

BACHELOR THESIS
ARTIFICIAL INTELLIGENCE
Radboud University Nijmegen

Diagnosis Systems for Dementia

YOUETTA KUNNEMAN

y.j.kunnneman@student.ru.nl

Supervisor:
LOUIS VUURPIJL

September 4, 2012

Abstract

In 2020 there will be about 42.3 million people worldwide with a form of dementia (Ferri et al., 2006). It seems likely that a fast and precise diagnoses can contribute to better care for the patient and more certainty for the caregiver. Medical diagnosis systems already exist and assist physicians and psychologists worldwide. Still no medical decision support system for diagnosing dementia exists.

The focus of my research will be the design, development and evaluation of decision support system for diagnosing dementia, based on four of the most prominent artificial intelligent techniques. These are:

- Logistic regression
- Nearest neighbor
- Neural networks
- Support vector machines

Because no diagnosis system exists for dementia, the accent of my research will be the exploration of suitable AI techniques for such a system. Furthermore I will look at the practical side of a diagnosis system for dementia by making use of the knowledge of an expert in the field.

Contents

1	Introduction	1
1.1	The diagnosis of dementia	1
1.1.1	Dementia	1
1.1.2	The diagnosis pipeline	3
1.2	Medical decision support systems	5
1.3	Research questions	7
1.4	Research plan and organization of the thesis	8
1.4.1	Preparations for this research	8
1.4.2	Data processing	8
1.4.3	Experimental rounds	8
1.4.4	Conclusion and discusion	9
2	Medical Decision Support Systems	10
2.1	Logistic regression	11
2.2	Nearest neighbor	11
2.3	Neural networks	12
2.4	Support vector machine	13
2.4.1	Sequential minimal optimization	13
2.5	Techniques used in this research	14
3	Methods	15
3.1	Neuropsychological tests	15
3.1.1	Cambridge Cognitive Examination	15
3.1.2	Mini-Mental State Examination	17
3.1.3	Other neuropsychological tests	17
3.2	Data acquisition	18
3.2.1	<i>UMC St. Radboud</i> data	18
3.2.2	<i>De Zorggroep</i> data	18
3.3	Pre-processing	19
3.3.1	Gender, age and education	21
3.3.2	CAMCOG test results attributes	22

3.3.3	CAMCOG norm attributes	25
3.3.4	Dementia types classification	25
3.4	The dataset for Round 1	27
3.5	The Weka workbench	30
3.6	Experiments Round 1	30
3.6.1	Experimental setting 1, cross-validation	30
3.6.2	Experimental setting 2, split	31
4	Results Round 1	32
4.1	Setting 1	32
4.2	Setting 2	32
5	Evaluation and Round 2	35
5.1	Discussion with expert	35
5.1.1	Evaluating the performance of the classifiers	35
5.1.2	Evaluating individual subjects	36
5.1.3	Recommendations	37
5.1.4	Consequences of incorrect diagnosing	38
5.1.5	Round 2A: new datasets	38
5.1.6	Results Round 2A	39
5.2	Round 2B: attribute reduction	41
5.2.1	Results Round 2B	43
6	Conclusion and Discussion	44
6.1	The results of the AI techniques	44
6.2	The artificial diagnosis system in practice	47
6.2.1	Final considerations	48
	References	50
	Appendix A	53
	Appendix B	54
	Appendix C	57
	Appendix D	60
	Appendix E	61
	Appendix F	63

Chapter 1

Introduction

Medical diagnosis systems already exist and assist physicians and psychologists worldwide. Medical diagnosis systems already exist and assist physicians and psychologists worldwide. Still no medical decision support system for diagnosing dementia exists. The focus of my research will be the design, development and evaluation of decision support system for diagnosing dementia, based on four of the most prominent artificial intelligent techniques.

In this introduction the use of decision support systems for medical problems is briefly discussed. The importance of a diagnosis system for dementia will be explained in Section 1.2. Also the epidemiology and diagnosis pipeline of dementia are described. I will discuss more about the psychological diagnosis of dementia in Section 1.1.

1.1 The diagnosis of dementia

1.1.1 Dementia

Dementia is a cluster of different phenomena, particularly the deterioration of memory and other cognitive functions. It is a very frequent condition that strongly increases with age (Jonker, Slaets, & Verhey, 2009).

The term dementia was already used in the Roman era, for patients suffering from ‘fever delirium remaining permanent insanity’ (Jonker et al., 2009 and Jonker, Verhey, & Slaets, 2001). Dementia was official classified as a syndrome by both Pinel (1818) and Esquirol (1838). Esquirol defined dementia as a psychological decay, caused by a chronic illness of the brain.

In the nineteenth century clinical symptoms and brain pathology became more associated (Jonker et al., 2009). At the time, one believed that arteriosclerosis is

the cause of mental deterioration at the end of life. In 1898 Alzheimer had his doubts about this theory and analyzed relative young patients with dementia. He concluded that not all the mental deterioration was caused by vascular changes (Jonker et al., 2009).

Since then a lot of research has been done about the causes of dementia. Nowadays dementia is a general name for a number of symptoms with different causes. According to the latest edition of the Diagnostic and Statistical Manual of Mental Disorders, DSM-IV-TR, the dementia disorders have “the development of multiple cognitive deficits (including memory impairment)” in common (American Psychiatric Association, 2000).

In the DSM-IV-TR dementia section dementia is categorized by cause in six different disorders:

- Dementia of the Alzheimer’s Type
- Vascular Dementia
- Dementia Due to Other General Medical Conditions
- Substance-Induced Persisting Dementia
- Dementia Due to Multiple Etiologies
- Dementia NOS (Not Otherwise Specified)

Some of these dementia categories could include more dementia types. For example ‘Frontotemporal Dementia’, is a dementia type categorized in the DSM-IV-TR under ‘Dementia Due to Other General Medical Conditions’.

In the practice of diagnosing dementia Dutch psychologists (e.g. from *de Zorggroep* and *UMC St. Radboud*) will not diagnose all the types of dementia. Distinguishing between all the types is difficult, because types do not exclude each other. For example a client could have Alzheimer’s Disease and Vascular Dementia (Jonker et al., 2009). Moreover for screening clients, the diagnosis of dementia alone could be satisfactory. When needed, a further diagnosis and examination will take place. In dealing with clients who have dementia diagnosing of dementia could be enough.

When a more specific diagnosis is needed, there will be a distinction between four types of dementia. These are the four most common types of dementia (*Dianostiek van dementie*, 2008):

- Alzheimer’s Disease ¹
- Vasculair Dementia
- Lewy Body Dementia ²
- Frontotemporal Dementia ²

¹Dementia of the Alzheimer’s Type

²The last two, ‘Lewy Body dementia’ and ‘Frontotemporal Dementia’, are types categorized in the DSM-IV-TR under ‘Dementia Due to Other General Medical Conditions’.

1.1.2 The diagnosis pipeline

“Diagnosis of dementia is a stepwise process that involves examination of patient history and early warning signs, as well as performance screening, assessment of daily functioning, behavioral problems, and caregiver status, with possible referral to specialist clinics for more thorough assessment” (Galvin & Sadowsky, 2012).

In order to come to a diagnosis, the psychologist goes through a diagnosis process. A general procedure is shown in Figure 1.1. The course of the procedure is for example adjusted to the client and his/her surroundings. The order of the different processing stages is variable and some stages are mergeable, replaceable or even removable.

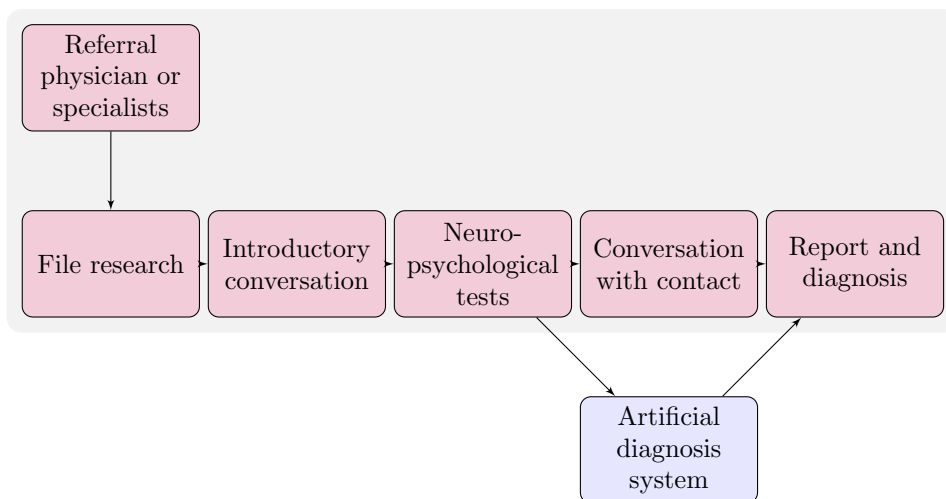


Figure 1.1: General diagnostic procedure and artificial diagnosis system

In the general procedure a client comes to the psychologist after referral. The psychologist begins with a research in the following files of the client: psychological, medical and care file³. This gives background information about the client: there might be physical or mental characteristics, for example a CVA, that are important to take into account in the diagnosis or in dealing with the client.

Next, the psychologist will give a self-introduction and tell more about the course of events. In this conversation the psychologist starts with observing. The psychologist could for example look at the punctuality and appearance. Afterward she neuropsychological tests take place. The psychologist or an assistant will do the

³Two or all files may be combined in a single care file

testing. In the Netherlands (e.g. *De Zorggroep*) the following tests are commonly used: Cambridge Cognitive Examination (CAMCOG) and Mini-Mental State Examination (MMSE). In section 3.1 some of these neuropsychological tests will be described in more detail.

Before making a diagnosis, a contact is consulted. The contact is the spouse, a child or other relative of the client. The contact provides additional information for the psychologist, for example background information about the past of the client. This is also necessary for the hetroanamnesis. When a client is a child, an elderly, a coma patient or mentally disturbed, the client could have a wrong insight in his/her illness and decay. With a hetroanamnesis, the psychologist asks someone else than the client (mostly a close relative) about the illness history of the clients' daily life.

On rare occasions this stage is skipped in the process, but it is not recommended. Most times the information, even if it is limited, helps to make a diagnosis.

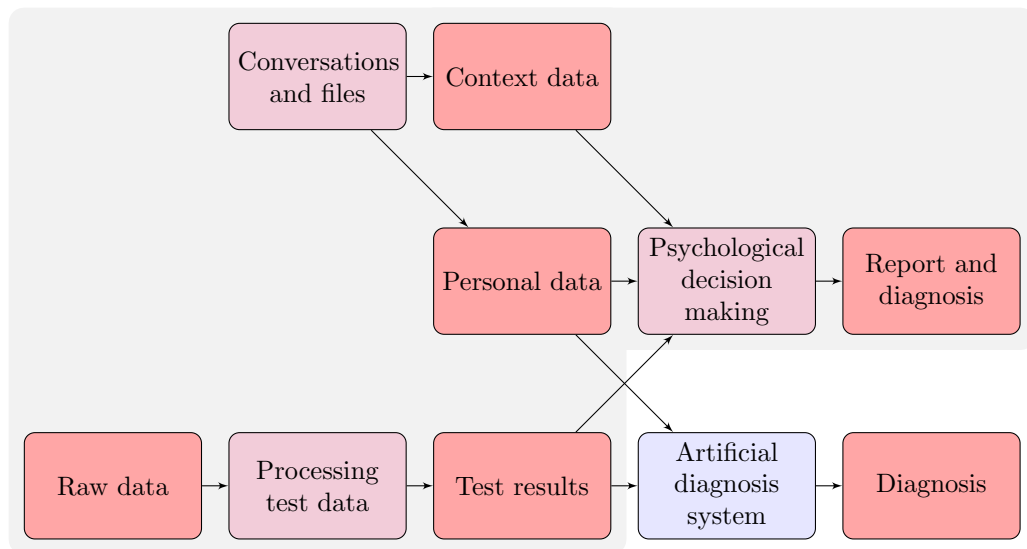


Figure 1.2: Data flow of the human and artificial decision making

At the end the psychologist makes up the report and the diagnosis (for more about the report see section 3.2.2). The conclusion of the report contains the diagnosis and sometimes also recommendations or advice for care staff, specialists or family. In making the diagnosis, the psychologist will look at context, impressions and observations, personal details and the test results. These factors are described in the report. The data flow of the diagnosis process is shown in Figure 1.2.

1.2 Medical decision support systems

In 2020 there will be 42.3 million people worldwide aged more than 60 years, expected with a form of dementia (Ferri et al., 2006). Of those people there will be 6.9 million living in Western Europe. Alzheimer's disease (the most common type of dementia) is one of the most disabling and burdensome health conditions worldwide (Ferri et al., 2006). Most patients with dementia live at home, where family caregiver(s) take care of them (Jonker, Verhey, & Slaets, 2001). However taking care of a demented person contributes to psychiatric morbidity (Schulz & Martire, 2004). It seems likely that a fast and precise diagnoses can contribute to better care for the patient and more certainty for the caregiver.

To come to a better diagnosis process, scientists can analyze, improve and test the human diagnosis process itself with the help of a computer. Many scientists analyzed the human reasoning process in medical diagnosis (Miller, 1994).

But computers can be of more assistance than improving the human diagnosis process. Computers can help the physicians with collecting and processing clinical information (Ledley & Lusted, 1959). Through the years more electronic data is stored and automatically processed. This makes medical decision support systems an addition in improving the accuracy of medical diagnosis (Mangiameli, West, & Rampal, 2004). The prediction is that the use of medical diagnosis systems will tenfold within the decade (Mangiameli et al., 2004).

In Figures 1.1 and 1.2 is the role of the artificial diagnosis system displayed.

Medical diagnosis systems are used to predict/diagnose all kinds of medical problems. Examples are listed in Table 1.1. For more information on the artificial intelligence techniques used for these medical diagnosis systems, see Chapter 2.

Medical problem	Classifiers (Mangiameli et al., 2004)	Classifiers (Kononenko, 2001)
Acute myocardial infarction	Logistic Regression, k-Nearest Neighbor, and Multilayer Perceptron	
Acute pulmonary embolism	Neural Network	
Blood transfusion costs	Neural Network	
Coronary artery disease	Logistic Regression	
Breast cancer	Logistic Regression, k-Nearest Neighbor, Kernel density, Multilayer Perceptron, and other Neural Networks	Symbolic Learning
Colorectal, hepatic and ovarian cancer	Multilayer Perceptron	
Cytomegalovirus retinopathy	Multilayer Perceptron	
Drug/plasma concentration	Multilayer Perceptron	
Gallstones	Logistic Regression	
Gynecologic cytology smears	Neural Network	
Heart diseases of newborn babies		Bayes
Liver metastases	Logistic Regression	
Lower back disorders	k-Nearest Neighbor, Multilayer Perceptron, other Neural Networks	
Lower urinary tract dysfunctions		Symbolic Learning
Lymphography		Symbolic Learning
Mortality risk for reactive airway disease	Logistic Regression	
Sepsis	Multilayer Perceptron	
Severe head injury	Kernel density	
Spondylarthropathy	Logistic Regression	
Survival of hepatitis		Symbolic Learning
Ulcers	Logistic Regression	

Table 1.1: Some medical problems use classifiers to predict/diagnose an medical problem

1.3 Research questions

The topic of my research will be the design, development and evaluation of a decision support system for diagnosing dementia based on four of these techniques.

There exists no medical decision support system for diagnosing dementia at present. However a number of promising artificial intelligence techniques may be pursued for the design of such a system. These techniques are shown in Table 1.1 and will be discussed in Chapter 2.

Because no diagnosis system exists for dementia, the focus of my research will be the exploration of suitable AI techniques for a diagnosis system. The techniques I will use for the diagnosis system, are selected for this exploration, although other techniques also might suit for this diagnosis problem.

Furthermore I will look at the practical side of a diagnosis system for dementia, by using knowledge and experience of an expert. When designing the diagnosis system and selecting the attributes, I will keep the practical side of the system in mind.

My research questions are as follows:

1. Can artificial intelligence techniques be used for the automated diagnosis of dementia?
 - (a) If so, which techniques are accurate and/or robust?
2. How can artificial intelligence techniques be used in practice?
 - (a) Which techniques are consistent with the experts?
 - (b) How much effort does it take to use the artificial diagnosis process in practice?

1.4 Research plan and organization of the thesis

In order to answer my research questions, I followed the next steps:

- literature study and contact with experts
- data acquisition and pre-processing
- running first round of experiments
- evaluation together with expert
- running second round of experiments
- conclusion and discussion

1.4.1 Preparations for this research

First I reviewed literature of the AI classifiers used in medical diagnosis systems. My conclusions are described in Chapter 2. Furthermore I interviewed experts in the field. These experts have specific knowledge about dementia and the diagnosing process and provided sufficient data for testing.

1.4.2 Data processing

The data that has been made available was submitted by two different health care institutions: *UMC St. Radboud* and *De Zorggroep*.

For combining of the two datasets, I applied relatively straight-forward transformations. Furthermore, I computed attributes, that were not present in one or both datasets, to make the datasets complete. Also attributes were removed from the dataset(s). These were irrelevant attributes that occur in only one of the datasets. In Chapter 3, the data analyzing and pre-processing are described.

After the data was analyzed and preprocessed, I prepared the data for application with the most prominent AI classifiers, in particular a support vector machine, neural network, k-nearest neighbor and logistic regression. I applied the experiments with these classifiers in the state-of-the-art machine learning workbench: Weka. The details of the classifiers and the Weka environment can also be found in chapter 3.

1.4.3 Experimental rounds

The first round of experiments is run to explore the diagnosing problem. After the evaluation of the results of this round, second round is designed to improve the results.

Round 1 In Round 1, the classifiers were tested in two experimental settings: the cross-validation setting and the split setting. These experimental settings are described in Section 3.6 and the results of the first round are displayed in Chapter 4.

Round 2A The performances and individual cases of the first round were evaluated with the expert. Round 2A was an experiment round, which anticipated on the evaluation together with the expert. It consisted of testing a dataset with additional attributes.

Round 2B The second round also contained experiment 2B. Round 2B has an attribute reduction of Round 1 with the setting cross-validation.

Round 2, which include the evaluation, the experiments 2A and 2B and their results, are described in Chapter 5.

1.4.4 Conclusion and discussion

In Chapter 6, the conclusions of the experiments and the answers on the research questions will be given.

Chapter 2

The Artificial Intelligence View: Medical Decision Support Systems

In this chapter the most prominent AI classifiers will be discussed. At the end the classifiers used for my research are described.

Machine learning could be the answer for what kind of system makes a good medical diagnosis system for dementia. Machine learning can solve problems in different medical domains with diagnostic and prognostic problems by providing methods, techniques and tools (Magoulas & Prentza, 2001). An advantage for the diagnosis system for dementia is that machine learning technology is suitable for small specialized problems in medical diagnosis (Kononenko, 2001). It is also adequate in cases in which algorithmic solutions are not available or there is a lack of formal models (Magoulas & Prentza, 2001). Kononenko (2001) claims that ‘machine learning algorithms were designed and used to analyze medical datasets from the very beginning’.

In his papers (2001, 1998) Kononenko made a comparison of some, as he calls it, “state-of-the-art systems” when applied to a couple of medical diagnostic tasks. The diagnostic tasks included the localization of the primary tumor, the predictions of recurrent breast cancer and the diagnostics of thyroid diseases and rheumatology (Kononenko, Bratko, & Kukar, 1998). An overview is shown in Table 2.1.

Below, the following artificial intelligence techniques are described:

- Logistic regression
- Nearest neighbor
- Neural networks
- Support vector machines

Technique	Classifier	Performance	Transparency	Explanation	Reduction	Missing data handling
Decision tree	Assistant-R Assistant-I	Good Good	Very Good Very Good	Good Good	Good Good	Acceptable Acceptable
LFC		Good	Good	Good	Good	Acceptable
Bayesian classifier	Naive Bayes Semi-naive Bayes	Very Good Very Good	Good Good	Very Good Very Good	No No	Very Good Very Good
Neural Network	Back-propagation	Very Good	Poor	Poor	No	Acceptable
k-Nearest Neighbor		Very Good	Poor	Acceptable	No	Acceptable

Table 2.1: The appropriateness of various algorithms for medical diagnosis. (Kononenko, 2011)

2.1 Logistic regression

With the ridge regression theory, standard linear regression can be extended to logistic regression (Cessie & Houwelingen, 1992). Logistic regression predicts the class by using a logistic sigmoid of a linear relationship between the class and the attributes. Logistic regression is a popular technique in biostatistics (Cessie & Houwelingen, 1992).

The logistic model is popular, because the logistic function provides a probability and an S-shape, which gives a description of the combined effect of several attributes (Kleinbaum, Klein, & Pryor, 2002).

The logistic model formula (Kleinbaum et al., 2002):

$$P(y = 1|x_i) = \frac{1}{1 + e^{-\alpha \sum_j \beta_j x_{i,j}}}$$

2.2 Nearest neighbor

Nearest neighbor is a machine learning technique that uses the idea that a subject is similar to the objects in the neighborhood of the subject. With k-nearest neighbor, the neighborhood is k objects big.

The k-nearest neighbor algorithm is very simple to implement and often performs quite good (Russell & Norvig, 2003). Despite that the k-nearest neighbor has a poor knowledge representation, physicians found that the k cases of nearest neighbor were shown as an explanation for the decision (Kononenko, 2001). The best matching case(s), which are the nearest neighbor(s), provide appropriate information for further diagnosis and examination.

A disadvantage of nearest neighbor is calculating distances in a large dimension space. The dimension space is the number of attributes. The methods for calculating the distances do not scale well with the dimensions, because the complexity increases rapidly with more attributes (Russell & Norvig, 2003).

The decision rule for nearest neighbor is as follows (Wu, 2012):

$$\begin{aligned} f(x_i) &= \theta(x'), \\ \theta(x_a) &= y_a, \\ x' &= \arg \min(\text{dist}(x_i, x')) \end{aligned}$$

And in case of k-nearest neighbor:

$$\begin{aligned} f(x_i) &= \theta(x'), \\ \theta(x_1, \dots, x_k) &= \text{mode}(y_1, \dots, y_k), \\ x' &= \{x' \mid \forall z : \text{dist}(x_i, z) \geq \text{dist}(x_i, v_k) \geq \text{dist}(x_i, v_{k-1}) \geq \dots \\ &\quad \geq \text{dist}(x_i, v_1), v_1 \neq v_2 \neq \dots \neq v_k \neq x_i, x' \in \{v_1, v_2, \dots, v_k\}\} \end{aligned}$$

2.3 Neural networks

Artificial neural networks are inspired by neural networks in the brain. A neural network contains input nodes, hidden nodes if necessary and output nodes. Input nodes are used to ‘sense’ the information, hidden nodes work a sort of extra information gain process and output nodes give the result of the whole process. These nodes are connected by weights, which indicate the strength of the connections between input and output nodes. In case of a multilayer perceptron (MLP) hidden layers, which contain hidden nodes, are added between the input and output layer.

For a large number of medical decision support systems, neural networks have been used, because they have a great predictive power (Mangiameli et al., 2004). Neural networks can map capabilities or associate patterns, generalize, is robust and has high fault tolerance and processes information parallel ant at a high high-speed (Ranjith & Khandelwal, 2011).

Neural networks are not transparent in knowledge representation and explanation of their decisions is not easy (Kononenko, 2001). Symbolic rules are more clear in their decisions. It is possible to extract symbolic rules from a neural network (Kononenko, 2001 and Saito & Nakano, 1988). This technique, symbolic knowledge extraction could be a promising approach to the knowledge acquisition problem (Saito & Nakano, 1988). However the extracted rules tend to be large

and complex and offer hardly a useful explanation (Kononenko, 2001).

The decision rule for a two layer neural network (Burges, 1996):

$$f(x_i) = \theta \left(\sum_{n=1}^{N_s} \alpha_n K(x, s_n) + \beta \right),$$

$$K = \tanh(\gamma(x \cdot s_n) + \delta)$$

2.4 Support vector machine

The support vector technique finds the optimal hyperplane separating two classes such that it can be used as a decision boundary (Vuurpijl, Schomaker, & Erp, 2003). The support vector machine is based on the support vector theory of Vapnik.

Support vector machines can handle non-linear solution surfaces using the idea of convolution of the dot-product (Cortes & Vapnik, 1995). Also it is robust against errors in the training set with the notion of soft margins.

Training of most support vector machines takes a lot of time, especially with a large dimension of space and they are also complex and hard to implement (Platt, 1998). Though by using the sequential minimal optimization algorithm the support vector machine has a better scaling properties, is often faster and easier to implement (Platt, 1998). The support vector machine is an extremely powerful and universal learning machine (Cortes & Vapnik, 1995).

The decision rule for a radial basis function machine (Burges, 1996):

$$f(x_i) = \theta \left(\sum_{n=1}^{N_s} \alpha_n K(x, s_n) + \beta \right),$$

$$K = \exp\left(\frac{-\|x - s_n\|^2}{\sigma^2}\right)$$

2.4.1 Sequential minimal optimization

SMO stands for a support vector machine that uses the sequential minimal optimization algorithm (Platt, 1998). When training a support vector machine, it solves a very large quadratic programming optimization problem. The SMO solves the quadratic programming optimization problem quickly without extra storage (Platt, 1998). Because the SMO scales dimensions very well, the SMO performs well on large datasets with many attributes (Platt, 1998).

2.5 Techniques used in this research

Nearest neighbor is not difficult to understand and is insightful in how it makes its decisions. A disadvantage of nearest neighbor is the difficulty of calculating useful distances in a large dimension space. In contrast to nearest neighbor, neural networks give no transparent explanation of the decision. Still neural networks are popular, because of their good pattern recognition. The logistic models can provide a useful probability and explain their decisions quite well. An advantage of the support vector machines is that they can handle non-linear solution surfaces. Furthermore the support vector machines is robust against errors in the training set with the notion of soft margins. With the SMO technique, the support vector machines are not slow.

The techniques described in this chapter are prominent and well-known AI techniques for classifying subjects. As shown in Table 1.1, they are also frequently used in solving medical problems. Therefore I decided to apply the following techniques:

- Logistic regression
- Neural network
- Nearest neighbors
- Support vector machine

Chapter 3

Methods

The data acquisition and pre-processing (Section 3.2 and 3.3) will be discussed in this chapter. Because the dataset contains a great part of test results, the neuropsychological tests that occur in my research are discussed (Section 3.1). Section 3.4 is about the first dataset used in this research. This chapter also contains a brief description of the machine learning environment Weka (Section 3.5).

3.1 Neuropsychological tests

The datasets that I used for my research contain attributes (derived) from the test results from neuropsychological tests used by *De Zorggroep* and *UMC St. Radboud* for the diagnosis of, and research on, dementia. The CAMCOG and the MMSE are described in Sections 3.1.1 and 3.1.2. Other neuropsychological tests that occur the *UMC St. Radboud* dataset are listed and described in Section 3.1.3.

3.1.1 Cambridge Cognitive Examination

The CAMCOG is the cognitive section of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX). The CAMDEX is “a comprehensive inventory of information relevant to the diagnoses of dementia” (Lindeboom, Horst, Hooyer, Dinkgreve, & Jonker, 1993). The aim of the CAMDEX is to design a single, standardized instrument for accurate clinical diagnosis of dementia and the dementia classifications: Alzheimer’s Disease, Vascular Dementia and Dementia Due to Other General Medical Conditions (Roth, Huppert, Mountjoy, & Tym, 2005).

Compared to other tests the CAMDEX is not as general as the CARE⁴ and

⁴the Comprehensive Assessment and Referral Evaluation

the GMS⁵, but also not as specific as the CERAD⁶ (Lindeboom et al., 1993).

A great advantage of the CAMCOG is that it uses short tasks similar to MMSE (Lindeboom et al., 1993). It takes administering the CAMCOG only 20 to 30 minutes (Koning, Dippel, Kooten, & Koudstaal, 2000). The CAMCOG total score is derived from 60 items from different components (questions and small tasks). The components are distributed over the memory section and the non-memory section. The sections contain two or more scales. Some scales have more subscales. For example the scale ‘orientation’:

Section	• Scale	◦ Subscale
Memory Section	• Orientation	◦ Time ◦ Place

For the all the scales and subscales see Table 3.1. The distribution of the scores are listed in Appendix A.

Memory Section	Non-memory Section
• Orientation	• Language
◦ Time	◦ Comprehension
◦ Place	◦ Expression
• Memory	• Concentration/Attention
◦ Remote	• Praxis
◦ Recent	◦ Constructie
◦ Learning	◦ Ideom/ideat
	• Calculation
	• Abstract thinking
	• Perception

Table 3.1: Sections, scales and subscales of the CAMCOG

The CAMDEX and CAMCOG can detect dementia at an early stage (Roth et al., 2005). An abnormal total score of the CAMCOG could point out dementia, if other causes are excluded. The valuation of the memory section score is a better indication for (early) dementia of the Alzheimers Type. Also an abnormal score on the non-memory section score could indicate a possible dementia syndrome.

⁵the Geriatric Mental State Schedule

⁶Consortium to Establish a Registry for Alzheimer’s Disease

3.1.2 Mini-Mental State Examination

The aim of the Mini-Mental State Examination (MMSE) is the detection of mental status changes, particularly in the elderly and thereby enhances patient care (Cummings, 1993). The MMSE is a structured approach to mental status testing that screens for intellectual impairment (Cummings, 1993).

The MMSE is usually tested together with the CAMCOG. Some components overlap in both tests, and both tests contain the same kind of tasks.

3.1.3 Other neuropsychological tests

A great part of the attributes in the *UMC St. Radboud* dataset describe the physical state of the client. The dataset also contains the results of psychological tests. I will describe the neuropsychological tests of the *UMC St. Radboud* dataset briefly:

Cambridge Cognitive Examination (CAMCOG)

The CAMCOG is a screening tool for dementia (Rijsbergen et al., 2011). See also Section 3.1.1.

Clinical Dementia Rating (CDR)

The CDR is a rating scale for dementia (Rijsbergen et al., 2011).

Fifteen-words test A and B (15-WT)

The 15-WT is a measurement for the short vs. long term memory (Deelman, 1972).

Lawton Instrumental Activities of Daily Living (IADL)

The IADL measures a persons ability to perform tasks in activities of daily life (Graf, 2008).

Mini-Mental State Examination (MMSE)

The MMSE is a measurement for the cognitive grading of patients (Folstein, Folstein, & McHugh, 1975). See also Section 3.1.2.

Neuropsychiatric Inventory (NPI)

The NPI measures the neuropsychiatric symptoms and behavior, that affect most subjects with dementia (Medeiros et al., 2010).

TrailMaking test A and B (TMT)

The TMT measures speed of cognitive processing and executive functioning (Sánchez-Cubillo et al., 2009).

Wechsler Adult Intelligence Scale-III, Digit Span A and B (WAIS-III)

The WAIS contain a series of subtests with different accents on a number of cognitive functions (Sherman & Blatt, 1968). The digit span is one of these subtests. This subset measures primarily the short-term verbal retention (Fenwick & Holmes, 1993).

Only the CAMCOG and MMSE are measures for the determination of dementia. The CDR tests the degree of the dementia and not determination of dementia itself. The 15-WT, IADL, TMT and WAIS-III Digit Span are neuropsychological tests for further or specific insight in the client. The NPI is a test that is applied after the diagnosis of dementia.

3.2 Data acquisition

During the first preparations, I came in contact with two different institutions: care institution *De Zorggroep* and academic hospital *UMC St. Radboud*. Both institutions were interested and assisted by providing data for my research.

3.2.1 *UMC St. Radboud* data

The University Medical Centre Sint Radboud is an academic hospital related to the Radboud University Nijmegen. In 2012, the UMC St. Radboud had 156.085 clients at the polyclinic, 32.163 intramural clients, 49.337 extramural clients, 22.882 elective operations and 24.505 emergency Department visits (*Jaardocument 2011 UMC St. Radboud*, 2011).

The dataset to which I was given access was a part of a dataset which was prepared for (ongoing) research at the UMC St. Radboud and Radboud University. The dataset contains 651 clients with 192 attributes. The list of all the original attributes can be found in Appendix B. A short description of the neuropsychological tests in the UMC St. Radboud dataset was listed in Section 3.1.3.

3.2.2 *De Zorggroep* data

Stichting Zorggroep Noord- en Midden-Limburg (trade name *De Zorggroep*) is a care institution in Limburg, a province in the Netherlands with several residential and nursing homes and care centra. In 2010 *De Zorggroep* had 2.672 intramural clients, 5.152 extramural clients and 72.137 clients in youth and maternity care (*Jaardocument 2010 Stichting Zorggroep Noord- en Midden-Limburg*, 2010).

In order to create the dataset, I was granted access to anonymous psychological reports of “screeningsonderzoeken”. These screening reports have the aim of gaining insight into the cognitive functioning of the client. The psychologist examines the cognitive functions of the client as described in Section 1.1. The reports describe inter alia the details of the neuropsychological test results of the client and the diagnosis. From these reports I could extract the attributes which were

important for my research. In Appendix C is an adapted version of a screening report. The general structure of a screening report:

- Personal details such as age and gender
- Examination question
- History of the client
- Heteroanamnese with contact
- Observation
- Diagnostic instrumentation
- Test results
- Conclusion and summary/recommendations

De Zorggroep made data available for research, but it should be noted that these data have been collected for non-research purposes. The screening reports are made for gaining insight into the cognitive functioning of the client for diagnosing dementia. Only the necessary neuropsychological tests to determine the diagnosis will be applied. Furthermore, I only extracted the attributes which were of interest for my research. This results in less attributes, compared to the attributes of *UMC St. Radboud* dataset (192 attributes). *De Zorggroep* dataset contains 87 attributes.

Only the CAMCOG and MMSE test results occur both in de *De Zorggroep* and *UMC St. Radboud* data.

For every client I had to read a screening rapport, and extract the attributes. Preparing the reports, for example anonimise, also took time for psychologist. Therefore *De Zorggroep* dataset contains far less subjects compared to *UMC St. Radboud* dataset.

3.3 Pre-processing

For combining the two datasets (*De Zorggroep* and *UMC St. Radboud*), the trivial attributes were made uniform. Furthermore, I applied the following adjustments:

0 - No adjustment, both present

Many attributes occur in both data sets. Twelve of these attributes are copied in the combined data set, which forms the basis of the experiments described in this thesis.

1 - Transformed

Six relatively straight-forward transformations have been performed. These transformations, described in Section 3.3.1, imply:

- categorization of age
- categorization of education

2 - Calculated

Twenty of the attributes were not present in one or both datasets and have been computed to make the datasets complete.

- normalization from date-of-birth to age (Section 3.3.1)
- categorization of education (Section 3.3.1)
- CAMCOG scales ‘attention’ and ‘calculation’ (Section 3.3.2)
- CAMCOG memory and non-memory section (Section 3.3.2)
- CAMCOG median and threshold of total score and most scales (Section 3.3.3)
- CAMCOG cumulative scores for the memory and non-memory sections (Section 3.3.3)

3 - Removed

The *UMC St. Radboud* data set original contained 192 attributes. From *De Zorggroep* screening reports, I extracted 87 attributes. I removed the attributes which were not relevant for my research. This were the attributes that occur in only one of the datasets.

The attributes that still remained although not present in both datasets, were necessary for the CAMCOG and could be recalculated in the other dataset.

For the visualization of the adjustments I will use flowcharts. These flowcharts will consist of different colors for the *UMC St. Radboud* attributes (blue), *De Zorggroep* attributes (red) and the new or combined attributes (purple). An example is given in Figure 3.1.

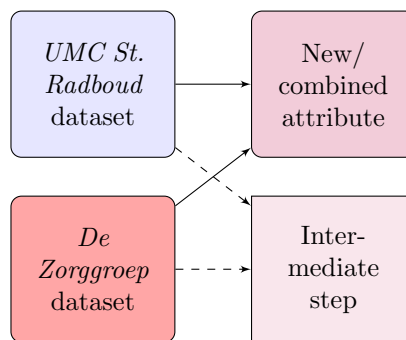


Figure 3.1: Prototype of a flowchart

3.3.1 Gender, age and education

In the CAMCOG, the population is distinguished by gender, age category and education level. The CAMCOG uses these characteristics to determine standard scores (Roth et al., 2005). In order to interpret the CAMCOG score, these categories need to be available for computing norms. For these categories, see Table 3.2.

Gender:		Male, Female
Age category:	I	65-69 years ⁷
	II	70-74 years
	III	75-79 years
	IV	80-84 jaar/years ⁷
Dutch education level:	1	Lagere school zes of minder klassen
	2	Voortgezet onderwijs, LO met meer dan 6 leerjaren, LBO, ULO, MULO
	3	Middelbaar en hoger onderwijs

Table 3.2: Distinction for norms and scores

The *De Zorggroep* dataset already consisted of an age and age category. But the *UMC St. Radboud* dataset contained none of these attributes. In Figure 3.2 the age and age category attributes are shown. The age and age category were calculated in Excel for all subjects in the *UMC St. Radboud* dataset.

I calculated the age with the following formula:

$$Leeftijd = opname\ datum - geboorte\ datum$$

Note that the date of birth is compared to the intake date, instead of the current date. For diagnosing, the age at the examination is important and not the current age.

The age category is calculated according to the CAMDEX-R/N⁸ manual (Roth et al., 2005):

$$Age\ category := \begin{cases} 65 - 69 & \text{if } age < 70 \\ 70 - 74 & \text{if } 70 \leq age < 75 \\ 75 - 79 & \text{if } 75 \leq age < 80 \\ 80 - 84 & \text{else} \end{cases}$$

⁷The category 65-69 years is for age ≤ 69 years, and the category 80-84 years is for age $80 \leq$ years

⁸The CAMCOG-R/N is the dutch version of the CAMCOG-R

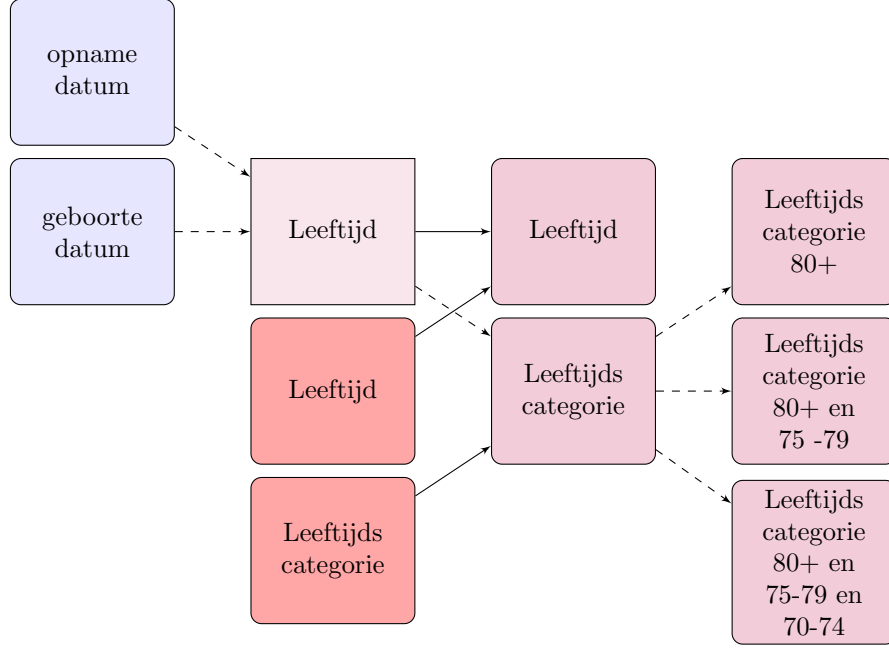


Figure 3.2: Age and Age category attributes evolution

No only the age-related attributes were calculated, the *UMC St. Radboud* dataset misses also the education level attributes. The educational attributes are displayed in Figure 3.3. For the ‘education level’ attribute, I used the following rule:

$$Education\ level := \begin{cases} 1 & \text{if } education = \text{lagere school zes of minder klassen} \\ 2 & \text{if } education = \text{voortgezet onderwijs, MULO,} \\ & \text{LO met meer dan 6 leerjaren, LBO, ULO} \\ 3 & \text{if } education = \text{middelbaar en hoger onderwijs} \end{cases}$$

The age attributes ‘Leeftijds categorie 80+’, ‘Leeftijds categorie 80+ en 75-79’ and ‘Leeftijds categorie 80+, 75-79 en 70-74’ are added separately because the age category is a nominal instead of an ordinal attribute. The same applies to the attribute of the education level, this attribute is split into the two nominal attributes ‘Opleidings niveau 1’ and ‘Opleidings niveau 2’.

3.3.2 CAMCOG test results attributes

Despite that both the datasets contain the test results of the CAMCOG and MMSE, the datasets needed a couple of processing steps. An overview with the evolution of the CAMCOG attributes is shown in Figure 3.4.

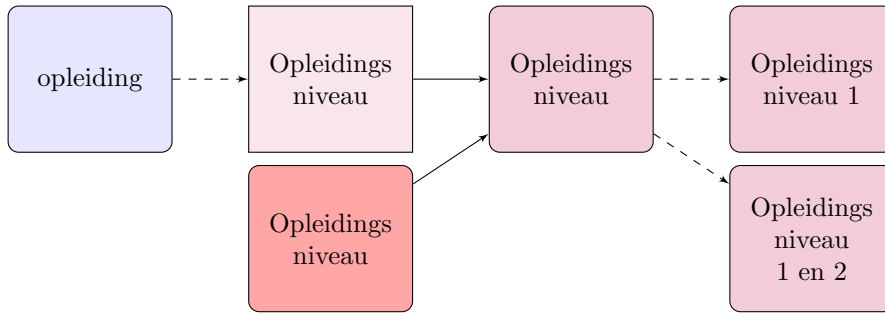


Figure 3.3: Education Level attribute evolution

The *Zorggroep* dataset contains test results of all the CAMCOG sections, scales and subscales. But the *UMC St. Radboud* data does not contain the subscales of the CAMCOG. Because the subscale can not be recalculated, these are removed.⁹

Most scales were present in both datasets and needed no processing. But in the *UMC St. Radboud* dataset, the scales ‘attention’ and ‘calculation’ are combined into one scale. It is not possible to separate this attribute and I can not recalculate them. Removing would be a waste of valuable data. Therefore the combined scales ‘attention’ and ‘calculation’ maintained, and were also combined in *De Zorggroep* dataset.¹⁰

$$CAMCOG\ SubAandachtEnRekenen = Aandacht + Rekenen$$

Also the memory section and non-memory section are missing in the *UMC St. Radboud* dataset. In contrast to the subscales, this section can be calculated. I used the CAMDEX manual (Roth et al., 2005) to calculate the memory section and non-memory section.¹¹

$$\begin{aligned}
 CAMCOG\ Geheugensectie &= CAMCOG\ SubGeheugen + CAMCOG\ SubOrientatie \\
 CAMCOG\ Niet\ geheugensectie &= CAMCOG\ SubTaal + CAMCOG\ SubPerceptie \\
 &\quad + CAMCOG\ SubPraxis + CAMCOG\ SubAbstractRedenen \\
 &\quad + CAMCOG\ SubAandachtEnRekenen
 \end{aligned}$$

⁹In Figure 3.4 are these attributes marked with ‘1’.

¹⁰The combining of scales ‘attention’ and ‘calculation’ is shown by number ‘2’ in Figure 3.4

¹¹The calculated memory and non-memory section are marked by number ‘3’ in Figure 3.4

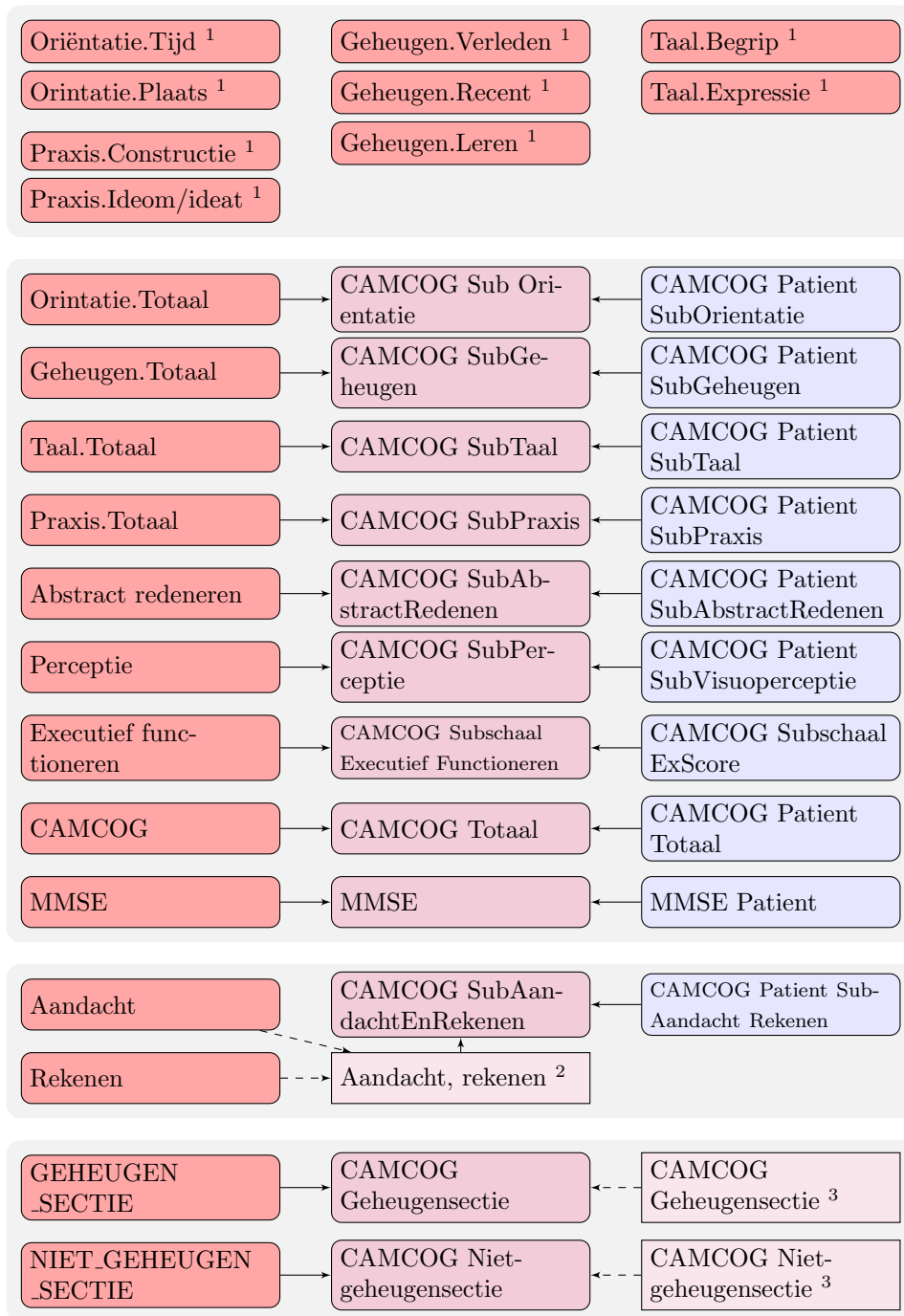


Figure 3.4: The CAMCOG attributes evolution

3.3.3 CAMCOG norm attributes

The CAMCOG uses a distinction in which the norms are categorized (Roth et al., 2005), see Table 3.2. These are used to calculate individual norms, thresholds and cumulative scores.

The following attributes have a median and a threshold:

- the total score and
- the scales:
 - orientation,
 - language,
 - memory,
 - attention,
 - abstract thinking,
 - and perception.

I calculated both the median and a threshold for the total score and those scales as described in the CAMDEX manual, based on individual characteristics as described above. Even though ‘attention’ and ‘calculation’ are combined, I added the norm for ‘attention’ by itself. For ‘calculation’ is no norm or threshold (Roth et al., 2005).

The scale ‘executive functioning’ has no norms for the absence of validation and standards (Roth et al., 2005). Also the scale ‘praxis’ has no norms calculated (Roth et al., 2005).

The sections memory and non-memory have no median or threshold, but a score. This is a cumulative percentage. The cumulative scores of the memory and non-memory section are also represented in the dataset by the attributes: ‘CAMCOG Geheugensectie Score’ and ‘CAMCOG Niet-geheugensectie Score’. These attributes are calculated according to the tables with the cumulative percentages of the CAMDEX manual (Roth et al., 2005). The table with the cumulative percentages of the memory section is given in Table 3.3.

3.3.4 Dementia types classification

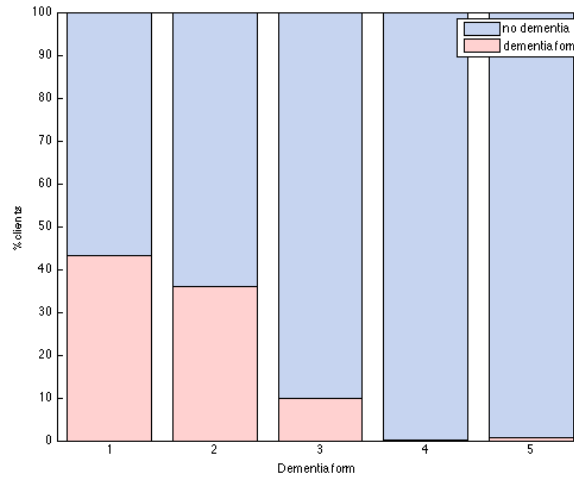
The class ‘dementia’ was available in both datasets and no further processing was necessary.

Initially the dataset contained also the classes for the four most common types of dementia: Alzheimer’s Disease (AD), Vascular Dementia (VS), Lewy Body

Test score	Education level 1	Education levels 2&3	Test score	Education level 1	Education levels 2&3
19-20	3	1	30	43	24
21-22	7	1	31	54	34
23	9	2	32	70	49
24-25	13	3	33	85	69
26	15	6	34	94	89
27	18	9	35	99	95
28	23	11	36	100	99
29	32	17	37		100

Table 3.3: The cumulative scores of the memory section determined for each test score and education level

Dementia (LB) and Frontotemporal Dementia (FT). These classes were present in the *UMC St. Radboud* dataset. I extracted more details about the diagnosis from the screening reports. I transformed this information into the four dementia type classes.



(1 = Total, 2 = Alzheimer's Disease , 3 = Vascular Dementia, 4 = Lewy Body Dementia, 5 = Frontotemporal Dementia)

Figure 3.5: Dementia types balances

As described above, the dataset contained the classes for the four types of dementia. But the balance, the ratio of the subjects of the smallest and largest group, is not sufficient for most of the types. Only the balance of Alzheimers Disease is above 0.3. The balances of the different dementia types are shown in Tabel 3.4 and Figure 3.5.

	TRUE/ total subjects	Total Dataset balance	UMC St. Radboud balance	De Zorggroep balance
AD	241 /670	0.3597	0.3687	0.0526
VS	67 /670	0.10	0.0968	0.1579
LB	1 /670	0.0015	0.0015	0.0
FT	4 /670	0.0060	0.0061	0.0

Table 3.4: The balances of the dataset for the different dementia types

As mentioned before *De Zorggroep* has giving me access to the screening reports. The determination of the dementia type is not relevant in these reports. In case of positive diagnosis of dementia, further examination could take place. So there could be more cases of AD, VS, LB or FT in the dataset.

The balance of the dementia types is too low and the dementia types could be incomplete. Therefore I decided to skip the classification on the dementia types.

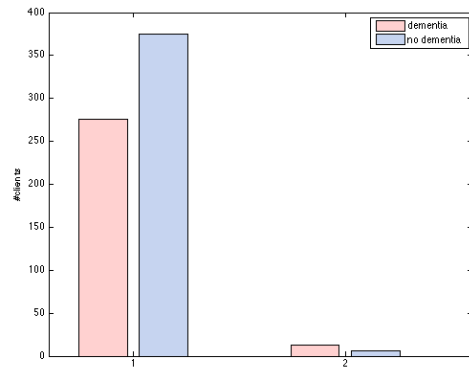
3.4 The dataset for Round 1

The first dataset contains both *De Zorggroep* and the *UMC St. Radboud* dataset. Of the 46 attributes I selected from the screening reports and *UMC St. Radboud* dataset, 38 attributes remained. This attributes are listed in Table 3.5. Some characteristics are shown in Figure 3.6.

Attribute/class	Scale of measure	Adjustment number
Geslacht	M,V	0
Leeftijd	numeric	2
'Leeftijdscategorie 80+'	TRUE, FALSE	1
'Leeftijdscategorie 80+ en 75 -79'	TRUE, FALSE	1
'Leeftijdscategorie 80+ en 75 -79 en 70-74'	TRUE, FALSE	1
Leeftijdscategorie	75-79, 80-84, 70-74, 65-69	1
'Opleidingsniveau 1'	TRUE, FALSE	1
'Opleidingsniveau 1 en 2'	TRUE, FALSE	1
Opleidingsniveau	numeric	2
'CAMCOG SubOrientatie'	numeric	0
'CAMCOG SubGeheugen'	numeric	0
'CAMCOG SubTaal'	numeric	0
'CAMCOG SubPraxis'	numeric	0
'CAMCOG SubAandachtEnRekenen'	numeric	2
'CAMCOG SubAbstractRedenen'	numeric	0
'CAMCOG SubPerceptie'	numeric	0
'CAMCOG SubschaalExecutiefFunctioneren'	numeric	0
'CAMCOG Totaal'	numeric	0
'CAMCOG Geheugensectie'	numeric	0
'CAMCOG Niet-geheugensectie'	numeric	2
MMSE	numeric	0
'CAMCOG Totaal Mediaan'	numeric	2
'CAMCOG Totaal Grensscore'	numeric	2
'CAMCOG Geheugensectie Score'	numeric	2
'CAMCOG Niet-geheugensectie Score'	numeric	2
'CAMCOG Orientatie Mediaan'	numeric	2
'CAMCOG Orientatie Grensscore'	numeric	2
'CAMCOG Taal Mediaan'	numeric	2
'CAMCOG Taal Grensscore'	numeric	2
'CAMCOG Geheugen Mediaan'	numeric	2
'CAMCOG Geheugen Grensscore'	numeric	2
'CAMCOG Aandacht Mediaan'	numeric	2
'CAMCOG Aandacht Grensscore'	numeric	2
'CAMCOG AbstractRedenen Mediaan'	numeric	2
'CAMCOG AbstractRedenen Grensscore'	numeric	2
'CAMCOG Perceptie Mediaan'	numeric	2
'CAMCOG Perceptie Grensscore'	numeric	2
Dementie	TRUE, FALSE	0

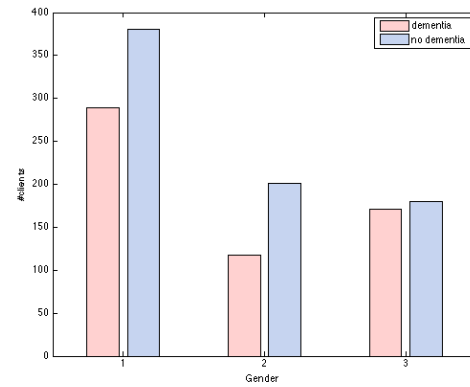
Adjustment numbers: [0] No adjustment, both present, [1] Transformed,
[2] Calculated and [3] Removed

Table 3.5: Attribute list of the first dataset



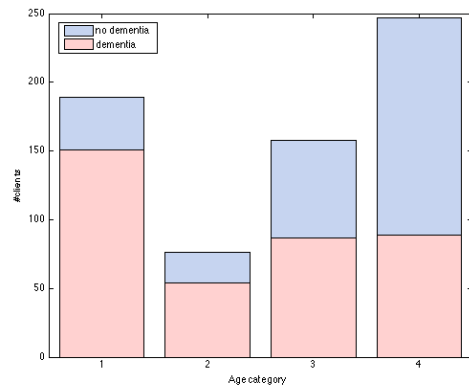
(1 = UMC St. Radboud, 2 = De Zorggroep)

(a) Dementia per health care centre



(1 = Total, 2 = Male, 3 = Female)

(b) Dementia versus Gender



(1 = 65 – 69, 2 = 70 – 74, 3 = 75 – 79, 4 = 80 – 85)

(c) Dementia versus Age category

	Total	RD	ZG
Original	-	192	87
#attributes			
#subjects	670	651	19
#subject with dementia	289	176	13
Balance	43 %	27%	68 %
Mean age	74,68	74,48	85,58
Gender M/V	319	308	11
	/351	/343	/8

(d) Some numbers of the datasets

Figure 3.6: Characteristics of the first dataset

3.5 The Weka workbench

For testing classifiers I used version 3.6.7 of the Weka workbench and the Weka manual (Bouckaert et al., 2012). The Weka workbench is a collection of state-of-the-art machine learning algorithms and data pre-processing tools (Frank et al., 2005). Weka supports the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically and visualizing both the input data and the result of learning (Frank et al., 2005).



Figure 3.7: Weka logo for version 3.6.7 for Mac

As discussed in Section 2.5, I decided to apply the following techniques: a logistic regression, a neural network, a nearest neighbors and a support vector machine. Weka has different classifiers for these techniques. The classifiers in Weka, which I applied, are: ‘Logistic’, ‘Multilayer Perceptron’, ‘NNge’ and ‘SMO’. The classifiers are described below.

Logistic Classifier for building and using a multinomial logistic regression model with a ridge estimator (Bouckaert et al., 2012).

Multilayer Perceptron A classifier that uses backpropagation to classify instances (Bouckaert et al., 2012).

NNge Nearest-neighbor-like algorithm using non-nested generalized exemplars (which are hyperrectangles that can be viewed as if-then rules) (Bouckaert et al., 2012).

SMO Implements John Platt’s sequential minimal optimization algorithm for training a support vector classifier (Bouckaert et al., 2012).

3.6 Experiments Round 1

In order to answer the research questions, this first round of experiments tests the classifiers on two different characteristics: robustness and comparing rate. The characteristics require a different experimental setting. The settings are described below.

3.6.1 Experimental setting 1, cross-validation

Experimental setting 1 is focused on the first research question:

Can artificial intelligence techniques be used for the automated diagnosis of dementia? If so, which techniques are accurate or robust?

In order to test the robustness of the classifiers, I applied ten-fold cross-validation.

Cross-validation is a technique that reduces variability in the performance of a classifier (Russell & Norvig, 2003). The use of cross-validation helps in assessing how the classifier would generalize against an unknown dataset, therefore increasing the reliability of the performance. With k -folds cross-validation, the experiment is run k times. Every time another $(\frac{1}{k})^{\text{th}}$ part is put aside for the test set. The performance of the experiment is with k -folds the average of the k performances.

3.6.2 Experimental setting 2, split

In the second experimental setting I will be able to compare the predictions of the four classifiers, because the test set will be the same for all classifiers. Not only I will be able to compare predictions, the expert can also evaluate prediction of individual subjects. This is necessary for research question 2:

How can artificial intelligence techniques be used in practice?
Which techniques are consistent with the experts?

The total dataset was split in a random training set, containing 66% of the subjects, and a test set, containing the other 34% subjects (440 versus 230 subjects). The balance is not held intact precisely, but has just slightly shifted. In the train set 43% of the subjects is from the dementia class and in the test set this percentage is 42%.

Chapter 4

Results Round 1

4.1 Setting 1

The results of setting 1, the ten-fold cross-validation, are listed in Table 4.1. The parameters of the classifiers used for these performances in my research, are shown in Appendix D. For more details of the results of setting 1, see Appendix E.

Scheme:	MLP ₁	SMO ₁	Logistic ₁	k-NN ₁
Correctly classified instances	533 ** 79.55 %	541 ** 80.75 %	534 ** 79.70 %	525 ** 78.36 %
Kappa statistic	0.58	0.61	0.59	0.56
Total number of instances	670	670	670	670

**p < 1 · 10⁻¹⁴

Table 4.1: Best results from the Weka classifiers: MLP, SMO, Logistic and NNge in Round 1 with setting 1: cross-validation

4.2 Setting 2

In this setting, the training and test set are separated. In Table 4.2, the first ten predictions and the performance of the four classifiers are displayed.

The classifiers were *all four* correct in 167 cases (69.58 %), three out of four classifiers were correct in 18 cases (7.50 %), only two correct predictions given by the four classifiers in 6 cases (2.50 %), only one classifier was correct in 11 cases (4.58 %) and in 33 cases (13.75 %) none of the classifiers could find the correct diagnosis. These percentages are displayed in Figure 4.1.

Inst	Actual class	Predicted MLP ₂	Predicted SMO ₂	Predicted Logistic ₁	Predicted k-NN ₂	Correct predictions
1	TRUE	TRUE	TRUE	TRUE	TRUE	4
2	FALSE	TRUE X	FALSE	FALSE	FALSE	3
3	TRUE	FALSE X	FALSE X	FALSE X	FALSE X	0
4	TRUE	TRUE	TRUE	FALSE X	FALSE X	2
5	TRUE	TRUE	TRUE	TRUE	TRUE	4
6	TRUE	TRUE	TRUE	TRUE	TRUE	4
7	FALSE	TRUE X	TRUE X	TRUE X	TRUE X	0
8	TRUE	TRUE	TRUE	TRUE	TRUE	4
9	TRUE	TRUE	TRUE	TRUE	TRUE	4
10	FALSE	FALSE	FALSE	FALSE	FALSE	4

Correctly classified instances	183 *	182 *	182 *	178 *	
Kappa statistic	79.57 %	79.13 %	79.13 %	77.39 %	
Total number of instances	0.58	0.20	0.57	0.52	
	230	230	230	230	

* $p < 1 \cdot 10^{-9}$, X = incorrect prediction

Table 4.2: Predictions of the first ten instances and performance of the four classifiers in Round 1 with setting 2: split

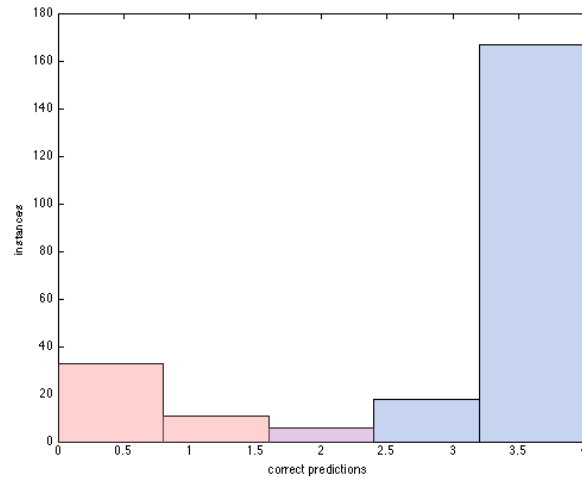


Figure 4.1: Histogram of the correct predictions of the four classifiers

The performance does not improve, when using plurality voting with the four or three best classifiers. See Table 4.3 for the performance, when using plurality voting. An explanation for this could be that in most cases the classifiers make the same prediction, even when they are incorrect (all correct: 68.58 % + all incorrect: 13.75 % = 83.33 %).

MLP ₂	SMO ₂	Logistic ₁	k-NN ₂
183/230 *	182/230 *	182/230 *	178/230 *
79.57 %	79.13 %	79.13 %	77.39 %
182/230 *			
79.13 %			
183/230 *			
79.57 %			

*p < 1 · 10⁻⁹

Table 4.3: Performance (correct/total instances) of the four classifiers, the three best classifiers combined and all classifiers combined

Chapter 5

Evaluation and Round 2

5.1 Discussion with expert

In the couple of interviews with the expert we discussed several topics, like dementia, the diagnosis process and neuropsychological tests. These were described in previous chapters. The evaluation of the performance and individual subjects was another topics. These and the consequences of incorrect diagnosing will be discussed in the next sections.

5.1.1 Evaluating the performance of the classifiers

The performance of the classifiers is near 80 %. The expert is not surprised to hear this result: the classifier s use only the test results and some personal details (shown in Figure 1.2) and not all the data that the psychologist uses.

The CAMCOG tells something about the cognitive functioning of the client. But low scores are not equal to dementia. As psychologist, you have to exclude other factors and interpret the score to the situation.

Because dementia is deterioration of the cognitive functions, the psychologist is looking for decay. The score of a client is compared to the standard of healthy subjects with the same age, education level, and so on as the clients. Nevertheless the client could have scored, in healthy state, higher or lower than the standard given in the CAMCOG. This means that a high score could be bad, if the client used to be developed above average.

Factors can influence the test results during the examination of the test. A psychologist takes into account whether the client has a handicap. In this way, visual, auditory, phatic or motor impairment issues can make the test more difficult. For example rheumatism can impede writing tasks. Factors like motivation or being tired influence the test results as well.

Some factors influence the CAMCOG specifically. Depression or even depressive symptoms decrease the CAMCOG score.

There are also other circumstances, like CVA, that decrease the cognitive functions. Then the low scores are not caused by dementia.

The psychologist also look at the illness awareness and illness insight. When a client has illness awareness but no insight, the client knows that he/she has a malfunction. However the client has no insight into the consequences of the illness. For example, a clients, who says:

“I’m forgetful but there is nothing wrong.”

is a client with illness awareness but no insight. However when a client says:

“my forgetfulness is a burden for my partner”

tells the clients has both illness awareness and illness insight. It is not possible to have illness insight but no awareness.

5.1.2 Evaluating individual subjects

The instances of the test set of the second setting are both *De Zorggroep* and *UMC St. Radboud* subjects. The expert can get enough details from the original screening reports to reconstruct the diagnosis of *De Zorggroep* subjects. This is not possible with the *UMC St. Radboud* subjects. Of the 230 instances, 9 came from *De Zorggroep*. These nine instances are shown in Table 5.1. The classifiers predicted most instances correct. Only in four cases one or more classifiers predicted incorrect. Instance 3 and 7 were classified incorrectly by all four classifiers. The MLP classifier was wrong with Instance 2. In case of instance 4, the classifiers Logistic and k-NN disagreed with the psychologist.

Inst	Actual class	Predicted MLP ₂	Predicted SMO ₂	Predicted Logistic ₁	Predicted k-NN ₂	ID1 in dataset	Sekse	Leeftijds-categorie	Opleidings niveau
1	TRUE	TRUE	TRUE	TRUE	TRUE	1	V	80-84	2
2	FALSE	TRUE	FALSE	FALSE	FALSE	2	V	80-84	2
3	TRUE	FALSE	FALSE	FALSE	FALSE	3	V	75-79	2
4	TRUE	TRUE	TRUE	FALSE	FALSE	4	M	80-84	2
5	TRUE	TRUE	TRUE	TRUE	TRUE	9	V	80-84	1
6	TRUE	TRUE	TRUE	TRUE	TRUE	11	V	80-84	1
7	FALSE	TRUE	TRUE	TRUE	TRUE	14	M	80-84	1
8	TRUE	TRUE	TRUE	TRUE	TRUE	15	M	80-84	2
9	TRUE	TRUE	TRUE	TRUE	TRUE	16	V	80-84	2

Table 5.1: The nine instances of *De Zorggroep* in the test set of setting 2

The expert reconstructed the diagnosis of these four instances. The first three were originally diagnosed by a colleague of the expert. Instance 7 was diagnosed by the expert.

Instance 2

Subject 2 was diagnosed with no dementia. The expert confirms this diagnosis, because the cognitive impairment was caused by a CVA.

The majority of the classifiers were indeed correct. The MLP failed because of the CVA. In the Section 5.1.1, the influence of the CVA on the diagnosis was discussed.

Instance 3

The psychologist was not thorough with the examination, for example the expert missed the heteroanamnese in this case. The expert assumed there was not a CVA or other medial deficits and concluded and concluded an early phase of dementia.

Although all classifiers predicted negatively, the diagnosis is still positive. It could be possible that the classifiers did not recognize the dementia, because it is still in an early phase.

Instance 4

Subject 4 was originally diagnosed with dementia. The classifiers did not collective agree: two classifiers agreed and two classifiers disagreed. The experts diagnosed subject 4 with no dementia, but the cognitive disorder NOS¹² (DSM IV).

The diagnosis was originally incorrect. The classifiers, which predicted initially correct, are now incorrect, and vice versa. This causes the performance of each classifier to change with 0.43%. The performance of the MLP and SMO decreased with 0.43%. The performance of the Logistic an k-NN increased with 0.43%.

Instance 7

The diagnosis of subject 7 is no dementia. The expert confirms this diagnosis. The subject has no dementia, because the cognitive impairment was caused by a CVA. Also the visual impairment (Hemianopsia) and motor impairment probably influenced the test results negatively.

The classifiers probably recognized the cognitive impairment, but were not capable of distinguishing between the CVA and dementia. This is also the case with subject 3 and discussed in Section 5.1.1.

5.1.3 Recommendations

As discussed previously, the CAMCOG is not always conclusive. Human observations could exclude other causes of abnormal cognitive functions. Based on the

¹²Not Otherwise Specified

interview with the expert, I listed the following recommended additional attributes:

- Motivated - Is the client motivated during the examination?
- Depressed - Has the client depressive symptoms or is the client depressed?
- Sense of failure - Is the client aware of failures during the examination?
- Visual impairment - Has the client visual impairment?
- Auditory impairment - Has the client auditory impairment?
- Phatic impairment - Has the client phatic impairment?
- Motor impairment - Has the client motor impairment?
- Illness awareness - Is the client aware that the he/she is ill?
- Illness insight - Has the client insight in the clients' illness?
- CVA - Has the client had a cerebrovascular accident (stroke)?

5.1.4 Consequences of incorrect diagnosing

When incorrect diagnosed with dementia, the consequences relate to the cause of the impaired cognitive functions. For example a depression can cause a low cognitive score. When the client is no longer depressed, the cognitive functions return. Client and relatives will still anticipate on dementia, which has negative consequences.

In case of a cognitive disorder, the influence of the diagnosis is not that large. Most cognitive disorders develop dementia and the diagnosis is just pre-mature. Nevertheless the client has no dementia, the cognitive functions are still abnormal. In other situations the client would still be dysfunctional. Worst case scenario is when the client is unnecessary transferred to a psychogeriatric department. Although this situation is not very likely to happen.

The consequences, when incorrectly diagnosed with no dementia, are less dramatic then incorrectly diagnosed with dementia. In case of uncertainty, the psychologist can do a second screening examination, for example a half year later. Because dementia is deterioration of the cognitive functions, the psychologist can compare the two timeframes and look for diminution of cognitive capabilities. The client and relatives will live longer in uncertainty, but are at least not misinformed about the diagnosis.

In conclusion, when it comes to incorrect diagnosing, false positives are less desirable false negatives.

5.1.5 Round 2A: new datasets

The screening reports of *De Zorggroep* contain more information than original extracted attributes. With the help of the expert, the recommended additional attributes were extracted from these reports. Most of the additional attributes

could not be found in the *UMC St. Radboud* dataset and attributes can not be added to the *UMC St. Radboud* dataset. Round 2A is based only on *De Zorggroep* dataset, because of the flexibility of this dataset. In Table 5.2, the list of the additional attributes is shown.

Gemotiveerd	TRUE, FALSE
Besef falen	TRUE, FALSE
Depressief	TRUE, FALSE
Visuele beperkingen	TRUE, FALSE
Gehoors beperkingen	TRUE, FALSE
Fatische beperkingen	TRUE, FALSE
Motorische beperkingen	TRUE, FALSE
Ziekte besef	TRUE, FALSE
Ziekte inzicht	TRUE, FALSE
CVA	TRUE, FALSE

Table 5.2: Additional attributes for Round 2A

De Zorggroep dataset used in Round 1 contained the 19 subjects. During the research, six other screening reports became available. These reports are not processed in Round 1, but only used in Round 2A¹³.

Of the 25 subjects in the new dataset, 17 subjects have dementia. The balance is different from the total dataset used for Round 1 with setting 1 and 2 ($17/25 = 0.68$ versus $289/670 = 0.43$).

For this round, the experimental settings were set according to setting 1 (described in Section 3.6.1). This means I used ten-fold cross-validation. In order to make a more fair comparison possible, not only the dataset with additional attributes was tested. The dataset with the same subjects was tested with the original 38 attributes of Round 1.

5.1.6 Results Round 2A

The results of ten-fold cross-validation with 38 attributes (without additional attributes) are shown in Table 5.3. In Table 5.4, the results with the 48 attributes (with additional attributes) are displayed.

The performance and confusion matrices of the four classifiers on both datasets are combined in Table 5.5.

¹³The reports are also not used in Round 2B

Scheme:	MLP ₃	SMO ₃	Logistic ₁	k-NN ₃
Correctly classified instances	17 [†] 68 %	18 [†] 72 %	20 [†] 80 %	12 [†] 48 %
Kappa statistic	0	0.38	0.52	-0.24
Total number of instances	25	25	25	25

[†]_p > 0.1

Table 5.3: Best results from the Weka classifiers: MLP, SMO, Logistic and NNge in Round 2A with the original 38 attributes

Scheme:	MLP ₄	SMO ₃	Logistic ₂	k-NN ₄
Correctly classified instances	17 [†] 68 %	17 [†] 68 %	10 [†] 40 %	13 [†] 52 %
Kappa statistic	0	0.26	-0.25	-0.27
Total number of instances	25	25	25	25

[†]_p > 0.1

Table 5.4: Best results from the Weka classifiers: MLP, SMO, Logistic and NNge in Round 2A with the original and additional attributes

Original 38 attributes

Scheme:	MLP ₃			SMO ₃			Logistic ₁			k-NN ₃		
	a	b	←	a	b	←	a	b	←	a	b	←
Confusion matrix	17	0	a = True	13	4	a	15	2	a	11	6	a
	8	0	b = False	3	5	b	3	5	b	7	1	b
	a	b	←	a	b	←	a	b	←	a	b	←
Confusion matrix	17	0	a = True	13	4	a	8	9	a	13	4	a
	8	0	b = False	4	4	b	6	2	b	8	0	b
Scheme:	MLP ₄			SMO ₃			Logistic ₂			k-NN ₄		

Original 38 attributes and additional attributes

Table 5.5: Performance and confusion matrices of the four classifiers based on the original 38 attributes dataset and the additional attributes dataset

5.2 Round 2B: attribute reduction

In Round 2B, instead of adding attributes, the performance may improve with a reduction of the attributes. With less attributes, the future space will be smaller.

The attributes with the lowest information gain and gain ratio are best candidates for removal. I used the Weka tools `InfoGainAttributeEval`, `GainRatioAttributeEval` and `Ranker` to calculate the information gain and gain ratio.

InfoGainAttributeEval Evaluates the worth of an attribute by measuring the information gain with respect to the class (Bouckaert et al., 2012).

$$InfoGain(Class, Attribute) = H(Class) - H(Class|Attribute)$$

GainRatioAttributeEval Evaluates the worth of an attribute by measuring the gain ratio with respect to the class (Bouckaert et al., 2012).

$$GainR(Class, Attribute) = (H(Class) - H(Class|Attribute)) / H(Attribute)$$

Ranker Ranks attributes by their individual evaluations. (Bouckaert et al., 2012).

I calculated the *sum* and the *distance* of the info gain and gain ratio to rank the attributes. Top attributes are selected for the new dataset. In Table 5.6 the removed attributes are indicated with a grey colour. The following twenty attributes remain:

- Attribute
- Leeftijd
- Leeftijdscategorie
- Leeftijdscategorie 80+
- Leeftijdscategorie 80+ en 75 -79
- MMSE
- CAMCOG Geheugensectie
- CAMCOG Geheugensectie Score
- CAMCOG Niet-geheugensectie
- CAMCOG Niet-geheugensectie Score
- CAMCOG Perceptie Grensscore
- CAMCOG SubAandachtEnRekenen
- CAMCOG SubAbstractRedenen
- CAMCOG SubGeheugen
- CAMCOG SubOrientatie
- CAMCOG SubPerceptie
- CAMCOG SubPraxis
- CAMCOG SubschaalExecutiefFunctioneren
- CAMCOG SubTaal
- CAMCOG Totaal

Attribute	Info gain (1)	Gain ratio (2)	Sum (1) & (2)	Distance (1) & (2)
MMSE	0,35	0,18	0,53	0,40
CAMCOG Totaal	0,32	0,17	0,50	0,37
CAMCOG Geheugensectie	0,31	0,18	0,49	0,36
CAMCOG SubOrientatie	0,28	0,15	0,43	0,32
CAMCOG SubGeheugen	0,28	0,14	0,42	0,31
CAMCOG Geheugensectie Score	0,24	0,14	0,38	0,28
CAMCOG Niet-geheugensectie	0,22	0,13	0,35	0,26
CAMCOG Niet-geheugensectie Score	0,17	0,11	0,29	0,21
CAMCOG SubTaal	0,16	0,08	0,24	0,18
CAMCOG SubschaalExecutiefFunctioneren	0,14	0,09	0,23	0,17
CAMCOG SubAandachtEnRekenen	0,13	0,10	0,23	0,16
CAMCOG SubPerceptie	0,14	0,08	0,22	0,16
Leeftijd	0,13	0,07	0,20	0,15
CAMCOG SubPraxis	0,12	0,08	0,20	0,14
CAMCOG SubAbstractRedenen	0,11	0,07	0,18	0,13
Leeftijdscategorie 80+ en 75 -79	0,08	0,09	0,17	0,12
Leeftijdscategorie	0,10	0,05	0,16	0,12
CAMCOG Perceptie Grensscore	0,08	0,08	0,16	0,12
Leeftijdscategorie 80+	0,08	0,08	0,15	0,11
CAMCOG AbstractRedenen Mediaan	0,09	0,06	0,14	0,10
CAMCOG Perceptie Mediaan	0,09	0,06	0,14	0,10
CAMCOG AbstractRedenen Grensscore	0,09	0,06	0,14	0,10
CAMCOG Totaal Mediaan	0,07	0,07	0,14	0,10
CAMCOG Totaal Grensscore	0,07	0,07	0,14	0,10
Leeftijdscategorie 80+ en 75 -79 en 70-74	0,07	0,08	0,14	0,10
CAMCOG Taal Mediaan	0,08	0,05	0,13	0,10
CAMCOG Taal Grensscore	0,08	0,05	0,13	0,10
CAMCOG Geheugen Mediaan	0,06	0,06	0,13	0,09
CAMCOG Geheugen Grensscore	0,03	0,06	0,09	0,07
Opleidingsniveau 1	0,03	0,04	0,06	0,05
CAMCOG Aandacht Mediaan	0,01	0,01	0,02	0,02
Opleidingsniveau 1 en 2	0,01	0,01	0,02	0,02
Geslacht	0,01	0,01	0,02	0,01
CAMCOG Aandacht Grensscore	0,01	0,01	0,01	0,01
Opleidingsniveau	0,01	0,01	0,01	0,01
CAMCOG Orientatie Grensscore	0,00	0,00	0,00	0,00
CAMCOG Orientatie Mediaan	0,00	0,00	0,00	0,00

(1) attributeSelection.InfoGainAttributeEval

with attributeSelection.Ranker -T -1.7976931348623157E308 -N -1

(2) attributeSelection.GainRatioAttributeEval

with attributeSelection.Ranker -T -1.7976931348623157E308 -N -1

$Sum = (1) + (2), \quad Distance = \sqrt{(1)^2 + (2)^2}$

Table 5.6: The attributes ranked on information gain and gain ratio

5.2.1 Results Round 2B

The results of Round 2B are shown in Table 4.1. For more details see Appendix F.

Scheme:	MLP ₅	SMO ₄	Logistic ₃	k-NN ₅
Correctly classified instances	533 **	540 **	542 **	530 **
	79.55 %	80.60 %	80.90 %	79.10 %
Kappa statistic	0.58	0.60	0.61	0.57
Total number of instances	670	670	670	670

**p < 1 · 10⁻¹⁴

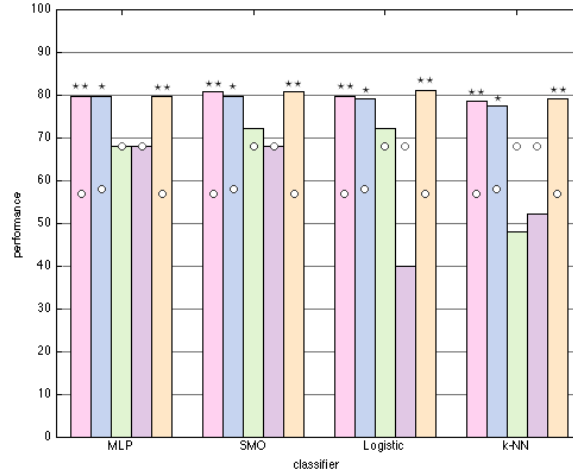
Table 5.7: Best results from the Weka classifiers: MLP, SMO, Logistic and NNge in Round 2B

Chapter 6

Conclusion and Discussion

6.1 The results of the AI techniques

In Figure 6.1 the performances of the four classifiers over all rounds are displayed. The results of all rounds are shown in Table 6.1.



○ = #class/#total; * $p < 1 \cdot 10^{-9}$, ** $p < 1 \cdot 10^{-14}$, else $p > 0.1$
Round 1 setting 1 & 2, Round 2A with & without additional attributes and Round 2B

Figure 6.1: Histogram of the correct predictions of the four classifiers in all rounds

The first results, Round 1 with setting 1, were promising:

MLP: 79.55 %, SMO: 80.75 %, Logistic: 79.70 % and k-NN: 78.36 %.

All classifiers scored significantly better than random at a succes rate of approximately 80% ($p < 1 \cdot 10^{-14}$). These results, while using ten-fold cross-validation, were not very different from setting 2 (with 66% split):

MLP: +0.03, SMO: -1.62, Logistic: -0.57 and k-NN: -0.97 ¹⁴.

Those results were also significant ($p < 1 \cdot 10^{-9}$). By using cross-validation when training the classifiers, the variability of the performances is reduced, increasing their robustness. The use of cross-validation helps in assessing how the classifier would generalize against an unknown dataset, so that the performance shows how robust the trained classifier is against unknown datasets.

None of the classifiers outperforms the others. I think the data is not a specific problem solved best by only one classifier technique. Other AI techniques could also be promising as well and should be tested in further research.

The four classifiers do not differ a lot in performance, and give most of the time the same predictions (83.33 %). Only in 12.08 % one classifier disagrees with the others, but in only 7.50 % these are correct.

The performance does not improve, when using plurality voting with the four or three best classifiers. An explanation for this could be that in most cases the classifiers make the same prediction, even when they are incorrect.

In the evaluation together with the expert, the performance and predictions of the classifiers were discussed. The expert stated that the input missed the observations and data to interpret the CAMCOG scores. For example all classifiers could predict dementia, but the client could have had a CVA. The CVA would cause low cognitive functions and therefore the client has no dementia.

It is likely that the classifiers learned to identify the low cognitive functions, but failed to learn all the causes of the CAMCOG score. This would explain why all four classifiers fail in predicting the same subjects. By adding the recommended attributes to the dataset, the performance should increase. The new dataset was small (25 subjects), and probably too small for training the classifiers. In this Round, 2A, the results are scattered and insignificant.

In order to improve the performance in another way than adding attributes, Round 2B was designed to remove attributes. These results were significant and generally the same as the results of Round 1. The differences are slightly better:

MLP: +0.0, SMO: -0.15, Logistic: +0.40 and k-NN: +0.74 ¹⁵.

The performance could be influenced by the variation in psychologists, which diagnosed the patients. But I think this influence could be insignificant. The subjects in the *UMC St. Radboud* dataset are multiple times diagnosed with dementia. Only *De Zorggroep* screening reports are made by one psychologist. This is 2.84 % of the total dataset.

¹⁴difference is 'Round 1, setting 1 – Round 1, setting 2',

¹⁵difference is 'Round 1, setting 1 – Round 2B',

Round 1, setting 1	MLP ₁	SMO ₁	Logistic ₁	k-NN ₁
Correctly classified instances	533 **	541 **	534 **	525 **
	79.55 %	80.75 %	79.70 %	78.36 %
Kappa statistic	0.58	0.61	0.59	0.56
Total number of instances	670	670	670	670
Round 1, setting 2	MLP ₂	SMO ₂	Logistic ₁	k-NN ₂
Correctly classified instances	183 *	182 *	182 *	178 *
	79.57 %	79.13 %	79.13 %	77.39 %
Kappa statistic	0.58	0.20	0.57	0.52
Total number of instances	230	230	230	230
Round 2A, without additional attributes	MLP ₃	SMO ₃	Logistic ₁	k-NN ₃
Correctly classified instances	17 †	18 †	20 †	12 †
	68 %	72 %	80 %	48 %
Kappa statistic	0	0.38	0.52	-0.24
Total number of instances	25	25	25	25
Round 2A, with additional attributes	MLP ₄	SMO ₃	Logistic ₂	k-NN ₄
Correctly classified instances	17 †	17 †	10 †	13 †
	68 %	68 %	40 %	52 %
Kappa statistic	0	0.26	-0.25	-0.27
Total number of instances	25	25	25	25
Round 2B	MLP ₅	SMO ₄	Logistic ₃	k-NN ₅
Correctly classified instances	533 **	540 **	542 **	530 **
	79.55 %	80.60 %	80.90 %	79.10 %
Kappa statistic	0.58	0.60	0.61	0.57
Total number of instances	670	670	670	670

†p > 0.1

*p < 1 · 10⁻⁹

**p < 1 · 10⁻¹⁴

Table 6.1: Results of the four classifiers in all rounds

6.2 The artificial diagnosis system in practice

The AI techniques for a diagnosis system for dementia used in this research were designed to use available data from two different institutions: care institution *De Zorggroep* and academic hospital *UMC St. Radboud*. In this way, the system would fit in the human diagnosis process pipeline and psychologists would not need to examine more tests than usual.

The attributes used in this research come from standard data in the human psychological diagnosis process. Therefore using a system based on these attributes would not require more effort than filling in a form (like a screening report). When the processing of the test data is automated, a diagnosis system could fit even better¹⁶. Only the observed data would need to be added, just by clicking some checkboxes.

In this research I skipped the classification of the four most common dementia forms. The ratio of dementia form versus non-dementia was too low and *De Zorggroep* dataset could be incomplete. But more important, in practice is the diagnosis in dementia form preserved for further examination. When a further diagnosis is necessary, another specialist system could examine the client in case of dementia.

It is possible to classify Alzheimer's Disease on solely the *UMC St. Radboud* dataset. Alzheimers Disease is present in 36.87% of the subjects. The other forms have balances lower than 0.10: Vascular Dementia (0.0968), Lewy Body Dementia (0.0015) and Frontotemporal Dementia (0.0061). I recommend you should acquire a new dataset, when looking at a new diagnosis system specialized in dementia forms.

Especially the screening reports of *De Zorggroep* contain valuable information for a diagnosis system for dementia, because they consist of all the relevant information on which the psychologist makes a diagnosis. The *UMC St. Radboud* contains a lot of other information used for academic research. The disadvantage of *De Zorggroep* dataset is that it takes a lot of work to extract the attributes, and the dataset contains only 25 subjects, which is actually too small for testing representatiely.

Therefor I recommend to use a larger dataset based on the screening reports for further research into a diagnosis system for dementia.

When using a diagnosis system in practice, 80% is not enough. One incorrect

¹⁶For the 'processing of the test data' and 'artificial diagnosis system' process steps look at Figure 1.2

diagnosis on every five cases would increase the trust in the diagnosis system. But I believe that with further research in possible AI techniques and with a bigger dataset of screening reports a diagnosis system could be trained that would assist and improve the diagnosis of dementia.

6.2.1 Final considerations

I am very grateful for the two institutes that have provided the data for my research. Since obtaining real patient data requires the full cooperation of the medical responsible personnel and careful attention for ethical procedures, the data acquisition and pre-processing process should not be underestimated. However, both the *UMC St. Radboud* and *De Zorggroep* were very interested in the outcomes of this research and therefore were able to provide me with this relevant data set.

Acknowledgments

First of all, I would like to give my special thanks to the expert, Josma Kunneman, who showed me the real world. A world where AI could contribute to. She, as an expert, gave me the possibility to get some insight in the psychological and practical side of my field. I am grateful for the work and effort she made to provide me relevant literature, contacts, screening reports, interviews and other information. Furthermore she inspired me to research this thesis topic.

Secondly, I want to say thanks to my supervisor, Louis Vuurpijl, for guiding me through the start of my thesis and helping me in doing the research needed for me to make this thesis. With that he showed me the importance and interesting side of research in my field.

I also want to thank Martijn Lappenschaar for providing the *UMC St. Radboud* dataset, as well as Marcel van Gerven for commenting on the draft version and being the second reader of this thesis.

As last I am grateful for everyone who supported me.

References

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)* (4th ed.). Washington, DC: American Psychiatric Publishing. Paperback.
- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., et al. (2012, may). Weka manual for version 3-6-7 [Computer software manual].
- Burges, C. J. (1996). Simplified support vector decision rules. In *13th international conference on machine learning* (p. 71-77). Morgan Kaufmann.
- Cessie, S. le, & Houwelingen, J. van. (1992). Ridge estimators in logistic regression. *Applied Statistics*, 41(1), 191-201.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cummings, J. L. (1993, may). Mini-mental state examination: Norms, normals, and numbers. *Journal of the American Medical Association*, 269(8), 2420-2421.
- Deelman, B. G. (1972). *Etudes in de neuropsychologie: psychologische onderzoeken bij patienten met gelocaliseerde cerebrale stoornissen*. Unpublished doctoral dissertation, Universiteit Groningen.
- Dianostiek van dementie*. (2008). Handreiking. NVVA.
- Fenwick, T. L., & Holmes, C. B. (1993). Effect of visual cues with the wais-r digit span subtest. *Perceptual and motor skills*, 76(3), 1025-1026.
- Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., et al. (2006). Global prevalence of dementia: a delphi consensus study. *The Lancet*, 366(9503), 2112-2117.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198.
- Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., et al. (2005). Weka. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (chap. WEKA, A Machine Learning Workbench for Data Mining). Springer US.
- Galvin, J. E., & Sadowsky, C. H. (2012). Practical guidelines for the recognition

- and diagnosis of dementia. *The Journal of the American Board of Family Medicine*, 25(3), 367-382.
- Graf, C. (2008, Apr). How to try this : The lawton instrumental activities of daily living scale - the lawton scale can help nurses to identify early functional decline and determine home care services. *American journal of nursing: professional journal of the American Nurses' Association*, 108(4), 52-63.
- Jaardocument 2010 stichting zorggroep noord- en midden-limburg. (2010).
- Jaardocument 2011 umc st. radboud. (2011).
- Jonker, C., Slaets, J., & Verhey, F. (2009). *Handboek dementie - laatste inzichten in diagnostiek en behandeling*. Houten: Bohn Stafleu van Loghum.
- Jonker, C., Verhey, F. R. J., & Slaets, J. (2001). *Alzheimer en andere vormen van dementie*. Houten: Bohn Stafleu Van Loghum.
- Kleinbaum, D. G., Klein, M., & Pryor, E. R. (2002). *Logistic regression: a self-learning text* (2nd edition ed.). New York: Springer.
- Koning, I. de, Dippel, D. W., Kooten, F. van, & Koudstaal, P. J. (2000). A short screening instrument for poststroke dementia : The r-camcog. *Stroke*(31), 1502-1508.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1), 89-109.
- Kononenko, I., Bratko, I., & Kukar, M. (1998). Application of machine learning to medical diagnosis. In R. Michalski, I. Bratko, & M. Kubar (Eds.), *Machine learning, data mining and knowledge discovery: methods and applications*. New York: Wiley.
- Ledley, R. S., & Lusted, L. B. (1959, July). Reasoning foundations of medical diagnosis. *Science*, 130(3366).
- Lindeboom, J., Horst, R. ter, Hooyer, C., Dinkgreve, M., & Jonker, C. (1993). Some psychometric properties of the camcog. *Psychological Medicine*, 23, 213-219.
- Magoulas, G. D., & Prentza, A. (2001). Machine learning in medical applications. *Machine Learning and Its Applications: advanced lectures*, 300-307.
- Mangiameli, P., West, D., & Rampal, R. (2004). Model selection for medical diagnosis decision support systems. *Decision Support Systems*, 36(3), 247 - 259.
- Medeiros, K. de, Robert, P., Gauthier, S., Stella, F., Politis, A., Leoutsakos, J., et al. (2010). The neuropsychiatric inventory-clinician rating scale (npi-c): reliability and validity of a revised assessment of neuropsychiatric symptoms in dementia. *International Psychogeriatrics*, 22(Special Issue 6), 984-994.
- Miller, R. A. (1994). Medical diagnostic decision support systems - past present and future. *Journal of the American Medical Informatics Association*, 1(1), 8-27.

- Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, & A. Smola (Eds.), *Advances in kernel methods - support vector learning*. MIT Press.
- Ranjith, P., & Khandelwal, M. (2011). Artificial neural network for prediction of air flow in a single rock joint. *Neural Computing and Applications, Online First*, 1-10.
- Rijsbergen, M. W. van, Oldenbeuving, A. W., Nieuwenhuis-Mark, R. E., Nys, G. M., Las, S. G., Roks, G., et al. (2011, July). Delirium in acute stroke: A predictor of subsequent cognitive impairment?: A two-year follow-up study. *Journal of the neurological sciences*, 306(1), 138-142.
- Roth, M., Huppert, F. A., Mountjoy, C., & Tym, E. (2005). *Camdex-r/n* (3rd ed.). Harcourt Test Publishers.
- Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd edition ed.). Prentice-Hall, Englewood Cliffs, NJ.
- Saito, K., & Nakano, R. (1988, jul). Medical diagnostic expert system based on pdp model. *IEEE International Conference on Neural Networks, Vol 1, 1*, 255-262.
- Sánchez-Cubillo, I., Periañez, J., Adrover-Roig, D., Rodríguez-Sánchez, J., Ríos-Lago, M., Tirapu, J., et al. (2009). Construct validity of the trail making test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438-450.
- Schulz, R., & Martire, L. M. (2004). Family caregiving of persons with dementia: Prevalence, health effects, and support strategies. *American Journal of Geriatric Psych*, 12(3).
- Sherman, A. R., & Blatt, S. J. (1968, Aug). Wais digit span, digit symbol, and vocabulary performance as a function of prior experiences of success and failure. *Journal of consulting and clinical psychology*, 32(4), 407-412.
- Vuurpijl, L., & Schomaker, L. (2000). Two-stage character classification: A combined approach of clustering and support vector classifiers. In L. Schomaker & L. Vuurpijl (Eds.), *Proceedings of the seventh international workshop on frontiers in handwriting recognition* (p. 423-432). International Unipen Foundation.
- Vuurpijl, L., Schomaker, L., & Erp, M. v. (2003). Architectures for detecting and solving conflicts: two-stage classification and support vector classifiers. *International Journal on Document Analysis and Recognition*, 5(4), 213-223.
- Wu, Y. (2012, Spring). *Nearest neighbor classification: Part i*. (Note)

Appendix A

In this appendix, the sections of CAMCOG are displayed with their maximum score.

Section	Max score
CAMCOG Total score	107
MMSE	30
Executive functioning	28

	Memory section	37
	Orientation	10
	Time	5
	Place	5
	Memory	27
	Remote	6
	Recent	4
	Learning	17
	Non-memory section	70
	Language	30
	Comprehension	9
	Expression	21
	Attention/Concentration	4
	Praxis	12
	Construction	6
	Ideom/ideat	6
	Calculation	5
	Abstract thinking	8
	Perception	11

Appendix B

This is complete list of all the original attributes of the *UMC St. Radboud* dataset (original file: *Geheugenpoli.xls*):

ID1	opname datum
Catogrie	geboortedatum
geslacht1	lengte
gewicht	woonsituatie
SamenstellingHuishouden	SamenstellingHuishoudenAnders
GeboorteLandPatient	burgerlijke staat
Woning	opleiding
opleidingAnders	aantal jaar onderwijs
HoogstebetrekkingPatient	Hoogstebetrekkingpartner
Thuiszorg	Dagopvang
PGDagopvang	SomDagbehandeling
Voorgescheidenis	Anamnese
LichtOZ	Medicatie
DementieDiagnosePatientMDO1	DementieDiagnosePatientMDO2
AD	VaD
DLB	FTD
Dementie	DementieDiagnoseMDO1Akkoord
DementieDiagnoseMDO2Akkoord	DefintieveDiagnose
DD2	DD3
DD4	DD1DementieAnders
DD1PsychiatrischeAndere	DD1NeurologischeAndere
DD2DementieAnders	DD2PsychiatrischeAndere
DD2NeurologischeAndere	DD3DementieAnders
DD3PsychiatrischeAndere	DD3NeurologischeAndere
CDRScoreMDO1	CDRScoreMDO2
SenielPreseniel	SenielPreseniel4
SenielPreseniel8	AnderDementieDiagnose
AnderePsychaitrischeDiagnose	AndereNeurologischDiagnose
AnderDementieDiagnoseMDO2	AnderePsychaitrischeDiagnoseMDO2
AndereNeurologischDiagnoseMDO2	WanneerZijnDeGeheugenProblemenBegonnen
Hypertensie	HypertensieSindsDag
HypertensieSindsMaand	HypertensieSindsJaar
Diabets	DiabetesSindsDag
:	:

:
 DiabetesSindsMaand
 CardiovasculaireZiekten
 RokenSinds
 RokenSindsMaand
 RokenStop
 RokenStopMaand
 RokenHoeveel
 RokenHoeveelperTijd
 HuidigeAlcoholinname
 AlcoholHoeveelheidPerTijd
 DepressieDag
 DepressieJaar
 Overgewicht
 HypercholHypertriDag
 HypercholHypertriJaar
 Vascular
 Respiratoir
 BovensteTractusDigestivus
 Lever
 UrogenitaleSysteem
 Neurologisch
 PsychiatrischeZiekten
 aantal catogorie
 aantal cat in 4
 Hart 4
 MMSEPatient
 GDS15Patient
 CAMCOGPatientTotaal
 CAMCOGPatientSubOrientatie
 CAMCOGPatientSubTaal
 CAMCOGPatientSubGeheugen
 CAMCOGPatientSubAandachtRekenen
 CAMCOGPatientSubPraxis
 CAMCOGPatientSubAbstractRedenen
 CAMCOGPatientSubVisuoperceptie
 CAMCOGSubschaalExScore
 15WoordVersie
 15WoordReeks2
 15WoordReeks4
 15WoordReeksTScore
 15WoordRecallTscorePatient
 15WoordRecognitionPatientAfggenomen
 TrailMakingAPatient
 TrailMakingBPatient
 CijferreeksenAPatient
 SchaalscoreCijferreeksen
 :

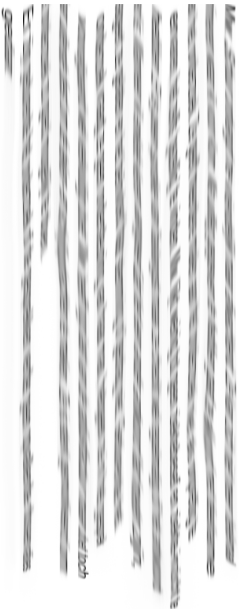
:
 DiabetesSindsJaar
 Roken
 RokenSindsDag
 RokenSindsJaar
 RokenStopDag
 RokenStopJaar
 RokenEenheid
 Alcohol
 AlcoholHoeveelheid
 Depressie
 DepressieMaand
 FamiliareBelasting
 HypercholHypertri
 HypercholHypertriMaand
 Hart
 Haematopoetisch
 OgenOrenNeusPharynxEnLarynx
 OndersteTractusDigestivus
 Nieren
 NeuromusculiareSysteem
 EndocrineEnMetaboleSystemMammae
 totaal score
 aantal cat in 3
 Hart 3
 poli
 MMSEPatientScore
 GDS15PatientScore
 CAMCOGPatientMaxScore
 CAMCOGPatientSubOrientatieMax
 CAMCOGPatientSubTaalMax
 CAMCOGPatientSubGeheugenMax
 CAMCOGPatientSubAandachtRekenenMax
 CAMCOGPatientSubPraxisMax
 CAMCOGPatientSubAbstractRedenenMax
 CAMCOGPatientSubVisuoperceptieMax
 CAMCOGSubschaalExMaxScore
 15WoordReeks1
 15WoordReeks3
 15WoordReeks5
 15WoordRecallPatient
 15WoordRecognitionPatient
 15woordPatientAfggenomenVersie
 TscoreTrailMakingAPatient
 TscoreTrailMakingBpatient
 CijferreeksenBPatient
 TinettiBalansPatient
 :

:
 TinettiBalansPatientMaxScore
 TinettiLopenPatientMaxScore
 BarthelPatiënt
 LawtonPatient
 BMIPatient
 RRDiaistolischPatient
 NaasteMantelzorger
 AnderemantelzorgerDANAasteWie
 GeslachteNaaste
 RelatiePatientNaasteMannelijk
 ZBINaaste
 NPI
 NPIEmo
 AanvullendeDiagnostiek
 KlachtenWanneerMaand
 PatientRoute
 MTARechts

:
 TinettiLopenPatient
 TimedUpAndGoPatient
 BarthelPatiëntMaxScore
 LawtonPatientMaxScore
 RRSystolischPatient
 NaasteInBeeld
 AndereMantelzorgerDanNaaste
 GeboortedatumNaaste
 RelatiePatientNaasteVrouwelijk
 RelatiePatientNaasteAnders
 ZBINaasteMaxScore
 NPIMaxScore
 NPIEmoMax
 KlachtenWanneerDag
 KlachtenWanneerJaar
 MTALinks
 TscoreTrailMakBtovApat

An example of a screening rapport of *De Zorggroep*. First, the rapport was made anonymous by the psychologist. Personal information about the psychologist is removed due to privacy reasons.

57



Diagnostisch Instrumentarium:

CAMDEX-R/N Deel B.

Testresultaten

Oriëntatie:

Tijd

Mw. weet de dag van de week en de datum te noemen. Ook het seizoen wordt correct benoemd. Het jaartal kan mw. niet vertellen, ze geeft als antwoord 2005.

Plaats

De provincie en stad worden correct benoemd. Mw. weet met enige moeite twee straten in de buurt te noemen en weet op welke verdieping van het gebouw zij zich bevindt.

Tijdsbesef

Het tijdsbesef is vrij goed. Mw. kan ongeveer benoemen hoe laat het is zonder op de klok te kijken en hoeveel tijd er verstreken is sinds de start van het onderzoek. Gedurende de dag kan zij door middel van een duidelijke dagsstructuur plaatsen hoe laat het ongeveer is.

Geheugen:

Hermenen (actief ophalen uit geheugen)

Het herkennen van visuele informatie gaat matig. Het herkennen van auditieve informatie op de korte termijn gaat goed. Het herkennen van een adres dat is voorgelezen (i.v.m. pijn aan de arm) gaat goed.

Herkennen

Het herkennen van visuele informatie gaat slecht.

Hermening van informatie uit het verleden

Het herkennen van informatie uit het verleden gaat matig.

Hermenen van recente informatie

Het herkennen van recente informatie gaat slecht.

Registratie

Het voor korte tijd opslaan van verbale informatie in het geheugen gaat goed.

Taal:

Begrip-motorische respons

Uit de motorische respons blijkt dat mw. de verbale opdracht goed begrijpt.

Begrip-verbale respons
Uit de verbale respons blijkt dat mw. de vraag goed begrijpt.
Expressie-benoemen
Het benoemen van voorwerpen gaat goed. Het benoemen van plaatjes gaat goed.
Expressie-woordloftheid
Het noemen van woorden in een bepaalde categorie uit het geheugen gaat matig.
Expressie-definities
Het geven van een definitie van een bepaald woord gaat matig.
Expressie-nazeggen/herhalen
Het nazeggen van een zin gaat goed.
Begrijpend lezen
Het leesbegrip is goed.

Aandacht/concentratie:

Mw. kan zich goed concentreren en haar aandacht is goed vast te houden.

Praxis:

Visuconstructie

Het naakten van figuren gaat goed. Het zelf tekenen van een klok gaat goed.

Ideationeel

Het vermogen om voorwerpen op de juiste wijze te hanteren door de juiste volgorde van handelen is goed.

Ideomotorisch

Het vermogen om bewegingen uit te beelden is goed.

Schoolsevaardigheden:

Schrijven/schrijven

Het schrijven van een zelf bedachte zin gaat goed ('Gezellig om hier met iemand te zitten'). Mw. legt uit dat ze 'ernand' opschrijft omdat ze zo snel niet weet hoe je psycholoog schrijft.

Rekenen

Eenvoudig rekenen gaat goed.

Executive functies:

Abstract denken/rekenen

Het abstract berekenen van verbale informatie gaat matig. Het abstract berekenen van visuele informatie gaat slecht.

Ideationele fluency

Het genereren van ideeën gaat matig.

Woordloftheid

Het noemen van woorden in een bepaalde categorie uit het geheugen gaat matig.

Perceptie:

Visueel

De visuele perceptie is goed.

Object-constantheid

Het herkennen van een object vanuit een ongebruikelijke hoek gaat slecht.

Herkennen

Het herkennen van een bekend persoon gaat goed.

Ziektebesef/ziekte-inzicht
Mw. heeft een gering ziektebesef. Het ziekte-inzicht is nagenoeg afwezig.

Overzicht testresultaten

CAMCOG		Maximum-score	Behaalde score	Mediaan/grens	Ondergrenscore
Onderdeel	Oriëntatie				
	Tijd (139 t/m 143)	5	4	5/4	
	Plaats (144 t/m 148)	5	5	5/3	
Geheugen	Totaal	10	9	10/8	
	Verleden (168 t/m 171)	6	3	5/3	
	Recent (172 t/m 175)	4	3	4/3	
	Leren (164,165,166)	17	10	13/10	
	Totaal	27	16	23/20	*
Geheugensectie		37	25	perc.	
Taal	Begrip (149 t/m 155, 181, 182)	9	8	9/8	*
	Expressie (156 t/m 163)	21	19		
	Totaal	30	27	27/25	
Aandacht Praxis	(178,179)	7	6	6/3	
	Constructie (183 t/m 186)	6	6	6/4	
	Idoom/ideat (189,191 t/m 193)	6	5	6/5	
Rekenen	Totaal	12	11		
	(194, 195)	2	2		
	Abstract redeneren (197 t/m 200)	8	6	7/4	
Perceptie	(201, 202)	8	3	5/3	
	Niet-geheugensectie	67	55	perc.	
	Executief functioneren (188,197,198, 199,200,200a,200b)	28	16		
CAMCOG		104	80	91/78	
MMSE		30	26		
normgroep: vrouw leeftijd: 75-79		opl.niv. 2			

Samenvatting en conclusie:

Uit het huidige onderzoek kan geconcludeerd worden dat er sprake is van een beginnend dementieel beeld.



Met vriendelijke groet,

Appendix D

The four classifiers are identified a index. In this appendix, the parameters of these classifiers are described.

MLP ₁	functions.MultilayerPerceptron -L 0.0050 -M 0.2 -N 750 -V 0 -S 0 -E 20 -H 2
MLP ₂	functions.MultilayerPerceptron -L 0.0050 -M 0.2 -N 900 -V 0 -S 0 -E 20 -H a,i
MLP ₃	functions.MultilayerPerceptron -L 0.05 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a,i
MLP ₄	functions.MultilayerPerceptron -L 0.05 -M 0.2 -N 100 -V 0 -S 0 -E 20 -H a,i
MLP ₅	functions.MultilayerPerceptron -L 0.01 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H i,a
SMO ₁	functions.SMO -C 1.0 -L 0.0010 -P 1.0E-14 -N 0 -V -1 1 -K functions.supportVector.PolyKernel -C 250007 -E 1.0
SMO ₂	functions.SMO -C 1.0 -L 0.0010 -P 1.0E-16 -N 0 -V -1 1 -K functions.supportVector.PolyKernel -C 250007 -E 1.1
SMO ₃	functions.SMO -C 1.0 -L 0.0010 -P 1.0E-12 -N 2 -V -1 1 -K functions.supportVector.PolyKernel -C 250007 -E 1.2
SMO ₄	functions.SMO -C 1.0 -L 0.0010 -P 1.0E-20 -N 0 -V -1 1 -K functions.supportVector.PolyKernel -C 250007 -E 1.3
Logistic ₁	functions.Logistic -R 1.0E-8 -M -1
Logistic ₂	functions.Logistic -R 1.0E-10 -M -2
Logistic ₃	functions.Logistic -R 1.0E-6 -M -3
Nnge ₁	rules.NNge -G 20 -I 10
Nnge ₂	rules.NNge -G 24 -I 6
Nnge ₃	rules.NNge -G 2 -I 1
Nnge ₄	rules.NNge -G 5 -I 1
Nnge ₅	rules.NNge -G 10 -I 3

Appendix E

This appendix contains the results of Round 1, setting 1. The full dataset with 38 attributes is classified with ten-folds cross-validation.

MLP: Multilayer Perceptron

Summary

Correctly Classified Instances	533	79.55 %
Incorrectly Classified Instances	137	20.45 %
Kappa statistic	0.58	
Mean absolute error	0.26	
Root mean squared error	0.38	
Relative absolute error	53.41 %	
Root relative squared error	77.27 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
221	68	a = TRUE
69	312	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.765	0.181	0.762	0.765	0.763	0.872	TRUE
	0.819	0.235	0.821	0.819	0.82	0.872	FALSE
Weighted Avg.	0.796	0.212	0.796	0.796	0.796	0.872	

SMO: Suport Vector Machine

Summary

Correctly Classified Instances	541	80.75 %
Incorrectly Classified Instances	129	19.25 %
Kappa statistic	0.61	
Mean absolute error	0.19	
Root mean squared error	0.44	
Relative absolute error	39.24 %	
Root relative squared error	88.60 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
218	71	a = TRUE
58	323	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.754	0.152	0.79	0.754	0.772	0.801	TRUE
	0.848	0.246	0.82	0.848	0.834	0.801	FALSE
Weighted Avg.	0.807	0.205	0.807	0.807	0.807	0.801	

Logistic: Logistic Regression**Summary**

Correctly Classified Instances	534	79.70 %
Incorrectly Classified Instances	136	20.30 %
Kappa statistic	0.59	
Mean absolute error	0.25	
Root mean squared error	0.37	
Relative absolute error	51.88 %	
Root relative squared error	74.22 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
219	70	a = TRUE
66	315	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.758	0.173	0.768	0.758	0.763	0.884	TRUE
	0.827	0.242	0.818	0.827	0.822	0.884	FALSE
Weighted Avg.	0.797	0.212	0.797	0.797	0.797	0.884	

k-NN: Nearest Neighbor**Summary**

Correctly Classified Instances	525	78.36 %
Incorrectly Classified Instances	145	21.64 %
Kappa statistic	0.56	
Mean absolute error	0.22	
Root mean squared error	0.47	
Relative absolute error	44.11 %	
Root relative squared error	9.93 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
208	81	a = TRUE
64	317	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.72	0.168	0.765	0.72	0.742	0.776	TRUE
	0.832	0.28	0.796	0.832	0.814	0.776	FALSE
Weighted Avg.	0.784	0.232	0.783	0.784	0.783	0.776	

Appendix F

This appendix contains the results of Round 2B. The full dataset with the reduced 20 attributes is classified with ten-folds cross-validation.

MLP: Multilayer Perceptron

Summary

Correctly Classified Instances	533	79.55 %
Incorrectly Classified Instances	137	20.45 %
Kappa statistic	0.58	
Mean absolute error	0.27	
Root mean squared error	0.38	
Relative absolute error	55.76 %	
Root relative squared error	77.01 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
220	69	a = TRUE
68	313	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.761	0.178	0.764	0.761	0.763	0.87	TRUE
	0.822	0.239	0.819	0.822	0.82	0.87	FALSE
Weighted Avg.	0.796	0.213	0.795	0.796	0.795	0.87	

SMO: Suport Vector Machine

Summary

Correctly Classified Instances	540	80.60 %
Incorrectly Classified Instances	130	19.40 %
Kappa statistic	0.60	
Mean absolute error	0.19	
Root mean squared error	0.44	
Relative absolute error	39.54 %	
Root relative squared error	88.94 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
220	69	a = TRUE
61	320	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.761	0.16	0.783	0.761	0.772	0.801	TRUE
	0.84	0.239	0.823	0.84	0.831	0.801	FALSE
Weighted Avg.	0.806	0.205	0.805	0.806	0.806	0.801	

Logistic: Logistic Regression**Summary**

Correctly Classified Instances	542	80.90 %
Incorrectly Classified Instances	128	19.10 %
Kappa statistic	0.61	
Mean absolute error	0.26	
Root mean squared error	0.37	
Relative absolute error	53.81 %	
Root relative squared error	75.00 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
220	69	a = TRUE
59	322	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.761	0.155	0.789	0.761	0.775	0.881	TRUE
	0.845	0.239	0.824	0.845	0.834	0.881	FALSE
Weighted Avg.	0.809	0.203	0.808	0.809	0.809	0.881	

k-NN: Nearest Neighbor**Summary**

Correctly Classified Instances	530	79.10 %
Incorrectly Classified Instances	140	20.90 %
Kappa statistic	0.57	
Mean absolute error	0.21	
Root mean squared error	0.46	
Relative absolute error	42.59 %	
Root relative squared error	92.30 %	
Total Number of Instances	670	

Confusion Matrix

a	b	← classified as
211	78	a = TRUE
62	319	b = FALSE

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.73	0.163	0.773	0.73	0.751	0.784	TRUE
	0.837	0.27	0.804	0.837	0.82	0.784	FALSE
Weighted Avg.	0.791	0.224	0.79	0.791	0.79	0.784	