Rough Set Rules (RSR) Predominantly Based on Cognitive Tests Can Predict Alzheimer's Related Dementia

Andrzej W. Przybyszewski[1,2(✉)], Kamila Bojakowska[3], Jerzy P. Nowacki[1], Aldona Drabik[1], and BIOCARD Study Team

[1] Polish-Japanese Academy of Information Technology, 02-008 Warszawa, Poland
{przy,jerzy.nowacki,drabik}@pjwstk.edu.pl
[2] Department of Neurology, UMass Medical School, Worcester, MA 01655, USA
[3] Department of Neurology, Central Clinical Hospital MSWiA, 02-507 Warszawa, Poland
kama.435@wp.pl

Abstract. Technology progress helped us to live better and longer, but aging is the major factor related to ND (neurodegenerative diseases) such as Alzheimer's or Parkinson's disease. Alzheimer's disease (AD) correlated neurodegenerative processes begin over 30 years, whereas cognitive changes begin about 15–11 years, before the first AD symptoms.

The purpose of our study was to predict if cognitive 'healthy' subjects might get AD dementia soon. We have analyzed Biocard data from the project that started with 349 normal subjects were followed over 20 years with over 150 different attributes. Subjects were evaluated every year by neurologists with the global score CDR (Clinical Dementia Rating) parameters to determine if a particular individual is normal, has Mild Cognitive Impairment, or has dementia. We have used classification based on CDRSUM (sum of boxes) as a more precise and quantitative general index than the global score to provide more information on patients with mild dementia. CDRSUM values for prodromal patients are: 0.0 normal; (0.5–4.0): questionable cognitive impairment; (0.5–2.5): questionable impairment; (3.0–4.0) very mild dementia; (4.5–9.0) – mild dementia. We have obtained rough set rules (RSR) from Model1: 149 patients classified as AD, MCI, and normal; and Model2: 40 patients with AD. By using Model1 classified by neurologists as 21 normal (CDR = 0) subjects, with our classification based on RSR, we have obtained 8 subjects with CDRSUM > 0: all 8 subjects were above 0.75, one subject between 0.75 and 1.25, and 5 subjects between 0.75 and 2.25, and two subjects were above 2.25. These subjects might have questionable cognitive impairment. Using Model2 we found with RSR that two subjects had CDRSUM between 4.5 and 6.5, which means they might have mild dementia (4.5–9.0). RSR consist of algorithms that might predict future cognitive AD-related impairments in individual, normal, healthy subjects.

Keywords: Neurodegeneration · Rough set theory · Intelligent predictions

N. T. Nguyen et al. (Eds.): ACIIDS 2022, LNAI 13757, pp. 129–141, 2022.
https://doi.org/10.1007/978-3-031-21743-2_11

1 Introduction

Cognitive changes are dominating in the most common neurodegenerative disease (ND) - Alzheimer's disease (AD). In most cases of AD, neurodegeneration starts in the hippocampus and frontal cortex, and it is related to memory and orientation problems. With the disease progression, other brain regions become also affected.

As each patient has dissimilar neurodegeneration development and compensation in consequence symptoms might be various and finding optimal treatment is an art for an experienced neurologist.

We have estimated disease progression with sets of psychophysical attributes found as the most meaningful in patients from the BIOCARD study [1, 2] and combined them with the results of the APOE.

The risk of AD increases and the age-at-onset decreases with the number of APOE4 alleles [3, 4]. A single APOE4 allele increases risk 2–4 fold and having two APOE4 alleles increases the risk about 8–12 fold [4]. The APOE4 allele also drives the age-at-onset down, APOE4 carriers are, on average, about 12 years younger than non-carriers [3, 4] However having a single APOE allele ε2 reduces the risk AD by about 40%, and being homozygous for APOE ε2 reduces the risk even more APOE2 homozygotes have a 66% reduction in AD risk compared to +2/+3 carriers, an 87% reduction in AD risk compared to APOE3 homozygotes, and a 99.6% reduction in AD risk when compared to APOE4 homozygotes [5].

This study is the continuation of the rough set theory application to follow predominantly the cognitive changes in neurodegenerative diseases (ND) such as Parkinson's [6] and now Alzheimer's diseases.

2 Methods

We have analyzed data from normal subjects (N), Mild Cognitive Impairment (MCI), and Alzheimer Disease (AD) patients divided into three main groups:

- **Model1** consists of 149 subjects with 40 normal (N), 40 AD, and 69 MCI
- **Model2** consists of 40 AD patients.
- **TestGr** consists of 21 normal (N) subjects

All subjects had the following neuropsychological tests performed every year: Logical Memory Immediate (LOGMEM1A), Logical Memory Delayed (LOGMEM2A), Trail Making, Part A (TrailA - connecting time in sec of randomly placed numbers), Trail Making Part B (TrailB - connecting time in a sec of randomly placed numbers and letters), Digit Symbol Substitution Test (DSST), Verbal Fluency Letter F (FCORR), Rey Figure Recall (REYRECAL), Paired Associate Immediate (PAIRED1), Paired Associate Delayed (PAIRED2), Boston Naming Test (BOSTON). In addition, we have subjects' age (years), APOE genotype; individuals who are *ApoE-4* carriers vs. non-carriers (digitized as 1 vs. 0), and CDRSUM (sum of boxes) as a precise and quantitative general index of the Clinical Dementia Rating [7].

*Data used in preparation of this article were derived from BIOCARD study, supported by grant U19 – AG033655 from the National Institute on Aging. The BIOCARD

study team did not participate in the analysis or writing of this report, however, they contributed to the design and implementation of the study. A listing of BIOCARD investigators can be found on the BIOCARD website (on the 'BIOCARD Data Access Procedures' page, 'Acknowledgement Agreement' document).

2.1 Theoretical Basis

Our analysis was performed with help of rough set theory (RST) invented by Zdzislaw Pawlak [8]).

In the standard RST procedure, our data were inserted in the decision table with rows that stand out the actual attributes' values for the different or for the same subject, and columns were linked to diverse attributes. Following [8] an information system is a pair $S = (U, A)$, where U, and A are nonempty finite sets. The set U is the universe of objects, and A is the set of attributes. If $a \in A$ and $u \in U$, the value $a(u)$ is a unique element of V (where V is a value set). The *indiscernibility relation IND(B)* of any subset B of A is defined after [8]: $(x, y) \in IND(B)$ *iff* $a(x) = a(y)$ for every $a \in B$ where the value of $a(x) \in V$. This relation divides A into *elementary granules* and it is the basis of RST. In the information system S set $B \subset A$ of is a reduct if $IND\ (B) = IND\ (A)$ and it cannot be further reduced. Other important RST properties such as *lower approximation* and *upper approximation* were defined and discussed in [8, 9].

In this work, we have used different intelligent algorithms implemented in RSES 2.2 software such as: exhaustive algorithm, genetic algorithm [10], covering algorithm, or LEM2 algorithm [11].

In addition to the classical [8] information system is its extension of the decision table to a triplet: $S = (U, C, D)$ where the set of attributes A is divided into C *as* condition and D as decision attributes [12]. In a single row, there are many conditions and only one decision attribute, all related to specific tests of the individual subject or patient. Names and values of classification attributes related to the value of the decision attribute give a unique rule. One difficult problem in the medical field is related to contradictory measurements (results). In this case, doctors often are using averaging techniques. RSR are considering and solving problems with contradictory rules, but the main feature of RSR is to generalize individual measurements (rules) and universal principles (knowledge) but rules have different confidence. There are always true rules related to the *lower approximation set* and rules that are only partly true associated with the *upper approximation set*. The difference between upper and lower approximation sets is the called border set. If this set is non-empty, it is related to uncertain rules.

We have based our approach on the mechanisms in the visual brain related to advanced processes of complex objects' recognition [13]. The processes in the higher visual brain areas that are related to different objects classification are using RSR to find upper and lower approximations of the retinal image [13]. These approximations are compared with the different objects' models (images) saved in the visual cortex. In the next steps of the object recognition (classification) lower visual areas are tunned to extract the properties of the chosen Model (the difference between upper and lower approximations becomes smaller). If the border set becomes empty, we recognize the object. We use this approach by proposing different Models to approximate the actual (future) cognitive state of tested normal subjects.

We have used RSR determined by RSES 2.2 [14] which generalizes rules from the decision table to process different patterns related to an individual patient sets of measurements. In our previous publication, related to Parkinson's disease patients, we demonstrated that the rough set theory application provides better results than other ML methods [6].

3 Results

As described above in the Methods section we had three groups of subjects: Model1 (149 subjects), Model2 (40 AD patients) and TestGr - test group (21 subjects).

3.1 Statistical Results

The subjects from group Model1 n = 149, Model2 n = 40, TestGr n = 21.

Table 1. Statistics of our data

P#	Age	Lgm1A	Lgm2A	TrailA	TrailB	DSST	Fcorr	CDRSUM
Model1	76.4 ± 8.6	16.1 ± 4.6	15.2 ± 5.3	39.6 ± 20	99.9 ± 59	46.5 ± 14	15.4 ± 5	1.3 ± 1.8
Model2	78.5 ± 12	12.2 ± 4.4	10.1 ± 5.2	5.03 ± 30	151 ± 78	37.3 ± 13	12.6 ± 6	3.5 ± 2.5
TestGr	76.6 ± 8.4	18.0 ± 3.0	17.5 ± 3.9	30.7 ± .12	64.2 ± 21	57 ± 13 1	8.2 ± 5	0 ± 0

Table 1 presents the statistical calculations for the Model1 and Model2 test normal subjects TestGr as mean ± SD. The age of subjects in different groups is similar, but other parameters show differences: Lgm1A (LOGMEM1A) is smallest for AD patients and largest for N, Lgm2A (LOGMEM2A) has similar changes as Lgm1A, execution functions: TrailA and TrailB are growing from N to AD, DSST is decreasing from N to AD in a similar way as Fcorr (FCORR). We did not show other parameters because of a lack of space. The changes in the CDRSUM are obvious as in normal subjects its values are 0 and significantly larger for AD patients. There are large differences between the values of individual subjects, so the mean values were not statistically significant.

3.2 RSR for Reference of Model1 Group

We have placed Model1 data in the following information table (Table 2):

The complete Table 2 has 149 rows, and 14 columns, there are shown the following condition attributes: P# - the number given to each patient, age –age of the subject, Lgm1A -Logical Memory Immediate, Lgm2A - Logical Memory Delayed, TrailA - Trail Making Part A, TrailB -Trail Making Part B, DSST - Digit Symbol Substitution

Table 2. Part of the decision table for Model1 subjects

P#	Age	Lgm1A	Lgm2A	Trail A	Trail B	DSST	Gcorr	Reycrl	APOE	…	CDRSUM
67643	74	9	8	40	208	35	14	18	1	…	0.5
70407	88	8	5	66	150	21	21	10	0	…	4.6
102541	71	15	25	25	202	52	52	23.5	0	…	1
119156	92	7	34	34	386	40	40	10.5	0	…	3.5
139134	81	6	51	51	60	49	49	6	1	…	2.5
142376	76	18	54	54	50	19	19	12	0	…	0

Table 3. Discretized table extract for above (Table 1) Model1 subjects

P#	Age	Lgm1A	Lgm2B	TrailA	TrailB	…APOE…	CDRSUM
67643	"(73.5,86.5)"	"(−5.5,16.5)"	"(5.5,16.5)"	"(31.5,48.0)"	"(137.5,Inf)"	…1…	"(−Inf,0.75)"
70407	"(86.5,Inf)"	"(−Inf,10.0)"	"(1.0,5.5)"	"(48.0,143.0)"	"(137.5,Inf)"	…0…	"(1.25,Inf)"
102541	"(−Inf,73.5)"	"(12.5,16.0)"	"(5.5,16.5)"	"(23.5,28.5)"	"(78.0,114.0)"	…0…	"(0.75,1.25)"
119156	"(86.5,Inf)"	"(−Inf,10.0)"	"(5.5,16.5)"	"(31.5,48.0)"	"(78.0,114.0)"	…0…	"(1.25,Inf)"
139134	"(73.5,86.5)"	"(−Inf,10.0)"	"(1.0,5.5)"	"(48.0,143.0)"	"(78.0,114.0)"	…1…	"(1.25,Inf)"
142376	"(73.5,86.5)"	"(16.0,20.5)"	"(5.5,16.5)"	"(48.0,143.0)"	"(137.5,Inf)"	…0…	"(−Inf,0.75)"

Test, Fcorr - Verbal Fluency Letter F, Reyrcl - Rey Figure Recall, APOE - *ApoE* genotype, … CDRSUM – the sum of boxes - index of the Clinical Dementia Rating.

Table 3 is a discretized table for six patients: 67643 to 142376. Significant condition attributes were age, Lgm1A (LOGMEM1A), Lgm2A (LOGMEM2A), TrailA, TrailB, APOE, and others not shown in Table 2 like DSST, Fcorr (FCORR), Reyrcl (REYRECA), PAIRED1, BOSTON. Not significant was: PAIRED2.

We have used RSES 2.2 for Model1 group discretization with the local cuts [RSES]. There were the following 3 ranges of the decision attribute CDRSUM: "(−Inf, 0.75)", "(0.75, 1.25)", "(1.25, Inf)".

We obtained 2581 rules using the exhaustive algorithm for Model1 subjects, and as an example, we present below 10 rules filled by the most cases:

$$(DSST = "(50.5, Inf)")\&(REYRECAL = "(15.75, 25.75)")\&(age = "(73.5, 86.5)") => (CDRSUM = "(-Inf, 0.75)"[16])16 \quad (1)$$

$$(LOGMEM1A = "(21.5, Inf)")\&(BOSTON = "(29.5, Inf)") => (CDRSUM = "(-Inf, 0.75)"[16])16 \quad (2)$$

$$(LOGMEMA = "(1.0, 5.5)") => (CDRSUM = "(1.25, Inf)"[8])8 \quad (3)$$

$(LOGMEM2A = "(5.5, 16.5)")\&(TRAILA = "(31.5, 48.0)")\&(REYRECAL$

$$= \text{"}(Inf, 15.75)\text{"})\&(BOSTON = \text{"}(-Inf, 26.5)\text{"}) => \Big(CDRSUM = \text{"}(1.25, Inf)\text{"}[7]\Big)7 \tag{4}$$

$$(TRAILA = \text{"}(31.5, 48.0)\text{"})\&(DSST = \text{"}(-Inf, 47.5)\text{"})\& (REYRECAL = \text{"}(-Inf, 15.75)\text{"})\&(BOSTON = \text{"}(-Inf, 26.5)\text{"}) => (CDRSUM = \text{"}(1.25, Inf)\text{"}[7])7 \tag{5}$$

$$(FCORR = \text{"}(-Inf, 10.5)\text{"})\&(REYRECAL = \text{"}(-Inf, 15.75)\text{"})\&(APOE = 1) => (CDRSUM = \text{"}(1.25, Inf)\text{"}[7])7 \tag{6}$$

$$(TRAILB = \text{"}(114.0, 137.5)\text{"})\&(APOE = 1) => (CDRSUM = \text{"}(1.25, Inf)\text{"}[6])6 \tag{7}$$

$$(LOGMEMIA = \text{"}(16.0, 20.5)\text{"})\&(BOSTON = \text{"}(-Inf, 26.5)\text{"})\& (age = \text{"}(73.5, 86.5)\text{"}) => (CDRSUM = \text{"}(0.75, 1.25)\text{"}[5])5 \tag{8}$$

$$(TRAILB = \text{"}(51.0, 78.0)\text{"})\&(BOSTON = \text{"}(-Inf, 26.5)\text{"})\& (age = \text{"}(73.5, 86.5)\text{"}) => (CDRSUM = \text{"}(0.75, 1.25)\text{"}[5])5 \tag{9}$$

There is the following interpretation of the above equations: Eq. 1 claims for 16 cases that if DSST is above 50.5 and REYRECAL is between 15.75 and 25.75 and the patient's age is between 73.5 and 86.5 years then CDRSUM is below 0.75 that means questionable impairment. The Eq. 3 if LOGMEM2A (Logical Memory Delayed) is between 1.0 and 5.5 then CDRSUM is larger than 1.25 which means that patient has the least questionable impairment (8 cases). Equation 10 states that if TrailB is between 51 and 78 s, and Boston naming test result is below 25.5, and the patient's age is between 73.5 and 86.5 years then CDRSUM is between 0.75 and 1.25, which means that patient has a questionable impairment, and it is fulfilled in 5 cases.

We have used the above general rules from Model1 to predict CDRSUM of the TestGr.

Table 4. Confusion matrix for CDRSUM of TestGr group by rules obtained from Model1 by *local cuts* [13]

Predicted				
Actual	"(−Inf, 0.75)"	"(1.25, Inf)"	"(0.75, 1.25)"	ACC
"(−Inf, 0.75)"	17. 0	2.0	2.0	0.81
"(1.25, Inf)"	0.0	0.0	0.0	0.0
"(0.75, 1.25)"	0.0	0.0	0.0	0.0
TPR	1.0	0.0	0.0	0.0

TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1.0 and the global accuracy was 0.81, the coverage for decision classes was 1.0, 0.0, 0.0.

Table 4 results are that for the 17 patients' prediction and tests results agreed. In two patients' predictions gave CDRSUM values above 1.25 and in two other patients, they were between 0.57 and 1.25 where tests results were 0 (below 0.75).

We were interested in those normal subjects who had predicted values of the CDRSUM > 0. From Table 4 there were two subjects with predicted values of CDRSUM = (0.75, 1.25)

$$(Pat = 284424)\&(LOGMEM\,1A = \text{"(16.0, 20.5)"}))\& (LOGMEM\,2A = \text{"(5.5, 16.5)"})\&(TRAILA = \text{"(48.0, 143.0)"})\& (TRAILB = \text{"(78.0, 114.0)"})\&(DSST = \text{"}(-Inf\,.47.5)\text{"})\& (FCORR = \text{"(21.5, }Inf\text{)"})\&(REYRECAL = \text{"(15.75, 25.75)"})\&(PAIRDI = \text{"(12.0, 21.5)"})\&(BOSTON = \text{"(26.5, 29.5)"})\&(age = \text{"(86.5, }Inf\text{)"}) \&(APOE = 0) => (CDRSUM = \text{"(0.75, 1.25)"} \quad (10)$$

$$(Pat = 558865)\&(LOGMEM\,1A = \text{"(16.0, 20.5)"}))\&(LOGMEM\,2A = \text{"(5.5, 16.5)"})\&(TRAILA = \text{"(31.5, 48.0)"})\&(TRAILB = \text{"(78.0, 114.0)"})\& (DSST = \text{"}(-Inf\,, 47.5)\text{"})\&(FCORR = \text{"(21.5, }Inf\text{)"})\&(REYRECAL = \text{"(15.75, 25.75)"})\& (PAIRDI = \text{"(12.0, 21.5)"})(BOSTON = \text{"(26.5, 29.5)"})\&(age = \text{"(73.5, 86.5)"})\& (APOE = 1) => (CDSUM = \text{"(0.75, 1.25)"} \quad (11)$$

In both patients, bad executive functions (TrialA and TrialB) play a significant role in the possible questionable impairment [7].

From Table 4 there were two subjects with predicted values of CDRSUM = (1.25, Inf)):

$$(Pat = 164087)\&(LOGMEM\,1A = \text{"(12.5, 16.5)"}))\&(LOGMEM\,2A = \text{"(5.5, 16.5)"}\&(TRAILA = \text{"(34.5, 42.0)"})\&(TRAILB = \text{"(83.0, 114.0)"})\& (DSST = \text{"(47.5, 50.5)"})\&FCORR = \text{"(10.5, }Inf\text{)"})\&(REYRECAL = \text{"(15.75, 26.5)"})\&(PAIRDI = \text{"(12.0, 21.5)"})\&(BOSTON = \text{"}(-Inf\,, 26.5)\text{"})\& (age = \text{"(73.0, 76.5)"})\& (APOE = 1) => (CDRSUM = \text{"(1.25, }Inf\text{))"} \quad (12)$$

$$(Pat = 401297)\&(LOGMEM\,1A = \text{"(12.5, 16.5)"})\&(LOGMEM\,2A = \text{"(5.5, 16.5)"}\&(TRAILA = \text{"}(-Inf\,, 21.0)\text{"})\&(TRAILB = \text{"}(-Inf\,, 50.0)\text{"})\& (DSST = \text{"(69.0, }Inf\text{)"}\&(FCORR = \text{"(10.5, }Inf\text{)"})\& (REYRECAL = \text{"}(-Inf\,, 15.75)\text{"})\&(PAIRDI = \text{"(12.0, 21.5)"})\& (BOSTON = \text{"(26.5, }Inf\text{)"})\text{"})\&(age = \text{"}(-Inf\,, 73.0)\text{"})\& (APOE = 1) => (CDRSUM = \text{"(1.25, }Inf\text{)"} \quad (13)$$

The first patient Eq. 13 has affected executive functions (TrialA and B), and the second patient (Eq. 14) has good executive functions, but bad the Rey Figure Recall REYRECAL.

Table 5. Confusion matrix for CDRSUM of TestGr group by rules obtained from Model1 by the *global cuts* [14]

Predicted				
Actual	"(−Inf, 0.75)"	"(2.25, Inf)"	"(0.75, 1.25)"	ACC
"(−Inf, 0.75)"	15. 0	2.0	4.0	0.71
"(2.25, Inf)"	0.0	0.0	0.0	0.0
"(0.75, 2.25)"	0.0	0.0	0.0	0.0
TPR	1.0	0.0	0.0	0.0

TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1.0 and the global accuracy was 0.714, the coverage for decision classes was 1.0, 0.0, 0.0.

In Table 5 we have also used RSES 2.2 for Model1 group discretization with the *global cuts* [14]. There were the following 3 ranges of the decision attribute CDRSUM: "(−Inf, 0.75)", "(0.75, 2.25)", "(2.25, Inf)". We have obtained 776 rules with the genetic algorithm for Model1 subjects. After removing rules related to single support cases, we have got 324 rules that in the confusion matrix gave only three subjects with CDRSUM = (0.75, 2.25) – we found the excluded subject and marked it below, but the same two subjects with CDRSUM = (2.25, Inf).

As above, we were interested in those normal subjects who had predicted values of the CDRSUM > 0. From Table 3 there were four subjects with CDRSUM = (0.75, 2.25) that with values between (0.5–2.5) might have a questionable impairment [7]:

$$(Pat = 204670)\&(LOGMEM\,1A = \text{"(15.5, 20.5)"})\&(LOGMEM\,2A = \text{"(16.5, } Inf\text{)"})\&(TRAILA = \text{"(}-Inf\text{, 23.5)"})\&(TRAILB = \text{"(}-Inf\text{, 74.5)"})\&(FCORR = \text{"(16.5, } Inf\text{)"})\&(REYRECAL = \text{"(25.25, } Inf\text{)"})\&(PAIRD2 = \text{"(6.5, } Inf\text{)"})\&(age = \text{"(}-Inf\text{, 76.5)"})\&(APOE = 1) => (CDRSUM = \text{"(0.75, 2.25)"} \quad (14)$$

$$(Pat = 463437)\&(LOGMEM\,1A = \text{"(15.5, 20.5)"}))\&(LOGMEM\,2A = \text{"(16.5, } Inf\text{)"})\&TRAILA = \text{"(23.5, 35.5)"})\&(TRAILB = \text{"(74.5, 153.0)"})\&(FCORR = \text{"(}-Inf\text{, 16.5)"})\&(REYRECAL = \text{"(15.75, 25.25)"})\&(PAIRD2 = \text{"(}-Inf\text{, 6.5)"})\&(age = \text{"(}-Inf\text{, 76.5)"})\&(APOE = 0) => (CDRSUM = \text{"(0.75, 2.25)"} \quad (15)$$

$$(Pat = 558865)\&(LOGMEM\,1A = \text{"(15.5, 20.5)"}))\&(LOGMEMA$$

$$= \text{"}(-Inf, 16.5)\text{"})\&(TRAILA = \text{"}(23.5, 35.5)\text{"})\&(TRAILB = \text{"}(74.5, 153.0)\text{"})\&$$
$$(FCORR = \text{"}(16.5, Inf)\text{"})\&(REYRECAL = \text{"}(15.75, 25.25)\text{"})\&$$
$$(PAIRD2 = \text{"}(-Inf, 6.5)\text{"})\&(age = \text{"}(76.5, Inf)\text{"})\&(APOE = 1)$$
$$=> (CDRSUM = \text{"}(0.75, 2.25)\text{"} \quad (16)$$

$$(Pat = 808698)\&(LOGMEM1A = \text{"}(15.5, 20.5)\text{"}))\&(LOGMEM2A$$
$$= \text{"}(16.5, Inf)\text{"}\&(TRAILA = \text{"}(-Inf, 23.5)\text{"})\&(TRAILB = \text{"}(-Inf, 74.5)\text{"}\&$$
$$(FCORR = \text{"}(-Inf, 16.5)\text{"})\&(REYRECAL = \text{"}(15.75, 25.25)\text{"})\&$$
$$(PAIRD2 = \text{"}(6.5, Inf)\text{"})\&(age = \text{"}(-Inf, 76.5)\text{"})\&(APOE = 1)$$
$$=> (CDRSUM = \text{"}(0.75, 2.25)\text{"} \quad (17)$$

As you may notice, rules for all 4 patients are similar, but with some significant differences, e.g., in Eqs. 16 and 17 both patients are slow in the executive actions; TrailA – connecting randomly place numbers, and TrialB - connecting numbers and letters – cognitive task. In the other two patients, it seems that the combination of different factors plays a major role in questionable impairment [7].

By using rules from Model1 group discretization with the *global cuts* [14] we got the following predictions:

$$(Pat = 284424)\&(LOGMEM1A = \text{"}(15.5, 20.5)\text{"}))\&(LOGMEM2A$$
$$= \text{"}(-Inf, 16.5)\text{"})\&(TRAILA = \text{"}(35.5, Inf)\text{"})\&(TRAILB = \text{"}(74.5, 153.0)\text{"})\&$$
$$(FCORR = \text{"}(16.5, Inf)\text{"})\&(REYRECAL = \text{"}(15.75, 25.25)\text{"})\&$$
$$(PAIRD2 = \text{"}(6.5, Inf)\text{"})\&(age = \text{"}(76.5, Inf)\text{"})\&(APOE = 0)$$
$$=> (CDRSUM = \text{"}(0.75, 2.25)\text{"} \quad (18)$$

In our new Model1 with global cuts there were two patients were classified as above *Pat = 204670 and Pat = 808698,* but we got a new patient *Pat = 284424* that we have not seen above (Eq. 19). He/she has problems with the executive actions (TrailA and B).

From Table 5 there were two subjects with CDRSUM = (2.25, Inf)):

$$(Pat = 164087)\&(LOGMEM1A = \text{"}(-Inf, 15.5)\text{"}))\&(LOGMEM2A$$
$$= \text{"}(-Inf, 16.5)\text{"})\&(TRAILA = \text{"}(35.5, Inf)\text{"})\&(TRAILB = \text{"}(74.5, 153.0)\text{"})\&$$
$$(FCORR = \text{"}(-Inf, 16.5)\text{"})\&(REYRECAL = \text{"}(15.75, 25.25)\text{"})\&(PAIRD2$$
$$= \text{"}(-Inf, 6.5)\text{"})\&(age = \text{"}(-Inf, 76.5)\text{"})\&(APOE = 1)$$
$$=> (CDRSUM = \text{"}(2.25, Inf)\text{"} \quad (19)$$

$$(\text{ Pat } = 401297)\&(\text{ LOGMEM1A } = \text{"}(-\text{Inf, } 15.5)\text{"})\&$$
$$(\text{ LOGMEM2A } = \text{"}(-\text{Inf, } 16.5)\text{"})\&(\text{ TRAILA } = \text{"}(35.5, \text{ Inf})\text{"})\&$$
$$(\text{ TRAILB } = \text{"}(74.5, 153.0)\text{"})\&(FCORR = \text{"}(-\text{Inf, } 16.5)\text{"})\&$$
$$(\text{ REYRECAL } = \text{"}(15.75, 25.25)\text{"})\&(\text{ PAIRD } = \text{"}(-\text{Inf, } 6.5)\text{"})\&$$
$$(\text{ age } = \text{"}(-\text{Inf, } 76.5)\text{"})\&$$

$$(APOE = 1) => (CDRSUM = \text{"}(2.25,\ \text{Inf})\text{"} \tag{20}$$

Notice that there are the same patients that were classified above (Eqs. 13, 14) with CDRSUM = (1.25, Inf). Parameters in the above equations (Eqs. 20, 21) have different values than in Eqs. 13, 14 but the interpretation is similar. The first patient *(Pat = 164087)* has affected executive functions (TrialA and B), especially TrialB related to the cognitive impairments. The second patient *(Pat = 401297)* has good (short times executions) executive functions, but bad the Rey Figure Recall REYRECAL.

3.3 RSR for Reference of Model2 Group

We have placed Model2 data in the following information table (Table 4):

Table 6. Part of the decision table for Model2 patients

P#	age	Lgm1A	Lgm2A	TrailA	TrailB	DSST	Fcorr	Reyrcl	APOE	…	CDRSUM
70407	88	8	5	66	150	21	21	10	0	…	4.5
119156	92	7	34	34	386	40	20	10.5	0	…	3.5
155699	94	9	3	76	119	14	9	7	1	…	4
265499	91	13	8	54	239	27	6	7	0	…	5
268713	79	10	5	29	189	32	12	11	1	…	3.5
299967	69	17	17	42	82	48	18	13	1	…	2

The complete Table 6 has 40 rows, and 14 columns, there are shown the following condition attributes: P# - the number given to each patient, age –age of the subject, Lgm1A -Logical Memory Immediate, Lgm2A - Logical Memory Delayed, TrailA - Trail Making Part A, TrailB -Trail Making Part B, DSST - Digit Symbol Substitution Test, Fcorr -Verbal Fluency Letter F, Reyrcl - Rey Figure Recall, APOE - *ApoE* genotype, … CDRSUM – the sum of boxes- index of the Clinical Dementia Rating.

Table 7. Discretized table extract for above (Table 6) Model2 AD patients

P#	Age	Lgm1A	TrailA	TrailB	DSST	Fcorr	Recyrcl	APOE	…	CDRSUM
70407	*	"(−Inf,11.5)"	*	*	"(−Inf,39.5)"	"(20.5,Inf)"	*	*	…	"(4.5,6.5)"
119156	*	"(−Inf,11.5)"	*	*	"(39.5,Inf)"	"(16.5,20.5)"	*	*	…	"(−Inf,4.5)"
155699	*	"(−Inf,11.5)"	*	*	"(−Inf,39.5)"	"(−Inf,12.5)"	*	*	…	"(−Inf,4.5)"
265499	*	"(11.5,13.5)"	*	*	"(−Inf,39.5)"	"(−Inf,12.5)"	*	*	…	"(4.5,6.5)"
268713	*	"(−Inf,11.5)"	*	*	"(−Inf,39.5)"	"(Inf,12.5)"	*	*	…	"(−Inf,4.5)"
299967	*	"(17.0,Inf)"	*	*	"(39.5,Inf)"	"(16.5,20.5)"	*	*	…	"(−Inf,4.5)"

Table 7 is a part of the discretized table for six (from all 40) AD patients: 70407 to 299967. Significant condition attributes were: Lgm1A (*L*OGMEM1A), DSST, Fcorr, and

others not shown in Table 7 like the BOSTON test. Not significant condition attributes were age, Lgm2A (LOGMEM2A), TrailA, TrailB, Reyrcl (REYRECAL), PAIRED1, PAIRED2, and APOE.

We had obtained 58 rules for Model2 subjects, and as an example, we present below 6 rules filled by the most cases:

$$(LOGMEM\,1A = \text{"}(17.0, Inf)\text{"}) => (CDRSUM = \text{"}(-Inf, 4.5)\text{"}[7])\,7 \tag{21}$$

$$\begin{aligned}&(FCORR = \text{"}(12.5, 16.5)\text{"})\&(BOSTON = \text{"}(24.5, Inf)\text{"})\\&=> (CDRSUM = \text{"}(-Inf, 4.5)\text{"}[5])\,5\end{aligned} \tag{22}$$

$$\begin{aligned}&(LOGMEM\,1A = \text{"}(-Inf, 11.5)\text{"})\&(DSST = \text{"}(39.5, Inf)\text{"})\\&\&(BOSTON = \text{"}(24.5, Inf)\text{"}) => (CDRSUM = \text{"}(-Inf, 4.5)\text{"}[4])\,4\end{aligned} \tag{23}$$

$$\begin{aligned}&(DSST = \text{"}(39.5, Inf)\text{"})\&(FCORR = \text{"}(20.5, Inf)\text{"})\\&=> (CDRSUM = \text{"}(-Inf, 4.5)\text{"}[2])\,2\end{aligned} \tag{24}$$

$$\begin{aligned}&(LOGMEM\,1A = \text{"}(13.5, 15.5)\text{"})\&(FCORR = \text{"}(12.5, 16.5)\text{"})\\&=> (CDRSUM = \text{"}(-Inf, 4.5)\text{"}[2])\,2\end{aligned} \tag{25}$$

$$\begin{aligned}&(LOGMEM\,1A = \text{"}(13.5, 15.5)\text{"})\&(FCORR = \text{"}(-Inf, 12.5)\text{"})\\&=> (CDRSUM = \text{"}(4.5, 6.5)\text{"}[2])\,2\end{aligned} \tag{26}$$

Interpretation: Eq. 22, if Logical Memory Immediate (LOGMEM1A) is above 17 then CDRSUM is below 4.5; Eq. 27 if Logical Memory Immediate (LOGMEM1A) is above between 13.5 and 15.1 and the Verbal Fluency (Letter F - FCORR) is poor (below 12.5) then CDRSUM is between 4.5 and 6.5, which means the mild dementia [7]. However, notice that this rule was fulfilled in only two cases in our 40 AD patients, whereas the rule Eq. 24 was fulfilled in 7 cases.

Table 8. Confusion matrix for CDRSUM of TestGr group by rules obtained from Model2-group

Predicted					
Actual	"(4.5,6.5)"	"(4.5,6.5)"	"(−Inf,4.5)"	"(6.5,Inf)"	ACC
	"(−Inf,4.5)"	0.0	0.0	0.0	0.0
	"(6.5,Inf)"	2.0	19.0	0.0	0.9
	"(6.5,Inf)"	0.0	0.0	0.0	0.0
	TPR	0.0	1.0	0.0	

TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1.0 and the global accuracy was 0.90, the coverage for decision classes was 0.0, 1.0, 0.0.

$$\begin{aligned}&(Pat = 164087)\&(LOGMEM\,1A = \text{"}(13.5, 15.5)\text{"})\&\\&(DSST = \text{"}(39.5, Inf)\text{"})\&(FCOR = \text{"}(-Inf, 12.5)\text{"})\&\\&(BOSTON = \text{"}(24.5, Inf)\text{"})\&(APOE = 1)\\&=> (CDRSUM = \text{"}(\text{"}(4.5,\ 6.5)\text{"}\end{aligned} \tag{27}$$

$$
\begin{aligned}
&(Pat = 776254)\&\,(LOGMEM\,1A = \text{"}(13.5, 15.5)\text{"})\& \\
&(DSST = \text{"}(39.5, Inf)\text{"})\&(FCOR = \text{"}(16.5, 20.5)\text{"})\& \\
&(BOSTON = \text{"}(24.5, Inf)\text{"})\&(APOE = 0) \\
&=> (CDRSUM = \text{"("}(4.5, 6.5)\text{"}
\end{aligned}
\tag{28}
$$

Results from Table 8 can be interpreted that from 21 'normal' subjects, 19 subjects have CDRSUM values below 4.5, which means that some of them may have very mild dementia or questionable impairment in the worst case [7]. However, two subjects had CDRSUM between 4.5 and 6, which means that they might have mild dementia [7]. The first patients, with his/her results described in Eq. 28 seems to be easier to interpret at least partly based on Eq. 27 with poor logical memory intermate and poor verbal fluency also with sensitive genetics with APOE = 1. It is more difficult to interpret Eq. 29.

4 Discussion

We have applied rough set theory and its rules (RSR) as the granular computing to estimate a possible disease progression in normal subjects from the BIOCARD study. We used intelligent granular computing with RSR to investigate test results set as granules for individual patients. To estimate their properties, we need to have a Model that has meaning and tells us what the importance of the pattern (granule) is. In fact, our granules are complex (c-granules) as they are changing their properties with the time of the neurodegeneration development till become like granules of patients with dementia or PD [15]. In this work, we have limited our test to the static granules (in one time moment) and we have tried to estimate what is the meaning of a particular, individual granule. We have used two models: Model1 has granules related to normal subjects, MCI, and AD patients. On this basis, we have obtained a large set of rules that have represented subjects' different stages of the disease from normal to dementia. We have tested several of such models mostly changing normal subjects and getting different rules, which we have applied to other normal subjects and estimated what 'normal' means. Also, rules can be created with different granularity and algorithms that might give different classifications. Therefore, we were looking for classifications that are universal e.g., they give similar results with different sets of rules. Model1 has given us rules that are subtle and determine the beginning of possible symptoms. In the next step, we used a more advanced model – Model2 that gave rules based on AD patients. We got higher values of the CDRSUM that gave us only classifications of the possible subjects with mild dementia. Looking into different rules, some of them are easy to interpret, but other patients' granules look relatively normal. As it is the first, to our knowledge, work that estimate distinctive complex pattern of individual patients' symptoms. These results open possibilities for early (preclinical) AD diagnosis based on neuropsychological testing that can be performed remotely.

References

1. Albert, M., et al.: the BIOCARD research team, cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to *ApoE* genotype. Curr. Alzheimer Res. **11**(8), 773–784 (2014)

2. Albert, M., et al.: Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years Brain. **141**(3), 877–887 (2018)
3. Corder, E.H., et al.: Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science **261**, 921–923 (1993)
4. Farrer, L.A., et al.: Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. JAMA **278**, 1349–1356 (1997)
5. Husain, M.A., Laurent, B., Plourde, M.: APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics REVIEW published, 17 February 2021. https://doi.org/10.3389/fnins.2021.630502
6. Przybyszewski, A.W., Kon, M., Szlufik, S., Szymanski, A., Koziorowski, D.M.: Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. Sensors **16**(9), 1498 (2016). https://doi.org/10.3390/s16091498
7. O'Bryant, S.E., Waring, S.C., Cullum, C.M., et al.: Staging dementia using clinical dementia rating scale sum of boxes scores: a Texas Alzheimer's research consortium study. Arch Neurol. **65**(8), 1091–1095 (2008)
8. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer, Dordrecht (1991)
9. Bazan, J., Nguyen, H.S., et al.: Desion rules synthesis for object classification. In: Orłowska, E. (ed.) Incomplete Information: Rough Set Analysis, pp. 23–57. Physica, Heidelberg (1998)
10. Bazan, J., Nguyen, H.S., et al.: Rough set algorithms in classification problem. In: Polkowski, L., Tsumoto, S., Lin, T. (eds.) Rough Set Methods and Applications, pp. 49–88. Physica, Heidelberg (2000)
11. Grzymała-Busse, J.: A new version of the rule induction system LERS. Fundamenta Informaticae **31**(1), 27–39 (1997)
12. Bazan, J.G., Szczuka, M.: The rough set exploration system. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets III. LNCS, vol. 3400, pp. 37–56. Springer, Heidelberg (2005). https://doi.org/10.1007/11427834_2
13. Przybyszewski, A.W.: The neurophysiological bases of cognitive computation using rough set theory. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 287–317. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89876-4_16
14. Bazan, J.G., Szczuka, M.: RSES and RSESlib - a collection of tools for rough set computations. In: Ziarko, W., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 106–113. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45554-X_12
15. Przybyszewski, A.W.: Parkinson's disease development prediction by C-granule computing. In: Nguyen, N.T., Chbeir, R., Exposito, E., Aniorté, P., Trawiński, B. (eds.) ICCCI 2019. LNCS (LNAI), vol. 11683, pp. 296–306. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28377-3_24