Assessing Significance of Cognitive Assessments for Diagnosing Alzheimer's Disease with Fuzzy-Rough Feature Selection

Tianhua Chen[1], Changjing Shang[2], Pan Su[3], Yinghua Shen[4], Mufti Mahmud[5], Raymond Moodley[6], Grigoris Antoniou[1], and Qiang Shen[2]
For the Alzheimer's Disease Neuroimaging Initiative*

[1] Department of Computer Science, University of Huddersfield, UK
[2] Department of Computer Science, Aberystwyth University, UK
[3] School of Control and Computer Engineering, North China Electric Power University, Baoding, China
[4] School of Economics and Business Administration, Chongqing University, China
[5] Department of Computer Science, Nottingham Trent University, UK
[6] Institute of Artificial Intelligence, De Montfort University, Leicester, UK

Abstract. Research in dementia diagnosis typically involves a range of data modalities and also, the use of cognitive assessments, aiming at the development of approaches that are non-invasive, time-saving and economical. Given the existing diversity of prevalent cognitive assessment factors it is useful to assess and exploit the effectiveness of such cognitive features, while working towards the establishment of a methodology for making informed choice of such factors in practical use. As an initial approach, this paper employs the powerful Fuzzy-Rough Feature Selection (FRFS) technique to support such an analysis, by varying the underlying similarity functions and search strategies employed by FRFS. Evaluated on a benchmark from the renowned Alzheimer's Disease Neuroimaging Initiative repository, experimental results demonstrate the significance and predictive capabilities of different cognitive assessments in working with a variety of popular classifiers.
Keywords: Dementia, cognitive measure, fuzzy-rough feature selection

1 Introduction

Dementia can be generally described as a condition that impairs the regular cognitive functions of the brain affecting memory, language, behaviour and the ability to carry out day-to-day tasks [1]. Primarily affecting the senior age group, between 60% and 70% of dementia cases are currently attributed to the Alzheimer's

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Disease (AD) [1]. Although no complete medical cure to this disease is currently possible, successful diagnosis at as early stage as possible would facilitate the use of symptom-slowing medication and enable the derivation of care plans to preserve as high a quality of life as possible [2]. Obviously, this has significant implications for economies, healthcare services and the public.

Research in dementia diagnosis typically covers a variety of data modalities, ranging from basic patient information (e.g., demographics and medical history), through cognitive measures that assesses cognitive functions (e.g., memory, learning and language), to neuroimaging and other biomarkers (e.g., magnetic resonance imaging (MRI) scans and cerebrospinal fluid (CSF)). However, biomarkers such as CSF, which are normally obtained with the classical method of lumbar puncture, are typically invasive and costly. The use of imaging also tends to be expensive [3] and time-consuming (with waiting lists up to 18 weeks in England [4]).

These observations have given rise to alternative considerations by the use of cognitive assessments. The acquisition of cognitive assessments is generally time-efficient, cost-saving and non-invasive, as it just involves a series of pen-and-paper tests and questions. Each of such assessment carries a score, which is interchangeably termed cognitive assessment factor or cognitive feature hereafter. Yet, among numerous cognitive assessments available, there is no global standard for what assessments are more appropriate to be applied to patients, as reflected by different practices observed in recent literature [5,6], while the separate use of such assessment factors individually also tends to perform poorly. For instance, when considering only the Mini-Mental State Examination (MMSE) [7], one of the most widely used assessment features, a mere less than 70% accuracy is attainable even with the powerful support vector machine and neural networks employed as the classifier. Fortunately, the use of multiple assessment factors has generally been shown to provide a good indicator of AD in a number of case studies [6].

In recognition of the potential of cognitive assessments, for effectively diagnosing dementia while being non-invasive, time-saving and economical, this paper investigates the significance of typical prevalent cognitive features. This is motivated with an aim to eventually offer a methodological approach to informing the choices of which assessments to be combined to work with powerful machine learning tools. It is carried out in an effort to improve the conduct of AD, inspired by the observation of recent advances in applying machine learning for successful applications in the healthcare industry [8,9]. In particular, the popular Fuzzy-Rough Feature Selection (FRFS) technique [10] is adopted herein as the basis upon which to perform such an investigation. Note that whilst having found successes in numerous problem domains, it is for the first time that FRFS is exploited in support of performing AD.

The core for the implementation of FRFS includes a similarity function and a search mechanism that it employs. As an initial examination in this important area, the exploration of the potential significance and predictive capabilities of cognitive assessment factors is therefore conducted by designing schemes that

focus on three similarity functions and two search strategies. In particular, the work evaluates the potential effects these core functionalities of FRFS may have on the resulting feature subsets (that best describe characteristics which may be utilised to perform cognitive assessment-based AD). To be complete, a brief outline of FRFS is presented in Section 2, with the problem case or dataset concerned being specified in Section 3. Then, experimental results are discussed and the conclusions drawn in Sections 4 and 5, respectively.

2 Fuzzy-rough Feature Selection for Significance Analysis

Fuzzy-Rough Feature Selection (FRFS) provides a means by which discrete and/or real-valued noisy data can be effectively reduced without requiring additional information such as thresholds or domain-dependent knowledge [11]. It is a natural extension to the original rough set-based feature selection methods ([12]). Being complementary to rough sets that are concerned with indiscernibility, fuzzy sets are concerned with vagueness. With many recent studies drawing conclusions about the complementary nature of the two methodologies [10], the hybridisation of both theories has established itself a popular choice in developing practically effective feature selection algorithms. Particularly, a fuzzy-rough set is defined by two fuzzy sets, i.e., a fuzzy lower approximation and a fuzzy upper approximation, obtained by extending the corresponding crisp rough set notions, resulting in greater flexibility in handling uncertainty by allowing membership of elements in the range [0,1] instead of being simply either absolute certainty or not at all exclusively.

For the present application problem of AD, without losing generality, let $IS = (\mathbb{U}, \mathbb{A})$ be an information system for the dementia data under study, where $\mathbb{U}$ is a nonempty finite set of patients (the universe) and $\mathbb{A}$ is a nonempty finite set of attributes such that $a : \mathbb{U} \rightarrow V_a$ for every $a \in \mathbb{A}$, with V_a being the set of values that the attribute a may take. Generally speaking, for decision systems, $\mathbb{A} = \{\mathbb{C} \cup \mathbb{D}\}$, where $\mathbb{C}$ is the set of conditional predictors and $\mathbb{D}$ contains a single attribute d standing for the decision variable that indicates the diagnostic outcome. The following defines the fuzzy lower and upper approximations:

$$\mu_{\underline{R_B}X}(x_i) = \inf_{x_j \in \mathbb{U}} I(\mu_{R_B}(x_i, x_j), \mu_X(x_j)) \tag{1}$$

$$\mu_{\overline{R_B}X}(x_i) = \sup_{x_j \in \mathbb{U}} T(\mu_{R_B}(x_i, x_j), \mu_X(x_j)), \tag{2}$$

where X is the (fuzzy) concept being approximated, I is a fuzzy implicator, T is a t-norm, and R_B is the fuzzy similarity relation induced by the subset of features B, and $x_i, x_j \in X$ are two arbitrary patients in X. In particular,

$$\mu_{R_B(x_i,x_j)} = T_{a \in B}\{\mu_{R_a}(x_i, x_j)\} \tag{3}$$

where $\mu_{R_a}(x_i, x_j)$ is the degree to which the patients x_i and x_j are similar for a certain feature $a \in \mathbb{A}$.

FRFS then employs a quality measure termed the fuzzy-rough dependence function $\gamma_B(Q)$ to gauge the dependency degree between two sets of attributes B and Q, which is defined by:

$$\gamma_B(Q) = \frac{\sum_{x \in \mathbb{U}} \mu_{POS_{R_B}(Q)}(x)}{|\mathbb{U}|}. \tag{4}$$

In this definition, the following concept of fuzzy positive region is introduced:

$$\mu_{POS_{R_B}(Q)}(x) = \sup_{X \in \mathbb{U}/Q} \mu_{\underline{R_B}X}(x), \tag{5}$$

which contains all patients of $\mathbb{U}$ that can be classified into classes of $\mathbb{U}/Q$ using the information in B. Therefore, $\gamma_B(Q)$ may be viewed as a measure of quality for a given feature subset $B \in \mathbb{C}$, with respect to the decision attribute $\{d\}$ (or more generally, the set of decision features $\mathbb{D}$): $0 \leq \gamma_B(Q) \leq 1$, with $\gamma_\emptyset(\mathbb{D}) = 0$. A fuzzy-rough reduct R can then be defined as a subset of features that preserves the dependency degree of the entire data set, i.e., $\gamma_R(\mathbb{D}) = \gamma_\mathbb{C}(\mathbb{D})$.

Much has been done in the literature in order to establish an efficient method for the evaluation of $\gamma_R(\mathbb{D})$ (or simply, $\gamma_R(\{d\})$ for the problem at hand). Amongst the different approaches is the popular hill climbing-based FRFS algorithm termed fuzzy-rough QuickReduct [10], which follows the original QuickReduct algorithm that is based on exploiting rough sets alone. The implementation of FRFS typically adopts a forward search strategy by adding to the current candidate feature subset a feature that leads to the highest fuzzy-rough dependency improvement. It terminates when the addition of any remaining feature does not result in an increase in the dependency measure. Being able to pick up features that always contribute most to the dependency, the forward approach is able to directly identify the most informative features, which could therefore be used to examine the significance of cognitive features. Instead of utilising forward search, backward search may also be used, which starts with the full set of features and iteratively removes a feature that does not affect the dependency measure until all features are examined. Either way, FRFS is utilised to assess the significance of cognitive features iteratively.

Note that the choice of a specific similarity relation $\mu_{R_a}(x_i, x_j)$, which measures the degree to which the patients x_i and x_j are deemed similar with respect to a certain feature $a \in \mathbb{A}$, may have a direct impact on the convergence rate of the QuickReduct algorithm, hence affecting what feature subset is to be finally selected and returned. Naturally, this will in turn, affect the discriminative capability of the selected features for use as the input to a predictive modelling algorithm or model subsequently. In general, when defining a similarity relation, three basic properties should be satisfied: (i) reflexivity iff $x_i \in \mathbb{U}, \mu_{R_a}(x_i, x_i) = 1$; (ii) symmetry iff $x_i, x_j \in \mathbb{U}, \mu_{R_a}(x_i, x_j) = \mu_{R_a}(x_j, x_i)$; and (iii) T-transitivity iff $x_i, x_j, x_k \in \mathbb{U}, \mu_{R_a}(x_i, x_j) \leq T(\mu_{R_a}(x_i, x_k), \mu_{R_a}(x_k, x_j))$, where T is a T-norm, e.g., a mapping $T(a, b) : [0, 1] \times [0, 1] \rightarrow [0, 1]$. As many functional relations exist satisfying these properties and hence, many can act as such a similarity measure. It is therefore, of great importance to examine the effectiveness and

robustness of any candidate similarity relations for the potential adoption in a given application. In particular, the following three computationally simple, and popular, similarity relations [10] (termed Sim-1, Sim-2, Sim-3, respectively) are herein to be explored for the AD problem:

$$\mu_{R_a}(x_i, x_j) = 1 - \frac{|a(x_i) - a(x_j)|}{a_{max} - a_{min}} \tag{6}$$

$$\mu_{R_a}(x_i, x_j) = \exp\left(- \frac{(a(x_i) - a(x_j))^2}{2\sigma_a^2} \right) \tag{7}$$

$$\mu_{R_a}(x_i, x_j) = \max\left(\min\left(\frac{a(x_j) - (a(x_i) - \sigma_a)}{a(x_i) - (a(x_i) - \sigma_a)}, \frac{(a(x_i) + \sigma_a) - a(x_j)}{(a(x_i) + \sigma_a) - a(x_i)} \right), 0 \right) \tag{8}$$

where σ_a^2 is the variance of the feature a, and $a(x_i)$ is the value of a for the object x_i.

In short, in adopting FRFS to aid in improving the performance of AD, the aforementioned two search strategies and three similarity functions will be explored below.

3 Benchmark Data Set

The benchmark data used in this study is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) repository. In order to lessen the potential impact from multiple data modalities such as imaging and other biological features, only patient information and cognitive assessment factors are considered. The patient information includes Age, Gender, level of education (Edu), ethnicity (Eth), race and marital status, as well as medical history of 19 items. Each of the 19 historical items is the answer to a binary question as to whether the patient has any specific historical issue in relation to psychological and neurological disorders, e.g., smoking, drugs, and alcohol abuse. Additionally, information on whether there is any close family member suffering from dementia (Father, Mother, Siblings) is included.

Seven popular cognitive assessment factors [13] are selected, including: Alzheimer's Disease Assessment Scale 13 (ADAS13) that adds onto the checks for concentration, planning and executive functions; Mini-Mental State Examination (MMSE) that covers the aspects of orientation, registration, attention, recall and language; Rey Auditory Verbal Learning Test (RAVLT) that is made up of several verbal trials considering various levels of memory and learning, including scores from the immediate (RAVLT_imdt), learning (RAVLT_learn), delayed memory (RAVLT_forget) and the percentage result (RAVLT_perc_forget); Functional Assessment Questionnaire (FAQ), that measures the levels of ability to

accomplish tasks beyond the basic day-to-day needs; Montreal Cognitive Assessment (MoCA) that assesses the cognitive domains including short-term memory recall, visuospatial abilities, language and abstract reasoning; and the means of the Logical Memory – Immediate Recall (LIMM) and Logical Memory – Delayed Recall (LDEL).

Table 1. Partial statistics on benchmark data set

	Age	Edu	ADAS13	MMSE	RAVLT_imdt	RAVLT_learn	RAVLT_forget	RAVLT_perc_forget	LDEL	FAQ	MoCA	LIMM
count	950	950	941	950	945	945	945	941	949	945	127	950
mean	74.68	15.57	17.59	26.96	33.43	3.88	4.29	60.62	6.17	4.54	24.20	8.61
std	7.10	3.00	9.04	2.60	11.60	2.67	2.39	34.59	5.15	6.31	2.89	4.71
min	54.4	4	1	18	0	-4	-5	-100	0	0	17	0
25%	70.6	13	10.67	25	25	2	3	33.33	2	0	23	5
50%	74.8	16	16.67	28	32	4	4	62.5	5	1	24	8
75%	79.8	18	23.33	29	41	6	6	100	9	7	26	12
max	90.9	20	54.67	30	69	12	13	100	22	30	30	22

In total, 950 instances are extracted from ADNI, involving 38 independent features, with the decision variable d representing the diagnosis. In particular, the value of $d = 229$ indicates the diagnostic outcome being clinically normal (CN), 188 being Alzhermier's disease (AD), 402 being Late Mild Cognitive Impairment (LMCI), and 131 being Early Mild Cognitive Impairment (EMCI). For illustration, Table 1 shows partial statistical information obtained from the full set of domain variables involved. They generally take values from the the range of 0 to 100, which makes domain normalisation unnecessary (that would otherwise potentially damage the interpretability of the resulting models) [14].

Note that a number of variables or features have missing values in certain instances (including ADAS13(9), RAVLT_imdt(5), RAVLT_learn(5), RAVLT_forget(5), RAVLT_perc_forget(9), LDEL(1), FAQ(5), and MoCA(823), where a bracketed figure indicates the number of missing values regarding the variable concerned), but those listed in Table 1 do not contain any missing values. The missing values associated with cognitive assessment factors reflect the real-world scenarios. That is, owing to the availability of numerous cognitive assessment features while lacking a universal agreement on the selection of which to use, only a certain subset of assessments are conducted for most patients, tailored by their local healthcare providers. Instead of directly filling missing values with the mean of the relevant variable, which would ignore any variations exhibited in other variables across different patients, in this study missing values are filled up using the standard k-Nearest Neighbours (kNN) algorithm. In so doing, each missing feature value is imputed with the mean of the values from its k ($k = 3$ empirically) nearest neighbours that have a value for the feature concerned.

4 Experimental Investigation

This section forms a major focus of the present work, presenting results of experimentally investigating how FRFS may contribute towards assessing the signifi-

cance of cognitive assessment factors. Being the first study of the use of FRFS in support of AD, it reports on the selection results of feature subsets with FRFS, followed by an evaluation of how effective the selected subsets may be in helping with effective AD (which is itself carried out subsequently with a certain machine learning-based classifier).

4.1 Feature Subsets Returned by FRFS

Following the generic approach as presented in Section 2, applying FRFS to the dataset will lead to a subset of features to be returned, depending on what combination of the similarity function and search strategy is to be exploited.

Table 2. Feature subsets by forward search (features listed in order of being selected)

Similarity Function	**Feature Subset**
Sim-1 (Total = 24)	MoCA, LDEL, MMSE, Age, Edu, FAQ, RAVLT_forget, LIMM, RAVLT_learn, ADAS13, RAVLT_imdt, MH10GAST, MH16SMOK, MH12RENA, MH9ENDO, MH13ALLE, FHQMOM, MH3HEAD, MH2NEURL, MH7DERM, RAVLT_perc_forget, MH8MUSCL, MH17MALI, MHPSYCH
Sim-2 (Total = 20)	MoCA, LDEL, MMSE, Age, RAVLT_forget, Edu, FAQ, RAVLT_learn, LIMM, RAVLT_imdt, ADAS13, FHQMOM, MH3HEAD, MH10GAST, MH2NEURL, MH13ALLE, RAVLT_perc_forget, MH4CARD, PTGENDER, Eth
Sim-3 (Total = 9)	MoCA, LDEL, MMSE, Age, Edu, RAVLT_forget, FAQ, RAVLT_learn, RAVLT_imdt

Table 2 lists the feature subsets obtained by searching in a forward manner, with respect to the use of different similarity measures. This search mechanism adds individual selected features, one at a time, on to the emerging feature subset that currently contributes the most to the dependency measure. The use of different similarity functions is expected to affect the selection of features and how quickly the fuzzy-rough QuickReduct algorithm converges, leading to eventual feature subsets of different sizes. In particular, Sim-3 only picks up nine out of the 38 conditional attributes, whilst Sim-1 and Sim-2 selects 20 and 24 features, respectively. Interestingly, note that the use of either of these three similarity functions results in the selection of exactly the same first four features, i.e., MoCA, LDEL, MMSE and Age. In addition, of the nine features selected by Sim-3, seven are cognitive assessment factors which form a subset of those returned by Sim-1 or Sim-2. Both feature subsets returned via the use of Sim-1 and Sim-2 include all available cognitive assessments, but they differ in terms of certain selected features that reflect medical histories (variables starting with 'MH' in the table). These results reflect the significance of cognitive features, given their being selected by FRFS while leading to the most gains in the dependency measure.

Similarly, Table 3 lists the feature subsets returned by FRFS through searching backwards (which works by eliminating features iteratively that do not affect

Table 3. Feature subsets by backward search (no order between features kept)

Similarity Function	Feature Subset
Sim-1 (Total = 26)	Age, Gender, Edu, Eth, PTMARRY, MHPSYCH, MH3HEAD, MH4CARD, MH5RESP, MH7DERM, MH8MUSCL, MH9ENDO, MH11HEMA, MH12RENA, MH13ALLE, MH16SMOK, MH17MALI, MH18SURG, MH19OTHR, FHQMOM, FHQSIB, ADAS13, MMSE, RAVLT_imdt, RAVLT_learn, LDEL
Sim-2 (Total = 23)	Age, PTGENDER, Edu, PTMARRY, MHPSYCH, MH2NEURL, MH3HEAD, MH4CARD, MH5RESP, MH7DERM, MH8MUSCL, MH9ENDO, MH10GAST, MH11HEMA, MH12RENA, MH13ALLE, MH14ALCH, MH16SMOK, MH17MALI, FHQMOM, FHQSIB, ADAS13, RAVLT_imdt
Sim-3 (Total = 20)	Age, PTGENDER, Edu, PTRACCAT, MHPSYCH, MH2NEURL, MH3HEAD, MH4CARD, MH5RESP, MH7DERM, MH8MUSCL, MH9ENDO, MH10GAST, MH11HEMA, MH12RENA, MH13ALLE, MH17MALI, MH18SURG, FHQMOM, FHQSIB

the overall dependency measure). Note that whilst the size of the selected feature subsets has increased slightly for cases where Sim-1 and Sim-2 are utilised, as compared to the use of forward search, it has expanded the size of selected subset significantly for Sim-3 from nine to 20. More importantly perhaps, the features selected by Sim-1 and Sim-2 in this case only include four and two cognitive assessment features respectively (which are all included in the subsets obtained by forward search). This performance is a deterioratation from that observed for the case using Sim-3, having none of the cognitive features included (but has seven selected using forward search). Such a result is obtained in despite of the fact that in implementation, cognitive assessment factors have been deliberately positioned at the rear of the full feature set so that they become the features to be first tested by the backwards approach. From these results, it may be conjectured that the removal of a single cognitive feature may not deteriorate the overall discrimination capability of an existing collection that contains a majority of other features.

Summarising the results discussed above, the following observation can be attained: Although feature subsets returned by backward search are computationally equivalent, in achieving the best dependency degree between the conditional attributes and the decision attribute d, to those obtained by forward search, the use of backward approach tends to keep non-cognitive features while the use of forward approach tends to return cognitive features. This empirically gained insight offers an opportunity to examine further the discriminative performances of the selected feature subsets as input to the predictive models that implement the task of AD, as to be reported next.

4.2 Performance Evaluation of Selected Feature Subsets for AD

In order to evaluate the effectiveness and robustness of the feature subsets returned by FRFS (six different ones in the problem currently investigated), a range of popular machine learning-based classification models are applied (which are obtained from the WEKA machine learning toolkit [15]), including: the Naive

Bayes (NB), the simple Logistic Regression (LR), the Multi-Layer Perceptron (MLP), the K-Nearest Neighbours (K=3), the crisp classification rule learning algorithm JRip, the Random Forest (RF), the C4.5 Decision Tree and the Support Vector Machine (SVM), all with their respective default parameter settings as commonly adopted in the literature.

In the absence of testing data for performance evaluation, stratified tenfold cross-validation (10-CV) is employed for result validation. In 10-CV, a given data set is partitioned into ten subsets. Of the ten, nine subsets are used to carry out training (to generate the required rule base for building the classification system), and the remaining single subset is used as the testing data for assessing the learned classifier's performance. This process is randomly repeated for ten times to lessen the impact of any random factors, with the results of the 10 $\times$ 10 cross-validations averaged to produce the final outcome.

Table 4. Performance comparison on accuracy (%) and training time (s)

	No FS		**Forward+Sim-1**		**Forward+Sim-2**		**Forward+Sim-3**	
Algorithms	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**
NB	82.66 ± 3.91	0.01	82.06 ± 3.92	0	82.62 ± 3.94	0	83.63 ± 3.77	0
LR	82.67 ± 3.81	0.86	82 ± 3.86	0.6	82.33 ± 4.26	0.59	82.08 ± 3.94	0.53
MLP	76.11 ± 4.34	4.66	78.82 ± 3.98	2.28	82.8 ± 4.23	1.79	85.47 ± 4.18	0.73
IBK-3	48.32 ± 4.97	0	51.58 ± 4.93	0	60.49 ± 4.1	0	80.14 ± 3.91	0
JRip	89.39 ± 3.08	0.17	89.43 ± 3.17	0.11	89.42 ± 2.99	0.1	89.78 ± 3.02	0.06
RF	86.53 ± 3.75	0.08	87.72 ± 3.2	0.07	88.35 ± 3.21	0.06	89.4 ± 3.06	0.05
C4.5	88.88 ± 3.29	0.04	88.87 ± 3.12	0.03	88.46 ± 3.17	0.03	89.32 ± 3.44	0.01
SVM	79.98 ± 3.59	0.19	79.39 ± 3.81	0.11	80.01 ± 3.81	0.07	81.17 ± 3.83	0.02
Average	79.318	0.751	79.984	0.400	81.810	0.330	**85.124**	**0.175**

Table 4 presents the performance of employing different feature subsets returned through forward search by FRFS. In comparison to the original benchmark that does not involve feature selection (FS), Sim-1 achieves four better and four worse performances out of eight classifiers. Notably, results by Sim-2 and Sim-3 beat the benchmark with six and seven better results, respectively. With all three subsets achieving a better averaged performance over the benchmark, these experimental studies clearly demonstrate the effectiveness of the use of fuzzy-rough QuickReduct algorithm implemented with forward search. Interestingly, both Sim-1 and Sim-2 have included all 10 cognitive assessment features. The most impressive case is however, the one returned by Sim-3, which consists of only nine features with seven of which being cognitive assessment factors; it beats the benchmark with a large margin. These observations collectively reflect the effectiveness of cognitive features in working with a variety of classifiers.

Showing a rather different trend, Table 5 presents the performance of features subsets returned through backwards search. It can be seen that these feature subsets significantly underperform in comparison to those attained by forward search. Despite there are more features included in each of the selected subset via the backward approach, fewer cognitive features are chosen. As such, the lack of cognitive assessment factors has remarkably deteriorated the resulting classification performance. Particularly, Sim-3 has been the top performer amongst

the alternatives with nine features when forward search is applied; now it only performs slightly better than random guess with around 20 features (but none of which are cognitive assessment factors).

Table 5. Performance comparison (cont'd) on accuracy (%) and training time (s)

	No FS		Backward+Sim-1		Backward+Sim-2		Backward+Sim-3	
Algorithm	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**
NB	82.66 ± 3.91	0.01	73.82 ± 4.24	0	62.68 ± 4.74	0	40.99 ± 3.76	0
LR	82.67 ± 3.81	0.86	78.75 ± 3.85	0.69	61.35 ± 4.16	0.54	42.44 ± 2.91	0.51
MLP	76.11 ± 4.34	4.66	71.27 ± 4.81	2.56	52.41 ± 5.03	2.08	36 ± 5.34	1.79
IBK-3	48.32 ± 4.97	0	46.22 ± 4.93	0	37.79 ± 4.4	0	33.98 ± 4	0
JRip	89.39 ± 3.08	0.17	83.31 ± 3.72	0.16	62.12 ± 4.5	0.19	42.52 ± 3.23	0.16
RF	86.53 ± 3.75	0.08	79.15 ± 3.61	0.08	60.25 ± 5.12	0.09	37.27 ± 4.56	0.1
C4.5	88.88 ± 3.29	0.04	82.2 ± 3.76	0.03	58 ± 4.84	0.04	34.43 ± 4.18	0.04
SVM	79.98 ± 3.59	0.19	75.34 ± 4.26	0.17	60.84 ± 4.75	0.21	42.32 ± 0.42	0.16
Average	79.318	0.751	73.758	0.461	56.930	0.394	38.744	0.345

Examining the results more closely, with just four cognitive features selected using the similarity function Sim-1, better classification performance can be achieved than using the subset returned by Sim-2 that involves just 2 cognitive features, whilst the latter beats the use of those features selected through the use of Sim-3 that does not include any cognitive features. Once again, these results demonstrate the effectiveness of cognitive features in offering discriminating power in support of the diagnosis of AD. From the perspective of runtime overheads, it is not surprising to observe that the training time required is reduced using a smaller feature subset, as consistently shown across all experiments. This brings forward an additional benefit of utilising FRFS in assessing the significance of cognitive assessment features.

Overall, the experimental investigations carried out so far demonstrate the effectiveness and stability of FRFS, implemented with the forward searching approach. This conforms to the findings of applying it to other problem domains (e.g., [16] for an engineering application), because it always aims to select the most informative features first (albeit one at a time). However, for FRFS with backward search, its working is not only affected by the order of the original features listed, the sequential test and removal strategy applied does not guarantee most effective features to be retained in the final subset (even though the resultant feature subset is able to achieve the equally best dependency measure as that produced by the version with forward search). Very importantly, the above experimental studies also highlight the ineffectiveness of those features expressed in terms of medical histories and personal demographics. This may be because such features are indicative of risk factors but not necessarily relating to AD. However, the significance of cognitive assessment factors is clearly shown: A feature subset of just nine factors selected contains seven cognitive ones, while it leads to robust and consistent performance, with best results achieved for seven out of eight learning classifiers examined.

5 Conclusion

In order to identify the significance of cognitive assessment factors in support of performing dementia diagnosis, which are non-invasive, time-saving and eco-

nomical, this paper has presented an initial exploring application of Fuzzy-Rough Feature Selection, by varying the similarity functions and search strategies that it may employ. It has been empirically shown that the use of different similarity functions and/or search strategies will result in different feature subsets, which in turn, may play different roles while working in conjunction with different learning classifiers for dementia diagnosis. Particularly, experimental studies have demonstrated the significance of using cognitive assessment features as informative features for the decision-making tasks concerned.

Whilst promising, much can be done to further improve this work. For instance, it would be interesting to examine in more details about the potential of involving multiple cognitive assessment factors in an effort to generate a standardised assessment mechanism that would reinforce their discriminating capability for practical use. Another significant piece of future work is to exploit advanced knowledge interpolation techniques [17], in an attempt to work with missing values that commonly exist in many real-world problem domains, including the problem of dementia diagnosis.

6 Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This work was partly supported by grants from the National Natural Science Foundation of China (No. 61906181 and 72001032), and partly by the Strategic Partner Acceleration Award (80761- AU201) under the Sêr Cymru II programme, UK.

References

1. World Health Organisation, "Dementia," https://www.who.int/news-room/fact-sheets/detail/dementia, 2019.
2. H. Kour, J. Manhas, and V. Sharma, "Evaluation of Adaptive Neuro-Fuzzy Inference System with Artificial Neural Network and Fuzzy Logic in Diagnosis of Alzheimer Disease," in *6th International Conference on Computing for Sustainable Global Development*. IEEE, 2019, pp. 1041–1046.
3. D. W. Young, "what does an mri scan cost?" *Healthcare Financial Management*, vol. 69, no. 11, pp. 46–49, 11 2015.
4. NHS, "Guide to NHS waiting times in England - NHS," 2019. [Online]. Available: https://www.nhs.uk/using-the-nhs/nhs-services/hospitals/guide-to-nhs-waiting-times-in-england/
5. F. Er, P. Iscen, S. Sahin, N. Çinar, S. Karsidag, and D. Goularas, "Distinguishing age-related cognitive decline from dementias: A study based on machine learning algorithms," *Journal of Clinical Neuroscience*, 2017.
6. F. Zhu, X. Li, D. McGonigle, H. Tang, Z. He, C. Zhang, G. U. Hung, P. Y. Chiu, and W. Zhou, "Analyze Informant-Based Questionnaire for the Early Diagnosis of Senile Dementia Using Deep Learning," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, 2020.
7. G. G. C. Lee, P. W. Huang, Y. R. Xie, and M. C. Pai, "Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Normal Based on Neuropsychological Data via Supervised Learning," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, oct 2019, pp. 1808–1812.
8. T. Chen, C. Shang, P. Su, E. Keravnou-Papailiou, Y. Zhao, G. Antoniou, and Q. Shen, "A decision tree-initialised neuro-fuzzy approach for clinical decision support," *Artificial Intelligence in Medicine*, vol. 111, p. 101986, 2020.
9. T. Chen, G. Antoniou, M. Adamou, I. Tachmazidis, and P. Su, "Automatic diagnosis of attention deficit hyperactivity disorder using machine learning," *Applied Artificial Intelligence*, pp. 1–13, 2021.
10. R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 4, pp. 824–838, 2009.
11. ——, "Fuzzy-rough sets assisted attribute selection," *Fuzzy Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 73–89, 2007.
12. A. Chouchoulas and Q. Shen, "Rough set-aided keyword reduction for text categorization," *Applied Artificial Intelligence*, vol. 15, no. 9, pp. 843–873, 2001.
13. B. Sheehan, "Assessment scales in dementia," *Therapeutic Advances in Neurological Disorders*, vol. 5, no. 6, pp. 349–358, 2012.
14. T. Chen, C. Shang, P. Su, and Q. Shen, "Induction of accurate and interpretable fuzzy rules from preliminary crisp representation," *Knowledge-Based Systems*, vol. 146, pp. 152–166, 2018.
15. M. A. H. Eibe Frank and I. H. Witten, *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Fourth Edition*. Morgan Kaufmann, 2016.
16. C. Shang, D. Barnes, and Q. Shen, "Facilitating efficient mars terrain image classification with fuzzy-rough feature selection," *International Journal of Hybrid Intelligent Systems*, vol. 8, no. 1, pp. 3–13, 2011.
17. T. Chen, C. Shang, J. Yang, F. Li, and Q. Shen, "A new approach for transformation-based fuzzy rule interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 12, pp. 3330–3344, 2020.