Ensemble Enhanced Evidential *k*-NN classifier through rough set reducts

Asma Trabelsi[1,2], Zied Elouedi[1], and Eric Lefevre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC , Tunisia
trabelsyasma@gmail.com,zied.elouedi@gmx.fr
[2] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France
eric.lefevre@univ-artois.fr

Abstract. Data uncertainty is seen as one of the main issues of several real world applications that can affect the decision of experts. Several studies have been carried out, within the data mining and the pattern recognition fields, for processing the uncertainty that is associated to the classifier outputs. One solution consists of transforming classifier outputs into evidences within the framework of belief functions. To gain the best performance, ensemble systems with belief functions have been well studied for several years now. In this paper, we aim to construct an ensemble of the Evidential Editing *k*-Nearest Neighbors classifier (EE*k*-NN), which is an extension of the standard *k*-NN classifier for handling data with uncertain attribute values expressed within the belief function framework, through rough set reducts.

Keywords: Evidential Editing *k*-Nearest Neighbors classifier, rough set reducts, belief function theory, uncertain attributes, ensemble classifier

1 Introduction

A multiple classifier system, also referred to as a classifier ensemble, has been proven to be an effective and efficient way for solving complex classification problems and achieving high performance [17]. The construction of an ensemble of classifiers consists mainly on two distinct levels: the generation of a set of base classifiers and the combination of their output predictions. It should be emphasized that the process of improving ensemble accuracy requires the best choice of the base classifiers and also the combination operator. In this paper, we focus only on the generation of good base classifiers for enhancing accuracies. Ensuring diversity between the base classifiers has been defended as a successful means for the production of a good ensemble of base classifiers. Although diversity can be achieved in several ways, the manipulation of the input feature space has been theoretically and experimentally proven to be one of the best methods for establishing high diversity between base classifiers [2, 29, 31]. In fact, it does not only allow the correlation reduction between the combined classifiers, but it also performs faster thanks to the reduced size of the input feature space

[2, 5, 9, 30]. The process of generating feature subsets with good predicting power is still under study. One commonly used solution is the random subspace method (RSM) offtenly called random subspacing. The major shortcoming of this latter technique is the random partition of the original input features. As a matter of fact, the random selection may potentially increase the risk of irrelevant and redundant features as part of the selected subsets.

The rough set theory, introduced by Pawlak [15], has been successfully applied in pattern recognition, data mining and machine learning domains, more particularly for attribute reduction problems. The reduced attribute set, representing the minimal subset of attributes that enables the discernation of objects with different decision values, is referred to as reduct. Since there have been usually multiple reducts for a given data set, the concept of ensemble classifiers through rough set reducts have been introduced and applied in a range of practical problems such as text classification [20], biomedical classification [21], tumor classification [32], web services classification [19], etc. It is important to emphasize that several real world application data suffer from some kinds of uncertainty, imprecision and also incompleteness that mainly pervade the attribute values. However, to the best of our knowledge, there are no rough set techniques allowing to obtain the possible reducts from data with uncertain attribute values.

In this paper, we aim to develop a classifier ensemble through rough set reducts (RSR) for dealing with uncertain data. More precisely where the uncertainty exists in the attribute values and is represented within the belief function theory, a flexible way for managing and representing all kinds of uncertainty. We therefore propose a new method for generating approximate reducts from such a kind of data. Since tens or hundreds of reducts may be generated, a selected subset of these reducts have to be used for constructing the base classifiers, notably the most diverse ones.

Herein, we have used the Evidential k-Nearest neighbors [28], an extension of the well known k-NN to handle the uncertainty that occurs in the attribute values within the belief function framework, as base classifiers. Given a query instance, the output beliefs of the base evidential classifiers will then be merged through a combination operator that is offered by the belief function framework [23].

The remaining of this paper is organized as follows: Section 2 is dedicated to recall some basic concepts of the belief function theory. Section 3 is committed to highlighting the fundamental concepts of the rough set theory. We describe, in Section 4, our proposed idea for constructing an ensemble of classifiers via rough set reducts for handling uncertain data. Our conducted experimentation on several synthetic databases is presented in Section 5. Finally, the conclusion and our main future work directions are reported in Section 6.

2 Belief function theory: Fundamental concepts

The belief function theory has been shown to be a convenient way for representing, managing and reasoning under uncertainty. In this Section, we briefly recall some fundamental concepts underlying this theory.

2.1 Information representation

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ be a frame of discernment with a finite non empty set of N elementary hypotheses that are assumed to be exhaustive and mutually exclusive. An expert's belief over a given subset of Θ has to be represented by the so-called basic belief assignment m (bba) as follows:

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

Each subset A of 2^Θ having fulfilled $m(A) > 0$ is called a focal element.

2.2 Combination operators

In several real-world problems, information has to be gathered from distinct sources. These latter have to be merged with the aim of obtaining the most accurate information possible. The belief function framework provides a set of combination rules for fusing such kinds of information. The conjunctive rule, proposed by Smets within the Transferable Belief Model (TBM) [25], is one of the best known rules. Given two information sources S_1 and S_2 with respectively m_1 and m_2 as bbas, the conjunctive rule, denoted by $\bigcirc\!\!\!\!\cap$, will be set as:

$$m_1 \bigcirc\!\!\!\!\cap\, m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Theta. \tag{2}$$

The belief fully involved to the empty set reflects the conflictual mass. With the aim of retaining the basic properties of the belief function theory, Dempster have proposed in [4], a normalized version of the conjunctive rule. This latter allows to manage the conflict by redistributing the conflictual mass over all focal elements. It is obtained as follows:

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Theta \tag{3}$$

where the conflictual mass K caused by the combination of the two bbas m_1 and m_2 through the conjunctive rule, is given as follows:

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{4}$$

2.3 Decision making

The pignistic probability $BetP$, proposed in [24], is an efficient and binding way for decision-making. It transforms beliefs into probability measures as follows:

$$BetP(A) = \sum_{B \cap A = \emptyset} \frac{|A \cap B|}{|B|} m(B), \quad \forall A \in \Theta \tag{5}$$

Making decision consists of selecting the most likely hypothesis, meaning the hypothesis H_s with the highest pignistic probability:

$$H_s = argmax_A BetP(A), \quad \forall A \in \Theta \tag{6}$$

2.4 Dissimilarity between bbas

Numerous measures have been introduced for computing the dissimilarity degree between two given bbas [7, 18, 26]. The Jousselme distance [7] is regarded as one of the well-known ones. Formally, the Jousselme distance, for two given bbas m_1 and m_2, is defined by:

$$dist(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D(m_1 - m_2)} \tag{7}$$

where the Jaccard similarity measure D is set to:

$$D(X, Y) = \begin{cases} 1 & \text{if } X{=}Y{=}\ \emptyset \\ \dfrac{|X \cap Y|}{|X \cup Y|} & \forall\ X{,}Y\ \in 2^{\Theta} \end{cases} \tag{8}$$

3 Rough set theory

The rough set theory, which is proposed by Pawlak [15], constitutes a valid mathematical solution for handling imperfect data for several machine learning applications. Examples include clustering [14], classification [6, 8] and attribute reduction [1, 11], etc. Attribute reduction within the rough set theory consists of discovering the minimal subsets of relevant features, also named reduct, from the original set. Authors in [22], have introduced the notation of discernibility matrix and function as a way for finding reducts for a given data T. Suppose that $T = \{x_1, \ldots, x_D\}$ is a data composed with D objects x_i ($i \in \{1, \ldots, D\}$) characterized by N attributes $A{=}\{a_1, \ldots, a_N\}$ having values $V = \{v_1^i, \ldots, v_N^i\}$ and a class label $Y_i \in C = \{c_A, \ldots, c_Q\}$ (i.e. Q is the number of classes). The discernibility matrix of T, denoted by DM, is a $|D| \times |D|$ matrix in which the element $DM(x_i, x_j)$ for an object pair (x_i, x_j) is defined as follows $\forall\ i, j = \{1, \ldots, D\}$ and $\forall\ n = \{1, \ldots, N\}$:

$$DM(x_i, x_j) = \{a_n \in A | v_n^i(x_i) \neq v_n^j(x_j) \text{ and } Y_i \neq Y_j\}$$

Each element $DM(x_i, x_j)$ represents the set of all condition attributes discerning objects x_i and x_j that have not the same class label. The notion of discernibility function can be defined from the discernibility matrix as follows:

$$f(DM) = \wedge\{\vee(DM(x_i, x_j)) | \forall x_i, x_j \in T, DM(x_i, x_j) \neq \emptyset\} \tag{9}$$

Reducts may be yielded by transforming the discernibility function from conjunctive normal form into disjunctive normal form. The major shortcoming of this solution is its costly computation which makes it impractical for large or even medium sized data sets. Therefore, several heuristics have been discussed to overcome this drawback. Johnson's heuristic algorithm and the hitting set approach are ones of the most known algorithms [3].

4 Classifier ensemble through rough set reducts

In this paper, we aim to construct an ensemble of classifiers from data characterized by uncertain attribute values expressed within the belief function framework. Particulary, we propose to construct an ensemble of the Enhanced Evidential k Nearest Neighbors (EEk-NN), an extended version of the classical k-NN for handling evidential data, through rough set reducts. The general structure of our proposed idea is depicted in Figure 1.

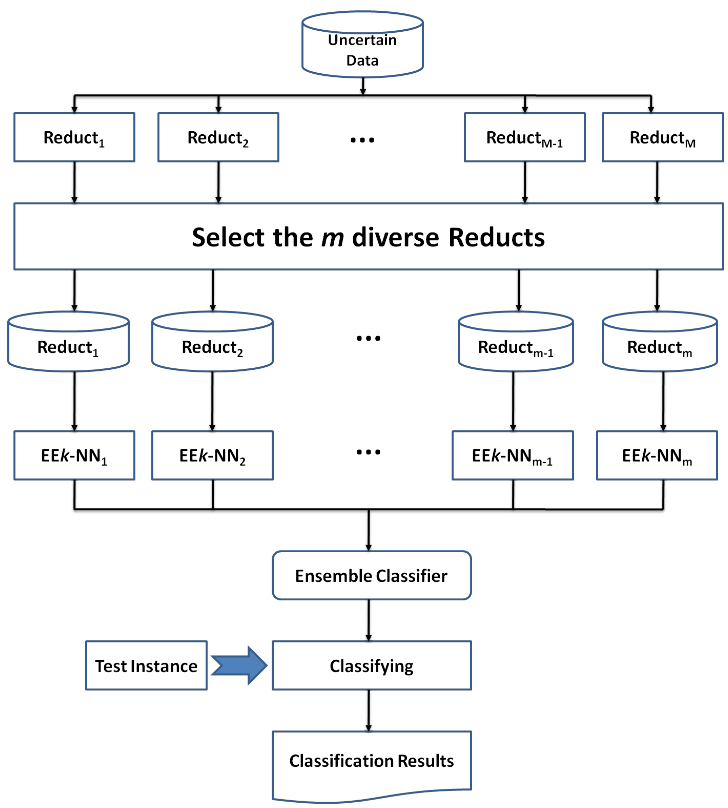


Fig. 1: **The general structure of the proposed method.**

Given a training data with uncertain attribute values, we have to generate firstly all possible rough set reducts $R = \{r_1, \ldots, r_M\}$. Subsequently, we have to choose the ones enabling the construction of a good ensemble of EEk-NNs. Mainly, we have to pick out the most diverse ones. The decision yielded by each individual classifier, for a given query instance, will be merged using the Dempster operator, one well used belief function combination rules for merging distinct classifiers.

Numerous reduct generation methods have been proposed in the literature and the commonly used ones are mainly based on the information entropy and the discernibility matrix. Examples include the Johnson algorithm and the hitting set approach. The former one consists of a greedy search technique for piking out a single reduct which is generally close to the optimal, while the latter one allows the computation of multiple reducts. Within the hitting set approach, a multiset ζ will contain the non empty sets of a given discernability matrix and the minimal hitting sets of ζ are exactly the reducts.

Since the computation of reducts using the hitting set approach is an NP-hard problem, genetic algorithms have been used for generating approximate hitting sets, meaning approximate reducts. One example is the SAVGenetic algorithm reducer [16], a Rosetta toolkit algorithm providing multiple reducts on the basis of the hitting set paradigm [10]. Since this algorithm is widely used, it has not the ability to handle uncertain data. Herein, we propose to extend the SAVGenetic algorithm reducer for handling data with uncertain attribute values represented by belief functions.

In analogy with the SAVGenetic algorithm, our proposed algorithm consists firstly of computing the discernability matrix for data with uncertain attributes. We have already proposed, in [27], a belief discernability matrix for discerning pairs of objects with uncertain attribute values expressed in terms of belief functions.

Given a data set T=$\{x_1,\ldots,x_D\}$ with a finite set of D objects x_i ($i \in \{1, \ldots, D\}$). Every object x_i is defined by a set of N uncertain attributes $A = \{a_1, \ldots, a_N\}$ with values $uV^i = \{uv_1^i, \ldots, uv_N^i\}$ and a certain class label $Y_i \in C = \{c_1, \ldots, c_Q\}$. Each uncertain attribute value uv_n^i relative to an instance x_i (with n in $\{1, \ldots, N\}$) will be expressed by a basic belief assignment $m_i^{\Theta_n}$ where Θ_n reflects the frame of discernment relative to the attribute n. Let S denotes a tolerance threshold (i.e. in this paper S is set to 0.1 for maximizing the search space) and $dist$ reflects the Jousselme distance. The entries of our proposed belief discernibility matrix, denoted by Λ', have been set as follows $\forall$ $i, j \in \{1, \ldots, D\}$ and $n \in \{1, \ldots, N\}$:

$$\Lambda'(x_i, x_j) = \{a_n \in A | dist(m_i^{\Theta_n}, m_j^{\Theta_n}) > S \text{ and } Y_i \neq Y_j\} \tag{10}$$

A multiset ζ' will then contain the non empty set of Λ' and the approximate hitting sets of ζ' correspond exactly to the approximate reducts. Our algorithm's fitness function corresponds exactly to that of the standard SAVGenetic algorithm. It consists of two main parts. The former one rewards subsets with shortest size, while the latter one rewards subsets that are hitting sets (i.e. meaning subsets having a non empty intersection with all elements of the discernability matrix). It is set as follows for each subset $B \in 2^N$:

$$f(B) = (1-\alpha) \times \frac{|A| - |B|}{|B|} + \alpha \times min\{\varepsilon, \frac{[F \in \zeta' | F \cap B = \emptyset|]}{|\zeta'|}\} \qquad (11)$$

where $\alpha \in [0, 1]$ reflects the adaptive weighting between the two terms and ε expresses the minimal hitting set fraction.

It is important to note that rough set approaches may generate tens or even hundreds of reducts and the most diverse ones have to be chosen for ensemble learning. One simplest algorithm for picking out diverse reducts is introduced in [3]. It consists of choosing randomly a reduct from the initial list and then adding progressively reducts that are diverse as much as possible from the chosen ones. The diversity degree is set as:

$$Div_s = 1 - \frac{\frac{\sum_{k \in L} Red_s \cap Red_k}{Red_s \cup Red_k}}{L} \qquad (12)$$

where L is the number of the chosen reducts and Red_s is the candidate reduct. The candidate reduct with the highest diversity will be added to the list of the chosen reducts.

5 Experimentations

More recently, we have introduced an ensemble of EEk-NNs through the random subspace method (RSM) and we have proven its performance compared with the individual EEk-NN that is trained in full feature space [28]. In this paper, we aim to evaluate the performance of the ensemble of EEk-NNs through rough set reducts (RSR). Thus, we propose to carry out a comparative study with the ensemble proposed in [28]. We relied mainly on the percentage of correct classification criterion (PCC). In what follows, we present our experimentation settings and results.

5.1 Experimentation settings

As we are handling data with uncertain attribute values expressed in terms of belief functions, we have proposed to construct synthetic databases by injecting uncertainty on real databases obtained from the the UCI machine learning repository [12]. We provide, in Table 1, a brief description of some categorical databases where #Instances, #Attributes and #Classes denote, respectively, the number of instances, the number of attributes and the number of classes. We have tackled different uncertainty levels P:

- Certain case: P=0
- Low uncertainty case: $(0 < P < 0.4)$
- Middle uncertainty case :$(0.4 \leq P < 0.7)$
- High uncertainty case $(0.7 \leq P \leq 1)$

Table 1: **Description of databases.**

Databases	#Instances	#Attributes	#Classes
Voting Records	435	16	2
Monks	432	7	2
Breast Cancer	286	9	2
Lymphography	148	18	4
Tic-Tac-Toa	958	9	2

Suppose that T is a database composed with D instances x_i $(i \in \{1, \ldots, D\})$. Each instance x_i is characterized by N uncertain attribute values uv_i^n $(n \in \{1, \ldots, N\})$. Suppose that Θ_n is the frame of discernment relative to the attribute n. Each attribute value uv_i^n relative to an instance x_i such that $uv_i^n \subseteq \Theta_n$ will be expressed in terms of belief functions as follows:

$$\begin{aligned} m^{\Theta_n}\{x_i\}(uv_i^n) &= 1 - P \\ m^{\Theta_n}\{x_i\}(\Theta_n) &= P \end{aligned} \tag{13}$$

It is important to note that some databases suffer from incompleteness. The belief function theory allows to represent and manage missing attribute values. In this paper, the missing attribute values will be modeled as follows:

$$\begin{aligned} m^{\Theta_n}\{x_i\}(uv_i^n) &= 0 \qquad \forall\, uv_i^n \subseteq \Theta_n \\ m^{\Theta_n}\{x_i\}(\Theta_n) &= 1 \end{aligned} \tag{14}$$

5.2 Experimentation results

In this experimentations, we have relied on the 10-fold cross validation strategy for learning the individual Enhanced k- Nearest Neighbors classifiers. One key issue which has to be addressed is the number of neighbors that may give satisfactory results. In our experimentation tests, we evaluate five values of k which respectively correspond to 1, 3, 5, 7 and 9. Another substantial key element when designing an ensemble of classifier is the number of individual classifiers used to get the final decision. The conclusion conducted following to the study of [13] proves that ensembles of 25 classifiers are sufficient for reducing the error rate and consequently for improving performance. Thus, in our paper, the number of the merged classifiers will be equal to 25. The final PCCs, which are obtained through the combination of the classifier outputs using the Dempster rule, will be given from Table 2 to Table 6, where RSM reflects the results yielded through the random subspace method and RSR reflects the results obtained with the rough set reducts method.

Table 2: **Results for Voting Records database (%).**

	$k=1$		$k=3$		$k=5$		$k=7$		$k=9$	
	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR
No	90.23	93.72	91.62	93.72	90.93	93.72	91.93	93.72	91.16	93.72
Low	90.46	93.95	91.62	93.95	91.16	93.95	90.06	93.95	91.39	93.95
Middle	91.39	93.95	91.39	93.95	91.86	93.95	91.62	93.95	90.93	93.95
High	88.37	95.53	89.53	89.76	89.76	89.76	89.76	90	89.30	90
Average	90.11	**95.03**	91.04	**92.84**	90.92	**92.84**	90.70	**93.65**	90.69	**93.65**

Table 3: **Results for Lymphography database (%).**

	$k=1$		$k=3$		$k=5$		$k=7$		$k=9$	
	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR
No	83.57	86.42	82.85	86.42	85.71	85	84.28	85	80.71	85
Low	81.42	77.85	90	77.85	81.42	73.85	78.85	78.57	79.28	77.85
Middle	83.57	81.42	82.85	81.42	83.57	80.71	81.42	80	82.85	80
High	62.24	72.85	62.85	73.75	61.42	73.57	62.85	73.57	61.42	73.57
Average	77.70	**79.62**	77.38	**79.81**	78.03	**79.28**	76.82	**79.10**	76.06	**79.10**

Table 4: **Results for Tic-Tac-Toa database (%).**

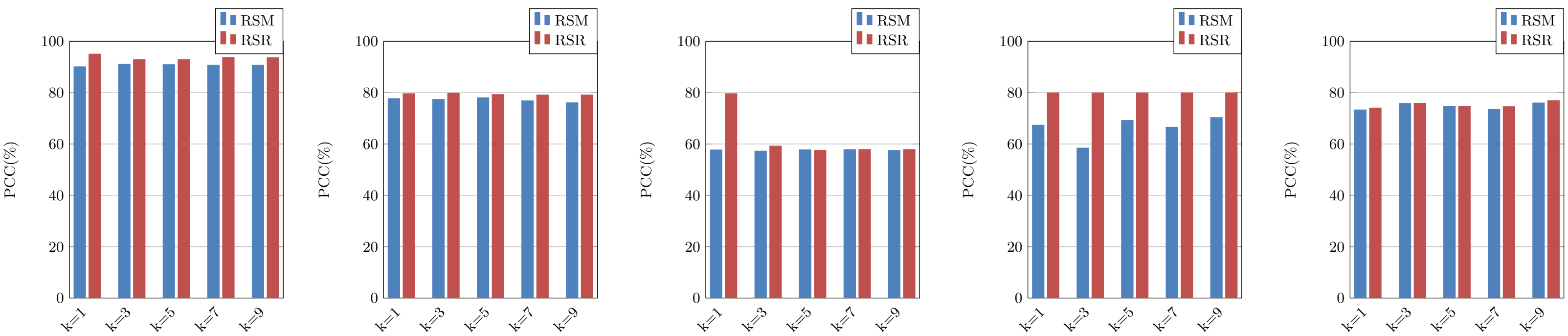
	$k=1$		$k=3$		$k=5$		$k=7$		$k=9$	
	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR
No	62.10	62	61.15	63.89	61.05	62.42	60.84	62	60	60.63
Low	55.36	56.73	55.78	57.05	55.89	54.52	55.89	55.26	55.57	55.68
Middle	55.57	57.57	56	57.47	56	55.47	56.21	56.10	56.21	56
High	57.78	57.89	57.68	58.10	58	57.89	58.31	58.21	50.31	59.15
Average	57.70	**58.54**	57.25	**59.2**	**57.73**	57.57	57.81	**57.89**	57.52	**57.86**

Table 5: **Results for Monks database (%).**

	$k=1$		$k=3$		$k=5$		$k=7$		$k=9$	
	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR
No	73.13	85.45	60.26	85.45	61.68	85.45	69.03	85.45	79.81	85.45
Low	71.01	84.36	59.49	84.36	94.16	84.36	70.65	84.36	76.54	84.36
Middle	69.85	85.27	60.26	85.27	68.9	85.27	72.84	85.27	70.72	85.27
High	56.14	64.72	53.68	64.72	52.03	64.72	53.72	64.72	54.18	64.72
Average	67.35	**79.95**	58.42	**79.95**	69.19	**79.95**	66.56	**79.95**	70.31	**79.95**

Table 6: **Results for Breast Cancer database (%).**

	$k=1$		$k=3$		$k=5$		$k=7$		$k=9$	
	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR	RSM	RSR
No	73,13	75.08	76.18	75.07	75.10	73.71	74.04	76.03	76.03	76.9
Low	73.10	73.57	76.10	76.18	74.89	75.10	73.91	74.04	75.8	76.9
Middle	73.92	73.75	75.59	76.18	74.32	75.10	73.80	74.04	76	76.9
High	73.01	73.75	75.45	76.18	74.62	75.10	72.12	74.04	76.13	76.9
Average	73.29	**74.03**	75.83	**75.90**	74.73	**74.75**	73.46	**74.53**	75.99	**76.9**



(a) Vote Records (b) Lymphography (c) Tic-Tac-Toa (d) Monks (e) Breast Cancer

Fig. 2: **Average accuracies.**

The PCC results given from Table 2 to Table 6 and the average classification accuracies depicted in Figure 2 have proven the efficiency of the ensemble classifiers that are obtained through the rough set reduct approach over that yielded using the random subspace method. In fact, the average accuracies achieved by the RSR method for the different values of k are almost always greater than those achieved by the RSM method. Taking the Voting Records database as example, the average accuracy done by the RSR approach are equal to 95.03, 92.84, 92.84, 93.65 and 93.65, while there are equal to 90.11, 91.04, 90.92, 90.70 and 90.69 for respectively k=1, k=3, k=5, k=7 and k=9. The conclusion derived from the carried out experimentation tests may be justified by the fact that random subspace methods may negatively affect the classification process as irrelevant and redundant features can part of the selected subsets .

6 Conclusion

The idea underlying this paper is to increase accuracy for a given classification system through ensemble systems. Herein, we have constructed an ensemble of the so-called Enhanced Evidential k-Nearest Neighbors for dealing with uncertain data, more precisely where the uncertainty pervades the attribute values and is represented with belief functions. With the aim of assessing the performance of our proposed technique, we have conducted a comparative study with ensemble constructed through random subspaces. The yielded results have shown the efficiency of the rough set reducts over random subspaces. As we combine distinct classifiers, in this paper, we have relied on the Dempster rule of combination. As there are other combination rules, in our future work, we intend to pick out the combination operator that yields the best classification results. With the aim of increasing accuracy, we look forward to take into consideration not only the diversity between reducts, but also the diversity between the merged classifiers and the accuracy of the individual classifiers to yield more performance.

References

[1] R. B. Bhatt and M. Gopal. On fuzzy-rough sets approach to feature selection. *Pattern recognition letters*, 26(7):965–975, 2005.

[2] R. Bryll, R. Gutierrez-Osuna, and F. Quek. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern recognition*, 36(6):1291–1302, 2003.

[3] E. Debie, K. Shafi, C. Lokan, and K. Merrick. Reduct based ensemble of learning classifier system for real-valued classification problems. In *IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL)*, pages 66–73. IEEE, 2013.

[4] A. P. Dempster. Upper and Lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[5] S. Günter and H. Bunke. Feature selection algorithms for the generation of multiple classifier systems and their application to handwritten word recognition. *Pattern recognition letters*, 25(11):1323–1336, 2004.

[6] R. Jensen and C. Cornelis. Fuzzy-rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42):5871–5884, 2011.

[7] A. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information fusion*, 2(2):91–101, 2001.

[8] L. Khoo, S. Tor, and L. Zhai. A rough-set-based approach for classification and rule induction. *International Journal of Advanced Manufacturing Technology*, 15(6):438–444, 1999.

[9] Y. Kim. Toward a successful crm: variable selection, sampling, and ensemble. *Decision Support Systems*, 41(2):542–553, 2006.

[10] J. Komorowski, A. Øhrn, and A. Skowron. The rosetta rough set software system. *Handbook of data mining and knowledge discovery*, pages 2–3, 2002.

[11] P. Kumar, P. Vadakkepat, and L. A. Poh. Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets. *Applied Soft Computing*, 11(4):3429–3440, 2011.

[12] P. Murphy and D. Aha. UCI repository databases. *http://www.ics.uci.edu/mlear*, 1996.

[13] D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

[14] D. Parmar, T. Wu, and J. Blackhurst. MMR: an algorithm for clustering categorical data using rough set theory. *Data & Knowledge Engineering*, 63(3):879–893, 2007.

[15] Z. Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, 1982.

[16] S. Phon-Amnuaisuk, S. Ang, and S. Lee, editors. *Multi-disciplinary Trends in Artificial Intelligence. In the proceedings of 11th International Workshop, MIWAI*, volume 10607. Springer, 2017.

[17] M. P. Ponti Jr. Combining classifiers: from the creation of ensembles to the decision fusion. In *24th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 1–10. IEEE, 2011.

[18] B. Ristic and P. Smets. The TBM global distance measure for the association of uncertain combat id declarations. *Information Fusion*, 7(3):276–284, 2006.

[19] S. Saha, C. Murthy, and S. K. Pal. Classification of web services using tensor space model and rough ensemble classifier. In *International Symposium on Methodologies for Intelligent Systems*, pages 508–513. Springer, 2008.

[20] L. Shi, X. Ma, L. Xi, Q. Duan, and J. Zhao. Rough set and ensemble learning based semi-supervised algorithm for text classification. *Expert Systems with Applications*, 38(5):6300–6306, 2011.

[21] L. Shi, L. Xi, X. Ma, M. Weng, and X. Hu. A novel ensemble algorithm for biomedical classification based on ant colony optimization. *Applied Soft Computing*, 11(8):5674–5683, 2011.

[22] A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In *Intelligent Decision Support*, pages 331–362. Springer, 1992.

[23] P. Smets. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

[24] P. Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

[25] P. Smets and R. Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994.

[26] B. Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315–329, 1993.

[27] A. Trabelsi, Z. Elouedi, and E. Lefevre. Feature selection from partially uncertain data within the belief function framework. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 643–655. Springer, 2016.

[28] A. Trabelsi, Z. Elouedi, and E. Lefevre. Ensemble enhanced evidential k-nn classifier through random subspaces. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 212–221. Springer, 2017.

[29] K. Tumer and J. Ghosh. Classi er combining: analytical results and implications. In *Proceedings of the National Conference on Artificial Intelligence*, pages 126–132. Citeseer, 1996.

[30] K. Tumer and N. C. Oza. Input decimated ensembles. *Pattern Analysis & Applications*, 6(1):65–77, 2003.

[31] K. Turner and N. C. Oza. Decimated input ensembles for improved generalization. In *In proceedings of International Joint Conference on Neural Network (IJCNN'99)*, volume 5, pages 3069–3074. IEEE, 1999.

[32] S.-L. Wang, X. Li, S. Zhang, J. Gui, and D.-S. Huang. Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction. *Computers in Biology and Medicine*, 40(2):179–189, 2010.