Diagnosis of Cognitive Impairment Compatible with Early Diagnosis of Alzheimer’s Disease

A Bayesian Network Model based on the Analysis of Oral Definitions of Semantic Categories

J. M. Guerrero1; R. Martinez-Tomás1; M. Rincón1; H. Peraita2
1Universidad Nacional de Educación a Distancia, Departamento de Inteligencia Artificial, Madrid, Spain; 2Universidad Nacional de Educación a Distancia, Departamento de Inteligencia Artificial, Madrid, Spain

Keywords
Bayesian networks, cognitive impairment, Alzheimer’s disease, neurodegenerative diseases, cognitive impairment of semantic content, linguistic corpus of oral definitions, lexical-semantic-conceptual deficit

Summary
Background: Early detection of Alzheimer’s disease (AD) has become one of the principal focuses of research in medicine, particularly when the disease is incipient or even prodromic, because treatments are more effective in these stages. Lexical-semantic-conceptual deficit (LSCD) in the oral definitions of semantic categories for basic objects is an important early indicator in the evaluation of the cognitive state of patients.

Objectives: The objective of this research is to define an economic procedure for cognitive impairment (CI) diagnosis, which may be associated with early stages of AD, by analysing cognitive alterations affecting declarative semantic memory. Because of its low cost, it could be used for routine clinical evaluations or screenings, leading to more expensive and selective tests that confirm or rule out the disease accurately. It should necessarily be an explanatory procedure, which would allow us to study the evolution of the disease in relation to CI, the irregularities in different semantic categories, and other neurodegenerative diseases. On the basis of these requirements, we hypothesise that Bayesian networks (BNs) are the most appropriate tool for this purpose.

Methods: We have developed a BN for CI diagnosis in mild and moderate AD patients by analysing the oral production of semantic features. The BN causal model represents LSCD in certain semantic categories, both of living things (dog, pine, and apple) and non-living things (chair, car, and trousers), as symptoms of CI. The model structure, the qualitative part of the model, uses domain knowledge obtained from psychology experts and epidemiological studies. Further, the model parameters, the quantitative part of the model, are learnt automatically from epidemiological studies and Peraita and Grasso’s linguistic corpus of oral definitions. This corpus was prepared with an incidental sampling and included the analysis of the oral linguistic production of 81 participants (42 cognitively healthy elderly people and 39 mild and moderate AD patients) from Madrid region’s hospitals. Experienced neurologists diagnosed these cases following the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA)’s Alzheimer’s criteria, performing, among other explorations and tests, a minimum neuropsychological exploration that included the Mini-Mental State Examination test.

Results: BN’s classification performance is remarkable compared with other machine learning methods, achieving 91% accuracy and 94% precision in mild and moderate AD patients. Apart from this, the BN model facilitates the explanation of the reasoning process and the validation of the conclusions and allows the study of uncommon declarative semantic memory impairments.

Conclusions: Our method is able to analyse LSCD in a wide set of semantic categories throughout the progression of CI, being a valuable first screening method in AD diagnosis in its early stages. Because of its low cost, it can be used for routine clinical evaluations or screenings to detect AD in its early stages. Besides, due to its knowledge-based structure, it can be easily extended to provide an explanation of the diagnosis and to the study of other neurodegenerative diseases. Further, this is a key advantage of BNs over other machine learning methods with similar performance: it is a recognisable and explanatory model that allows one to study irregularities in different semantic categories.

Correspondence to:
Rafael Martinez-Tomás
Universidad Nacional de Educación a Distancia
Departamento de Inteligencia Artificial
Calle Juan del Rosal 16
28040 Madrid
Spain
E-mail: rmtomas@dia.uned.es

Methods Inf Med 2016; 55: 42–49
http://dx.doi.org/10.3414/ME14-01-0071
received: July 3, 2014
accepted: April 6, 2015
epub ahead of print: April 30, 2015

© Schattauer 2016

Methods Inf Med 1/2016
1. Introduction

Early detection of Alzheimer’s disease (AD) is of utmost interest to achieve maximum effectiveness in pharmacological treatments and cognitive therapies. Standard techniques for diagnosing AD are currently based, among others, on clinical evaluation and medical judgment. Biological markers or biomarkers have made enormous progress and can replace previous clinical procedures. In fact, biomarkers in combination with diagnostic techniques (neuroimaging, cephalorachidian fluid, and genetic tests) can already diagnose AD without any margin of error and without waiting for the post-mortem anatomopathological analysis (see for example, the works of Nordbert et al. [1], Johnson et al. [2], Hu et al. [3] or Maddalena et al. [4]).

However, unfortunately, it will take a while for these tests to become routine in clinical practice. These technologies are not available to the general population since they are costly and time consuming. Moreover, the diagnosis of AD usually requires considering the complete neurological history (neuropsychological and neurological examinations, neuroimaging, neurobehaviour, and general analysis), as it is stated in Peña-Casanova [5] and Domínguez-Orozco [6].

Cognitive impairment (CI) is detected in the very early stages of AD and other neurodegenerative diseases (NDs). One of the evidences of CI in AD patients is semantic impairment of specific categories, which also worsens with time. Differential semantic impairment between the categories of living and non-living things, or between specific semantic features, in people affected with certain NDs is an important finding for early AD diagnosis (as indicated by the work of McRae et al. [7], Capitani et al. [8], or Rogers & Paut [9]).

Recent investigations also study verbal utterances for CI diagnosis by using machine learning techniques, e.g. Garrard et al. use naïve Bayes [10], Orimaye et al. use support vector machines [11], and Williams et al. use different machine learning algorithms that combine additional neuropsychological and demographic measures [12].

The fundamental problem in machine learning AD diagnosis, as in most neurological studies, is the absence of a training dataset large enough to build a reliable AD diagnosis system using supervised learning. It is certainly possible to build a classifier with good classification performance on the training dataset by selecting a reduced subset of variables, but this classifier demonstrates limited generalisation capabilities because the sample is not fully representative of the population. Therefore, a model that is based on not only the training data but also all the available domain knowledge and that the expert may recognise and understand is a better solution than a ‘black box’ classifier. This model must also facilitate the explanation of the reasoning process and the validation of the conclusions and, apart from this general behaviour, it may serve for analysing progressive semantic memory impairment caused by AD and for studying uncommon cases. We believe all these advantages are provided by the Bayesian network (BN) model.

BNs have already been successfully used in some CI diagnosis methods; for example, Seixas et al. [13] and Sun et al. [14] use BNs for diagnose DCL and AD taking as inputs the final assessments of various neuropsychological tests; Evanthia and Tripoliti use BNs to diagnose AD based on functional magnetic resonance (fMRI) [15] and Chen and Kerskovits propose a BN for early diagnosis of mild cognitive impairment (MCI) using variables that represent clinical and cognitive functions and combining them again with data from structural magnetic resonance (sMRI) [16].

Other non-conventional diagnostic methods proposed to evaluate the cognitive state of patients or even to detect motor deficiencies caused by a brain haemorrhage from monitoring daily activity. In particular, Matic et al. analyse the act of getting dressed [17] and Kearns et al. analyse the tortuosity in movement paths – irregular movements – of elderly people with cognitive impairment [18].

In this work, we assume that an early AD diagnosis method based on oral production analysis for identifying semantic memory impairment is simple, cheap, and accessible to the general population as a first filter for posterior more expensive and precise tests. Besides, our hypothesis is that BNs are an appropriate tool to achieve this objective because they combine the expert’s a priori knowledge (BN qualitative model) with the knowledge learnt from a data set of solved cases (BN quantitative model).

Two sources of information are used in the design of our BN. On one hand, Peraita and Grassós’s corpus of oral definitions [19] provides information about the differences in semantic production between healthy people and mild and moderate AD sufferers. In this corpus, 1) they analyse how certain semantic categories are represented mentally using theoretical models of semantic features or attributes obtained from explicit linguistic tasks and 2) they detail the results of the compiled cases. On the other hand, Fernández et al.’s epidemiological study [20] relates age, sex, and educational level with AD. In this way, we have the linguistic production related with CI and the epidemiological studies related with AD, and because CI is an early indicator of AD, combining these two sources allow us to attain our objectives.

The remainder of the paper is organised as follows: Section 2 describes our objectives clearly and concisely. Sections 3 and 4 describe and evaluate the BN method proposed to diagnose and analyse progressive semantic memory impairment. Finally, Section 5 presents our concluding remarks.

2. Objectives

The objective of this work is to essentially confirm two hypotheses: 1) It is possible to define a reliable and economical method for early diagnosis of CI compatible with AD based on identifying declarative semantic memory impairment and 2) BNs are the appropriate techniques to construct this intelligent diagnosis system because they combine an expert’s a priori knowledge (BN causal model) with the knowledge learnt from a set of solved cases.

We shall confirm that it is possible to analyse with artificial intelligence tech-
techniques the cognitive alterations associated with AD, thereby providing a tool to complement the expert's clinical criteria in an easy, friendly, and economical way. This tool also provides a means for examining how the organisation of categorical information in semantic memory is affected by progressive cognitive impairment in AD patients.

3. Methods

The method proposed in this research is constructed from the fields of cognitive psychology and artificial intelligence (AI). On one hand, from cognitive psychology, the linguistic corpus of oral definitions provides information about the differences in semantic production between healthy people, and mild and moderate AD sufferers. On the other hand, from AI, intelligent analysis techniques allow the modelling of complex systems by combining domain knowledge with case data. In subsection 3.1, we define the characteristics of the linguistic corpus of oral definitions, and in subsections 3.2 and 3.3 we describe the qualitative and quantitative parts of the BN model.

A software application has been developed in Java to manage the entire BN design process. Bayesian inference is performed through the inference modules developed in Elvira [21], an open-source Java tool for the implementation of BNs.

3.1 Corpus Description

In Peraita and Grasso’s linguistic corpus [19], individuals were asked to define orally six semantic categories, three living things (apple, dog, and pine) and three non-living things (car, chair, and trousers), within a time limit. The following is a transcript of an oral definition of dog: ‘They bark a lot, and they bite you if you aren’t careful, as they did to me once, and what else, that’s all. What can I tell you about dogs? I don’t know. As I don’t have one, or anything, I don’t know, I don’t remember anything, just look at the state I’m in.’

After the oral definitions were analysed and interpreted from the cognitive semantic theoretical framework, the resulting linguistic production was segmented into 11 basic semantic features, considered to be the conceptual components underlying the organisation and representation of object categories: ‘taxonomic,’ ‘types,’ ‘parts,’ ‘functional,’ ‘evaluative,’ ‘places/habitat,’ ‘behaviour,’ ‘cause/generate,’ ‘procedural,’ ‘life cycle,’ and ‘others.’ Some of these semantic features are common to all the semantic categories, as taxonomic or types, while others, such as life cycle or behaviour, are not.

The sample selected from the entire corpus, hereinafter referred to as the corpus, consisted of 81 people, of whom 42 were cognitively healthy and 39 were mild or moderate AD sufferers from various Madrid region’s hospitals. Severe AD sufferers were excluded from this study because our main focus was early AD diagnosis. As the corpus was oriented to the study of the relation between oral linguistic production of semantic categories and AD, neurologists selected patients without other neurological comorbidities. Neurologists also diagnosed mild or moderate AD patients following the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) clinical criteria, which included the Folstein Mini-Mental Status Examination criteria [22]. Table 1 shows the demographic characteristics of the sample (age, sex, and educational level).

The corpus was developed as part of a study that was approved by a double Institutional Review Board (IRB), first by UNED IRB as research coordinator institution and then by each Madrid region’s hospital IRB that provided patients for the study. Written informed consent was obtained from all the subjects.

3.2 BN Structure for the Diagnosis

Figure 1 shows the BN structure designed from the domain knowledge with the cooperation of psychology experts. In this causal model, we can find four types of variables:

1) Context or risk factor variables: Previous investigations and official reports (e.g. Fernández-Martínez et al. [20], 2013 European Commission Report [23], 2013 Alzheimer’s Association Report [24], Stern et al. [25]) have established a certain degree of association of age, sex, and educational level with AD. Variable age can take values of ‘more than 85,’ ‘between 80 and 84,’ ‘between 75 and 79,’ ‘between 70 and 74,’ ‘between 65 and 69,’ or ‘between 0 and 65,’ which are the ranges used in the epidemiological study that we rely on. Variable sex can take the values of ‘man’ or ‘woman.’ Further, variable educational level can take the values of ‘primary,’ ‘secondary,’ or ‘university.’

2) Information variables representing the symptoms: They represent the deficit in the oral production of semantic features or lexical-semantic-conceptual deficit.

<table>
<thead>
<tr>
<th>Age</th>
<th>Healthy</th>
<th>Mild and moderate AD</th>
<th>Overall N = 81</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between 55 and 64</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Between 65 and 69</td>
<td>8</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Between 70 and 74</td>
<td>6</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>Between 75 and 79</td>
<td>10</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Between 80 and 84</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>More than 84</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>12</td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td>Secondary</td>
<td>14</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>University</td>
<td>16</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>23</td>
<td>19</td>
<td>42</td>
</tr>
<tr>
<td>Woman</td>
<td>19</td>
<td>20</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 1

Demographic characteristics of the individuals

© Schattauer 2016
These variables contain knowledge about feature frequencies in the oral definitions of living and non-living things. These variables, particularised for each category, are LSCD [taxonomic| types| functional| evaluative| places/habitat| behaviour| cause/generate| procedural| life cycle| others]. They are Boolean variables: true (present) when the person presents LSCD, and false (absent) when s/he does not.

3) Intermediate variables: These variables cannot be directly observed, and their a priori probabilities are not of immediate interest, but they play an important role in achieving the correct conditional dependence of the properties and therefore, efficient inference. These variables represent the deficit in the oral production of semantic features for each category. These variables are LSCD [lt| nlt| apple| dog| pine| car| chair| trousers], where lt refers to ‘living thing’, and nlt refers to ‘non-living thing’. These variables are also Boolean. The psychological studies of Bentler [26] and Bollen [27] justify using latent or intermediate variables in causal models in this context.

4) Variables of interest or hypothesis: This group includes variables on which we want to compute their a posteriori probabilities from the findings.

- Alzheimer’s disease (AD). This variable is the objective of this research and can take values of ‘true’ (mild and moderate AD) or ‘false’ (healthy). Note that our model only considers the types included in the corpus; however, it could be easily extended to the diagnosis of other NDs that cause semantic memory impairment.
- Cognitive impairment (CI). This Boolean variable is of immediate interest, and by introducing it into the causal model, we are able to represent differential impairment in the LT and NLT semantic domains.
It is important to consider age as a context factor in the oral production of semantic features because the processing time of semantic memory slows down with age. For this reason, the numerical variables of the linguistic corpus are discretised by age groups [28, 29], thus, taking into account the age factor without increasing the BN structure. Within each group, the discretisation to consider whether LSCD is present or absent is realised in unsupervised mode, using k-means++ cluster analysis, because of the irregular decrease in the oral production in the different semantic categories and features.

We have chosen causal links with abductive reasoning between the semantic features and categories, and between the semantic categories and their respective semantic domains because with this type of reasoning, we can simplify the quantitative model. If we choose deductive reasoning, each of the conditional probability tables (CPTs) in each semantic category would need to adjust $2^n$ parameters. Consequently, it would be impossible to automatically adjust these parameters without a much wider sample. Intermediate variables are of immediate interest, and they play an important role in achieving the correct conditional dependence and independence of the properties and thus, efficient inference. The relations between nodes are modelled as follows:

- **Relation between context information and AD**: The epidemiological studies previously referenced show that age, sex, and educational level (context information) have some correlation with AD. This justifies the links between these variables, but we do not model an active pathway between the context factors and intermediate variables. With this technique, we can analyse the a posteriori probabilities of the intermediate variables without any direct influence from the context factors. We consider it appropriate at this stage of research where we attempt to study the predictive nature of each semantic category independently of the context factors.

- **Relation between AD and CI**: We consider semantic memory impairment as an aspect of CI; because CI may be associated with the early stages of AD, we model it as a causal link between AD and CI.

- **Relation with semantic features**: Our model considers all the semantic features proposed in the corpus. No attribute selection is performed because we want to analyse all the semantic categories that we have considered extensively.

The main advantage of this causal model is that it can be easily extended. For example, with a sufficiently wide sample, the ability of new semantic categories to predict AD could be studied.

### 3.3 BN Quantitative Model

The probabilistic weights of the BN links are automatically learnt from the linguistic corpus and epidemiological study. In the first level, the variables considered as context or risk factors are deterministic inputs. Next, the CPT for the interest variable AD is calculated using the naive Bayes simplifier (Equation 1). With this method, CPT values can be learnt from a small number of cases. The rest of the a priori conditional probabilities are calculated directly as the proportion of cases in the corpus that fulfil the condition of the node. The resulting CPTs are provided at simda.uned.es/Alzheimer/.

The CPT for variable AD is calculated using Equation 1 as follows:

$$P(AD \mid A, S, L, CI) = \alpha \times P(AD) \times P(A \mid AD) \times P(S \mid AD) \times P(L \mid AD) \times P(CI \mid AD)$$

(1)

where

- $\alpha$: Corrective factor.
- A: Age. Possible values in [0 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, over 84]
- S: Sex. Possible values in {man, woman}
- L: Educational level. Possible values in {primary, secondary, university}
- CI: Cognitive Impairment. Possible values in {true, false}.

The CPT for variable CI is calculated using Equation 2 as follows:

$$P(CI \mid LSCD_{LT}, LSCD_{NL}) = \frac{N(CI, LSCD_{LT}, LSCD_{NL})}{N(LSCD_{LT}, LSCD_{NL})}$$

(2)

where

$$N(x):$$ Number of cases that satisfy the conditions expressed by $x$.

$LSCD_{LT}$: LSCD in the living thing domain. Possible values in {true, false}.

$LSCD_{NL}$: LSCD in the non-living thing domain. Possible values in {true, false}.

The CPTs for variables $LSCD_{LT}$ and $LSCD_{NL}$ are calculated using Equations 3a and 3b as follows:

$$P(LSCD_{LT}) = \frac{N(LSCD_{LT})}{N}$$

(3a)

$$P(LSCD_{NL}) = \frac{N(LSCD_{NL})}{N}$$

(3b)

where $N$ is the total number of cases in the corpus.

CPTs of the causal relations between the intermediate variables representing each of the semantic categories and the variables representing the features are calculated according to Equation 4 as follows:

$$P(LSCD_{LT} \mid LSCD_{CAT}) = \frac{N(LSCD_{LT}, LSCD_{CAT})}{N(LSCD_{CAT})}$$

(4)

where

$LSCD_{CAT}$: LSCD in semantic category CAT. Possible values in {true, false}.

$LSCD_{SE,CAT}$: LSCD in semantic category CAT and semantic feature SF. Possible values in {true, false}.

CAT: Semantic Category, in {Apple, Dog, Pine, Car, Chair, Trousers}.

SF: Semantic Feature, in {Taxonomic, Types, Parts, Functional, Evaluative, Place/Context, Behaviour, Causal, Procedural, Life Cycle, Others}.

### 4. Results

We compare the probability of suffering AD inferred by the proposed BN model with the diagnosis carried out by neurologists, who followed the NINCDS-ADRDA Alzheimer’s Criteria. Leave-one-
out cross-validation (LOOCV) [30] has been used to validate the model given the small number of cases available in the corpus. LOOCV, which is a particular case of n-fold cross-validation, divides the set of n cases into n partitions, so that (n – 1) partitions are used for training and one for testing. The process is repeated n times so that a different partition is used for validation in each iteration. The n validation results obtained are then averaged to produce the average performance estimation. LOOCV allows the usage of the largest possible number of cases for training and at the same time obtains an accurate evaluation of the diagnosis method.

Taking the AD cases as positive and healthy cases as negative, and following the method proposed in Fawcett’s work [31], we obtain the ROC curve shown in Figure 2. This ROC curve can be used for estimating the optimal cut-off according to the compromise between true positives and false positives. Given that the objective of the system is the early diagnosis of AD, we would want the system to detect all the AD cases although the number of healthy persons classified as AD is high. Therefore, the optimal threshold in the continuous posterior probability distribution would be the one that maximised the number of true positives and kept the number of false negatives sufficiently low; however, to obtain the optimal threshold, a cost analysis is necessary. Instead, we consider our BN as one more method for supporting AD diagnosis along with other complementary tests, and select the optimal cut-off as the point on the ROC curve with maximum accuracy. Figure 2 shows the performance metrics for this threshold.

The method also analyses the effectiveness of the intermediate variables – semantic categories – in the early diagnosis of AD. Table 3 shows the performance metrics on intermediary variables for the selected threshold. The semantic category that is the best predictor of AD is Apple, with 87.18% precision, 87.65% accuracy, and an AUC of 0.9341; the semantic category that is the worst predictor of AD is the semantic category Dog, as can be seen in Table 4.

Finally, other automatic learning algorithms were tested, which included both supervised (C4.5, support vector machines (SVM), naive Bayes, and logistic regression) and unsupervised (k-means) algorithms, using different configurations [29]. Table 5 shows the best results obtained with each of these methods. The initial input feature vector contained 69 features (11 × 6 semantic features + 3 context features) that were normalised and standardised. The best results were obtained using attribute selection techniques, reducing the classifier to the analysis of just a few features in some cases, and in no case were context variables selected. For example, only three variables (‘car_type’, ‘dog_other’, and ‘pine_evaluative’) were used in logistic regression. Therefore, in addition to not obtaining better results, a large amount of valuable intermediate information was lost. SVMs, using 23 features, obtained slightly better results than our BN model; however, no explanation of the results could be given to the clinicians, which was one of our goals.

5. Discussion

The motivation of this research was to demonstrate that AI techniques can offer solutions to support the early diagnosis of AD in analysing free oral production of semantic features for basic objects. With our Bayesian model, we achieved 94.44% precision and 91.36% accuracy in the diagnosis of declarative semantic memory impairment compatible with AD. The system is rapid and economical, and it can be used in a preliminary screening of the population before conducting more expensive and/or complex tests. Moreover, with the resulting methodology and application, it is possible to detect patterns of impairment in the categorisation processes of the semantic memory; that is, analyse how the organisation of the categorical information in the semantic
memory is affected by progressive cognitive impairment in AD patients and study how differential impairment is produced in some semantic categories and not in others.

5.1 Assessment of Results

The statistical parameters in Table 3 are noteworthy; these parameters reveal the difference in the discriminating power of semantic categories. The semantic category Dog produces the most semantic features and yet is the worst predictor of AD. Conversely, the semantic category Apple is the best predictor of AD. Although in our research we have used six semantic categories, with this methodology, it is possible to analyse a larger number of semantic categories at distinct stages of the disease. Yet, in the evaluation of each particular case, this set can be reduced in accordance with distinct criteria, such as the occupation or educational level, to a more adjusted and sufficiently discriminative subset. If the result inferred by the BN does not provide a clear diagnosis, the BN can be used to measure the utility of evaluating other semantic categories or providing new findings.

5.2 Study Limitations

We are aware that we have few cases to examine in-depth how progressive cognitive impairment experienced by AD sufferers affects the categorical organisation of information in semantic memory, and how a differential impairment occurs in some semantic categories and not in others; however, we are convinced of the feasibility of the proposal, which we have shown by considering a sample of 81 cases.

5.3 Conclusions and Further Work

We can conclude that most of the advantages of the method stem from the BN modelling, and we therefore believe that they are the most appropriate models for the diagnosis system and for studying the disease from the analysis of the oral production of semantic features. Apart from what has already been stated earlier, here, we break down the benefits of the BN modelling, which can also be used in future lines of work:

- Primarily, the BN model, its structure and parameters, is recognisable by the expert. Furthermore, it is a quantitative model that can be adjusted from a database of cases and studies. In other words, the model allows the collection both the expert’s knowledge and the knowledge implicit in the available data.
- The BN that we present in this article along with the automatic learning algorithm can be generally used with a wide set (indeterminate) of semantic categories, as wide as possible or desired (logically, the corpus must have sufficient cases).
- New variables could also be incorporated into our BN model, for example,

### Table 3 Performance values obtained with the BN for each intermediary variable

<table>
<thead>
<tr>
<th></th>
<th>Apple</th>
<th>Dog</th>
<th>Pine</th>
<th>Car</th>
<th>Chair</th>
<th>Trouser</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>34</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>TN</td>
<td>37</td>
<td>32</td>
<td>37</td>
<td>34</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>FP</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>FN</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>TP rate</td>
<td>0.8718</td>
<td>0.8205</td>
<td>0.8205</td>
<td>0.9231</td>
<td>0.8462</td>
<td>0.8205</td>
</tr>
<tr>
<td>FP rate</td>
<td>0.119</td>
<td>0.2381</td>
<td>0.119</td>
<td>0.1905</td>
<td>0.2143</td>
<td>0.119</td>
</tr>
<tr>
<td>Precision</td>
<td>0.8718</td>
<td>0.7619</td>
<td>0.8649</td>
<td>0.8182</td>
<td>0.7857</td>
<td>0.8649</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.8765</td>
<td>0.7901</td>
<td>0.8519</td>
<td>0.8642</td>
<td>0.8148</td>
<td>0.8519</td>
</tr>
<tr>
<td>AUC</td>
<td>0.9341</td>
<td>0.8651</td>
<td>0.8974</td>
<td>0.9127</td>
<td>0.8822</td>
<td>0.8791</td>
</tr>
</tbody>
</table>

### Table 4 Mean and typical deviation of the oral production for each semantic category

<table>
<thead>
<tr>
<th></th>
<th>Healthy</th>
<th>AD</th>
<th>Healthy</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>14.2857</td>
<td>5.2821</td>
<td>5.6018</td>
<td>2.982</td>
</tr>
<tr>
<td>Dog</td>
<td>15.2143</td>
<td>7.4615</td>
<td>7.7066</td>
<td>5.2957</td>
</tr>
<tr>
<td>Pine</td>
<td>13.4762</td>
<td>5.8462</td>
<td>5.9356</td>
<td>3.6673</td>
</tr>
<tr>
<td>Car</td>
<td>13.881</td>
<td>5.8205</td>
<td>4.9396</td>
<td>4.0515</td>
</tr>
<tr>
<td>Chair</td>
<td>12.7857</td>
<td>4.7949</td>
<td>5.7277</td>
<td>3.9013</td>
</tr>
<tr>
<td>Trouser</td>
<td>12.3571</td>
<td>5.5385</td>
<td>5.2117</td>
<td>4.1729</td>
</tr>
</tbody>
</table>

### Table 5 Comparison with other automatic learning algorithms

<table>
<thead>
<tr>
<th></th>
<th>BN</th>
<th>SVM</th>
<th>C4.5</th>
<th>Naïve Bayes</th>
<th>K-means</th>
<th>Logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td>31</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>TN</td>
<td>40</td>
<td>41</td>
<td>33</td>
<td>39</td>
<td>31</td>
<td>35</td>
</tr>
<tr>
<td>FP</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>FN</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>TP rate</td>
<td>0.87</td>
<td>0.87</td>
<td>0.90</td>
<td>0.79</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>FP rate</td>
<td>0.05</td>
<td>0.03</td>
<td>0.23</td>
<td>0.20</td>
<td>0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>Precision</td>
<td>0.94</td>
<td>0.97</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
<td>0.90</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.91</td>
<td>0.93</td>
<td>0.84</td>
<td>0.86</td>
<td>0.85</td>
<td>0.89</td>
</tr>
<tr>
<td>AUC</td>
<td>0.96</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>–</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Our method can be extended with a new explanation mechanism, for which it would be really useful to measure the sensitivity of the model to changes in variables. With the sensitivity analysis of the parameters, it is possible to study the influence of each of the variables in the result, which determines whether to include new variables or not in our BN model.

Our method can be extended with a decision analysis (influence diagram), to maximise the expected utility of several open options for a decision. For example, the expected utility of the recommendation to do the Mini-Mental State Examination test, a neuropsychological evaluation, and complementary explorations such as biochemical analysis, PET, etc., could be maximised.

Finally, we believe that this method could be extended with new tools to provide a large amount of knowledge on irregular semantic memory impairment and contribute to the early diagnosis of this severe disease in an extremely economical and accessible way.

References