Information Sciences

A belief rule-based classification system using fuzzy unordered rule induction algorithm

--Manuscript Draft--

Manuscript Number:	INS-D-23-9027R1
Article Type:	Full length article
Keywords:	rule-based system; pattern classification; belief rule-based classification system; FURIA; evidential reasoning
Corresponding Author:	Yangxue Li University of Granada SPAIN
First Author:	Yangxue Li
Order of Authors:	Yangxue Li
	Ignacio Javier Pérez
	Francisco Javier Cabrerizo
	Harish Garg
	Juan Antonio Morente-Molinera
Abstract:	A rule-based system is a widely used artificial intelligence system that employs a set of rules to make decisions. The belief rule-based (BRB) classification system is an extension of fuzzy rule-based (FRB) system that handles uncertainty and imprecision in classification tasks by incorporating Dempster-Shafer evidence theory and fuzzy set theory. However, the BRB classification system suffers from the combinatorial explosion and high time complexity problems. To solve these issues, the fuzzy unordered rule induction algorithm (FURIA) is introduced into the BRB classification systems to design a novel BRB classification system in this paper. FURIA is computationally less complex and has superior classification performance. The proposed system outperforms BRB classification systems in terms of classification performance and significantly reduces the number of rules and conditions. Furthermore, the proposed system offers a more effective solution than FURIA. To evaluate the validity and superiority of the proposed system, we conduct two classification experiments, comparing it with five traditional classifiers and seven rule-based systems, respectively, in terms of classification performance and interpretability.
Suggested Reviewers:	Yong Deng University of Electronic Science and Technology of China dengentropy@uestc.edu.cn Prof.Deng has published many works about belief functions.
	Xiaozhuan Gao Northwest A&F University gaoxiaozhuan@nwafu.edu.cn Prof. Gao has published many works about belief functions.

UNIVERSIDAD DE GRANADA

January 24, 2024

Dear editor,

Enclosed please find the manuscript entitled as “A belief rule-based classification system using fuzzy unordered rule induction algorithm” (by Yangxue Li, Ignacio Javier Pérez, Francisco Javier Cabrerizo, Harish Garg, Juan Antonio Morente-Molinera*, INS-D-23-9027), which is considered for the possible publication in *Information Sciences*.

According to your and the reviews’ comments and suggestions, we try our best to revise the paper. The revisions are detailed in three parts named as “Reply to reviewer #1”, “Reply to reviewer #2” and “Reply to reviewer #3”, respectively. We hope the revised version can be accepted by *Information Sciences* this time.

Finally, we greatly appreciate your encouragement and the anonymous reviewers for their valuable comments and suggestions to improve this paper. If you have further question, please do not hesitate to point out and we would like to revise them according to your comments.

Sincerely yours,

The authors of the paper.

RESPONSES TO THE COMMMENTS

Reply to Handling Editor

Thank you for your valuable comments and suggestions to improve this paper. We do our best to revise the paper according to your comments and hope the revised version can be accepted. If you think there are some other issues to be addressed, we will revise again based on your suggestions and comments.

Comment 1: The reviewers appreciate the contribution presented. However, there are a numnber of issues, which need to be rectified in a thorough revision. In particular, some definitions need to be clarified, based on which the superiority of the proposed approach must be demonstrated more clearly.

Response: Thank you for your valuable comments to improve our paper. We have carefully reviewed the comments and are committed to addressing the identified issues in a comprehensive revision. Specifically, in the revised manuscript, we clarified the definitions of complexity and interpretability. We demonstrated the superiority of the method from four perspectives: accuracy, F1 score, running time, and interpretability. Specific revisions can be found in the following responses to the reviewers' comments.

Reply to Reviewer #1

Thank you for your valuable comments and suggestions to improve this paper. We do our best to revise the paper according to your comments and hope the revised version can be accepted. If you think there are some other issues to be addressed, we will revise again based on your suggestions and comments.

Comment 1: In the rule generation step, what is the significance of the order in which the antecedents of the rule are generated? The position of the antecedents seems to have no impact on the final classification result?

Response: Thank you for your valuable comments to improve our paper. Rule generation begins with an empty rule, the antecedents in the rule are incrementally expanded. At each step, the antecedent that yields the maximum information gain is chosen, and the current rule is set as the default rule for the subsequent selections. Therefore, the selection of each antecedent is intricately linked with the antecedents that came before it in the rule. We also mentioned it in the revised manuscript. The revisions are listed as follows.

Page 8, lines 45-47	In function $findrule(X_{nc}, \theta_k)$, the split point v_j and operator o_j indicates that if o_j is '$\geq$', the selector is 'F is $[v_j, +\infty)$'; if o_j is '$\leq$', the selector is 'F is $[-\infty, v_j]$'; if o_j is '=', the selector is 'F is v_j'. And if there are two selectors $(\underline{v}_j, o_j)$ and $(\overline{v}_j, o'_j)$ on a same feature F, $\underline{v}_j < \overline{v}_j$, o_j is '$\geq$' and o'_j is '$\leq$', these two selectors can be seemed as one selector 'F is $[\underline{v}_j, \overline{v}_j]$'. Rule generation begins with an empty rule, the selectors in the rule are incrementally expanded. At each step, the selector that yields the maximum information gain is chosen, and the current rule is set as the default rule for the subsequent selections.

Comment 2: The consequent part in the belief rules only considers the situation of singleton subsets. What is different from the output of traditional classification tasks, and how does it reflect the advantages of the belief degrees framework?

Response: Thank you for your valuable comments to improve our paper. The belief degrees framework in belief rule-based classification systems is primarily manifested through the utilization of ER approach to combine the consequent parts of belief rules. In the combination process, there exist not only belief degrees for singleton subsets but also belief degrees for the full set. We also mentioned it in the revised manuscript. The revisions are listed as follows.

Page 12, lines 35-37	*3.3. New pattern classification* Identifying the rules that cover a new pattern is the first step towards determining its class label. Next, the ER approach is employed to combine the BPAs of the consequent parts of these rules to obtain a BPA about the class of the new pattern. Suppose that L belief rules have been learned, $\hat{RS} = \{\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_L\}$ and $\hat{R}_l$, $l = 1, 2, \ldots, L$ is: Rule $\hat{R}_l$: If $F_{\rho_l(1)}$ is $\widetilde{A}_{\rho_l(1)}$ and $F_{\rho_l(2)}$ is $\widetilde{A}_{\rho_l(2)}$ and $\cdots$ and $F_{\rho_l(M)}$ is $\widetilde{A}_{\rho_l(M)}$ then the consequent is $C = \{(\theta_1, \beta_{l1}), \ldots, (\theta_c, \beta_{lc}), (\Theta, \beta_{l\Theta})\}$. (22) We can observe that the BPA in the consequent part of the rule $\hat{R}_l$ primarily consists of singleton sets and the full set with their belief degrees, which aligns with the context of ER approach. ER approach can address the counter-intuitive problem associated with Dempster's rule.A new pattern is denoted as $x = \{x_1, x_2, \ldots, x_d\}$. If ~~the coverage of~~ a belief

Comment 3: How to consider the information loss caused during the rule conversion process?

Response: Thank you for your valuable comments to improve our paper. The antecedents of fuzzy rules and the conversed belief rules derived from these fuzzy rules are identical; the difference lies in their consequent parts. The consequent part of a fuzzy rule is a single class label, whereas the consequent part of a belief rule is a BPA, providing more information than a single class label. Thus, fuzzy rules entail information loss, while the conversed belief rules do not suffer from information loss. We also mentioned it in the revised manuscript. The revisions are listed as follows.

Page 11, lines 51-55	Step 3: **Convert the fuzzy rules to belief rules.** We can see that both crisp and fuzzy rules have a singular consequent class. Nevertheless, in practice, the patterns covered by these rules may not all belong to the same class, and a few patterns from other classes may be included, resulting in uncertainty in the information in the consequent part. Representing this uncertainty with a singular consequent class would overlook a few patterns from other classes, resulting in information loss. Therefore, we use belief degrees framework to describe this type of uncertain information, aiming to mitigate information loss. The idea is to extend the consequent part of a fuzzy rule by computing the belief degree for

Reply to Reviewer #2

Thank you for your valuable comments and suggestions to improve this paper. We do our best to revise the paper according to your comments and hope the revised version can be accepted. If you think there are some other issues to be addressed, we will revise again based on your suggestions and comments.

Comment 1: FOIL's information gain is adopted to measure how much the rule improves upon the default rule for the target class. It would be helpful to review other measures of information gain and justify the use of FOIL's information gain.

Response: Thank you for your valuable comments to improve our paper. We added a review of a commonly used information gain measures and analyzed the advantages of FOIL's information gain. The revisions are detailed in the responses below.

Page 5, lines 3-28	A dataset may encompass a multitude of features, among which numerous may be irrelevant or redundant to the classification task. Considering all features in the rules would lead to a combinatorial explosion, thereby escalating algorithmic time costs and diminishing interpretability. Hence, it is imperative to undertake features selection during rule generation. Information gain is commonly employed as a metric for sorting and filtering features. The commonly used measures of information gain include mutual information gain and gain ratio. **Definition 7.** *The mutual information gain is a metric used to quantify the correlation between two variables, which is defined as follows.* $$IG_M(R) = I(p,n) - E(R),$$ $$E(R) = -\frac{p_R + n_R}{p+n} I(p_R, n_R), \quad (10)$$ $$I(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}.$$ *Where p_R and n_R are the number of positive and negative patterns covered by rule R while p and n are the number of positive and negative patterns covered by default rule.* **Definition 8.** *The gain ratio is an extension of information gain, designed to overcome the bias of information gain towards features with a larger number of values.* $$GR(R) = \frac{IG_M(R)}{-\sum_{i=1}^{c}\frac{|X_{\theta_i}|}{|X|}\times\log_2\frac{|X_{\theta_i}|}{|X|}}, \quad (11)$$ *where $|X|$ is the number of patterns in training data and $|X_{\theta_i}|$ is the number of patterns in training data belong to class θ_i.*
Page 5, lines 43-46	The objective of FOIL is to generate rules with high generality, meaning rules that can generalize to new instances. Therefore, FOIL's information gain is utilized to assess the generality of rules rather than evaluating the classification ability on a specific dataset. The uniqueness of FOIL lies in its approach of guiding the learning process by obtaining information gain from positive examples not mentioned in the rules.

Comment 2: Is there any information loss from converting the fuzzy rules to belief rules.

Response: Thank you for your valuable comments to improve our paper. The antecedents of fuzzy rules and the conversed belief rules derived from these fuzzy rules are identical; the difference lies in their consequent parts. The consequent part of a fuzzy rule is a single class label, whereas the consequent part of a belief rule is a BPA, providing more information than a single class label. Thus, fuzzy rules entail information loss, while the conversed belief rules do not suffer from information loss. We also mentioned it in the revised manuscript. The revisions are listed as follows.

Page 11, lines 51-55	Step 3: **Convert the fuzzy rules to belief rules.** We can see that both crisp and fuzzy rules have a singular consequent class. Nevertheless, in practice, the patterns covered by these rules may not all belong to the same class, and a few patterns from other classes may be included, resulting in uncertainty in the information in the consequent part. Representing this uncertainty with a singular consequent class would overlook a few patterns from other classes, resulting in information loss. Therefore, we use belief degrees framework to describe this type of uncertain information, aiming to mitigate information loss. The idea is to extend the consequent part of a fuzzy rule by computing the belief degree for

Comment 3: It looks from Table 3 that the proposed method outperforms SVM on many datasets, while SVM has superior performance on some other datasets, e.g., Balance. It would be interesting to discuss the potential reasons.

Response: Thank you for your valuable comments to improve our paper. The performance of the

method is influenced by domain knowledge and problem characteristics. The proposed method may be overly flexible in certain situations, leading to overfitting. Furthermore, to compare the generalization ability of the method, we did not perform parameter tuning on different datasets. The parameters currently employed may not be optimal for some datasets. We also mentioned it in the revised manuscript. The revisions are listed as follows.

Page 15, lines 56-57	also obtains the best average accuracy and average F1 score over all datasets. The proposed method demonstrates good classification performance on the majority of datasets; however, it may exhibit excessive flexibility leading to overfitting on certain datasets, such as Balance. To further detect significant differences between the proposed method

Comment 4: In addition, some datasets, e.g., Optdigits involve a large number of features, and will the use of rule-based models still lead to combinatorial explosion of rules and incur overfitting issue.

Response: Thank you for your valuable comments to improve our paper. We conducted feature selection during rule generation, where the antecedents in the rules are incrementally expanded from an empty rule, selecting at each step the antecedent that provides the maximum information gain. Therefore, even in datasets with a large number of features, this approach mitigates the risk of combinatorial explosion and overfitting. We also listed the number of rules generated for each dataset, the number of conditions, and the number of fired rules. From the results, it can be observed that, even for datasets with a large number of features, the number of rules remains at a relatively low level. The revisions are listed as follows.

Page 8, lines 45-47	seemed as one selector 'F is $[\underline{v}_j, \overline{v}_j]$'. Rule generation begins with an empty rule, the selectors in the rule are incrementally expanded. At each step, the selector that yields the maximum information gain is chosen, and the current rule is set as the default rule for the subsequent selections.
Page 20 Table 10	Table 10: Interpretability measures for the eight rule-based systems.

	Numbers of rules								Numbers of conditions								Number of fired rules							
Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	95.4	3.3	28.3	4.7	2.6	180	3.4	4.3	7	1.4	7	4.5	6	2	7	1.5	95.4	1.3	1	1	1	77.2	1.2	1.5
Balance	562.5	24.3	66.2	3	4.2	180	4	23.1	4	3.1	4	0	3.2	2	4	3.1	562.5	2.6	1	1	1	51.3	2.1	2.5
Banknote	1234.8	12.5	30.8	6.9	3.4	180	3.9	12.5	4	2.3	4	1.2	3.4	2	4	2.4	1234.8	2.0	1	1	1	79.9	2.2	2.0
Bupa	310.5	9.8	43.2	9	3.3	180	3	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.6	1.2
Car	1554.3	83.5	690.5	27.5	1.6	180	3.3	78.7	6	4.3	6	3.6	6	2	6	4.3	1554.3	1.3	1	1	1	40.3	1.5	1.3
Column3C	279	7.8	26.8	10.8	3.5	180	3.1	8.1	6	2.5	6	1.9	5.3	2	6	2.6	279	1.4	1	1	1	79.5	1.9	1.4
Ecoli	302.4	18	63	29.7	8.9	180	3.9	17	7	2.9	7	2.0	1.3	2	7	2.9	302.4	1.9	1	1	1	58.0	1.6	1.8
German	900	9.4	885.4	22.5	2.4	180	3	10.1	20	2.3	20	3.1	19.8	2	20	2.3	900	1.4	1	1	1	41.7	1.2	1.4
Haberman	275.4	4	16.5	3.7	2.9	180	5.3	3.7	3	1.7	3	0.9	1.4	2	3	1.3	275.4	1.4	1	1	1	63.3	1.8	1.3
Hayes-roth	144	9.1	39.1	4.5	5.3	180	4.9	9.3	4	2.9	4	1.8	0.8	2	4	3.0	144	0.9	1	1	1	43.0	2.3	0.9
Heart-statlog	243	9.1	217.5	11.8	2.1	180	3	8.2	13	2.8	13	2.8	12	2	13	2.6	243	1.7	1	1	1	43.5	1.6	1.7
Housevotes	208.8	5.4	146.5	3	2	180	3	5.5	16	2.2	16	3	16	2	16	2.1	208.8	1.5	1	1	1	20	1.5	1.5
HTRU_2	16108.2	12.9	89.5	11.5	5.9	180	3.5	14.2	8	2.8	8	3.4	5.9	2	8	2.9	16108.2	3.2	1	1	1	80.0	1.1	3.7
Ilpd	524.7	9.4	86	10.9	2.1	180	3	9.6	10	2.6	10	4.0	9.6	2	10	2.6	524.7	1.5	1	1	1	71.2	1.1	1.7
Iris	135	4.3	14.7	5.3	3.9	180	5.3	4.7	4	1.8	4	1.6	2.9	2	4	1.9	135	1.2	1	1	1	78.0	2.4	1.3
Liver-disorders	310.5	9.8	43.2	9	3.3	180	4.1	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.8	1.2
Magic	17118	29.8	312.8	12.1	2.8	180	3.2	26.9	10	3.2	10	2.9	9.2	2	10	3.2	17118	3.2	1	1	1	80.0	1.8	2.7
Mfeat-zernike	1800	10.8	1679.9	7.4	2.2	180	3	10.9	47	2.6	47	4.3	45.8	2	47	2.5	1800	3.4	1	1	1	79.9	1.0	3.4
New-thyroid	193.5	6.6	18.4	12.2	3.3	180	4.7	7.1	5	2.5	5	2.4	3.3	2	5	2.4	193.5	1.9	1	1	1	79.0	1.8	2.0
Nursery	11664	274.1	4222.6	18.4	5	180	3	276.7	8	5.0	8	4.0	5	2	8	5.0	11664	1.5	1	1	1	36.8	1.8	1.5
Occupancy	2398.5	12.7	25	3.8	2.8	180	4.1	13.2	5	2.7	5	2.9	3.2	2	5	2.7	2398.5	2.0	1	1	1	70.8	1.6	1.8
Optdigits	5058	35.1	5055.6	15	2	180	3	34.7	64	4.0	64	4.9	64	2	64	4.0	5058	4.6	1	1	1	45.2	1.0	4.6
Pima	691.2	7.8	101.3	8.6	4.1	180	3.6	8.3	8	2.4	8	2.6	6.1	2	8	2.4	691.2	1.8	1	1	1	69.6	1.8	1.8
QSAR	949.5	20.1	436.4	15.4	2.8	180	3	22.1	41	3.4	41	4.6	39.2	2	41	3.4	949.5	2.1	1	1	1	48.0	1.2	2.2
Rice-OC	3429	5.8	95.8	7.8	3.3	180	3.6	4.4	7	1.9	7	3.6	5.7	2	7	1.8	3429	2.5	1	1	1	80.0	1.7	1.9
Seed	189	8.5	53.2	9.6	3.8	180	3.2	8.4	7	2.2	7	3.0	5	2	7	2.2	189	1.8	1	1	1	79.3	2.1	1.7
Sonar	186.3	11.2	186.3	7.6	2.8	180	3	11	60	2.6	60	4.4	60	2	60	2.6	186.3	1.7	1	1	1	78.6	1.0	1.7
Spambase	4140.9	33.3	356.2	7.8	2	180	3	34.9	57	4.2	57	5	56	2	57	4.3	4140.9	3.5	1	1	1	29.5	1.1	3.6
Tic-tac-toe	862.2	21.6	862.2	9.3	2	180	3	21	9	3.5	9	2.0	8	2	9	3.5	862.2	1.1	1	1	1	35.4	1.2	1.1
Titanic	1980.9	5.8	9.9	7.2	3	180	4.2	5.7	3	2.2	3	1.1	3	2	3	2.2	1980.9	2.1	1	1	1	26.1	2.1	2.1
Transfusion	673.2	5	12.2	4	2.7	180	4.7	5	4	1.8	4	1.3	3.7	2	4	1.8	673.2	1.5	1	1	1	71.2	1.8	1.6
Vehicle	761.4	23.6	227.6	22.6	6.2	180	3	24.1	18	3.3	18	3.3	15.4	2	18	3.4	761.4	1.3	1	1	1	79.0	1.7	1.2
Mean	2352.6	23.3	504.5	10.7	3.4	180.0	3.6	23.2	14.9	2.8	14.9	2.9	13.7	2.0	14.9	2.7	2352.6	1.9	1.0	1.0	1.0	61.7	1.6	1.8

Reply to Reviewer #3

Thank you for your valuable comments and suggestions to improve this paper. We do our best to revise the paper according to your comments and hope the revised version can be accepted. If you think there are some other issues to be addressed, we will revise again based on your suggestions and comments.

Comment 1: However, the key concepts underlying these statements, such as "interpretability" and "complexity", were only vaguely defined.

Response: Thank you for your valuable comments to improve our paper. The complexity here mainly refers to time complexity. Interpretability is primarily composed of three criteria: the number of rules, the number of conditions, and the number of fired rules. To avoid confusion, we clarified their definitions. The revisions are listed as follows.

Page 1, line 21	theory. However, the BRB classification system suffers from the combinatorial explosion and high time complexity
Page 2, line 13	such as increased time complexity and unstable classification performance. EBRB classification system is a solution
Page 2, line 18	To solve the problems of high time complexity and combinatorial explosion of traditional BRB classification
Page 2, line 45	(1) To overcome the problems of BRB classification system, FURIA is introduced into BRB classification system, which can improve the classification performance and reduce the timecomplexity of the BRB system.
Page 13, lines 45-52	*3.4. Timecomplexity analysis* The timecomplexity of the proposed BRB classification system can be analyzed by considering the timecomplexity of modified RIPPER algorithm, fuzzification and belief degree assignment. For training dataset X with d features, the timecomplexity of modified RIPPER algorithm is $O(\|X\|(\log\|X\|)^2)$, with $\|X\|$ the number of patterns in X. The timecomplexity of fuzzification procedure is $O(\|X\|d^2)$, with d the number of features [19]. The timecomplexity of belief degree assessment is $O(\|X\|)$. Therefore, the overall timecomplexity is $O(\max(\|X\|(\log\|X\|)^2, \|X\|d^2, \|X\|)) = O((\|X\|\log\|X\|)^2)$.
Page 18, lines 52-58 Page 19, lines 43-49	~~In addition to classification performance , our method significantly reduces the numbers of rules and conditions compared to EBRB and FRB systems. The average~~ Besides accuracy, F1 score and running time, interpretability is also a crucial criterion for rule-based systems. Ref. [18] provides a detailed survey of interpretability measures for rule-based systems. In this experiment, we adopt three commonly used interpretability measures suggested in Ref. [18] to perform a quantitative analysis. • Number of rules: The number of rules is an important interpretability measure for rule-based systems. This is based on the principle of Occam's razor, which states that the best model is the simplest one that fits the system behavior well. In other words, the set of fuzzy rules should be as small as possible while still preserving the model's performance to a satisfactory level. • Number of conditions: The number of conditions should be minimized to enhance the readability of the rules. • Number of fired rules: To ensure semantic interpretability, it is desirable to minimize the number of rules used in the reasoning process for a given input.

Comment 2: In the analysis of classification performance, the measure of accuracy has been used, whereas, when it comes to classification, other more informative measures are expected, such as precision, recall, F-measure.

Response: Thank you for your valuable comments to improve our paper. To comprehensively compare the classification performance of methods, we added a comparison of the F1 scores for each method across datasets. The revisions are listed as follows.

Page 15, Section 4.2

4.2. Comparison with traditional classifiers

In this section, we select five traditional classifiers, including C45 decision tree (C45) [28], support vector machine (SVM) [11], backpropagation (BP) neural network [29], hybrid decision tree/genetic algorithm (DTGA)[5], and k-nearest neighbors (KNN) [12] classifiers, to compare their classification performance with the proposed method. Table 2 shows the key parameter settings for each of these classifiers. Although tuning the parameters of each method for each specific problem could potentially lead to better results, we prefer to use the default parameter values for each method. This is because a classifier's generalization ability is also an important factor to consider. A classifier with good generalization capability should be able to perform well on different datasets. All the algorithms are implemented on the KEEL software platform [1].

Table 3 presents the classification accuracy and F1 score rates of six classifiers on each dataset. The best ~~classification accuracy~~ result for each dataset is highlighted in boldface, and the last row displays the average ~~accuracy~~ values of all datasets for each classifier. It can be observed that the proposed BRB classification system achieves the best ~~results on 17~~ accuracy on 16 out of 32 datasets and the best F1 score on 12 out of 32 datasets, respectively. It also obtains the best average accuracy and average F1 score over all datasets. The proposed method demonstrates

Table 3: Classification accuracy and F1 score rates (%) of six classifiers on 32 datasets.

	Accuracy						F1 Score					
Dataset	C45	SVM	BP	DTGA	KNN	Our method	C45	SVM	BP	DTGA	KNN	Our method
Appendicitis	84.09	**86.91**	73.73	84.09	81.91	84.82	65.48	70.03	69.11	65.48	**71.74**	67.80
Balance	78.07	**95.68**	68.32	75.66	78.24	83.05	55.19	**90.98**	47.75	53.32	56.74	58.95
Banknote	98.40	**100.00**	93.66	98.10	99.85	99.13	98.37	**100.00**	93.61	98.07	99.85	99.11
Bupa	66.97	63.50	54.82	64.32	63.14	**71.31**	64.81	60.56	54.35	62.88	61.94	**69.62**
Car	92.88	95.02	50.26	79.79	86.39	**95.48**	84.35	**91.46**	31.75	41.12	63.99	83.95
Column3C	**81.61**	76.13	68.39	79.03	76.77	**81.61**	74.01	69.70	59.34	71.04	72.70	**75.46**
Ecoli	79.47	61.96	42.70	77.09	80.70	**80.99**	68.05	37.23	21.80	62.46	65.65	**70.91**
German	70.70	65.80	61.90	69.60	69.30	**73.40**	61.97	46.64	54.26	57.84	62.36	**63.69**
Haberman	70.91	69.94	70.24	70.91	67.63	**72.55**	57.96	54.53	**61.79**	57.96	57.40	61.14
Hayes-roth	80.00	78.75	45.63	74.38	37.50	**80.63**	**83.30**	80.20	37.91	76.22	42.77	81.98
Heart-statlog	75.19	59.63	80.37	77.41	75.56	**81.11**	74.62	55.30	79.92	76.46	74.96	**80.59**
Housevotes	96.11	96.11	90.02	96.11	91.32	**96.56**	96.08	96.08	90.00	96.08	91.25	**96.54**
HTRU_2	**97.94**	96.64	86.81	97.90	97.14	97.88	**93.56**	89.41	74.17	93.39	91.37	93.37
Ilpd	67.07	**73.07**	62.77	68.59	63.45	69.97	56.51	**61.38**	54.69	58.77	56.94	51.35
Iris	96.00	**97.33**	61.33	96.00	93.33	94.00	96.01	**97.31**	57.40	96.01	93.31	93.95
Liver-disorders	66.97	63.50	54.82	64.32	63.14	**71.31**	64.81	60.56	54.35	62.88	61.94	**69.62**
Magic	**85.08**	68.03	70.38	84.98	80.78	84.72	**82.97**	57.70	68.83	82.61	78.40	82.38
Mfeat-zernike	98.35	90.00	49.90	98.35	**99.40**	98.95	95.43	47.37	43.39	95.45	**98.33**	97.00
New-thyroid	93.98	93.96	81.34	91.65	**97.21**	94.46	91.30	91.23	70.17	88.24	**96.20**	92.23
Nursery	96.98	93.20	28.21	94.07	82.08	**98.47**	**87.64**	84.80	21.17	83.55	73.27	76.45
Occupancy	98.24	97.90	97.94	98.31	**98.72**	98.54	98.11	97.73	97.80	98.19	**98.62**	98.43
Optdigits	97.74	93.29	38.36	97.72	89.82	**98.20**	93.67	73.43	33.49	93.69	47.32	**95.22**
Pima	**75.40**	63.80	70.33	74.61	69.78	75.13	**72.78**	50.48	68.84	72.12	66.28	71.06
QSAR	83.51	**86.26**	69.66	82.00	83.12	85.50	81.40	**84.73**	66.44	80.22	81.47	83.26
Rice-OC	92.36	68.82	83.25	92.36	88.79	**92.73**	92.20	62.58	83.14	92.20	88.56	**92.57**
Seed	92.38	91.43	80.95	89.52	**94.29**	92.86	92.36	91.39	79.31	89.47	**94.21**	92.82
Sonar	73.86	80.64	64.24	75.31	**86.93**	81.17	73.40	80.08	63.93	75.21	**86.66**	81.10
Spambase	92.57	86.57	74.48	91.74	90.11	**93.65**	92.21	86.09	62.86	91.42	89.64	**93.33**
Tic-tac-toe	84.55	81.63	60.23	77.98	73.07	**99.17**	82.13	77.42	58.96	71.34	59.46	**99.06**
Titanic	**79.06**	78.24	76.06	77.83	60.16	78.51	70.06	**71.99**	70.11	70.85	55.83	69.05
Transfusion	78.61	71.25	69.78	78.47	61.64	**79.41**	66.95	55.13	63.91	66.27	53.87	**67.82**
Vehicle	**75.06**	52.14	38.31	68.80	69.62	71.16	**74.69**	55.12	32.12	68.16	69.32	69.11
Mean	84.38	80.53	66.22	82.72	79.72	**86.14**	79.45	72.77	60.21	76.53	73.82	**80.59**

Table 4: Wilcoxon signed-rank tests of the accuracy and F1 score for comparing our method to five traditional classifiers.

	Classifier	Statistic value	p-value	Critical value	Hypothesis
Accuracy	C45	63	0.00028	159	Rejected
	SVM	82	0.00036	159	Rejected
	BP	0	4.65661e-10	159	Rejected
	DTGA	20	1.72760e-07	159	Rejected
	KNN	48	1.11940e-05	159	Rejected
F1 score	C45	159	0.04976	159	Rejected
	SVM	123	0.00733	159	Rejected
	BP	18	1.17812e-07	159	Rejected
	DTGA	94	0.00098	159	Rejected
	KNN	95	0.00105	159	Rejected

good classification performance on the majority of datasets; however, it may exhibit excessive flexibility leading to overfitting on certain datasets, such as Balance. To further detect significant differences between the proposed method and other classifiers, we conducted the Wilcoxon signed-rank test on the performance of the classifiers on 32 datasets. This nonparametric statistical test was utilized to determine whether the proposed method outperformed other classifiers significantly on the given datasets. The test results are presented in Table 4. The test aims to determine whether there is a significant difference in performance between two classifiers being compared. The null hypothesis is that there is no significant difference between the two classifiers. The test statistic is the sum of the ranks of the positive or negative differences, depending on which is smaller. The p-value is then calculated by comparing the test statistic to a critical value from a reference table. If the p-value is below the significance level (0.05 here), it implies rejecting the null hypothesis and indicates a statistically significant distinction in performance between the two models.

As shown in Table 4, all null hypotheses are rejected, demonstrating that the proposed method statistically outperformed five traditional classifiers: C45, SVM, BP neural network, DTGA, and KNN, on all 32 datasets.

Pages 17-18, Section 4.3

4.3. Comparison with rule-based classifiers

In this section, we compare the proposed BRB classification system with seven other rule-based classification systems: the EBRB classification system [23], FURIA [19], FRB classification system [10], fuzzy rule-based decision tree (FRDT) [31], steady-state genetic algorithm for extracting fuzzy classification rules from data (SGERD) [25], tree (FRDT) [31], steady-state genetic algorithm for extracting fuzzy classification rules from data (SGERD) [25], ensemble belief rule-based model (En-BRB) [38], and parallel multipopulation optimization for BRB (PMP-BRB) [41]. We evaluate these systems based on classification accuracy and running time. Table 6 presents some key parameter settings for each rule-based classification system. All the algorithms are implemented on the KEEL software platform [1].

Table 7 presents the classification accuracy and F1 score rates obtained by each classifier for each dataset, with the best ~~accuracy~~ results highlighted in boldface. The last row of the table lists the average ~~accuracy rate~~ results across all datasets for each classifier. We can see that the proposed method achieved the best ~~results~~ accuracy on 20 out of 32 datasets and ~~had the best~~ the best F1 score on 16 out of 32 datasets. It also has the best average classification accuracy and the best average F1 score across all 32 datasets, indicating superior performance compared to other rule-based methods for most of the datasets. To statistically compare the results, we used the Wilcoxon signed-rank test for multiple comparisons between the proposed method and other methods. Table 8 presents the test results of the comparison between the proposed method and each rule-based system. As shown in the table, all null hypotheses were rejected, indicating that the proposed BRB classification system outperformed the EBRB classification system, FURIA, FRB classification system, FRDT, SGERD, En-BRB, and PMP-BRB on these 32 datasets.

Table 9 displays the average running time, including the training and testing time, for each classifier shown in Table 4 over 10CV. Our method demonstrates good performance in terms of running time, with an average running time of ~~1881.3~~ 5440.0 milliseconds across the 32 datasets. Among these classifiers, the proposed method stands at a moderate performance level, significantly outperforming classifiers with longer running time (such as EBRB, ~~DRB,~~ En-BRB and PMP-BRB), while also exhibiting comparability with classifiers that have shorter running time (such as FURIA).

Table 7: Classification accuracy and F1 score rates (%) of eight rule-based classification systems on 32 datasets.

	Accuracy								F1 Score							
Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	84.91	83.09	85.82	75.64	86.91	86.91	**87.73**	84.82	71.45	65.37	71.80	68.30	70.35	**73.27**	70.91	67.80
Balance	33.11	82.08	**90.39**	46.08	75.66	85.92	78.41	83.05	27.30	58.51	49.22	21.03	46.32	**59.64**	54.46	58.95
Banknote	98.62	98.91	94.97	87.90	83.75	93.44	**99.13**	**99.13**	98.60	98.89	94.93	87.81	80.56	93.34	99.11	**99.11**
Bupa	61.18	65.81	58.54	63.50	58.84	63.19	62.39	**71.31**	50.95	63.48	37.64	63.31	44.47	51.28	58.07	**69.62**
Car	65.90	94.27	78.00	72.04	58.01	70.01	80.89	**95.48**	45.47	82.68	54.81	40.75	16.14	20.59	39.60	**83.95**
Column3C	68.71	**82.58**	59.03	81.29	68.39	77.10	77.10	81.61	58.08	75.15	44.99	**75.72**	61.90	65.06	66.44	75.46
Ecoli	69.33	79.77	72.02	61.30	74.69	63.73	68.78	**80.99**	55.76	69.22	60.10	44.38	61.67	34.71	38.37	**70.91**
German	64.90	73.20	18.80	70.50	65.30	70.00	72.40	**73.40**	58.50	62.06	16.54	**67.54**	28.90	41.18	58.39	63.69
Haberman	71.90	72.23	73.87	62.56	71.92	72.87	**74.19**	72.55	45.62	58.74	44.26	53.52	46.30	46.82	58.26	**61.14**
Hayes-roth	63.13	**80.63**	58.75	49.38	45.00	59.38	68.75	**80.63**	61.19	**82.05**	41.77	43.15	30.61	61.13	59.62	81.98
Heart-statlog	72.22	80.37	52.59	**83.70**	75.19	80.37	77.78	81.11	71.23	79.61	42.01	**83.59**	60.57	79.04	76.14	80.59
Housevotes	79.69	96.54	41.36	**96.97**	87.88	89.20	95.24	96.56	79.35	96.53	37.91	**96.97**	61.34	89.07	95.20	96.54
HTRU_2	96.26	97.85	95.68	97.50	95.75	97.26	97.68	**97.88**	86.37	93.21	83.65	92.36	81.25	90.77	92.63	**93.37**
Ilpd	68.27	68.96	70.32	55.39	71.01	**71.19**	69.29	69.97	45.88	52.24	38.56	**55.25**	40.18	42.15	44.48	51.35
Iris	95.33	95.33	92.67	94.67	95.33	**97.33**	94.00	94.00	95.30	95.30	92.47	94.63	95.30	**97.32**	93.97	93.95
Liver-disorders	61.18	65.81	58.54	63.50	58.84	64.95	64.24	**71.31**	50.95	63.48	37.64	63.31	44.47	53.93	59.03	**69.62**
Magic	76.17	84.68	76.78	74.62	71.30	73.73	81.88	**84.72**	69.16	82.15	69.68	72.21	59.76	62.63	76.32	**82.38**
Mfeat-zernike	**99.15**	98.80	98.80	96.00	97.55	90.50	89.05	98.95	**97.61**	96.48	90.36	90.55	92.46	51.93	32.93	97.00
New-thyroid	82.36	94.46	84.24	93.53	87.03	**94.94**	91.65	94.46	63.74	91.94	66.24	92.18	73.87	91.58	85.82	**92.23**
Nursery	75.22	97.05	84.49	53.24	64.40	73.75	69.99	**98.47**	63.71	**88.13**	62.51	36.59	41.01	53.83	49.94	76.45
Occupancy	93.85	98.27	93.09	97.86	96.96	93.17	97.90	**98.54**	93.32	98.15	92.54	97.72	87.15	92.56	97.75	**98.43**
Optdigits	**98.29**	98.75	45.41	56.32	89.79	89.82	90.20	98.20	73.52	**96.48**	39.02	49.74	47.31	47.32	42.39	95.22
Pima	72.13	74.86	73.17	73.05	70.70	69.66	73.69	**75.13**	63.75	**71.16**	60.01	70.83	55.58	53.64	68.92	71.06
QSAR	83.80	84.27	78.39	53.09	68.34	70.51	75.36	**85.50**	82.24	81.49	52.08	51.93	40.38	52.54	60.42	**83.26**
Rice-OC	91.97	92.57	88.19	92.62	92.20	91.73	92.65	**92.73**	91.70	92.39	87.58	92.46	86.00	91.48	92.48	**92.57**
Seed	90.95	92.38	91.90	90.95	86.19	**93.33**	90.48	92.86	90.88	92.34	91.86	90.75	86.09	**93.27**	90.19	92.82
Sonar	73.52	79.71	61.48	74.90	71.52	74.43	39.19	**81.17**	53.94	79.52	47.43	74.69	69.29	73.06	25.17	**81.10**
Spambase	73.68	93.24	71.81	85.26	69.42	67.64	62.20	**93.65**	68.64	92.86	43.16	83.85	41.06	54.03	35.46	**93.33**
Tic-tac-toe	98.33	98.54	0.00	70.78	69.94	65.87	73.18	**99.17**	98.13	98.37	0.00	69.29	67.00	41.33	63.07	**99.06**
Titanic	77.56	78.33	78.28	75.37	71.97	77.92	78.46	**78.51**	66.97	68.66	70.04	**71.24**	43.52	69.66	69.72	69.05
Transfusion	75.94	78.74	76.61	71.52	76.34	77.27	78.48	**79.41**	44.14	**67.88**	45.42	64.14	46.23	48.95	64.71	67.82
Vehicle	68.45	70.58	61.59	48.70	53.44	57.33	51.79	**71.16**	66.74	68.28	58.48	44.21	50.87	52.85	45.49	**69.11**
Mean	77.69	85.40	70.80	74.05	75.30	78.26	78.25	**86.14**	68.44	80.09	57.02	68.88	58.06	63.44	64.55	**80.59**

Table 8: Wilcoxon signed-rank tests of the accuracy and F1 score rates for comparing our method to seven rule-based classification systems.

	Classifier	Statistic value	p-value	Critical value	Hypothesis
Accuracy	EBRB	15	6.37955e-08	159	Rejected
	FURIA	58.5	0.00034	159	Rejected
	FRB	29	8.09784e-07	159	Rejected
	FRDT	19	1.42958e-07	159	Rejected
	SGERD	14	5.12227e-08	159	Rejected
	En-BRB	46	8.74791e-06	159	Rejected
	PMP-BRB	31	1.10594e-06	159	Rejected
F1 Score	EBRB	18	1.17812e-07	159	Rejected
	FURIA	104	0.00207	159	Rejected
	FRB	6	6.51925e-09	159	Rejected
	FRDT	74	0.00017	159	Rejected
	SGERD	3	2.32830e-09	159	Rejected
	En-BRB	24	3.54833e-07	159	Rejected
	PMP-BRB	17	9.63918e-08	159	Rejected

Comment 3: It is not always clear what is meant by "complexity". Sometimes it is mentioned specifically computational complexity, and at other points, it seems that the term refers to the rule base complexity, measured by criteria such as number of rules and conditions. When it comes to computational complexity, the analysis of run time shows that the proposed method is better than SVM only.

Response: Thank you for your valuable comments to improve our paper. The complexity here mainly refers to time complexity. Interpretability is primarily composed of three criteria: the number of rules, the number of conditions, and the number of fired rules. To avoid confusion, we clarified their definitions. When comparing the running time with traditional classifiers, the proposed method only outperforms SVM. However, in comparison with rule-based classifiers, the proposed method exhibits significantly lower runtime compared to EBRB, Ensemble BRB, and PMP-BRB classification systems. The revisions are listed as follows.

Page 1, line 21	theory. However, the BRB classification system suffers from the combinatorial explosion and high time complexity
Page 2, line 13	such as increased time complexity and unstable classification performance. EBRB classification system is a solution
Page 2, line 18	To solve the problems of high time complexity and combinatorial explosion of traditional BRB classification

Page 2, line 45

(1) To overcome the problems of BRB classification system, FURIA is introduced into BRB classification system, which can improve the classification performance and reduce the timecomplexity of the BRB system.

Page 13, lines 45-52

3.4. Timecomplexity analysis

The timecomplexity of the proposed BRB classification system can be analyzed by considering the timecomplexity of modified RIPPER algorithm, fuzzification and belief degree assignment. For training dataset X with d features, the timecomplexity of modified RIPPER algorithm is $O(|X|(\log|X|)^2)$, with $|X|$ the number of patterns in X. The timecomplexity of fuzzification procedure is $O(|X|d^2)$, with d the number of features [19]. The timecomplexity of belief degree assessment is $O(|X|)$. Therefore, the overall timecomplexity is $O(\max(|X|(\log|X|)^2, |X|d^2, |X|)) = O((|X|\log|X|)^2)$.

Page 18, lines 52-58 Page 19, lines 43-49

~~In addition to classification performance , our method significantly reduces the numbers of rules and conditions compared to EBRB and FRB systems. The average~~ Besides accuracy, F1 score and running time, interpretability is also a crucial criterion for rule-based systems. Ref. [18] provides a detailed survey of interpretability measures for rule-based systems. In this experiment, we adopt three commonly used interpretability measures suggested in Ref. [18] to perform a quantitative analysis.

- Number of rules: The number of rules is an important interpretability measure for rule-based systems. This is based on the principle of Occam's razor, which states that the best model is the simplest one that fits the system behavior well. In other words, the set of fuzzy rules should be as small as possible while still preserving the model's performance to a satisfactory level.
- Number of conditions: The number of conditions should be minimized to enhance the readability of the rules.
- Number of fired rules: To ensure semantic interpretability, it is desirable to minimize the number of rules used in the reasoning process for a given input.

Page 16, Table 5 Page 17, lines 15-16

Table 5 lists the average running time, including the training time and testing time, of each classifier shown in Table 2 over 10CV. It is evident that our method runs faster than SVM.

Table 5: Running time (milliseconds) of six classifiers on 32 datasets.

Dataset	C45	SVM	BP	DTGA	KNN	Our method
Appendicitis	45.2	30.1	193.3	58.2	37.8	94.9
Balance	74.9	73	202.3	57.7	53.2	303.3
Banknote	97.9	66.6	213.1	549.5	83.7	336.7
Bupa	60	68.1	223.3	505.1	48.4	217.8
Car	76.4	281.8	318.9	219.6	98	1082
Column3C	59.6	65.2	216	212	50.1	257.8
Ecoli	67.6	90.3	228.2	198.9	52.8	295.2
German	118.2	251.7	329.9	211.3	97.2	575.5
Haberman	45.1	70.8	177.4	1335.9	42.6	181.5
Hayes-roth	45.1	39.6	188.8	1239	40	140.6
Heart-statlog	65.1	61	227.8	218.5	50.2	240.9
Housevotes	48.5	40.9	235.5	259.2	47	131.3
HTRU_2	4507.2	28254.4	620.5	506.6	2683.7	22588.7
Ilpd	129.4	133	234.3	496.6	87.5	367.5
Iris	56.9	34	210.6	427.2	54.5	113.5
Liver-disorders	82.5	74	217	431.2	69.8	227.2
Magic	9073.8	235534.2	610.2	67.3	2791.7	47012.7
Mfeat-zernike	679.8	1745.4	406.9	61.6	327.7	1372.1
New-thyroid	49.8	48.7	210.5	108.3	46.1	116
Nursery	365.6	14823.4	552.2	129.3	1368.5	80088
Occupancy	134.5	543.1	265.3	196.9	185.5	597.1
Optdigits	4158.3	33767.9	580.2	207.4	1362.2	4708.8
Pima	91.2	223.3	251.4	66.5	84.1	499
QSAR	358.8	314.1	312.2	66.5	174.3	1379.2
Rice-OC	321.4	6242.6	322.2	7362.2	306.2	1915
Seed	55.5	38.4	229.6	8839.7	60.6	155.8
Sonar	135	68.2	328.4	292.7	99.9	422.5
Spambase	5852.1	7168	515.6	265.9	984.5	7084.1
Tic-tac-toe	72.6	171	278.8	60.4	103	254
Titanic	123.9	438.6	245.6	88.8	143.3	284.9
Transfusion	63.1	118.8	225.7	202.7	76.4	237.7
Vehicle	170.3	250.9	316.4	224	108.9	798.4
Mean	874.7	10673.6	302.3	804.6	377.8	5589.7

Page 18, lines 47-51 Page 19, Table 9

Table 9 displays the average running time, including the training and testing time, for each classifier shown in Table 4 over 10CV. Our method demonstrates good performance in terms of running time, with an average running time of ~~1881.3~~ 5440.0 milliseconds across the 32 datasets. Among these classifiers, the proposed method stands at a moderate performance level, significantly outperforming classifiers with longer running time (such as EBRB, ~~DRB,~~ Ensemble-BRB and PMP-BRB), while also exhibiting comparability with classifiers that have shorter running time (such as FURIA).

Table 9: Running time (milliseconds) of eight classifiers on 32 datasets.

Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	117.7	60.6	84.3	103.1	69.3	43401.6	13515.3	94.9
Balance	353.5	245.2	140.2	116.8	100.8	39506.5	12414.8	303.3
Banknote	1097.7	262	163.1	157.3	163.1	63239.8	20188.2	336.7
Bupa	333.4	158	108.8	120.6	92.4	67455.5	13121	217.8
Car	1584.4	889.5	637	266.9	205.6	39802.2	44905.2	1082
Column3C	380.2	154.6	138.6	149.6	125.3	59696.8	13264.6	257.8
Ecoli	352.9	155.7	121.9	158.5	97.4	54334.7	22522.4	295.2
German	2636.8	334.9	1652.2	239.4	246.7	47893.9	114454.6	575.5
Haberman	203.6	93.7	108.3	114.7	90.5	44382.1	7002.1	181.5
Hayes-roth	132.6	71.6	94.5	103.2	70.8	34095.5	6939	140.6
Heart-statlog	469.6	136.9	311.7	153.7	120.3	29973.9	40976.8	240.9
Housevotes	428.2	67.3	291.1	125.8	120.1	11779	51138.8	131.3
HTRU_2	137517.5	14844.3	7813.1	707.3	4142.6	66482.8	228874.5	22588.7
Ilpd	1215.4	278.1	212.5	117.6	126.1	37310.2	30279.2	367.5
Iris	242.1	68.3	126.4	84	84.1	50877.8	6989.5	113.5
Liver-disorders	548.4	165.8	148.6	88	95.3	58893.7	13936.1	227.2
Magic	181762	37410.5	15065.6	675.9	3857.6	66989.6	303463.3	47012.7
Mfeat-zernike	26758.3	1087.9	8351.4	329.2	661.5	78510.4	512651.9	1372.1
New-thyroid	250.5	80.5	82.6	86.1	89.9	73102.5	9851.3	116
Nursery	55048.3	68715	15440.8	1352.9	3080.7	25265.3	316697.7	80088
Occupancy	3829	476.7	284.7	157.3	318.5	17236.1	34671.6	597.1
Optdigits	180856.3	3178.4	101451	1214	2194.8	31604.6	1456326.2	4708.8
Pima	949.5	372.6	265.2	154.7	192.5	48019.8	26997	499
QSAR	6174.4	996.1	1233.1	292	357.7	56135.1	424913.9	1379.2
Rice-OC	8376.3	1377.6	825.7	335.7	550.2	27048.6	62251.1	1915
Seed	239.1	95.8	117.4	86.6	91	72200.3	14518.8	155.8
Sonar	1593.9	294.7	773.5	142.9	229.8	84155.6	713764.8	422.5
Spambase	123357.5	4887.7	5412.9	760.1	1617.5	20230.1	989565.7	7084.1
Tic-tac-toe	940.2	202.1	736.8	144.5	179.5	33018.8	39015.6	254
Titanic	1323.6	228.4	199.9	165.6	234.1	26465.9	28303	284.9
Transfusion	546.5	186.1	97.3	94.5	117.5	46330.7	14520.3	237.7
Vehicle	2429.6	650.2	480.3	207.8	243	61014.8	90847.4	798.4
Mean	23189.0	4319.6	5092.8	281.4	623.9	47389.2	177465.1	5440.0

Comment 4: When it comes to rule base complexity, as long as the number of rules and conditions is mentioned as part of the contributions of the paper, the definitions and analysis must be more precise and complete.

Response: Thank you for your valuable comments to improve our paper. We added a comparison of interpretability among rule-based systems, including the number of rules, the number of conditions, and the number of fired rules. The revisions are listed as follows.

Page 18, lines 52-58
Page 19, lines 43-58
Page 20, lines 3-28

~~compared to EBRB and FRB systems. The average~~ Besides accuracy, F1 score and running time, interpretability is also a crucial criterion for rule-based systems. Ref. [18] provides a detailed survey of interpretability measures for rule-based systems. In this experiment, we adopt three commonly used interpretability measures suggested in Ref. [18] to perform a quantitative analysis.

- Number of rules: The number of rules is an important interpretability measure for rule-based systems. This is based on the principle of Occam's razor, which states that the best model is the simplest one that fits the system behavior well. In other words, the set of fuzzy rules should be as small as possible while still preserving the model's performance to a satisfactory level.
- Number of conditions: The number of conditions should be minimized to enhance the readability of the rules.
- Number of fired rules: To ensure semantic interpretability, it is desirable to minimize the number of rules used in the reasoning process for a given input.

Based on the three interpretability measures, we tested the six rule-based methods. We evaluated them using the thirty-two real-world problems shown in Table 1, and the 10-CV model was used to calculate the average results. Table 10 presents the results of the three interpretability measures for the different methods.

Based on the measure of the number of rules, the proposed BRB classification system is found to have higher interpretability than EBRB, FRB and ensemble BRB classification systems, but lower interpretability than FRDT and SGERD methods. The interpretability of the proposed BRB classification system and FURIA was found to be close based on the number of rules ~~and conditions for our methodare 20.6 rules and 2.8 conditions, respectively, while for EBRB, it's 2007.1 rules and 14.6 conditions, and for FRB, it's 520.0 rules and 24.6 conditions. These two measuresare~~

Table 10: Interpretability measures for the eight rule-based systems.

	Numbers of rules								Numbers of conditions								Number of fired rules							
Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	95.4	3.3	28.3	4.7	2.6	180	3.4	4.3	7	1.4	7	4.5	6	2	7	1.5	95.4	1.3	1	1	1	77.2	1.2	1.5
Balance	562.5	24.3	66.2	3	4.2	180	4	23.1	4	3.1	4	0	3.2	2	4	3.1	562.5	2.6	1	1	1	51.3	2.1	2.5
Banknote	1234.8	12.5	30.8	6.9	3.4	180	3.9	12.5	4	2.3	4	1.2	3.4	2	4	2.4	1234.8	2.0	1	1	1	79.9	2.2	2.0
Bupa	310.5	9.8	43.2	9	3.3	180	3	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.6	1.2
Car	1554.3	83.5	690.5	27.5	1.6	180	3.3	78.7	6	4.3	6	3.6	6	2	6	4.3	1554.3	1.3	1	1	1	40.3	1.5	1.3
Column3C	279	7.8	26.8	10.8	3.5	180	3.1	8.1	6	2.5	6	1.9	5.3	2	6	2.6	279	1.4	1	1	1	79.5	1.9	1.4
Ecoli	302.4	18	63	29.7	8.9	180	3.9	17	7	2.9	7	2.0	1.3	2	7	2.9	302.4	1.9	1	1	1	58.0	1.6	1.8
German	900	9.4	885.4	22.5	2.4	180	3	10.1	20	2.3	20	3.1	19.8	2	20	2.3	900	1.4	1	1	1	41.7	1.2	1.4
Haberman	275.4	4	16.5	3.7	2.9	180	5.3	3.7	3	1.7	3	0.9	1.4	2	3	1.3	275.4	1.4	1	1	1	63.3	1.8	1.3
Hayes-roth	144	9.1	39.1	4.5	5.3	180	4.9	9.3	4	2.9	4	1.8	0.8	2	4	3.0	144	0.9	1	1	1	43.0	2.3	0.9
Heart-statlog	243	9.1	217.5	11.8	2.1	180	3	8.2	13	2.8	13	2.8	12	2	13	2.6	243	1.7	1	1	1	43.5	1.6	1.7
Housevotes	208.8	5.4	146.5	3	2	180	3	5.5	16	2.2	16	3	16	2	16	2.1	208.8	1.5	1	1	1	20	1.5	1.5
HTRU_2	16108.2	12.9	89.5	11.5	5.9	180	3.5	14.2	8	2.8	8	3.4	5.9	2	8	2.9	16108.2	3.2	1	1	1	80.0	1.1	3.7
Ipd	524.7	9.4	86	10.9	2.1	180	3	9.6	10	2.6	10	4.0	9.6	2	10	2.6	524.7	1.5	1	1	1	71.2	1.1	1.7
Iris	135	4.3	14.7	5.3	3.9	180	5.3	4.7	4	1.8	4	1.6	2.9	2	4	1.9	135	1.2	1	1	1	78.0	2.4	1.3
Liver-disorders	310.5	9.8	43.2	9	3.3	180	4.1	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.8	1.2
Magic	17118	29.8	312.8	12.1	2.8	180	3.2	26.9	10	3.2	10	2.9	9.2	2	10	3.2	17118	3.2	1	1	1	80.0	1.8	2.7
Mfeat-zernike	1800	10.8	1679.9	7.4	2.2	180	3	10.9	47	2.6	47	4.3	45.8	2	47	2.5	1800	3.4	1	1	1	79.9	1.0	3.4
New-thyroid	193.5	6.6	18.4	12.2	3.3	180	4.7	7.1	5	2.5	5	2.4	3.3	2	5	2.4	193.5	1.9	1	1	1	79.0	1.8	2.0
Nursery	11664	274.1	4222.6	18.4	5	180	3	276.7	8	5.0	8	4.0	5	2	8	5.0	11664	1.5	1	1	1	36.8	1.8	1.5
Occupancy	2398.5	12.7	25	3.8	2.8	180	4.1	13.2	5	2.7	5	2.9	3.2	2	5	2.7	2398.5	2.0	1	1	1	70.8	1.6	1.8
Optdigits	5058	35.1	5055.6	15	2	180	3	34.7	64	4.0	64	4.9	64	2	64	4.0	5058	4.6	1	1	1	45.2	1.0	4.6
Pima	691.2	7.8	101.3	8.6	4.1	180	3.6	8.3	8	2.4	8	2.6	6.1	2	8	2.4	691.2	1.8	1	1	1	69.6	1.8	1.8
QSAR	949.5	20.1	436.4	15.4	2.8	180	3	22.1	41	3.4	41	4.6	39.2	2	41	3.4	949.5	2.1	1	1	1	48.0	1.2	2.2
Rice-OC	3429	5.8	95.8	7.8	3.3	180	3.6	4.4	7	1.9	7	3.6	5.7	2	7	1.8	3429	2.5	1	1	1	80.0	1.7	1.9
Seed	189	8.5	53.2	9.6	3.8	180	3.2	8.4	7	2.2	7	3.0	5	2	7	2.2	189	1.8	1	1	1	79.3	2.1	1.7
Sonar	186.3	11.2	186.3	7.6	2.8	180	3	11	60	2.6	60	4.4	60	2	60	2.6	186.3	1.7	1	1	1	78.6	1.0	1.7
Spambase	4140.9	33.3	3562	9.8	2	180	3	34.9	57	4.2	57	5	56	2	57	4.3	4140.9	3.5	1	1	1	29.5	1.1	3.6
Tic-tac-toe	862.2	21.6	862.2	9.3	2	180	3	21	9	3.5	9	2.0	8	2	9	3.5	862.2	1.1	1	1	1	35.4	1.2	1.1
Titanic	1980.9	5.8	9.9	7.2	3	180	4.2	5.7	3	2.2	3	1.1	3	2	3	2.2	1980.9	2.1	1	1	1	26.1	2.1	2.1
Transfusion	673.2	5	12.2	4	2.7	180	4.7	5	4	1.8	4	1.3	3.7	2	4	1.8	673.2	1.5	1	1	1	71.2	1.8	1.6
Vehicle	761.4	23.6	227.6	22.6	6.2	180	3	24.1	18	3.3	18	3.3	15.4	2	18	3.4	761.4	1.3	1	1	1	79.0	1.7	1.2
Mean	2352.6	23.3	504.5	10.7	3.4	180.0	3.6	23.2	14.9	2.8	14.9	2.9	13.7	2.0	14.9	2.7	2352.6	1.9	1.0	1.0	1.0	61.7	1.6	1.9

~~EBRB, it's 2007.1 rules and 14.6 conditions, and for FRB, it's 520.0 rules and 24.6 conditions. These two measuresare negatively correlated with~~ measure. According to the number of conditions measure, the proposed method, along with FURIA, FRDT and ensemble BRB classification system, are found to have higher interpretability than EBRB, FRB, PMP-BRB classification system, and SGERD. Based on the number of fired rules measures, the interpretability of the ~~rule-based system [18]. The smaller their numbers, the higher the~~ FRB classification system, FRDT, and SGERD is found to be similar and higher than that of the proposed method. The interpretability of FURIA and PMP-BRB is comparable to the proposed method. However, the interpretability of EBRB and En-BRB is lower than that of the proposed method. Overall, the interpretability of the ~~system~~proposed method was found to be similar to that of FURIA and much higher than that of the EBRB, FRB, ensemble BRB classification systems.

Comment 5: The meaning of "cover" must also be precisely defined.

Response: Thank you for your valuable comments to improve our paper. We added the definition of "cover" in the revised manuscript. The revisions are listed as follows.

Page 6, line 25

Definition 11. *For a fuzzy rule $\widetilde{R}$, its purity is defined as follows:*

$$Pur_{\widetilde{R}} = \frac{\sum_{x_i \in X^R_{y_R}} \phi_{\widetilde{R}}(x_i)}{\sum_{x_i \in X^R} \phi_{\widetilde{R}}(x_i)}, \tag{14}$$

where X^R is the set of patterns in the training set X that are covered by rule R, $X^R = \{x_i \in X | \phi_{\widetilde{R}}(x_i) > 0\}$. $X^R_{y_R}$ is the subset of X^R that contains only the positive patterns covered by the rule R, $X^R_{y_R} = \{x_i \in X | \phi_{\widetilde{R}}(x_i) > 0$ and x_i belongs to class $y_R\}$. $\phi_{\widetilde{R}}(x_i)$ is the coverage of the rule $\widetilde{R}$ covering pattern x_i computed by the following formula.

$$\phi_{\widetilde{R}}(x_i) = \prod_{m=1}^{M} \mu_{\widetilde{A}_{\rho(m)}}(x_{i\rho(m)}), \tag{15}$$

where $\mu_{\widetilde{A}_{\rho(m)}}$ is the membership function of fuzzy number $\widetilde{A}_{\rho(m)}$. If $\phi_{\widetilde{R}}(x_i) > 0$, the rule $\widetilde{R}$ covers pattern x_i.

Comment 6: The algorithm proposed here presents a strong similarity with another algorithm described in the article introduced in another submission by the same authors, entitled "An Algorithm for Belief Rule Induction with Partial Ignorance". It is also mandatory that the authors explain the relation between both algorithms.

Response: Thank you for your valuable comments to improve our paper. the article "An Algorithm for Belief Rule Induction with Partial Ignorance" is a follow-up to this paper. In this paper, we transformed the fuzzy rules generated by FURIA into belief rules and employed evidence reasoning methods to combine the consequent BPAs. To leverage the advantages of DS evidence theory to the fullest extent, we improved the belief rules by introducing belief degrees for partial ignorance. This work was proposed in the article "An Algorithm for Belief Rule Induction with Partial Ignorance", where Dempster's rule was utilized to combine the consequent BPAs. We also mentioned this in the revised manuscript. The revisions are listed as follows.

Page 20, lines 45-47

The current belief rules only consider residual support and global ignorance while overlooking partial ignorance, thus not fully leveraging the advantages of evidence theory. In future work, we will continue to explore partial ignorance information within belief rules.

A belief rule-based classification system using fuzzy unordered rule induction algorithm

Yangxue Li[a], Ignacio Javier Pérez[a], Francisco Javier Cabrerizo[a], Harish Garg[b], Juan Antonio Morente-Molinera[a,*]

[a]*Andalusian Research Institute in Data Science and Computational Intelligence, Dept. of Computer Science and AI, University of Granada, 18071 Granada, Spain.*

[b]*School of Mathematics, Thapar Institute of Engineering and Technology (Deemed University), Patiala 147004, Punjab, India*

Abstract

A rule-based system is a widely used artificial intelligence system that employs a set of rules to make decisions. The belief rule-based (BRB) classification system is an extension of fuzzy rule-based (FRB) system that handles uncertainty and imprecision in classification tasks by incorporating Dempster-Shafer evidence theory and fuzzy set theory. However, the BRB classification system suffers from the combinatorial explosion and high time complexity problems. To solve these issues, the fuzzy unordered rule induction algorithm (FURIA) is introduced into the BRB classification systems to design a novel BRB classification system in this paper. FURIA is computationally less complex and has superior classification performance. The proposed system outperforms BRB classification systems in terms of classification performance and significantly reduces the number of rules and conditions. Furthermore, the proposed system offers a more effective solution than FURIA. To evaluate the validity and superiority of the proposed system, we conduct two classification experiments, comparing it with five traditional classifiers and seven rule-based systems, respectively, in terms of classification performance and interpretability.

Keywords:
rule-based system, pattern classification, belief rule-based classification system, FURIA, evidential reasoning

1. Introduction

A rule-based system is a type of artificial intelligence system that employs a set of rules to make decisions. The rules are typically expressed in an "if–then" format, where the "if" part specifies the conditions for the rule to be applied, and the "then" part specifies the action or conclusion to be taken if the conditions are met. The rule-based system is used for a wide range of applications, including expert systems [4], decision support systems [6, 21], and automated control systems [33, 17]. It is particularly useful in situations where there is a large amount of knowledge to be processed and where the knowledge is well-structured and can be expressed in a set of rules.

The BRB classification system [3, 20, 8] is an extension of FRB system that incorporates Dempster-Shafer (DS) evidence theory and fuzzy set theory to handle uncertainty and imprecision in classification tasks. The BRB classification system is composed of a collection of belief if-then rules that map input variables to output class labels, similar to traditional rule-based systems. However, instead of using a single crisp number or label to represent the consequent part of an if–then rule, belief rules use basic probability assignments (BPAs) to represent the uncertainty associated with the output variables. The BRB system utilizes the evidential reasoning (ER) approach [36] for inference, setting it apart from other rule-based inference methodologies. The BRB model offers two significant advantages in the areas of classification and prediction. Firstly, the process of modeling and reasoning based on if–then rules is transparent and traceable, which allows for the incorporation of expert knowledge to refine the model. Secondly, the BRB model is capable of effectively integrating both nominal and numerical information under conditions of uncertainty and incompleteness in the modeling process. Additionally, an extended belief rule has been developed that incorporates

*Corresponding authors.
Email addresses: liyangxueugr@gmail.com (Yangxue Li), jamoren@decsai.ugr.es (Juan Antonio Morente-Molinera)

belief degrees throughout all possible antecedent terms of the rule [37, 39]. This innovation enables the handling of vagueness, incompleteness, and uncertainty in a more flexible and straightforward manner. The utilization of extended belief rules in systems is referred to as the extended belief rule-based (EBRB) system.

When constructing a BRB model, it is essential to establish a set of referential values for each antecedent feature. This requirement may cause the problem of combinatorial explosion as a result of multiple referential values across several features. The determination of referential values, rule weights, and belief degrees is also an issue within the context of belief rule-based systems. Optimization methods are extensively employed as a means to address this issue, such as, differential evolution (DE) algorithm [38, 41], genetic algorithm (GA) [32, 7], particle swarm optimization (PSO) [9]. However, using optimization algorithms to train a large number of parameters can lead to additional issues, such as increased time complexity and unstable classification performance. EBRB classification system is a solution to these problems. But for EBRB model, each training pattern is employed to generate a rule, leading to numerous rules being stored in the knowledge base. Additionally, the EBRB inference is susceptible to the rule inconsistency problem, where almost all extended belief rules are activated for any input pattern. Therefore, large numbers of rules and fired rules lead to lower interpretability of the model.

To solve the problems of high time complexity and combinatorial explosion of traditional BRB classification system, we introduce the FURIA in BRB classification systems. As a state-of-the-art fuzzy rule learner, FURIA is an advanced derivative of the repeated incremental pruning to produce error reduction (RIPPER) algorithm [19]. Specifically, FURIA adopts a different approach to learning fuzzy rules compared to traditional methods. Instead of using pre-defined referential values and fuzzy partitions, FURIA employs a technique that maximizes information gain to generate fuzzy antecedents. This approach is computationally less complex and has been shown to enhance performance. This paper introduces a novel BRB classification system that leverages the FURIA. The rule generation process comprises three primary stages: (i) crisp rule generation, (ii) fuzzification of the crisp rules, and (iii) converting the fuzzy rules to belief rules. Additionally, the proposed BRB classification system employs the ER approach for inference. Unlike traditional BRB systems, the proposed system employs activation weight that is directly correlated with the coverage of the rule covering pattern. Compared to the traditional BRB classification system, the novel approach not only yields superior classification performance but also significantly reduces the number of rules, conditions, and fired rules. This implies that the proposed system is more interpretable while still achieving superior performance. Moreover, the developed system builds on the strengths of the DS evidence theory, enabling the integration of nominal and numerical information in modeling processes under uncertainty and incompleteness. In contrast to FURIA, the proposed system provides a more effective solution.

In order to substantiate the efficacy and superiority of the proposed system, we conducted two classification experiments on 32 benchmark datasets. The first experiment entailed a comparison of our method against five conventional classifiers, such as the C4.5 decision tree [28], support vector machine (SVM) [11], and k-nearest neighbors (KNN) classifiers [12], on the basis of their respective classification performance. The second experiment entailed a comparison of our method against seven widely-used rule-based classification systems, such as the FRB classification system [10], FURIA [19] and EBRB classification system [23] with emphases on classification performance and interpretability.

In summary, this paper makes several key contributions as follows.

(1) To overcome the problems of BRB classification system, FURIA is introduced into BRB classification system, which can improve the classification performance and reduce the time complexity of the BRB system.

(2) To generate belief rules for the proposed BRB classification system, the method of converting fuzzy rules into belief rules is proposed. The DS evidential framework is used to describe the consequent part of the fuzzy rule.

(3) The effectiveness and superiority of the proposed method in handling classification tasks have been validated by comparing it to some state-of-the-art models.

The structure of this paper is outlined as follows: In Section 2, we provide background information on Dempster-Shafer evidence theory and FURIA. Next, in Section 3, we introduce the BRB classification system based on FURIA. In Section 4, we present two experiments that evaluate the accuracy and running time of the proposed system. Finally, we conclude the paper in Section 5.

2. Preliminaries

In this section, we review some background knowledge of DS evidence theory and FURIA related to this paper.

2.1. Dempster-Shafer evidence theory

DS evidence theory is a mathematical framework for reasoning with uncertain information [13, 30]. It is a probabilistic theory that allows for the representation of uncertainty in a more flexible and intuitive manner than traditional probability theory. In this framework, uncertainty is represented using BPAs, which assign a belief degree to each possible outcome of a given event. These BPAs are then combined to make inferences and decisions based on available evidence. The DS evidence theory has found applications in various fields such as information fusion [40, 14], decision-making [15, 16], and data analysis [24, 22]. Some definitions of evidence theory are shown below.

Definition 1. *The DS evidence theory defines the frame of discernment as the set of all possible hypotheses or propositions that are mutually exclusive and exhaustive, which may be true or false about a specific event or phenomenon. This set is denoted as follows [13, 30].*

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_c\}. \tag{1}$$

The central idea of the DS theory is the concept of the "basic probability assignment (BPA)" which assigns a belief degree to each element of the power set of the frame of discernment 2^Θ defined as following [13, 30].

Definition 2. *A BPA is a mapping $m(\cdot) : 2^\Theta \to [0, 1]$ such that $m(\varnothing) = 0$ and*

$$\sum_{A \subseteq \Theta} m(A) = 1. \tag{2}$$

Definition 3. *For a BPA $m(\cdot)$, the belief function $Bel(A)$ for $\forall A \subseteq \Theta$ is defined as follows [13, 30].*

$$Bel(A) = \sum_{B \subseteq A} m(B). \tag{3}$$

$Bel(A)$ represents the total belief assigned to all subsets of Θ that are contained within A. This value is interpreted as the lower limit of the probability of A.

Definition 4. *For a BPA m, the plausibility function $Pl(A)$ for $\forall A \subseteq \Theta$ is defined as follows [13, 30]:*

$$Pl(A) = \sum_{B \cap A \neq \varnothing, B \subseteq \Theta} m(B). \tag{4}$$

$Pl(A)$ represents the total belief assigned to all subsets of Θ that are compatible with A. This value is interpreted as the upper limit of the probability of A.

The Dempster's combination rule is defined as follows for combining two or more BPAs.

Definition 5. *Given two BPAs $m_1(\cdot)$ and $m_2(\cdot)$ defined on the same frame of discernment Θ, their combined BPA $m(\cdot)$ is defined as follows [13].*

$$m(A) = \begin{cases} 0, & A = \varnothing, \\ \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K}, & A \neq \varnothing, \end{cases} \tag{5}$$

where $B, C \subseteq \Theta$, and $K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$ is the conflict coefficient. When $K = 1$, the Dempster's combination rule is invalid. Dempster's rule is commutative and associative.

To deal with the counter-intuitive problem of Dempster's rule, Yang et al. established ER approach to combine multiple pieces of independent evidence conjunctively with weights and reliabilities [36]. In ER approach, the assessment of all possible hypotheses is regarded as a piece of evidence. Given two BPAs m_1 and m_2 with their weights w_1

and w_2 defined on the same frame of discernment $\Theta = \{\theta_1, \theta_2, \ldots, \theta_c\}$, such that $w_1, w_2 \in [0, 1]$ and $w_1 + w_2 = 1$. First, two new BPA m_j^w, $j = 1, 2$ is generated as weighted belief degrees as follows [36].

$$\begin{aligned} m_j^w(\theta_i) &= w_j m_j(\theta_i), i = 1, 2, \cdots, c \\ m_j^w(\Theta) &= w_j m_j(\Theta), \\ m_j^w(P(\Theta)) &= 1 - w_j, \end{aligned} \tag{6}$$

where $m_j^w(P(\Theta))$ is the remaining support left uncommitted by m_j, coined as the residual support of m_j. Then the weighted BPAs m_1^w and m_2^w are combined into a joint probability masses m'_{12} for θ_i, Θ and $p(\Theta)$ using the following equations [36]:

$$\begin{aligned} m'_{12}(\theta_i) &= K(m_1^w(\theta_i)m_2^w(\theta_i) + m_1^w(\theta_i)(m_2^w(\Theta) + m_2^w(P(\Theta))) + m_2^w(\theta_i)(m_1^w(\Theta) + m_1^w(P(\Theta)))), \\ m'_{12}(\Theta) &= K(m_1^w(\Theta)m_2^w(\Theta) + m_1^w(\Theta)m_2^w(P(\Theta)) + m_2^w(\Theta)m_1^w(P(\Theta))), \\ m'_{12}(P(\Theta)) &= K(m_1^w(P(\Theta))m_2^w(P(\Theta))), \end{aligned} \tag{7}$$

where $K = (1 - \sum_{i=1}^{c} \sum_{t=1, t\neq i}^{c} m_1^w(\theta_i)m_2^w(\theta_t))^{-1}$. If there are more than two BPAs to be combined, the process described in Eq. (7) can be repeated to combine the third BPA with the previously-combined assessment $m'_{12}(\cdot)$, and so on, until all BPAs are recursively combined. Finally, to obtain the combined BPA without loss of generality, the $m'_{12}(P(\Theta))$ is reassigned back to all focal elements as follows [36].

$$\begin{aligned} m_{12}(\theta_i) &= \frac{m'_{12}(\theta_i)}{1 - m'_{12}(P(\Theta))}, i = 1, 2, \ldots, c, \\ m_{12}(\Theta) &= \frac{m'_{12}(\Theta)}{1 - m'_{12}(P(\Theta))}, \end{aligned} \tag{8}$$

where the degree of $P(\Theta)$ is reassigned to the focal elements.

2.2. Fuzzy unordered rule induction algorithm

FURIA is an algorithm derived from the RIPPER algorithm, with several enhancements that improve its performance in rule induction tasks [19]. It produces fuzzy rules that offer a greater degree of flexibility in modeling classification boundaries as opposed to conventional rules. The fuzzy rules are generated by replacing fuzzy intervals through trapezoidal fuzzy numbers in combination with the highly sophisticated rule induction technique utilized by the original RIPPER algorithm. FURIA employs an unordered set of rules for each class in a one-vs-rest scheme. Below, we provide a detailed explanation of some key definitions and algorithms that underpin the FURIA algorithm.

FURIA's fuzzy rules are generated through the process of fuzzifying the final crisp rules obtained from a modified version of the RIPPER algorithm. This modified algorithm skips the pruning step of the original RIPPER and instead learns the initial ruleset directly from the entire training dataset. Suppose we have a training dataset X comprising n patterns, $X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$, each with d features, $x_i = \{x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{id}\}$, where x_{ij} is the value of x_i with respect to feature F_j, $F_j \in \{F_1, F_2, \ldots, F_d\}$. The dataset X contains c distinct output classes, which are denoted as $\Theta = \{\theta_1, \theta_2, \ldots, \theta_c\}$. The modified RIPPER algorithm can be divided into two phases: building and optimization. During the rule building phase, the IREP* [2] algorithm is used without the pruning step.

Definition 6. *A crisp rule R induced by the modified RIPPER takes the following form:*

$$\textit{Rule R: If } F_{\rho(1)} \textit{ is } A_{\rho(1)} \textit{ and } F_{\rho(2)} \textit{ is } A_{\rho(2)} \textit{ and } \cdots \textit{ and } F_{\rho(M)} \textit{ is } A_{\rho(M)} \textit{ then Class } y_R, \tag{9}$$

where $\rho(\cdot)$ is an index function taken from set $\{1, \ldots, d\}$ and indicate the index of the feature considered in rule. M is the number of considered features. $A_{\rho(1)}, \ldots, A_{\rho(M)}$ represent the antecedents or selectors of the rule. If feature $F_{\rho(m)}$ is nominal, then $A_{\rho(m)}$ takes the value of the corresponding feature value. If $F_{\rho(m)}$ is numerical, then $A_{\rho(m)}$ takes the form of an interval $[\underline{v}_m, \overline{v}_m]$, $\underline{v}_m \leq \overline{v}_m$ and $\underline{v}_m, \overline{v}_m \in \mathcal{R}$. The output or consequent of the rule, denoted by $y_R \in \Theta$, represents the class that the rule is predicting.

A dataset may encompass a multitude of features, among which numerous may be irrelevant or redundant to the classification task. Considering all features in the rules would lead to a combinatorial explosion, thereby escalating algorithmic time costs and diminishing interpretability. Hence, it is imperative to undertake features selection during rule generation. Information gain is commonly employed as a metric for sorting and filtering features. The commonly used measures of information gain include mutual information gain and gain ratio.

Definition 7. *The mutual information gain is a metric used to quantify the correlation between two variables, which is defined as follows.*

$$\begin{aligned} IG_M(R) &= I(p,n) - E(R), \\ E(R) &= -\frac{p_R + n_R}{p+n} I(p_R, n_R), \\ I(p,n) &= -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}. \end{aligned} \tag{10}$$

Where p_R and n_R are the number of positive and negative patterns covered by rule R while p and n are the number of positive and negative patterns covered by default rule.

Definition 8. *The gain ratio is an extension of information gain, designed to overcome the bias of information gain towards features with a larger number of values.*

$$GR(R) = \frac{IG_M(R)}{-\sum_{i=1}^{c} \frac{|X_{\theta_i}|}{|X|} \times \log_2 \frac{|X_{c_i}|}{|X|}}, \tag{11}$$

where $|X|$ is the number of patterns in training data and $|X_{\theta_i}|$ is the number of patterns in training data belong to class θ_i.

To learn a new rule on the training data, FURIA employs a propositional version of the FOIL algorithm [26]. For each class, the algorithm begins with an empty conjunction of antecedents and gradually adds more antecedents until the rule can no longer cover any negative instances. The next antecedent to be added is chosen in such a way that it maximizes FOIL's information gain (IG) criterion, which is defined as follows [26].

Definition 9. *FOIL's information gain is a measure of how much the rule improves upon the default rule for the target class [26]. It is calculated using the following formula.*

$$IG_F(R) = p_R \times (\log_2(\frac{p_R}{p_R + n_R}) - \log_2(\frac{p}{p+n})), \tag{12}$$

where p_R and n_R are the number of positive and negative patterns covered by rule R while p and n are the number of positive and negative patterns covered by default rule.

The objective of FOIL is to generate rules with high generality, meaning rules that can generalize to new instances. Therefore, FOIL's information gain is utilized to assess the generality of rules rather than evaluating the classification ability on a specific dataset. The uniqueness of FOIL lies in its approach of guiding the learning process by obtaining information gain from positive examples not mentioned in the rules.

Subsequently, optimization is applied to the generated rules. The optimization process involves re-examining the ruleset RS produced in the building phase. For each rule R in the ruleset, two alternative rules R_{rep} and R_{rev} are created. R_{rep} is an empty replacement rule, which is grown in a way that minimizes the error of the modified ruleset $(RS \cup \{R_{rep}\})/\{R\}$. R_{rev} is the revision rule that is created in the same way, but it starts from R instead of the empty rule. The MDL (Minimum Description Length) criterion is used to determine which version of the rule to retain [26].

After obtaining the final crisp rules from the modified RIPPER algorithm, FURIA fuzzify these crisp rules for numerical features using training dataset.

Definition 10. *After fuzzification, a crisp rule is transformed into a fuzzy rule $\widetilde{R}$ with the following form:*

$$\textit{Rule } \widetilde{R}\textit{: If } F_{\rho(1)} \textit{ is } \widetilde{A}_{\rho(1)} \textit{ and } F_{\rho(2)} \textit{ is } \widetilde{A}_{\rho(2)} \textit{ and } \cdots \textit{ and } F_{\rho(M)} \textit{ is } \widetilde{A}_{\rho(M)} \textit{ then Class } y_R, \tag{13}$$

where $\rho(\cdot)$ is an index function from set $\{1,\ldots,d\}$ and indicate the index of feature considered in rule. $\widetilde{A}_{\rho(1)},\ldots,\widetilde{A}_{\rho(M)}$ are a conjunction of antecedents (selectors). If feature $F_{\rho(m)}$ is nominal, $\widetilde{A}_{\rho(m)}$ takes a value of feature $F_{\rho(m)}$. If $F_{\rho(m)}$ is numerical, $\widetilde{A}_{\rho(m)}$ is a trapezoidal fuzzy number $\widetilde{A}_{\rho(m)} = (\underline{v}_m^L, \underline{v}_m, \overline{v}_m, \overline{v}_m^U)$. $y_R \in \Theta$ is a consequent class.

FURIA fuzzifies the interval antecedents $A_{\rho(m)}$ in Eq. (9) into the fuzzy antecedents $\widetilde{A}_{\rho(m)}$ in Eq. (13).

Purity is a measure of the quality of a fuzzy rule, which is defined as the proportion of coverage of correctly classified patterns covered by the rule over that of all covered patterns. A higher purity indicates better classification performance of the rule on the training set.

Definition 11. *For a fuzzy rule $\widetilde{R}$, its purity is defined as follows:*

$$Pur_{\widetilde{R}} = \frac{\sum_{x_i \in X_{y_R}^R} \phi_{\widetilde{R}}(x_i)}{\sum_{x_i \in X^R} \phi_{\widetilde{R}}(x_i)}, \tag{14}$$

where X^R is the set of patterns in the training set X that are covered by rule R, $X^R = \{x_i \in X | \phi_{\widetilde{R}}(x_i) > 0\}$. $X_{y_R}^R$ is the subset of X^R that contains only the positive patterns covered by the rule R, $X_{y_R}^R = \{x_i \in X | \phi_{\widetilde{R}}(x_i) > 0$ and x_i belongs to class $y_R\}$. $\phi_{\widetilde{R}}(x_i)$ is the coverage of the rule $\widetilde{R}$ covering pattern x_i computed by the following formula.

$$\phi_{\widetilde{R}}(x_i) = \prod_{m=1}^{M} \mu_{\widetilde{A}_{\rho(m)}}(x_{i\rho(m)}), \tag{15}$$

where $\mu_{\widetilde{A}_{\rho(m)}}$ is the membership function of fuzzy number $\widetilde{A}_{\rho(m)}$. If $\phi_{\widetilde{R}}(x_i) > 0$, the rule $\widetilde{R}$ covers pattern x_i.

Definition 12. *For a new pattern x, the support degree of class θ_k is defined as follows.*

$$sup_k(x) = \sum_{\widetilde{R} \in \widetilde{RS}, y_R = \theta_k} \phi_{\widetilde{R}}(x) \cdot CF(\widetilde{R}), \tag{16}$$

where $CF(\widetilde{R})$ is the certain factor of rule $\widetilde{R}$, which is defined as follows.

$$CF(\widetilde{R}) = \frac{2\frac{|X_{y_R}|}{|X|} + \sum_{x_i \in X_{y_R}} \phi_{\widetilde{R}}(x_i)}{2 + \sum_{x_i \in X} \phi_{\widetilde{R}}(x_i)}, \tag{17}$$

where $|X_{y_R}|$ is the number of patterns in training data belong to class y_R and $|X|$ is the number of patterns in training data.

3. A novel belief rule-based classification system

To highlight the benefits of belief rules in representing and reasoning with uncertain information, as well as the exceptional classification performance and interpretability of FURIA, we extend the fuzzy rules generated by FURIA to the belief degrees framework. Through this extension, we develop a new BRB classification system, which can provide superior classification performance while still maintaining interpretability. The proposed BRB classification system, as depicted in Fig. 1, is composed of two main components: the knowledge base and the inference engine. The knowledge base stores a collection of belief if–then rules, which are generated from training data. The training model is responsible for generating these rules, and in the figure, we can see that a collection of crisp rules is generated by the modified RIPPER algorithm from the training data. These rules are then fuzzified to obtain fuzzy rules, and finally, belief rules are derived by calculating belief degrees. The inference engine is used to obtain the output classes of the testing patterns based on the rules stored in the knowledge base.

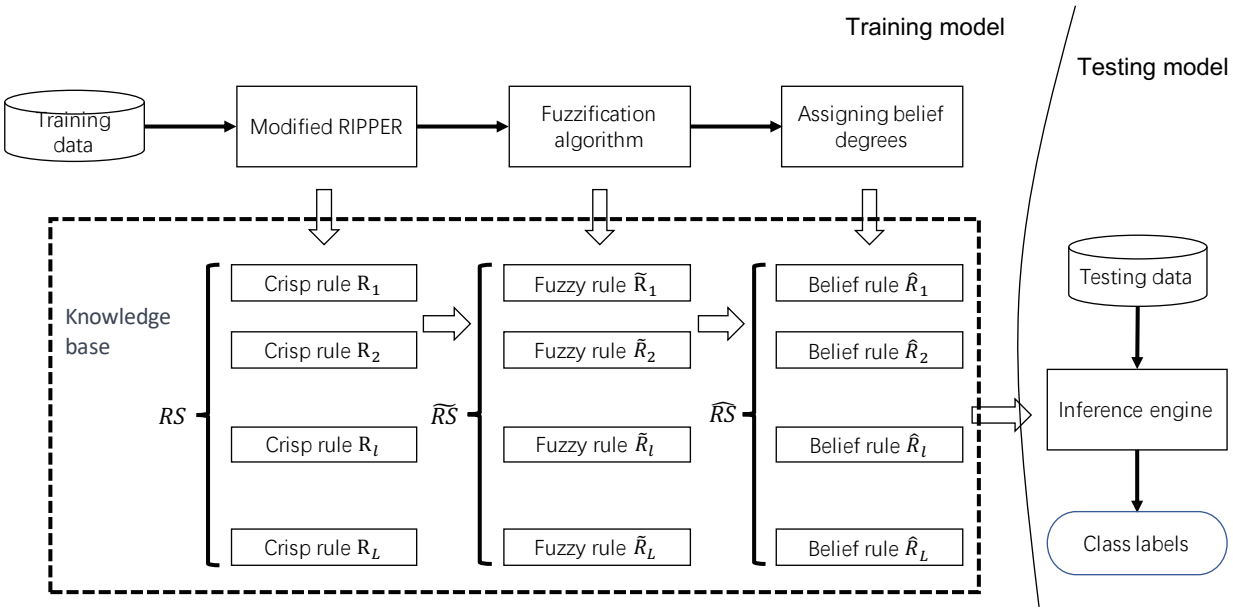

Figure 1: The framework of the proposed BRB classification system.

3.1. Belief rule representation

Belief rules are critical components of a BRB expert system, serving as an extension of the conventional fuzzy if-then rule. Such a rule takes into account the belief degrees of consequent classes, providing a more comprehensive approach to classification. In a traditional belief rule, each antecedent feature assumes multiple referential values, while all possible consequent classes are associated with belief degrees.

Definition 13. *A belief rule is defined in Ref. [35] as follows:*

$$\textit{Rule } \hat{R}\textit{: If } F_1 \textit{ is } a_1 \textit{ and } F_2 \textit{ is } a_2 \textit{ and } \cdots \textit{ and } F_d \textit{ is } a_d \textit{ then the consequent is } C = \{(\theta_1,\beta_1),\ldots,(\theta_c,\beta_c)\}, \tag{18}$$

where $F_1, F_2, \ldots, F_d$ are the antecedent features. a_j, $j = 1, 2, \ldots, d$, is the referential value of feature F_j. $\theta_1, \theta_2, \ldots, \theta_c$ are consequent classes and β_k is the support degree to believe θ_k is true, satisfying $\beta_k \in [0, 1]$ and $\sum_{k=1}^{c} \beta_k \leq 1$.

Here's an example of a belief rule in the Iris classification problem:

$\hat{R}$: If *sepal length* is *medium* and *sepal width* is *medium* and *petal length* is *long* and *petal width* is *short* then *specie* is $\{(setosa, 0), (versicolor, 0.3), (virginica, 0.7)\}$.

The extended belief rule is more commonly used in classification problems. The rule in Eq. (18) is extended by embedding belief degrees in all possible antecedent terms of each rule. This extension enables the rule to handle vagueness, incompleteness, and uncertainty simultaneously, while also providing a more flexible and simpler knowledge base generation scheme.

Definition 14. *An extended belief rule is of the form:*

$$\begin{aligned}&\textit{Rule } \hat{R}\textit{: If } F_1 \textit{ is } \{a_1,\alpha_1\} \textit{ and } F_2 \textit{ is } \{a_2,\alpha_2\} \textit{ and } \cdots \textit{ and } F_d \textit{ is } \{a_d,\alpha_d\}\\ &\qquad \textit{then the consequent is } C = \{(\theta_1,\beta_1),\ldots,(\theta_c,\beta_c)\},\end{aligned} \tag{19}$$

where $\{a_j, \alpha_j\}$ is the packet antecedent with belief distribution, $\{a_j, \alpha_j\} = \{(a_{jt}, \alpha_{jt}), t = 1, 2, \ldots, T_j\}$. T_j is the number of referential values of antecedent feature F_j.

Here's an example of an extended belief rule in the Iris classification problem:

$\hat{R}$: If *sepal length* is $\{(short, 0.18), (medium, 0.82), (long, 0)\}$ and *sepal width* is $\{(short, 0), (medium, 0.77), (long, 0.23)\}$ and *petal length* is $\{(short, 0), (medium, 0.38), (long, 0.62)\}$ and *petal width* is $\{(short, 0.85), (medium, 0.15), (long, 0)\}$ then *specie* is $\{(setosa, 0), (versicolor, 0.3), (virginica, 0.7)\}$.

This extended belief rule includes belief degrees embedded in each possible antecedent term and each possible class label. It can handle the vagueness and uncertainty present in the classification problem.

The referential value can be assigned either as a fuzzy membership function [35] or a numerical value through a utility-based equivalence transformation technique [34] and optimization methods. The above two forms show that traditional methods typically consider all features and do not perform feature selection. In this paper, we employ FURIA to generate belief rules that minimize both the number of rules and conditions. Moreover, rather than relying on numerical values and traditional triangular fuzzy numbers as referential values, we utilize trapezoidal fuzzy numbers generated by FURIA.

Definition 15. *A belief rule used in this paper is of the form:*

$$\begin{aligned} &\textit{Rule } \hat{R}\textit{: If } F_{\rho(1)} \textit{ is } \widetilde{A}_{\rho(1)} \textit{ and } F_{\rho(2)} \textit{ is } \widetilde{A}_{\rho(2)} \textit{ and } \cdots \textit{ and } F_{\rho(M)} \textit{ is } \widetilde{A}_{\rho(M)} \\ &\textit{then the consequent is } C = \{(\theta_1, \beta_1), \ldots, (\theta_c, \beta_c), (\Theta, \beta_\Theta)\}, \end{aligned} \tag{20}$$

where β_k, $k = 1, 2, \ldots, c$ is the support degree to believe θ_k is the consequent class label for belief rule $\hat{R}$, satisfying $\sum_{k=1}^{c} \beta_k \leq 1$ and $\beta_\Theta = 1 - \sum_{k=1}^{c} \beta_k$.

The consequent part of a belief rule is a BPA on the frame of discernment $\Theta = \{\theta_1, \theta_2, \ldots, \theta_c\}$. For example, an example of the belief rule about Iris classification problem: $\hat{R}$: If *petal length* is $(1.9cm, 3.0cm, 5.0cm, 5.1cm)$ and *petal width* is $(-\infty, -\infty, 1.7cm, 1.8cm)$ then *specie* is $\{(setosa, 0.02), (versicolor, 0.96), (virginica, 0.02), (\Theta, 0)\}$.

3.2. Rule generation

Belief rules are generated from numerical data and stored in the knowledge base for used by the inference engine. Rule generation comprises three stages: (i) crisp rule induction, (ii) fuzzification of the crisp rules, and (iii) converting the fuzzy rules to belief rules.

Step 1: **Generate crisp rules using the modified RIPPER algorithm.**

The modified RIPPER algorithm omits the pruning step of the original method, retaining all the generated crisp rules, and instead of learning on the growing data by training on the entire training data [19]. The main idea of the modified RIPPER algorithm is to search for a rule that attains maximum accuracy for a class, continuing the search until no further patterns belonging to the class are present in the remaining data. The main algorithm of modified RIPPER algorithm is shown in Algorithm 1. The set of obtained crisp rules is denoted as RS. The function $findrule(X_{nc}, \theta_k)$ is to find a rule with maximum accuracy for class θ_k on data X_{nc}, which is shown in Algorithm 2.

In function $findrule(X_{nc}, \theta_k)$, the split point v_j and operator o_j indicates that if o_j is '$\geq$', the selector is 'F is $[v_j, +\infty)$'; if o_j is '$\leq$', the selector is 'F is $[-\infty, v_j]$'; if o_j is '=', the selector is 'F is v_j'. And if there are two selectors $(\underline{v}_j, o_j)$ and $(\overline{v}_j, o'_j)$ on a same feature F, $\underline{v}_j < \overline{v}_j$, o_j is '$\geq$' and o'_j is '$\leq$', these two selectors can be seemed as one selector 'F is $[\underline{v}_j, \overline{v}_j]$'. Rule generation begins with an empty rule, the selectors in the rule are incrementally expanded. At each step, the selector that yields the maximum information gain is chosen, and the current rule is set as the default rule for the subsequent selections.

The function $optimization(X, RS, nF)$ is employed for rule optimization. Firstly, the dataset X is randomly partitioned into nF folds, with the first $nF - 1$ folds serving as growing dataset and the final fold utilizing as pruning dataset. For each rule R in the ruleset, two alternative rules R_{rep} and R_{rev} are created. The MDL (Minimum Description Length) criterion is used to determine which version of the rule to retain [27]. For a detailed algorithmic description of $optimization(X, RS, nF)$, reference can be made to FURIA [19].

Example 1. *The Iris dataset consists of 150 patterns, each with four feature values: sepal length, sepal width, petal length, and petal width. These patterns belong to three distinct classes: setosa, versicolor, and virginica. We utilize 90% of the dataset, which corresponds to 135 patterns, for training the rules. Based on the Algorithm 1 and 2, here is the process to obtain a crisp rule for class versicolor. The default rule is the entire training dataset.*

Algorithm 1 The modified RIPPER algorithm.

Input: The training dataset $X = \{(x_i, y_i)|y_i \in \Theta\}$; the number of optimization nOP; the number of folds nF.
Output: Crisp ruleset $RS = \{R_1, \ldots, R_L\}$.

$l \leftarrow 1$
for each class $\theta_k \in \Theta$ **do**
  $X_{nc} \leftarrow X$, $X_c \leftarrow \varnothing$
  while X_{nc} has patterns belonging to θ_k **do**
    $R_l \leftarrow findrule(X_{nc}, \theta_k)$
    $l \leftarrow l + 1$
    $X_c \leftarrow$ patterns covered by R_l, $X_{nc} \leftarrow X/X_c$
  end while
end for
$nO \leftarrow 1$
for $nO \leq nOP$ **do**
  $optimization(X, RS, nF)$
  $nO \leftarrow nO + 1$
end for

Algorithm 2 $findrule(X, \theta_t)$.

Input: The dataset $X = \{(x_i, y_i)|y_i \in \Theta\}$; the target class $\theta_t \in \Theta$.
Output: Crisp rule R with antecedents $A_{\rho(1)}, \ldots, A_{\rho(M)}$ and consequent class y_R.

$y_R = \theta_t$, $IG_{max} = 0$, $r = null$, $R = null$, $X_c = X$, $F = null$
while X_c has negative patterns **do**
  for each feature F_j **do**
    find the split point v_j and operator o_j, $o_j \in \{\geq, \leq, =\}$ to obtain the maximum information gain, IG_j, in X_c.
    if $IG_j > IG_{max}$ **then**
      $IG_{max} \leftarrow IG_j$, $r \leftarrow (v_j, o_j)$, $F = F_j$
    end if
  end for
  R adds (F is r)
  $X_c \leftarrow$ patterns covered by R, $X_{nc} \leftarrow X/X_c$
end while

For class versicolor, we individually search for the maximum information gain split point and operator for each attribute (where 1 *is used to avoid division by zero errors).*

sepal length $\geq 5.5cm$ *covers* 89 *patterns and out of which* 40 *belong to class versicolor.*

$$IG_1 = 40 \times (\log_2(\frac{45+1}{135+1}) - \log_2(\frac{40+1}{89+1})) = 17.18.$$

sepal width $\leq 2.9cm$ *covers* 52 *patterns and out of which* 31 *belong to class versicolor.* $IG_2 = 25.92$.

petal length $\geq 3.0cm$ *covers* 90 *patterns and out of which* 45 *belong to class versicolor.* $IG_3 = 26.09$.

petal with $\geq 1.0cm$ *covers* 90 *patterns and out of which* 45 *belong to class versicolor.* $IG_4 = 26.09$.

$IG_{max} = IG_3$*, then petal length* $\geq 3.0cm$ *is the first antecedent,* $A_{\rho(1)}$ = *petal length is* $[3.0cm, +\infty)$*. Afterward, from covered* 90 *patterns by the first antecedent, we continue selecting the next antecedent using the same method,* $A_{\rho(2)}$ = *petal width* $\leq 1.7cm$*. Repeat this process iteratively until the last antecedent does not cover any negative patterns to obtain a crisp rule R.*

R: If petal length is $[3.0cm, +\infty)$ *and petal width is* $(-\infty, 1.7cm]$ *and petal length is* $(-\infty, 5.0cm]$ *and then specie is versicolor.*

In which, the first antecedent and the third antecedent are on the same feature, therefore rule R also can be written as:

R: If petal length is $[3.0cm, 5.0cm]$ *and petal width is* $(-\infty, 1.7cm]$ *then specie is versicolor.*

Step 2: **Fuzzify crisp rules to obtain fuzzy rules.**

The generation of crisp rules identifies the pure regions associated with each class, while the process of fuzzification aims to identify its boundary regions. Despite the fact that all rules are unordered, the antecedents of the rules are ordered. Each antecedent further filters the patterns covered by the previous antecedent in order to determine the next antecedent. Thus, the fuzzification process begins with the first antecedent over the entire training patterns. Subsequent antecedents are fuzzified on the patterns selected by the previous antecedent the final antecedent of the rule is completely fuzzified. It must be pointed out that only numerical features will be subjected to fuzzification. The fuzzification processes can be shown in Algorithm 3. The collection of obtained fuzzy rules is denoted as $\widetilde{RS}$.

Algorithm 3 The antecedent fuzzification algorithm for a crisp rule R

Input: The training dataset $X = \{(x_i, y_i), y_i \in \Theta\}$, a crisp rule R with antecedents $A_{\rho(1)}, \ldots, A_{\rho(M)}$.

Output: A fuzzy rule $\widetilde{R}$ with antecedents $\widetilde{A}_{\rho(1)}, \ldots, \widetilde{A}_{\rho(M)}$.

```
for i = 1 to M do
    if feature F_ρ(i) is numerical then
        Compute the covered patterns X_base by rule with antecedents A_ρ(1), ..., A_ρ(i−1), A_ρ(i+1), ..., A_ρ(m).
        Compute the values of v_i^L and v̄_i^U to maximum the purity of rule with antecedent A_ρ(i) based on X_base.
        Ã_ρ(i) ← (v_i^L, v_i, v̄_i, v̄_i^U).
    else
        Ã_ρ(i) ← A_ρ(i).
    end if
end for
```

Example 2. *Sequentially fuzzify the antecedents in the crisp rule R obtained from Example 1.*

For the first antecedent, petal length is $[3.0cm, +\infty)$*, of crisp rule R, there are* 87 *patterns covered by* $A_{\rho(2)}$ *and* $A_{\rho(3)}$*. Within the* 87 *patterns, the petal length values of the patterns that are less than* $\underline{v}_1 = 3.0cm$ *are considered as boundary values* $\underline{v}_1^L$*, and their purity are calculated. The boundary value with the highest purity is selected as the final boundary value,* $\underline{v}_1^L = 1.9cm$ *with maximum purity* $Pur_{\widetilde{A}_{\rho(1)}} = 0.9555$*.* $\overline{v}_1^U = +\infty$ *because of* $\overline{v}_1 = +\infty$*. Therefore,* $\widetilde{A}_{\rho(1)} = (1.9cm.3.0cm, +\infty, +\infty)$.

For the second antecedent petal width is $(-\infty, 1.7cm]$ *of crisp rule R, there are* 49 *patterns covered by* $A_{\rho(1)}$ *and* $A_{\rho(3)}$*. Within the* 49 *patterns, the petal width values of the patterns that are greater than* $\overline{v}_2 = 1.7cm$ *are considered*

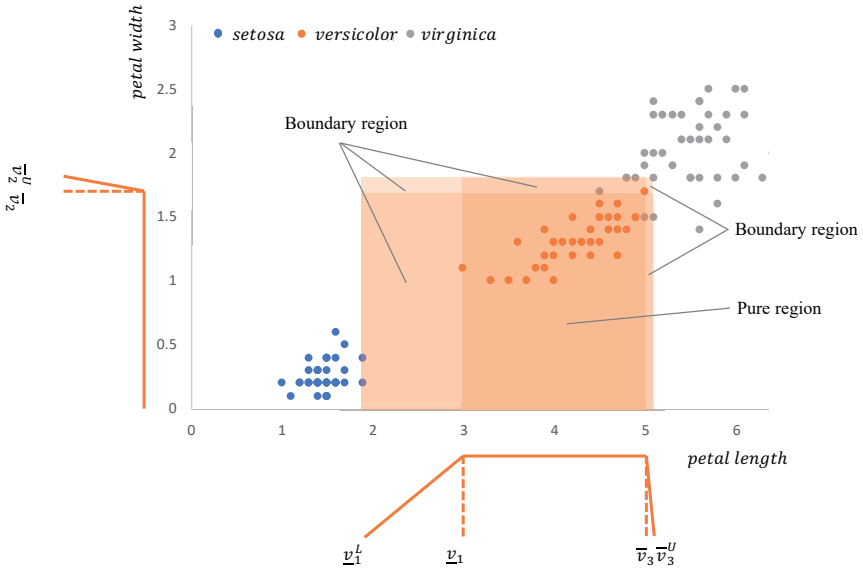

Figure 2: The fuzzy rule $\widetilde{R}$: If *petal length* is $(1,9cm,3.0cm,5.0cm,5.1cm)$ and *petal width* is $(-\infty,-\infty,1.7cm,1.8cm)$ then *specie* is *versicolor*.

as boundary values $\overline{v}_2^U$*, and their purity are calculated. The boundary value with the highest purity is selected as the final boundary value,* $\overline{v}_2^U = 1.8cm$ *with maximum purity* $Pur_{\widetilde{A}_{\rho(2)}} = 0.9555$. $\underline{v}_2^L = -\infty$ *because of* $\underline{v}_2 = -\infty$*. Therefore,* $\widetilde{A}_{\rho(2)} = (-\infty,-\infty,1.7cm,1.8cm)$.

For the third antecedent petal width is $(-\infty,5.0cm]$ *of crisp rule R, there are* 46 *patterns covered by* $A_{\rho(1)}$ *and* $A_{\rho(2)}$*. Within the* 46 *patterns, the petal width values of the patterns that are greater than* $\overline{v}_3 = 5.0cm$ *are considered as boundary values* $\overline{v}_3^U$*, and their purity are calculated. The boundary value with the highest purity is selected as the final boundary value,* $\overline{v}_3^U = 5.1cm$ *with maximum purity* $Pur_{\widetilde{A}_{\rho(2)}} = 0.9555$. $\underline{v}_3^L = -\infty$ *because of* $\underline{v}_3 = -\infty$*. Therefore,* $\widetilde{A}_{\rho(3)} = (-\infty,-\infty,5.0cm,5.1cm)$.

The previous rule R became a fuzzy rule $\widetilde{R}$ *after being fuzzified, as shown in Fig. 2.*

$\widetilde{R}$*: If petal length is* $(1,9cm,3.0cm,5.0cm,5.1cm)$ *and petal width is* $(-\infty,-\infty,1.7cm,1.8cm)$ *then specie is versicolor.*

In comparison to the conventional fuzzy rules based on triangular fuzzy linguistic numbers, the approach of fuzzifying crisp rules into fuzzy rules allows for a better fitness of the fuzzy rules to the training patterns and leave room for flexibility at the edges. Each of four parameters of the trapezoidal fuzzy number antecedent is designed to maximize the coverage of positive patterns by the rules. As depicted in Fig. 2, the pure region of intersection between the two crisp intervals $[\underline{v}_1,\overline{v}_2]$ *and* $[-\infty,\overline{v}_3]$ *covers only the patterns belonging to versicolor (the deepest-colored region). This implies that any patterns appearing within this area must necessarily belong to versicolor. The fuzzy (boundary) regions (the lighter-colored areas) signifies that patterns appearing within this region might potentially belong to versicolor.*

Step 3: **Convert the fuzzy rules to belief rules.**

We can see that both crisp and fuzzy rules have a singular consequent class. Nevertheless, in practice, the patterns covered by these rules may not all belong to the same class, and a few patterns from other classes may be included, resulting in uncertainty in the information in the consequent part. Representing this uncertainty with a singular consequent class would overlook a few patterns from other classes, resulting in information loss. Therefore, we use belief degrees framework to describe this type of uncertain information, aiming to mitigate information loss. The idea is to extend the consequent part of a fuzzy rule by computing the belief degree for each class, resulting in a belief rule as Definition 15.

Definition 16. *For a fuzzy rule $\widetilde{R}$, the belief degree for consequent class θ_k, $k = 1, 2, \ldots, c$ is defined as follows.*

$$\beta_k = \frac{\sum_{x_i \in X_{\theta_k}} \phi_{\widetilde{R}(x_i)}}{\sum_{x_i \in X} \phi_{\widetilde{R}}(x_i)}, \tag{21}$$

where X_{θ_k} is the pattern set that belongs to class θ_k.

Performing this operation on each rule in the fuzzy ruleset $\widetilde{RS}$, we obtain a set of belief rules, $\hat{RS}$. Afterward, these rules are used to determine the class labels of the patterns to be classified in the inference engine.

Example 3. *Assign the belief degrees for classes as the consequent part of the fuzzy rule $\widetilde{R}$ obtained from Example 2.*

$$\beta_1 = \frac{\sum_{x_i \in X_{\theta_1}} \phi_{\widetilde{R}(x_i)}}{\sum_{x_i \in X} \phi_{\widetilde{R}}(x_i)} = \frac{0.0}{45} = 0.0;$$

$$\beta_2 = \frac{\sum_{x_i \in X_{\theta_2}} \phi_{\widetilde{R}(x_i)}}{\sum_{x_i \in X} \phi_{\widetilde{R}}(x_i)} = \frac{43}{45} = 0.9555;$$

$$\beta_3 = \frac{\sum_{x_i \in X_{\theta_3}} \phi_{\widetilde{R}(x_i)}}{\sum_{x_i \in X} \phi_{\widetilde{R}}(x_i)} = \frac{2}{45} = 0.0445.$$

After assigning belief degrees on the consequent part, the previous fuzzy rule $\widetilde{R}$ becomes a belief rule $\hat{R}$: $\hat{R}$: If petal length is $(1, 9cm, 3.0cm, 5.0cm, 5.1cm)$ and petal width is $(-\infty, -\infty, 1.7cm, 1.8cm)$ then specie is $\{(setosa, 0.0), (versicolor, 0.9555), (virginica, 0.0445), (\Theta, 0)\}$.

3.3. New pattern classification

Identifying the rules that cover a new pattern is the first step towards determining its class label. Next, the ER approach is employed to combine the BPAs of the consequent parts of these rules to obtain a BPA about the class of the new pattern. Suppose that L belief rules have been learned, $\hat{RS} = \{\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_L\}$ and $\hat{R}_l$, $l = 1, 2, \ldots, L$ is:

$$\begin{aligned} &\text{Rule } \hat{R}_l\text{: If } F_{\rho_l(1)} \text{ is } \widetilde{A}_{\rho_l(1)} \text{ and } F_{\rho_l(2)} \text{ is } \widetilde{A}_{\rho_l(2)} \text{ and } \cdots \text{ and } F_{\rho_l(M)} \text{ is } \widetilde{A}_{\rho_l(M)} \\ &\text{then the consequent is } C = \{(\theta_1, \beta_{l1}), \ldots, (\theta_c, \beta_{lc}), (\Theta, \beta_{l\Theta})\}. \end{aligned} \tag{22}$$

We can observe that the BPA in the consequent part of the rule $\hat{R}_l$ primarily consists of singleton sets and the full set with their belief degrees, which aligns with the context of ER approach. ER approach can address the counter-intuitive problem associated with Dempster's rule. A new pattern is denoted as $x = \{x_1, x_2, \ldots, x_d\}$. If a belief rule $\hat{R}_l$, where $l = 1, 2, \ldots, L$, covers this pattern with a value greater than zero, then this rule can be considered activated. The coverage can be considered as the activation weight w_l of $\hat{R}_l$. The activation weight of the l-th rule, which measures its degree of activation and importance, can be calculated using the following formula.

$$w_l = \prod_{j=1}^{m} \mu_{\widetilde{A}_{\rho_l(m)}}(x_{i\rho_l(m)}). \tag{23}$$

Next, the activated belief rules' consequent BPAs are combined using the ER approach. First, the consequent BPA is weighted as follows [36]:

$$\begin{aligned} \beta^w_{lk} &= w_l \beta_{lk}, \\ \beta^w_{l\Theta} &= w_l \beta_{l\Theta}, \\ \beta^w_{l(P(\Theta))} &= 1 - w_l, \end{aligned} \tag{24}$$

Then, the weighted consequent BPAs of L activated belief rules are combined in turn. The joint probability assessment of the weighted BPAs of the first and second belief rules, denoted as $m'_{12}(\cdot)$, can be got as follows [36]:

$$\begin{aligned} m'_{12}(\theta_k) &= K(\beta^w_{1k}\beta^w_{2k} + \beta^w_{1k}(\beta^w_{2\Theta} + \beta^w_{2P(\Theta)}) + \beta^w_{2k}(\beta^w_{1\Theta} + \beta^w_{1P(\Theta)})), k = 1, 2, \ldots, c, \\ m'_{12}(\Theta) &= K(\beta^w_{1\Theta}\beta^w_{2\Theta} + \beta^w_{1\Theta}\beta^w_{2P(\Theta)} + \beta^w_{2\Theta}\beta^w_{1P(\Theta)}), \\ m'_{12}(p(\Theta)) &= K(\beta^w_{1P(\Theta)}\beta^w_{1P(\Theta)}), \end{aligned} \tag{25}$$

where $K = (1 - \sum_{k=1}^{c} \sum_{t=1, t\neq k}^{c} \beta_{1k}^{w} \beta_{2t}^{w})$. To calculate the final joint probability mass, we use the same method to combine $m'_{12}(\cdot)$ and the weighted consequent BPA of the third belief rule into joint probability mass m'_{13}. Then, we combine them further to obtain the final joint probability mass m'_{1L}. After calculating $m'_{1L}(P(\Theta))$, we can reassign it back to all focal elements using the following procedure [36].

$$\begin{aligned} m_{1L}(\theta_k) &= \frac{m_{1L}(\theta_k)}{1 - m'_{1L}(P(\Theta))}, k = 1, 2, \ldots, c, \\ m_{1L}(\Theta) &= \frac{m_{1L}(\Theta)}{1 - m'_{1L}(P(\Theta))}. \end{aligned} \tag{26}$$

After calculating the BPA m_{1L} for a given pattern, we can determine its class label y^* using either the class label with the maximum lower bound probability $Bel_{1L}(y^*)$, upper bound probability $Pl(y^*)$ or belief degree $m_{1L}(y^*)$.

$$\begin{aligned} y^* &= argmax_{\theta_k \in \Theta} m_{1L}(\theta_k) \\ &= argmax_{\theta_k \in \Theta} Bel_{1L}(\theta_k) \\ &= argmax_{\theta_k \in \Theta} Pl_{1L}(\theta_k). \end{aligned} \tag{27}$$

Due to the fact that only the singletons, $\theta_1, \theta_2, \ldots, \theta_c$, and the universal set, Θ, possess belief degrees, the results of these three approaches are identical.

Algorithm 4 New patterns classification.

Input: The testing dataset $T = \{x_1, x_2, \ldots, x_N\}$; the set of belief rules $\hat{RS} = \{\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_L\}$.
Output: The class labels of testing patterns $Y_T = \{y_1, y_2, \ldots, y_N\}$, $y_i \in \Theta$, $i = 1, 2, \ldots, N$.

```
Y_T ← null
for i = 1 to N do
    m ← {(θ_1, 0), (θ_2, 0), ..., (θ_c, 0), (Θ, 1)}.
    for l = 1 to L do
        Compute activation weight w_l of belief rule R̂_l using Eq. (23).
        if w_l > 0 then
            Weight the consequent BPA of belief rule R̂_l using Eq. (24) to obtain the weighted BPA m_l^w(·).
            m ← combination of m and m_l^w(·) using Eq. (25).
        end if
    end for
    Reassign the joint assessment m(·) to a BPA m_i(·) using Eq. (26).
    Y_T adds y* with maximum belief degree in m_i(·).
end for
```

To illustrate the classification process, we can refer to Algorithm 4, which outlines the steps involved in determining the class label of a given pattern.

3.4. Time complexity analysis

The time complexity of the proposed BRB classification system can be analyzed by considering the time complexity of modified RIPPER algorithm, fuzzification and belief degree assignment. For training dataset X with d features, the time complexity of modified RIPPER algorithm is $O(|X|(\log|X|)^2)$, with $|X|$ the number of patterns in X. The time complexity of fuzzification procedure is $O(|X|d^2)$, with d the number of features [19]. The time complexity of belief degree assessment is $O(|X|)$. Therefore, the overall time complexity is $O(\max(|X|(\log|X|)^2, |X|d^2, |X|)) = O((|X|\log|X|)^2)$.

4. Performance evaluation

This section conducts two experiments to evaluate the effectiveness and superiority of the proposed BRB classification system. The first one compares the performance of the proposed method with five traditional classifiers,

including C45 decision tree [28], SVM [11], KNN [12], backpropagation neural network [29] and hybrid decision tree/genetic algorithm [5]. The second one compares the method with seven common rule-based classification systems, including FURIA [19], EBRB [23] and FRB [10] classification systems, fuzzy rule-based decision tree [31], the steady-state genetic algorithm [25], ensemble belief rule-based model [38] and parallel multipopulation optimization for BRB system [41] in terms of classification accuracy and running time.

Table 1: The description of used datasets.

Dataset	# Patterns	# Features	# Classes	Dataset	# Patterns	# Features	# Classes
Appendicitis	106	7	2	Magic	19020	10	2
Balance	625	4	3	Mfeat-zernike	2000	47	2
Banknote	1372	4	2	New-thyroid	215	5	3
Bupa	345	6	2	Nursery	12690	8	5
Car	1728	6	4	Occupancy	2665	5	2
Column3C	310	6	3	Optdigits	5620	64	2
Ecoli	336	7	8	Pima	768	8	2
German	1000	20	2	QSAR	1055	4	2
Haberman	306	3	2	Rice-OC	3810	7	2
Hayes-roth	160	4	3	Seed	210	7	3
Heart-statlog	270	13	2	Sonar	208	60	2
Housevotes	232	16	2	Spambase	4601	57	2
HTRU_2	17898	8	2	Tic-tac-toe	958	9	2
Ilpd	583	10	2	Titanic	2201	3	2
Iris	150	4	3	Transfusion	748	4	2
Liver-disorders	345	6	2	Vehicle	846	18	4

Table 2: Parameters setting of five traditional classifiers.

Algorithm	Parameters	
C45	Confidence: 0.25	The minimum patterns per leaf: 2
SVM	Kernel type: radial basis function	
	C: 100	Epsilon: 0.001
	Degree: 1	Gamma: 0.01
	Coef0: 0.0	nu: 0.1
	p:1.0	Shrinking: 1.0
BP	Number of hidden layers: 2	Cctivation function: Htan
	Number of neurons in each layer: 15	Number of epochs: 1000
DTGA	Confidence: 0.25	Number of total generations: 50
	The minimum patterns per leaf: 2	Number of chromosomes:200
	Genetic algorithm: GA-LARGE-SN	Crossover probability: 0.8
	Threshold: 10	Mutation probability: 0.01
KNN	K: 1	Distance function: Euclidean

4.1. Experimental set-up

The proposed method is evaluated through two experiments using 32 benchmark datasets with variations in the number of classes, features, and patterns. The description of these datasets is provided in Table 1, where '# Patterns' represents the number of patterns, '# Features' represents the number of features, and '# Classes' represents the number of classes.

To evaluate the generalization ability of the proposed method, we employ the 10-fold cross-validation (10CV) technique. In this method, the datasets are divided into ten folds, each containing 10% of the patterns. Nine of these

folds are used for training the model, while the remaining one is used for testing it. The classification performance of the model is evaluated based on the average accuracy of the ten folds. The number of optimization is 2 and number of folds is 3 in the proposed method for all datasets.

4.2. Comparison with traditional classifiers

In this section, we select five traditional classifiers, including C45 decision tree (C45) [28], support vector machine (SVM) [11], backpropagation (BP) neural network [29], hybrid decision tree/genetic algorithm (DTGA)[5], and k-nearest neighbors (KNN) [12] classifiers, to compare their classification performance with the proposed method. Table 2 shows the key parameter settings for each of these classifiers. Although tuning the parameters of each method for each specific problem could potentially lead to better results, we prefer to use the default parameter values for each method. This is because a classifier's generalization ability is also an important factor to consider. A classifier with good generalization capability should be able to perform well on different datasets. All the algorithms are implemented on the KEEL software platform [1].

Table 3: Classification accuracy and F1 score rates (%) of six classifiers on 32 datasets.

	Accuracy						F1 Score					
Dataset	C45	SVM	BP	DTGA	KNN	Our method	C45	SVM	BP	DTGA	KNN	Our method
Appendicitis	84.09	**86.91**	73.73	84.09	81.91	84.82	65.48	70.03	69.11	65.48	**71.74**	67.80
Balance	78.07	**95.68**	68.32	75.66	78.24	83.05	55.19	**90.98**	47.75	53.32	56.74	58.95
Banknote	98.40	**100.00**	93.66	98.10	99.85	99.13	98.37	**100.00**	93.61	98.07	99.85	99.11
Bupa	66.97	63.50	54.82	64.32	63.14	**71.31**	64.81	60.56	54.35	62.88	61.94	**69.62**
Car	92.88	95.02	50.26	79.79	86.39	**95.48**	84.35	**91.46**	31.75	41.12	63.99	83.95
Column3C	**81.61**	76.13	68.39	79.03	76.77	**81.61**	74.01	69.70	59.34	71.04	72.70	**75.46**
Ecoli	79.47	61.96	42.70	77.09	80.70	**80.99**	68.05	37.23	21.80	62.46	65.65	**70.91**
German	70.70	65.80	61.90	69.60	69.30	**73.40**	61.97	46.64	54.26	57.84	62.36	**63.69**
Haberman	70.91	69.94	70.24	70.91	67.63	**72.55**	57.96	54.53	**61.79**	57.96	57.40	61.14
Hayes-roth	80.00	78.75	45.63	74.38	37.50	**80.63**	**83.30**	80.20	37.91	76.22	42.77	81.98
Heart-statlog	75.19	59.63	80.37	77.41	75.56	**81.11**	74.62	55.30	79.92	76.46	74.96	**80.59**
Housevotes	96.11	96.11	90.02	96.11	91.32	**96.56**	96.08	96.08	90.00	96.08	91.25	**96.54**
HTRU_2	**97.94**	96.64	86.81	97.90	97.14	97.88	**93.56**	89.41	74.17	93.39	91.37	93.37
Ilpd	67.07	**73.07**	62.77	68.59	63.45	69.97	56.51	**61.38**	54.69	58.77	56.94	51.35
Iris	96.00	**97.33**	61.33	96.00	93.33	94.00	96.01	**97.31**	57.40	96.01	93.31	93.95
Liver-disorders	66.97	63.50	54.82	64.32	63.14	**71.31**	64.81	60.56	54.35	62.88	61.94	**69.62**
Magic	**85.08**	68.03	70.38	84.98	80.78	84.72	**82.97**	57.70	68.83	82.61	78.40	82.38
Mfeat-zernike	98.35	90.00	49.90	98.35	**99.40**	98.95	95.43	47.37	43.39	95.45	**98.33**	97.00
New-thyroid	93.98	93.96	81.34	91.65	**97.21**	94.46	91.30	91.23	70.17	88.24	**96.20**	92.23
Nursery	96.98	93.20	28.21	94.07	82.08	**98.47**	**87.64**	84.80	21.17	83.55	73.27	76.45
Occupancy	98.24	97.90	97.94	98.31	**98.72**	98.54	98.11	97.73	97.80	98.19	**98.62**	98.43
Optdigits	97.74	93.29	38.36	97.72	89.82	**98.20**	93.67	73.43	33.49	93.69	47.32	**95.22**
Pima	**75.40**	63.80	70.33	74.61	69.78	75.13	**72.78**	50.48	68.84	72.12	66.28	71.06
QSAR	83.51	**86.26**	69.66	82.00	83.12	85.50	81.40	**84.73**	66.44	80.22	81.47	83.26
Rice-OC	92.36	68.82	83.25	92.36	88.79	**92.73**	92.20	62.58	83.14	92.20	88.56	**92.57**
Seed	92.38	91.43	80.95	89.52	**94.29**	92.86	92.36	91.39	79.31	89.47	**94.21**	92.82
Sonar	73.86	80.64	64.24	75.31	**86.93**	81.17	73.40	80.08	63.93	75.21	**86.66**	81.10
Spambase	92.57	86.57	74.48	91.74	90.11	**93.65**	92.21	86.09	62.86	91.42	89.64	**93.33**
Tic-tac-toe	84.55	81.63	60.23	77.98	73.07	**99.17**	82.13	77.42	58.96	71.34	59.46	**99.06**
Titanic	**79.06**	78.24	76.06	77.83	60.16	78.51	70.06	**71.99**	70.11	70.85	55.83	69.05
Transfusion	78.61	71.25	69.78	78.47	61.64	**79.41**	66.95	55.13	63.91	66.27	53.87	**67.82**
Vehicle	**75.06**	52.14	38.31	68.80	69.62	71.16	**74.69**	55.12	32.12	68.16	69.32	69.11
Mean	84.38	80.53	66.22	82.72	79.72	**86.14**	79.45	72.77	60.21	76.53	73.82	**80.59**

Table 3 presents the classification accuracy and F1 score rates of six classifiers on each dataset. The best result for each dataset is highlighted in boldface, and the last row displays the average values of all datasets for each classifier. It can be observed that the proposed BRB classification system achieves the best accuracy on 16 out of 32 datasets and the best F1 score on 12 out of 32 datasets, respectively. It also obtains the best average accuracy and average F1 score over all datasets. The proposed method demonstrates good classification performance on the majority of datasets; however, it may exhibit excessive flexibility leading to overfitting on certain datasets, such as Balance. To

Table 4: Wilcoxon signed-rank tests of the accuracy and F1 score for comparing our method to five traditional classifiers.

	Classifier	Statistic value	p-value	Critical value	Hypothesis
Accuracy	C45	63	0.00028	159	Rejected
	SVM	82	0.00036	159	Rejected
	BP	0	4.65661e-10	159	Rejected
	DTGA	20	1.72760e-07	159	Rejected
	KNN	48	1.11940e-05	159	Rejected
F1 score	C45	159	0.04976	159	Rejected
	SVM	123	0.00733	159	Rejected
	BP	18	1.17812e-07	159	Rejected
	DTGA	94	0.00098	159	Rejected
	KNN	95	0.00105	159	Rejected

Table 5: Running time (milliseconds) of six classifiers on 32 datasets.

Dataset	C45	SVM	BP	DTGA	KNN	Our method
Appendicitis	45.2	30.1	193.3	58.2	37.8	94.9
Balance	74.9	73	202.3	57.7	53.2	303.3
Banknote	97.9	66.6	213.1	549.5	83.7	336.7
Bupa	60	68.1	223.3	505.1	48.4	217.8
Car	76.4	281.8	318.9	219.6	98	1082
Column3C	59.6	65.2	216	212	50.1	257.8
Ecoli	67.6	90.3	228.2	198.9	52.8	295.2
German	118.2	251.7	329.9	211.3	97.2	575.5
Haberman	45.1	70.8	177.4	1335.9	42.6	181.5
Hayes-roth	45.1	39.6	188.8	1239	40	140.6
Heart-statlog	65.1	61	227.8	218.5	50.2	240.9
Housevotes	48.5	40.9	235.5	259.2	47	131.3
HTRU_2	4507.2	28254.4	620.5	506.6	2683.7	22588.7
Ilpd	129.4	133	234.3	496.6	87.5	367.5
Iris	56.9	34	210.6	427.2	54.5	113.5
Liver-disorders	82.5	74	217	431.2	69.8	227.2
Magic	9073.8	235534.2	610.2	67.3	2791.7	47012.7
Mfeat-zernike	679.8	1745.4	406.9	61.6	327.7	1372.1
New-thyroid	49.8	48.7	210.5	108.3	46.1	116
Nursery	365.6	14823.4	552.2	129.3	1368.5	80088
Occupancy	134.5	543.1	265.3	196.9	185.5	597.1
Optdigits	4158.3	33767.9	580.2	207.4	1362.2	4708.8
Pima	91.2	223.3	251.4	66.5	84.1	499
QSAR	358.8	314.1	312.2	66.5	174.3	1379.2
Rice-OC	321.4	6242.6	322.2	7362.2	306.2	1915
Seed	55.5	38.4	229.6	8839.7	60.6	155.8
Sonar	135	68.2	328.4	292.7	99.9	422.5
Spambase	5852.1	7168	515.6	265.9	984.5	7084.1
Tic-tac-toe	72.6	171	278.8	60.4	103	254
Titanic	123.9	438.6	245.6	88.8	143.3	284.9
Transfusion	63.1	118.8	225.7	202.7	76.4	237.7
Vehicle	170.3	250.9	316.4	224	108.9	798.4
Mean	874.7	10673.6	302.3	804.6	377.8	5589.7

further detect significant differences between the proposed method and other classifiers, we conducted the Wilcoxon signed-rank test on the performance of the classifiers on 32 datasets. This nonparametric statistical test was utilized to determine whether the proposed method outperformed other classifiers significantly on the given datasets. The test results are presented in Table 4. The test aims to determine whether there is a significant difference in performance between two classifiers being compared. The null hypothesis is that there is no significant difference between the two classifiers. The test statistic is the sum of the ranks of the positive or negative differences, depending on which is smaller. The p-value is then calculated by comparing the test statistic to a critical value from a reference table. If the p-value is below the significance level (0.05 here), it implies rejecting the null hypothesis and indicates a statistically significant distinction in performance between the two models.

As shown in Table 4, all null hypotheses are rejected, demonstrating that the proposed method statistically outperformed five traditional classifiers: C45, SVM, BP neural network, DTGA, and KNN, on all 32 datasets.

Table 5 lists the average running time, including the training time and testing time, of each classifier shown in Table 2 over 10CV. It is evident that our method runs faster than SVM.

4.3. Comparison with rule-based classifiers

In this section, we compare the proposed BRB classification system with seven other rule-based classification systems: the EBRB classification system [23], FURIA [19], FRB classification system [10], fuzzy rule-based decision tree (FRDT) [31], steady-state genetic algorithm for extracting fuzzy classification rules from data (SGERD) [25], ensemble belief rule-based model (En-BRB) [38], and parallel multipopulation optimization for BRB (PMP-BRB) [41]. We evaluate these systems based on classification accuracy and running time. Table 6 presents some key parameter settings for each rule-based classification system. All the algorithms are implemented on the KEEL software platform [1].

Table 6: Parameters setting of seven rule-based systems.

Algorithm	Parameters	
EBRB	Number of utility values per feature: 3	Rule weight: Rule inconsistency
	Transformation method: utility-based transformation method	
FURIA	Number of optimization: 2	Number of folds: 3
FRB	Number of linguistic labels per feature: 3	T-norm: product
	Rule weight: penalized certainty factor	Fuzzy reasoning method: winning rule
FRDT	Maximum number of features: 5	
	Threshold: 0.6	α: 0.02
SGERD	Number of Q rules per class: 0	Rule evaluation criteria: 2
En-BRB	Number of reference values: 4	Number of antecedent attributes for each weak BRB: 2
	Number of weak BRB: 20	Optimization technique: DE
	Iteration number: 200	Population size: 50
	Mutation probability: 0.5	Cross probability: 0.8
PMP-BRB	Optimization technique: DE	Range of rules: [3,6]
	Iteration number: 500	Population size: 100
	Mutation probability: 0.8	Cross probability: 0.8

Table 7 presents the classification accuracy and F1 score rates obtained by each classifier for each dataset, with the best results highlighted in boldface. The last row of the table lists the average results across all datasets for each classifier. We can see that the proposed method achieved the best accuracy on 20 out of 32 datasets and the best F1 score on 16 out of 32 datasets. It also has the best average classification accuracy and the best average F1 score across all 32 datasets, indicating superior performance compared to other rule-based methods for most of the datasets. To statistically compare the results, we used the Wilcoxon signed-rank test for multiple comparisons between the proposed method and other methods. Table 8 presents the test results of the comparison between the proposed method and each rule-based system. As shown in the table, all null hypotheses were rejected, indicating that the proposed BRB classification system outperformed the EBRB classification system, FURIA, FRB classification system, FRDT, SGERD, En-BRB, and PMP-BRB on these 32 datasets.

Table 7: Classification accuracy and F1 score rates (%) of eight rule-based classification systems on 32 datasets.

	Accuracy								F1 Score							
Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	84.91	83.09	85.82	75.64	86.91	86.91	**87.73**	84.82	71.45	65.37	71.80	68.30	70.35	**73.27**	70.91	67.80
Balance	33.11	82.08	**90.39**	46.08	75.66	85.92	78.41	83.05	27.30	58.51	49.22	21.03	46.32	**59.64**	54.46	58.95
Banknote	98.62	98.91	94.97	87.90	83.75	93.44	**99.13**	**99.13**	98.60	98.89	94.93	87.81	80.56	93.34	99.11	**99.11**
Bupa	61.18	65.81	58.54	63.50	58.84	63.19	62.39	**71.31**	50.95	63.48	37.64	63.31	44.47	51.28	58.07	**69.62**
Car	65.90	94.27	78.00	72.04	58.01	70.01	80.89	**95.48**	45.47	82.68	54.81	40.75	16.14	20.59	39.60	**83.95**
Column3C	68.71	**82.58**	59.03	81.29	68.39	77.10	77.10	81.61	58.08	75.15	44.99	**75.72**	61.90	65.06	66.44	75.46
Ecoli	69.33	79.77	72.02	61.30	74.69	63.73	68.78	**80.99**	55.76	69.22	60.10	44.38	61.67	34.71	38.37	**70.91**
German	64.90	73.20	18.80	70.50	65.30	70.00	72.40	**73.40**	58.50	62.06	16.54	**67.54**	28.90	41.18	58.39	63.69
Haberman	71.90	72.23	73.87	62.56	71.92	72.87	**74.19**	72.55	45.62	58.74	44.26	53.52	46.30	46.82	58.26	**61.14**
Hayes-roth	63.13	**80.63**	58.75	49.38	45.00	59.38	68.75	**80.63**	61.19	**82.05**	41.77	43.15	30.61	61.13	59.62	81.98
Heart-statlog	72.22	80.37	52.59	**83.70**	75.19	80.37	77.78	81.11	71.23	79.61	42.01	**83.59**	60.57	79.04	76.14	80.59
Housevotes	79.69	96.54	41.36	**96.97**	87.88	89.20	95.24	96.56	79.35	96.53	37.91	**96.97**	61.34	89.07	95.20	96.54
HTRU_2	96.26	97.85	95.68	97.50	95.75	97.26	97.68	**97.88**	86.37	93.21	83.65	92.36	81.25	90.77	92.63	**93.37**
Ilpd	68.27	68.96	70.32	55.39	71.01	**71.19**	69.29	69.97	45.88	52.24	38.56	**55.25**	40.18	42.15	44.48	51.35
Iris	95.33	95.33	92.67	94.67	95.33	**97.33**	94.00	94.00	95.30	95.30	92.47	94.63	95.30	**97.32**	93.97	93.95
Liver-disorders	61.18	65.81	58.54	63.50	58.84	64.95	64.24	**71.31**	50.95	63.48	37.64	63.31	44.47	53.93	59.03	**69.62**
Magic	76.17	84.68	76.78	74.62	71.30	73.73	81.88	**84.72**	69.16	82.15	69.68	72.21	59.76	62.63	76.32	**82.38**
Mfeat-zernike	**99.15**	98.80	98.80	96.00	97.55	90.50	89.05	98.95	**97.61**	96.48	90.36	90.55	92.46	51.93	32.93	97.00
New-thyroid	82.36	94.46	84.24	93.53	87.03	**94.94**	91.65	94.46	63.74	91.94	66.24	92.18	73.87	91.58	85.82	**92.23**
Nursery	75.22	97.05	84.49	53.24	64.40	73.75	69.99	**98.47**	63.71	**88.13**	62.51	36.59	41.01	53.83	49.94	76.45
Occupancy	93.85	98.27	93.09	97.86	96.96	93.17	97.90	**98.54**	93.32	98.15	92.54	97.72	87.15	92.56	97.75	**98.43**
Optdigits	**98.29**	98.75	45.41	56.32	89.79	89.82	90.20	98.20	73.52	**96.48**	39.02	49.74	47.31	47.32	42.39	95.22
Pima	72.13	74.86	73.17	73.05	70.70	69.66	73.69	**75.13**	63.75	**71.16**	60.01	70.83	55.58	53.64	68.92	71.06
QSAR	83.80	84.27	78.39	53.09	68.34	70.51	75.36	**85.50**	82.24	81.49	52.08	51.93	40.38	52.54	60.42	**83.26**
Rice-OC	91.97	92.57	88.19	92.62	92.20	91.73	92.65	**92.73**	91.70	92.39	87.58	92.46	86.00	91.48	92.48	**92.57**
Seed	90.95	92.38	91.90	90.95	86.19	**93.33**	90.48	92.86	90.88	92.34	91.86	90.75	86.09	**93.27**	90.19	92.82
Sonar	73.52	79.71	61.48	74.90	71.52	74.43	39.19	**81.17**	53.94	79.52	47.43	74.69	69.29	73.06	25.17	**81.10**
Spambase	73.68	93.24	71.81	85.26	69.42	67.64	62.20	**93.65**	68.64	92.86	43.16	83.85	41.06	54.03	35.46	**93.33**
Tic-tac-toe	98.33	98.54	0.00	70.78	69.94	65.87	73.18	**99.17**	98.13	98.37	0.00	69.29	67.00	41.33	63.07	**99.06**
Titanic	77.56	78.33	78.28	75.37	71.97	77.92	78.46	**78.51**	66.97	68.66	70.04	**71.24**	43.52	69.66	69.72	69.05
Transfusion	75.94	78.74	76.61	71.52	76.34	77.27	78.48	**79.41**	44.14	**67.88**	45.42	64.14	46.23	48.95	64.71	67.82
Vehicle	68.45	70.58	61.59	48.70	53.44	57.33	51.79	**71.16**	66.74	68.28	58.48	44.21	50.87	52.85	45.49	**69.11**
Mean	77.69	85.40	70.80	74.05	75.30	78.26	78.25	**86.14**	68.44	80.09	57.02	68.88	58.06	63.44	64.55	**80.59**

Table 8: Wilcoxon signed-rank tests of the accuracy and F1 score rates for comparing our method to seven rule-based classification systems.

	Classifier	Statistic value	p-value	Critical value	Hypothesis
Accuracy	EBRB	15	6.37955e-08	159	Rejected
	FURIA	58.5	0.00034	159	Rejected
	FRB	29	8.09784e-07	159	Rejected
	FRDT	19	1.42958e-07	159	Rejected
	SGERD	14	5.12227e-08	159	Rejected
	En-BRB	46	8.74791e-06	159	Rejected
	PMP-BRB	31	1.10594e-06	159	Rejected
F1 Score	EBRB	18	1.17812e-07	159	Rejected
	FURIA	104	0.00207	159	Rejected
	FRB	6	6.51925e-09	159	Rejected
	FRDT	74	0.00017	159	Rejected
	SGERD	3	2.32830e-09	159	Rejected
	En-BRB	24	3.54833e-07	159	Rejected
	PMP-BRB	17	9.63918e-08	159	Rejected

Table 9 displays the average running time, including the training and testing time, for each classifier shown in Table 4 over 10CV. Our method demonstrates good performance in terms of running time, with an average running time of 5440.0 milliseconds across the 32 datasets. Among these classifiers, the proposed method stands at a moderate performance level, significantly outperforming classifiers with longer running time (such as EBRB, En-BRB and PMP-BRB), while also exhibiting comparability with classifiers that have shorter running time (such as FURIA).

Besides accuracy, F1 score and running time, interpretability is also a crucial criterion for rule-based systems. Ref. [18] provides a detailed survey of interpretability measures for rule-based systems. In this experiment, we adopt three commonly used interpretability measures suggested in Ref. [18] to perform a quantitative analysis.

- Number of rules: The number of rules is an important interpretability measure for rule-based systems. This is based on the principle of Occam's razor, which states that the best model is the simplest one that fits the system

Table 9: Running time (milliseconds) of eight classifiers on 32 datasets.

Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	117.7	60.6	84.3	103.1	69.3	43401.6	13515.3	94.9
Balance	353.5	245.2	140.2	116.8	100.8	39506.5	12414.8	303.3
Banknote	1097.7	262	163.1	157.3	163.1	63239.8	20188.2	336.7
Bupa	333.4	158	108.8	120.6	92.4	67455.5	13121	217.8
Car	1584.4	889.5	637	266.9	205.6	39802.2	44905.2	1082
Column3C	380.2	154.6	138.6	149.6	125.3	59696.8	13264.6	257.8
Ecoli	352.9	155.7	121.9	158.5	97.4	54334.7	22522.4	295.2
German	2636.8	334.9	1652.2	239.4	246.7	47893.9	114454.6	575.5
Haberman	203.6	93.7	108.3	114.7	90.5	44382.1	7002.1	181.5
Hayes-roth	132.6	71.6	94.5	103.2	70.8	34095.5	6939	140.6
Heart-statlog	469.6	136.9	311.7	153.7	120.3	29973.9	40976.8	240.9
Housevotes	428.2	67.3	291.1	125.8	120.1	11779	51138.8	131.3
HTRU_2	137517.5	14844.3	7813.1	707.3	4142.6	66482.8	228874.5	22588.7
Ilpd	1215.4	278.1	212.5	117.6	126.1	37310.2	30279.2	367.5
Iris	242.1	68.3	126.4	84	84.1	50877.8	6989.5	113.5
Liver-disorders	548.4	165.8	148.6	88	95.3	58893.7	13936.1	227.2
Magic	181762	37410.5	15065.6	675.9	3857.6	66989.6	303463.3	47012.7
Mfeat-zernike	26758.3	1087.9	8351.4	329.2	661.5	78510.4	512651.9	1372.1
New-thyroid	250.5	80.5	82.6	86.1	89.9	73102.5	9851.3	116
Nursery	55048.3	68715	15440.8	1352.9	3080.7	25265.3	316697.7	80088
Occupancy	3829	476.7	284.7	157.3	318.5	17236.1	34671.6	597.1
Optdigits	180856.3	3178.4	101451	1214	2194.8	31604.6	1456326.2	4708.8
Pima	949.5	372.6	265.2	154.7	192.5	48019.8	26997	499
QSAR	6174.4	996.1	1233.1	292	357.7	56135.1	424913.9	1379.2
Rice-OC	8376.3	1377.6	825.7	335.7	550.2	27048.6	62251.1	1915
Seed	239.1	95.8	117.4	86.6	91	72200.3	14518.8	155.8
Sonar	1593.9	294.7	773.5	142.9	229.8	84155.6	713764.8	422.5
Spambase	123357.5	4887.7	5412.9	760.1	1617.5	20230.1	989565.7	7084.1
Tic-tac-toe	940.2	202.1	736.8	144.5	179.5	33018.8	39015.6	254
Titanic	1323.6	228.4	199.9	165.6	234.1	26465.9	28303	284.9
Transfusion	546.5	186.1	97.3	94.5	117.5	46330.7	14520.3	237.7
Vehicle	2429.6	650.2	480.3	207.8	243	61014.8	90847.4	798.4
Mean	23189.0	4319.6	5092.8	281.4	623.9	47389.2	177465.1	5440.0

behavior well. In other words, the set of fuzzy rules should be as small as possible while still preserving the model's performance to a satisfactory level.

- Number of conditions: The number of conditions should be minimized to enhance the readability of the rules.
- Number of fired rules: To ensure semantic interpretability, it is desirable to minimize the number of rules used in the reasoning process for a given input.

Based on the three interpretability measures, we tested the six rule-based methods. We evaluated them using the thirty-two real-world problems shown in Table 1, and the 10-CV model was used to calculate the average results. Table 10 presents the results of the three interpretability measures for the different methods.

Based on the measure of the number of rules, the proposed BRB classification system is found to have higher interpretability than EBRB, FRB and ensemble BRB classification systems, but lower interpretability than FRDT and SGERD methods. The interpretability of the proposed BRB classification system and FURIA was found to be close based on the number of rules measure. According to the number of conditions measure, the proposed

Table 10: Interpretability measures for the eight rule-based systems.

	Numbers of rules								Numbers of conditions								Number of fired rules							
Dataset	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method	EBRB	FURIA	FRB	FRDT	SGERD	En-BRB	PMP-BRB	Our method
Appendicitis	95.4	3.3	28.3	4.7	2.6	180	3.4	4.3	7	1.4	7	4.5	6	2	7	1.5	95.4	1.3	1	1	1	77.2	1.2	1.5
Balance	562.5	24.3	66.2	3	4.2	180	4	23.1	4	3.1	4	0	3.2	2	4	3.1	562.5	2.6	1	1	1	51.3	2.1	2.5
Banknote	1234.8	12.5	30.8	6.9	3.4	180	3.9	12.5	4	2.3	4	1.2	3.4	2	4	2.4	1234.8	2.0	1	1	1	79.9	2.2	2.0
Bupa	310.5	9.8	43.2	9	3.3	180	3	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.6	1.2
Car	1554.3	83.5	690.5	27.5	1.6	180	3.3	78.7	6	4.3	6	3.6	6	2	6	4.3	1554.3	1.3	1	1	1	40.3	1.5	1.3
Column3C	279	7.8	26.8	10.8	3.5	180	3.1	8.1	6	2.5	6	1.9	5.3	2	6	2.6	279	1.4	1	1	1	79.5	1.9	1.4
Ecoli	302.4	18	63	29.7	8.9	180	3.9	17	7	2.9	7	2.0	1.3	2	7	2.9	302.4	1.9	1	1	1	58.0	1.6	1.8
German	900	9.4	885.4	22.5	2.4	180	3	10.1	20	2.3	20	3.1	19.8	2	20	2.3	900	1.4	1	1	1	41.7	1.2	1.4
Haberman	275.4	4	16.5	3.7	2.9	180	5.3	3.7	3	1.7	3	0.9	1.4	2	3	1.3	275.4	1.4	1	1	1	63.3	1.8	1.3
Hayes-roth	144	9.1	39.1	4.5	5.3	180	4.9	9.3	4	2.9	4	1.8	0.8	2	4	3.0	144	0.9	1	1	1	43.0	2.3	0.9
Heart-statlog	243	9.1	217.5	11.8	2.1	180	3	8.2	13	2.8	13	2.8	12	2	13	2.6	243	1.7	1	1	1	43.5	1.6	1.7
Housevotes	208.8	5.4	146.5	3	2	180	3	5.5	16	2.2	16	3	16	2	16	2.1	208.8	1.5	1	1	1	20	1.5	1.5
HTRU_2	16108.2	12.9	89.5	11.5	5.9	180	3.5	14.2	8	2.8	8	3.4	5.9	2	8	2.9	16108.2	3.2	1	1	1	80.0	1.1	3.7
Ilpd	524.7	9.4	86	10.9	2.1	180	3	9.6	10	2.6	10	4.0	9.6	2	10	2.6	524.7	1.5	1	1	1	71.2	1.1	1.7
Iris	135	4.3	14.7	5.3	3.9	180	5.3	4.7	4	1.8	4	1.6	2.9	2	4	1.9	135	1.2	1	1	1	78.0	2.4	1.3
Liver-disorders	310.5	9.8	43.2	9	3.3	180	4.1	9.3	6	2.6	6	3.1	5.6	2	6	2.5	310.5	1.2	1	1	1	79.2	1.8	1.2
Magic	17118	29.8	312.8	12.1	2.8	180	3.2	26.9	10	3.2	10	2.9	9.2	2	10	3.2	17118	3.2	1	1	1	80.0	1.8	2.7
Mfeat-zernike	1800	10.8	1679.9	7.4	2.2	180	3	10.9	47	2.6	47	4.3	45.8	2	47	2.5	1800	3.4	1	1	1	79.9	1.0	3.4
New-thyroid	193.5	6.6	18.4	12.2	3.3	180	4.7	7.1	5	2.5	5	2.4	3.3	2	5	2.4	193.5	1.9	1	1	1	79.0	1.8	2.0
Nursery	11664	274.1	4222.6	18.4	5	180	3	276.7	8	5.0	8	4.0	5	2	8	5.0	11664	1.5	1	1	1	36.8	1.8	1.5
Occupancy	2398.5	12.7	25	3.8	2.8	180	4.1	13.2	5	2.7	5	2.9	3.2	2	5	2.7	2398.5	2.0	1	1	1	70.8	1.6	1.8
Optdigits	5058	35.1	5055.6	15	2	180	3	34.7	64	4.0	64	4.9	64	2	64	4.0	5058	4.6	1	1	1	45.2	1.0	4.6
Pima	691.2	7.8	101.3	8.6	4.1	180	3.6	8.3	8	2.4	8	2.6	6.1	2	8	2.4	691.2	1.8	1	1	1	69.6	1.8	1.8
QSAR	949.5	20.1	436.4	15.4	2.8	180	3	22.1	41	3.4	41	4.6	39.2	2	41	3.4	949.5	2.1	1	1	1	48.0	1.2	2.2
Rice-OC	3429	5.8	95.8	7.8	3.3	180	3.6	4.4	7	1.9	7	3.6	5.7	2	7	1.8	3429	2.5	1	1	1	80.0	1.7	1.9
Seed	189	8.5	53.2	9.6	3.8	180	3.2	8.4	7	2.2	7	3.0	5	2	7	2.2	189	1.8	1	1	1	79.3	2.1	1.7
Sonar	186.3	11.2	186.3	7.6	2.8	180	3	11	60	2.6	60	4.4	60	2	60	2.6	186.3	1.7	1	1	1	78.6	1.0	1.7
Spambase	4140.9	33.3	356.2	7.8	2	180	3	34.9	57	4.2	57	5	56	2	57	4.3	4140.9	3.5	1	1	1	29.5	1.1	3.6
Tic-tac-toe	862.2	21.6	862.2	9.3	2	180	3	21	9	3.5	9	2.0	8	2	9	3.5	862.2	1.1	1	1	1	35.4	1.2	1.1
Titanic	1980.9	5.8	9.9	7.2	3	180	4.2	5.7	3	2.2	3	1.1	3	2	3	2.2	1980.9	2.1	1	1	1	26.1	2.1	2.1
Transfusion	673.2	5	12.2	4	2.7	180	4.7	5	4	1.8	4	1.3	3.7	2	4	1.8	673.2	1.5	1	1	1	71.2	1.8	1.6
Vehicle	761.4	23.6	227.6	22.6	6.2	180	3	24.1	18	3.3	18	3.3	15.4	2	18	3.4	761.4	1.3	1	1	1	79.0	1.7	1.2
Mean	2352.6	23.3	504.5	10.7	3.4	180.0	3.6	23.2	14.9	2.8	14.9	2.9	13.7	2.0	14.9	2.7	2352.6	1.9	1.0	1.0	1.0	61.7	1.6	1.9

method, along with FURIA, FRDT and ensemble BRB classification system, are found to have higher interpretability than EBRB, FRB, PMP-BRB classification system, and SGERD. Based on the number of fired rules measures, the interpretability of the FRB classification system, FRDT, and SGERD is found to be similar and higher than that of the proposed method. The interpretability of FURIA and PMP-BRB is comparable to the proposed method. However, the interpretability of EBRB and En-BRB is lower than that of the proposed method. Overall, the interpretability of the proposed method was found to be similar to that of FURIA and much higher than that of the EBRB, FRB, ensemble BRB classification systems.

5. Conclusions

This paper presented a novel BRB classification system that leverages FURIA. In this BRB system, belief rules and ER approach provides the ability of handling both nominal and numerical information under conditions of uncertainty and incompleteness in the modeling process. Meanwhile, FURIA provided a rule induction approach with less computational complexity and superior performance in classification tasks. Compared to the traditional BRB classification system, the proposed approach not only provided better classification performance but also enhanced interpretability. Furthermore, the proposed system is more effective and comprehensive than FURIA. We evaluated the new model's effectiveness and superiority by conducting two classification experiments on 32 benchmark datasets, comparing it to five traditional classifiers and seven common rule-based classification systems. The experiment results indicated that the proposed BRB classification system outperforms five traditional classifiers and seven other rule-based classification systems in terms of classification performance. In conclusion, the incorporation of FURIA can improve the performance, and reduce the computational complexity of BRB classification systems. It even has the potential to improve the interpretability of BRB systems.

The current belief rules only consider residual support and global ignorance while overlooking partial ignorance, thus not fully leveraging the advantages of evidence theory. In future work, we will continue to explore partial ignorance information within belief rules.

Acknowledgements

This work was supported by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Proyecto B-TIC-590-UGR20, by the China Scholarship Council (CSC), and by the project PID2019-103880RB-I00 funded by MCIN/AEI/10.13039/501100011033.

References

[1] Alcalá-Fdez, J., Sanchez, L., Garcia, S., del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., et al., 2009. Keel: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13, 307–318.

[2] Boström, H., 2004. Pruning and exclusion criteria for unordered incremental reduced error pruning. Proceedings of the ECML/PKDD Workshop on Advances in Inductive Rule Learning , 17–29.

[3] Calzada, A., Liu, J., Wang, H., Kashyap, A., 2014. A new dynamic rule activation method for extended belief rule-based systems. IEEE Transactions on knowledge and data engineering 27, 880–894.

[4] Cao, Y., Zhou, Z., Hu, C., He, W., Tang, S., 2020. On the interpretability of belief rule-based expert systems. IEEE Transactions on Fuzzy Systems 29, 3489–3503.

[5] Carvalho, D.R., Freitas, A.A., 2004. A hybrid decision tree/genetic algorithm method for data mining. Information Sciences 163, 13–35.

[6] Casalino, G., Castellano, G., Castiello, C., Pasquadibisceglie, V., Zaza, G., 2019. A fuzzy rule-based decision support system for cardiovascular risk assessment, in: Fuzzy Logic and Applications: 12th International Workshop, WILF 2018, Genoa, Italy, September 6–7, 2018, Revised Selected Papers, Springer. pp. 97–108.

[7] Chang, L., Zhang, L., Fu, C., Chen, Y.W., 2021. Transparent digital twin for output control using belief rule base. IEEE Transactions on cybernetics 52, 10364–10378.

[8] Chang, L., Zhou, Z., You, Y., Yang, L., Zhou, Z., 2016. Belief rule based expert system for classification problems with new rule activation and weight calculation procedures. Information Sciences 336, 75–91.

[9] Chang, L.L., Zhou, Z.J., Liao, H., Chen, Y.W., Tan, X., Herrera, F., 2019. Generic disjunctive belief-rule-base modeling, inferencing, and optimization. IEEE Transactions on Fuzzy Systems 27, 1866–1880.

[10] Chi, Z., Yan, H., Pham, T., 1996. Fuzzy algorithms: with applications to image processing and pattern recognition. volume 10. World Scientific.

[11] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20, 273–297.

[12] Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory 13, 21–27.

[13] Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. The annals of mathematical statistics , 325–339.

[14] Deng, Y., 2020a. Information volume of mass function. International Journal of Computers Communications & Control 15, 3983. doi:https://doi.org/10.15837/ijccc.2020.6.3983.

[15] Deng, Y., 2020b. Uncertainty measure in evidence theory. Science China Information Sciences 63, 210201.

[16] Deng, Y., 2022. Random permutation set. International Journal of Computers Communications & Control 17, 4542. doi:https://doi.org/10.15837/ijccc.2022.1.4542.

[17] Ding, J., Li, L., Peng, H., Zhang, Y., 2019. A rule-based cooperative merging strategy for connected and automated vehicles. IEEE Transactions on Intelligent Transportation Systems 21, 3436–3446.

[18] Gacto, M.J., Alcalá, R., Herrera, F., 2011. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences 181, 4340–4360.

[19] Hühn, J., Hüllermeier, E., 2009. Furia: an algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery 19, 293–319.

[20] Jiao, L., Pan, Q., Denoeux, T., Liang, Y., Feng, X., 2015. Belief rule-based classification system: Extension of frbcs in belief functions framework. Information Sciences 309, 26–49.

[21] Kong, G., Xu, D.L., Yang, J.B., Wang, T., Jiang, B., 2020. Evidential reasoning rule-based decision support system for predicting icu admission and in-hospital death of trauma. IEEE Transactions on Systems, Man, and Cybernetics: Systems 51, 7131–7142.

[22] Liu, D., Wang, S., Tomovic, M.M., Zhang, C., 2020. An evidence theory based model fusion method for degradation modeling and statistical analysis. Information Sciences 532, 33–60.

[23] Liu, J., Martinez, L., Calzada, A., Wang, H., 2013. A novel belief rule base representation, generation and its inference methodology. Knowledge-Based Systems 53, 129–141.

[24] Liu, Z.G., Liu, Y., Dezert, J., Cuzzolin, F., 2019. Evidence combination based on credal belief redistribution for pattern classification. IEEE Transactions on Fuzzy Systems 28, 618–631.

[25] Mansoori, E.G., Zolghadri, M.J., Katebi, S.D., 2008. Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data. IEEE Transactions on Fuzzy Systems 16, 1061–1071.

[26] Quinlan, J.R., 1990. Learning logical definitions from relations. Machine learning 5, 239–266.

[27] Quinlan, J.R., 1995. Mdl and categorical theories (continued), in: Machine Learning Proceedings 1995. Elsevier, pp. 464–470.

[28] Quinlan, J.R., 2014. C4. 5: programs for machine learning. Elsevier.

[29] Rojas, R., 2013. Neural networks: a systematic introduction. Springer Science & Business Media.

[30] Shafer, G., et al., 1976. A mathematical theory of evidence. volume 1. Princeton university press Princeton.

[31] Wang, X., Liu, X., Pedrycz, W., Zhang, L., 2015. Fuzzy rule based decision trees. Pattern Recognition 48, 50–59.

[32] Wang, Y.M., Yang, L.H., Fu, Y.G., Chang, L.L., Chin, K.S., 2016. Dynamic rule adjustment approach for optimizing belief rule-base expert system. Knowledge-Based Systems 96, 40–60.

[33] Woźniak, M., Połap, D., 2019. Intelligent home systems for ubiquitous user support by using neural networks and rule-based approach. IEEE Transactions on Industrial Informatics 16, 2651–2658.

[34] Yang, J.B., 2001. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. European journal of operational research 131, 31–61.

[35] Yang, J.B., Liu, J., Wang, J., Sii, H.S., Wang, H.W., 2006. Belief rule-base inference methodology using the evidential reasoning approach-rimer. IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans 36, 266–285.

[36] Yang, J.B., Singh, M.G., 1994. An evidential reasoning approach for multiple-attribute decision making with uncertainty. IEEE Transactions on systems, Man, and Cybernetics 24, 1–18.

[37] Yang, L.H., Liu, J., Wang, Y.M., Martínez, L., 2018. A micro-extended belief rule-based system for big data multiclass classification problems. IEEE Transactions on Systems, Man, and Cybernetics: Systems 51, 420–440.

[38] You, Y., Sun, J., Chen, Y.w., Niu, C., Jiang, J., 2021. Ensemble belief rule-based model for complex system classification and prediction. Expert Systems with Applications 164, 113952.
[39] Zhang, A., Gao, F., Yang, M., Bi, W., 2020. A new rule reduction and training method for extended belief rule base based on dbscan algorithm. International Journal of Approximate Reasoning 119, 20–39.
[40] Zhang, H., Deng, Y., 2020. Weighted belief function of sensor data fusion in engine fault diagnosis. Soft computing 24, 2329–2339.
[41] Zhu, W., Chang, L., Sun, J., Wu, G., Xu, X., Xu, X., 2021. Parallel multipopulation optimization for belief rule base learning. Information Sciences 556, 436–458.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: