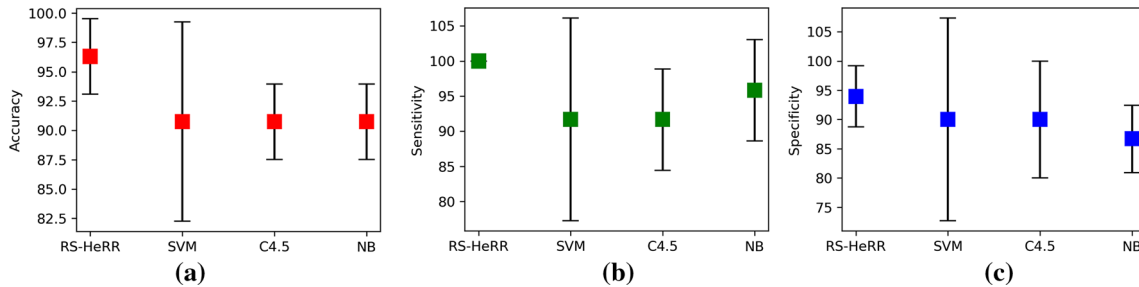


Fig. 12 Classification results on the ovarian cancer dataset (statistics over 3 cross-fold validation)

Table 12 Classification results on the ovarian cancer dataset, compared with the QuickReduct, QuickReduct II and without rough set reduction (statistics over 3 cross-fold validation)

Models	Ovarian cancer dataset								
	Accuracy %		No. of rules			No. of features			
	Mean	Std.	Mean	Std.	Dec %	Mean	Std.	Dec %	
HeRR without RS	70.37	6.41	36.00	0.00	N/A	14.95	58.20	N/A	
RS-HeRR	96.30	3.21	12.33	5.77	65.75	2.67	0.58	99.82	
HeRR + QuickReduct	96.30	3.21	17.67	9.29	50.92	3.33	1.53	99.78	
HeRR + QuickReductII	87.04	3.21	26.33	5.86	26.86	4.67	0.58	99.69	

the accuracy of C4.5 and NB is the same. Table 12 shows attribute selection comparison results. RS-HeRR selects only 2.67 ± 0.58 features out of the whole 1536 features for classification. This indicates the ability to represent the knowledge using about only three attributes (which corresponds to only 3 genes) and about only 12 rules (12 genetic combinations characterizing the presence or absence of ovarian cancer). However, the accuracy is enhanced from 70.37 ± 6.41 to 96.30 ± 3.21 . This is because of the elimination of the attributes that affect classification accuracy adversely. The performance of the RS-HeRR and HeRR with QuickReduct in the terms of testing accuracy is the same. However, the RS-HeRR selects and derives less number of features and less number of rules than that of HeRR with QuickReduct. The proposed attribute selection algorithm outperforms the QuickReduct, as the proposed algorithm makes the rule set more compact.

5 Conclusion

This paper proposes RS-HeRR, the hybridization of Hebbian rule reduction fuzzy system and rough set theory, for generating fuzzy rules for pattern classification problems. The rule set is initially generated using the Hebbian-based rule reduction algorithm. A post-processing step is used to further remove redundant attributes based on the rough set theory. In this step, a rough-set-based attribute selection algorithm is proposed to reduce and simplify the rule set without affecting the classification accuracy. It uses both the system’s performance and partial dependency as guides to determine suitable attributes subset. The key strengths of the proposed methods are summarized as follows:

1. Compact, effective and interpretable rule set: a compact, effective and interpretable rule set is generated from the trained fuzzy system. The redundancy is

removed from the rule set using rough-set-based attribute selection algorithm, while the classification performance is maintained and boosted. Membership functions in each dimension are clearly separated and the size of the resultant rule set is small. It makes the rule set readily understandable by human.

- No additional control parameter: HeRR needs only two control parameters, namely initial rule set generation threshold θ and the membership function merge threshold λ (see Sect. 2). RS-HeRR does not require any additional parameter.

The performance of the proposed rough-set-based Hebbian rule reduction system is evaluated using four datasets: (1) the Pima Indian diabetes dataset; (1) the urban water treatment plant monitoring dataset; (3) the sonar dataset; (4) the ovarian cancer dataset. The experimental results show good performance by the proposed method, when benchmarked against with other well established classifiers. In certain challenging classification problems, such as skewed dataset of the urban water treatment plant monitoring characterized with very few negative sample and the ovarian cancer dataset characterized by huge number of features, RS-HeRR shows a huge advantage. For instance, RS-HeRR can reduce the knowledge from the data characterized by 1536 features in the ovarian cancer dataset to identification of 3 genes and 12–14 genetic combinations that characterize the presence of ovarian cancer.

There are a few future avenues of the proposed method. First, we will apply a similar mechanism in the image classification domain. Next, we will consider extending the method with a combination of deep neural networks. This may open up new possibilities and challenges in the domain of interpretable and explainable artificial intelligence.

Acknowledgements Open Access funding provided by UiT The Arctic University of Norway.

Compliance with ethical standards

Conflict of interest Authors declare no conflict of interest

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aurangzeb AM, Carly E, Ankur T (2018) Interpretable machine learning in healthcare. In: Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics, pp 559–560. ACM
- Arras L, Horn F, Montavon G, Müller K-R, Samek W (2017) “What is relevant in a text document?”: an interpretable machine learning approach. *PLoS One* 12(8):e0181142
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci* 116(44):22071–22080
- Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG (2018) An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 46(4):547–553
- Antonelli M, Bernardo D, Hagrass H, Marcelloni F (2016) Multiobjective evolutionary optimization of type-2 fuzzy rule-based systems for financial data classification. *IEEE Trans Fuzzy Syst* 25(2):249–264
- Gorzalczany MB, Rudziński F (2016) A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability. *Appl Soft Comput* 40:206–220
- Manescu P, Lee YJ, Camp C, Cicerone M, Brady M, Bajcsy P (2017) Accurate and interpretable classification of microspectroscopy pixels using artificial neural networks. *Med Image Anal* 37:37–45
- Rudziński F (2016) A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers. *Appl Soft Comput* 38:118–133
- Koczy LT, Medina J, Reformat M, Wong KW, Yoon JH (2019) Computational intelligence in modeling complex systems and solving complex problems. *Complexity* 2019:1–6
- Kostikova AV, Tereliansky PV, Shuvaev AV, Parakhina VN, Timoshenko PN (2016) Expert fuzzy modeling of dynamic properties of complex systems. *ARN J Eng Appl Sci* 11(17):10601–10608
- Zadeh Lotfi A (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans Syst Man Cybern* 1:28–44
- Krzywanski J, Nowak W (2016) Modeling of bed-to-wall heat transfer coefficient in a large-scale cfb by fuzzy logic approach. *Int J Heat Mass Transf* 94:327–334
- Sahu A, Patel US (2017) Modelling & simulation of fuzzy logic based controller for energy storage system. *J Electron Design Technol* 8(2):9–15
- Teodorescu H-NL, Kandel A, Jain LC (2017) Fuzzy logic and neuro-fuzzy systems in medicine and bio-medical engineering: a historical perspective. In: *Fuzzy and neuro-fuzzy systems in medicine*. CRC Press, Taylor & Francis Group, pp 1–18
- Khan E (2020) Neural fuzzy based intelligent systems and applications. In: *Fusion of neural networks, fuzzy systems and genetic algorithms*. CRC Press, Taylor & Francis Group, pp 105–140
- Olivas F, Valdez F, Castillo O, Gonzalez CI, Martinez G, Melin P (2017) Ant colony optimization with dynamic parameter adaptation based on interval type-2 fuzzy logic systems. *Appl Soft Comput* 53:74–87
- Xiao F (2019) EFMCDM: evidential fuzzy multicriteria decision making based on belief entropy. *IEEE Trans Fuzzy Syst*. <https://doi.org/10.1109/TFUZZ.2019.2936368>
- Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L (2017) A knowledge-based system for breast cancer classification using fuzzy logic method. *Telemat Inform* 34(4):133–144

19. Duan L, Wei H, Huang L (2019) Finite-time synchronization of delayed fuzzy cellular neural networks with discontinuous activations. *Fuzzy Syst Systs* 361:56–70
20. Zhang W-R (2017) From equilibrium-based business intelligence to information conservational quantum-fuzzy cryptography—a cellular transformation of bipolar fuzzy sets to quantum intelligence machinery. *IEEE Trans Fuzzy Syst* 26(2):656–669
21. Liu H, Cocea M (2017) Fuzzy rule based systems for interpretable sentiment analysis. In: 2017 9th international conference on advanced computational intelligence (ICACI). IEEE, pp 129–136
22. Setnes M, Babuska R, Kaymak U, van Nauta Lemke HR (1998) Similarity measures in fuzzy rule base simplification. *IEEE Trans Syst Man Cybern B (Cybern)* 28(3):376–386
23. Herrera-Semenets V, Andrés Pérez-García O, Hernández-León R, den Berg J, Doerr C (2018) A data reduction strategy and its application on scan and backscatter detection using rule-based classifiers. *Expert Syst Appl* 95:272–279
24. Yang L-H, Wang Y-M, Lan Y-X, Chen L, Yang-Geng F (2017) A data envelopment analysis (dea)-based method for rule reduction in extended belief-rule-based systems. *Knowl Based Syst* 123:174–187
25. Liu F, Quek C, Ng GS (2007) A novel generic hebbian ordering-based fuzzy rule base reduction approach to mamdani neuro-fuzzy system. *Neural Comput* 19(6):1656–1680
26. Yin Y, Yuan H, Zhang B (2017) Dynamic behavioral assessment model based on hebb learning rule. *Neural Comput Appl* 28(1):245–257
27. Kumar SS, Inbarani HH, Azar AT, Polat K (2017) Covering-based rough set classification system. *Neural Comput Appl* 28(10):2879–2888
28. Mitra S, Pal SK, Mitra P (2002) Data mining in soft computing framework: a survey. *IEEE Trans Neural Netw* 13(1):3–14
29. Liao H, Ding S, Wang M, Ma G (2016) An overview on rough neural networks. *Neural Comput Appl* 27(7):1805–1816
30. Peters G, Crespo F, Lingras P, Weber R (2013) Soft clustering-fuzzy and rough approaches and their extensions and derivatives. *Int J Approx Reason* 54(2):307–322
31. de José JR (2009) Sofmls: online self-organizing fuzzy modified least-squares network. *IEEE Trans Fuzzy Syst* 17(6):1296–1309
32. Ganivada A, Dutta S, Pal SK (2011) Fuzzy rough granular neural networks, fuzzy granules, and classification. *Theor Comput Sci* 412(42):5834–5853
33. Zanaty EA (2012) Support vector machines (SVMS) versus multilayer perception (MLP) in data classification. *Egypt Inform J* 13(3):177–183
34. Aquino G, Rubio JDJ, Pacheco J, Gutierrez GJ, Ochoa G, Balcazar R, Cruz DR, Garcia E, Novoa JF, Zacarias A (2020) Novel nonlinear hypothesis for the delta parallel robot modeling. *IEEE Access* 8:46324–46334
35. Meda-Campaña JA (2018) On the estimation and control of nonlinear systems with parametric uncertainties and noisy outputs. *IEEE Access* 6:31968–31973
36. Elias I, Rubio JJ, Cruz DR, Ochoa G, Novoa JF, Martinez DI, Muñoz S, Balcazar R, Garcia E, Juarez CF (2020) Hessian with mini-batches for electrical demand prediction. *Appl Sci* 10(6):2036
37. Mitra S, Mitra P, Pal SK (2001) Evolutionary modular design of rough knowledge-based network using fuzzy attributes. *Neurocomputing* 36(1–4):45–66
38. Pal SK, Mitra S, Mitra P (2003) Rough-fuzzy MLP: modular evolution, rule generation, and evaluation. *IEEE Trans Knowl Data Eng* 1:14–25
39. Pal SK, Meher SK (2013) Natural computing: a problem solving paradigm with granular information processing. *Appl Soft Comput* 13(9):3944–3955
40. Shen Q, Jensen R (2004) Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognit* 37(7):1351–1363
41. Saltos R, Weber R (2016) A rough-fuzzy approach for support vector clustering. *Inf Sci* 339:353–368
42. Lozowski A, Cholewo TJ, Zurada JM (1996) Crisp rule extraction from perceptron network classifiers. In: Proceedings of international conference on neural networks, pp 94–99
43. Pal SK, Skowron A (1999) Rough-fuzzy hybridization: a new trend in decision making. Springer, Berlin
44. Kai KA, Chai Q (2005) RSPOP: rough set-based pseudo outer-product fuzzy rule identification algorithm. *Neural Comput* 17(1):205–243
45. Das RT, Ang KK, Quek C (2016) IERSPOP: a novel incremental rough set-based pseudo outer-product with ensemble learning. *Appl Soft Comput* 46:170–186
46. Iyer DKP, Quek CH (2018) Pie-RSPOP: a brain-inspired pseudo-incremental ensemble rough set pseudo-outer product fuzzy neural network. *Expert Syst Appl* 95:172–189
47. Ang KK, Quek C, Pasquier M (2003) POPFNN-CRI (S): pseudo outer product based fuzzy neural network using the compositional rule of inference and singleton fuzzifier. *IEEE Trans Syst Man Cybern B (Cybern)* 33(6):838–849
48. Quek C, Zhou RW (1999) POPFNN-AAR (S): a pseudo outer-product based fuzzy neural network. *IEEE Trans Syst Man Cybern B (Cybern)* 29(6):859–870
49. Hassanien AE, Abraham A, Peters JF, Schaefer G, Henry C (2009) Rough sets and near sets in medical imaging: a review. *IEEE Trans Inf Technol Biomed* 13(6):955–968
50. Kumar DA, Meher SK, Kanhar D, Kumari KP (2016) Unified granular neural networks for pattern classification. *Neurocomputing* 216:109–125
51. Guerbai Y, Chibani Y, Hadjadj B (2015) The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. *Pattern Recognit* 48(1):103–113
52. Hssina B, Merbouha A, Ezzikouri H, Erritali M (2014) A comparative study of decision tree id3 and c4. 5. *Int J Adv Comput Sci Appl* 4(2):0–0
53. Diab DM, El Hindi KM (2017) Using differential evolution for fine tuning naïve bayesian classifiers and its application for text classification. *Appl Soft Comput* 54:183–199
54. Choubey DK, Paul S, Kumar S, Kumar S (2017) Classification of pima indian diabetes dataset using naïve bayes with genetic algorithm as an attribute selection. In: Communication and computing systems: proceedings of the international conference on communication and computing system (ICCCS 2016), pp 451–455
55. Khishe M, Mosavi MR, Kaveh M (2017) Improved migration models of biogeography-based optimization for sonar dataset classification by using neural network. *Appl Acoust* 118:15–29
56. Tsai M-H, Wang H-C, Lee G-W, Lin Y-C, Chiu S-H (2016) A decision tree based classifier to analyze human ovarian cancer cdna microarray datasets. *J Med Syst* 40(1):21