Clinical Decision Support System for Early Detection and Diagnosis of Dementia

Mohamed Bennasar

Cardiff University

School of Engineering

This thesis is submitted in fulfilment of the requirement of the degree of

Doctor of Philosophy

September 2014

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed............................... (candidate) Date ....

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed .... (candidate) Date ....

STATEMENT 2

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit reference.

Signed .... (candidate) Date ....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed .... (candidate) Date ....

STATEMENT 4

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, after expiry of a bar on access approved by the Graduate Development Committee.

Signed................................ (candidate) Date ....

ABSTRACT

Dementia is a syndrome caused by a chronic or progressive disease of the brain, which affects memory, orientation, thinking, calculation, learning ability and language. Until recently, early diagnosis of dementia was not a high priority, since the related diseases were considered untreatable and irreversible. However, more effective treatments are becoming available, which can slow the progress of dementia if they are used in the early stages of the disease. Therefore, early diagnosis is becoming more important. The Clock Drawing Test (CDT) and Mini Mental State Examination (MMSE) are well-known cognitive assessment tests. A known obstacle to the wider usage of the CDT assessments is the scoring and interpretation of the results.

This thesis introduces a novel diagnostic Clinical Decision Support System (CDSS) based on CDT which can help in the diagnosis of three stages of dementia. It also introduces the advanced methods developed for the interpretation and analysis of CDTs. The data used in this research consist of 604 clock drawings produced by dementia patients and healthy individuals. A comprehensive catalogue of 47 visual features within CDT drawings is proposed to enhance the sensitivity of the CDT in diagnosing the early stages of dementia. These features are selected following a comprehensive analysis of the available data and the most common CDT scoring systems reported in the medical literature. These features are used to build a new digitised dataset necessary for training and validating the proposed CDSS.

In this thesis, a novel feature selection method is proposed for the study of CDT feature significance and to define the most important features in diagnosing dementia.

A new framework is also introduced to analyse the temporal changes in the CDT features corresponding to the progress of dementia over time, and to define the first onset symptoms.

The proposed CDSS is designed to differentiate between four cognitive function statuses: (i) normal; (ii) mild cognitive impairment or mild dementia; (iii) moderate or severe dementia; and (vi) functional. This represents a new application of the CDT, as it was previously used only to detect the positive dementia cases.

Diagnosing mild cognitive impairment or early stage dementia using CDT as a standalone tool is a very challenging task. To address this, a novel cascade classifier is proposed, which benefits from combining CDT and MMSE to enhance the overall performance of the system.

The proposed CDSS diagnoses the CDT drawings and places them into one of three cognitive statuses (normal or functional, mild cognitive impairment or mild dementia, and moderate or severe dementia) with an accuracy of 78.34 %. Moreover, the proposed CDSS can distinguish between the normal and the abnormal cases with accuracy of 89.54 %.

The achieved results are good and outperform most of CDT scoring systems in discriminating between normal and abnormal cases as reported in existing literature. Moreover, the system shows a good performance in diagnosing the CDT drawings into one of the three cognitive statuses, even comparing well with the performance of dementia specialists.

The research has been granted ethical approval from the South East Wales Research Ethics Committee to employ anonymised copies of clock drawings and copies of Mini Mental State Examination made by patients during their examination by the memory team in Llandough hospital, Cardiff.

ACKNOWLEDGEMENT

First, I am grateful to Allah (My Lord) the all high, the all great who made it possible for me to finish this work.

Second, I would like to thank my supervisors, Professor Rossitza Setchi, Doctor Yulia Hicks, and Professor Antony Bayer not only for their constant assistance, excellent guidance, and thoughtful advices, but also for their constant support and encouragement. The high standard of their research has always been an inspiration to me.

I would like also to thank my colleagues in Knowledge Engineering Systems (KES) group, Dr. Todorov, Dr. Fadzli, Dr. Shi, Dr. Stankov, Mrs. Romli, Mrs. Anuar, Mr. Saadi, and Mrs. Harbi for stimulating discussion. I would like also to thank Miss Cotterill for her help with collecting data.

Finally, I would like to express my sincere gratitude and appreciation to my late father, and my beloved family: my mother, wife, brother, sister, and grandmother, for their endless love, support and prayers. Thanks also to my beloved kids Malek, Juri, and Jenna.

Mohamed Bennasar

September 2014

Table of Contents

List of figures

List of tables

LIST OF PUBLICATIONS

Journal articles:

Bennasar, M., Setchi, R., and Hicks, Y. (2013). Feature interaction maximisation. *Pattern Recognition Letters, 15*, pp.1630-1635.

Bennasar, M., Setchi, R., and Hicks, Y. (2014). Feature Selection Using Joint Mutual Information Maximisation. *Submitted to Expert Systems with Applications*.

Conference proceedings:

Bennasar, M., Setchi, R., Bayer, A., and Hicks, Y. (2014). *Cascade Classification for Diagnosing Dementia.* IEEE SMC 2014 International Conference on Systems, Man and Cybernetics, San Diego, California, USA, pp. 2535-2540.

Bennasar, M., Setchi, R., Bayer, A., and Hicks, Y. (2013). *Feature Selection Based on Information Theory in the Clock Drawing Test.* 17th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Kitakyushu, Japan. pp. 902 – 911.

Bennasar, M., Setchi, R., and Hicks, Y. (2012). *Unsupervised Discretization Method based on Adjustable Intervals*. 16th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, San Sebastian, Spain. pp. 79-87.

Articles under preparation

Bennasar, M., Setchi, R., Bayer, A., and Hicks, Y. (2015). Clinical Decision Support System for Early Assessment and Detection and of Dementia (CDSS-DD). Will be submitted to *IEEE Transactions on Biomedical Engineering*.

Chapter: 1

Introduction

1.1 Motivation

The term ‘dementia’ is used to describe loss of cognitive ability, which is usually progressive and eventually severe and which affects memory, attention, problem-solving and communication skills. People of any age can be affected by dementia, but it is most common among elderly people. One in six people over 80, and one in fourteen people over 65, suffer from some type of dementia (Knapp and Prince, 2005).

Dementia is considered to be the most prominent disability among elderly people (Milne et al., 2008). Alzheimer’s disease (AD) and vascular dementia (VaD) are the first and second most common causes of dementia respectively. They are usually accompanied by memory loss and disturbances in executive cognitive functioning, leading to difficulty performing everyday activities (July et al., 2002). With the average age of the population steadily increasing, dementia has become an important issue; according to a 2012 report of the Alzheimer's Society the estimated number of the people suffering from dementia in the UK is about 800,000 (1.2 % of the entire UK population). One third of these people live alone in their own homes and need much support to carry out their everyday lives. The number of people with dementia is expected to increase to 1 million by 2021, which represents an increase of 25 % over the next 7 years. Dementia costs the NHS, local authorities and families 23 billion pounds per year, and this number is expected to grow to 27 billion pounds by 2018 (Lakey, et al., 2012).

Until recently, early diagnosis of dementia was not a high priority, since the related diseases were considered untreatable and irreversible. However, more effective treatments are becoming available, which can slow the progress of dementia if they are used in the early stages of the disease. For this reason, early diagnosis is becoming more important as it is the first step in understanding and managing the condition (Wild et al., 2008). Early diagnosis of dementia could therefore generate a significant positive effect on public health (Grober et al., 2008).

Regular assessment of dementia is one of the most effective approaches for the early detection of dementia. Mini Mental State Examinations (MMSE) and Clock Drawing Tests (CDT) are two of the most widely used instruments for assessing the degree of cognitive impairment, because they are simple to implement and have a reasonable sensitivity and specificity (Mittal et al., 2010).

The MMSE was introduced in 1975. It takes the form of a questionnaire which includes 30 items covering the five test areas of cognitive function: (i) registration, (ii) orientation, (iii) recall, (iv) language, (v) attention and calculation. The maximum score which can be attained is 30. A score of lower than 27 indicates cognitive impairment. The MMSE usually takes only 5-10 minutes to administer and is used repeatedly and routinely (Ismail et al., 2010).

CDTs are also used as an assessment tool for cognitive impairment and dementia. They are a measure of spatial dysfunction. Although the clock drawing test would appear at face value to be a relatively simple task, it is one which requires a wide range of perceptual and intellectual skills, making it a versatile screening instrument for the assessment of: comprehension; planning; visuospatial ability; visual memory; motor programming and execution; concentration; abstraction; and response inhibition (Ismail et al., 2010). During the test, the participant is asked to draw the face of a clock, mark in the hours and then draw the clock hands to indicate a specific time (for example, five minutes to three). The drawings are then assessed using various scoring systems with different degrees of complexity. Research shows a good correlation between the CDT and other more detailed (and more time-consuming) cognitive assessment methods. The CDT test has the advantage of being well accepted by patients (Shulman et al., 1986). The nature of the drawing task requires contributions from a diverse range of cerebral brain regions. Therefore, in the case of brain injury or disease, some of these regions are often compromised. The deficits in the produced clock drawings vary greatly according to the location of the pathology in the brain (Freedman et al., 1994)

However, even given the benefits of these routine measures of cognition, some methods such as MMSE can fail to detect executive dysfunction, which always occurs before memory decline (Mittal et al., 2010) and is hence a good indicator. To increase the sensitivity and specificity of the screening, CDT is often used in parallel with MMSE because of its ability to reveal the person's visual-spatial, constructional, and higher-order cognitive abilities, in addition to executive aspects, numerical knowledge, and the concept of time (Kim et al., 2008).

In this manner the two tests complement each other, with CDT helping to assess the visuoconstructional and executive function while MMSE assesses orientation, memory, and language (Hayley and John, 2002). Nonetheless, a known obstacle to the wider usage of the CDT screening method is the scoring and interpretation of the results. The CDT method is used in conjunction with numerous administration and scoring systems, each with different degrees of complexity. These can range from a simple binary rating to more complex qualitative and quantitative systems which capture the wide variety of errors in the drawn clock. Yet, none of the scoring systems have been accepted universally as the most effective system (Ismail et al., 2010).

It is now widely accepted that the early diagnosis of dementia can provide significant medical, social, and practical benefits. General practitioners (GPs) are in a pivotal position in dealing with people who have dementia, especially during the early stages (Alisoun et al., 2004). However, due to the absence of suitable validated instruments and the inadequacy of training in assessment practices, more than half of all patients with dementia are never diagnosed (Milne et al., 2008). In this context, complex screening tools are not practical and are not accepted by GPs.

It is generally agreed that if the CDT is to be used as a assessment tool then it should employ a brief, generalisable, highly predictive and 'quick and simple' scoring system (Shulman, 2000). To develop such a scoring system, the significance of all possible errors in the CDT drawings should be analysed in order to find the dominant errors which relate to dementia diagnosis.

Clinical Decision Support Systems (CDSS) are software algorithms designed to assist clinicians and other health specialists with decision making regarding diagnosis. It is defined by Van Bemmel and Musen (1997) as "*any piece of software that takes as input information about a clinical situation and produces as output inferences that can assist practitioners in their decision making and that would be judged as intelligent by the program users*".

The introduction of computer-based assessment tests could hence facilitate the early detection and diagnosis of dementia. It could also provide an advanced level of understanding of the test results, and can be administrated by health care associates other than neuropsychologists (while the tasks of interpretation and diagnosis can still be performed by specialists). Moreover, the computerised system can be administered via the internet in order to make the cognitive assessment available to a large number of people.

A computerised system could also facilitate study into the importance of the CDT errors, establishing which errors correspond to which disease (and stage thereof), while also enabling the validation of past research data, and highlighting the importance of the range of possible errors. Selecting the important CDT features might improve the sensitivity and specificity of the test. Moreover, the computer-based screening tool would enable easy comparison of historic results and identification of causes of concern.

1.2 **Aims and Objectives**

The scope of this project is the novel interpretation of CDT results. The project aims to assist clinicians at the point of care in the detection of early symptoms of dementia. This will be achieved by developing an intelligent clinical decision support system based on image processing and Artificial Intelligence (AI) algorithms. The clinical decision support system could provide more accurate and more consistent diagnoses. It could also provide deep insight into the CDT, and the relation between particular CDT errors and particular diagnoses.

The specific objectives necessary to achieve the aim are identified as

1. Creation of a conceptual model and architecture of a clinical decision support system, which can assist clinicians in the early diagnosis of dementia;

2. Building a comprehensive catalogue of CDT image features, both new geometric detailed and previously identified, with the propose of producing a new digitised dataset from hard copy drawing data;

3. Development and validation of a new subset feature selection methods for assessment of the importance of different features for medical diagnosis;

4. Development of a new framework for the analysis of temporal changes in the CDT features corresponding to the progress of dementia;

5. Design of a new cascade classifier for automatic dementia diagnosis on the basis of the features extracted from the CDT drawings;

6. Integration of the MMSE test with CDT in order to enhance the overall performance of the system.

1.3 Thesis Outline

This thesis is organised into the following structure:

- Chapter 1 has provided an introduction to the work.

- Chapter 2 presents a review of the background information surrounding dementia and the clock drawing test, and discusses related work in the field. It also reviews the relevant literature regarding: clinical decision support systems, machine learning, cascade classification, discretisation of continuous data, and dimensionality reduction.

- Chapter 3 proposes a conceptual model of the clinical decision support system for the early diagnosis of dementia; it also provides an overview of the process of collecting clock data.

- Chapter 4 introduces a comprehensive catalogue of CDT features with the purpose of producing a new digitised dataset from hard copy drawing data.

- Chapter 5 proposes novel subset feature selection methods for assessment of the importance of different features for medical diagnosis; it also presents the validation of them using benchmarking data.

- Chapter 6 employs one of the devolved feature selection methods to define the most significant features in the clock drawings. It also proposes a new framework for analysing the temporal changes in the CDT features corresponding to the progress of dementia.

- Chapter 7 proposes a diagnosis stage to classify the clock drawings into a number of categories. A new cascade classifier is proposed and evaluated.

- Chapter 8 highlights the contributions, limitation, and conclusions of this thesis, and proposes further work.

Chapter: 2

Literature Review

This chapter reviews previous studies relevant to the work presented in this thesis. The chapter provides a general background regarding dementia, as well as the common assessment tools and CDT scoring systems, and also investigates the relevant techniques that are capable of tackling the problem. This chapter is organised as follows: Section 2.1 introduces dementia and the diseases related to it; Section 2.2 discusses mild cognitive impairment and its sub-types; Section 2.3 reviews the CDT and its scoring systems. The MMSE test is discussed in Section 2.4; the focus then turns to reviewing the techniques relevant to each stage of the proposed conceptual model. The clinical decision support systems are discussed in section 2.5; Section 2.6 reviews the supervised machine learning techniques; Section 2.7 reviews the cascade classification; Section 2.8 reviews the image enhancement; Section 2.9 presents some common computerised cognitive assessment tools; Section 2.10 presents the current computer-based CDT systems; Section 2.11 gives an overview of the various discretisation techniques; Section 2.12 focuses on the dimensionality reduction, and introduces information theory as a foundation for further analysis; Finally Section 2.13 summaries the chapter and concludes the findings.

2.1 **Dementia**

Dementia is a syndrome caused by a chronic or progressive disease of the brain. These diseases cause disruption of cortical functions including memory, orientation, thinking, calculation, learning ability, and language. These impairments are usually combined with problems with emotional control and social behavior (Allbon et al., 2007).

The symptoms of dementia become increasingly severe over time, making it an irreversible and progressive condition. Dementia is not just one disease, and can therefore affect people in many different ways. This means the impact on the everyday life skills can be very different.

The number of people suffering from dementia in the UK is around 800,000, with over 17000 of them being below 66 years old. 84 % of UK citizens with dementia live in England, 8 % in Scotland, 5 % in wales, and 2 % in Northern Ireland (Lakey et al., 2012). The rate of increase in the number of people with dementia is not uniform across a range of ages. This is evident in Figure 2-1, which illustrates the increase in the number of people with dementia for each age group. Internationally 24.3 million people suffer from dementia, and one new case is diagnosed every 7 seconds. These numbers are expected to double every 20 years (Ferri et al., 2005). Dementia costs the UK economy 23 billion pound per year as unpaid careers, and Health care costs, and social care health service (Fernandez et al., 2010).

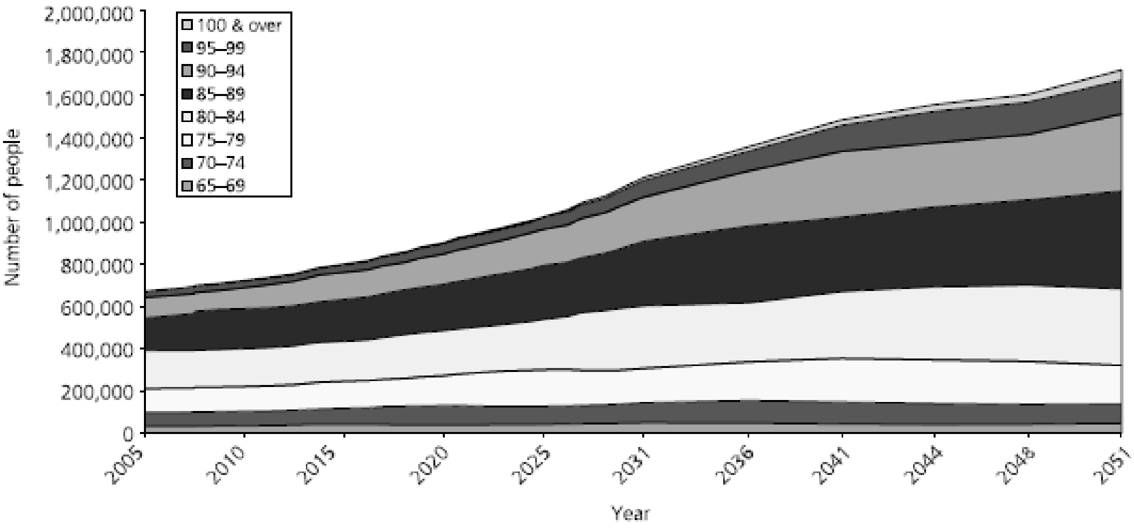


Figure 2-1: Projected number of people with late onset dementia by age group (UK) (adopted from Lakey, et al., 2012).

Dementia can be caused by a diverse range of diseases, but, Alzheimer's disease (AD) and vascular dementia (VaD) are the most common forms. 62 % of people with dementia suffer from AD, while 27 % suffer from VaD and mixed dementia (VaD and AD together) (Lakey, et al., 2012).

AD is the most common cause of dementia, with about 496,000 people in the UK suffering from this disease. It affects the brain cells, changes the chemistry and the brain structure, and eventually leads to death. It is a progressive disease, over time affecting more and more brain regions. The onset symptoms of this disease begin with lapses of memory and difficulty in finding suitable words. As the disease becomes more severe over the time, people with Alzheimer's disease may experience symptoms such as becoming confused, feeling sad or angry or scared, becoming more withdrawn, and having difficulty carrying out everyday activities. Gradually, the body becomes unable to perform all functions, ultimately leading to death (Waldemar et al., 2007).

VaD is the second most common cause of dementia. It is initiated due to improper blood supply to the brain. Strokes and small vessel disease are the most common causes of this form of dementia. This dementia doesn't progress gradually like Alzheimer's; the progression is usually more stepped. Vascular dementia shares some symptoms with Alzheimer's disease. In addition however, people with this form of dementia may experience: difficulty with speed thinking; problems with concentration; changes in behavior; depression and anxiety; memory problems; hallucinations; believing things which are not true; and becoming more obsessive.

There are other forms of dementia, but they are much less prevalent than AD and VaD. One of the other forms is Frontotemporal dementia (FTD), which affects the front of the brain. In the early stage of this form of dementia, the memory may stay intact while the other cognitive functions might be affected (McKhann et al., 2001). Dementia with Lewy Bodies (DLB) is another rare form of dementia, caused by tiny spherical protein deposits which develop within the nerve cells of the brain, leading to impaired memory, concentration and language skills (McKeith, 2004).

2.2 **Mild Cognitive Impairment (MCI)**

MCI is a condition associated with the exhibition of symptoms of cognitive decline which is slightly more rapid than would be expected at a given age. This decline usually appears as a minor memory complaint. Individuals with MCI are still able to carry out daily activities without any noticeable impairment. MCI is therefore generally considered as a risk state of dementia (Levely et al., 2006; Petersen et al., 2001; Gauthier et al., 2006), being between normal aging and AD. There is indeed an overlap in the boundaries between the normal ageing process and AD (Gauthier et al., 2006).

Several dementia severity scales have been proposed. The Clinical Dementia Rating (CDR) system (Hughes et al., 1982) classifies the stages of dementia into 5 categories: CDR 0 (Healthy), CDR 0.5 (Questionable dementia), CDR 1 (Mild dementia), CDR 2 (Moderate dementia), and CDR 3 (Severe dementia). In this scaling system there is no stage corresponding to MCI; this term was first used in the Global Deterioration Scale for ageing and dementia (GDS) (Roth et al., 1986). The latter scale divides the development of dementia into 7 levels: level 1 (No cognitive impairment), level 2 (Very mild cognitive decline), level 3 (Mild cognitive decline), level 4 (Moderate cognitive decline), level 5 (Moderately severe cognitive decline), level 6 (Severe cognitive decline), level 7 (Very severe cognitive decline). Some researchers have suggested that CDR 0.5 is a broad category which can include mild dementia and MCI (Gauthier et al., 2006).

There are several diverse subtypes of MCI. These are: (i) Amnestic MCI (a-MCI), which is characterised by memory problems; (ii) Multiple Domain MCI (md-MCI), where several cognitive domains are impaired such as executive function, visuospatial skills, and language. An individual with md-MCI may also suffer from memory impairment; (iii) Impairment in one single cognitive domain other than memory. This type is the least common.

MCI subtypes are caused by multiple diseases, and it is not necessary for all of them to develop to AD. Table 2.1 shows the etiological classification of each subtype. All of these MCI clinical subtypes do not affect the functional activities and do not show a significant change in the individual activity.

Cognitive assessment tools such as CDT and MMSE are important for the facilitation of early diagnosis of dementia. However, only 24 % of GPs assess the cognitive abilities of their patients regularly, the main reason for this being the lack of fast and simple assessment tools (Pinto and Peters, 2009).

Table 2-1: Classification of clinical subtypes of mild cognitive impairment with presumed etiology (adopted from Petersen, 2004).

Subtype	Etiology		
	Degenerative	Vascular	Psychiatric
Amnestic MCI (a-MCI)	AD		Depression
Multiple-domain MCI with amnesia (md-MCI +)	AD	VaD	Depression
Without amnesia (md-MCI -)	DLB	VaD	-
MCI affecting single non-memory domain	FTD, DLB	-	-

2.3 Clock Drawing Test (CDT)

CDT is originally designed to test for constructional apraxia, which is an inability or difficulty in drawing or assembling objects, or even a difficulty in understanding the task (Freedman et al., 1994). CDT has been employed as a tool for the assessment of cognitive abilities and neurological disorders for around 30 years (Lessig et al., 2008; Mendez et al., 1992; Shulman et al., 1993; Sunderland et al., 1989; Tuokko et al., 1992). It is also widely used as a component within neuropsychological assessment batteries such as: the 7 minute Neurocognitive Screening (Solomon et al., 1998); the Cambridge Cognitive Examination (CAMCOG) (Schmand et al., 1998); and the Mini-Cog screening (Brson et al., 2000). Researches show a high correlation between the results obtained via CDT and those acquired by other more detailed and time-consuming cognitive assessment tools. The CDT has the additional advantage of being well accepted by patients (Shulman et al., 1986).

During the test the individual is asked to draw a clock face, and he or she might also be asked to set the hands to a specific time. The task of drawing the clock seems to be a simple one. However, when analysed, it becomes clear that it is a complex task which requires contributions from various brain regions and also requires diverse neurological functions. Figure 2-2 shows an example of clock drawings drawn by patients with different types of dementia.

According to (Freedman et al., 1994), the process of drawing the clock starts when the command “draw a clock” is given. Auditory and language functions are needed to interpret the instruction, and this process is controlled in the brain by the temporal lobe (Gazzaniga et al., 2002). Then the long term visual memory system is activated to retrieve a visual layout of a typical clock, and the associated retrieval mechanism is also activated. This subtask is associated with the Limbic system and prefrontal cortex, sensory association cortex, and the temporal lobe (Purves et al., 2004; Hart, 2002; H. Markowitsch, 1995; Markowitsch, and Staniloiu, 2012). Figure 2-3 shows the structure of the human brain.

Visuomotor skills are required for the translation of the visual mental representation into a list of movements which draw the clock; this process is related with the motor cortex and sensory cortex (Hart, 2002). Executive functions are needed to assist with planning and organising the steps of the drawing process, and they also monitor and correct any errors. These functions are associated with frontal lobe (Purves et al., 2004). Finally if the test includes setting the clock hands at a specific time, working memory is involved in retaining the instruction until the individual needs that information. Most of the working memory is located in the frontal lobe (Hart, 2002). Therefore, if one or more of these cognitive functions is not working probably due to impairment in a related brain region, the resulting clock will reflect the deficits. In this way a clock is an appropriate subject matter, which requires more areas of the brain than simply drawing an object such as a house or a car.

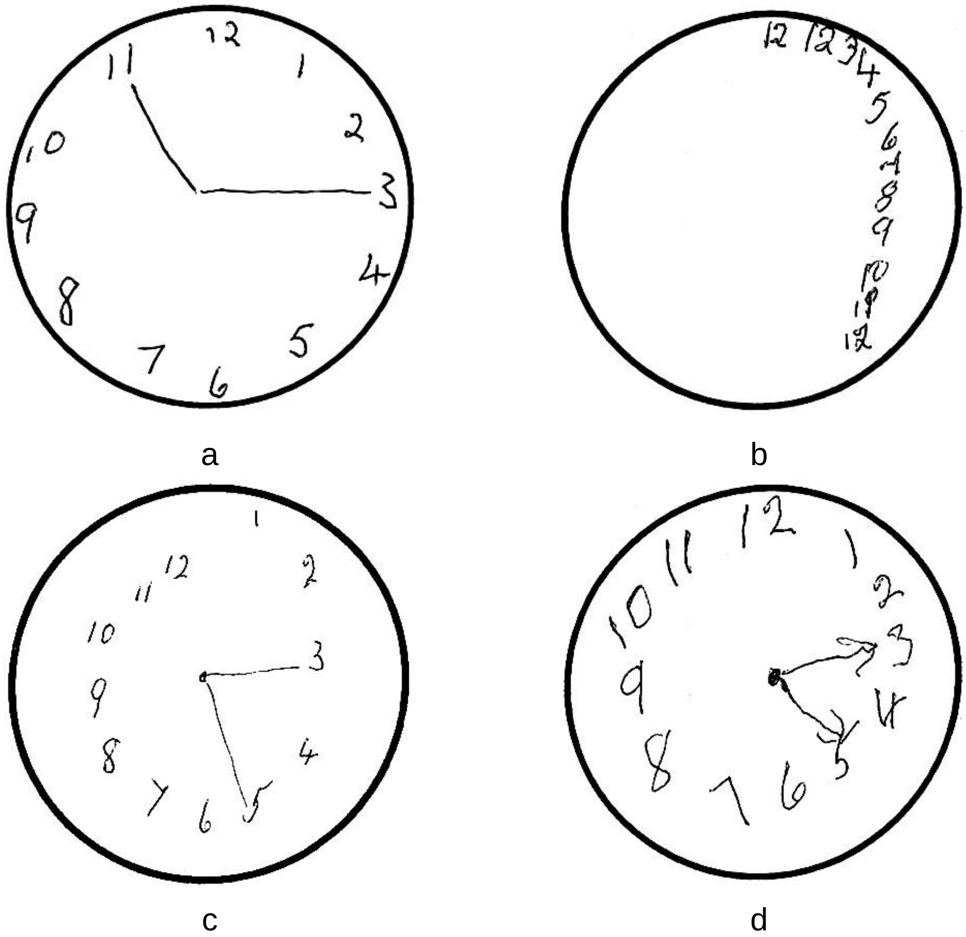


Figure 2-2: Examples of clock drawings drawn by patients at the Llandough hospital, Cardiff, UK: (a) Normal, (b) Alzheimer's disease, (c) Mild Dementia, (d) Vascular Dementia.

There are two ways in which individuals may be asked to perform the test. This could be either (i) a verbal command to draw the clock and set the hands to a specific time; or (ii) an instruction to copy a given template. There is a difference between these both forms in terms of the cognitive functions required to understand the instructions and perform the test. Command form places demand on language skills, long term memory, working memory, and executive functions. Therefore this form of the test is sensitive to any temporal or frontal lobe dysfunctions. On the other hand, the copying form of the test depends on perceptual functions, and therefore this form is sensitive to the impairments of the parietal lobe (Freedman et al., 1994).

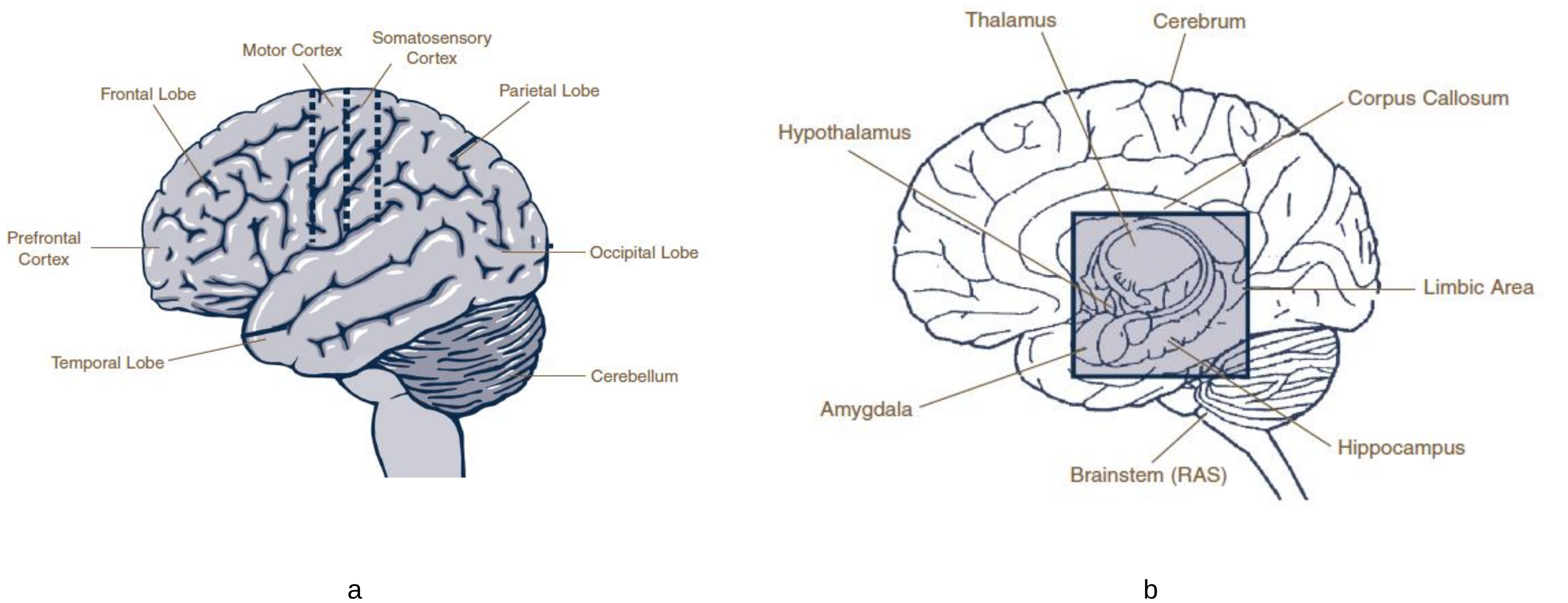


Figure 2-3: (a). The major exterior regions of the brain, (b) cross section of the human brain. (adopted from Hart, 2002).

There are several administration procedures which can be used for the verbal command form of the test. The variation between them ranges from slight differences in the command to obvious differences in the requested time setting. Almost all the proposed CDTs fall into one of the following three categories, which are illustrated in Figure 2-4:

1. Free-hand CDT, in which the individual is instructed to draw the face of the clock on a blank sheet of paper.

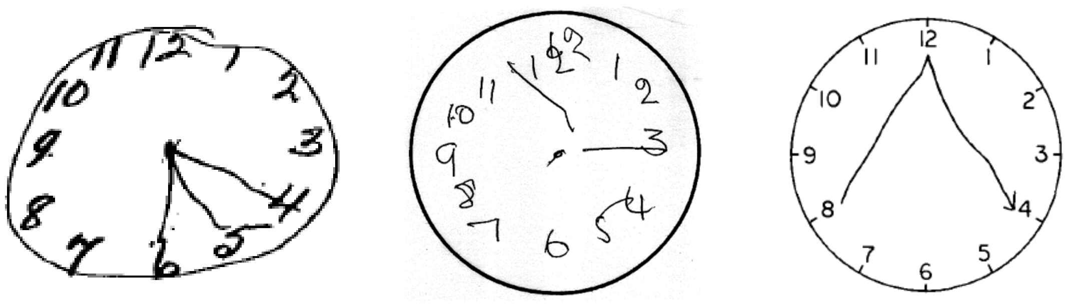
2. Pre-drawn CDT; in this type the individual is provided with a sheet of paper with a pre-drawn circle as a contour of the clock, and he or she is asked to complete the clock (write the numbers and set the time).

3. Examiner clock, where the individual is given a clock drawing with all the numbers written on it, then he or she is asked to set the hands to a specific time.

Free-hand CDT has been criticised for the influence that clock contour might have on the rest of the test, for example if the circle is drawn too small or distorted (Pinto and Peters, 2009).

The required time setting can also be varied in CDT. Most of the administration procedures ask for a time setting, however, some don't require any time settings (Pinto and Peters, 2009). The most sensitive time settings to neurocognitive dysfunction and hence the most widely used are (in descending order) "10 past 11", "20 past 8", and "3 o' clock". For the first two time settings the hands are drawn in both the right and left visual fields, which calls upon the functionality of both hemispheres of the brain. The difference between them is that "10 past 11" draws the hands in the upper half of the clock, placing demand on the temporal lobe and executive functions, while "20 past 8" draws the hands in the lower half of the clock, placing demand on the parietal lobe.

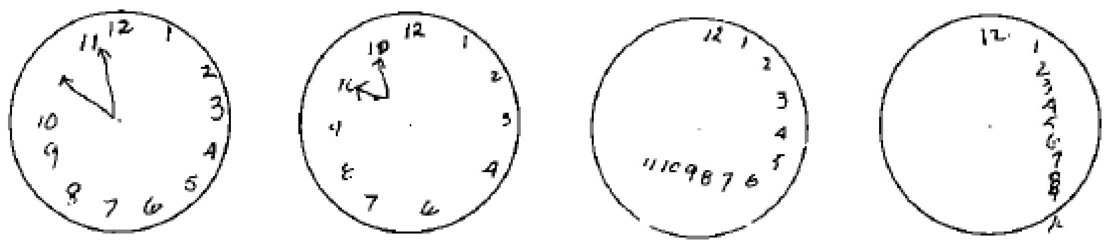
The setting “10 past 11” has been reported as being more sensitive for the diagnosis of dementia (Freedman et al., 1994).



a b c

Figure 2-4: Examples of CDT administration: (a) free-hand CDT, (b) pre-drawn CDT, (c) examiner CDT.

Dementia is a progressive syndrome; cognitive impairment degrades over time, and therefore the patient faces more difficulty in drawing the clock as more cognitive functions become affected. Figure 2-5 shows four successive clock drawings over a period of 18 months, and illustrates that the produced clocks are becomes more deteriorated over the time.



Initial After 6 months After 9 months After 18 months

Figure 2-5: Clock drawings showing the cognitive deterioration over time (adopted from Freedman et al., 1994).

2.3.1 **CDT Scoring Systems**

There are many scoring systems which have been developed to interpret clock drawings and to diagnose cognitive impairments from the drawings. Each of these systems places an emphasis on a subset of cognitive functions. The systems have different complexities, ranging from a simple binary rating to more complex qualitative and quantitative assessments (Ismail et al., 2010).

More than fifteen common scoring systems can be identified from the existing literature. These are validated using a varied number of patients and healthy people, and most of them use the verbal command method (Pinto and Peters, 2009). Some of these systems use free-hand CDT administration (Sunderland et al., 1989; Mendez et al.,1992; Royall et al.,1998; Kanchanatawan et al.,2006), while others employ the pre-drawn CDT demonstration (Shulman et al., 1986; Tuokko; Freedman et al., 1994; Manos, 1999).

The scoring systems provide qualitative or quantitative criteria for evaluation of the drawing errors related to cognitive impairments. Each system uses a cut off value to differentiate between the normal and abnormal cases.

During the development of the scoring systems, most of the researchers propose their own methods to assess errors, which could be qualitative and/or quantitative. The methods are usually proposed based on the personal experience of the researchers using the CDT. The above explains why most of the scoring systems employ different lists of assessed errors (features) (Shulman et al, 1986; Mendez et al, 1992).

After proposing a list of features, researchers validate the robustness of the criteria in diagnosing the positive dementia cases. Statistical measures (sensitivity and specificity) are two measures used to describe the performance of the system. Sensitivity is the percentage of the positive cases which are correctly diagnosed, meaning the ability of the test to identify the people with dementia.

Specificity is the percentage of negative cases that are correctly diagnosed as healthy. Equations (2.1) and (2.2) show how the values of sensitivity and specificity are calculated (Choi, 1992).

$$Sensitivity = \frac{number\ of\ true\ positive}{number\ of\ true\ positive + number\ of\ false\ negatives} \quad (2.1)$$

$$Specificity = \frac{number\ of\ true\ negative}{number\ of\ true\ negative + number\ of\ false\ positive} \quad (2.2)$$

Different numbers of patients and control samples are used in the process of validating scoring systems, ranging from 37 to 648 individuals (Pinto and Peters, 2009). The generalisability of some scoring systems is questionable because of the small sample size (Kouk, 2010). It has been reported that there are only two scoring systems which are proposed after a systematic study has been conducted, those of: Freedman et al. (1994); and Tuokko et al., (1992). They studied the clock feature (errors) to decide which features should be included as criteria in the scoring systems. However, there is no consensus in the literature about which scoring system is the best (Aprahamian et al., 2009).

Comparisons between the scoring systems are difficult because of the diverse patient samples used to validate the systems, and also because of differences in the scoring criteria (Aprahamian et al., 2009; Pinto and Peters, 2009).

Often the CDT scoring systems produce lower sensitivity and specificity when they are compared to the original validation by the authors who proposed the systems

(Aprahamian et al., 2009).Several comparative studies in literature compare a variety of scoring systems. They report systems proposed by Shulman et al. (1986), Sunderland et al. (1989), and Mendez et al. (1992) as having the best diagnostic accuracy (Aprahamian et al., 2009). The scoring system proposed by Tuokko et al., (1992) has been said to have good discriminatory power (Pinto and Peters, 2009), but it is criticised for its complexity (Jouk and Tuokko, 2012).

Much research has been conducted to test the capability of the available CDT scoring systems to diagnose MCI and the early stages of dementia. These studies employ some of the available scoring systems described in literature (Pinto and Peters, 2009; Ehreke et al., 2010; Ehreke et al., 2011).

Ehreke et al. (2011) report that no existing CDT scoring system is suitable to be used as an assessment tool for MCI. They suggest that more focus should be placed on the clock hands and the numbers and detailed features should be used. This supports previous studies which demonstrate that more complex scoring systems are more sensitive to the early stage of dementia (Mainland, and Shulman, 2013).

2.3.2 **CDT Errors (Features) Significance**

Experts in the field of clock drawing tests have agreed that the best scoring system should use only a small number of errors which are easy to score and give a reliable assessment of cognitive impairment (Jouk and Tuokko, 2012). Only a few studies have examined the significance of the elements (i.e. features) of the clock drawings for the correct diagnosis of dementia. Lessing et al. (2008) have studied the features which indicate errors in the clock drawings in an attempt to reduce the number of features required and to select the most important features that indicate dysfunction. Their study, combining three scoring systems (Mendez, Tuokko, and Shulman) identifies a list of 24 features. The six most important features, as reported in Lessing et al. (2008) are: wrong time; no hands; missing numbers; number substitution; repetition; and refusal to draw a clock. Jouk and Tuokko (2012) have studied Tuokko's scoring system to find the most important features out of the 24 employed. The experiment uses 356 clock drawings composed of 80 classified as dementia cases and 276 classified as normal. A binary value represents the dementia status (0 = normal, 1 = dementia). The authors employ logistic regression to find the significant features in the clock drawings. The experiment highlights the five most significant features: missing numbers; repeated numbers; number orientation; extra marks; and number distance. Both studies assume that the features are independent and they study the relation between dementia status and each feature individually. However, neither of these studies includes MCI data. MCI cases have been neglected based on the argument that CDT is not an appropriate tool for diagnosing MCI. This may be true if the current scoring systems are used, but it does not mean the test is inappropriate for MCI if more appropriate, detailed features could be developed.

2.3.3 **Main Categories of Clock Features (Deficits)**

The most common errors occurring within the produced clock drawings are grouped into six categories by Roulrau et al. (1992): (1) size of the clock; (2) graphic difficulties; (3) stimulus-bound response; (4) conceptual deficits; (5) spatial and/or planning deficits; and finally (6) preservation. Since the data that is used for the present research consists only of clocks drawn onto pre-drawn faces, the size of clock category is not applicable here. The five other categories are described in more detail as follows:

I. Graphical difficulties

Graphical difficulties are present if the clock hands are not straight, numbers are difficult to read, or if the produced clock is distorted in general, but overall the produced drawing can still be recognised as a clock.

This type of error is more common among the moderate VaD cases than the AD cases, and it becomes worse as VaD progress. Graphical Difficulties occur because of secondary disruption of the frontostriatal circuits which coordinate motor control and planning (Eknoyan et al., 2012).

II. Stimulus-bound response

Stimulus-bound response is the tendency of the clock drawings to be dominated or guided by a single stimulus. For example, in the case of "10 past 11" time settings, a stimulus-bound response appears as setting the time as "10 to 11". In the present research the time setting used is "5 to 3", and so the patient might be attracted to the strong stimulus source, which is 5. Therefor the time may be set as "25 past 3". This type of error also includes writing the time using letters or numbers beside 3 or 5, and is also related with the conceptual error category.

Stimulus-bound response errors are more common in AD than in the other types of dementia (Eknoyan et al., 2012).

III. Conceptual deficit

Conceptual deficit is the loss of (or difficult in accessing) knowledge relating to the general features and even the meaning of clocks. This category includes a wide range of errors such as misrepresentation of the clock (the clock does not look like a clock) or misrepresentation of the time (the hands are absent, the time written on the clock). Time setting represents the real meaning of the clock: a tool for communicating the time.

Conceptual deficit is related more with AD than with other types of dementia, though it may also occur in MCI cases and the early stage of AD. Conceptual deficits are caused by an impairment of the semantic memory in the temporal lope, which stores conceptual knowledge (Eknoyan et al., 2012).

IV. Spatial and/or planning deficits

The spatial and planning errors may be reflected in the produced clock as: neglect of the left hemispace; gaps before the numbers 12 3, 6, or 9; numbers written outside the clock; numbers written counterclockwise; deficits in number spacing.

This category of errors has been found to be common in most types of dementia. However, they are present in VaD more than AD. The errors in this category are largely due to an impairment of the right parietal lope. Another cause for this type of error is an impairment of the circuits that communicate between the parietal lobe and the frontal lobe, and also the circuits that communicate between the frontal lobe and the subcortical area (Eknoyan et al., 2012).

V. Preservation

Preservation is the continuance or repetition of the same activity without any stimulus. This category includes drawing more than two hands, writing numbers beyond "12" or repeating the same numbers.

Preservation errors are most common in AD. They occur due to an impairment of prefrontal area that can affect executive functions (Eknoyan et al.,2012).

2.3.4 **Longitudinal and Error Analysis of CDT**

Several researchers have studied the longitudinal analysis of the CDT (Shulman et al., 1993; Rouleau et al., 1996; Lee et al., 2011). The first study (Shulman et al., 1993) includes 183 subjects, with 4 assessments being carried out in 6 month intervals, and analysis being conducted using the modified Shulman scoring system. The clocks are scored and the result of the CDT shows strong correlation with other cognitive assessment tools. 33 AD patients participated in the second study (Rouleau et al., 1996), with three drawing tests taking place in consecutive years. Results are assessed using the 10 point revised scale system. Qualitative criteria are used to analyse the effects of the progress of dementia on the clock score and on the category of the qualitative error. The third study (Lee et al., 2011) is conducted with 370 patients suffering from different types of dementia. The study employs Mano's scoring system along with qualitative analysis. It reports that qualitative error analysis can help to differentiate between dementia subtypes.

One research study in the literature has conducted a qualitative analysis of drawings with errors at different levels of severity (Kitaqbayashi et al., 2001). 67 probable AD patients, 44 probable VaD patients, and 8 healthy control subjects participated. The study shows that in very mild VaD cases the frequency (how often a certain error is found in the drawings) of all the errors is below 20 %. In the mild VaD cases the spatial/planning deficit had a frequency of 68 % and the conceptual deficits had 63 % frequency. In moderate VaD the conceptual deficit frequency is 90 %. In AD there is no significant difference in the frequency, with conceptual deficits being 60 % and spatial/planning deficits being 40-60 % for all levels of severity.

2.3.5 **Diagnostic Accuracy of the CDT**

Many scoring systems have been developed to capture the deficit in clock drawings and to diagnose abnormalities in cognitive function. The sensitivity and specificity are used as measures of the performance rather than the diagnostic accuracy. The developers of most scoring systems report a good sensitivity and specificity in diagnosing the demented cases. However, lower figures are reported by other researchers who conducted comparable studies (Berger et al., 2008; Lee et al., 1996; Aprahamian et al., 2009). All of these scoring systems are used to classify the drawings into normal/abnormal groups.

The performance of the CDT has been rated by dementia specialists as reported by Nair et al. (2010). The authors report the results of dichotomous rating. The specialists discriminate between the drawings of MCI and healthy subjects with a sensitivity and specificity of 47 % and 81 % respectively, while the sensitivity and specificity of diagnosing AD from healthy cases is 75 % and 81 %. The specialists discriminate between the drawing of AD and MCI with sensitivity and specificity of 75 %, and 53 % respectively. Finally the reported sensitivity and the specificity of the CDT in discriminating between abnormal (AD and MCI), and healthy cases are 61 % and 84 % respectively.

2.4 **MMSE**

Mini Mental State Examination (MMSE) is another popular dementia test. This test is used by practitioners such as neuropsychologists and GPs to assess cognitive functions. The MMSE consists of a list of 30 questions which assess orientation, registration, attention and calculation, recall capability, and language abilities (Folstein et al., 1975).

In this test the patient scores one point for answering each question correctly. Therefore, the maximum score is 30, when all the questions are answered correctly. The individual is considered normal if a score of 27 or above is attained. However, this does not necessarily mean that a score below 27 indicates dementia. The original cutoff of the MMSE is 23.

Many studies have been conducted to test the performance of combining the CDT with the MMSE in diagnosing dementia. It has been reported that combining the two tests can enhance the accuracy of diagnosing dementia (Heinikl et al., 2003; Aprahamian et al., 2010; Kato et al., 2013).

2.5 **Clinical Decision Support System (CDSS)**

The Clinical Decision Support System (CDSS) is a piece of software which takes a set of input information about a patient's clinical situation and produces an output that can help practitioners to make decisions. A CDSS is employed to reduce human error, to automate routine tasks, to address information overload, and to make the clinical guidelines available and accessible to a wide range of medical staff (Bemmel and Musen, 1997). CDSSs can provide practitioners, patients, and other individuals with information that is filtered and processed as it is needed (Zheng, 2010).

The use of CDSSs has increased steadily over the last 10 years. In the existing literature, there are over 5000 articles relating to CDSSs, and they suggest that all CDSSs can be classified into one of the following categories: (1) systems of general diagnosis, which suggest differential diagnoses and work-up protocols, (2) systems for a limited number of clinical diagnoses, which produce a specific diagnosis among them, (3) specific systems which are designed to interpret a certain category of images, such as digitised Xray images or pathology slides (Miller, 2009).

There are a wide range of applications for CDSSs within health care. Around 100 CDSSs have been reviewed and classified in this context by Garg et al. (2005), who defined the following categories:

1. Systems for disease management (40 %)

2. Systems for drug dosing and prescribing (29 %)

3. Reminder systems for prevention (21 %)

4. Systems for diagnosis (10 %)

The focus of the vast majority of the proposed CDSSs is on managing already diagnosed cases, and managing drug dosing or drug prescribing, rather than assisting with diagnoses.

Metzger et al. (2002) also classify CDSSs according to the timing at which the CDSS provides the support (before, during, or after the practitioners make the decision).

Based on internet technology, CDSSs can be classified as stand-alone or web-based systems. Based on the target clinical domain, CDSSs can also be classified into different clinical areas (Kong, 2011).

With regard to the computer algorithms, CDSSs can be categorised into two types: knowledge-based systems, and non-knowledge-based systems (Berner and Lande, 2007; Prasath et al., 2013; Musen and Middleton, 2014). The knowledge-based systems consist of three parts: the knowledge base; the inference engine; and the mechanism for communication with the user.

The knowledge-based systems are the most common type of CDSSs. They rely on medical information compiled in the form of IF-THEN rules. The following is an example of a system providing support to laboratory test ordering: IF new specific test is ordered AND IF the same test was previously ordered within the last two days, THEN alert the practitioner. In this situation the rules are prepared to avoid duplicating the same test. Another example is the diagnostic CDSS, which provides suggestions about the diagnosis to the physicians. The knowledge base contains information about the symptoms of various diseases. The inference engine employs the necessary formulae to combine the knowledge base with the patient's data. Finally, the communication part of the system is used to interface with the user, to show results and to receive input data (Berner and Lande, 2007).

Many different types of CDSS employ rule-based techniques such as: alerts and reminders, diagnostic assistance, therapy critiquing and planning, prescribing DSS, information retrieval, and image recognition and interpretation (Coria, 2003).

One of the earliest CDSS that is classified as knowledge-based is the MYCIN system (Shortliffe, 1976). In this system the knowledge of infectious diseases is represented as 600 rules, which are formed based on consultation with medical experts.

Unlike knowledge-based systems, the non-knowledge-based variety does not contain a knowledge base, and do not rely on the medical literature or expert physicians' knowledge. They also do not need IF-THEN rules; instead they employ data mining algorithms such as neural networks, classifiers, or genetic algorithms. This type of CDSS can learn from the past data of known diagnoses. It does not need any prior knowledge from medical literature or from experts in the medical area. The system makes decisions by studying the patterns within the data to find the relationships between the input features (signs and symptoms) and the diagnoses. After the system is trained it can be used to diagnose the new cases based on their input features (Berner and Lande, 2007). Non-knowledge-based systems are the preferred choice when relevant prior medical knowledge is limited or does not exist (Hardin and Chieng, 2007).

The main advantage of using non-knowledge-based systems is that they do not need to employ IF-THEN rules and do not rely on any medical information from expert clinicians. However, this type of CDSS cannot explain or justify the chosen decision (Berner and Lande, 2007).

The knowledge-based systems are called evidence-adaptive CDSSs when they are designed to use a clinical knowledge base which is derived from, and continually reflects, the most up-to-date evidence from the research literature and practice-based sources (Sim et al., 2001).

With the development of computational power and medical technology, large medical datasets and classification algorithms have become available. Consequently, data mining has gained considerable interest. It has begun to be used in many CDSSs for various applications such as: medical imaging recognition and interpretation systems, gene and protein expression analysis, education systems, laboratory systems, acute care systems, and other miscellaneous systems.

2.6 Supervised Machine Learning

The area of Machine Learning (ML) is a subfield of data mining, concerning algorithms and programs that are able to improve their performance by learning from data. The knowledge among the dataset is transferred to the system via a process called training. Many research studies have been conducted to develop supervised classification algorithms. The algorithms are trained using training data, with every observation in the data being represented by the same number of features (attributes). These features could be binary, continuous, or categorical. As the volume of training data increases, the pattern recognition will become more accurate. The produced predictive model is built to diagnose the future case, value, or entity.

Generally, according to Kotsiantis (2007), the process of ML consists of five steps. It begins with collecting the data, which sometimes requires an expert in the area of interest to select the feature that should be measured. If there is no expert available, measurement of all the possible features may take place, though this could lead to measurement of irrelevant or redundant features.

The second step is data pre-processing. During this step issues such as missing values, discretisation, and noise removal are resolved by employing pre-processing algorithms.

The third step is to perform feature selection to identify and remove the irrelevant and redundant features. This can speed up the learning algorithm and enhance its performance. Selecting the learning algorithm is a critical step, and a diverse range of algorithms will be discussed in the flowing subsection.

The final step is the evaluation. There are at least three approaches to perform this step, all of which are based on dividing the available data into a training set and an unlabeled testing set. The only difference is how the division is between these data sets is made. Cross-validation, and leave one out, are examples of validation procedures.

If the accuracy rate is not satisfactory, the previous steps are analysed to establish the reason, as a wide range of factors can lead to poor accuracy and should be checked. These include factors such as: relevant features are not included, imbalanced data, inappropriate learning algorithm, or high dimensionality. In the following section only classification learning will be reviewed. Supervised learning classification refers to the creation of a model, known as a classifier that is built based on the training data; this model predicts the class of undiagnosed input features.

Decision Trees

Decision trees are a classifier that characterises the observations according to the value of the features. Each feature is represented by a node in the tree; with each branch from the node representing a value that the feature can assume. The classification starts at the root node. The features that best divide the data are selected as nodes. Different measures are used to find the best features such as: information gain, gini index, and refliefF algorithms. The data division continues until the data is split into two subsets belonging to the same class. Many decision trees classification algorithms have been proposed (e.g. CART (Duda et al., 2001), ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993)). Decision trees are employed by Gerald et al. (2002) to develop a CDSS that assisted health workers in a tuberculin skin test.

Neural Networks

The original idea behind the neural network is inspired by the mechanism of patterns recognition in the brain. A neural network can approximate any relation between the class label and the features, and they can also deal with multi-class data. An example of employing neural networks in CDSSs is the Computer-Aided Diagnosis of Solid

Breast Nodules (Joo et al., 2004), where the authors have reported a sensitivity of 99%.

K-Nearest Neighbor Classifier

This classifier is based on the principle that the observations in the data are generally close to the other similar observations which belong to the same class. KNN assigns to an unlabeled observation the dominant class among the K nearest neighbors. Many different metrics have been used to calculate the distance between the observations within n-dimensional space, where n is the number of features in the dataset (Kotsianits, 2007).

The value of K influences the performance of the classifier. When the data is noisy, choosing a large K might improve the performance. When the some classes in the data are located in small regions, a small k is preferred so as to not include observations from the other classes that surround the targeted class (Kotsianits et al., 2006).

This classifier has been employed by Burroni et al. (2004) for the CDSSs that assist clinicians with early diagnosis of melanoma based on the analysis of digitised epiluminescence microscopy images.

Support Vector Machines

SVMs are statistical classifiers that use a discriminant hyper-plane to distinguish between the classes. The selected hyper-plane is the one that maximises the margins between the classes. Maximising the margins minimises the risk of over-fitting the training data (Cortes, and Vapnik, 1995).

SVM is a linear classifier suitable for binary classification problems. However, it has the capability to work within high dimensional feature spaces without any extra computational complexity. This is achieved by mapping the data into a higher dimensional space by using Kernel functions (Ben-Hur and Weston, 2010), though it does not always improve the performance. A linear kernel has been reported to provide the best performance in many applications, and requires only one parameter to be tuned (Ben-Hur et al., 2010).

SVMs have been employed in many CDSSs to help clinicians to make decisions about diagnosis of disease and interpretation of medical tests. For example, in (Yu et al., 2010) an approach based on SVM techniques is proposed for the classification of persons with and without common diseases. Several studies use SVM for diagnosing neurological and psychiatric disorders by utilising a diverse range of neuroimaging techniques (Orru et al., 2012). The SVM is used in automatic computer-aided diagnosis systems for early diagnosis of AD by the means of SPECT imaging (Ramirez et al., 2013).

Random Forest

RF is an ensemble classifier that consists of many decision trees. RF seeks to address the problem of instability associated with single trees and their sensitivity to the training data. The output class of RF is the statistical model of the output of individual trees. RFs combine the “Bagging” concept (Breiman, 2001), with a random subset of features (Ho, 1995).

Each tree is built on a separate bootstrapped sample, and only a randomly selected feature subset is used at each node. In this case a variation among the trees is obtained. The performance of the RF is not sensitive to the values of their parameters [Yeh et al., 2012].

RFs are a widely used classifier in some applications because they are easy to use, only two parameters need to be set, and they make no distributional assumptions. RFs have been applied in many domains such as image classification, and biomedical problems. Some comparative studies in the literature have reported a good performance of RFs compare with other classifiers, sometimes as good as SVMs (Caruana et al., 2008; Cutler et al., 2007; Bhattacharyya et al., 2011).

RF is used within many CDSSs to diagnose a variety of diseases. For example, Asaoka et al. (2014) have used RF classification to diagnose glaucoma in highly myopic and emmetropic eyes. RF is also used by Gray et al. (2013) to diagnose AD based on combining different types of feature data.

2.7 Cascade Classification

The cascade classifier, or multi-stage classifier, is a concatenation of multiple classifiers; the output of each classifier forming the input of the next. The first stage is trained using the entire dataset, while the following stages are trained using specific regions of interest. Cascade classification is used to reduce the computation cost of the classification process. There are two types of cascade classification. The first is the multi-stage rejection classifier (MSRC), which can deal with multi-class data and is able to make a classification decision at any stage of the cascade. In contrast to MSRC, the second type which is detection cascade deals only with binary classification problems. At each stage a partial decision is taken, and the final classification is delayed until the final stage. Detection cascades are capable of dealing with unbalanced data (Trapeznikov et al., 2012). Cascade classifiers have been widely used in many applications, such as handwriting recognition, face recognition and medical diagnoses (Zhang et al., 2007; Saatci and Town, 2006; Trapeznikov et al., 2012).

Several research projects have employed cascade classifiers in medical diagnoses. For example, cascade classification has been used to diagnose the degree of fibrosis in patients with chronic hepatitis C infections (Hashem et al., 2012). The study reports better classification accuracy of the cascade classifier compared to the single stage classifier. Cascade classification is also used in diagnosing diabetes (Polat et al., 2008), where the cascade system consisted of two stages, and the classification accuracy achieved was 82.05 %. A recent study (Zhang et al., 2013) uses cascade classification to diagnose breast cancer. The system is evaluated using 361 images and produces 99.25 % classification accuracy.

2.8 Image Enhancement

The aim of image enhancement is to improve the visibility of the image or to provide better representation for image processing tasks such as: image segmentation, detection, and feature extraction. During the image enhancement process one or more features of the image is transformed into a new range of values. The process also includes image cleaning, contrast enhancement, image smoothing, and morphological processing (Maini and Aggarwal, 2010).

The methods of image enhancement can be classified into spatial domain methods, where the operations are performed directly on the image pixels, and frequency domain methods, which process the transformed images using techniques such as Fourier, wavelet, and cosine transformation (wang and Tan, 2011).

Contrast enhancement (also called intensity adjustment) employs a contrast stretching technique to map the intensity values of an original image into new values which have increased contrast (Gonzalez and Wood, 2010). This process expands the range of the intensity levels. Image cleaning or noise removal is used to solve the impulse (background) noise. A median filter of neighborhood size 3-by-3 (Gonzalez and Wood, 2010) is employed to remove this type of noise before applying morphological processing, which would otherwise serve to amplify the noise.

Morphological operations are techniques which are applied to the input images to create output images of the same size. The value of each pixel in the output image depends on the comparison of the pixel in the input image against its neighbors; this process is performed by applying structuring elements to the input image. Techniques such as erosion, dilation, opening, and closing are used to perform the morphological processing (Sreedhar et al., 2012).

2.9 Computer Based Cognitive Assessment Tools

For more than 30 years, research projects have examined the development of computerised tools to perform neuropsychological assessment. These assessment tools fall into one of three categories: (1) computerisation of an existing test by using the computer to perform the administration, (2) an entirely newly-developed computer-based test to assess cognitive abilities, and (3) tools which fall between the previous two categories. These tools benefit from the capability of computers to administrate and analyse the existing tests in new innovative ways (Wild et al., 2008).

An Automated Neuropsychological Assessment Matrix (ANAM) is a computerised tool which is developed originally for the department of defence. It examines memory, attention, psychomotor, language and reaction skills (Rice et al., 2011). Some researchers have studied the capabilities of the tool, reporting good correlation with various other assessment tests. The test takes approximately 30 minutes to administer (Wild et al., 2008).

The Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment (CANS-MCI) is another computerised tool, which is designed for the diagnosis of MCI by assessing language capability, memory, and executive function. It has been reported that this test also takes approximately 30 minutes to be completed (Tornatore et al., 2005).

The Cambridge Neuropsychological Test Automated Battery (CANTAB) is a tool which focuses on working memory, planning, attention, and visuospatial memory (Robbins and Sahakian, 2002).

The Central Nervous System Vital Signs (CNS - Vital Signs) computerised tool includes seven cognitive assessments covering memory, psychomotor speed, reaction time, cognitive flexibility, and complex attention. This test takes approximately 30 minutes to be completed (Bojar et al., 2012).

Another tool known as a Computerised Neuropsychological Test Battery (CNTB) is one of the earliest computerised tools. It performs an assessment via 11 subtests in order to measure the cognitive ability domains of memory, attention, verbal and spatial memory, language, motor speed, information processing, and spatial abilities (Veroff et al., 1991).

The Cognitive Drug Research Computerised Assessment System (COGDRAS) is designed to detect the effect of drugs on cognitive functions of the elderly, and has been employed to detect dementia (COGDRAS-D). The tool consists of eight subtests: immediate word recognition; simple and choice; reaction time; digit vigilance; delayed picture recognition; delayed word recognition; delayed face recognition; and memory scanning. The system takes approximately 15 to 20 minutes to administer (Simpson, et al., 1991).

CogState is a Neuropsychological test for measuring simple, choice, and complex reaction, continuous monitoring, working memory, matching, incidental learning, and associative learning. It takes approximately 15 to 20 minutes to complete (Maruff et al., 2009).

The Cognitive stability Index (CSI) is a computerised test battery designed for screening for general cognitive impairment in individuals. This tool reportedly takes approximately 30 minutes to administer, and employs 10 subtests which focus on memory, attention, response speed, and processing speed (Lapshin et al., 2012).

2.10 Computer-Based CDT

Two studies report the use of computers to analyse and diagnose CDT drawings (Heinik et al., 2010; Kim, 2013). Their focus however is on the capturing of the clock drawing process using a digitiser or tablet computer, rather than on analysing the visual features in the drawings and their significance to the diagnosis. The first of these studies (Heinik et al., 2010) examines the importance of the kinematic features in the diagnosis of mild Major Depressive Disorder (MDD). The focus group includes 20 patients and 20 healthy individuals. The study explores seven features: the mean number of drawn segments, mean time to complete the task, mean pressure on the writing surface, mean angle between the pen projection and the north line, segments' width, segments' height, and segments' length. The achieved accuracy in classifying mild MDD is 81.1 %, with the most important factors for the classification being the pressure, the segment width, mean segment length, angle between the pen and the north line, and the segment height (in that order).

The second study involves a computerised CDT assessment (Kim, 2013) using a system called ClockME. This method employs a tablet computer for recording and playing back the drawing process so that clinicians can study the planning strategy of the patients. The study introduces air time as a new feature, which is defined as the time period for which the individual would stop drawing before continuing again. This feature is introduced as it may give an indication of memory problems apparent from the patient's struggling to recall some information related to the clock. The author also focuses on measuring the pressure of the stylus on the screen during handwriting of the numbers. The clock drawings are analysed using the scoring system proposed by Freedman et al. (1994). This research involves 45 patients and healthy volunteers, and three practitioners to evaluate the proposed system. The usability of the system from the perspective of both the subjects and the practitioners is examined. The average accuracy in recognising the numbers drawn is 84 %. However, the study does not specify the accuracy of the system in differentiating between normal and abnormal cases. There is also no comparison between the proposed system and the paper-based CDT.

2.11 Discretisation

Discretisation is a process that transforms continuous data into data with discrete attributes without significant loss of information. When a ‘many to one’ transformation is performed, each value in the original data is mapped onto a new discrete value. Discretisation is a long-standing problem and has been studied extensively (Liu et al., 2002; Kotsiantis and Kanellopoulos, 2006; Yang and Webb, 2002). It can be considered as a data reduction technique as it maps the attributes from a wide range of values into a limited set of discreet values (Garcia et al., 2013).

There are several reasons to use discretisation, for example many machine learning algorithms require discrete data (Dougherty et al., 1995). Though algorithms do exist which can deal with continuous data, learning is more effective with discretised data. Most real-world data is continuous in nature, and efficient algorithms cannot be applied if the data is not first discretised. Therefore, discretisation is often used in advance of the learning stage. The use of discretisation has several advantages: discrete data requires less memory space; discretisation brings the data closer to a knowledge-level representation, which facilitates the understanding of the data; and discretisation improves the accuracy of the learning algorithm, accelerating the learning process (Dash et al., 2011).

Discretisation is usually performed off-line as pre-processing stage before the learning stage. A typical discretisation task consists of the following steps: (1) the continuous values of each feature are sorted; (2) a cut-off point for splitting or merging the adjacent bin “intervals” is calculated; (3) based on some criterion, splitting or merging of bins is carried out; and (4) the process is terminated according to a stopping criterion (Liu et al., 2002).

Numerous discretisation methods have been reported in the literature. Generally, these can be classified as: (1) supervised or unsupervised, (2) global or local, (3) direct or indirect, (4) static or dynamic, (5) top-down or bottom-up.

Discretisation generally causes some degree of loss of information from the continuous data. Several proposed methods aim to reduce this loss as much as possible by employing various techniques (Garcia et al., 2013). Equal Width Discretisation (EWD) and Equal Frequency Discretisation (EFD) (Dougherty et al., 1995) are considered to be the simplest discretisation techniques, as they are classed as unsupervised binning methods. These two methods do not use any class label information.

Robect (1993) developed a supervised binning method known as 1R, which sorts the continuous values of the attributes into a number of intervals. The interval boundaries are then adjusted according to the relation between the attribute and the class, to ensure each interval has a minimum number of observations.

ChiMerge is a supervised merging method (Kerber, 1992). It begins by putting each distinct instance into a separate interval, and then it merges intervals based on the x^2 statistic. Using a large value of x^2 can cause over-discretisation, while a small value of x^2 can lead to under-discretisation.

Several discretisation methods that employ information theory have been proposed in the literature. Catlett (1991) introduced a supervised splitting method called D2, while Fayyad and Irani (1993) proposed a new discretisation method based on information theory and it is called Minimum Description Length (MDL). The authors employed the MDL criterion.

Ho and Scott, 1997 proposed a discretisation method called Zeta, which measures the strength between the class and the attribute. The method stops when a predefined number of intervals is reached.

Determining the most appropriate discretisation method for a given situation is very challenging. However, some comparative studies have been conducted in literature. Dougherty et al. (1995) compared the MDL, EWD, 1R methods, and the binary discretisation method. Two classifiers are used to assess the performance of these methods, and the authors found that MDL outperforms the others.

Liu et al. (2002) conducted another comparative study comparing the EWD, EFD, 1R, D2 MDL, Zeta, ChiMerge, Mantaras, and Chi2 methods. Thirteen datasets are used in the experiment. The C4.5 classifier is used to find the classification accuracy of each method. The result showed that MDL is the best choice.

2.12 Dimensionality Reduction

High-dimensional data is a significant problem in both supervised and unsupervised learning (Janecek et al., 2008). For instance, it has been shown that increasing the dimensionality of the dataset considerably retards the learning process. The presence of irrelevant features in the dataset may cause over-fitting, which leads to degradation of the classification accuracy (Yu and Liu, 2004). When a limited amount of data is used, the irrelevant features can disguise the distributions of the relevant features within the dataset, which leads to poor performance of the classification algorithms (Brown, 2009; Cheng et al., 2011).

The main motivation for reducing the dimensionality of the data and keeping the number of features as small as possible is to decrease the training time and the enhance the classification accuracy. This avoids over-fitting, and improves the classification accuracy (Guyon and Elisseeff, 2003; Jain et al., 2000; Liu and Yu, 2005).

Dimensionality reduction methods can be divided into two main groups: those based on feature extraction and those based on feature selection. Feature extraction methods transform existing features into a new feature space of lower dimensionality. During this process, new features are created based on linear or nonlinear combinations of features from the original set. Principal Component Analysis (PCA) (Bajwa et al., 2009; Turk and Pentland, 1991) and Linear Discriminant Analysis (LDA) (Tang et al., 2005; Yu and Yang, 2001) are two examples of such algorithms. Feature selection methods reduce the dimensionality by selecting a subset of features that can minimise certain cost functions (Jain et al., 2000; Guyon et al., 2006). Unlike feature extraction, feature selection does not alter the data and, as a result, it is the preferred choice when an understanding of the underlying physical process and data interpretation is essential. Feature extraction could be the better choice when discrimination only is needed (Jain et al., 2000).

Feature selection is normally used at the pre-processing stage before training a classifier. This process is also known as variable selection, feature reduction or variable subset selection. In terms of evaluation strategy, feature selection methods are categorised as classifier dependent (‘wrapper’ and ‘embedded’ methods) or classifier independent (‘filter’ methods).

Wrapper methods search the feature space, and test all possible subsets of feature combinations by using the prediction accuracy of a classifier as a measure of the selected subset's quality, without modifying the learning function. Therefore, wrapper methods can be combined with any learning machine (Guyon et al., 2006). They perform well because the selected subset is optimised for the classification algorithm. On the other hand, wrapper methods may suffer from over-fitting of the learning algorithm. This means that any changes in the learning model may reduce the usefulness of the subset. In addition, these methods are very expensive in terms of computational complexity, especially when handling extremely high-dimensional data (Brown et al., 2012; Cheng et al., 2011; Ding and Peng, 2005; Karegowda et al., 2010).

The feature selection stage in the embedlded methods is combined with the learning stage. These methods are less expensive in terms of computational complexity and less prone to over-fitting; however, they are limited in terms of generalisation, because they are very specific to the used learning algorithm (Guyon et al., 2006).

Classifier-independent methods rank features according to their relevance to the class label in the supervised learning. The relevance score is calculated using distance, information, correlation and consistency measures. Many techniques have been proposed to compute the relevance score, including Pearson correlation coefficients (Rodgers and Nicewander, 1988), Fisher's discriminate ratio "F score" (Lin et al., 2004), the Scatter criterion (Duda et al., 2001), Single Variable Classifier SVC (Guyon, 2003), Mutual Information (Battiti, 1994), and the Relief Algorithm (Kira and Rendell, 1992; Liu and Motoda, 2008).

The main advantages of the filter methods are their computational efficiency, scalability in terms of the dataset dimensionality, and independence from the classifier (Saeys et al., 2007). A common drawback of these methods is the lack of information about the interaction between the features and the classifier. In addition, selection is performed based on univariate filter methods, where each feature is considered individually and any dependency between features is ignored. This is unlikely to provide the optimal subset of features when there is a strong correlation between them and can lead to the production of a subset with redundant features (Fleuret, 2004). To overcome this problem, several multivariate filter methods, which incorporate feature dependencies, have been proposed (Battiti, 1994; Yang and Moody, 1999; Peng et al., 2005).

Information theory (Cover and Thomas, 2006) has been widely applied in filter methods, where information measures such as mutual information (MI) are used as a measure of the features' relevance and redundancy (Battiti, 1994). MI does not make any assumption of linearity between the variables, and can deal with categorical and numerical data with two or more class values (Meyer et al., 2008). There are several alternative measures in information theory that can be used to compute the relevance of features, namely mutual information, interaction information, conditional mutual information, and joint mutual information.

2.12.1 Feature Selection Methods Based on Information Theory

Information theory (Cover and Thomas, 2006) has been employed by several filter feature selection methods, all of which attempt to measure the significance of a feature or a subset of features for the purposes of classification. Information Gain (IG) (Guyon and Elisseeff, 2003) is the simplest of these methods. It is classified as a univariate feature selection method, as it ranks features based on the value of their mutual information with the class label. Simplicity and low computational costs are the main advantages of this method. However, it does not take into consideration the dependency between the features, rather, it assumes independency, which is not always the case. Therefore some of the selected features may carry redundant information. To tackle this problem new methods have been proposed for selecting relevant features, which are non-redundant with respect to each other.

For a feature set $F = \{f_1, f_2, \dots\dots, f_N\}$, the feature selection process identifies a subset of features S with dimension k where $\mathrm{k} \leq \mathrm{N}$, and $\mathrm{S} \subseteq \mathrm{F}$. In theory, the selected subset S should maximise the joint mutual information between the class label C and the subset S of a fixed size k.

$$I(S; C) = I(f_1, f_2, \dots\dots, f_k; C) \qquad (2.3)$$

However, such an approach is impractical, due to the number of calculations and the limited number of observations available for the calculation of the high-dimensional probability mass function. As a result, many methods use a heuristic approach to approximate the ideal solution.

Generally, criteria based on information theory concepts, such as feature relevance, redundancy and complementarity can be split into two groups: linear criteria, which are linear combinations of MI terms; and nonlinear criteria, which use maximum or minimum operations or normalised MI in their goal functions (Brown et al., 2012).

Battiti (1994) has introduced a first-order incremental search algorithm, known as the Mutual Information Feature Selection (MIFS) method, for selecting the most relevant k features from an initial set of n features. A greedy selection method is used to build the subset. Instead of calculating the joint MI between the selected features and the class label, Battiti (1994) studies the MI between the candidate feature and the class, and the relationship between the candidate and the already-selected features.

Kwak and Choi (2002) propose the MIFS-U method which improves the performance of the MIFS method by making a better estimation of the MI between the input feature and the class label. Another variant of MIFS, the mRMR method is proposed by Peng et al. (2005). The redundancy term in mRMR is divided over the cardinality $|S|$ of the selected subset S to balance the magnitude of this term, and to avoid it growing very large as the subsets expand. As reported in the existing literature (Brown et al., 2012; Peng et al., 2005), this modification allows mRMR to outperform the conventional MIFS and MIFS-U methods.

Estevez et al. (2009) propose an enhanced version of MIFS, MIFS-U and mRMR, called Normalised Mutual Information Feature Selection (NMIFS). It uses normalised MI (instead of MI) in the redundancy term. The normalisation of MI prevents bias towards multivalued features and limits the value of MI to the range of zero to unity (Estevez et al., 2009).

Hoque et al. (2014) propose a method called MIFS-ND. This method calculates the mutual information between the candidate feature and the class label, and the average of the mutual information between the candidate feature and the features within the selected subset. A genetic algorithm is employed to select the feature that maximises the mutual information with the class, and minimises the average mutual information with the other selected features.

Other proposed criteria (Yang et al., 1999; Fleuret, 2004; Meyer and Bontempi, 2006; Vidal-Naquet and Ullman, 2003) use the MI between the candidate feature and the class label in the context of the selected subset features. They utilise conditional mutual information, joint mutual information or feature interaction. Some of them apply cumulative summation approximations (Yang et al., 1999; Meyer and Bontempi, 2006), while others use the ‘maximum of the minimum’ criterion (Fleuret, 2004; Vidal-Naquet and Ullman, 2003).

Yang et al. (1999) propose a feature selection method called Joint Mutual Information (JMI). In this method, the candidate feature that maximises the cumulative summation of Joint Mutual Information with features of the selected subset is chosen and added into the subset. This method is reported to perform well in terms of classification accuracy and stability (Brown et al., 2012). Meyer and Bontempi (2006) introduce a similar method known as Double Input Symmetrical Relevance (DISR). The joint mutual information in the goal function of this method is substituted with symmetrical relevance.

Other methods that employ the ‘maximum of the minimum’ criterion have been proposed. Vidal-Naquet and Ullman (2003) introduce a method called Information

Fragment (IF), while Fleuret (2004) proposes Conditional Mutual Information Maximisation. This technique is typically the same as the IF method.

There are also several other proposed methods that use Feature Interaction. These methods select relevant features that maximise the interaction. For example, Jakulin (2005) proposes the Interaction Capping (IC) method, while El Akadi et al. (2008), propose a method which uses feature interaction, known as Interaction Gain Based Feature Selection (IGFS). However, this is typically the same as JMI.

Finally, Brown et al. (2012) study an MI-based feature selection criterion. They propose a general formula based on conditional likelihood, which can be used to derive many of the methods mentioned in this section. Generally, most of the methods which consist of linear combinations of mutual Information can be derived from this formula. However, the authors state that the goal function of the nonlinear method cannot be generated by this formula.

2.12.2 **Limitations of the Current Feature Selection Criteria**

In general, most of the methods listed in the previous section use the criteria consisting of two elements: the relevancy term and the redundancy term. The methods attempt to simultaneously maximise the relevancy term whilst minimising the redundancy term. It has been noted in literature that such feature selection methods have a number of limitations (Estevez et al. 2009; Peng et al. 2005).

For example, MIFS and MIFS-U share a common problem: when the number of selected features grows, the redundancy term grows in magnitude with respect to the relevancy term. In this case some irrelevant features may be selected. This problem has been partly solved in the mRMR, NMIFS, MIFS-ND methods by dividing the redundancy term over the cardinality of the subset.

Another problem shared by all above methods (MIFS, MIFS-U, mRMR, NMIFS, and MIFS-ND) is that the redundancy term is calculated based on the value of the MI between the candidate feature and the features within the selected subset, without any consideration of the class label. The features may share information between each other, but that does not mean they are redundant; they may in fact share different information with the class.

Yet another problem particular to the methods employing cumulative summation and forward search to approximate the solution of Eq. (14) (such as MIFS, NMIFS, mRMR, NMIFS, MIFS-ND, DISR, IGFS, and JMI) is the overestimation of the significance of some candidate features. For example, this can occur when the candidate feature is in complete correlation with one or several pre-selected features, but at the same time is almost independent from the majority of the subset. In such situation, the value of the goal function will be high despite the redundancy of the candidate feature to some features within the subset.

In practice, the significance of each of the above problems depends on the data and the characteristics of each particular data set.

2.13 **Summary**

This chapter has provided a background of dementia symptoms, and the use of clock drawing tests as an assessment tool for determining cognitive abilities. It has also presented a review of the techniques relevant to the system proposed for solving the problem. The main findings of the literature review are as follows:

1. Due to the lack of a fast and simple assessment tool, only 24 % of patients are assessed regularly. If the general practitioner were provided with a quick, simple, computer-based tool for diagnosis, a greater percentage of cases could be diagnosed in the early stage.

2. More than 15 scoring systems have been reported in the literature, each system using a set of specific features (clock errors) to assess the patient's cognitive function. However, there is no consensus about which features are most significant for the final diagnosis. Defining a list of significant features can lead to better interpretation of the CDT and also an enhancement in the performance of the CDT in diagnosing the abnormality of cognitive functions. However, a very limited number of research studies have addressed this issue.

3. Current CDT scoring systems are not suitable for diagnosing MCI and very early stages of dementia. However, focusing on the clock hands and numbers in more detail can improving the sensitivity of the test to diagnose these cases.

4. Feature selection techniques do not alter the data, and can therefore help to understand the underlying physical process and to interpret the data. They can be employed to define the significant clock features (errors) which enable differentiation between different cognitive statuses.

5. A new feature selection method is needed to overcome the primary drawback of existing methods, the overestimation of the significance of features. This may happen if the cumulative summation criterion employed. The new method would benefit from the strength of using joint mutual information whilst resolving the drawback of cumulative summation. Furthermore, this method could be employed to study the significance of CDT features.

6. There is a lack of studies explaining the temporal changes in the CDT features corresponding to the progress of dementia, and how they develop from MCT into the more severe stages.

7. Combining CDT and MMSE can enhance the overall accuracy of dementia diagnoses. Using both the CDT features and MMSE questions for the patient can improve the sensitivity of diagnosing dementia.

8. ML algorithms are being used in many applications including CDSSs. The Use of these algorithms can facilitate patient diagnosis from CDT analysis without relying on scoring systems.

9. Cascade classification techniques are used to enhance the performance of the classification task. These techniques have been used in many applications including medical diagnosis. The process of this type of classification is very similar to the mechanism that doctors follow when they make decisions during the diagnostic process. This type of classification can be used in the clinical decision support system to diagnose the CDT drawings.

Next chapter introduces the conceptual model of the proposed system and explains the medical data analysis conducted.

Chapter: 3

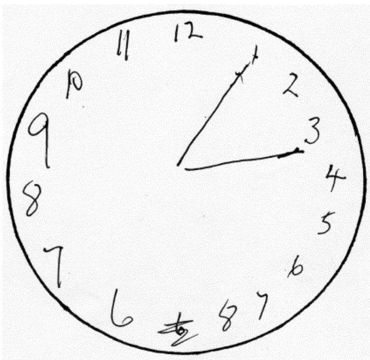
Research Methodology and Design

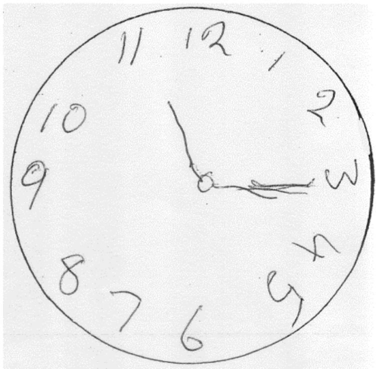
This chapter presents and discusses the conceptual model applied in this study, the contribution of proposing a new conceptual model to automate the diagnosis of the CDT, the proposed conceptual model also facilitate the in depth analysis of the CDT. In contrast to the only one existing system, the proposed system employs a sophisticated machine learning techniques to enhance the diagnostic performance of the CDT and analyses the CDT drawings. This chapter is organised as follows: the collection and characteristics of the CDT data used in the research is explained in section 3.1; Section 3.2 presents the data analysis; Section 3.3 introduces conceptual model of proposed system CDSS-DD; Finally, section 3.4 summarises the chapter.

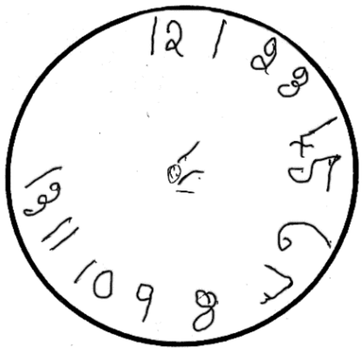
3.1 CDT Data

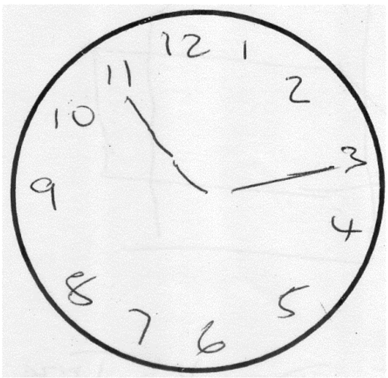
This research examines data provided by the Memory Clinic at the Llandough Hospital in Cardiff, UK. The data have been collected during the patients' examination procedures in the period between 1999 and 2009. The total number of clock drawings is 648. Each of the drawings is accompanied by the MMSE sheet, diagnosis, patient's age, and gender.

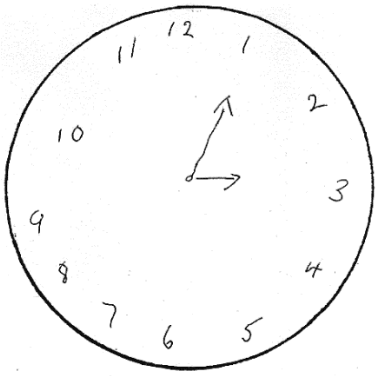
Working with real patients' data in this research can raise some ethical concerns, such as obtaining the patients permission to participate in the research, and whether the study may have any unforseen effect on the patients. Moreover the eligibility of the dementia patients to decide to take part in the research is questionable. Using data from a patient's medical record may also expose some confidential and personal information. To resolve these issues, the decision is taken to not ask patients to take part, and also to not have access to any medical records. The data used in this study are hence collected from patients' records by the staff in the memory clinic at the Llandough Hospital, and are then photocopied and anonymised before being delivered to the research lab. Based on these criteria an ethical approval is granted for this research by the South East Wales Research Ethics Committee. The size of the data used in the research is the size of the sample received from the Hospital. Figure 3.1 shows an example of the CDT data for various diagnoses.


A


b


c


D


e

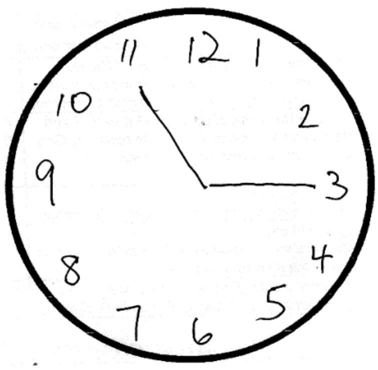
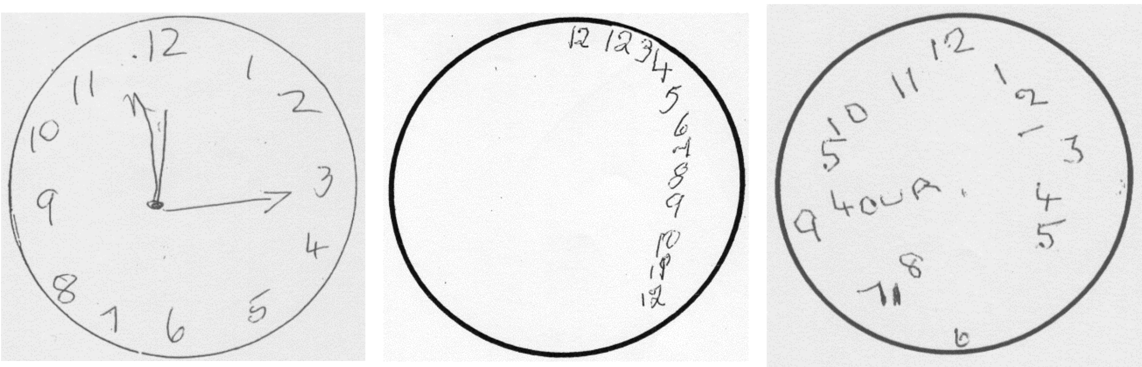

f

Figure 3-1: examples of the data received from the memory clinic at the Llandough Hospital, (a) AD, (b) VaD, (c) Dementia with Lewy Bodies, (d) Functional, (e) MCI, (f) Normal.

CDT is not a standalone test in the diagnosing process, the diagnosis should be made using standard research criteria and after a full clinical assessment. Another issue with the data is that no score is combined with the drawings and no scoring system is mentioned in the criteria used to evaluate the drawings.

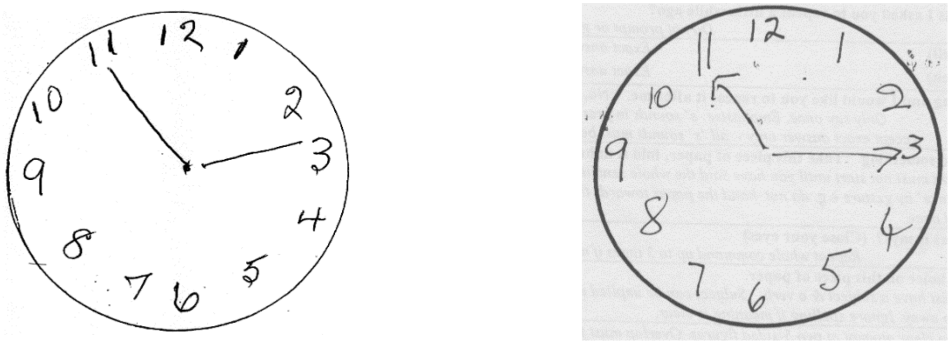
In addition, CDT is designed to reflect only impairment in the brain. It is not able to diagnose the disease behind that impairment. Therefore, patients who suffer from the same disease may produce clock drawings with different degrees of deterioration. Figure 3.2 shows three clock drawings drawn by three AD patients: drawing (a) by a 64 year old female; drawing (b) by an 84 year old female; and drawing (c) by a 79 year old male. Although the three patients suffer from the same disease, they produce drawings of varying quality.



A b c

Figure 3-2: Three clock drawings from three different patients who suffer from AD.

On the other hand, drawings by patients affected by different diseases can look very similar or even almost intact, as shown in figure 3.3. The reason behind this is that the quality of the produced drawing depends on which brain regions are affected by the disease and its severity level. Moreover the type of errors exhibited in the clock drawings can vary vastly different from one patient to another, even for those afflicted with the same disease. Some drawings may omit numbers, have the wrong time setting, wrong number positioning, omitted hands, or the patient may even show a complete inability to draw the clock.



A b

Figure 3-3: Two clock drawings, (a) 85 year old male suffering from AD, (b) 85 year old female suffering from VaD.

3.2 **Analysis**

The hard copy images data provided for the research is scanned. Image processing techniques are employed to enhance the quality of the images. One of the aims of this research is to classify any unclassified drawings based on the available information. Therefore, the overall task can be defined as image recognition and image classification.

The total number of clock drawings is 648. Each of the drawings is accompanied by the MMSE sheet, diagnosis, patient’s age, and gender. The number of drawings that are drawn by male patients are 263, and 385 drawings are drawn by female patients. Patient's age varies between 21 and 103 years, however, the vast majority of patients are between 56 and 82 years old.

There are many diseases which could cause dementia. Seventeen different diagnoses have been found within the CDT data. The number of observations (drawings) for some types of these diagnoses is very low. Following medical advice, the diagnoses in this study are grouped into seven distinctive classes based on the cause of the dementia. These are: (i) VaD (including VaD, mixed dementia and stroke); (ii) AD, (iii) Normal, (iv) MCI, (v) Functional (depression and anxiety), (vi) Other Degenerative Dementia (ODD) (such as Parkinson's disease, dementia with Lewy bodies and fronto-temporal dementia); and (vii) any other forms of dementia (e.g. tumour, alcoholism, head injury, etc.), Figure 3.4 shows the distribution of the grouped diagnoses in the CDT drawings. It is clear that even after grouping, some of the class sizes are very small in comparison to others. For this reason, the data for group v (ODD) and group vii (Others) will not be included in the study.

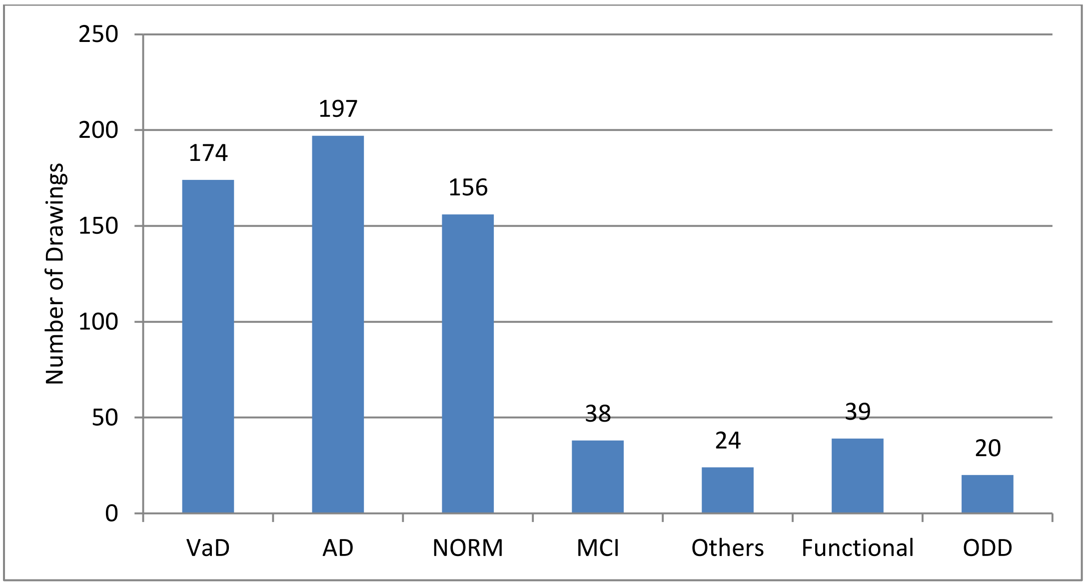


Figure 3-4: Data distribution of the grouped diagnosis in the CDT drawings.

In literature, there are several approaches for performing this image recognition process (Zhao et al., 2003). Due to the characteristics of the CDT data, a Geometric Feature Based (GFB) approach is employed. This is also beneficial because it gives the ability to study the significance of clock drawing errors made by patients. This aids in the diagnosis procedure.

In the GFB approach, local features within the image are measured in order to produce a digitised dataset. Each image in a dataset is represented by one observation, and each observation consists of a set of features.

3.3 Conceptual Model

3.3.1 Definition of Clinical Decision Support System for Early Diagnosis of Dementia (CDSS-DD)

CDSS_DD can be defined as follows:

CDSS_DD is a computerised tool for diagnosing dementia, via automated analysis and evaluation of the CDT. The system can expedite the assessment of cognitive abilities by reducing the labor required to interpret the produced drawings in CDT.

From a research prospective, it is defined as follows:

CDSS_DD is a machine learning algorithm which consists of several stages. To build this system, medical research is conducted to make decisions about the characteristics, and the outcome of the system. It is also built to be a platform of more medical analysis and provide additional dimensions in understanding the relation between the cognitive impairment and the defect in the produced clock. In this algorithm the CDT drawings are digitised, whereby each one is converted into an observation represented by a set of features. The algorithm learns from the available diagnosed data to classify (diagnose) the undiagnosed drawings.

Figure 3-5 shows the conceptual model of the CDSS-DD. It includes two phases: firstly the training phase; and secondly the diagnosis phase. Each phase consist of five steps to perform the following tasks: image enhancement; feature extraction; discretisation of continuous features; feature selection; and diagnostic stage. feature extraction is explained in chapter 4, feature selection is described in chapters 5, and 6, and d diagnosis stage is introduced in chapter 7.

The process is begun by the collection of data from patients' records in the hospital. The drawings are then scanned and stored as grey-scale images before being provided to the proposed system. During the first phase the diagnosis stage is trained using the available classified (diagnosed) drawings. Discretisation is performed during the training phase, and the same produced bins are used to discretise features during the diagnosis phase. The same process is repeated during the feature selection stage. The importance of the features is found during the training phase, and the selected features are used in the diagnosis phase. The process is performed to maintain consistency between the training data and testing data for the prediction stage. The five stages are described in the following sections.

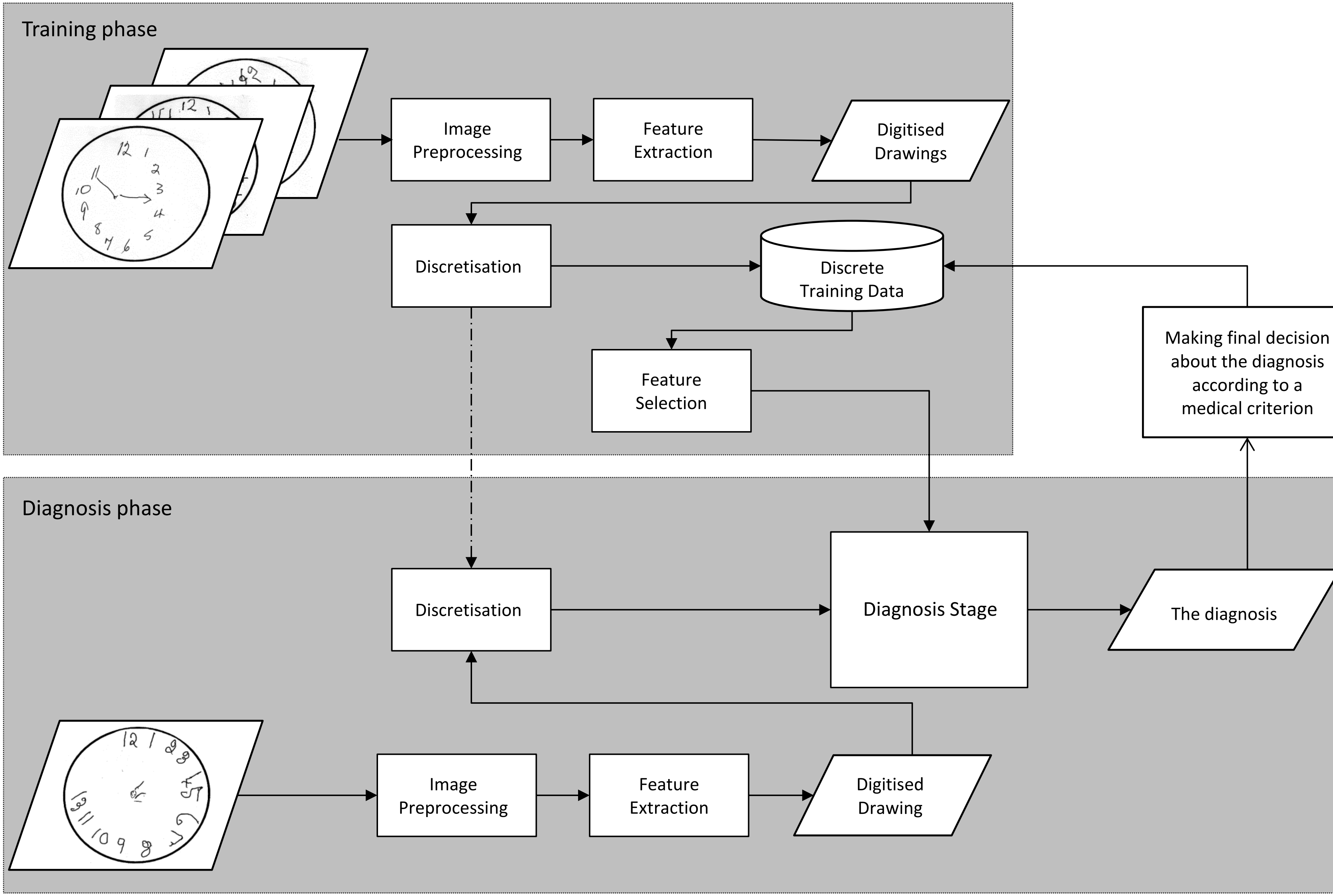


Figure 3-5 Conceptual Model of the CDSS-DD.

3.3.2 **Clock Image Enhancement**

Most of the drawings have been produced more than 10 years ago, and hence the quality of many of them has been degraded due to aging and multiple photocopying of the original drawings. The noise in the background is high in some drawings, and there are other problems such as disconnected numbers and hands, and low intensity of clock components. Figure 3-6 shows some examples of image quality problems.

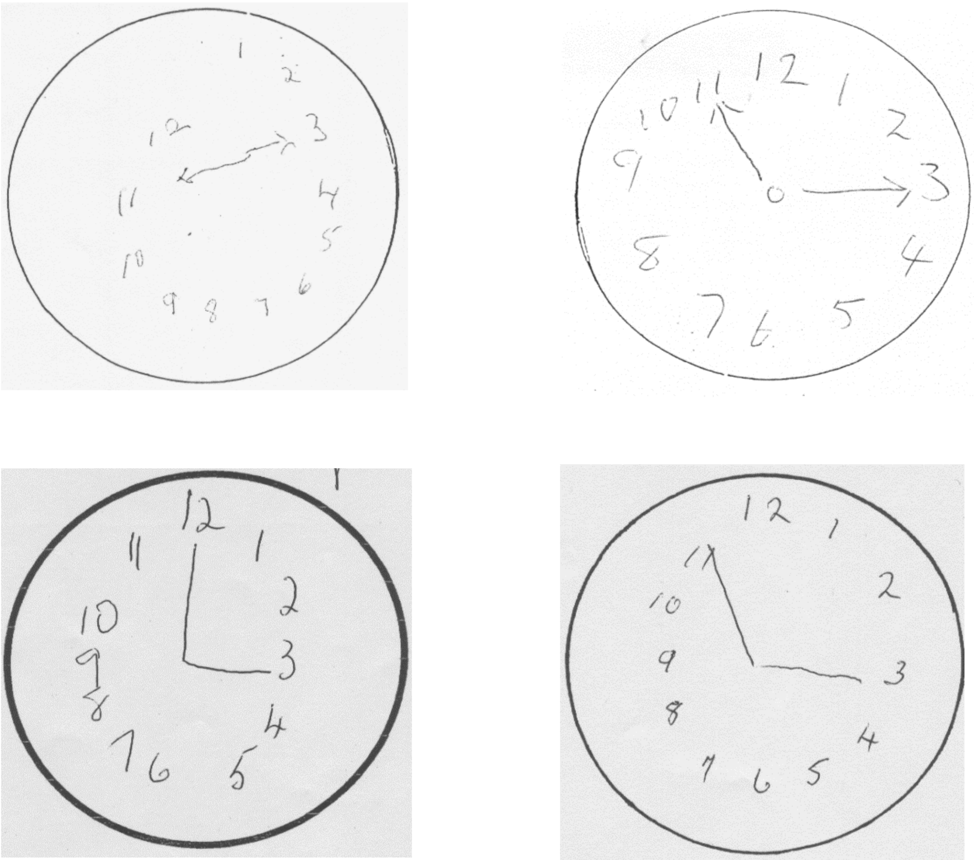


Figure 3-6: Examples of image quality problems.

The drawing enhancement is performed in several steps starting with filtering, then thresholding, and finally processing by morphological techniques to connect the disconnected numbers and hands. The following block diagram describes this process flow.

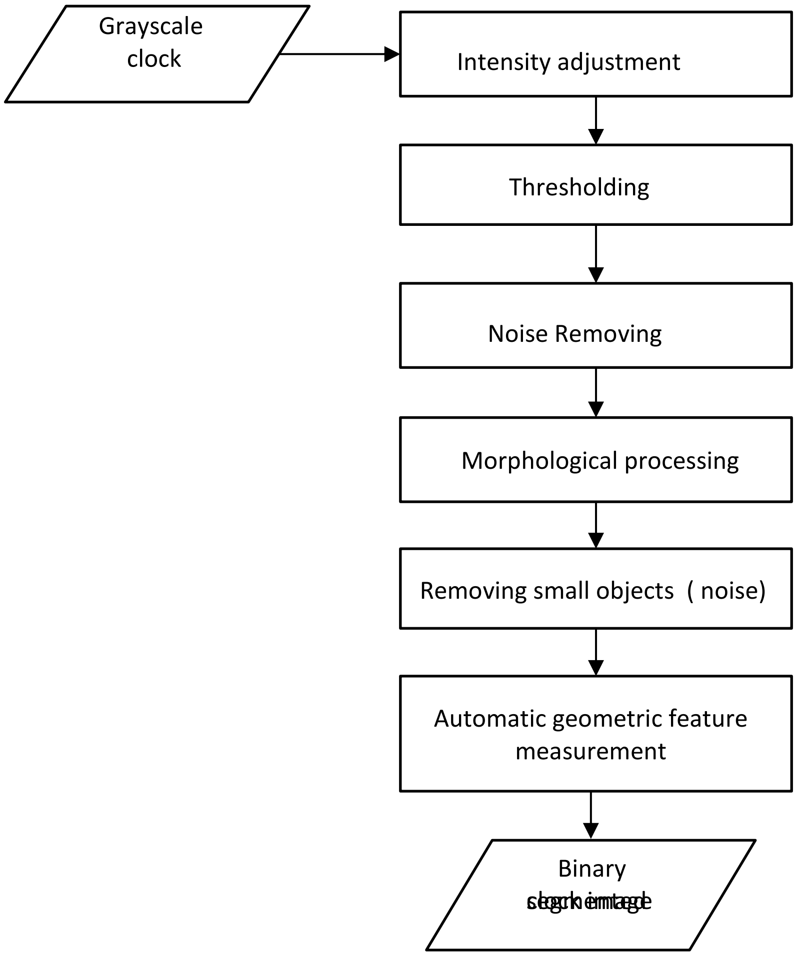


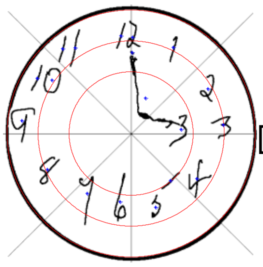
Figure 3-7: Block diagram of clock image enhancement stage.

3.3.3 **Feature Extraction**

This stage is needed to produce a digitised CDT dataset, which forms objective number two. At this stage a comprehensive list of 47 different types of qualitative visual feature is extracted from the drawing images. This list is based on a review of the most common CDT scoring systems in the medical literature, and a comprehensive analysis of the CDT data. The process is not fully automated, with some of the features being measured manually.

The enhanced clock images are processed further to locate the centroid of each clock number. The clock circle is grammatically divided into three parts after two new circles are drawn to capture new features.

The clock is divided into four quadrants, and divided further into eight sectors. These divisions are used to capture the deficit in spacing among the clock. Figure 3-8 shows the process of extracting the features and digitising the CDT drawings. This stage is discussed in detail in chapter 4.



drawings	F1	F2	F3	...	...	...	...	F47
Clock1	12	0	0					3
Clock2	3	3	3					-1
. .								
. .								
Clock648	12	0	0					1.5

Figure 3-8: The process of extracting the clock features.

3.3.4 **Discretisation**

Discretisation is a process that transforms continuous data into discrete data attributes without significant loss of information so that when ‘many to one’ transformation is performed, each value in the original data is mapped onto a new discrete value.

After extracting the features in the previous stage, some of the features are continuous, and some of them are discrete. Since one of the objectives is defining the most significant features within the clock, all the features should be discrete in order to avoid any bias during the feature selection stage (Meyer and Bontempi, 2006).

3.3.5 **Feature Selection**

Feature selection is a powerful tool for reducing the dimensionality of datasets by selecting the most informative features with the greatest discriminatory power. Feature selection is normally used at the pre-processing stage before training a classifier. Feature selection does not alter the data so it is the best choice when data interpretation is essential and it is the preferred choice for studying the clock features.

The novel filter method which is developed as part of this research, is used in this system. This adds the significant features one by one to the selected subset provided each feature is not redundant to any features already selected, or irrelevant to the class lable.

This stage is used in the system for two reasons: firstly to enhance the performance of the diagnostic stage; and secondly to study in depth the significance of the clock features when the CDT is used to distinguish between different degrees of cognitive impairment. The proposed novel feature selection method and the important features in the clock are presented in chapter 5 and chapter 6.

3.3.6 **Diagnosis Stage**

This is the final stage, in which the system makes decisions about the diagnosis. The prediction stage is built based on a hierarchical classifier, with each classifier trained on the training data. The new, undiagnosed drawing is submitted to the diagnostic stage, within which the new drawings are classified into one of the classes (diagnoses) within the training data.

The cascade classifier consists of three classifiers, which are connected sequentially. Figure 3-9 shows the block diagram of this system in which the classification is performed in two stages of cascade classification. CDT drawings are classified into one of the four classes (diagnoses): (I) Normal, (II) Functional, (III) MCI+, (IV) Severe and moderate dementia. This stage is presented in detail in chapter 7.

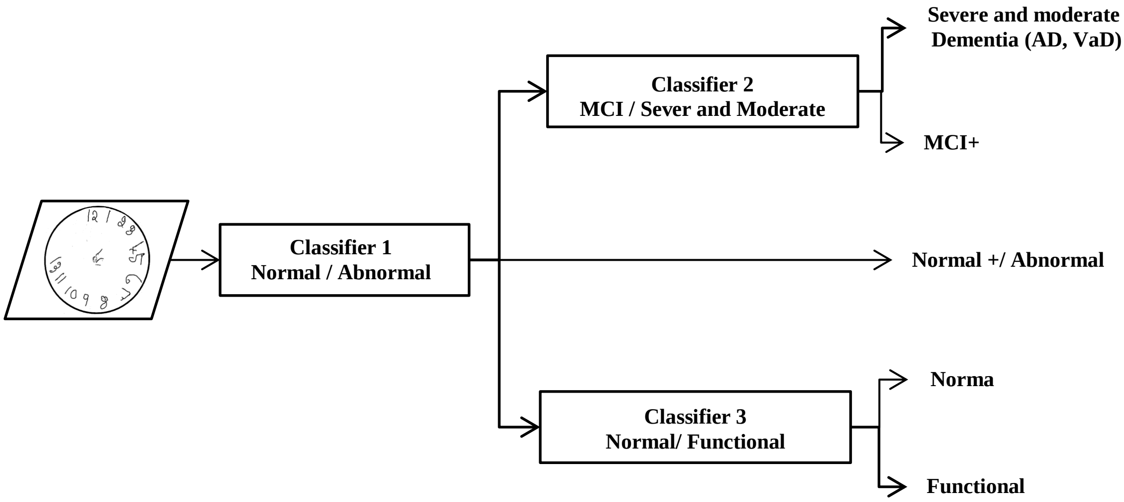


Figure 3-9: Block diagram of the prediction stage.

3.4 Summary

This chapter has discussed the CDT data that is employed in this research. The new conceptual model of the proposed CDSS-DD system is also introduced in this chapter. The chapter proposed a new conceptual model that can facilitate the analysis of CDT, and examine the abilities of the test to differentiate between different levels of severity of dementia. In contrast to the previous research the proposed conceptual model employs sophisticated machine learning techniques to analyse the importance of the produced clock features for dementia diagnosis, and also to diagnose the CDT drawings, not using a scoring system criteria. The next chapter discusses the extraction of the clock features and outlines the conducted comparative study between clock scoring systems.

Chapter: 4

CDT Feature Extraction

The development of an automated system for the diagnosis of dementia based on CDT requires a digitised CDT dataset for the training and validation of the proposed system. When the present research took place, there was no electronic CDT dataset available. For this reason the available 'hard-copy' CDT drawings are scanned and a number of important visual features are extracted from the digitised drawings using a combination of automatic image processing techniques and manual measurement. This chapter presents the newly proposed comprehensive list of important visual features which can exist within clock drawings, and describes in detail the process of image enhancement and feature extraction. The output of this process is a fully digitised list of 47 features extracted from 604 scanned drawings.

The contribution of this chapter is two-fold:

1. A comprehensive catalogue of 47 CDT image features is produced. The list of the features is constructed based on an in-depth review of previous research in the area of dementia assessment. It also includes new geometric features which have never been used for CDT analysis before. These features are proposed based on the analysis of the CDT drawings that are available for this research.

2. A new, digitised CDT dataset is produced, which consists of 604 drawings produced by patients with five different cognitive diagnoses. This dataset is according to the catalogue of 47 CDT features, and can facilitate the validation of any proposed CDT-based diagnostic system. It also enables further research into, and analysis of, the capabilities of CDTs.

This chapter is organised as follows: Section 4.1 presents a comparative study of the four most effective CDT scoring systems reported in previous literature; Section 4.2 describes the clock image enhancement process; Section 4.3 presents the comprehensive list of features used, explaining how the new proposed features are extracted and the typical values they can take; Finally, section 4.4 summarises this chapter.

4.1 Comparative Study of the Best Scoring Systems

It is noted in Chapter 2 that many scoring systems have been proposed in the literature. Four of these systems are reported to produce good sensitivity, specificity, and diagnostic accuracy (Aprahamian et al., 2009; Pinto and Peters, 2009). These systems are proposed by (i) Shulman et al. (1986), (ii) Sunderland et al. (1989), (iii) Mendez et al. (1992), and (iv) Tuokko et al. (1992). Each system will be referred to in the rest of this thesis by the primary author’s name.

A comparative study is conducted using a sample of the CDT data. The sample consisted of cases which are deemed not to be severe, cases of very mild dementia, and MCI cases. The diagnosis of the latter two types is very challenging. These are considered to be the early stages of cognitive disorder, and so although it is difficult to diagnose disorder in these stages, it is critical that the symptoms are detected early.

To test the effectiveness of each system in diagnosing positive dementia cases, a comparative study is conducted and the four aforementioned scoring systems are used to analyse 61 clock drawings to provide diagnoses. The sample contained CDTs from patients with four different diagnoses (AD, VaD, MCI and Normal). The ages of participants ranged from 35 to 93 years old. 35 of the drawings are produced by males and 26 by females. Figure 4.1 shows the distribution of the final diagnoses of these samples.

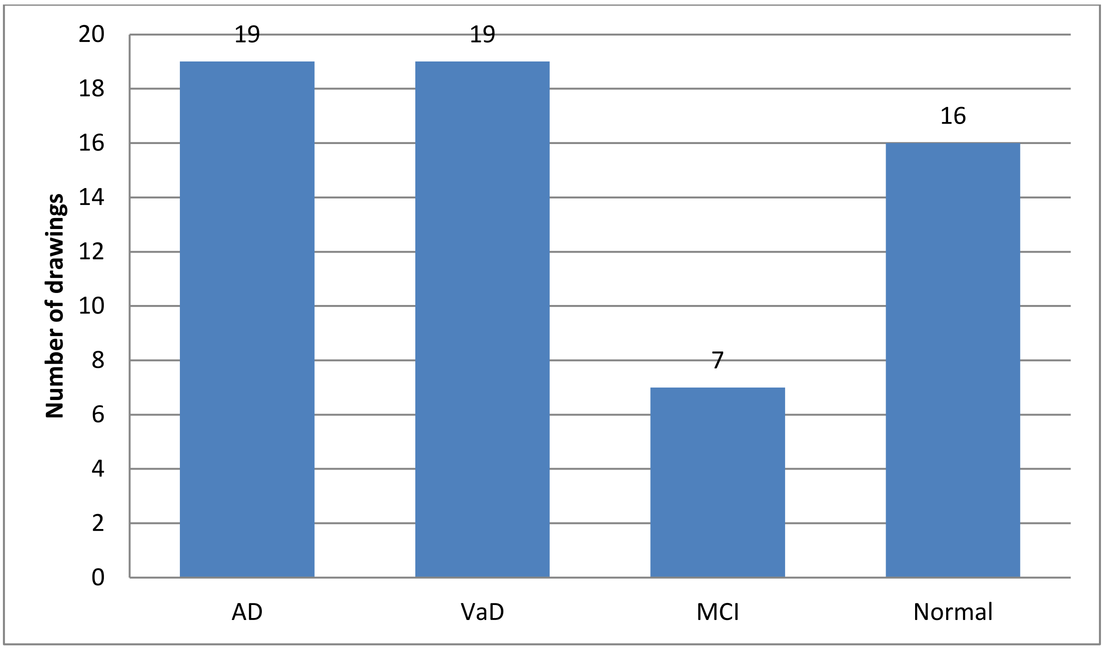


Figure 4-1: Distribution of diagnoses from the drawings used in the comparative study.

The selection of the positive cases is based on the score of the MMSE; the score is between 15 and 29. The scores relating to some of these drawings are well within the normal range, producing an average score of 22.5. For this reason, the majority of the drawings used in this comparative study presented a challenge, since many of them displayed only subtle errors.

4.1.1 **Results**

Table 4.1 presents the results of this study, showing that the Tuokko scoring system produce the most reliable performance. The accuracy of this system (percentage of drawings diagnosed correctly) is 65.57 %, while the sensitivity (ability of the system to diagnose the positive cases) is 57.78 %, and finally the specificity (the accuracy of distinguishing the negative cases) is 87.50 %.

The other scoring systems produce a poorer performance as shown in Table 4.1. The accuracy of the other systems is between 36.07 % and 44.26 %, while the sensitivity is between 13.33 % and 26.67 %, and the specificity is between 93.75 % and 100 %.

Table 4-1: Results of the comparative study.

	Shulman	Sunderland	Mendez	Tuokko
Correctly Identified	42.62 %	36.07 %	44.26 %	65.57 %
Sensitivity	24.44 %	13.33 %	26.67 %	57.78 %
Specificity	93.75 %	100.00 %	93.75 %	87.50 %

Table 4.2 presents the percentage of correct diagnoses for each of the four scoring systems, within each of the cognitive impairment categories. While all the systems are able to recognise the majority of the ‘normal’ cases, Sunderland’s system is the only one to achieve 100 % accuracy in distinguishing these cases. For the abnormal cases with dementia, the table shows that all the scoring systems achieve their best diagnostic accuracy when diagnosing AD. However, this accuracy is still below 50 %, except for Tuokko’s system which predicts the AD cases with an accuracy of 63.16 %.

The results also show that all the scoring systems produce low accuracy in diagnosing VaD and MCI cases. Two systems did not diagnose any of the patients suffering from the MCI disease.

Table 4-2: Performance of the scoring systems in correctly diagnosing each cognitive impairment.

Diagnosis	Shulman	Sunderland	Mendez	Tuokko
AD	42.11 %	26.32 %	36.84 %	63.16 %
VaD	10.53 %	5.26 %	26.32 %	31.58 %
MCI	14.29 %	0.00 %	0.00 %	42.86 %
Normal	93.75 %	100.00 %	93.75 %	87.50 %

4.1.2 **Discussion**

The aim of this study is to test the robustness of the scoring systems that are reported as being reliable in the literature. The study shows that Tuokko's system is superior to the other systems in identifying the positive dementia cases because it produces the best sensitivity and the best trade-off between sensitivity and specificity. The study also shows that Sunderland's system produces the worst accuracy in identifying the positive dementia cases. However, none of the four scoring systems produced high diagnostic accuracy.

These results are far inferior to the results reported by the developers of these scoring systems. In the literature, Shulman's system is reported to produce a sensitivity of 86 %, and specificity of 72 %. The sensitivity Sunderland's system is reported as 76 % and the specificity as 81 %, while Mendez' system sensitivity is reported as 73 % and the specificity as 77 %. Toukko's system is reported to produce the highest diagnostic accuracy among the four techniques, with a sensitivity of 92 % and a specificity of 86 %. The reason for the difference between the published results and the results of this study can be attributed to the data sample used. As noted earlier, most of the chosen samples are CDT drawings produced by MCI and mild dementia patients. The early diagnosis of these cognitive impairment stages is very important, as it allows medical interventions to slow the progress of the disease and treat the causes. However, it is very challenging to perform a diagnosis in these circumstances based on CDT alone.

The results of this study agree with the results of the comparative studies presented elsewhere, which concluded that the available CDT systems are not capable of diagnosing MCI and early stage dementia in the majority of cases.

In conclusion, the results obtained from the study show that changing the cut-off point of the scoring systems may improve the performance in diagnosing the cases of MCI and early stage dementia. Moreover, extracting new detailed CDT features may increase the robustness of the test in diagnosing the challenging cases. This is because the new features could reveal more information about clock drawing errors which are specific to the early stages of dementia.

4.2 **Clock Image Enhancement**

The Image Processing Toolbox in Matlab is used in this stage to perform all the image processing operations. This stage consists of 5 steps. In each step the clock drawing is subjected to a set of image processing operations. The five processing steps fall into two groups:

I. Image pre-processing which includes: (i) intensity adjustment; (ii) thresholding; (iii) cleaning and noise removal; and (iv) morphological operations to fill the gaps and connect disconnected objects.

II. Image analysis (v), where the clock circle is divided into sub-regions in order to capture defects in the spacing, and to find the centroid of the objects within the clock image.

Figure 4.2 shows the effect of the above image enhancement operations on an example CDT image.

4.3 **Clock Feature Extraction**

To enhance the sensitivity of the CDT in diagnosing the early symptoms of dementia, a suite of new features are extracted in addition to the qualitative features of the common scoring systems. A comprehensive analysis of the CDT data is performed to identify new errors in the clock drawings which have not been covered in existing scoring systems. These additional features are included in the feature list used in this research. As a result a list of 47 visual features is established, as shown in Table 4.3. The list includes the majority of the features employed in the existing scoring systems, along with an additional new geometric features.

The features can be divided into three groups: (i) features related to the clock numbers; (ii) features related to the clock hands, the time setting; (iii) miscellaneous features which are not included in the first three groups; and (iv) features related to the clock center. The features are each explained in the Appendix A. only the new geometrical features are described in the following section along with the criteria for their measurement.

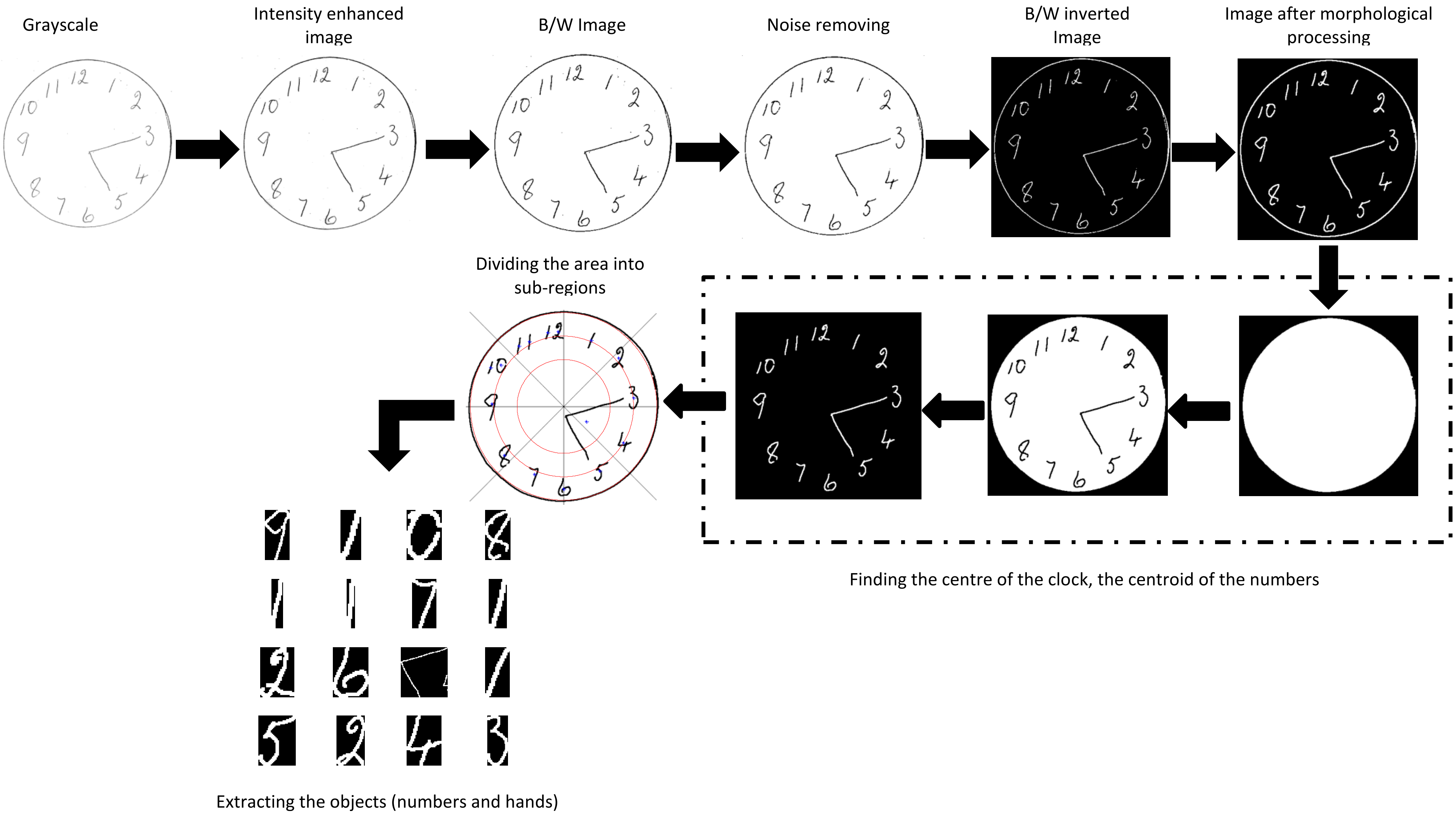


Figure 4-2: Steps of clock image enhancement and preparation for feature extraction.

Table 4-3: List of clock features.

No	Feature	Data type
1	Count of numbers within area 1.	Discrete
2	Count of numbers within area 2.	Discrete
3	Count of numbers within area 3.	Discrete
4	Count of numbers within quadrant 1.	Discrete
5	Count of numbers within quadrant 2.	Discrete
6	Count of numbers within quadrant 3.	Discrete
7	Count of numbers within quadrant 4.	Discrete
8	Minimum size of the numbers mm^2.	Continuous
9	Maximum size of the numbers mm^2.	Continuous
10	Ratio between the maximum number size and minimum size.	Continuous
11	Count of numbers outside the contour.	Discrete
12	Minimum angle between numbers.	Continuous
13	Maximum angles between numbers.	Continuous
14	Count of numbers whose rotation is over 25 degrees.	Discrete
15	Count of numbers left out from the drawing.	Discrete
16	Count of duplicated numbers.	Discrete
17	Sequential numbers are written following 12 (e.g. 13, 14, 15)	Binary
18	Numbers not in sequence.	Binary
19	Numbers 3 and 11 not present.	Binary
20	Arabic only numbers used.	Binary
21	Direction of written numbers.	Binary
22	Self-correction of numbers.	Binary
23	Minute hand is present.	Binary
24	Hour hand is present.	Binary
25	More than two hands are drawn.	Binary
26	Self-correction of hands.	Binary
27	Time is correct.	Discrete
28	Time is indicated by writing minute number next to 3 or next to 11.	Binary
29	Straight line is used between the two numbers.	Binary
30	Displacement of hour hand or mark from the target number.	Discrete
31	Displacement of minute hand or mark from the target number.	Discrete
32	Hands connected with target number.	Discrete
33	Arrows on hands.	Discrete
34	Displacement of arrows less than 4 mm.	Discrete
35	Arrows are pointing in the wrong direction.	Discrete
36	Presence of superfluous.	Binary
37	Hands are joint or within 12 mm.	Discrete
38	Position of minute hand.	Discrete
39	Position of hour hand.	Discrete
40	Angle between clock hands.	Continuous
41	Ratio between hands.	Discrete
42	Presence of stem of clock hands (near to the center) is left out.	Discrete
43	Time is written across the clock.	Binary
44	Time is written outside the clock.	Binary
45	Picture of a human face is drawn on clock.	Binary
46	Presence of written words.	Binary
47	Distance between the position of hands intersection and the center of the clock	Continuous

4.3.1 **New Geometric CDT features:**

These include features derived from the clock numbers, namely the spacing, position, size, orientation, position of clock hands, and angle between clock hands.

Features 1, 2, and 3 (count of numbers within areas 1, 2, and 3)

Geometric data analysis has shown that individuals who suffer from dementia have a tendency to write the numbers far from the outer contour of the clock (Freedman et al., 1994). To capture this behaviour, two new circles are drawn to divide the area of the clock into three parts (figure 4.3). As shown in figure 4.2, the first new circle is drawn with a diameter of $0.75D$, and the second one with a diameter of $0.5D$, where D is the diameter of the clock. The count of the numbers in the areas 1, 2, and 3 are used as the value of the feature 1, 2, and 3 respectively.

Any numbers whose centroid is located precisely on any of the circular boundaries are counted with the outer area of that circle. In the case of the normal drawings all 12 numbers are likely to be written in area 1 (the outer area).

Some scoring systems have taken into account the writing of numbers away from the clock contour (Mendez, and Tuokko), although only qualitatively. The criteria proposed here have never been used before. Figure 4.3 shows examples of measuring features 1, 2, and 3.

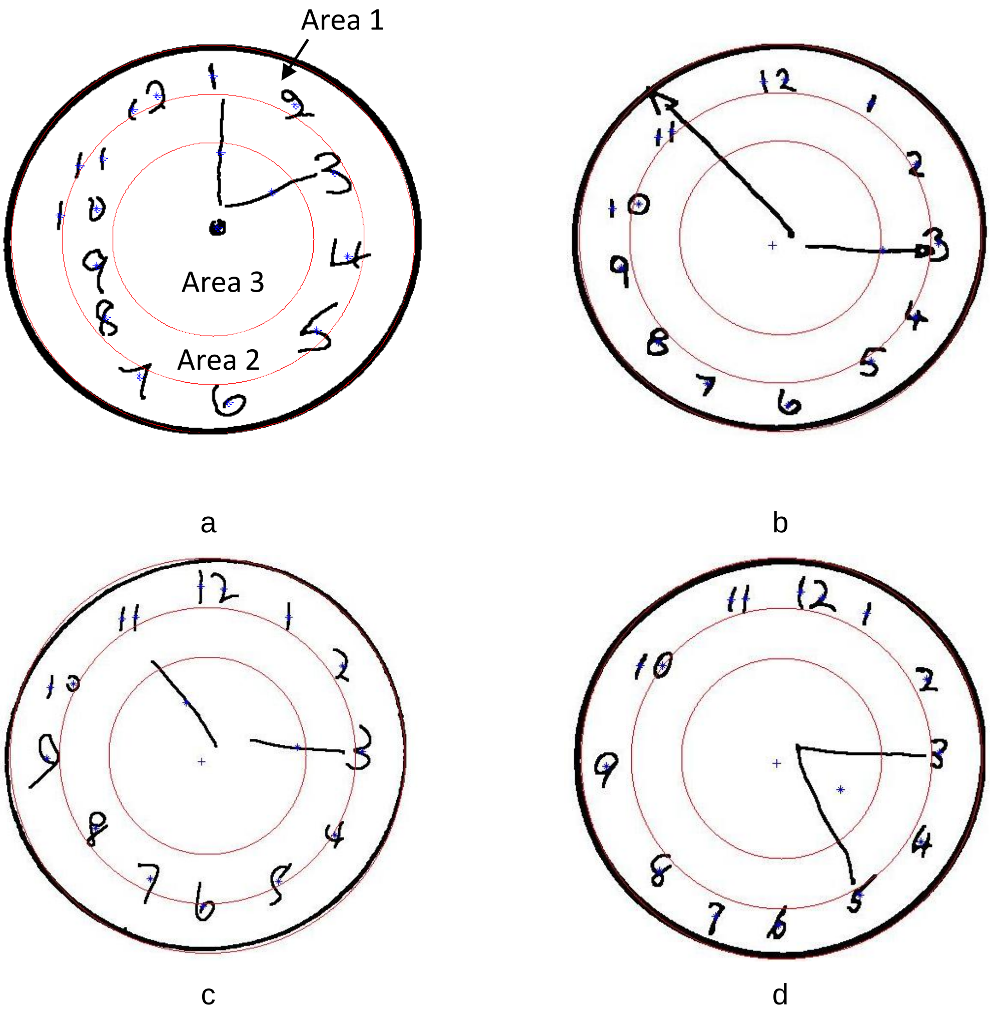


Figure 4-3: Examples of measuring features 1, 2, and 3: (a) Drawing by an 81 year old male VaD patient. The values of features 1, 2, and 3 are 5, 7, and 0 respectively; (b) Drawing by a 57 year old healthy female. The values of features 1, 2, and 3 are 12, 0, and 0 respectively; (c) Drawing by an 80 year old male MCI patient. The values of features 1, 2, and 3 are 9, 3, and 0 respectively; (d) Drawing by an 88 year old male AD patient. The values of features 1, 2, and 3 are 12, 0, and 0 respectively.

Features 8, 9, and 10 (related to the size of numbers)

The analysis of the data showed that some dementia patients have a tendency to write numbers of unusually large size. Therefore, three new features are proposed to capture this deficit. Figure 4.4 shows several examples of clock drawings by dementia patients in which large numbers are used.

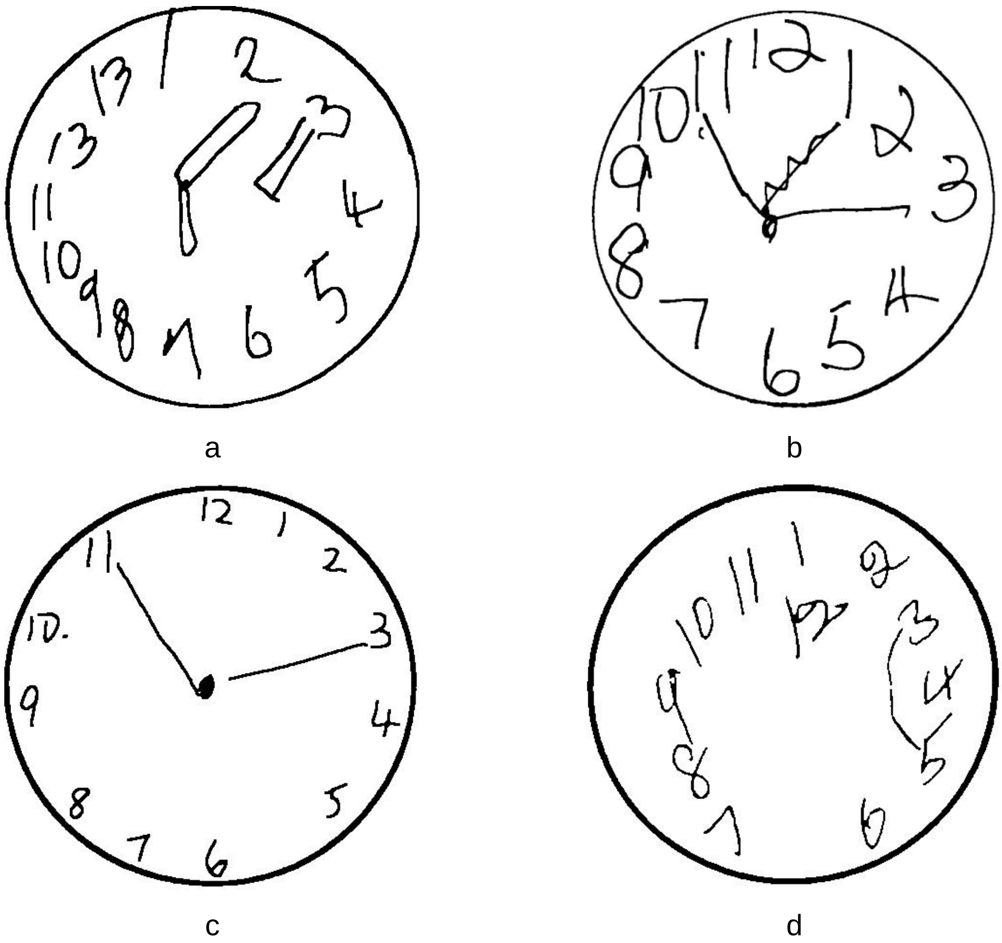


Figure 4-4: Examples of number size in clock drawings: (a) Drawing by a 91 year old female AD patient; (b) Drawing by an 82 year old female VaD patient; (c) Drawing by a 64 year old healthy individual, (d) Drawing by a 73 year old female VaD patient.

The size of each number is calculated by measuring the area that it occupies in the clock face (in mm^2). The area is calculated as shown in Figure 4.5. The minimum number size and maximum number size are defined as features 8 and 9 respectively. The ratio between the maximum and the minimum is also calculated to represent the variation in number size. This is defined as feature 10.

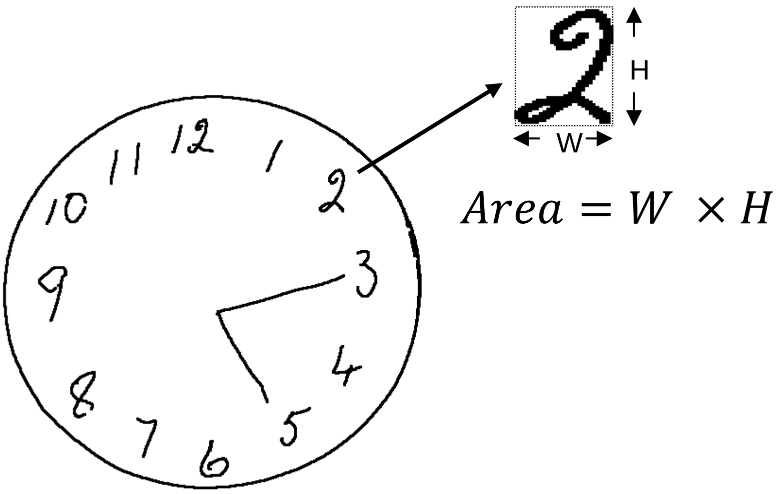


Figure 4-5: Measuring the area of the number.

Features 12 and 13 (the angle between numbers)

These features are used to capture the deficit in the spacing of numbers. To achieve this, the angular separation between each two consecutive numbers is quantified by connecting the centroid of each number to the clock face center, and measuring the angle between each two adjacent lines. Dementia patients have a tendency to struggle with distributing the numbers evenly around the clock face, causing the angles between numbers to be very small and/or very large. The maximum angle and the minimum angle are used for features 12 and 13. Figure 4.6 shows how the angles are measured: the maximum angle of this clock is 39°, which corresponds to the angle between the numbers 8 and 9 or between 3 and 4, while the minimum angle is 24°, which is the angle between numbers 6 and 7. These two features are continuous.

Feature 14 (the count of numbers whose rotation is over 25 degrees)

This is the count of digits whose rotation angle in any direction is greater than 25°. This can occur because dementia patients sometimes rotate the paper while drawing the clock, and as a result the orientation of the numbers deviates from the horizontal axis. A similar feature is used in the Tuokko scoring system but with an angle threshold of 45°. The Freedman scoring system also proposes a qualitative feature related with number rotation. The angle of 25° is chosen after preliminary analysis of the data, and allows the feature to be more sensitive to the early symptoms of dementia.

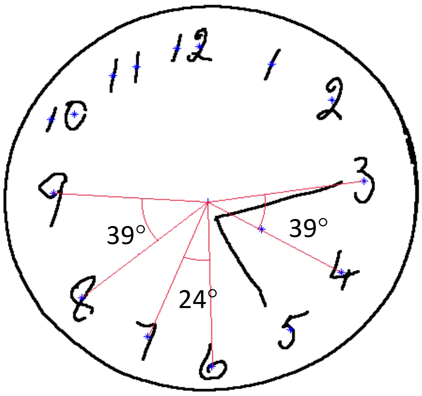


Figure 4-6: The angle between numbers.

Figure 4.7 shows how the rotation is measured. Here a clock is drawn by a 77 year old female individual with functional problems, and the rotation of some numbers (8, 9, 10, and 11) is greater than 25°. Therefore the value taken for this drawing is 4.

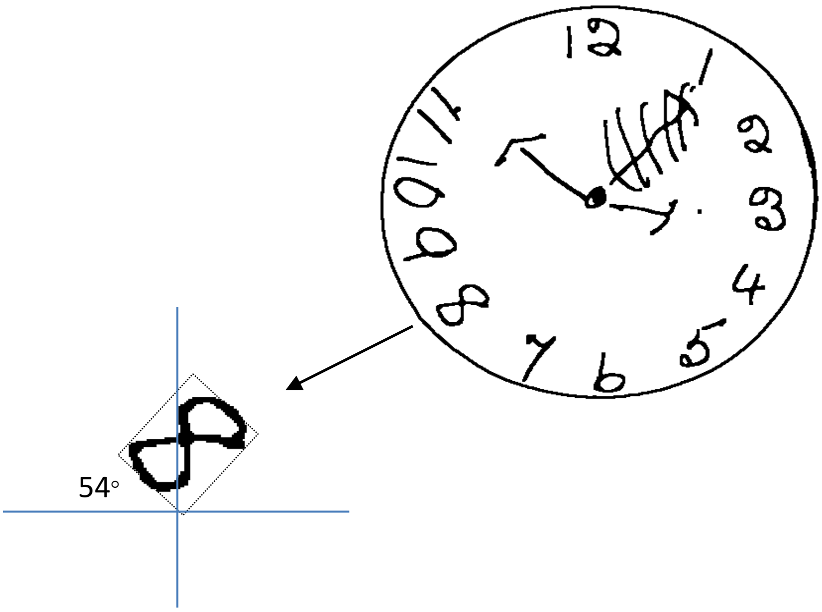


Figure 4-7: Measuring the rotation angle of clock numbers.

Feature 26 (self-correction of hands)

This is a discrete feature. It is proposed based on the preliminary analysis of the available data. The value is "1" if there is self-correction, "0" if there is no self-correction, and "-1" if the hands are missed.

Feature 32 (hands connected to the target number)

This is a discrete feature, which can take five different values: "1" if the minute hand only is connected to its target number; "2" if the hour hand only is connected to its target number; "3" if both are connected to their target numbers; "0" if neither of the hands is connected; and finally "-1" if the hands are missed.

Features 38, and 39 (position of clock hands)

Two features are proposed to capture the deficit in the position of the clock hands. They are discrete features. The area of the clock is divided into eight sectors as shown in Figure 4.8. The number of the sector in which each hand is located is used as a feature. In the normal case the position of the minute hand is in eight 1 (E1) and the hour hand is in eight 3 (E3). These two features are set to "0" if the hands are missing.

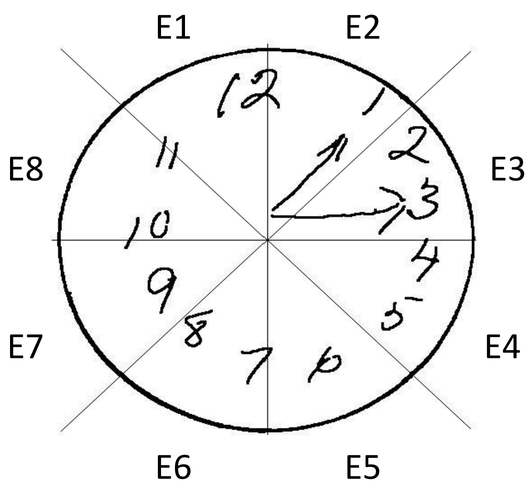


Figure 4-8: Clock divided into eight sectors to find the position of each hand. This drawing is by an 89 year old female AD patient.

Feature 40 (angle between clock hands)

The angle between the hands (if the hands are present) is measured and used as a feature. The value is equal to the angle between the hands starting from the hour hand (the hand closest to the number “3” is considered as the hour hand). If the hands are not connected they are extended until they intersect, and if the hands are not straight then new best fit lines are drawn. Figure 4.9 shows examples of measuring this quantity. The value is set to “0” when one or both hands are missed.

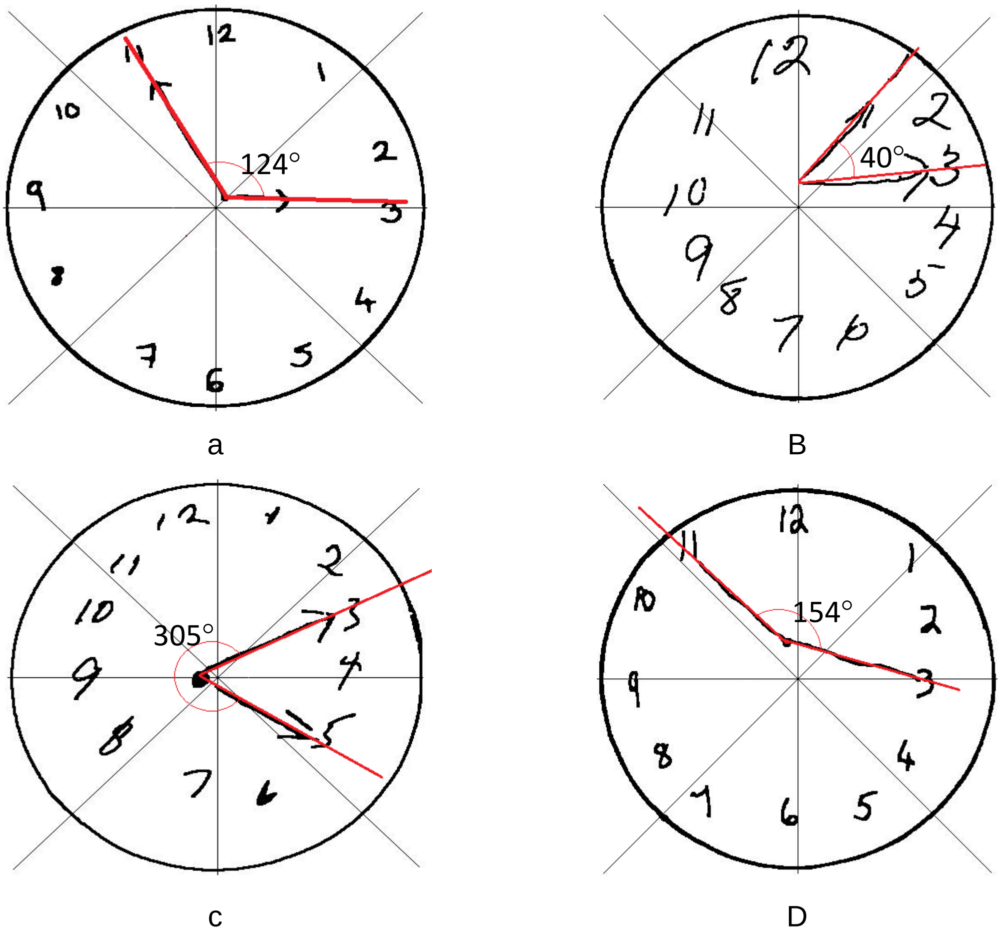


Figure 4-9: Examples of measuring the angle between the hands: (a) drawing by a 22 year old normal female; (b) drawing by an 89 year old female AD patient; (c) drawing by an 86 year old male AD patient; (d) drawing by an 82 year old female MCI patient.

Feature 47 (distance between the position of hands intersection and the center of the clock)

It is proposed based on the preliminary analysis of the data. The normative data shows that the healthy individual is likely to start drawing the hands from a point close to the clock face center, while dementia patients start drawing from a point away from the center. The distance between the center of the clock and the intersection of the clock hands is measured in mm. In the case when the hands are not connected they are extended until they intersect. The distance from the center to the point of intersection is measured and used as a feature. Figure 4.10 shows examples of measuring this.

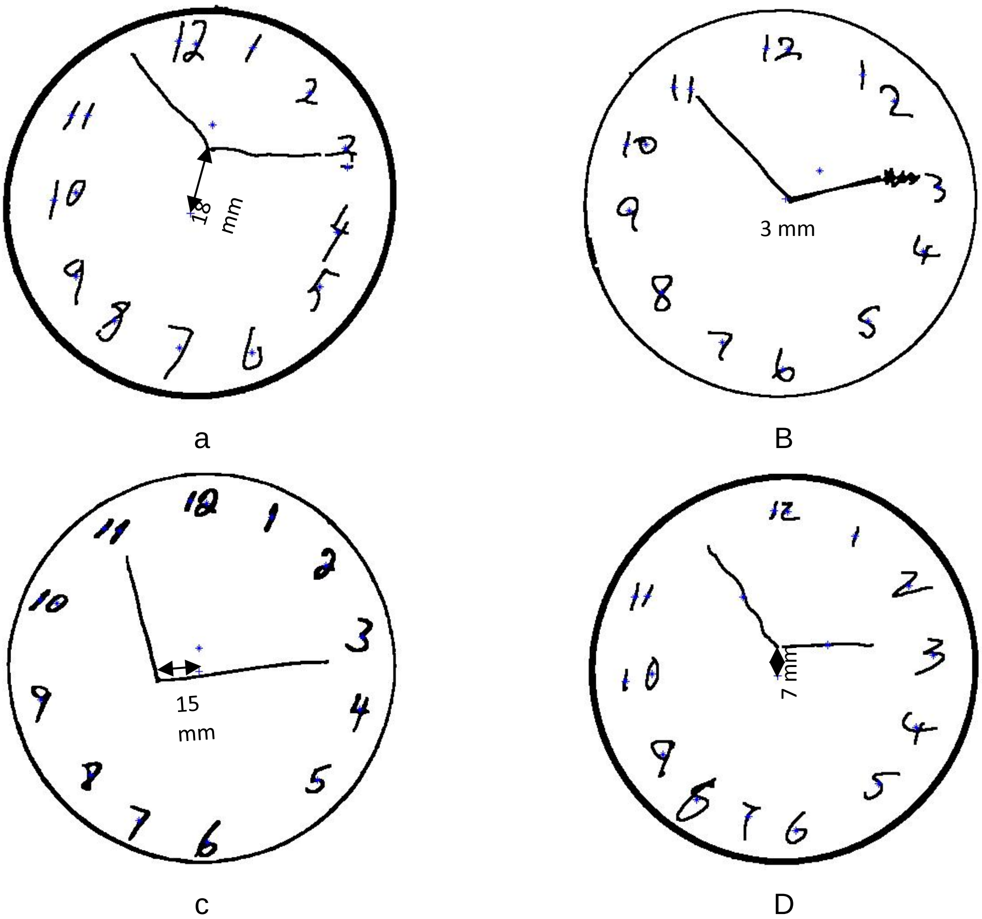


Figure 4-10: Examples of measuring the distance between the center of the clock and the intersection of the hands: (a) drawing by a 77 year old male VaD patient; (b) drawing by a 63 year old healthy male; (c) drawing by 78 year old male with functional problems; (d) drawing by 65 year old male MCI patient.

4.4 Discretisation Stage

Seven of the extracted 47 features are continuous and the rest are binary or discrete. Since one of the objectives of this research is to define the significant clock features to the discriminative task, feature selection techniques will be employed. Some techniques are vulnerable to the continuous features, and so a discretisation method is needed to convert these features into discrete variables. The MDL discretisation method (Fayyad, and Irani, 1993) has been reported in the literature as one of the best discretisation methods (Dougherty et al 1995; Liu et al., 2002). This method is employed to discretise the continuous features whenever it is applicable. MDL method tends to remove features when it fails to find appropriate cut-off points to discretise them. In the instances where MDL is not applicable, EWD and EFD are used. Although EWD and EFD are simple and easy to implement they are reported to produce a good performance and even close to more sophisticated methods (Dougherty et al 1995).

4.5 Summary

This chapter presented a comparative study to test the robustness of the best CDT scoring systems reported in literature for the diagnosis of MCI and early stage dementia. The study shows a superior performance by the Tuokko scoring system, with an accuracy of 65.57 %. However, none of the four scoring systems produced sufficient performance in diagnosing the positive cases. This indicates the need to include more detailed features to capture the defects in the clock drawings which are indicative of early stage dementia. Increasing the sensitivity of some features may also be necessary in order to enhance the performance of the CDT in diagnosing the early stage dementia and MCI cases.

In response a new comprehensive catalogue of CDT image features is introduced. 47 features are presented along with the meaning of each feature, the mechanism of measurement, and the values that each feature may take. These features are used to digitise the clock drawings to provide a novel CDT dataset. In this chapter also the clock image enhancement stage is explained with an illustrative example. In the next chapter two new feature selection methods are proposed and validated.

Chapter: 5

Feature Selection Based on Joint Mutual Information

Although a large variety of clock features have been used previously to test for a range of disorders, at the time of this research there is comparatively little research into the significance of clock features in dementia assessment. As outlined in Chapter 2, only two such studies have been conducted in the literature. However, using a sophisticated statistical method to determine the significance of features can reveal new information about the relationship between the clock features and the stages of dementia. It can also provide medical practitioners with a more relevant list of features to focus their attention on during the dementia assessment.

The feature selection technique is chosen because it is the preferred choice when an understanding of the underlying physical process and data interpretation is desired.

The main contribution of this chapter is two new methods of feature selection which are developed based on information theory. The proposed methods employ joint mutual information, and use the ‘maximum of the minimum’ approach and forward greedy search algorithm. The proposed methods aim to address the problem of overestimating the significance of some features, which occurs in the Joint Mutual Information (JMI) and Double Input Symmetrical Relevance (DISR) methods when a cumulative summation approximation is employed. The new methods are evaluated and compared with the current state of the art feature selection methods.

This chapter is organised as follows; Section 5.1 introduces the proposed methods; Section 5.2 presents the evaluation of the proposed methods; Section 5.3 presents the stability performance of the methods; Section 5.4 discusses the results; Finally, section 5.4 concludes and summarises the chapter.

5.1 Proposed Methods for Feature Selection

In this chapter, two new methods for feature selection are proposed. The methods employ joint mutual information, and use the 'maximum of the minimum' approach. The proposed methods aim to address the problem of overestimating the significance of some features, which occurs when cumulative summation approximation is employed.

For a feature set $F = \{f_1, f_2, \ldots\ldots, f_N\}$ of a data set D of dimension N, the feature selection process identifies a subset of features S with dimension K where $K \leq N$, and $S \subseteq F$. The subset S should produce equal or better classification accuracy compared to feature set F. In other words feature selection defines the subset of features that maximises mutual information with the class label $I(S, C)$.

In the past, a number of alternative definitions of feature relevance have been used (Battiti, 1994; Estevez et al. 2009; Brown et al. 2012) In this work, the following definition is used.

Definition 1 (Feature relevance). Feature f_i is more relevant to the class label C than feature f_j in the context of the already selected subset S when $I(f_i, S; C) > I(f_j, S; C)$.

Definition 2 (Minimum joint mutual information): Let F be the full set of features, and let S be the subset of features that are selected already. Let $f_i \in F - S$, and $f_s \in S$. The m-Joint MI is the minimum value of joint mutual information that the candidate feature f_i shares with the class label C when it is joined with every feature within the subset S individually, hence $\min_{s=1,2,\ldots,k} I(f_i, f_s; C)$,

Lemma 1. For a feature f_i , if the m-Joint MI is larger than that of all other features f_j, where f_i and $f_j \in \mathrm{F}-\mathrm{S}$ $(\mathrm{i} \neq \mathrm{j})$, then it is the most relevant feature to the class label C in the context of the subset S.

Proof: Let $S = \{f_1, f_2, \dots\dots, f_K\}$. The joint mutual information of f_i and each feature in S with C is calculated. The minimum value of this mutual information (m-Joint*)* is the lowest amount of new information that the feature f_i adds to the shared information between S and C. The feature that produces the maximum m-Joint is the feature that adds maximum information to that shared between S and C, which means it is the feature which is the most relevant to the class label C in the context of the subset S according to the Definition 1.

Definition 3. Candidate feature f_i is redundant to the selected features within the subset S if f_i does not share new information with the class C.

Lemma 2. Let F be the full set of features, let S be the subset of features that are selected already, and $f_i \in F - S, f_s \in S$. If the feature f_i is highly correlated with a feature f_s in the subset then $I(f_i; C) \cong I(f_s; C) \cong I(f_i, f_s; C)$.

Proof: If the feature f_i is highly correlated with a feature f_s, then the probability mass function of f_i , f_s , and (f_i , f_s) are equal, $p(f_i) \cong p(f_s) \cong p(f_s, f_i)$.

Since the definition of the entropy is $H(\mathrm{X}) = -\sum_{i=1}^{N} p(x_i) log\big(p(x_i)\big)$

$then\ H(f_i) \cong H(f_s) \cong H(f_s, f_i)$,

Since the definition of the mutual information is

$I(X; C) = H(X) + H(C) - H(X, C) then\ I(f_i; f_s) \cong H(f_s) \cong H(f_i)\ and\ I(f_i; \mathrm{C}) \cong I(f_s; \mathrm{C})$

$I(f_i,\ f_s; C) = H(f_i,\ f_s) + H(C) - H(\ f_i, f_s, C)$, according to the definition, which can be simplified to: $I(f_i,\ f_s; C) = H(f_i) + H(C) - H(f_i, C)\ , I(f_i,\ f_s; C) \cong I(f_i; C) \cong\ I(f_s; C)$

5.1.1 **Joint Mutual Information Maximisation (JMIM)**

All methods mentioned in the section 2.12 attempt to optimise the relationship between relevancy and redundancy when selecting features by approximating the solution of Eq. (2.16). The JMI method is reported in existing literature as being the method which selects the most relevant features (Brown et al., 2012). It studies relevancy and redundancy, and takes into consideration the class label when calculating MI. However, the method still allows overestimation of the significance of some features, for example, when the candidate feature is in complete correlation with one or a few pre-selected features, but at the same time is almost independent from the majority of the subset. In such a situation, the value of the JMI goal function will be high despite the redundancy of the candidate feature to some features within the subset. This drawback is evident in almost all methods that use the cumulative sum approximation.

For this reason, a new method called Joint Mutual Information Maximisation (JMIM) is proposed in this research. JMIM employs joint mutual information and the ‘maximum of the minimum’ approach. Features are selected by JMIM according to the following criterion:

$$f_{JMIM} = \arg\ max_{f_i \in F-S}(\ min_{f_s \in S}(I(f_i, f_s; C))), \tag{6.1}$$

where

$$I(f_i, f_s; C) = I(f_s; C) + I(f_i; C/f_s), \tag{6.2}$$

$$I(f_i, f_s; C) = H(C) - H(C/f_i, f_s), \tag{6.3}$$

$$I(f_i, f_s; C) = \left[-\sum_{c \in C} p(c) \log\big(p(c)\big)\right] - \left[\sum_{c \in C} \sum_{f_i \in F-S} \sum_{f_s \in S} \log\left(\frac{p(f_i\, f_s, c/f_s)}{p(f_i/f_s) p(\ c/f_s)}\right)\right] \tag{6.4}$$

The method uses the following iterative forward greedy search algorithm to find the relevant feature subset of size *k* within the feature space:

Algorithm 6.1: Forward greedy search

1. (Initialisation) Set $\mathrm{F} \leftarrow$ "initial set of n features"; $S \leftarrow$ "empty set."
2. (Computation of the MI with the output class) For $\forall f_i \epsilon F$ compute $I(C;\ f_i)$.
3. (Choice of the first feature) Find a feature f_i that maximises $I(C;\ f_i\ )$; set $F \leftarrow F\backslash\{f_i\}$; set $S \leftarrow \{f_i\}$.
4. (Greedy selection) Repeat until$|S| = k$: (Selection of the next feature) Choose the feature $f_i = arg\ max_{f_i \in F-S}(\ min_{f_s \in S}(I(f_i, f_s; C)))$; set $F \leftarrow F \setminus \{f_i\}$; set $S \leftarrow S \cup \{f_i\}$.
5. (Output) Output the set S with the selected features.

5.1.2 **Advantage over Existing Alternative Methods**

The Venn diagrams in Figure 5.1 show different scenarios for the relationship between the candidate feature f_i, the selected feature f_s, and the class label C. Figure 5.1a illustrates the case in which methods like MIFS, NMIFS or mRMR will fail to select f_i because it is redundant to f_s, although each of them shares different information about C, and the correlation is not in the context of C.

The goal function of JMIM is similar to the goal function of CMIM (section 2.12.2), as CMIM also uses the 'maximum of the minimum' approach. The main difference is that CMIM maximises the amount of information the candidate feature f_i contributes given the pre-selected feature f_s (i.e. f_i is selected for any complementing f_s), whereas JMIM selects the feature that maximises the joint mutual information with f_s. Figures 5.1b and 5.1c are used to explain this difference further. The figures represent two candidate features f_i and f_j , and the subsequent selection of one of them. $I(f_i, f_s; C)$ is the union of areas 1, 2, and 3; $I(f_i; C/f_s)$ is area 1 in Figure 5.1b. The CMIM method would select f_i in Figure 5.1b, even though its complementing feature $\mathrm{f_s}$ from the subset does not carry as much information as the feature f_j in Figure 5.1c. Conversely, JMIM would select the feature that maximises JMI, so it would select feature f_i in Figure 5.1c. Therefore, the joint mutual information between the candidate feature and at least one of the pre-selected features will be high, which can increase the discrimination power of the selected subset.

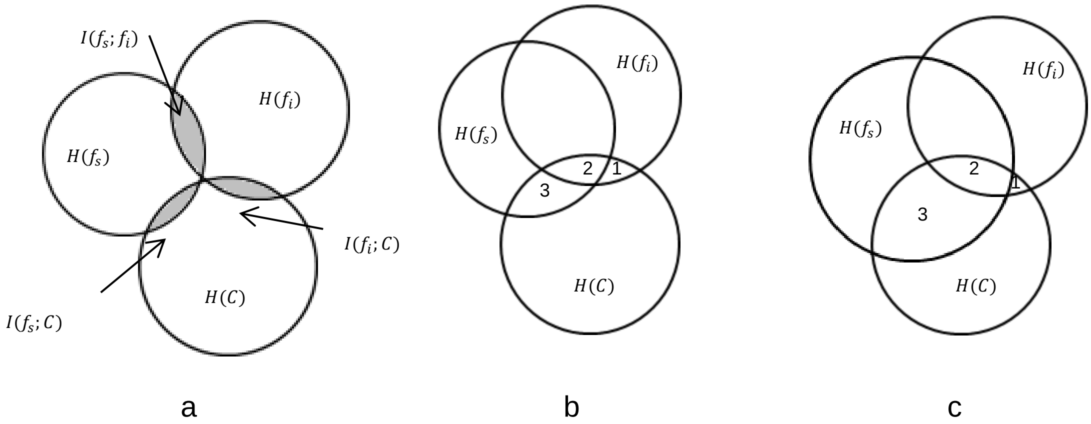


Figure 5-1: Venn diagrams illustrating the relation between features and class

5.1.3 **Normalised Joint Mutual Information Maximisation (NJMIM)**

The second method proposed in this chapter uses a goal function which is very similar to the one used in JMIM, the difference being that symmetrical relevance is used as an alternative to MI. This method is called Normalised Joint Mutual Information Maximisation (NJMIM). It is proposed in order to study the effect of using normalised MI instead of MI. The feature is selected according to the following equation:

$$F_{NJMIM} = \arg\ max_{f_i \in F-S}(\ min_{f_s \in S}(SR(f_i, f_s; C))), \quad (6.5)$$

where

$$Symmetrical\ Relevance = SR(F;C) \quad = \frac{I(F;C)}{H(F,C)}. \quad (6.6)$$

The method selects features according to the criterion:

$$F_{NJMIM} = \arg\ max_{f_i \in F-S}\left(min_{f_s \in S}\left(\frac{I(f_i,f_s;C)}{H(f_i,f_s,C)}\right)\right). \quad (6.7)$$

The same iterative forward greedy search algorithm is used to find the subset of features within the candidate feature space.

5.2 **Evaluation**

The performance of the two proposed methods in this chapter, JMIM and NJMIM, is compared with the results produced by five other methods: CMIM, DISR, mRMR, JMI, and IG. These methods are chosen for the following four reasons: (i) these methods are reported in the literature to provide good performance (Brown et al., 2012); (ii) the choice of these methods allows the comparison of the 'maximum of the minimum' approach used by JMIM and NJMIM with the cumulative summation used by JMI and DISR; (iii) it enables the analysis of the effect of using the symmetrical relevance instead of MI on the algorithm's performance; (iv) it allows the comparison of the effects of using joint mutual information and conditional mutual information, which are employed in JMIM and , CMIM respectively.

The seven methods are applied to data from different domains such as: life sciences, physical sciences, engineering, business, handwriting recognition, and gene microarray. The features within these datasets have different characteristics, being binary, discrete or categorical, or continuous. The continuous features are discretised into 10 equal intervals, using the Equal Width Discretisation (EWD) method (Dougherty et al., 1995).

Two classifiers are used to evaluate the quality of the selected subsets. These are Naïve-Bayes with kernel density estimation, and 3-Nearest Neighbours. Both classifiers are available in the Matlab Statistics Toolbox. The average classification accuracy is used as a measure of the quality of the selected features. Five-fold cross-validation is employed when processing feature selection and feature validation; therefore each fold is used for validation once. This means that 80% of the data is used for feature selection and classification training, whilst 20% is used for validation. This is repeated five times, using the whole dataset for validation over the course of five experiments. Overall, five different subsets of samples are used to generate five different subsets of features. Discretisation is performed as a pre-processing step for all data prior to the feature selection step.

Figure 5.2 shows the evaluation framework used in this experiment. To test the impact of adding each feature to the subset on the classification accuracy, training and validation are performed after the selection of each feature in the subset.

5.2.1 **Data**

Eight datasets from the UCI Repository (Frank and Asuncion, 2010) are used in the experiment (Table 5.1). These datasets have been previously used in similar research (El Akadi et al., 2008; Brown et al., 2012; Cheng et al., 2011). They have different characteristics in terms of the number of classes, features, instances and feature type. An example-feature ratio (Brown et al., 2012) is used as an indication of the difficulty of the feature selection task for the dataset. This ratio is computed using $\frac{N}{mC}$, where N is the number of instances, m is the median number of values that the features have, and C is the number of classes. The most challenging feature selection tasks are those performed using datasets with a small example-feature ratio. *Libra movement* is the most challenging dataset.

Table 5-1: UCI datasets used in the experiment.

No	Data Set	Number of features	Number of instances	Number of classes	Ratio
1	Credit approval	15	690	2	54
2	Gas sensor	128	13874	6	198
3	Libra movement	90	483	15	3
4	Parkinson	22	195	2	11
5	Breast	30	569	2	28
6	Sonar	60	208	2	10
7	Musk	166	7074	2	354
8	Handwriting	649	2000	10	20

To test the behaviour of the methods with an extremely small sample, datasets from Peng et al. (2005) are also used in the evaluation process, and these are shown in Table 5.2.

Table 5-2: Additional datasets used in the experiment (Peng et al., 2005).

No	Data Set	Number of features	Number of instances	Number of classes	Ratio
1	Colon	2000	62	2	10
2	Leukemia	7070	72	2	12
3	Lymphoma	4026	96	9	4

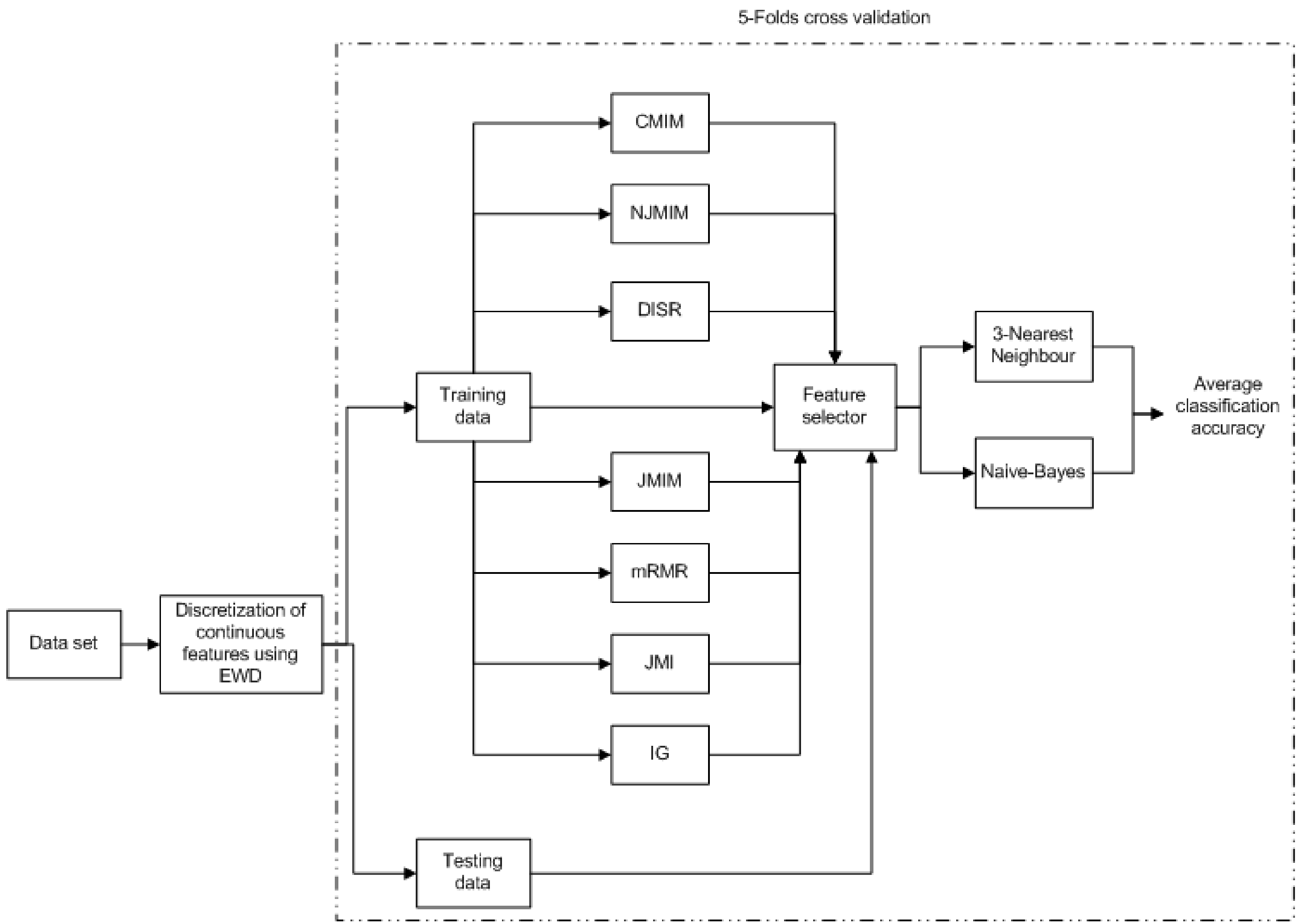


Figure 5-2: Evaluation framework

5.2.2 **Performance Analysis on Low Dimensional Datasets**

Figures 5.3 - 5.5 show the average classification accuracy of the three datasets with low numbers of features (*Parkinson, credit approval and breast*). The classification is computed over the whole size of the selected subset, from 1 feature up to 20 features (or all features of the dataset in the case of the *credit approval* dataset).

As shown in Figure 5.3, which illustrates the experiment with the first dataset, JMIM achieves the highest average accuracy (90.77 %) with just 8 features, which is higher than the accuracy of CMIM (90.26 %) and JMI (88.97 %). On the other hand, methods that use normalised MI, such as NJMIM and DISR, perform less well than JMIM and JMI, which use MI. This is expected for datasets with discrete features, because the normalisation may reduce the significance of the feature when it has high entropy and shares a high amount of information with the class label. The mRMR and IG methods perform poorly on this dataset.

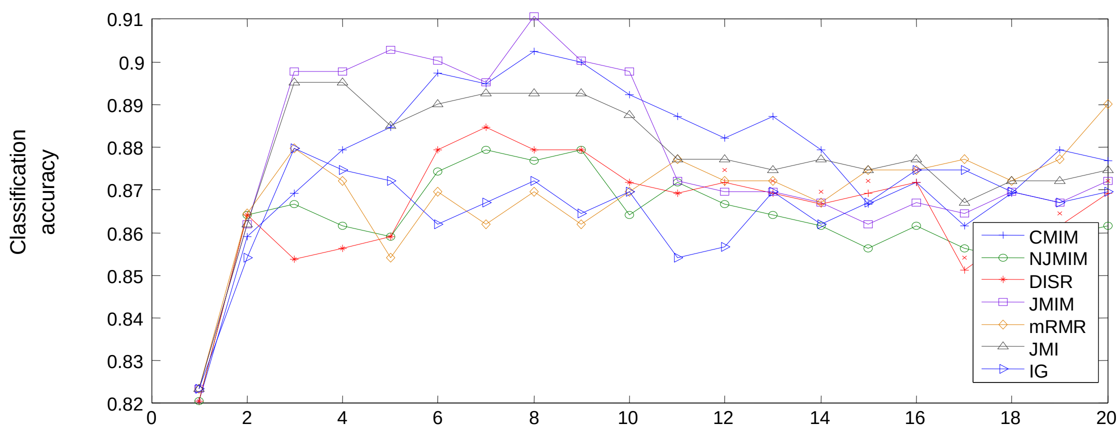


Figure 5-3: Average classification accuracy achieved with the *Parkinson* dataset

JMIM and JMI again achieve the highest classification accuracy, using only 4 features to reach an accuracy of 82.92 %. The accuracy of CMIM is 79.17 % with the same number of features. The other methods perform poorly compared to JMIM and JMI with the same number of features. The figure 5.4 also shows that the methods using normalised MI do not perform as well as those which use MI. Features selected by the JMIM and JMI methods have a higher discriminative power than the features which are selected by NJMIM and DISR. NJMIM performs better than DISR, yet both perform poorly.

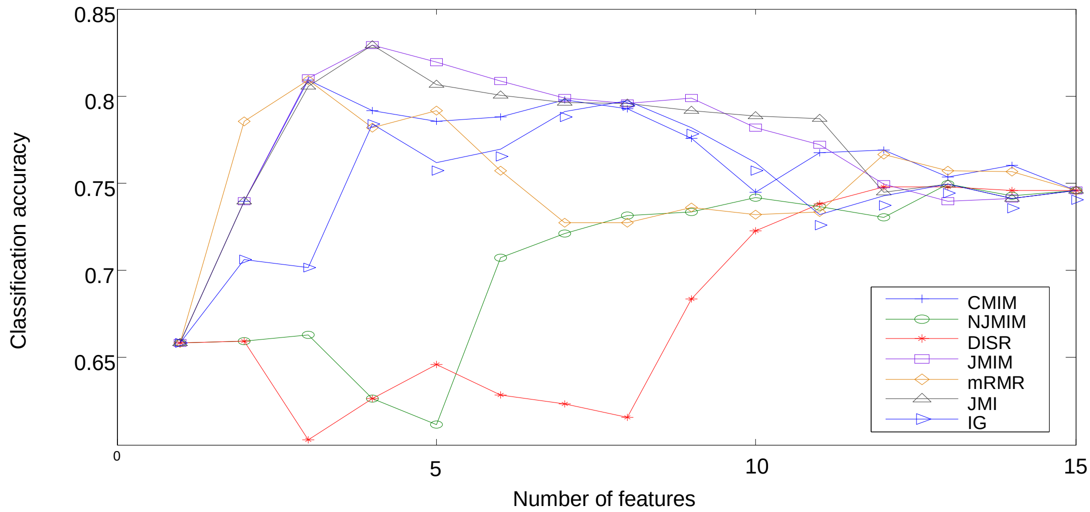


Figure 5-4: Average classification accuracy achieved with the *credit approval* dataset

The *breast* dataset has 20 features selected. As seen in Figure 5.5, JMIM does not achieve the highest classification accuracy. However, it produces a high accuracy (95.87 %) with only 5 features, while mRMR requires 14 features to achieve the same accuracy. JMIM performs better in comparison with JMI and CMIM. The performance of NJMIM and DISR is not as good as JMIM and JMI, as with 4 features their classification accuracies are 87.61 % and 89.28 % respectively.

5.2.3 **Performance Analysis on High Dimensional Datasets**

The second experiment involves high dimensional data (*musk, sonar, gas sensor, and handwriting* datasets. The experiment with the *gas sensor* and *sonar* datasets includes the selection of 50 features, with JMIM achieving high classification accuracy with a relatively small number of features. The other methods require more features to achieve this level of accuracy (Figures 5.6-5.7).

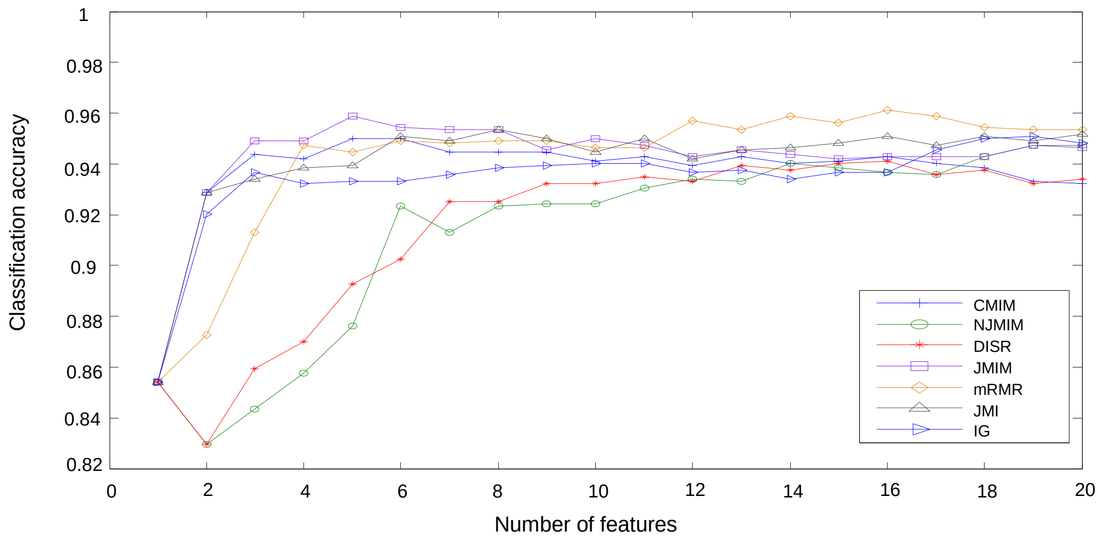


Figure 5-5: Average classification accuracy achieved with the *breast* dataset

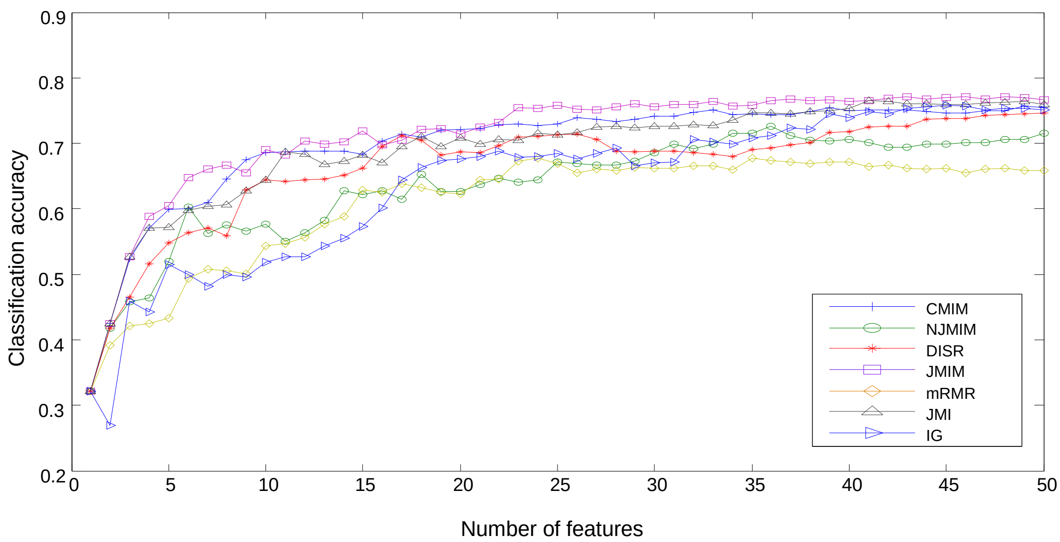


Figure 5-6: Average classification accuracy achieved with the *gas sensor* dataset

Figure 5.8 shows the results for the *handwriting* dataset. 50 features are selected. JMIM performs well, but is inferior to JMI and mRMR. In terms of classification accuracy of the selected subset JMI performed better than JMIM, in the subset with 11-21 features, by a maximum difference in accuracy of 0.5 %. The mRMR method also performs well with this dataset; however JMIM produces the highest accuracy (97.68%) with the selected subset of 33 features.

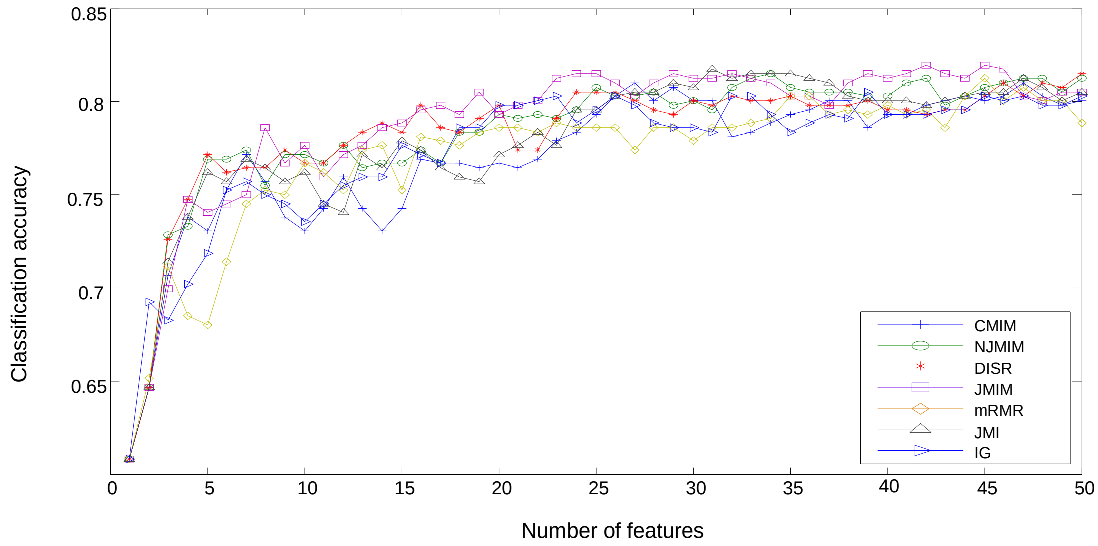


Figure 5-7: Average classification accuracy achieved with the *sonar* dataset

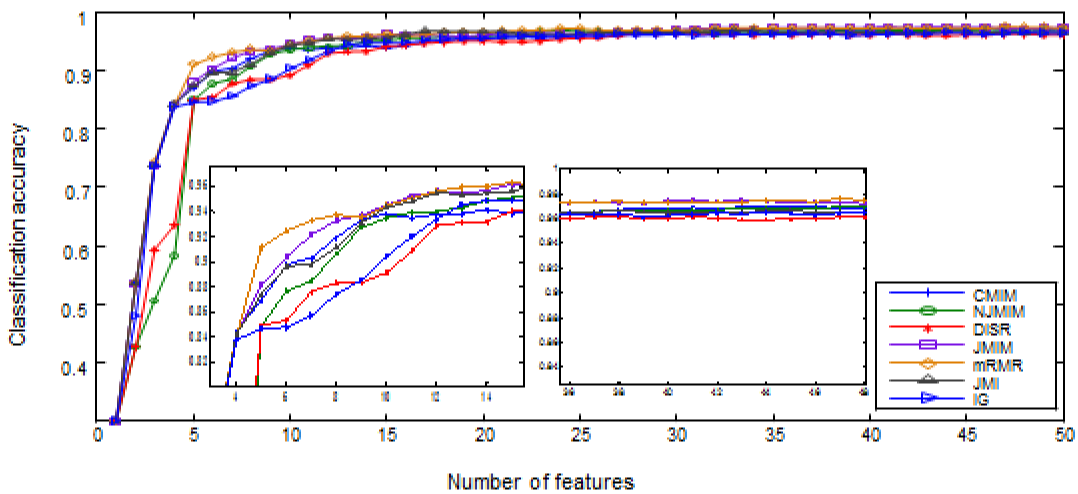


Figure 5-8: Average classification accuracy achieved with the *handwriting* dataset

The experimental results using the *libra movement* dataset are shown in Figure 5.9, when 50 features are selected. JMIM is the best method with this dataset with almost any number of selected features, followed by NJMIM. JMIM outperforms JMI by up to 3 % in terms of classification accuracy. NJMIM also outperforms DISR for all of the selected subsets.

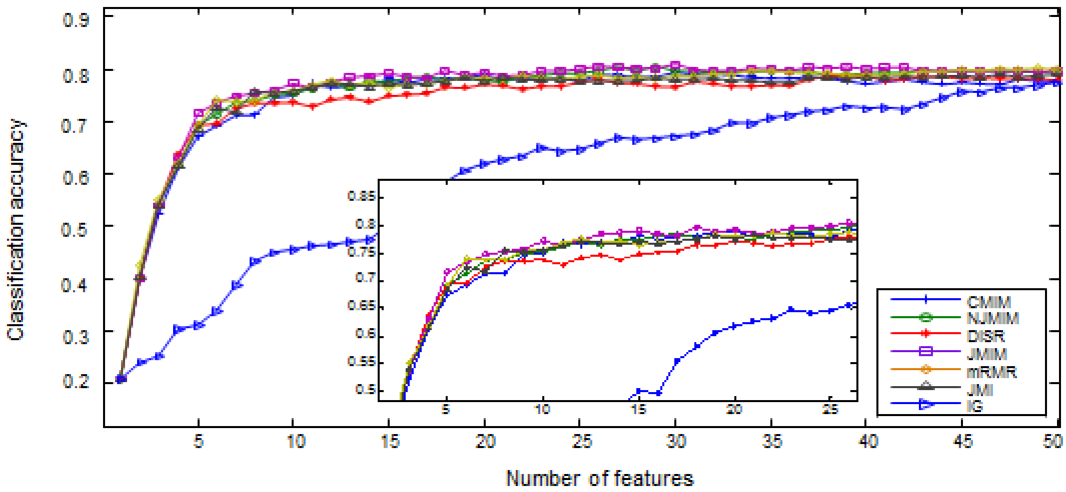


Figure 5-9: Average classification accuracy achieved with the *libra movement* dataset.

The methods are also applied to the *musk* dataset. Figure 5.10 shows the result when 50 features are selected. With this dataset, JMIM selects the best subset and outperforms the other methods in terms of classification accuracy. NJMIM does not perform as well as JMIM, but produces better accuracy than DISR and mRMR for most of the features selected.

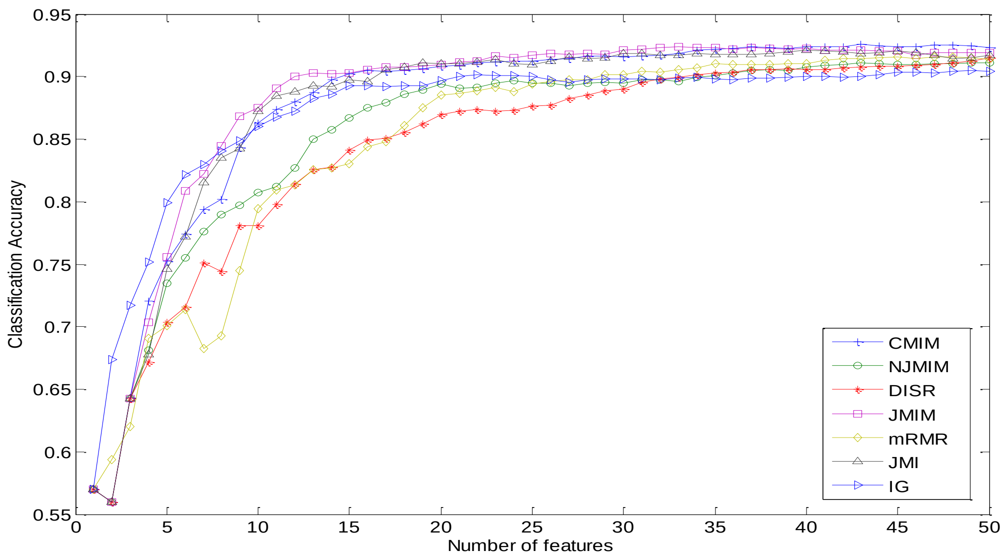


Figure 5-10: Average classification accuracy achieved with the *musk* dataset

5.2.4 **Performance Analysis on Peng ET. AL. (2005) Datasets**

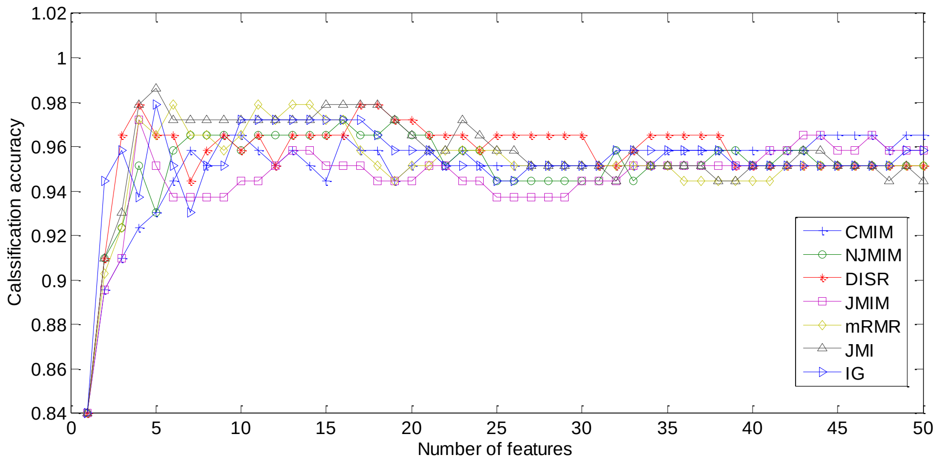
The results using the three datasets employed by Peng et al. (2005) are shown in Figure 5.11. The leukemia dataset (Figure 5.11a) has a small number of samples. The results show that none of the feature selection methods perform particularly well, confirming the findings reported in the review article by Brown et al. (2012). The *colon* dataset, which is the least challenging dataset of the three in terms of the number of classes and features, is shown in Figure 5.11b. The results indicate the better performance of JMIM and JMI compared to the other methods, especially CMIM, which performs poorly. However, CMIM is the method that provides the best accuracy with the *lymphoma* dataset, while JMIM, JMI and mRMR also perform well, with JMIM being the best of these. NJMIM performs better than DISR with all of the subsets below 34 features.

5.3 **Stability of the Methods**

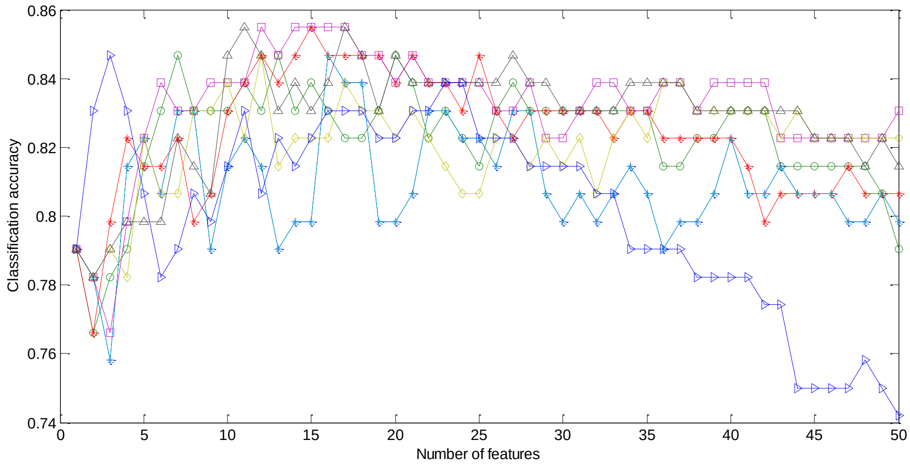
This section focuses on the stability of the feature selection methods discussed. The selected subset features are dependent on the datasets provided, and therefore any change to the data might lead to different selected features. In this context, the present study investigates the influence of changes in the data on the features selected. Kuncheva's measure of stability (Kuncheva, 2007), known as the consistency index, uses Eq. (6.8) to compute the consistency between two selected feature subsets, S^1 and S^2:

$$(S^1, S^2) = \frac{rn - k^2}{k(n-k)}, \qquad (6.8)$$

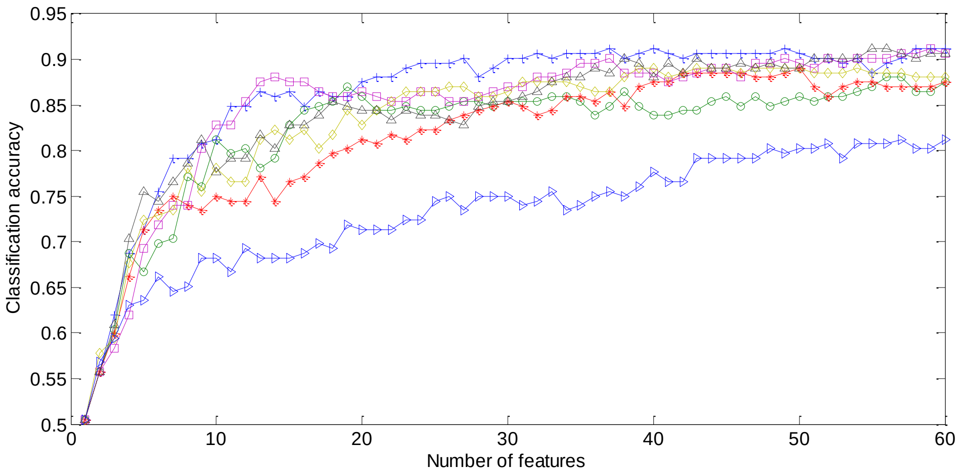
where S^1 and S^2 are selected feature subsets using different groups of dataset samples, i.e. $S^1, S^2 \in F$ where F is the total set of the feature, $|S^1| = |S^2| = k, |F| = n$, and $r = |S^1 \cap S^2|$. However, this method does not take into consideration the correlation between features.



a-*Leukemia*



b-*Colon*



c-*Lymphoma*

Figure 5-11: Average classification accuracy with the additional datasets

Yu et al. (2008) proposed a method for measuring stability based on similarity. This method takes into account the correlation between features. It calculates the weight between each pair of features from the subsets S^1 and S^2, computes the similarity between S^1 and S^2, and constructs a bipartite graph. If f_i is a feature belonging to S^1 and f_j is a feature belonging to S^2, the value of the weight can be the correlation coefficient, or any other similarity measure. This article uses symmetrical uncertainty (Yu and Liu, 2004) to calculate the weight w:

$$w\left(s_i^1, s_j^2\right) = 2\left[\frac{I\left(s_i^1, s_j^2\right)}{H\left(s_i^1\right) + H\left(s_j^2\right)}\right], \tag{6.9}$$

where $0 \leq w\left(s_i^1, s_j^2\right) \leq 1.0$. To find the maximum weighted bipartite matching, the Hungarian Algorithm (Kuhn, 1955) is used to find the optimal solution.

This experiment uses the eight UCI datasets, as shown in Table 5.1. Each dataset is divided into 5 folds; 4 of which are used for feature selection using the CMIM, NJMIM, DISR, JMIM, mRMR, JMI, and IG methods. Eq. (6.9) is used to calculate the weight between the features within each pair of selected subsets from each dataset. The final cost is divided over the cardinality of the subset used, and therefore the magnitude of the final cost should be less than or equal to 1.0 (it is 1.0 if all selected subsets are the same).

The relationship between accuracy and stability is computed by comparing the average classification accuracy and the average stability with different numbers of features.

Table 5.3 shows the average accuracy/stability for each method in no particular order. It is worth noting that the methods employing the ‘maximum of the minimum’ criterion (JMIM, NJMIM and CMIM) tend to have lower stability than the methods using the cumulative summation approximation (JMI and DISR). The best method in terms of stability is IG. JMIM has the best compromise between accuracy and stability. Moreover, it demonstrates the best average classification accuracy among all methods.

Table 5-3: Average stability, average accuracy and the compromise between accuracy and stability.

Method	Accuracy	Stability	Accuracy/ stability
CIMIM	0.8488	0.8598	0.9197
NJMIM	0.8264	0.8344	0.8954
DISR	0.8129	0.9054	0.8807
JMIM	0.8578	0.8598	0.9294
mRMR	0.8278	0.8868	0.8969
JMI	0.8490	0.8838	0.9199
IG	0.8226	0.9228	0.8913

5.4 **Discussion**

During the evaluation JMIM performs well in comparison with state of the art methods. It produces the best accuracy for datasets with a low number of features, such as the *Parkinson, credit approval and breast* datasets. In these experiments, the maximum average classification accuracy achieved by JMIM with the *Parkinson* dataset (figure 5.3) is 90.77 %. JMIM and JMI achieved an accuracy of 82.92 % with the *credit approval* dataset whilst JMI and CMIM achieved 93.83 % and 95.22 % respectively. The JMIM method also performed well on high dimensional datasets, such as the *musk, sonar, gas sensor* and *handwriting* datasets.

JMIM and JMI also outperform the other methods on extremely small sample datasets with a large number of features, such as the *colon* dataset. However, CMIM produces the best performance with the *lymphoma* dataset. JMIM, JMI, and mRMR also perform better than the other three methods.

In addition to the quantitative assessment of the accuracy of the proposed methods, several experiments are conducted to enable an in-depth comparison of different feature selection methods, according to several specific criteria. For example, the nonlinear approach, which uses the 'maximum of the minimum' criterion, is compared to the linear approach that employs cumulative summation approximation. In particular, JMIM is compared to JMI, with the results showing that the non-linear approach performed better than the linear approach when tested with most of the datasets.

The goal function based on joint mutual information is compared to the goal function based on conditional mutual information, with the result showing better performance of joint mutual information in combination with the non-linear criterion.

Finally, the effect of using normalised mutual information instead of mutual information is tested by comparing the performance of JMIM and JMI with NJMIM and DISR. The results show that, with the discretised datasets, the methods employing non normalised mutual information perform better than those using normalised mutual information. This suggests that division of the mutual information over the joint entropy does not improve performance.

In addition, the methods are compared in terms of their stability, as described in detail in Section 6. The results demonstrate that the methods employing the 'maximum of the minimum' criterion, such as CMIM, JMIM, and NJMIM, show a lower average stability than the methods which employ cumulative summation, although there is no dominant method.

5.5 **Summary**

This chapter presents two novel feature selection methods based on information theory: JMIM and NJMIM. The methods employ a forward search algorithm and 'maximum of the minimum' criterion to produce a significant feature subset.

The performance of the proposed methods is compared against that of five other feature selection methods: Joint Mutual Information (JMI), Conditional Mutual Information Maximisation (CMIM), Maximum Relevancy Minimum Redundancy (mRMR), Double Input Symmetrical Relevance (DISR), and Information Gain (IG). They are compared in terms of their ability to select features with high discriminative power, and their stability. Eleven datasets are used to evaluate the comparison. The JMIM method shows a good performance, and it outperforms the other methods with most of the datasets.

Overall, the results show that the JMIM method outperforms the other methods in terms of classification accuracy when applied to most of the tested datasets. Moreover, this method produces the best trade-off between accuracy and stability.

Feature selection is employed in this research to define the significant clock features. Defining theses feature can improve the diagnostic accuracy of CDT and also give more information about the relation between the features and the various tasks of discriminating between different cognitive statuses. In the next chapter JMIM is employed to define the clock features which are significant to discriminating between different dementia types and stages. The chapter also discusses the temporal changes in the clock as dementia gets more severe.

Chapter: 6

Application of JMIM to Identify Significant Clock Features

Although there a large variety of clock features have been used to identify various neurological conditions, at the time of this research there had been little study into the significance of clock drawing features in dementia assessment. As outlined in Chapter 2, only two such studies have been identified in the literature, and these are broad studies which defined the significant features for discriminating between normal and abnormal diagnoses. This study uses more sensitive analysis to relate clock features specific cognitive status by using sophisticated statistical methods to determine the significance of the features. This analysis can also enable medical practitioners to use a more relevant, reduced list of features during dementia assessment.

This chapter proposes to investigate the significance of each of the 47 features proposed in this research. The aim is to determine which clock features provide the maximum discrimination between cognitive statuses. The JMIM method is one of two methods proposed in Chapter 5 is used to select the most significant features for five discriminative tasks as this method outperform the other methods. This set of features is the one which is considered to be the most informative, which leads to a more accurate diagnosis in the classification stage.

According to the algorithm of the JMIM method (Section 5.1), the most important feature is first selected and added to the empty subset at the first step. The method continues by adding features to the subset based on their contribution to the joint mutual information between the subset and the diagnosis label (class label).

This chapter also proposes a new framework for analysis of the correlation between the temporal changes in the features of CDT drawings and the progress of dementia. The aim is to determine how the severity of dementia affects the abnormality pattern among the clock features.

The contributions of this chapter are:

1. Defining the significant feature subsets in the clock drawings by utilising the JMIM method on 604 drawings. The significant subsets for five different discriminative tasks are defined.

2. A new framework for analysing the correlation between the temporal changes in the features of CDT drawings and the progress of dementia.

The chapter is organised as follows: Section 6.1 presents the preparation of the CDT data for the feature selection experiment; Section 6.2 defines the significant CDT features for five different discriminative tasks; Section 6.3 proposes a framework for the analysis of temporal changes in the CDT features corresponding to the progress of dementia; finally, section 6.4 summarises this chapter.

6.1 **Data Preparation**

As specified in Chapter 4, 648 CDT drawings are used in this study. The distribution of this data shows that there are an insufficient number of drawings for the two diagnoses known as “other degenerative dementia” and “other types of dementia”. These two diagnoses are therefore not included in the research because of the low volume of their samples and also because they are not common causes of dementia. It is mentioned in Chapter 2 that AD and VaD are the cause of almost 90 % of dementia cases, and so these are focused on. Besides these two diagnoses the MCI and non-organic cases (normal and functional) are also included. The MCI stage is important as this diagnosis is considered as the risk state of dementia. A functional problem is not an organic problem (the brain is still physiologically healthy) so in part of this research the functional problems are grouped with the normal diagnoses to construct a new group named Normal+.

One of the aims of this research is to diagnose dementia in the early stages, in order to take the advantage of early medical intervention. For this reason, and following advice from medical specialists, the cases that are diagnosed as AD or VaD are divided into 3 groups based on the severity level (Severe, Moderate, or Mild). The cases in which the MMSE score is over 23 are considered as mild AD and mild VaD dementia cases. 23 is chosen as the threshold since it is the original cut-off point for MMSE to indicate normal cognitive abilities. The mild dementia cases are added to the MCI group to construct a new group: MCI+. Diagnosing the cases that belong to this class (group) is important, since medical intervention is most effective at this early stage. Figure 6.1 shows the data distribution of each diagnosis.

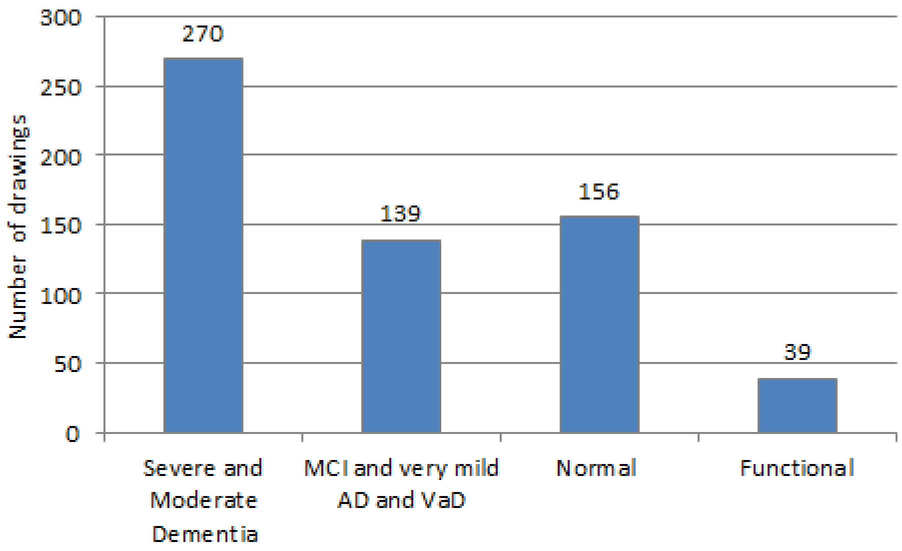


Figure 6-1: Data distribution of the classes in the data set.

6.2 **Significant Clock Features Using JMIM Method.**

The experimental algorithm depicted in Figure 6.2 is employed to determine the significant features. Since seven of the extracted features are continuous, discretisation is needed to convert them into discrete variables. The MDL method is employed for this task as it is reported in comparative studies in the literature to be one of the most effective methods (Dougherty et al., 1995; Liu et al., 2002). In some cases MDL fail to discretise the majority of the features because it could not find the cut-point in the continuous range of these features. In these cases the EWD unsupervised method is employed. Although EWD is a simple and easy to implement it is reported to produce good performance comparable to the performance of some more sophisticated methods (Dougherty et al 1995).

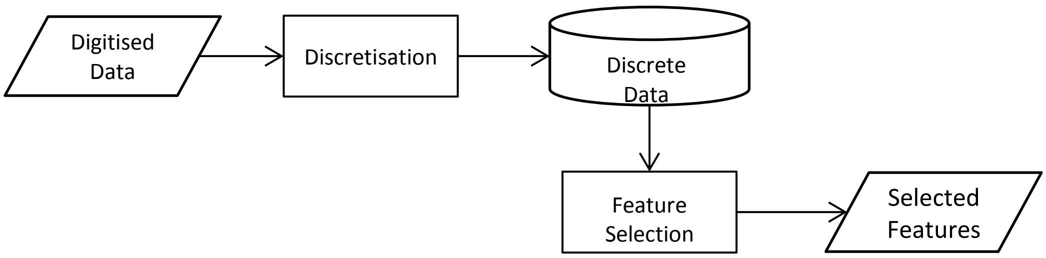


Figure 6-2: Feature selection algorithm.

The primary outcome of this analysis is the identification of the significant feature subsets that produce high discrimination between cognitive statuses. The mutual information shared between each individual feature and the class label is also calculated. Mutual information describes the amount of information that the feature contains about the class label. The experiment is repeated five times in order to define the features that are significant in discriminating between the following cognitive statuses:

1. Normal+ / MCI+
2. Severe and moderate / MCI+
3. Normal+ / Abnormal
4. Normal / Functional
5. AD / VaD

These five discriminative tasks are chosen according to a medical advice because they are medically significant, and also because the feature subsets of some of these discriminative tasks could be employed to enhance the performance of the diagnosis stage (next chapter).

During the experiment, the CDT dataset is reorganised to construct five new sub-sets of data, each of them containing only the two classes for that specific diagnosis task.

6.2.1 Results

In this section, the results of each discriminative task are described. The full results are shown in Appendix B, with only the most informative 20 features for each diagnosis task being listed in this chapter.

Significant Features for Distinguishing between Normal+ and MCI+ Cases

As explained earlier in this chapter, discriminating between Normal+ and MCI+ classes is the most important task in assessing dementia from a medical prospective. The diagnosis of MCI cases and the early stage of dementia is a very challenging task which even the best CDT scoring systems are likely to fail as it is shown in the comparative study in Section 4.1. In response to this, a greater number of features are extracted from the CDT in order to improve their performance and to increase the sensitivity of the test in capturing the deficits within the clock at this stage

The JMIM method is applied to the feature set described in Section 4.3 to determine the order of significance among the features for this task. Table 6.1 shows the 20 most significant features, including those related to the clock hands, time setting and the center, and also features related to the numbers. This result agrees with the conclusions of Ehreke, et al., 2011 who found that using more features related to the clock hands and setting of the time increased the performance of CDT in distinguishing between MCI and Normal cases. The table also shows that eight of the new proposed features are within the most significant 20 features for this diagnosis task. The table which illustrates the rank of all the features is shown in appendix B.

Table 6.1 also shows that patients in the early stage of dementia or MCI are likely to be unable to set the hands correctly, and find it difficult to write numbers in their correct positions. This is consistent with a previous study conducted by Parsey, et al. (2011), whereby errors in hand position and number spacing are seen to be more sensitive to the early stages of cognitive impairment.

Figure 6.3 shows the amount of mutual information shared between each feature and the class label individually. The figure shows that the angle between the clock hands is the most important of the 47 features. It also shows that the features related to the time setting share a large amount of information with the diagnosis, especially the setting of the minute hand. The experiment also demonstrates that writing the numbers far from the perimeter is another important sign of early dementia.

Table 6-1: Twenty significant features for discriminating between MCI+ and normal+.

No.	Feature
1	Angle between clock hands
2	Count of numbers within Area 1
3	Maximum angle between numbers
4	Displacement of minute hand or mar k from the target number
5	Count of numbers within Area 2
6	Position of min hand
7	Ratio between hands
8	Time is correct
9	Maximum size of numbers
10	Distance between the position of hands intersection and the center
11	Position of Hour hand
12	Count of numbers within quadrant4
13	Arrows on hands
14	Minimum angles between numbers
15	Hands are joint or within 12 mm
16	Hands connected with target number
17	Count of numbers within Area 3
18	Stem of clock hands (near to the center) is left out
19	Displacement of arrows less than 4 mm
20	Count of numbers within quadrant2

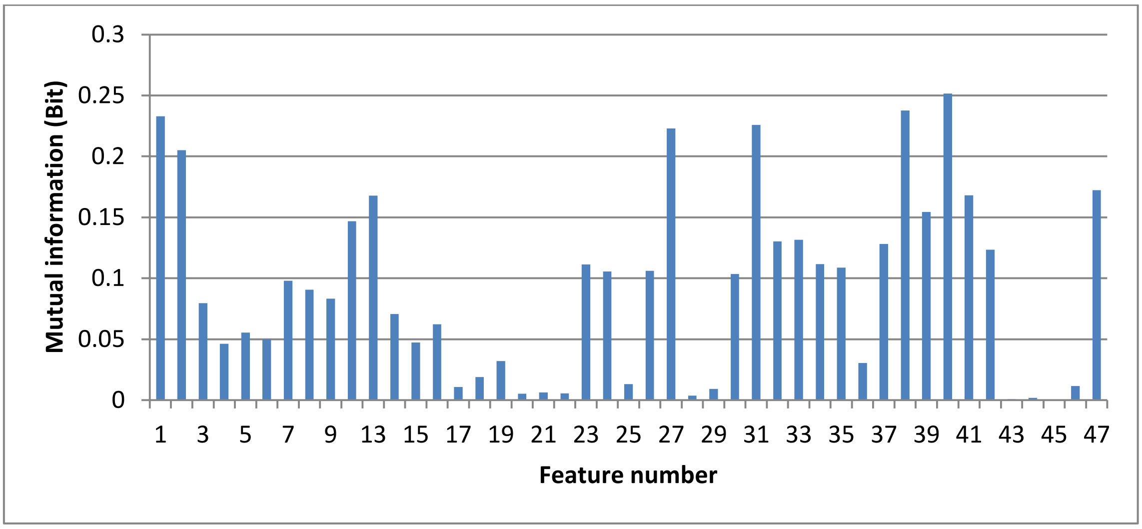


Figure 6-3: Mutual information between the features and the diagnosis (Normal+ / MCI+).

The Significant Features for Distinguishing between MCI+ and Severe/ Moderate Cases

This task involves the discrimination between MCI+ and the moderate / severe stages of dementia. Table 6.2 lists the 20 most significant features for this discrimination task. Six features (maximum size of the numbers; ratio between the maximum number size and the minimum size; count of numbers outside the perimeter; count of numbers whose rotation is over 25 degrees; count of numbers left out from the drawing; and count of duplicated numbers) are removed during the discretization stage because the MDL method failed to find optimal cut-off points to discretise them due to the low mutual information between these features and the class label.

Table 6-2: Twenty significant features for discriminating between MCI+ and Severe or moderate dementia.

No.	Feature
1	Count of numbers within Area 1
2	Angle between clock hands
3	Count of numbers within Area 3
4	Displacement of minute hand or mar k from the target number'
5	Count of numbers within Area 2
6	Position of min hand
7	Time is correct
8	Count of numbers within quadrant 4
9	Ratio between hands
10	Count of numbers within quadrant 2
11	Count of numbers within quadrant 1
12	Count of numbers within quadrant 3
13	minimum angles between numbers
14	Hands are joint or within 12 mm
15	Stem of clock hands (near to the center) is left out
16	Maximum angle between numbers
17	Minute hand is present
18	Position of Hour hand
19	Distance between the position of hands intersection and the center
20	Hands self-correction

The significant subset for this task is very similar to that of the normal vs. MCI+ diagnostic task, with only four features being replaced..

The four new features which are within the top 20 for this discrimination task but which were not selected in the previous one is: count of numbers within quadrant 2; count of numbers within quadrant 1; minute hand is present; and hands self-correction. These features indicate that severe and moderate dementia patients might face more difficulty in number spacing, and also more difficulty in setting the hands, with an additional tendency to omit the minute hand.

The mutual information between the remaining 41 features and the diagnosis label is shown in the Figure 6.4. The figure shows that the count of numbers in area 1 is the most important feature, followed by features related with the clock hands and the time setting. The figure also shows that the amount of shared information is generally lower than the information that is shared between the class and the features in the case of Normal+ vs. MCI+ which means the difference between the two classes is not as great as it is for the previous discriminative task.

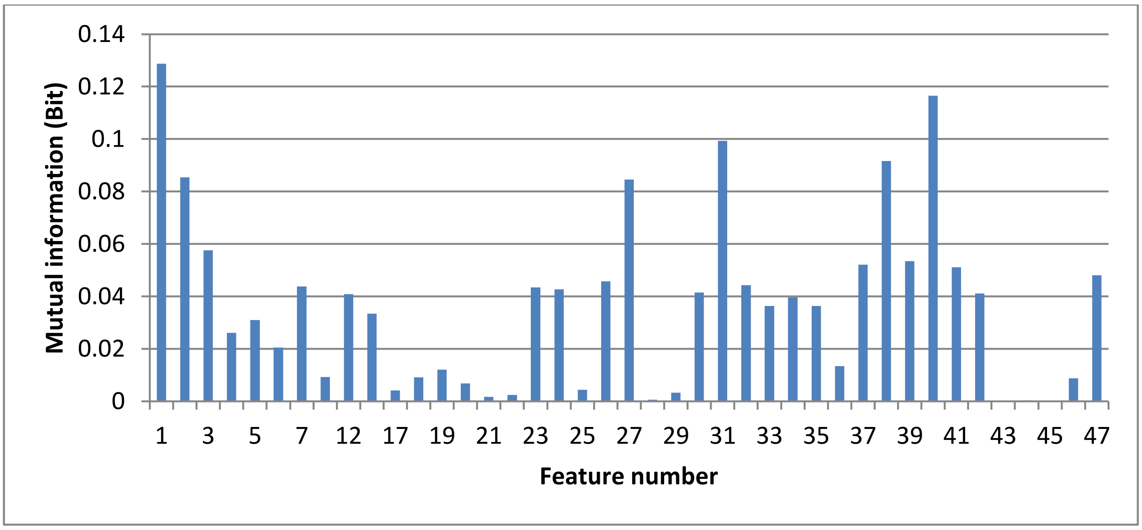


Figure 6-4: Mutual information between the features and the diagnosis (MCI / Severe or moderate).

Significant Features for Distinguishing between Normal+ and Abnormal Cases

The discrimination between normal and abnormal cognitive functions is the primary task for CDT. All the proposed CDT Scoring systems are designed to perform an assessment of the drawings whereby the drawings are classified as normal or abnormal based on a predetermined cut-off point. For this diagnosis task, the MDL could not find the optimal points to discretise the features ‘Count of numbers outside the perimeter’; therefore this feature is removed from the list.

Table 6.3 lists the 20 most significant features selected by JMIM. The subset is very similar to the subset selected in the case of Normal+ / MCI+, differing only in the two features: “count of numbers with orientation value larger than 25” and “arrows are pointing to the wrong direction”. Figure 6.5 shows the mutual information between the features and the diagnosis label (Normal+ vs. Abnormal). The amount of information shared is the highest compared with the previous cases. The reason for this is that when severe and moderate cases are included with MCI+, the difference between the classes of Normal+ and Abnormal is increased.

The results for the three diagnosis tasks show that features related to the time setting and clock hands, and to the number positioning and spacing, are significant for discriminating the dementia classes. Some errors which are reported in literature (lessig et al., 2008; Jouk and Tuokko, 2012) as significant clock features (such as number substitution, missing numbers, repetition, number orientation, extra marks, etc.) do not appear here to be important features in terms of the CDT’s diagnostic power (although number orientation is among the top 20 features in the Normal+ vs. Abnormal diagnostic task).

This difference is due to the creation of the MCI+ group in the current research. It is also attributed to the low number of very severe drawings among the data available for this study.

Table 6-3: Twenty significant features for discriminating between Normal+ and Abnormal and moderate dementia cases.

No	Feature
1	Angle between clock hands
2	Ratio between hands'
3	Minimum angles between numbers
4	Distance between the position of hands intersection and the center
5	Minimum size of numbers
6	Count of numbers within Area 1
7	Maximum angle between numbers
8	Time is correct
9	Position of minute hand
10	Arrows on hands
11	Count of numbers within quadrant4
12	Displacement of minute hand or mark from the target number
13	Position of Hour hand
14	Hands connected with target number
15	Count of numbers within Area 2
16	Stem of clock hands (near to the center) is left out
17	Maximum size of numbers
18	Count of numbers which its orientation is More 25
19	Arrows are pointing to the wrong direction
20	Hands are joint or within 12 mm

Significant Feature for Distinguishing between Normal and Functional Cases

This task diagnosis the functional cases from the normal healthy cases. The influence of functional problems on the CDT have not been widely studied, and the researches which have been conducted shows inconsistency. Some studies report a difference between the healthy patients and those with functional problems, while other researchers reported no difference between the two groups (Heinik et al., 2010). In this section the JMIM method is applied to the Functional vs. Normal data to determine the significant feature subset. Table 6.4 shows the 20 most significant features selected by JMIM.

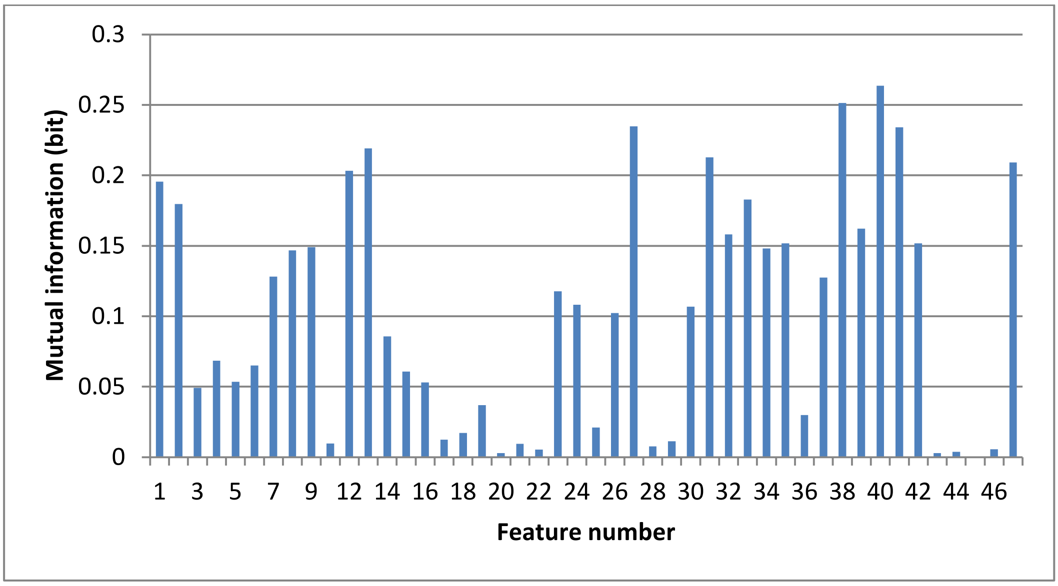


Figure 6-5: Mutual information between the features and the diagnosis (normal+ / abnormal).

The table shows that the features are almost the same as the significant subset for the Normal+ vs. MCI+ discriminative task, but with a different ranking order. The "Maximum size of numbers" feature is shown to be the most informative feature for this task followed by the "Maximum angle between numbers". "Angle between hands" and "Count of numbers in area 1" are not the two most significant in this case, but they are still among the 20 most significant features.

The mutual information between the diagnosis label and each feature individually for the discriminative task (Functional vs. Normal) is shown in Figure 6.5. The figure shows that just a few features share a reasonable amount of information with the diagnosis label and the rest are almost irrelevant. The feature "Maximum size of numbers" is sharing the maximum amount of information, significantly more that the rest of the features. The features "Maximum angle between numbers", "Minimum size of numbers", "Angle between clock hands", and "Ratio between hands" are shown to share relatively high information with the diagnosis label compared to the rest of the features.

Table 6-4: Twenty significant features for discriminating between Normal and Functional dementia.

No	Feature
1	Maximum size of numbers
2	Maximum angle between numbers
3	Distance between the position of hands intersection and the centre
4	Minimum angles between numbers
5	Angle between clock hands
6	Count of numbers within Area 2
7	Minimum size of numbers
8	Ratio between hands
9	Count of numbers within quadrant2
10	Position of min hand
11	Count of numbers which its orientation is More 25
12	Position of Hour hand
13	Arrows on hands
14	Numbers within Area 1
15	Stem of clock hands (near to the centre) is left out
16	Count of numbers within quadrant 4
17	Time is correct
18	Ratio between max and min size
19	Hands are joint or within 12 mm
20	Displacement of arrows less than 4 mm

Significant Features for Distinguishing between VaD and AD Cases

In this section the significant features for the VaD vs. AD (severe and moderate) discriminative task are defined. The literature does not show a significant difference between the CDT drawings of AD patients and VaD patients (Brian et al., 2013). Freedman et al. (1996) also concluded that CDTs are not a diagnostic tool for defining the specific type of disease. Regardless, the most 20 significant features for this task are identified here to verify the previous studies' results. Table 6.5 shows the subset of 20 most significant features.

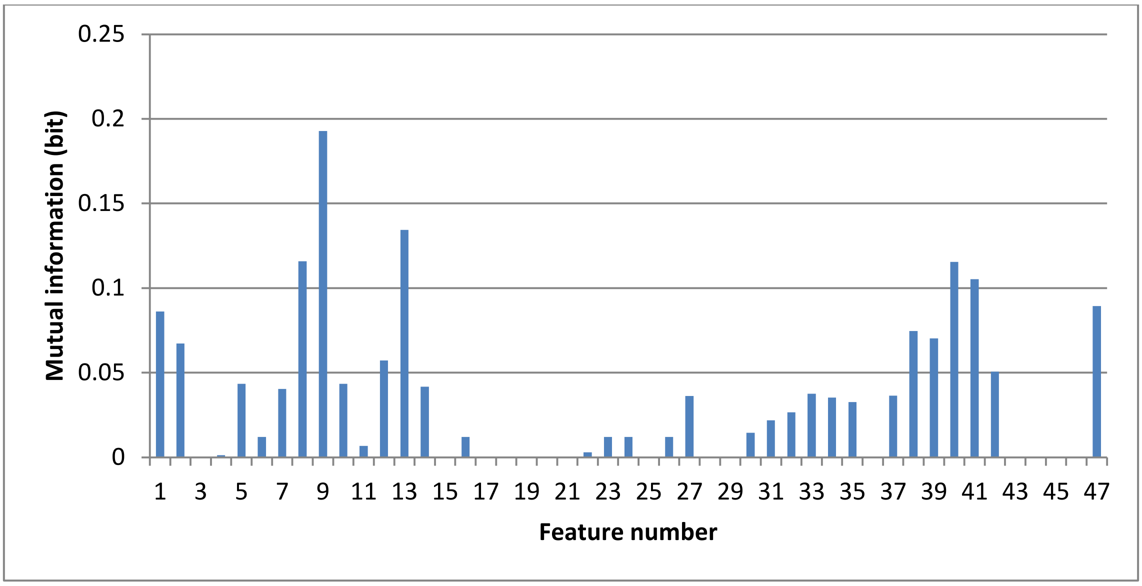


Figure 6-6: Mutual information between the features and the diagnosis (functional/ normal).

Table 6.5 shows the features that discriminate between AD and VaD. However, as it is shown in Figure 6.7, not all the features in the subset share a large amount of information with the diagnosis label, supporting the results reported in the literature about the difficulty in discriminating between the different types of dementia.

6.3 Temporal Changes in the CDT Features Corresponding to the Progress of Dementia

This experiment aims to assess the progress of deficits in the clock drawings as the dementia develops. Over time the clock produced by a given patient is likely to become more deficient as the dementia becomes more severe. However, this experiment is not a longitudinal study because, as explained in Chapter 4, the data used in this research is anonymous, and hence it is not possible to follow the development of a specific case. Instead, the characteristics of the CDT drawings for each diagnosis group are analysed and compared.

Table 6-5 : Subset of twenty significant features for discriminating between AD and VaD (severe and moderate cases).

No	Feature
1	Maximum angle between numbers
2	Distance between the position of hands intersection and the centre
3	Angle between clock hands
4	Count of numbers within Area 1
5	maximum size of numbers
6	Count of numbers which its orientation is More 25
7	Position of min hand
8	Count of numbers within Area 2
9	minimum size of numbers
10	Count of numbers within quadrant 1
11	Count of numbers within quadrant 3
12	Count of numbers within quadrant2
13	Count of numbers within quadrant4
14	Position of Hour hand
15	numbers within Area 3
16	Arrows on hands
17	hands connected with target number
18	repeated or duplicated numbers
19	Number of numbers left out
20	Ratio between max and min size

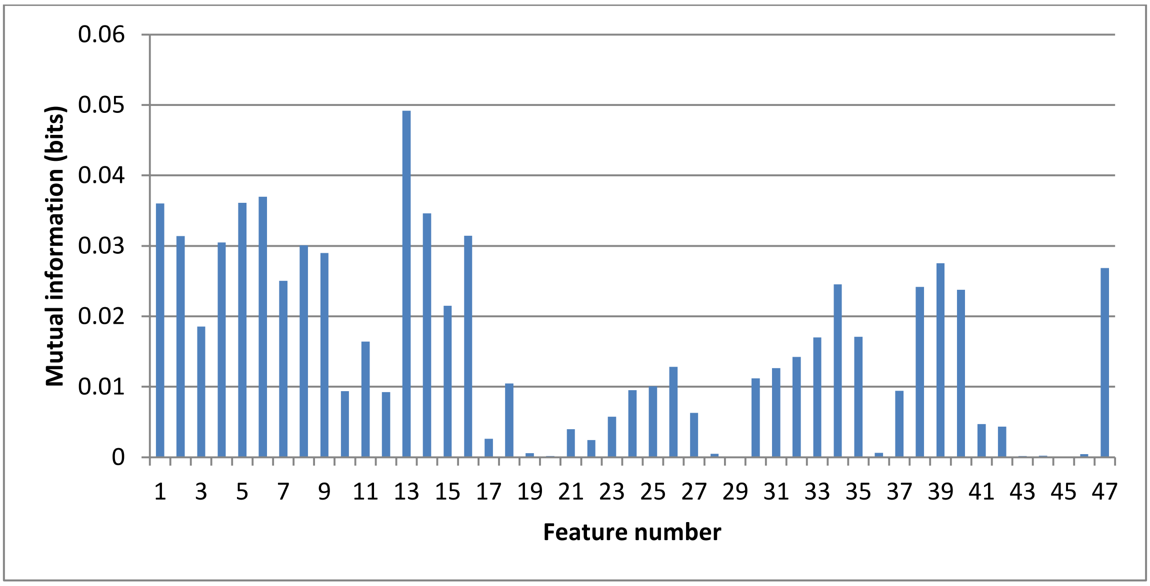


Figure 6-7: Mutual information between the features and the diagnosis (AD vs. VaD).

In this study, the MMSE test is used as an indication of the severity level of the drawing errors. The experiment is conducted using 409 clock drawings by VaD, AD, and MCI patients. The MMSE score of these cases generally remains between 5 and 30. Figure 6.8 shows the distribution of the drawings within each MMSE score 'bin' for AD, VaD, and MCI. The figure shows that the MCI drawings are within the MMSE range of 23-30.

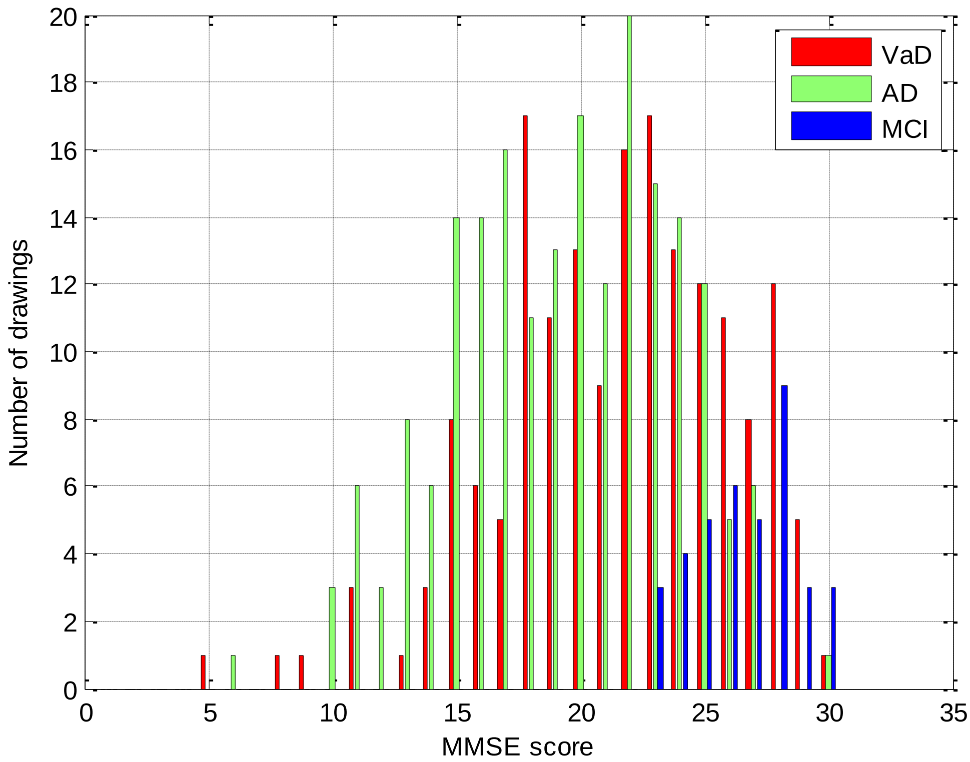


Figure 6-8: Distribution of the drawings over the MMSE score.

A total of 156 normative clock drawings are used to find the normative base (range of values) for each of the 47 features. Figure 6.9 shows a flowchart of the framework that is used to analyse the progression of symptoms.

The first step is to define the normative range of each feature by excluding the outliers. The Boxplot method (Tukey, 1977) is employed to find the outliers among each feature in the normative data. Boxplots are reported to perform well on data which is not highly skewed (Seo, 2006), which is the case for the CDT features. The normative range of each feature is given in Appendix C.

The second step is to find, for a given feature, the number of drawings whose feature value lies outside the normative range. Consequently the percentage of the drawings is calculated and used as indication of how likely it is for the patient to fail to draw that particular feature normally (the frequency of feature abnormality). This task is then repeated for all the features.

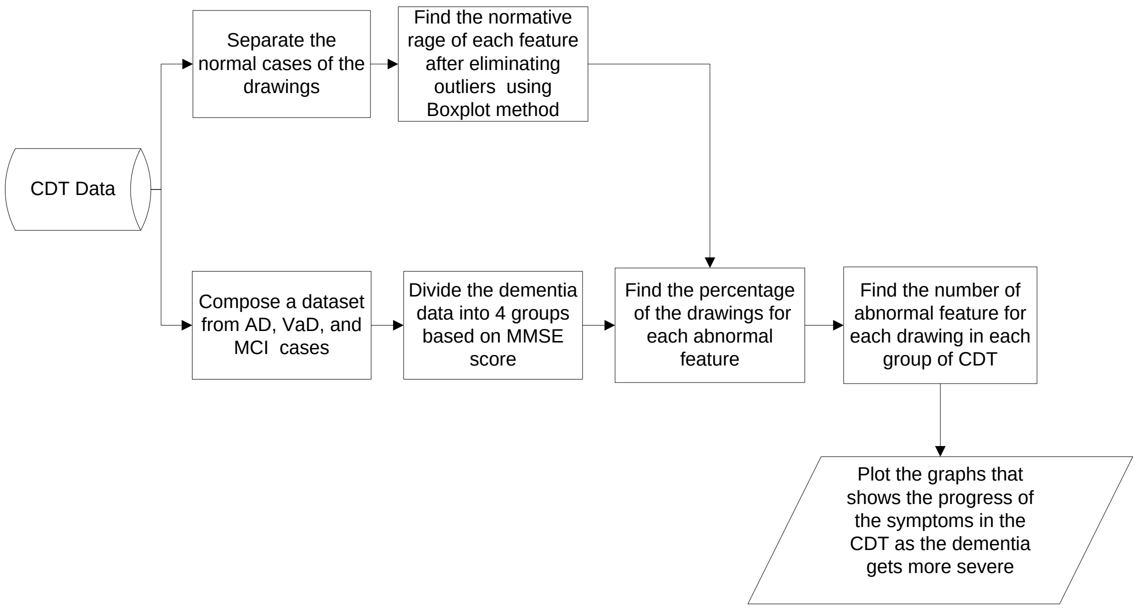


Figure 6-9: Framework of the CDT symptom progression analysis.

Due to the low number of drawings within each individual MMSE score 'bin', instead of repeating the process at each MMSE score, the range of scores is divided into four groups according to the cutoff levels of the NICE guidelines (27-30: Healthy; 21-26: Mild dementia; 11-20: Moderate dementia; and 0-10: Severe dementia). Because these guidelines are used to group dementia or MCI cases, the healthy group will be named 'very mild dementia and MCI'. Figure 6.10 shows the distribution of the drawings over these four groups. The percentage of the drawings whose features are abnormal is calculated for each group, however, the severe group is not included in further analysis due the low number of drawings attributed meaning that the findings for this group cannot be generalised. Most of the cases which are in the MMSE range of 'Normal' belong to VaD and MCI, with only 7 drawings from AD.

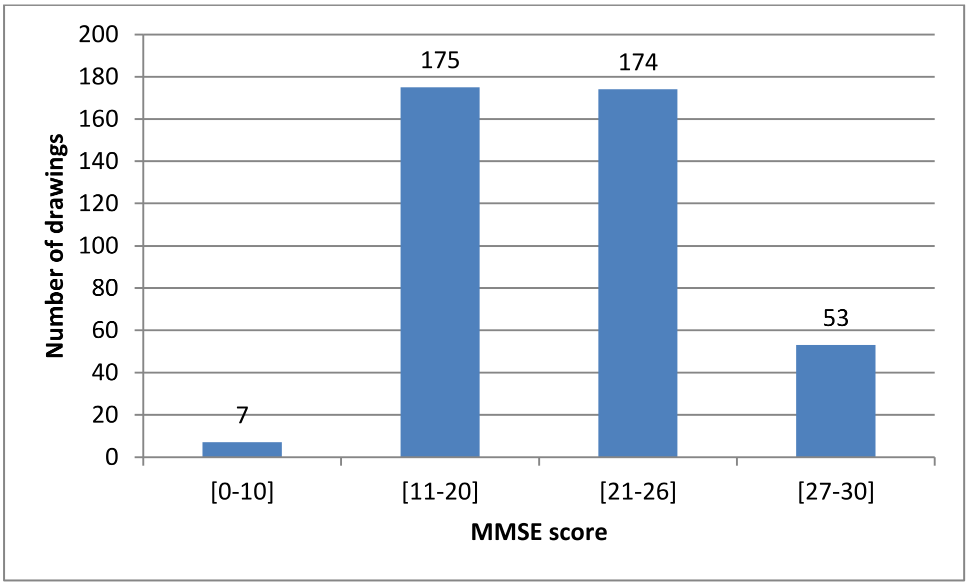


Figure 6-10: Distribution of all abnormal drawings over the MMSE groups

Finally, the number of abnormal features per single drawing is calculated for each drawing in each MMSE group. The results of this analysis framework are discussed in the following section.

6.3.1 **Results for Temporal Changes in the CDT Features Analysis**

The experiment is conducted using all the abnormal CDT drawings. Figures 6.11, 6.13, and 6.15 show the frequency (percentage) of each feature occurring within the normal and abnormal groups. The bars in the figure indicate the frequency of abnormal features.

Kitabayashi et al. (2001) studies the effect of the progress of dementia on clock drawings. The authors used data consisting of AD and VaD drawings only. The qualitative errors proposed by Rouleau, et al. (Section 2.3.4) are used. The authors divided the data in terms of severity based on intervals of MMSE score different of the ones used in this research. The study has been conducted on a low number of drawings, (45 drawings) and therefore the generalisation of the findings is questionable.

Figure 6.11 shows that most of the features are within the normative range for most of the drawings of the MMSE group 27-30 (very mild dementia). However, the figure still shows that some features are abnormal, with reasonable frequency.

‘Arrows on hands’, and ‘presence of stem of clock hands (near to the center) is left out’ are shown as abnormal features with a frequency of 62 % and 52 % respectively. Features such as ‘the count of numbers within area 1’, and ‘the count of numbers within area 2’ are also shown to be abnormal with a frequency greater than 40 %. Besides these features, those related with the count of numbers within quadrants 3 and 4 are also abnormal, with a frequency of more than 30 %.

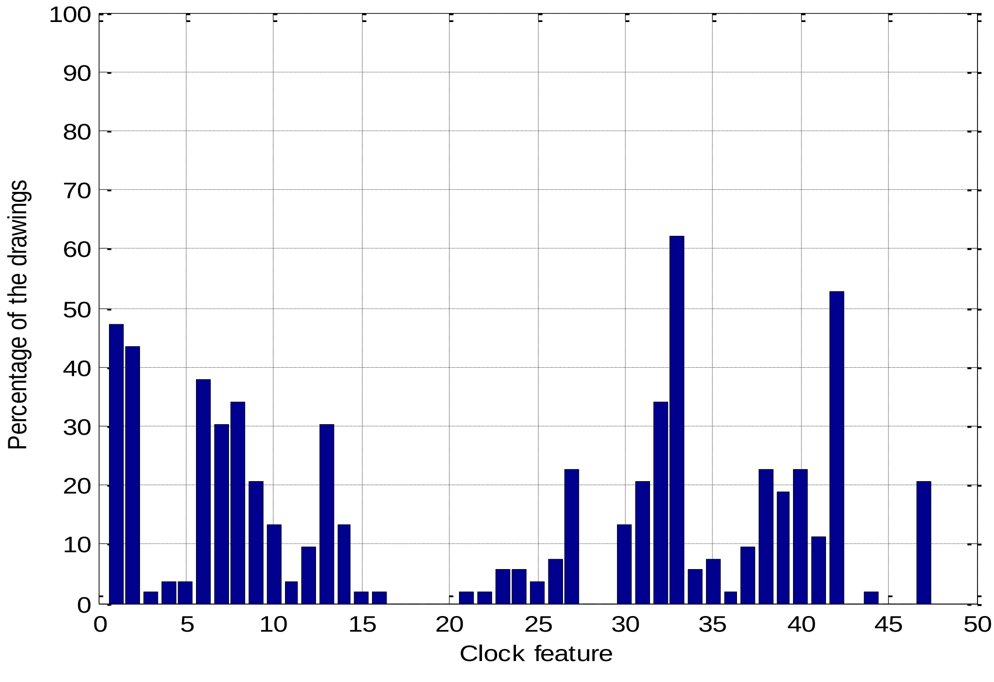


Figure 6-11: Percentage of drawing features being normal and abnormal MMSE group (27-30).

Figure 6.12 shows the number of abnormal features in each drawing. The figure shows that the majority of the drawings contain between 4 and 9 abnormal features, with few drawings containing more than 9 or fewer than 4 abnormal features.

It can be concluded that patients in the early stage of dementia tend to not draw the arrows on the hands, and are also likely to draw the hands without connecting them at the center. Moreover, these patients might face problems with the number spacing, such as writing the numbers far from the perimeter of the clock. The number of incorrect features that these patients are likely to make is generally between 4 and 9 for each drawing.

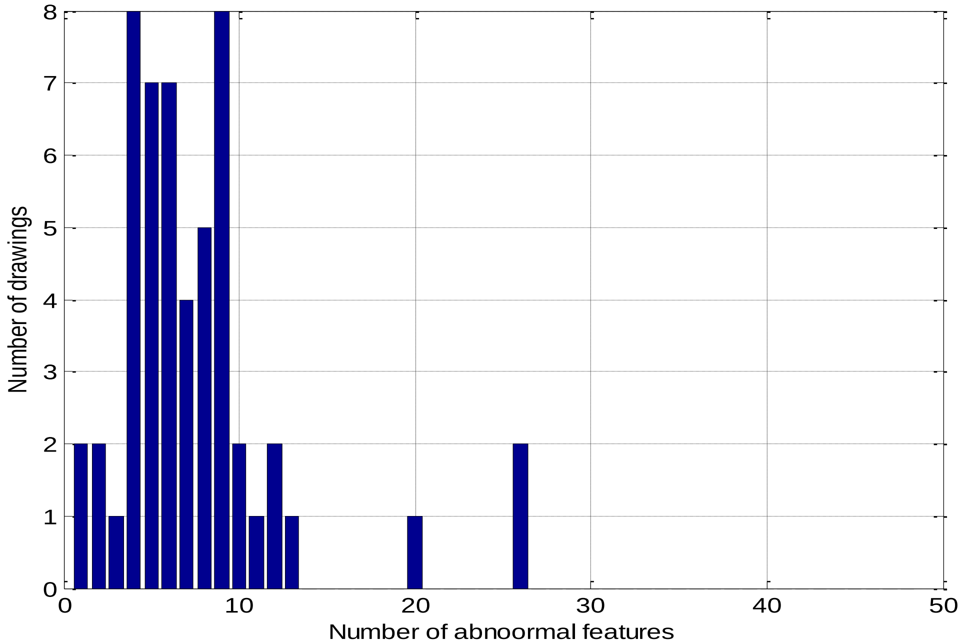


Figure 6-12: Distribution of the numbers of drawings containing certain quantities of abnormal features for MMSE group (27-30).

As the dementia develops the clock starts to show more abnormal features. Figure 6.13 shows that the frequency of abnormal features begins to increase for the MMSE group 21-26 (Mild dementia).

Besides increasing the frequency of the features that are shown as abnormal (as mentioned for the previous group), features related with time setting and the clock center also began to be abnormal for more than 50 % of the drawings in this group. These features include: angle between clock hands; distance between the position of hands intersection and the center of the clock; and position of minute hand.

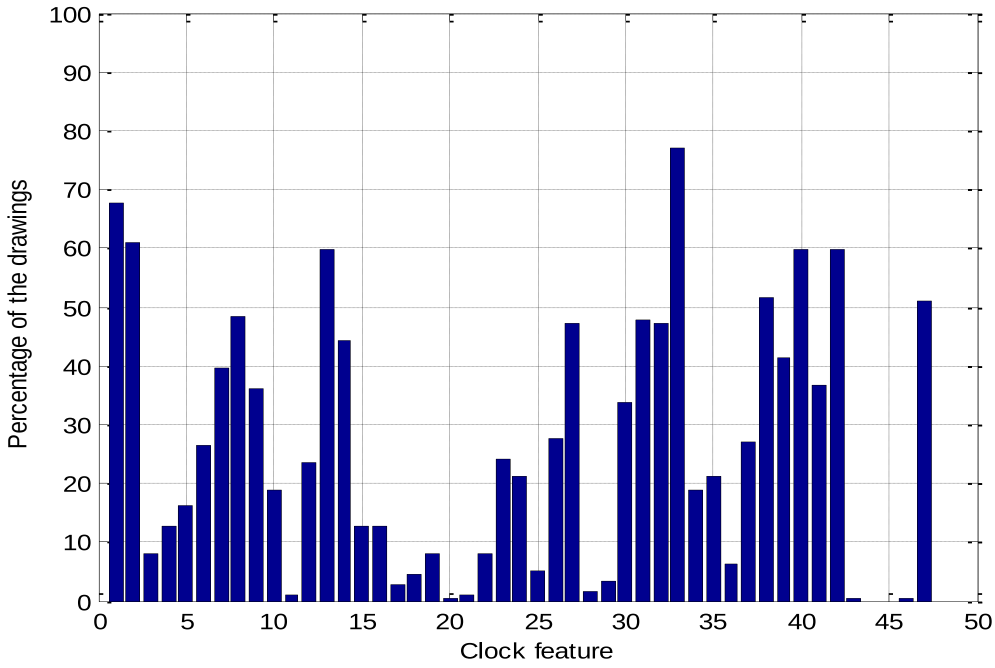


Figure 6-13: Percentage of drawings with normal and abnormal features MMSE group (21-26).

Figure 6-13 also shows that patients can experience difficulties writing the numbers in their correct positions. Features such as ‘count of numbers within area 1’, ‘count of numbers within area 2’, and ‘maximum angle between numbers’ are shown as abnormal in more than 60 % of the drawings. The figure shows that more than 47 % of patients have difficulty in setting the time correctly because of the difficulty in setting the minute hand at the right number. In addition, more than 40 % of the drawings contain numbers with abnormal size and rotated numbers.

Figure 6.14 shows the number of abnormal features in each drawing. The figure shows that as the dementia progresses the number of abnormal features per drawing increases. The majority of the drawings contain between 4 and 16 errors. The figure also shows that some drawings contain significantly more than that.

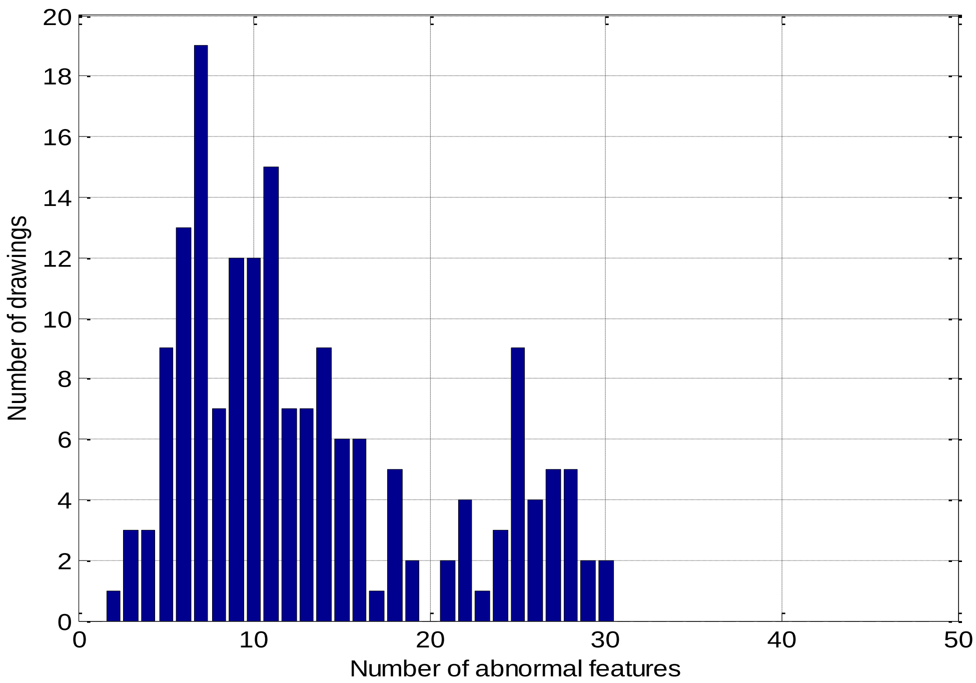


Figure 6-14: Distribution of the number of abnormal feature per drawing for MMSE group (21-26).

In conclusion, at this stage of dementia the defects in the clock start to be more recognisable, and the frequency is also increased. Besides the increase in the frequency of the features which had become abnormal at the previous stage, the patient might also fail to set the time correctly because of difficulty in setting the minute hand to “11”. The number spacing has also become more abnormal, numbers can be written with abnormal size, and the patient might tend to rotate the paper, resulting in rotated numbers.

As the dementia becomes more moderate, additional features become abnormal in a greater number of drawings. Figure 6.15 shows the frequency of abnormal / normal features in the drawings for the MMSE range 11-20 (Moderate dementia).

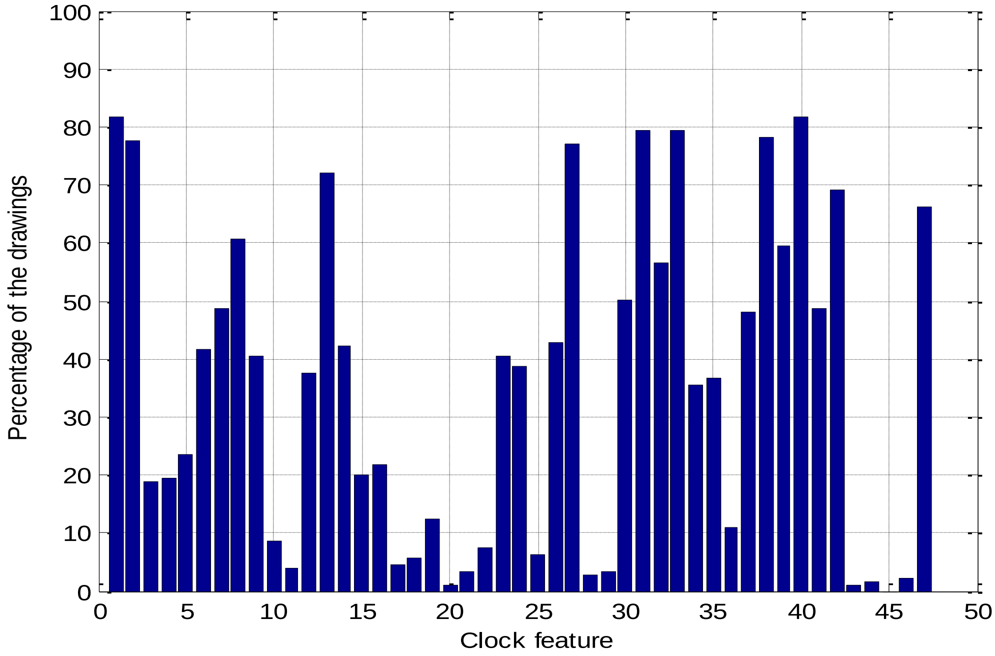


Figure 6-15: Percentage of drawings with normal and abnormal features MMSE group (11-20).

The figure shows that the same features which are seen to be abnormal in the previous MMSE group are still shown as abnormal but in a greater number of drawings. 81 % of the patients in this group fail to write the numbers in the correct position, close to the perimeter. About 80 % of the patients face problems in setting the time correctly. Of these, about 40 % is due to the absence of hands. 60 % of the drawings contain numbers of abnormal size.

Figure 6.16 shows the number of abnormal features in each drawing for MMSE groups 11-20. The figure shows that as the dementia progresses the number of abnormal features per drawing is increased. The majority of the drawings contain between 7 and 29 abnormal features.

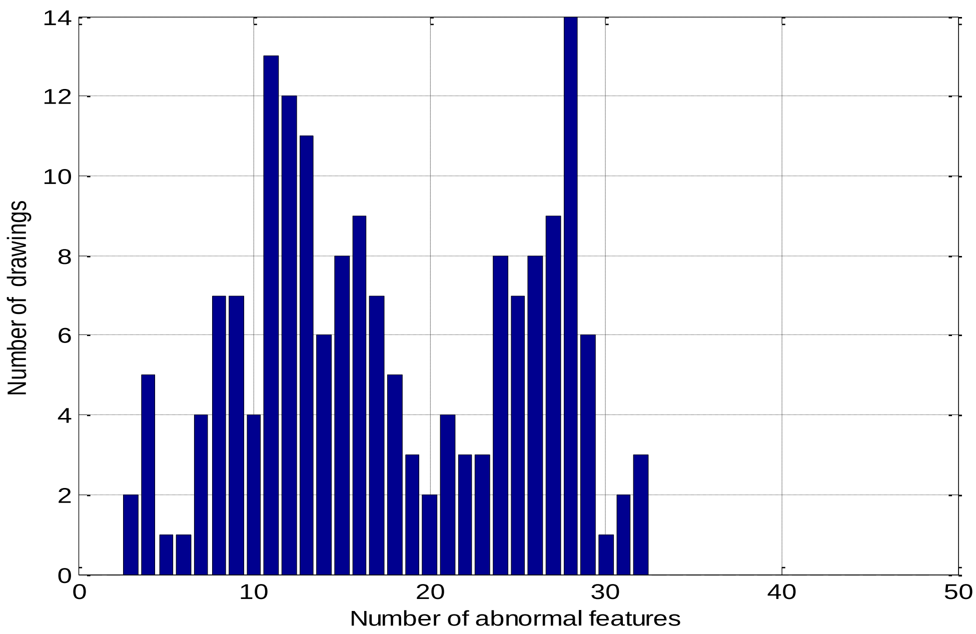


Figure 6-16: Distribution of the number of drawings over the number of abnormal features per drawing for MMSE group (11-20).

To summarise, Figure 6.19 portrays the common symptoms visible in the clock drawings which indicate the progress of dementia. The symptoms might start at the very early stage, with missing arrows and unconnected hands at the center, and might also include numbers written too far from the perimeter.

As the stage of severity develops into mild dementia, in addition to the previous symptoms, the patient might also have a problem in drawing the hands normally. For example, an abnormal angle or incorrect minute hand position might be shown, however the time is still likely to be largely correct. The clock might also show abnormal number sizes, and the defect in the number positioning can become more frequent. Writing of numbers far from the perimeter becomes more frequent, and the number spacing is likely to be abnormal (e.g. the numbers are crowded close to each other).

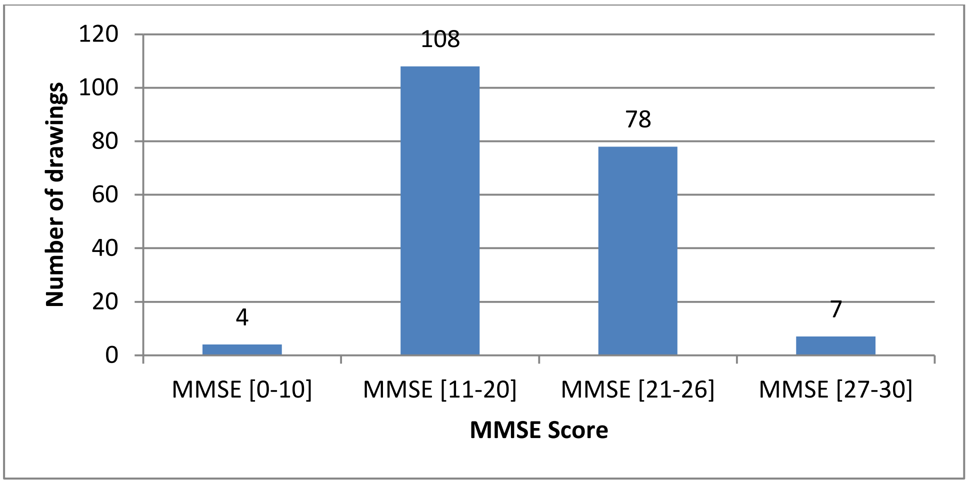
The symptoms become more frequent at the moderate stage. Besides increasing the number of abnormal features per drawing, some new symptoms are shown such as: an inability to set the time correctly; missing hands; more frequent defects in number spacing, especially writing numbers close to the center; and the presence of abnormal number size.

6.3.2 Temporal Changes in the CDT Features Corresponding to the Progress in AD and VaD Dementia

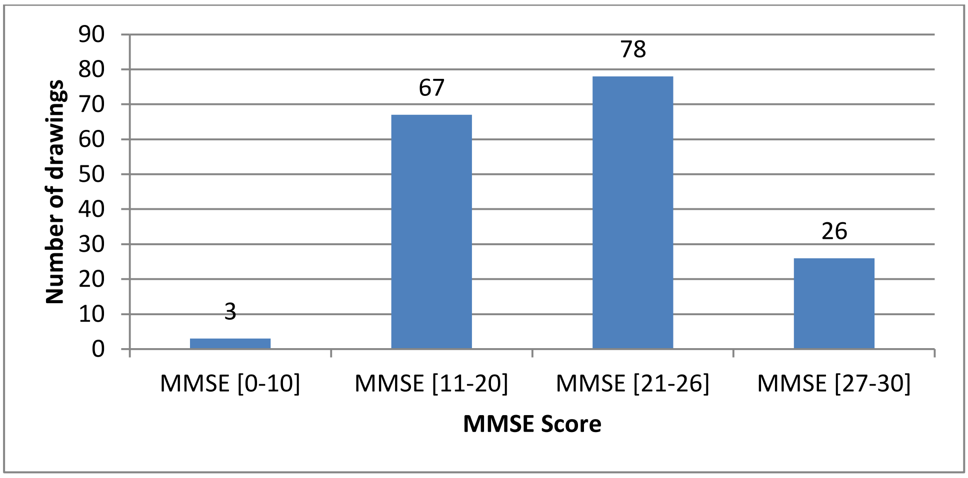
In this section, the progress of the symptoms in the clock drawings by sufferers of AD and VaD dementia is analysed. Previously, Kitabayashi, et al(2001) have reported that the VaD patients tend to perform better than AD patients during very mild stage especially with conceptual and spatial / planning deficits . During the mild stage, the performance of VaD patients deteriorates steeply for both deficits, while the performance of the AD patient on the spatial / planning deficits improves. Surprisingly their results show that during the moderate stage VaD patients show a large improvement in spatial / planning deficits, and deterioration in conceptual deficits. AD patients at this stage perform worse than at the other stages.

Employing a relatively large number of drawings to analyse the progress of the abnormality of the features as AD and VaD diseases develop will give the opportunity to compare the results to those reported in the literature.

The distribution of the drawings over the MMSE groups for AD, and VaD dementia is shown in figure 6.17.



a



b

Figure 6-17: Distribution of drawings over the MMSE groups (a) AD dementia, (b) VaD dementia**.**

Due to the low number of drawings by patients with very mild and severe cases, the analysis will be conducted on the mild and moderate dementia groups only. Figure 6.18 shows the frequency of the abnormal features for mild cases of AD and VaD diseases.

The figure shows almost no difference between the two types of dementia. However, the frequency of some abnormal features is slightly higher for AD than for VaD. This is because the CDT is not capable of diagnosing the specific cause of dementia; rather, it is designed to demonstrate the deficits that happen due to the impairment (Freedman et al., 1994; Mainland, and Shulman, 2013). The result is very similar to that presented in the previous section for all types of abnormal drawings.

The results do not agree with those reported by Kitabayashi, et al. (2001) as the frequency of some abnormal features considered as spatial / planning features increased from 70 % to 83 % while the previous authors state that they are decreasing.

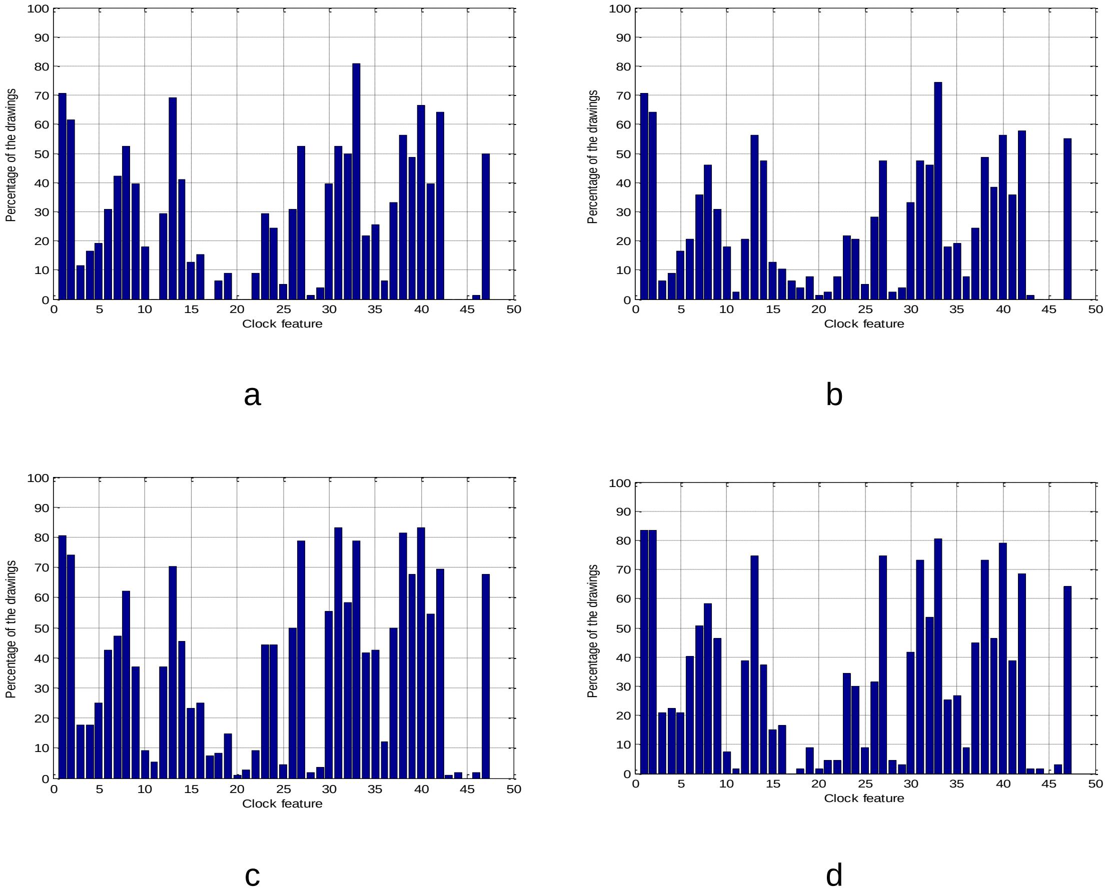


Figure 6-18: Percentage of drawings with normal and abnormal features, (a) mild AD, (b) mild VaD, (c) moderate AD, (d) moderate VaD.

MCI and very early stage of dementia

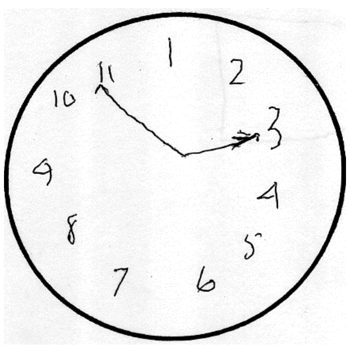
1- Arrows are not present.
2- Presence of stem of clock hands.
3- Writing numbers far from the perimeter

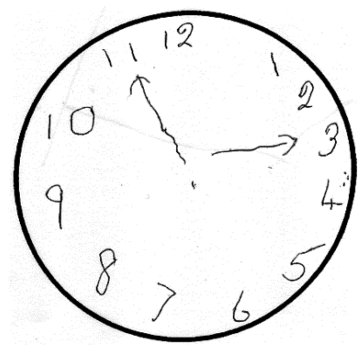
Early stage of dementia

1- Problem with features related to time setting (angle between hands, or position of hands), but likely still able to set the time correctly.
2- Obviously writing numbers far from the perimeter.
3- Obvious problem in spacing the numbers.
4- Rotating some numbers.
5- Size of the numbers could be larger than normal.
6- An increasing number of abnormal features per drawing.

Moderate stage of dementia

1- The time is more likely to be wrong (Could be because of missing one hand or both, wrong hand position or angle is wrong).
2- More obviously writing numbers far from the contour.
3- More obvious problem in spacing the numbers.
5- The intersection of hands far from the clock center.
6- An increasing number of abnormal features per drawing.





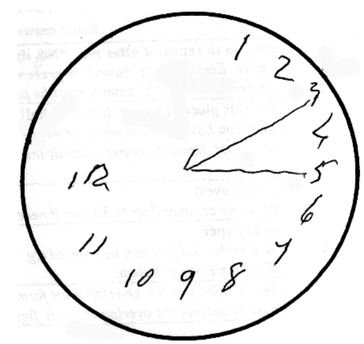


Figure 6-19: Summary of the progress of the symptoms on the clock, with examples of clock drawings

6.4 **Summary**

This chapter proposes a new approach to enable the determination of a list of features which is significant for discriminating between different cognitive states. In contrast to the previous two research studies (Lessing et al. 2008; Jouk and Tuokko, 2012) which defined the significant CDT features for only the normal vs. abnormal discriminative task, the proposed approach (following medical experts' advice), allows one to determine the significant feature subset for discriminating between Normal+ vs. MCI+, Severe and moderate vs. MCI+, Normal+ vs. Abnormal, Normal vs. Functional, and AD vs. VaD. These significant subsets will be used in the diagnostic stage (next chapter) to enhance the performance of the classification process. The subsets also help to explain the relationships between the abnormality of certain features and the discrimination between cognitive statuses. The degree of relation between each individual feature and the diagnosis is defined via calculation of the mutual information shared between each feature and the diagnosis for each discriminative task. The selected subsets show that the new proposed features are significant in terms of discriminative power. They also show that features related to time setting, clock hands and number spacing are significant.

This chapter further proposes a framework for analysis of the temporal changes in the CDT features corresponding to the progress of dementia from very mild +MCI to moderate. In contrast to the previous research the proposed framework uses detailed features and a dataset of 604 clock drawings. The framework also determines the normative range of each feature which is used to identify the abnormal features whose values lie outside the normative range. The previous research involved a low number of drawings and very broad qualitative features with no normative base for these features.

The results show that missing arrows on hands, unconnected clock hands at the center, and numbers written far from the clock perimeter could be signs of early dementia or MCI. They also show that as dementia develops the frequency of abnormal features increases accordingly. Furthermore, there is no significant difference between the deficit frequency of AD sufferers and VaD sufferers, and As the dementia develops, the frequency of deficits increases, which is in disagreement with previous literature (Kitabayashi et al. 2001).

The diagnosing stage is discussed in the next chapter; the result of the feature selection stage is employed to enhance the classification accuracy. The chapter also introduces a new cascade classifier which is used to classify the CDT drawings into one of three classes.

Chapter: 7

Diagnosis Stage

The clock drawing test is classified to diagnose abnormalities in a subject's cognitive abilities. The produced drawing is classified into one of the two diagnoses (classes), normal or abnormal. Diagnosing early stage dementia or MCI is one of the aims of this research. Using CDT alone to diagnose these two cognitive status is a very challenging task. The test is reported in literature as a tool which is not suitable for diagnosing MCI (Pinto and Peters, 2009). However, this conclusion is made based on results collected using the available scoring systems. To make the CDT more sensitive in diagnosing MCI and early dementia cases, a very detailed comprehensive list of features is used to digitise the CDT drawings (Chapter 4).

The contribution of this chapter is a novel classification algorithm based on cascade classification, which is introduced as a diagnosis stage in the proposed conceptual model. The classification algorithm aims to assign the CDT drawings to one of four diagnoses (Normal, Functional, MCI+, Moderate and severe). Classifying CDT drawings into one of these diagnoses would provide a new scope for the CDT as it has not been used before in this manner. The performance of the proposed classifier in discriminating between the diagnoses is also considered as an appropriate evaluation of the whole proposed CDSS-DD, as explained in Chapter 3.

The decision to propose this cascade classifier is based on:

1. The classification process of cascade classifiers is very similar to the process that doctors follow to make diagnoses. It is a multistage decision process, in which doctors reject the possible diagnoses, one by one, until the most probable diagnosis remains.

2. The cascade structure is also proposed as it benefits from the significant feature subsets which are defined in Chapter 6 for each discriminative task. Each classifier within the cascade structure will be employed to discriminate between two cognitive states, with only the significant features for that discriminative task being used for the training and classification task. Using fewer features can enhance the performance of the learning algorithm and reduce the elapsed time.

Many supervised classification algorithms have been discussed in Chapter 2. The Support Vector Machine (SVM) and Random Forest (RF) algorithms have gained attention recently as they reportedly produce a remarkable performance with diverse applications (Bhattacharyya et al., 2011). These two algorithms are hence nominated for use in the cascade classifier. K Nearest Neighbors (KNN) is an instance-based learning algorithm, which is selected as a baseline algorithm in the comparative study between SVM and RF

A comparative study is conducted to choose the best classification algorithm for each classifier in the cascade system. The Support Vector Machine (SVM), and Random Forest (RF) algorithms are tested separately. Classification accuracy, sensitivity and specificity are used as performance measures.

The performance of the proposed cascade algorithm is compared against a single stage classification and also against the performance of the dementia specialists in the literature.

This chapter is organised as follows. Section 7.1 presents the results of the single stage classification in diagnosing CDT drawings; Section 7.2 proposes new cascade configurations to diagnose the cognitive status based on CDT. Finally, Section 7.3 summarises the chapter.

7.1 Single Stage Clock Drawing Classification

One of the aims of this research is to diagnose unlabeled clock drawings based on previous experience from the labeled data. Classification algorithms are utilised for this task. The classification algorithms are trained and tested using the clock drawing data that has been extracted from the CDT drawings. The drawings are classified into one of the four diagnoses (Normal, Functional, MCI+, and Moderate and severe) as mentioned in Chapter 6.

In the single stage approach SVM, RF, and KNN algorithms are employed individually. Five folds cross-validation is chosen to train and test the algorithms. The performance of the classifiers is assessed when the drawing is classified into one of three diagnoses (Normal+, MCI+, and Moderate and severe), and when classified into one of four diagnoses (Normal, Functional, MCI+, and Moderate and severe).

The parameters of each classifier are tuned to get the best performance. The KNN algorithm is evaluated using three different values of K ($K = 3$, $K = 5$, and $K = 7$). The RF is also evaluated with a different number of sub-trees (100, 200, and 500). The linear kernel and Gaussian kernel are used with the SVM. The Matlab Statistics Toolbox is used for the KNN classifiers, while the Matlab code of Abhishek (2009) is employed for the RF classifier. The Lib-SVM code (Chang and Lin, 2011) is employed for SVM classifier. Lib-SVM is an implementation of SVM that is able to deal with multiclass datasets.

The experiment is conducted twice: once using data from three classes (MCI+, severe/ moderate dementia, and normal+); and another time using data from four classes (MCI+, severe/moderate dementia, normal, and functional).

Results

Table 7.1 shows the classification accuracy of the KNN classifier. The table shows the results of classifying the drawing into one of the three diagnoses, and into one of the four classes. The table shows that best classification accuracy is at K=7 with accuracy of 68.91 % for the three, and 65.65 % the four classes.

Table 7-1: Classification accuracy of single stage classification using KNN classifier.

Number of classes	Classification accuracy		
	K=3	K=5	K=7
Three diagnoses	66.85 %	66.88 %	68.91 %
Four diagnoses	64.88 %	65.56 %	65.65 %

The results of the RF classifier show that the performance of RF is superior to KNN. Table 7.2 shows that the maximum average accuracy is 71.42 % when classifying the drawings into three classes, and 68.54 % for four classes. The table also shows that increasing the number of sub-trees in the RF does not significantly improve the performance, which agrees with the findings of Oshiro et al. (2012).

Table 7-2: Classification accuracy of single stage classification using RF classifier.

Number of classes	Classification accuracy		
	Sub-trees = 100	Sub-trees = 200	Sub-trees = 500
Three diagnoses	70.90 %	71.30 %	71.42 %
Four diagnoses	68.62 %	68.73 %	68.54 %

The table 7.3 shows the average classification accuracy of the SVM classifier with the kernels: Linear, and Gaussian. It can be seen that the produced accuracy is better that for RF. However, Lib-SVM is a very slow algorithm. Moreover, optimisation of the classifier parameters is needed during every training iteration, which leads to further delay. Classification using SVM with linear kernel takes 10 times the duration of RF classification with 100 sub-trees, while SVM with a Gaussian kernel takes 100 times that duration. For these reasons SVM is not the preferred choice for this task.

Table 7-3: Classification accuracy of single stage classification using SVM classifier.

Number of classes	Classification accuracy	
	Linear kernel	Gaussian Kernel
Three diagnoses	74.11 %	72.61 %
Four diagnoses	69.63 %	72.18 %

7.2 **The Proposed Cascade Classification**

In this section a novel cascade classification system is presented. it consists of three classifiers, which are connected sequentially. Figure 7.1 shows the block diagram of this system in which the classification is conducted in two stages. In the first stage, Classifier 1 discriminates the drawings into normal+ and abnormal cases. This classifier is trained using the entire dataset after it has been rearranged as binary classified data in which all normal and functional cases are considered as one class (nonorganic cases) called normal+. The other organic diseases are isolated as MCI+ and severe / moderate dementia. Regardless of the cause of the dementia (AD or VaD), these are considered to be one class, referred to as ‘abnormal’.

The second stage employs two more classifiers. The first (classifier 2) is used to differentiate the MCI+ diagnoses from severe and moderate dementia. Following advice from medical consultation, the scores from questions 16, 17 and 18 in the MMSE dataset are added to the clock features of the training and testing data of this classifier to enhance its performance. The questions are the last three of five attention and calculation questions, in which the patient is asked to subtract 7 from 100, 5 times consecutively (i.e. 93, 86, 79, 72, 65) (Folstein et al., 1975). Classifier 3 is used to differentiate between normal and functional cases. This classifier is trained using the part of the dataset containing only these two classes.

For each discriminative task in the cascade, only the significant feature subsets are used in the training and testing instead of including all 47 features. Using fewer features can improve the performance and reduce the classification complexity and time.

Discretisation is required during the training phase because some of the clock features are not discrete. The same bins employed during the training are used for discretisation during testing in order to maintain consistency among the continuous features. The training of the classifiers, with 5-fold cross-validation, is performed according to the following algorithm:

Algorithm 7.1: Training of the cascade classifier

```
INPUT: dataset Tr with classes: MCI+, Normal, Functional, Severe
Split into 5 equal sets, Tr1, Tr2, Tr3, Tr4, Tr5
SFc1 is the significant features subset for discriminative task of the classifier 1
SFc2 is the significant features subset for discriminative task of the classifier 2
SFc3 is the significant features subset for discriminative task of the classifier 3
Tc1 is the training data for classifier 1
Tc2 is the training data for classifier 2
Tc3 is the training data for classifier 3
Tri is the testing fold at each iteration
t is an observation from training folds  (Tr\ Tri)
for i=1 to 5
   Tc1=Tr(SFc1)\Tri(SFc1)
   for all t ∈Tr\ Tri do
      If the class of the observation t is Severe and Moderate or MCI+ then
            Tc2=Tc2∪t(SFc2)
            Tc2=Tc2∪MMSE(Q16,Q17Q18)
      else
           Tc3=Tc3∪t(SFc3)
    end if
 end for
Train a classifier 1 using Tc1
Train a classifier 2 using Tc2
Train a classifier 3 using Tc3
end for
```

The proposed cascade initially classifies the data as normal+ and abnormal. It is then also able to classify the abnormal data into two classes, namely MCI+, severe/moderate dementia, as well as classifying normal+ data into normal and functional dementia cases.

7.2.1 **Comparative Study between Classification Algorithms**

A comparative study is conducted to choose the best classification algorithm for each classifier in the cascade system. Least squares Support Vector Machine (LS-SVM), and Random Forest (RF) are tested separately. Classification accuracy, sensitivity and specificity are used as performance indicators.

The RF is evaluated with a different number of sub-trees (100, 200 and 500). The Matlab Statistics Toolbox is used for the SVM classifier this time as all the classifiers in the system dealt with binary classified data.

The training and the testing is performed using the significant feature subsets which were selected in the previous chapter. The task of training and testing is done after adding every new feature to the subset; it starts with one feature until all the features are selected and added into the subset.

To select the proper discretisation method for each classifier in the structure, another comparative study is conducted in parallel with the previous study. Three discretisation methods are compared, including two unsupervised methods (Equal Width Discretisation (EWD) and Equal Frequency Discretisation (EFD)) and a state-of-the-art supervised method (Minimum Description Length (MDL)).

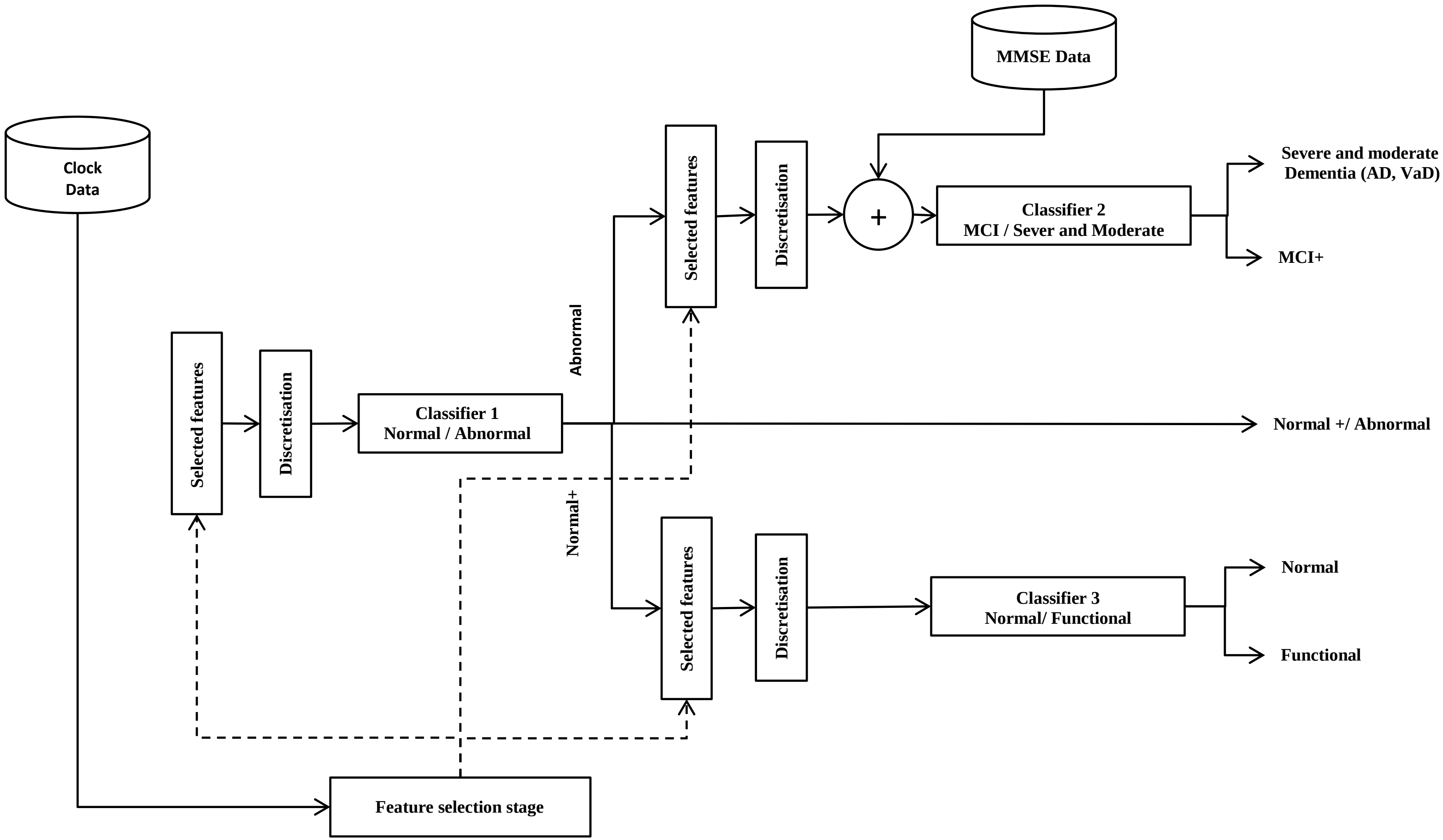


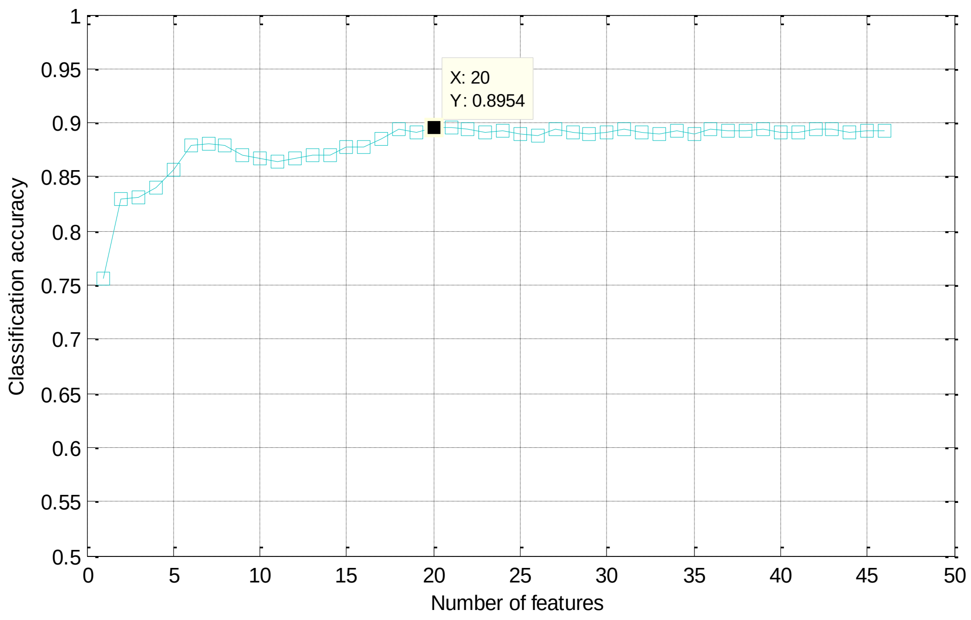
Figure 7-1: Two stage cascade classification system classifying the CDT drawings.

The discretisation method which produced the best classification accuracy, sensitivity, and specificity for a particular classifier test is chosen for that classifier. 10 intervals are used for EWD and EFD. Therefore the experiment is run almost 12 times to select the best classification algorithm and the best discretisation method for each classifier in the cascade structure. The performance of the classification algorithms and the discretisation methods is evaluated using 5-fold cross-validation 10 times.

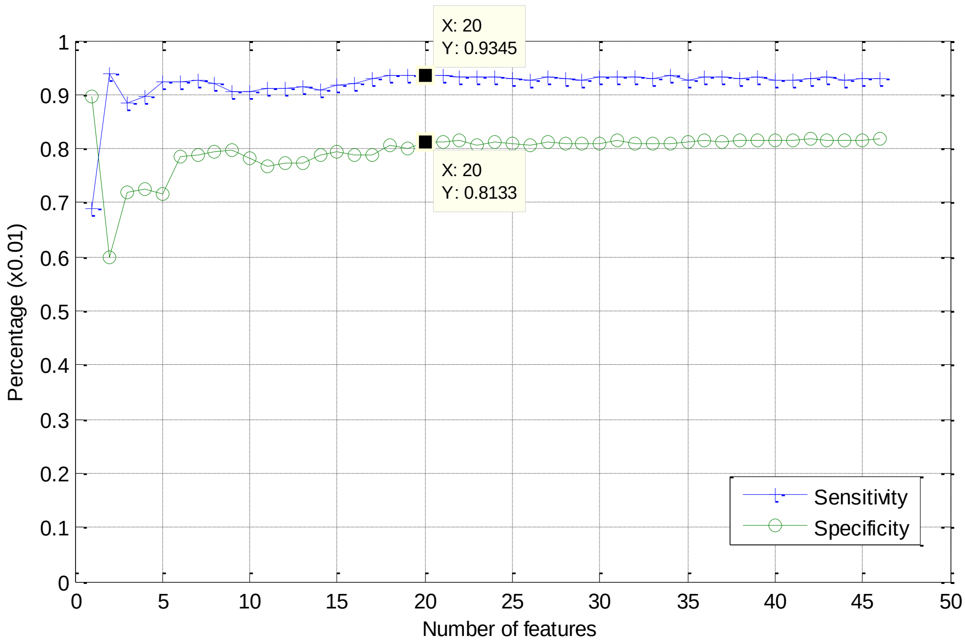
Results

The results of the comparative study for Classifier 1 show good performance of SVM,. However, the RF classifier produces the best performance with MDL discretisation method. The results of RF are consistent with the earlier explanation of the effect of the number of sub-trees on the performance of this algorithm. RF outperforms SVM in classification accuracy by at least 1.54 % and in sensitivity by at least 3.72 %. Figure 7.2 shows the classification accuracy, sensitivity, and specificity for classifier 1 when the RF algorithm with 100 sub-trees and MDL method is employed. The results of the rest of the comparative study are shown in appendix D.

The figure 7.2a shows the classification accuracy after adding the features to the subset one by one until all 46 features are included in the training and testing of the classifier. The accuracy of discriminating between normal and abnormal cases using all 46 features is 89.32 %. The figure also shows that using a sub-set of the most significant 20 features produces slightly improved accuracy, at 89.54 %, even though more than half the features are omitted. Figure 7.2b shows the sensitivity and the specificity of this classifier. The figure shows that the ability of the classifier to identify the positive cases correctly is 92.91 % using the 46 features, and 93.54 % using only the subset of the 20 most significant features. It also shows that the performance of the classifier to identifying the negative cases correctly is 81.79 % using 46 features, and 81.33 % using only the sub-set of 20 features.



a



b

Figure 7-2: Performance of classifier 1 in the cascade using RF with 100 sub-trees, and MDL: (a) classification accuracy, (b) sensitivity and specificity.

The performance of classifier 1, which classifies the drawings into normal+ / abnormal groups is much better than the performance of most of the scoring systems that are currently being used to score CDT and it is competitive even against some more detailed scoring systems. Table 7.4 shows the sensitivity and specificity of the scoring systems which have been reported as the best in literature. However, in literature theses scoring system showed a lower performance than the original studies (Pinto, Peters, 2009).

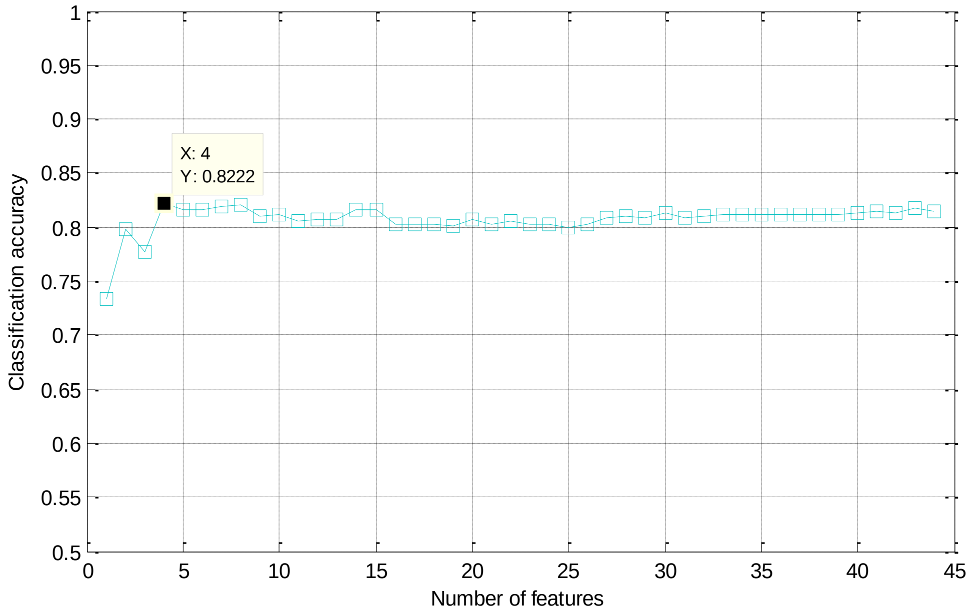
Table 7-4: Sensitivity and specificity of the best four scoring systems and Classifier 1 in the cascade.

The Scoring system	Sensitivity	Specificity	Number of cases
Sunderland et al	76 %	81 %	41
Shulman et al	86 %	72 %	75
Mends et al	73 %	77 %	46
Toukko et al	92 %	86 %	72
Classifier1	93.54 %	81.33 %	604

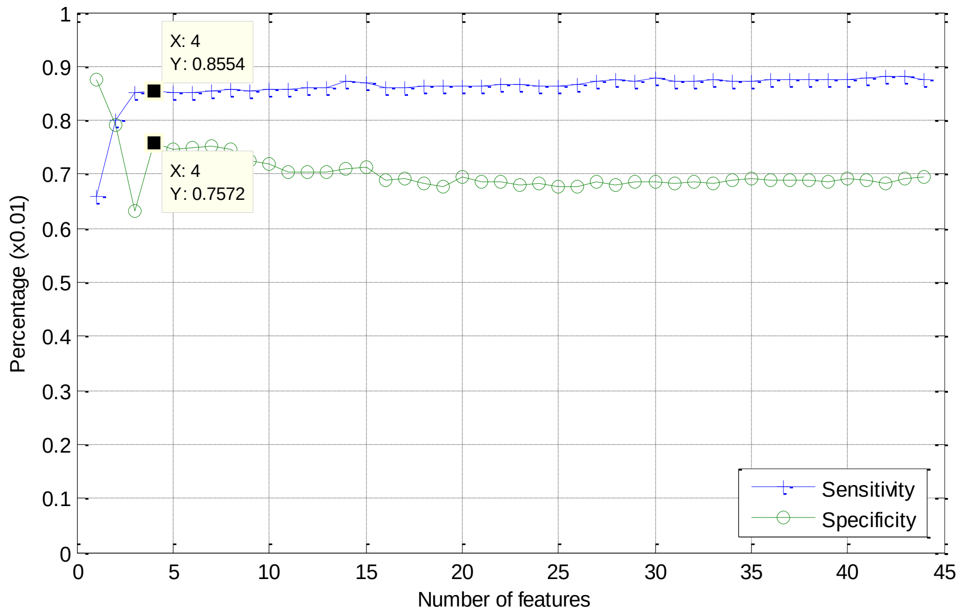
Table 7.4 shows that classifier 1 produces a better sensitivity than all the other scoring systems, which means a better performance in diagnosing the abnormal cases correctly. The specificity is higher than all of the systems except Toukko's system. However, the number of the cases that are employed in this research is much higher than the number of cases used in developing the scoring systems.

The results are then compared with the assessment of dementia specialists. The sensitivity and the specificity of this classifier 1 outperform the sensitivity and the specificity of the dementia specialists (Nair et al, 2010) by 32.45 %, 0.33 % respectively. The RF outperforms the SVM classifier when tested using the sub-dataset of classifier 2 in the cascade structure (MCI+, severe and moderate).

Figure 7.3 shows the classification accuracy of the RF classifier with 100 sub-trees along with the sensitivity and specificity. The figure shows that the produced accuracy using all 44 features is 81.44 % and the sensitivity and the specificity are 87.49 %, 69.57 % respectively. However, the classier achieves better accuracy with a subset of just 4 features, 82.22 %. The sensitivity (performance of the classifier in discriminating the positive moderate and severe cases) is 85.54 %, and the specificity (the performance of the classifier in diagnosing MCI+ cases correctly) is 75.72 %.



a



b

Figure 7-3: Performance of classifier 2 in the cascade using RF with 100 sub-trees, and the MDL discretisation method: (a) classification accuracy, (b) sensitivity and specificity.

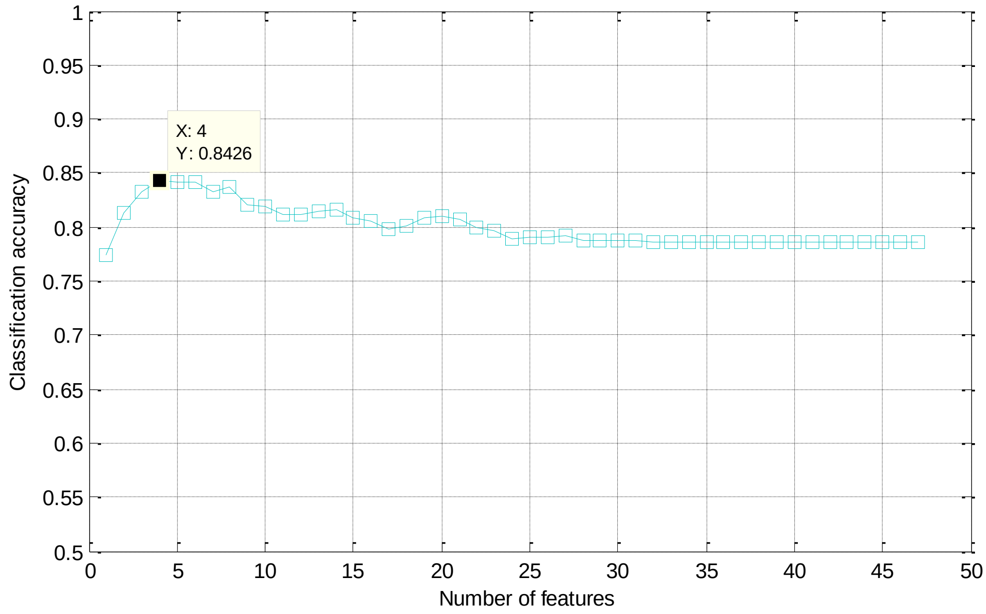
Existing CDT scoring systems are designed to discriminate between the normal and abnormal drawings, but are not able to differentiate the MCI and the early stages of dementia. However, the results of classifier 2 in the cascade show that the performance in discriminating between an MCI+ diagnosis, and a moderate /severe diagnosis is better than that of the dementia specialists which have been reported by Nair et al, (2010) (75 %, and 53 % for the average sensitivity and the average specificity).

Finally, the performance of Classifier 3 is tested. This classifier is designed to differentiate between normal and functional cases. MDL is not included in this comparison, since it fails to find appropriate cut-off points to discretise features, and hence removes many features.

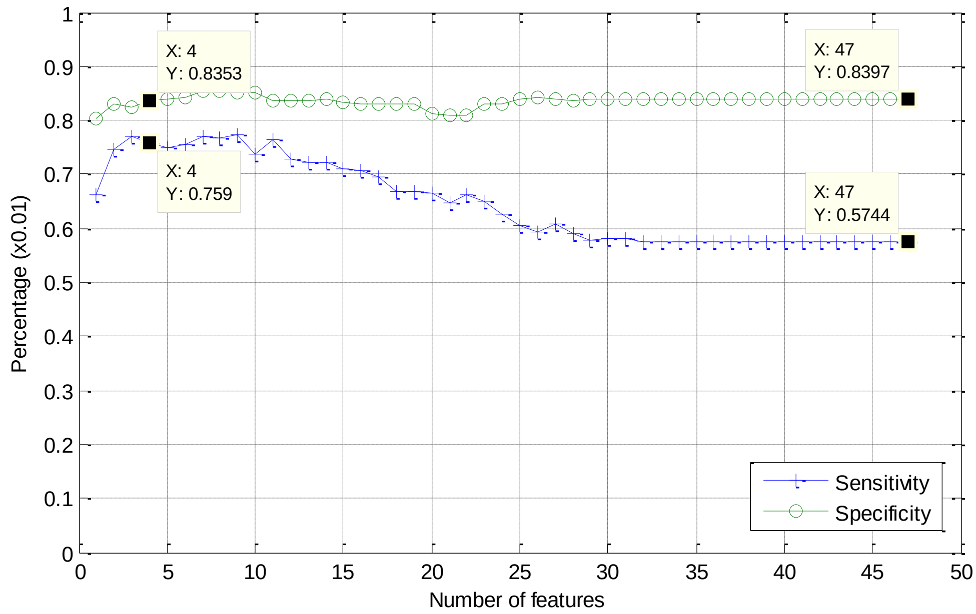
This part of the data is unbalanced, as the number of functional observations is low. Therefore, the minority class is given more weight than the majority, a ratio of 3 to 1 being used for this classifier. With this part of the data the SVM with EWD method produces the best performance, and it outperforms the RF in terms of the best tradeoff between sensitivity and specificity. The best performance of SVM is achieved by employing the most significant 4 features, producing a classification accuracy of 84.26 %, and sensitivity and specificity of 75.9 %, 83.53 % respectively. Figure 7.4 shows the results of the SVM with linear kernel and EWD method. Although the produced classification accuracy using all the features is 82.77 %, using all the features for this classifier produces lower sensitivity (46.41 %) and high specificity (91.68 %).

In conclusion, the comparative study shows that the RF classifier produces the best performance for classifiers 1 and 2 in the cascade structure. It also shows that the performance of the MDL discretisation method is better than EWD and EFD methods for these two classifiers. It also shows that SVM produces the best performance when it is employed for classifier 3 in the cascade with the EWD discretisation method.

Therefore, the RF classifier and MDL method will be used for classifiers 1 and 2, and the SVM classifier and EWD method will be used for classifier 3.



a



b

Figure 7-4: the performance of classifier 3 in the cascade using SVM, and EWD discretisation method: (a) classification accuracy, (b) sensitivity and specificity.

7.2.2 **Results of the Proposed Cascade Configurations**

In order to measure the performance of the proposed cascade classifier, it is trained and tested using the entire dataset, and two experiments are conducted: one using the data from three classes (MCI+, severe / moderate dementia, and normal+); and another using the data from four classes (MCI+, severe / moderate dementia, normal, and functional). The experiments are run using 10 times 5-fold cross-validation. The same folds of data in the case of four diagnoses are used to train and test the RF single stage classifier. This experiment is also repeated 10 times. The result shows that the proposed classifier differentiates between normal+ and abnormal cases with 89.32 % accuracy.

Table 7.5 shows the average classification accuracy of the single stage classifier and the proposed classifier in classifying the drawings into three diagnoses (MCI+, severe/moderate dementia, and normal+), and the accuracy of classifying them into four diagnoses (MCI+, severe/moderate dementia, normal, and functional).

Table 7-5: Classification accuracy of single stage using SVM classifier.

Number of classes	Classification accuracy	
	Single stage	Cascade classifier
Three diagnoses	71.0 %	78.34 %
Four diagnoses	68.58 %	75.07 %

7.2.3 **ANOVA Statistical Testing**

To determine the significance of the achieved improvement over the single stage, the results of ten runs of the cascade classifier and the single stage classifier are submitted to ANOVA statistical testing. Table 7.6 shows the ANOVA results, where P-value is the probability of the improvement to occur by chance, and MS is the mean square error.

The improvement is significant if the P-value is less than 0.05, which means that the improvement is unlikely to happen by chance. As the P-value is almost zero in both cases, the improvement of the proposed classifier over the single stage is significant when used to classify the clock drawing to one of three classes (MCI+, severe/moderate dementia, and normal+), and also when they are used for classifying the clock drawings in four classes (MCI+, severe/moderate dementia, normal, and functional).

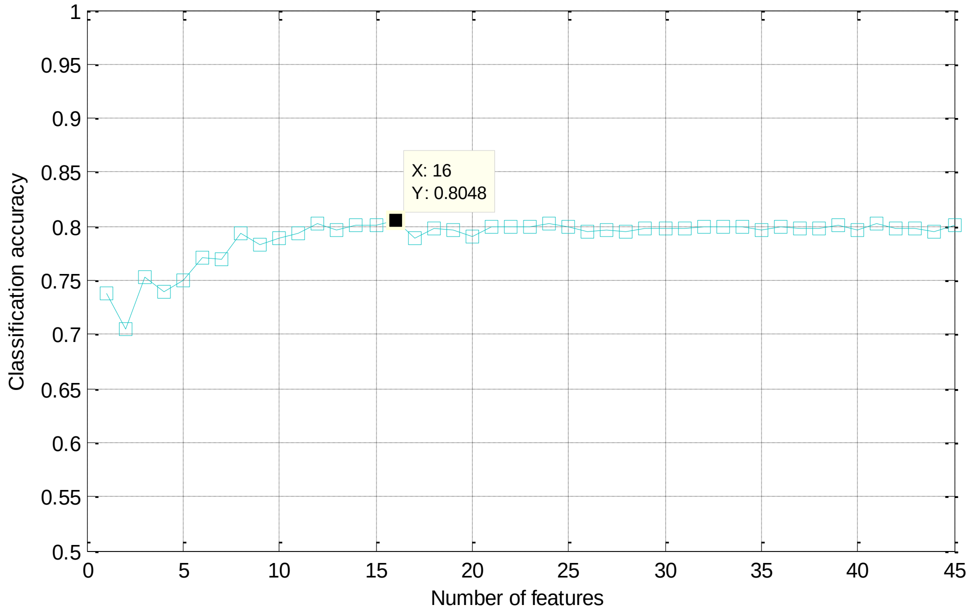
Table 7-6: ANOVA test**.**

Number of classes	Classification accuracy		
	F	P-value	MS
Three diagnoses	3306.57	7.43E-22	269.48
Four diagnoses	2261.47	2.22E-20	210.07

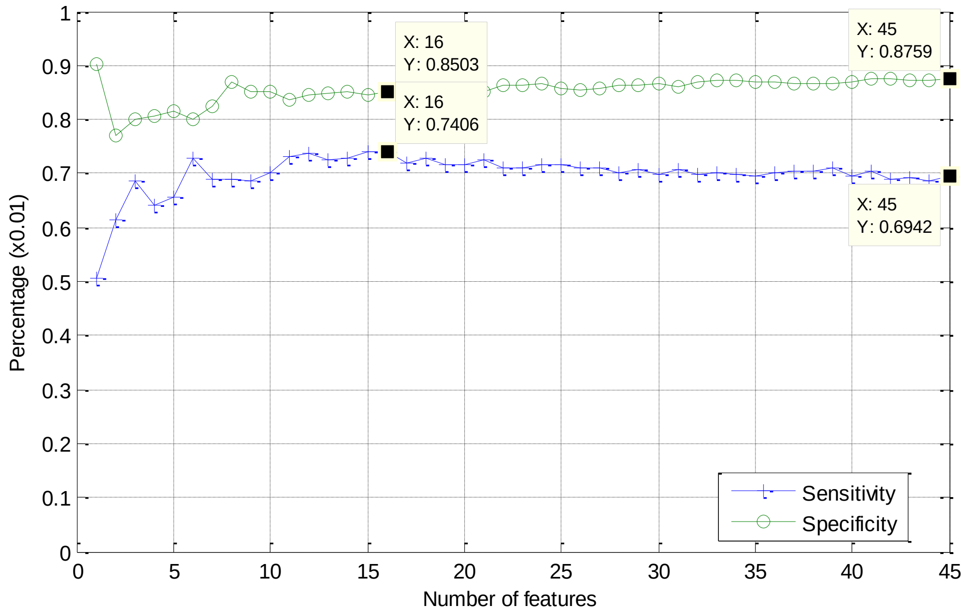
7.3 CDT Performance for Special Discriminative Tasks

This section evaluates the performance of the proposed conceptual model in discriminating between some cognitive statuses which are not included in the competitive study of the cascade classifier. A single stage RF classifier with 100 sub-trees is employed to test the performance of CDT in discriminating between MCI+ and Normal+, and also between AD and VaD in moderate and severe cases.

The first discriminative task is important because distinguishing subjects with early stage cognitive impairment from healthy individuals is one of the most challenging tasks. The second task is included in this experiment to assess whether the CDT is a good tool to distinguish between AD and VaD. In literature the CDT has been reported to be unable to discriminate between these two diseases.



a



b

Figure 7-5: Performance of discrimination between MCI+ and Normal+ using RF with 100 sub-trees, and MDL discretisation method: (a) classification accuracy, (b) sensitivity and specificity.

The classifier is evaluated using a dataset which contains the two classes which are subject to the discriminative task. 5-fold cross-validation is used in this evaluation. The MDL discretisation method is employed for the MCI+ / Normal+ discriminative task, and

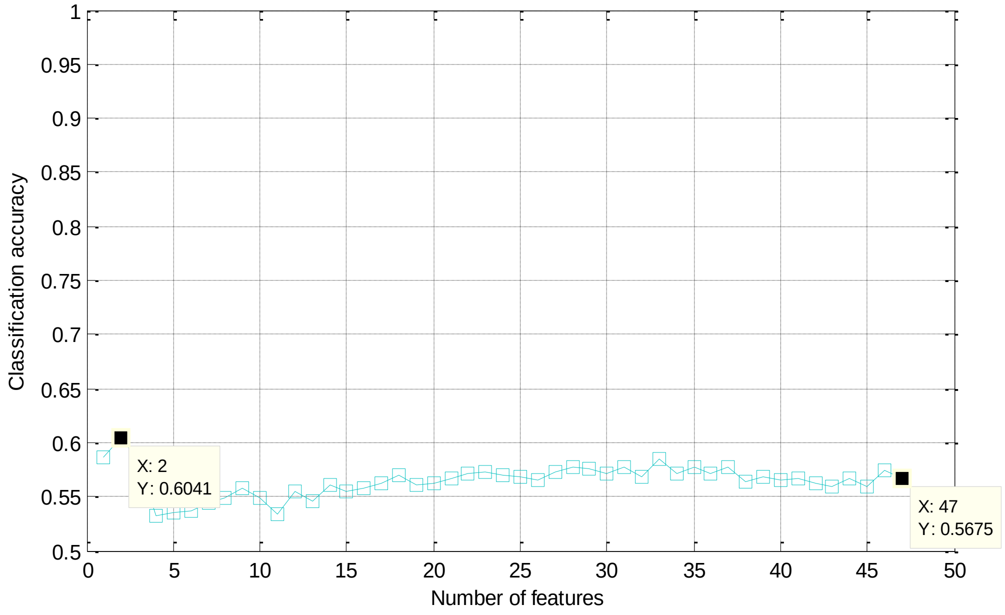
EFD for the AD / VaD discriminative task. The significant feature subsets are used in the classification. The training and the testing is performed after selecting every new feature and adding it into the subset until the full feature set has been included.

Figure 7.5 shows the classification accuracy, sensitivity, and specificity for the discrimination between MCI+ and Normal+ cases. The figure also shows that the best classification accuracy is produced with a subset of the 16 most significant features. The accuracy is 80.48 %. The sensitivity and the specificity are 74.06 %, and 85.03 % respectively.

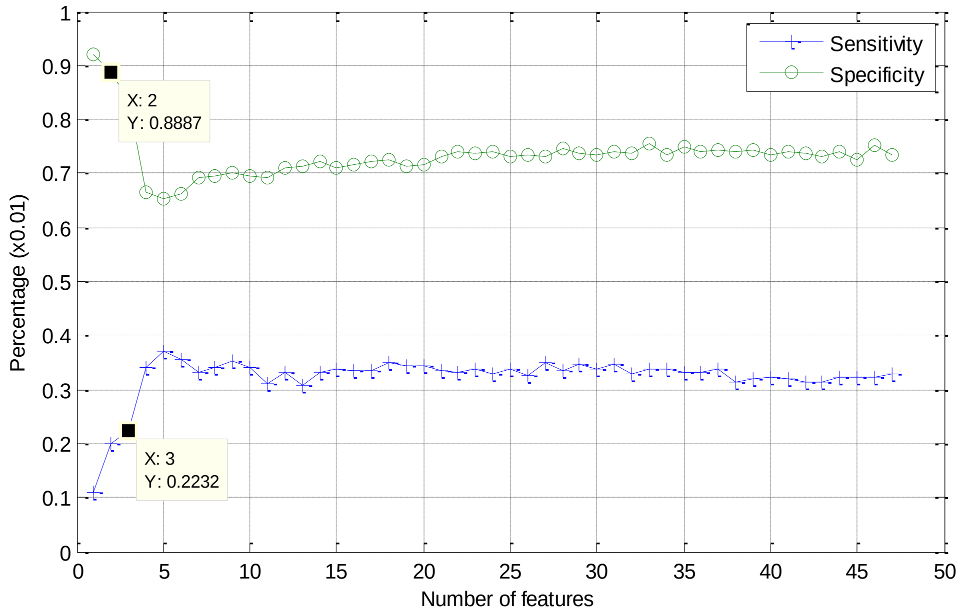
The approach that is used to classify the CDT drawing began with defining a comprehensive list of features, and go on to employ machine learning techniques. This approach outperforms even dementia specialists, with the reported sensitivity and specificity of specialists discriminating between MCI and Normal cases being 47 %, and 81 % respectively (Nair et al., 2010). This is below the performance of the computer based approach.

The performance of the CDT in discriminating between AD and VaD is tested using the same approach. An RF with 100 sub-trees is trained and tested using moderate and severe clock drawings datasets. EWD with 10 intervals is used to discretise the data.

Figure 7.6 shows the performance of the CDT on the AD / VaD discriminative task. The best accuracy (60.41 %) has been produced with the two most significant features. However, the sensitivity (the ability to diagnose VaD correctly) is very low (36.96 %), while the specificity (the ability to diagnose AD correctly) is relatively good (65.22 %).



a



b

Figure 7-6: Performance of the discrimination between AD and VaD using RF with 100 sub-trees, and the EWD discretisation method: (a) classification accuracy, (b) sensitivity and specificity.

The results show that the performance of the CDT in differentiating VaD from AD is very limited, a conclusion consistent with the literature (Freedman et al., 1994; Niures and Caramelli, 2010).

7.4 **Summary**

This chapter presents a novel cascade classifier for diagnosing dementia by classifying CDT drawings in the diagnosis stage of the proposed conceptual model for this research. The classifier is validated using 604 drawings which have been drawn by patients and healthy individuals. This number of drawings is relatively high compared with other similar studies. The proposed comprehensive list of features (Chapter 4) is used to digitise the clock drawing and produce a novel digitised CDT. In contrast to the previous research, the proposed system can assess the cognitive functions status with very good accuracy, including the MCI+ stage which is traditionally very challenging. It also employs a supervised learning algorithm to classify the drawings and make decisions about the abnormality, rather than more traditional scoring criteria.

Single stage classification is used to discriminate between four diagnoses (Normal, Functional, MCI+, Moderate and severe), and also between three diagnoses (Normal+, MCI+, Moderate and severe).

Three classification algorithms are employed individually to perform the single stage classification task. The results demonstrate that RF produces a good performance, and is as good as SVM, which is very slow due to the optimisation of the parameters. Furthermore, the results show that increasing the number of sub-trees in the RF does not increase the performance significantly.

A comparative study is conducted to choose the proper algorithm for each classifier in the cascade. The results for this comparative study show that RF and MDL produce the best performance for classifier 1 in the cascade. It is also shown that the performance is better that most of the CDT scoring system, and is also better than the ratings of some dementia specialists. RF and MDL are also the most suitable for classifier 2, with the performance of this classifier outperforming the dementia specialists' ratings for the CDT. Finally, SVM and EFD are found to produce the best performance for classifier 3.

The proposed cascade classifier benefits from the significant feature subsets which are defined in Chapter 6. The performance of the classifiers using these subsets showed an improvement over using the whole set of 47 features.

The proposed cascade classifier is compared with a single stage classifier. The results show a significant improvement in the classification accuracy. The improvement in classifying the drawings into one of three classes (Normal+, MCI+, Moderate and severe) is found to be 7.34 %, and 6.49 % when the drawings are classified into four classes (Normal, Functional, MCI+, Moderate and severe.

Chapter: 8

Conclusions and Future work

This chapter concludes the thesis. Section 8.1 lists the main contributions of thie research work; Section 8.2 provides a conclusion to the work that has been described in this thesis; Section 8.3 discusses the limitations of the research; and Section 8.4 discusses potential future work.

8.1 Contributions

The main contributions presented in this thesis are as follows:

- A conceptual model of a clinical decision support system is proposed for the early detection of dementia (CDSS-DD). The model is designed to support clinicians at the diagnosis stage of dementia. The system is based on the clock drawing test, and employs machine learning and image processing techniques to improve the sensitivity of the clock drawing test in diagnosing the cognitive disorder.

- Two new feature selection methods which are based on information theory are proposed and validated. The methods employ joint mutual information, symmetrical relevance, and 'maximum of the minimum' criteria to select a subset of significant clock features. One of these methods which has the better performance and outperforms the state of the art methods is used as part of the proposed CDSS.

- A new cascade (multistage) classification scheme which employs three classifiers in two stages, whereby RF and SVM algorithms are employed. The cascade classifier is used to enhance the performance of clock drawings tests by combining the MMSE with CDT to discriminate between MCI+ cognitive status and severe to moderate dementia. It uses a different subset of significant features for each classifier.

- A new comprehensive catalogue of 47 clock features is proposed, which includes new geometric features. The list of features is detailed to enhance the performance of the CDT in diagnosing the early symptoms of dementia.

- A new electronic CDT data set is prepared using the catalogue of 47 clock features. This is used to train and validate the proposed system. Furthermore, the dataset can facilitate future research around the use of clock drawing tests for diagnosis of dementia.

- The significant CDT features are defined for five discriminative tasks using the JMIM method which is proposed in this thesis.

- A new framework is proposed for analysing the temporal changes in the CDT features corresponding to the progress of dementia. It is used to define the first symptoms of dementia apparent in the clock drawings. The framework also defines the progress of the symptoms for AD and VaD individually.

- This research has verified some conclusions from literature which are related with the CDT. For example it confirms that CDTs are capable of distinguishing between AD and VaD, and that the use of a detailed feature set can improve the sensitivity of the test in diagnosing MCI.

8.2 Conclusions

Dementia is a progressive and irreversible disease, with around 800,000 people suffering from dementia in the UK. Dementia costs the UK economy 23 billion pounds per year including unpaid carers, health care costs, and social care health services.

To benefit from medical intervention it is very important to diagnose dementia very early, as new medications are regularly emerging which can delay or stop the impairment from progressing.

In this research the general problem of dementia is investigated with the main focus lying on the assessment tools used to assess cognitive functions, especially CDT and MMSE. CDT is widely used as a cognitive assessment tool for its simplicity, because it is quick to administer and because it is well accepted by both dementia patients and practitioners.

In the first step, CDT and MMSE data are collected from the Memory Clinic at the Llandough Hospital in Cardiff, UK. The data is collected during patients' examination procedures in the period from 1999 to 2009. The CDT drawings are then scanned, and image processing techniques are employed to enhance the quality of drawings.

For the purpose of enhancing the sensitivity of CDTs in detecting MCI and very early dementia cases a new comprehensive catalogue of 47 CDT features is prepared. This catalogue consists of new geometrical features which have never been used before, and is employed here to digitise 604 drawings and thereby construct a new electronic CDT dataset.

In this thesis the field of image feature selection is investigated, with special focus on information theory based methods. A research gap is identified, and a new feature selection method is proposed to resolve the drawback of some methods that have been reported in the literature. The method is validated using benchmarking data, and the JMIM shows good performance when compared against state of the art methods such as JMI, CMIM, mRMR, and DISR. The JMIM method is employed to study the significance of the clock features in terms of their discriminative power. The lists of significant CDT features for five different discriminative tasks are defined. No previous research has studied the CDT features using these feature selection techniques, and no studies have examined discriminative tasks other than discriminating between normal and abnormal cases. The findings show that most of the newly proposed CDT features are significant to the task of discriminating between different cognitive statuses.

There is no previous study in the research literature which analyses temporal changes in the CDT features corresponding to the progress of dementia. In response to this a new framework is proposed from MCI to the moderate stage dementia. The first features that are likely to be abnormal are defined, and as the disease progresses the number of features which are expected to be abnormal increases.

A new cascade (multistage) classifier scheme is proposed as a diagnosing stage in the proposed CDSS-DD. The CDT drawing is used to classify the patient into one of four cognitive statuses (Normal, Functional problem, MCI and early dementia, and Moderate and severe dementia). The cascade classifier is proposed to enhance the performance of the classification stage. The proposed classifier benefits from the feature selection stage by using different significant set of features for each classifier. This also enables the CDT and MMSE tests to be combined in order to enhance the performance in discriminating between the two diagnoses (MCI and early dementia vs. Moderate and severe dementia). It is further proposed that the classification process in the cascade classifiers is very similar to the diagnostic process that doctors follow to make diagnoses.

The proposed CDSS-DD shows good performance in diagnosing CDT drawings. The CDT is used currently as an assessment tool to diagnose patients as normal or abnormal. The proposed system could diagnose the CDT drawings into one of three statuses (Normal+, MCI+, Moderate and severe) with an accuracy of 78.34 %, and into one of four statuses (Normal, Functional, MCI+, Moderate and severe) with an accuracy of 71.0 %. The improvement in classifying the drawings compared with a single stage classifier is found to be 7.34 % when the drawings are classified into three classes, and 6.49 % when the drawings are classified into four classes. Moreover, the proposed CDSS-DD can distinguish between the normal+ and the abnormal cases with an accuracy of 89.54 %, which is better than the performance of most of the scoring systems that are currently being used to score CDT, and is competitive even against some more detailed scoring systems.

One of the cognitive statuses that the proposed CDSS-DD can classify drawings as is MCI+. It is very challenging to diagnose patients in the MCI+ category using the CDT scoring systems in the existing literature. This is in agreement with reports in the literature of the need for detailed features to be identified which can enhance the sensitivity of the CDT in diagnosing MCI. The results also show that CDT is not able to distinguish between AD and VaD, again in agreement with existing literature.

8.3 **Limitations**

This section discusses the limitations of the presented research. The main limitations are: Not all the listed features could be extracted automatically; Lack of information regarding the educational background of the patients, which may influence the outcome of the CDT; The anonymised drawings cannot be used to perform a longitudinal study on the CDTs; A further limitation related to unbalance data is due to the fact that not all the diagnoses are considered in the study, it is not possible to analyse in depth the performance of the CDT in diagnosing MCI cases due also to the limited number of MCI drawings available to this study.

8.4 **Future Work**

In future studies, the performance and behavior of the CDT in diagnosing MCI should be explored by gathering additional CDT drawings by MCI patients. A greater amount of data can also improve the overall performance of the system. The inclusion of background features such as the level of patients' education would facilitate a deeper analysis of the CDT drawings.

Another research path which deserves to be followed is the use of more sophisticated image processing techniques and handwriting recognition algorithms to fully automate the feature extraction tasks. The use of an online CDT via tablet computers or digitisers would enable the capture of clock drawings during the drawing task to reveal more information about the patients' planning process. This could also facilitate the extraction of new dynamic features such as: stylus presser, stylus velocity, air time, angle between the stylus and the surface. These dynamic features can lead to improved system performance, while online capturing can ease the image processing task.

Multiple assessment tools such as MMSE, problem-solving, trail making tests, cube-copying tests, and picture naming, can be included in one comprehensive CDSS package. This will lead to a better general understanding of dementia and the progression of the disorder, improving diagnostic performance and improving the prospects of patients.

To enhance the performance of the proposed feature selection methods a backward search can be used at the same time as the forward search. The performance of the diagnosis stage could be improved if more than one classifier is used at each stage of the cascade, and majority voting is employed to find the final classification.

Finally, a web-based implementation of the system can provide a very flexible tool for the early assessment of dementia.

References

Akadi, A., Ouardighi, A., and Aboutajdine, D. (2008). A Powerful Feature Selection approach based on Mutual Information. *IJCSNS International Journal of Computer Science and Network Security, 8*, 116-121.

Aprahamian, I., Martinelli, J., Neri, A., and Yassuda, M. (2009). The Clock Drawing Test: A Review of its Accuracy in Screening for Dementia. *Dementia and Neuropsychologia*, 3(2), 74-80.

Aprahamian, I., Martinelli, J., Neri, A., and Yassuda, S. (2010). The accuracy of the Clock Drawing Test Compared to that of Standard Screening Tests for Alzheimer's Disease: Results from a Study of Brazilian Elderly with Heterogeneous Educational Backgrounds. *International Psychogeriatrics, 22*(1), 64–71.

Asaoka, R., Iwase, A., Tsutsumi, T., Saito, H., Otani, S., Miyata, K., Murata, H., Mayama, C., and Araie, M. (2014). Combining Multiple HRT Parameters Using the 'Random Forests' Method Improves the Diagnostic Accuracy of Glaucoma in Emmetropic and Highly Myopic Eyes. *Investigative Ophthalmology and Visual Science, 55*(4), 2842-2850.

Australian Institute of Welfare, Health Australia Department of Health Ageing. (2007). Dementia in Australia: National Data Analysis and Development: Australian Institute of Health and Welfare.

Bache, K., and Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Bajwa, I., Naweed, M., Asif, M., and Hyder, S. (2009). Feature Based Image Classification by using Principal Component Analysis. *ICGST International Journal on Graphics, Vision and Image Processing, 9*(2), 11-17.

Battiti, R. (1994). Using Mutual Information for Selecting Features in Supervised Neural Net Learning. *IEEE Transactions on Neural Networks, 5*, 537-550.

Ben-Hur, A., and Weston, J. (2010). A User's Guide to Support Vector Machines. *Data Mining Techniques for the Life Sciences, 609*, 223-239.

Berger, G., Frolich, L., Weber, B., and Pantel, J. (2008). Diagnostic Accuracy of the Clock Drawing Test: The Relevance of ‘‘Time Setting’’ in Screening for Dementia. *Journal of Geriatric Psychiatry and Neurology,* 21, 250-260.

Berner, E., and Lande, T. (2007). Overview of Clinical Decision Support Systems. *Clinical Decision Support Systems.* New York: Springer

Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J. (2011). Data Mining for Credit Card Fraud: A comparative study. *Decision Support Systems, 50*, 602-613.

Bojar, I., Wójcik-Fatla, A., Owoc, A., and Lewiński, A. (2012). Polymorphisms of Apolipoprotein E Gene and Cognitive Functions of Postmenopausal Women, Measured by Battery of Computer Tests – Central Nervous System Vital Signs. *Activitas Nervosa Superior Rediviva, 50*, 385-392.

Borson, S., Scanlan, J., Brush, M., Vitaliano, P., And Dokmak, A. (2000). The Mini-Cog: a Cognitive 'Vital Signs' Measure for Dementia Screening in Multi-Lingual Elderly. *International Journal of Geriatric Psychiatry,* 15, 1021-1027.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32.

Brown, G., Pocock, A., Zhao, M., and Luj´an, M. (2012). Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. *Journal of Machine Learning Research, 13*, 27-66.

Burroni, M., Corona, R., Dell'Eva, G., Sera, F., Bono, R., Puddu, P., Perotti, R., Nobile, F., Andreassi, L., and Rubegni P. (2004). Melanoma Computer-Aided Diagnosis: Reliability and Feasibility Study. *Clinical Cancer Research, 10*, 1881–1886.

Catlett, J. (1991). On Changing Continuous Attributes Into Ordered Discrete Attributes. Paper presented at the In Pro. Fifth European Working Session On Learning.

Cheng, H., Qin, Z., Feng, C., Wang, Y., and Li F. (2011). Conditional Mutual Information-Based Feature Selection Analysing for Synergy and Redundancy. *ETRI Journal, 33*, 210-218.

Choi, B. (1992). Sensitivity and Specificity of a Single Diagnostic Test in the Presence of Work-up Bias. *Journal of Clinical Epidemiology*, 45, 581-558.

Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. *Machine Leaming, 20*, 273-297.

Cover, T., and Thomas, J. (2006). *Elements of Information Theory*. Hoboken, New Jersey: John Wiley and Sons.

Cutler, D., Edwards, T., Beard, K. Cutler, A., Hess, K., Gibson, J., and Lawler, J. (2007). Random Forests for Classification in Ecology. *Ecological Society of America*, 88(11), 2783-2792.

Dash, R., Paramguru, R., and Dash, R. (2011). Comparative Analysis of Supervised and Unsupervised Discretization Techniques. *International Journal of Advances in Science and Technology, 2*(3), 29-37.

Ding, C., and Peng, H. (2005). Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *Journal of Bioinformatics and Computational Biology, 3*(2), 185–205.

Dougherty, J., Kohavi, R., and Sahami, M. (1995). *Supervised and Unsupervised Discretization of Continuous Features.* Paper presented at the Machine Learning: Proceedings of the Twelfth International Conference, San Francisco, pp. 194-202.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern Classification*. New York: John Wiley and Sons.

Ehreke, L., Luck,T., Luppa, M., Konig, H., Villringer, A., and Riedel-Heller, S. (2011). Clock Drawing Test - Screening Utility for Mild Cognitive Impairment According to Different Scoring Systems: Results of the Leipzig Longitudinal Study of the Aged (LEILA 75+). *International Psychogeriatrics*, 23(10), 1592–1601.

Ehreke, L., Luppa, M., Konig, H., and Riedel-Heller, S. (2010). Is the Clock Drawing Test a Screening Tool for The Diagnosis of Mild Cognitive Impairment? A Systematic Review. *International Psychogeriatrics*, 22(1), 56-63.

Eknoyan, D., Hurley, R., and Taber, K. (2012). The Clock Drawing Task: Common Errors and Functional Neuroanatomy. *Windows to the Brain*, 24(3), 260-265.

Estévez, P., Tesmer, M., Perez, C., and Zurada, M. (2009). Normalized Mutual Information Feature Selection. *IEEE Transactions on Neural Networks, 20*, 189-201.

Fayyad, U., and Irani, K. (1993). *Multi-interval discretization of continuous valued attributes for classification learning.* Paper presented at the In Proceedings of the 13th International Joint Conference on Artificial Intelligence, Los Altos, pp. 1022-1027.

Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P. R., Rimmer, E., and Scazufca, M. (2005). Global Prevalence of Dementia: a Delphi Consensus Study. *Lancet*, 366, 211-217.

Fleuret, F. (2004). Fast Binary Feature Selection with Conditional Mutual Information. *Journal of Machine Learning Research, 5*, 1531–1555.

Folstein, M., Folstein, S., and Mchugh, P. (1975). "Mini-Mental State" a Practical Method For Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research, 21*, 189–198.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. I., and Delis, D. C. (1994). *Clock Drawing: A Neuropsychological Analysis*. New York, Oxford University Press.

Garcıa, S., Luengo, J. Saez, J., Lopez, V., and Herrera, F. (2013). A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering, 25*(4), 734-750.

Garg, A., Adhikari, N., McDonald, H., Rosas-Arellano, M., Devereaux, P., Beyene, J., Sam, J., and Haynes, R. (2005). Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes A Systematic Review. *American Medical Association, 293*(10), 1223-1338.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. Whitehouse, P., and Winblad, B. (2006). Mild Cognitive Impairment. *Lancet*, 367, 1262-1270.

Gazzaniga, M., Ivry, R., and Mangun, G. (2009). *Cognitive Neuroscience: The Biology of the Mind*: W.W. Norton.

Gerald, L. Tang, S., Bruce, F., Redden, D., Kimerling, M. Brook, N., Dunlap, N., and Bailey W. (2002). A Decision Tree for Tuberculosis Contact Investigation. *American journal of respiratory and critical care medicine, 166*(8), 1122-1127.

Gonzalez, R., Woods, R., and Eddins, S. (2010). *Digital Image Processing Using MATLAB*: Gatesmark Publishing.

Gray, K., Aljabar, P., Heckemann, R., Hammers, A., and Rueckert, D. (2013). Random Forest-Based Similarity Measures for Multi-Modal Classification of Alzheimer's Disease. *NeuroImage, 65*, 167–175.

Grober, E., Hall, C., Lipton R. B., and Teresi, J. A. (2008). Primary Care Screen for Early Dementia. *The American Geriatrics Society*, 56, 206-213.

Guyon, I., and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research, 3*, 1157-1182.

Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L. (2006). *Feature Extraction Foundations and Applications*.

Hardin, J., and Chhieng, D. (2007). Data Mining and Clinical Decision Support Systems. *Clinical Decision Support Systems Theory and Practice*. New York: Springer

Hart, L. (2002). *Human Brain and Human Learning* (3 ed.): Books for Educators.

Hashem, A., Rasmy, M., Wahba, K., and Shaker, O. (2012). Single Stage and Multistage Classification Models for the Prediction Of Liver Fibrosis Degree In Patients With Chronic Hepatitis C Infection. *Computer Methods and Programs in Biomedicine, 105*, 194-209.

Heinik, j., Solomesh, I., Bleich, A., and Berkman, P. (2003). Are the Clock-Drawing Test and the MMSE Combined Interchangeable With CAMCOG as a Dementia Evaluation Instrument in a Specialized Outpatient Setting?. *Journal of Geriatric Psychiatry and Neurology, 16*, 74-79.

Heinik, J., Werner, P., Dekel, T., Gurevitz, I., and Rosenblum, S. (2010). Computerized Kinematic Analysis of the Clock Drawing Task in Elderly People with Mild Major Depressive Disorder: An Exploratory Study. *International Psychogeriatrics, 22:3*, 479-488.

Ho, T. (1995). *Random decision forests.* Paper presented at the In Document Analysis and Recognition, Proceedings of the Third International Conference, pp. 278-282.

Ho, K., and Scott, P. (1997). Zeta: A Global Method for Discretization of Cotitinuous Variables. Paper presented at the 3rd International Conference on Knowledge Discovery and Data Mining (KDD99), NewPort Beach, USA.

Hoque, N., Bhattacharyya, D., and Kalita, J. (2014). MIFS-ND: A Mutual Information-Based Feature Selection Method. *Expert Systems with Applications, 41*, 6371-6385.

Hughes, C., Berg, L., Danziger, W., Coben, L., and Martin., R. (1982). A New Clinical Scale for the Staging of Dementia. *The Royal College of Psychiatrists*, 140, 566-572.

Ismail, Z., Rajji, T, K., and Shulman, K. I. (2010). Brief Cognitive Screening Instruments: An Update. *International Journal of Geriatric Psychiatry*, 25, 111-120.

Jaiantilal, A. (2009). Classification and Regression by Random Forest-matlab: http://code.google.com/p/randomforest-matlab.

Jain, A. Duin, R., and Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 4-37.

Jain, A., Duin, R., and Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 4-37.

Jakulin, A. (2003). *Attribute Interactions in Machine Learning.* Master, University of Ljubljana, Sezana.

Jakulin, A. (2005). *Machine Learning Based on Attribute Interactions.* PhD, University of Ljubljana, Sezana.

Janecek, A. Gansterer, W. Demel, M., and Ecker, G. (2008). On the Relationship between Feature Selection and Classification Accuracy. *JMLR: Workshop and Conference Proceedings, 4*, 90-105.

Joo, S., Yang, Y., Moon, W., and Kim, C. (2004). Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial Neural Network Based on Multiple Sonographic Features. *IEEE TRANSACTIONS ON MEDICAL IMAGING, 23*(10), 1292-1300.

Jouk, A., and Tuokko, H. (2012). A reduced scoring system for the Clock Drawing Test Using a Population-Based Sample. *International Psychogeriatric*, 24(11), 1738-1748.

Juby, A., Tench, S., Baker, V. (2002). The Value of Clock Drawing in Identifying Executive Cognitive Dysfunction in People with a Normal Mini-Mental State Examination Score. *Canadian Medical Association Journal*, 167, 859-864.

Kanchanatawan, B., Jitapunkul, S., Supapitiporn, S., and Chansirikarnjana, S. (2006). Validity of Clock Drawing Test (CDT), Scoring by Chula Clock-Drawing Scoring System (CCSS) in Screening Dementia among Thai Elderly in Community. *Journal of the Medical Association of Thailand*, 89(8), 1150-1156.

Karegowda, A., Jayaram, M., and Manjunath, A. (2010). Feature Subset Selection Problem using Wrapper Approach in Supervised Learning. *International Journal of Computer Applications, 1*, 13-17.

Kato, Y., Narumoto, J., Matsuoka, T., Okamura, A., Koumi, H., Kishikawa, Y., Terashima, S., and Kenji, F. (2013). Diagnostic Performance of a Combination of Mini-Mental State Examination and Clock Drawing Test in Detecting Alzheimer's Disease. *Neuropsychiatric Disease and Treatment, 9*, 581–586.

Kerber, R. (1992). ChiMerge: Discretisation of numeric attributes. Paper presented at the AAAI-92, Ninth National Confrence Artificial Intelligence

Kim, H. (2013). *The Clockme System: Computer-Assisted Screening Tool for Dementia.* PhD, Goorgia Institute of Technology, Georgia.

Kim, Y., Lee, K., Choi, B., Sohn, E., Lee, A. (2009). Relation Between the Clock Drawing Test (CDT) and Structural Changes of Brain in Dementia. *Archives of Gerontology and Geriatrics*, 48, 218-221.

Kira, K., and Rendell, L. (1992). *A practical approach to feature selection. In A practical approach to feature selection.* Paper presented at the ninth international workshop on Machine learning, San francisco, CA, USA.

Kitabayashi, Y., Ueda, H., Narumoto, J., Nakamura, K., Kita, H., and Fukui K. (2001). Qualitative Analyses of Clock Drawings in Alzheimer's Disease and Vascular Dementia. *Psychiatry and Clinical Neurosciences*, 55, 485–491.

Knapp, M., Prince, M. 2005. *Dementia UK- The Full Report.* (2005). Alzheimer's Society.

Kong, G. (2011). *An Online Belief Rule-Based Group Clinical Decision Support System*. Phd, University Of Manchester, Manchester.

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica, 31*, 249-268.

Kotsiantis, S., and Kanellopoulos, D. (2006). Discretization Techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering, 32 (1)*, 47-58.

Kotsiantis, S., Zaharakis, I., and Pintelas, P. (2006). Machine learning: a Review of Classification and Combining Techniques. *Artificial Intelligence Review, 26*, 159–190.

Kouk, A. (2010). A Reduced Scoring System for the Clock Drawing Test Using a Population Based Sample. Master of Science, University of Victoria.

Kuhn, H. (1955). The Hungarian Method for the Assignment Problem. *Naval Research Logistic Quarterly*, 2, 83-97.

Kuncheva, L. (2007). *A Stability Index for Feature Selection* Paper presented at the 25th Lasted international Multi-Conference Artificial intelligence and Applications, Innsbruck, Austria, pp. 2783-2792.

Kwak, N., and Choi, C. (2002). Input Feature Selection for Classification Problems. *IEEE Transactions on Neural Networks, 13*, 143-159.

Lakey, L., Chandaria, K., Quince, C., Kane M., and Saunders, T. (2012). *Dementia 2012: A National Challenge*. Alzheimer's Society.

Lapshin, H., O'Connor, P., Lanctot, K., and Feinstein, A., (2012). Computerized Cognitive Testing for Patients with Multiple Sclerosis. *Multiple Sclerosis and Related Disorders, 1*, 196-201.

Lee, H., Swanwick, G. Coen, R., and Lawlor, B. (1996). Use of the Clock Drawing Task in the Diagnosis of Mild and Very Mild Alzheimer's Disease. *International Psychogeriatrics*, 8(3), 469-476.

Lee, J., Oh, E., Jeong, S., Sohn, E., Lee, Y., Lee, A. (2011). Longitudinal changes in Clock Drawing Test (CDT) Performance According to Dementia Subtypes and Severity. Archives of Gerontology and Geriatrics, 53, e179-e182.

Lessig , M., Scanlan , J., Nazemi, H., and Borson, S. (2008). Time that Tells: Critical Clock-Drawing Errors for Dementia Screening. *International Psychogeriatrics*, 20, 459 – 470.

Levey, A. Lah, J., Goldstein, F., Steenland, K., and Bliwise, D. (2006). Mild Cognitive Impairment: An Opportunity to Identify Patients at High Risk for Progression to Alzheimer's Disease. *Clinical Therapeutics*, 28, 991-1001.

Lin, T., Li, H., and Tsai, K. (2004). Implementing the Fisher's Discriminant Ratio in a k-Means Clustering Algorithm for Feature Selection and Data Set Trimming. *Journal of Chemical Information and Computer Sciences*, 44, 76-87.

Liu, H., and Motoda, H. (2008). *Computational Methods of Feature Selection*. New York: Chapman and Hall/ CRC.

Liu, H., and Yu, L. (2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering, 17*, 491-502.

Liu, H., Hussain, F., TAN, C., and DASH, M. (2002). Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery, 6*, 393–423.

Luengo-Fernandez, R., Leal, J., and Gray, A. (2010). Dementia 2010 the Economic Burden of Dementia and Associated Research Funding In the United Kingdom.

Maini, R., and Aggarwal, (2010). A Comprehensive Review of Image Enhancement Techniques. *Journal of Computing*, 2(3), 8-13.

Manos, P. (1999). Ten-Point Clock Test Sensitivity for Alzheimer's Disease In Patients With MMSE Scores Greater Than 23. *International Journal of Geriatric Psychiatry*, 14, 454-458.

Markowitsch, H. (1995). Which Brain Regions are Critically Involved in the Retrieval of Old Episodic Memory?. *Brain Research Reviews*, 21, 117-127.

Markowitsch, H., and Staniloiu, A. (2012). Amnesic disorders. *Lancet*, 380, 1429–1440.

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., and Pietrzak, R. (2009). Validity of the CogState Brief Battery: Relationship to Standardized Tests and Sensitivity to Cognitive Impairment in Mild Traumatic Brain Injury, Schizophrenia, and AIDS Dementia Complex. *Archives of Clinical Neuropsychology, 24*, 165–178.

Matioli, M., and Caramelli, P. (2010). Article Limitations in Differentiating Vascular Dementia from Alzheimer's Disease with Brief Cognitive Tests. *Arquivos de Neuro-Psiquiatria, 68*(2), 185-188.

McKeith, I. (2004). Dementia with Lewy Bodies. *Dialogues in Clinical Neuroscience*, 6, 333-341.

Mckhann, G. M., Albert, M. S., Grossman, M. Miller, B., Dickson, D., and Trojanowski., J. Q. (2001). Clinical and Pathological Diagnosis of Frontotemporal Dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease. *American Medical Association*, 58, 1803-1809.

Mendez, M., Ala, T., and Underwood, K. (1992). Development of Scoring Criteria for the Clock Drawing Task in Alzheimer's Disease. *Journal of the American Geriatrics Society*, 40, 1095 – 1099.

Metzger, J., and MacDonald, K. (2002). *Clinical Decision Support for the Independent Physician Practice.* California Healthcare Foundation.

Meyer, P., and Bontempi, G. (2006). *On the Use of Variable Complementarity for Feature Selection in Cancer Classification.* Paper presented at the in Applications of Evolutionary Computing, pp. 91–102.

Meyer, P., Schretter, C., and Bontempi, G. (2008). Information-Theoretic Feature Selection in Microarray Data Using Variable Complementarity. *IEEE Journal of Selected Topics in Signal Processing, 2*, 261-274.

Miller, R. (2009). Computer-Assisted Diagnostic Decision Support: History, Challenges, and Possible Paths Forward. *Advances in Health Sciences Education, 14*, 89–106.

Milne, A., Culverwell, A., Guss, R., Tuppen, J., and Whelton, R. (2008). Screening for Dementia in Primary Care: A Review of the Use, Efficacy and Quality of Measures. *International Psychogeriatrics*, 20(5), 911–926.

Milne, A., Hamilton-west, K., and Hatzidimitriadou, E. (2005). GP Attitudes to Early Diagnosis of Dementia: Evidence of Improvement. *Aging and Mental Health*, 9(5), 449-455.

Mittal, C., Gorthi, C., and Rohatgi, M. (2010). Early Cognitive Impairment: Role of Clock Drawing Test. *Canadian Medical Association*, 167(8), 859-864.

Musen, M., Shahar, Y., And Shortliffe, E. (2014). Clinical Decision-Support Systems. *Biomedical informatics*, 643-674.

Nair, A., Gavett, B. Damman, M., Dekker, W., Green, R., Mandel, A., Auerbach, S., Steinberg, E., Hubbard, E., Jefferson, A., and Stern, R. (2010). Clock Drawing Test Ratings by Dementia Specialists: Interrater Reliability and Diagnostic Accuracy. *Journal of Neuropsychiatry and Clinical Neurosciences*, 22, 85-92.

Orrù, G., Pettersson-Yeo, W., Marquand, A., Sartori, G., and Mechelli, A. (2012). Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review. *Neuroscience and Biobehavioral Reviews, 36*, 1140–1152.

Oshiro, T., Perez, P., and Baranauskas, J. (2012). *How Many Trees in a Random Forest?,* Paper presented at the 8th international conference on Machine Learning and Data Mining in Pattern Recognition, pp. 154-168.

Peng, H., Long, F., and Ding, C. (2005). Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*, 1226-1238.

Petersen, R. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183–194.

Petersen, R., Stevens, J., Ganguli, M., Tangalos, E., Cummings, J., DeKosky, DeKosky, S. (2001). Practice Parameter: Early Detection of Dementia: Mild Cognitive Impairment (An Evidence-Based Review). *Neurology*, 56, 1133-1142.

Pinto, E., and Peters, R. (2009). Literature Review of the Clock Drawing Test as a Tool for Cognitive Screening. *Dementia and Geriatric Cognitive Disorders*, 27, 201-213.

Prasath, V., Lakshmi, N., Nathiya, M., Bharathan, N., Neetha, N. (2013). A Survey on the Applications of Fuzzy Logic in Medical Diagnosis. *International Journal of Scientific & Engineering Research*, 4(4), 1199-1203.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A., Mcnamara ,J., Williams, S. (2004). *Neuroscience*. Sunderland: Sinauer Associates.

Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning, 1*, 81-106.

Quinlan, J. (1993). *C4.5: programs for machine learning*. San Francisco: Morgan Kaufmann.

Ramírez, J., M.Górriz, J., Salas-Gonzalez, D., Romero, A., López, M., Álvarez, I., and Gómez-Río, M. (2013). Computer-Aided Diagnosis of Alzheimer's Type Dementia Combining Support Vector Machines and Discriminant Set of Features. *Information Sciences, 237*, 59-72.

Richardson, H., and Glass, N. (2002). A Comparison of Scoring Protocols on the Clock Drawing Test in Relation to Ease of Use, Diagnostic Group, and Correlations with Mini-Mental State Examination. *Journal of the American Geriatrics Society*, 50, 169–173.

Robbins, T., and Sahakian, B. (2002). Computer Methods of Assessment of Cognitive Function. *Principles and Practice of Geriatric Psychiatry, 28*, 147-151.

Rodgers, J., and Nicewander, W. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician, 42*, 59-66.

Rokach, Lior. (2005). Decision Trees. In L. Rokach, and Oded Maimon (Ed.), *Data Mining and Knowledge Discovery Handbook* (pp. 165-192): Springer US.

Roth, M., Tym, E., Mountjoy, C., Huppert, F., Hendrie, H., Verma, S., and Goddard, R. (1986). CAMDEX: A Standardised Instrument for the Diagnosis of Mental Disorder in the Elderly with Special Reference to the Early Detection of Dementia. *British Journal of Psychiatry,* 149, 698-709.

Rouleau, I., Salmon, D., And Butters, N. (1996). Longitudinal Analysis of Clock Drawing in Alzheimer's Disease Patients. *Brain and Cognition*, 31, 17–34.

Roulrau, I., Salmon, D., Butters, N., Kennedy, C., and Mcguire, K., (1992). Quantitative and Qualitative Analyses of Clock Drawings in Alzheimer's and Huntington's Disease. *Brain and Cognition*, 18, 70-87.

Royall, D. (1998). CLOX: an executive clock drawing task. *Journal of Neurology, Neurosurgery & Psychiatry,* 64, 588-594.

Saatci, Y., and Town, C. (2006). *Cascaded Classification of Gender and Facial Expression using Active Appearance Models.* Paper presented at the the 7th International Conference on Automatic Face and Gesture Recognition.

Saeys, Y., Inza, I., and Larranaga, P. (2007). A Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics, 23*, 2507–2517.

Seo, S. (2006). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. Master, University of Pittsburgh, Pittsburgh.

Sezgin, M., and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging, 13*(1), 146–165.

Shortliffe, E. (1976). *Computer-based medical consultations: MYCIN*. Amsterdam: Elsevier.

Shulman, k. 2000. Clock-Drawing: is it the Ideal Cognitive Screening Test?. *International Journal of Geriatric Psychiatry*, 4, 437- 450.

Shulman, K. Gold, D., Cohen, C., and Zucchero, C. (1993). Clock-Drawing and Dementia in the Community: A Longitudinal Study. *International Journal of Geriatric Psychiatry*, 8, 487-496.

Shulman, K., Shedletsky, R., Silver, I. (1986). The Challenge of Time: Clock-Drawing Cognitive Function in the Elderly. *International Journal of Geriatric Psychiatry*, 1, 135-140.

Sim, I., Gorman, P., Greenes, R. Haynes, B., Kaplan, B., Lehmann, H., Tang P. (2001). Clinical Decision Support Systems for the Practice of Evidence-based Medicine. *Journal of the American Medical Informatics Association*, 8(6), 527-534.

Simpson, P., Surmonk, J., Wesnes, A., and Wilcock, G. (1991). The Cognitive Drug Research Computerized Assessment System for Demented Patients: A Validation Study. *International Journal of Geriatric Psychiatry, 6*, 95-102.

Solomon, P., Hirschoff, A., Kelly, B., Relin, M., Brush, M., DeVeaux, R., and Pendlebury, W. (1998). A 7Minute Neurocognitive Screening Battery Highly Sensitive to Alzheimer's Disease. *American Medical Association*, 55, 349-355.

Sreedhar K., and Panlal, B. (2012). Enhancement of Images Using Morphological Transformations. *International Journal of Computer Science & Information Technology*, 4(1), 33-50.

Sunderland, T., Hill, J., Mellow, A., Lawlor, B., Gundersheimer, J., Newhouse, PA., and Grafman, J. (1989). Clock Drawing in Alzheimer's Disease: A Novel Measure of Dementia Severity. *Journal of the American Geriatrics Society*, 37, 725– 729.

Tang, E., Suganthana, P., Yao, X., and Qina, A. (2005). Linear Dimensionality Reduction Using Relevance Weighted LDA. *Pattern Recognition, 38*, 485 – 493.

Tornatore, J., Hill, E., Laboff, J., and McGann, M. (2005). Self-Administered Screening for Mild Cognitive Impairment: Initial Validation of a Computerized Test Battery. *The Journal of Neuropsychiatry and Clinical Neurosciences, 17*, 98-105.

Trapeznikov, K., Saligrama, V., and Castanon, D. (2012). Multi-Stage Classier Design. *Journal of Machine Learning Research, 25*, 459-474.

Tukey, J. (1977). Exploratory data analysis.

Tuokko, H., Hadjistavropoulos, T., Miller, JA., and Beattie, BL. (1992). The clock test: A Sensitive Measure to Differentiate Normal Elderly from Those with Alzheimer Disease. *Journal of the American Geriatrics Society*, 40, 579 – 584.

Turk, M., and Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience, 3*, 72-86.

Van Bemmel, J., and Musen, M. (1997). *Handbook of Medical Informatics*. New York: Springer.

Veroff, A., Cutler, N. Sramek, J., Prior, P. Mickelson, W., and Hartman, J. (1991). A New Assessment Tool for Neuropsychopharmacologic Research: The Computerized Neuropsychological Test Battery. *Journal of Geriatric Psychiatry and Neurology, 4*, 211-217.

Vidal-Naquet, M., and Ullman, S. (2003). *Object recognition with informative features and linear classification.* Paper presented at the Ninth IEEE International Conference on Computer Vision, pp. 281-288.

Waldemar, G., Dubois, B., Emre, M., Georges, J., McKeith, I. G., Rossor, M., Scheltens, P., Tariska, P., and Winblad., B. (2007). Recommendations for the Diagnosis and Management of Alzheimer's Disease and other Disorders Associated with Dementia: EFNS Guideline. *European Journal of Neurology*, 14, e1-e26.

Wang, J., and Tan, Y. (2011). *Morphological Image Enhancement Procedure Design By Using Genetic Programming*. Paper presented at the 13th annual conference on Genetic and evolutionary computation.

Wild, K., Howieson, D., Webbe, F., Seelye, A., and Kaye, J. (2008). The Status of Computerized Cognitive Testing in Aging: A Systematic Review. *Alzheimer and Dementia*, 4, 428-437.

Yang, H., and Moody, J. (1999). *Feature Selection Based on Joint Mutual information* Paper presented at the Advances in Intelligent Data Analysis (AIDA),Computational Intelligence Methods and Applications (CIMA),International Computer Science Conventions, New York, USA, pp. 22-25.

Yang, Y., and Webb, G. (2002). *A Comparative Study of Discretization Methods for Naive-Bayes Classifiers*. Paper presented at the 2002 Pacific Rim Knowledge Acquisition Workshop, Tokyo, Japan,

Yeh, C., Lin, F., and Hsu, C. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33, 166–172.

Yu, H., and Yang, J. (2001). A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition. *Pattern Recognition, 34*, 2067-2070.

Yu, L., and Liu, H. (2004). Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research, 5*, 1205–1224.

Yu, L., Ding, C., and Loscalzo, S. (2008). *Stable feature selection via dense feature groups*. Paper presented at the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Yu, W., Liu, T., Valdez, R., Gwinn, M., and Khoury, M. (2010). Application of Support Vector Machine Modelling for Prediction of Common Diseases: The Case of Diabetes and Pre-Diabetes. *BMC Medical Informatics and Decision Making, 10*(16), 1472-6947.

Zheng, K. (2010). Clinical Decision Support Systems." Management, Types and Standards”*. Encyclopaedia of Library and Information Sciences, Third Edition, 36, 501-509.*

Zhang, P., Bui, T., and Suen, C. (2007). Anovel Cascade Ensemble Classifier System With a High Recognition Performance on Handwritten Digits. *Pattern Recognition, 40*, 3415 – 3429.

Zhang, Y., Zhang, B., Coenen, F., and Lu, W. (2013). Breast Cancer Diagnosis from Biopsy Images with Highly Reliable Random Subspace Classifier Ensembles. *Machine Vision and Applications, 24*, 1405–1420.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. (2003). Face recognition: A literature survey. *Association for Computing Machinery computing Surveys, 35*(4), 399-458.

Appendix A

Catalogue of CDT image features

Features related to clock numbers:

These include features derived from the clock numbers, namely the spacing, position, size, omission, preservation, orientation, direction of writing, sequence, and type of numbers.

Features 1, 2, and 3 (count of numbers within areas 1, 2, and 3)

Features 1, 2 and 3 are used to capture the deficit in the spacing of the numbers. Since some dementia patients tend to write the numbers far from the outer contour of the clock. The clock area is divided into three parts by drawing new two circles (explained in section 4.3).

Any numbers whose centroid is located precisely on any of the circular boundaries are counted with the outer area of that circle. In the case of the normal drawings all 12 numbers are likely to be written in area 1 (the outer area).

Features 4, 5, 6, and 7 (count of numbers within quadrants 1, 2, 3, and 4)

Features 4, 5, 6 and 7 are used to capture the deficit in the spacing of the numbers. Since some dementia patients tend to distribute the numbers unevenly around the clock (in some cases the numbers are clustered in one area of the clock while the rest of the clock is empty) the clock face is divided into 4 quadrants to capture this deficit in spacing (Figure A1).

The count of the numbers in each quadrant is defined as a feature. If the centroid of the number lies precisely on one of the boundaries between the quadrants, it is counted with the quadrant which is positioned immediately counter-clockwise to the boundary. Figure A1 shows examples of measuring features 4, 5, 6, and 7 from a 77 year old female AD patient's clock drawing. The value of these features for this drawing are: 3, 3, 2, and 4 respectively. These four features are adopted from the Watson scoring system. Mendez et al. also employ similar qualitative features.

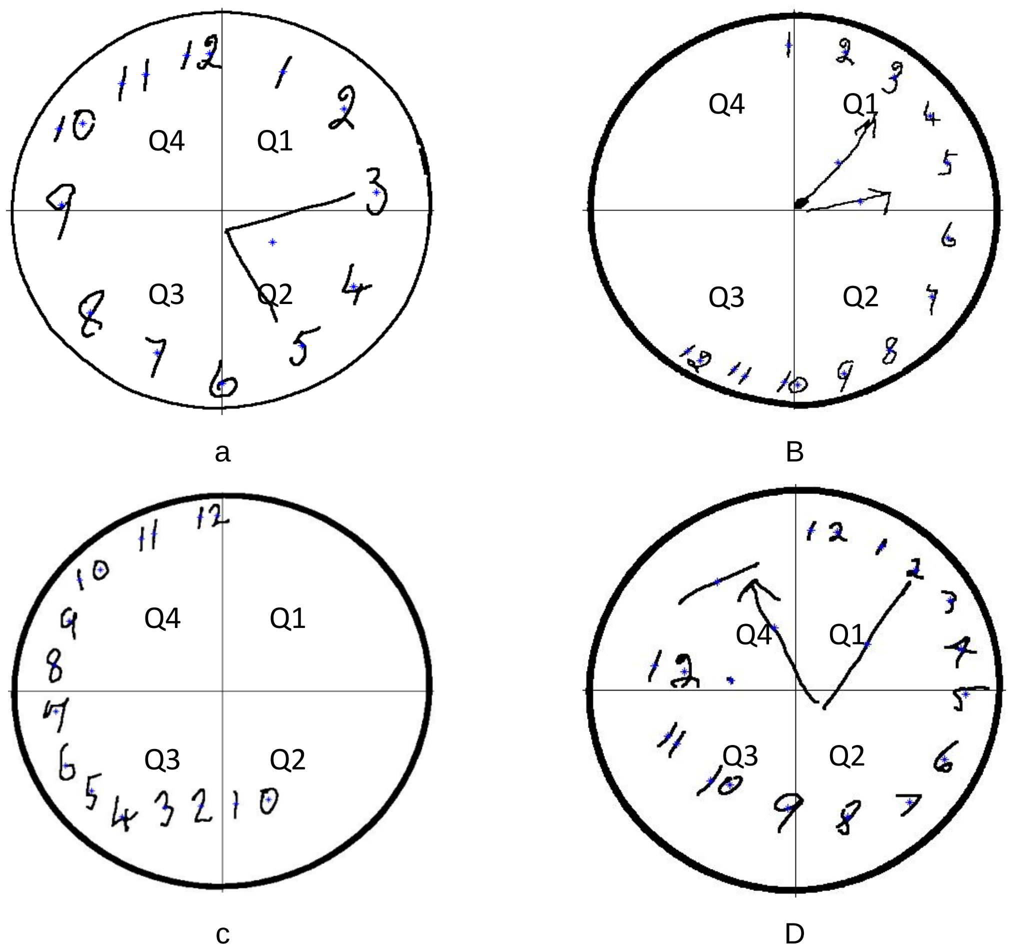


Figure A1: Examples of measuring features 1, 2, 3, and 4: (a) Drawing by a 77 year old female AD patient. The values of features 4, 5, 6 and 7 are 3, 3, 2, and 4 respectively, (b) Drawing by a 74 year old male AD patient. The values of features 1, 2,3 and 4 are 4, 4, 3, and 1 respectively, (c) Drawing by a 76 year old male VaD patient. The values of features 1, 2,3 and 4 are 0, 2, 6, and 5 respectively , (d) Drawing by an 84 year old female AD patient. The value of features 1, 2, 3 and 4 are 5, 4, 3, and 1 respectively.

Features 8, 9, and 10 (related with the size of numbers)

These three features are related with the size of the numbers, the area of each number is calculated as shown in section 4.3.

The minimum number size and maximum number size are defined as features 8 and 9 respectively. The ratio between the maximum and the minimum is also calculated to represent the variation in number size. This is defined as feature 10.

Feature 11 (the count of numbers outside the contour)

This is a count of the numbers written completely or partially outside the contour of the clock face. Similar features are used in the scoring systems by Mendez, Tuokko, Freedman, and Royall.

Features 12 and 13 (the angle between numbers)

These new features are the maximum angle and the minimum angle between each two consecutive numbers they are used to capture the deficit in the spacing of numbers. These two features are explained more in section 4.3.

Feature 14 (the count of numbers whose rotation is over 25 degrees)

This is the count of digits whose rotation angle in any direction is greater than 25°. This feature is explained more in section 4.3.

Feature 15 (the count of numbers left out from the drawing)

This is a count of the numbers that are omitted and not written by the individual. This measure has been employed widely by most of the common scoring systems: Shulman, Mendez, Tuokko, Freedman, Royall, and Manos.

Feature 16 (the count of duplicated numbers)

This is a count of duplicated numbers. Similar features have been used within some other scoring systems for example, by Mendez and Tuokko.

Feature 17 (sequential numbers are written following 12 (13, 14, 15 …))

This is a binary feature. It takes a value of “1” if numbers beyond 12 are written; otherwise it is equal to “0”. This technique is also employed within the Shulman, Mendez, Tuokko, and Freedman scoring systems.

Feature 18 (numbers not in sequence)

This is another binary feature, which takes the value “1” if the sequence is incorrect, and otherwise has a value of “0”. This is adopted from the Tuokko scoring system.

Feature 19 (the numbers 3 and 11 not present)

This is a binary feature. It is equal to unity if both numbers are present, and “0” if one or both of them are missing. Similar features are employed by the Mendez scoring system.

Feature 20 (arabic only numbers used)

This is a binary feature, which has a value of “1” if only Arabic numbers are used, and is otherwise equal to “0”. Similar features are used within the Mendez and Royall scoring systems.

Feature 21 (direction of written numbers)

This is another binary feature, whose value is “1” if the numbers are written in the clockwise direction, and is otherwise equal to “0”. Similar features are used within the Mendez, Tuokko, and Freedman scoring systems.

Feature 22 (self-correction of numbers)

This binary feature takes a value of “1” if one or more numbers have been corrected after they were first written, otherwise the feature is equal to “0”. This is adopted from a similar feature in the Freedman scoring system.

Features related to time setting:

These features include data related to the presence of clock hands, position of hands, time accuracy, angle between hands, and ratio between the lengths of the hands.

Features 23, 24 (the presence of clock hands)

Two binary features are used to express the hands' present states, one for the minute hand and another for the hour hand. The value of these two features is "1" if the hand is drawn and zero if not. Similar features are used within the following scoring systems: Sunderland, Mendez, Tuokko, Freedman, and Royall.

Feature 25 (more than two hands are drawn)

This takes a binary value of "1" when more than two hands are drawn, and otherwise is equal to "0". There are several scoring systems that use similar features, including Tuokko, and Royall.

Feature 26 (self-correction of hands)

This is a new discrete feature. It is proposed based on the preliminary analysis of the available data. The value is "1“ if there is self-correction, "0“ if there is no self-correction, and "-1“ if the hands are missed.

Feature 27 (time is correct)

This describes whether the time is correct or not. The value is "1" if the time is correct, "0" if the time is wrong, and "-1" if the hands are missing. This feature is also employed by the Sunderland, and Tuokko scoring systems.

Feature 28 (time is indicated by writing minute number next to 3 or 11)

This is a binary feature. It is similar to a feature used in the Freedman scoring system. The value of the feature is "1" if the time is written close to 3 or 11, and is otherwise "0".

Feature 29 (straight line is used between the two numbers)

This is a binary feature, whose value is “1” if the hand is drawn as a straight line between 3 and 11, and is otherwise “0”. This feature is adopted from the Shulman scoring system.

Features 30, 31 (displacement of hour and minute hands from the target number)

These are discrete features. The value of each feature is “1” if the hand is not pointing to the target number, even when the time is considered in general as correct, “0” if the hand is pointing to the target number, and “-1” if the hand is missed. A similar feature is used within the Tuokko scoring system.

Feature 32 (hands connected to the target number)

This is a new discrete feature, which can take five different values: “1” if the minute hand only is connected to its target number; “2” if the hour hand only is connected to its target number; “3” if both are connected to their target numbers; “0” if neither of the hands is connected; and finally “-1” if the hands are missed.

Feature 33 (arrows on hands)

This is also a discrete feature, which takes five different values as well: “1” if the arrow is drawn only on the hour hand; “2” if the arrow is drawn only on the minute hand; “3” if both of the hands have arrows; “0” if neither of the hands have arrows; and finally “-1” if the hands are missed. This feature is adopted from the Freedman scoring system.

Feature 34 (displacement of arrows less than 4mm)

This is a discrete feature, which takes four different values: “1” if there is no displacement of any of the arrows; “0” if at least one arrow is displaced from the hand by more than or equal to 4 mm; “-1” if there are no arrows on the hands; and “-2” if the hands are missed. This is adopted from the Freedman scoring system.

Feature 35 (arrows are pointing in the wrong direction)

This is another discrete feature. Its value is “1” when at least one arrow is pointing in the wrong direction; “0” if both of them are pointing in the right direction, “-1” if there are no arrows drawn, and “-2” if the hands are missed. This technique is adopted from a similar feature in the Freedman scoring system.

Feature 36 (presence of superfluous)

This is a binary feature to capture whether there are any additional marks such as spokes of a wheel, or other marks like “Christmas trees”. The value is “1” if there is superfluous marking, and is otherwise “0”. This is adopted from the Freedman scoring system.

Feature 37 (hands are joint or within 12 mm)

This is a discrete feature, whose value is “1” if the hands are joint to within 12 mm, “0” if the hands are not connected, and “-1” if one hand or both are missed. This is adopted from a similar feature in the Freedman scoring system.

Features 38, and 39 (position of clock hands)

Two new features are proposed to capture the deficit in the position of the clock hands. They are discrete features. Their values are the number of the sector in which each hand is located, after The area of the clock is divided into eight sectors as explained in section 4.3.

Feature 40 (angle between clock hands)

This is also a new feature. The angle between the hands (if the hands are present) is measured and used as a feature. It is discriped more in section 4.3.

Feature 41 (ratio between hands)

This is a continuous feature to measure the ratio between the minute hand length and the hour hand length. For normal cases the value of this is likely to be greater than “1”. A similar feature is used in the Mendez scoring system.

Feature 42 (presence of stem of clock hands near to the center

This is a discrete variable. Its value is “1” when the hands are not connected close to the center, “0” when the hands are connected close to the centre, and “-1” if one or both of the hands are missing. A similar feature is used in the Tuokko scoring system.

Features related to distortions and substitutions:

This list includes features related to writing or drawing irrelevant letters, words or figures.

Feature 43 (time is written across the clock)

This is a binary feature. Its value is “1” when there is a time written on the clock face, and is otherwise is “0”. This is adopted from similar features used in the scoring systems by Tuokko and Freedman.

Feature 44 (time is written outside the clock)

Another binary feature is adopted from the Freedman scoring system. Its value is “1” when there is any word written outside the clock face, and otherwise is “0”.

Feature 45 (picture of a human face is drawn on the clock face)

The value of this feature is binary. It is set to “1” when there is a picture of a human face drawn on the clock face, and otherwise it is “0”. This is adopted from the Shulman, Tuokko, Freedman, and Royall scoring systems.

Feature 46 (presence of written words)

The value of this is binary. It is set to “1” when there is a word written on the clock face, and is otherwise “0”. This feature is adopted from the Shulman, Tuokko, Freedman, and Royall scoring systems.

Features related to the clock center

Feature 47 (distance between the position of hands intersection and the center of the clock)

This is a new feature. It is proposed based on the preliminary analysis of the data. The normative data shows that the healthy individual is likely to start drawing the hands from a point close to the clock face center, while dementia patients start drawing from a point away from the center. The distance between the center of the clock and the intersection of the clock hands is measured in mm. In the case when the hands are not connected they are extended until they intersect. The distance from the center to the point of intersection is measured and used as a feature. Figure 4.11 shows examples of measuring this.

Appendix B

CDT Feature Selection Experiment Results

Table B1: Results of feature selection for Normal / MCI+ discriminative task.

No.	Feature
1.	Angle between clock hands
2.	Count of numbers within Area 1
3.	Maximum angles between numbers
4.	Displacement of minute hand or mar k from the target number
5.	Count of numbers within Area 2
6.	Position of min hand
7.	Ratio between hands
8.	Time is correct
9.	Maximum size of numbers
10.	Distance between the position of hands intersection and the center
11.	Position of Hour hand
12.	Count of numbers within quadrant4
13.	Arrows on hands
14.	Minimum angles between numbers
15.	Hands are joint or within 12 mm
16.	Hands connected with target number
17.	Count of numbers within Area 3
18.	Stem of clock hands (near to the center) is left out
19.	Displacement of arrows less than 4mm
20.	Count of numbers within quadrant2
21.	Minute hand is present
22.	Count of numbers which its orientation is more 25
23.	Minimum size of numbers
24.	Count of numbers within quadrant 3
25.	Hour hands is present'
26.	Arrows are pointing to the wrong direction
27.	Count of numbers within quadrant 1
28.	Hands self-correction
29.	Displacement of hour hand or mar k from the target number
30.	Count of repeated or duplicated numbers
31.	Count of numbers left out
32.	There are superfluous
33.	Sequence of numbers not in sequence
34.	Number (11,3) aren't present
35.	More than two hands are used
36.	Presence of written words
37.	Sequential numbers are written following 12
38.	Straight line is used between the two numbers
39.	Direction of numbers written
40.	Number representation
41.	Number self-correction
42.	Time is indicated by writing minute number
43.	Time is written across the clock
44.	Time is written outside the clock
45.	Picture of human face is drawn on clock

Table B2: Results of feature selection for Severe/ MCI+ discriminative task.

No.	Feature
1.	Count of numbers within Area 1
2.	Angle between clock hands
3.	Count of numbers within Area 3
4.	Displacement of minute hand or mar k from the target number'
5.	Count of numbers within Area 2
6.	Position of min hand
7.	Time is correct
8.	Count of numbers within quadrant 4
9.	Ratio between hands
10.	Count of numbers within quadrant 2
11.	Count of numbers within quadrant 1
12.	Count of numbers within quadrant 3
13.	minimum angles between numbers
14.	Hands are joint or within 12 mm
15.	Stem of clock hands (near to the center) is left out
16.	Maximum angles between numbers
17.	Minute hand is present
18.	Position of Hour hand
19.	Distance between the position of hands intersection and the center
20.	Hands self-correction
21.	Hour hands is present
22.	Displacement of hour hand or mar k from the target number
23.	Hands connected with target number
24.	Arrows are pointing to the wrong direction
25.	Displacement of arrows less than 4mm
26.	There are superfluous
27.	Arrows on hands
28.	Number (11,3) aren't present
29.	Sequence of numbers not in sequence
30.	Number representation
31.	Minimum size of numbers
32.	Presence of written words
33.	Sequential numbers are written following 12
34.	Direction of numbers written
35.	More than two hands are used
36.	Number self-correction
37.	Straight line is used between the two numbers
38.	Time is indicated by writing minute number
39.	Time is written across the clock
40.	Time is written outside the clock
41.	Picture of human face is drawn on clock

Table B3: Results of feature selection for Normal+ / abnormal discriminative task.

No.	Feature
1.	Angle between clock hands
2.	Ratio between hands'
3.	Minimum angles between numbers
4.	Distance between the position of hands intersection and the center
5.	Minimum size of numbers
6.	Count of numbers within Area 1
7.	Maximum angles between numbers
8.	Time is correct
9.	Position of minute hand
10.	Arrows on hands
11.	Count of numbers within quadrant4
12.	Displacement of minute hand or mark from the target number
13.	Position of Hour hand
14.	Hands connected with target number
15.	Count of numbers within Area 2
16.	Stem of clock hands (near to the center) is left out
17.	Maximum size of numbers
18.	Count of numbers which its orientation is more 25
19.	Arrows are pointing to the wrong direction
20.	Hands are joint or within 12 mm
21.	Count of numbers within quadrant 1
22.	Displacement of arrows less than 4mm
23.	Minute hand is present
24.	Count of numbers within quadrant 2
25.	Count of numbers within quadrant 3
26.	Displacement of hour hand or mar k from the target number
27.	Hour hands is present
28.	Hands self-correction
29.	Count of numbers of numbers left out
30.	Count of numbers within Area 3
31.	Count of numbers of repeated or duplicated numbers
32.	There are superfluous
33.	Number (11,3) aren't present
34.	More than two hands are used
35.	Sequence numbers not in sequence
36.	Straight line is used between the two numbers
37.	Sequential numbers are written following 12
38.	Ratio between max and min size
39.	Direction of numbers written
40.	Time is indicated by writing minute number
41.	Presence of written words
42.	Number self-correction
43.	Time is written outside the clock
44.	Time is written across the clock
45.	Numbers representation
46.	Picture of human face is drawn on clock

Table B4: Results of feature selection for Normal / Functional discriminative task.

No.	Feature
1.	Maximum size of numbers
2.	Maximum angles between numbers
3.	Distance between the position of hands intersection and the centre
4.	Minimum angles between numbers
5.	Angle between clock hands
6.	Count of numbers within Area 2
7.	Minimum size of numbers
8.	Ratio between hands
9.	Count of numbers within quadrant2
10.	Position of min hand
11.	Count of numbers which its orientation is More 25
12.	Position of Hour hand
13.	Arrows on hands
14.	Count of numbers within Area 1
15.	Stem of clock hands (near to the centre) is left out
16.	Count of numbers within quadrant 4
17.	Time is correct
18.	Ratio between max and min size
19.	Hands are joint or within 12 mm
20.	Displacement of arrows less than 4mm
21.	Hands connected with target number
22.	Displacement of minute hand or mar k from the target number
23.	Arrows are pointing to the wrong direction
24.	Count of numbers within quadrant 3
25.	Hands self-correction'
26.	Count of numbers within quadrant 1
27.	Displacement of hour hand or mar k from the target number
28.	Count of repeated or duplicated numbers
29.	Hour hands is present
30.	Minute hand is present
31.	Number self-correction
32.	Count of numbers outside the contour
33.	Presence of written words
34.	Picture of human face is drawn on clock
35.	Time is written outside the clock
36.	Time is written across the clock
37.	There are superfluous
38.	Straight line is used between the two numbers
39.	Time is indicated by writing minute number
40.	More than two hands are used'
41.	Direction of numbers written
42.	Numbers representation
43.	Number (11,3) aren’t present
44.	Sequence numbers not in sequence
45.	Sequential numbers are written following 12
46.	Number of numbers left out
47.	Count of numbers within Area 3

Table B5: Results of feature selection for AD / VaD discriminative task.

No.	Feature
1.	Maximum angles between numbers
2.	Distance between the position of hands intersection and the centre
3.	Angle between clock hands
4.	Count of numbers within Area 1
5.	maximum size of numbers
6.	Count of numbers which its orientation is More 25
7.	Position of min hand
8.	Count of numbers within Area 2
9.	minimum size of numbers
10.	Count of numbers within quadrant 1
11.	Count of numbers within quadrant 3
12.	Count of numbers within quadrant2
13.	Count of numbers within quadrant4
14.	Position of Hour hand
15.	numbers within Area 3
16.	Arrows on hands
17.	hands connected with target number
18.	repeated or duplicated numbers
19.	Number of numbers left out
20.	Ratio between max and min size
21.	Minimum angles between numbers
22.	Displacement of arrows less than 4mm
23.	Hands self-correction
24.	Count of numbers outside the contour
25.	Displacement of hour hand or mar k from the target number
26.	Time is correct
27.	Displacement of minute hand or mar k from the target number
28.	Arrows are pointing to the wrong direction
29.	Sequence numbers not in sequence
30.	Hands are joint or within 12 mm
31.	Stem of clock hands (near to the centre) is left out
32.	More than two hands are used
33.	Hour hands is present
34.	Direction of numbers written
35.	Minute hand is present
36.	Ratio between hands
37.	Number self-correction
38.	Sequential numbers are written following 12
39.	Number (11,3) aren't present
40.	There are superfluous
41.	Time is indicated by writing minute number
42.	Presence of written words
43.	Numbers representation
44.	Time is written outside the clock
45.	Time is written across the clock
46.	Straight line is used between the two numbers
47.	Picture of human face is drawn on clock

Appendix C

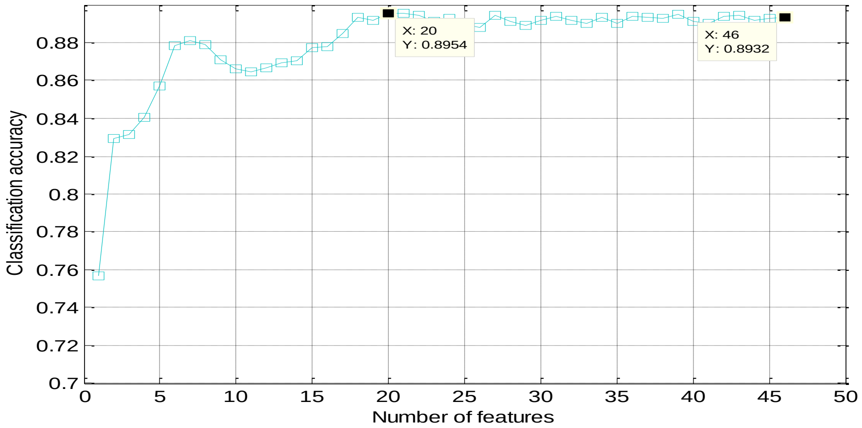
Normative Range of the Clock Features

Table C1: The normative range of the CDT features.

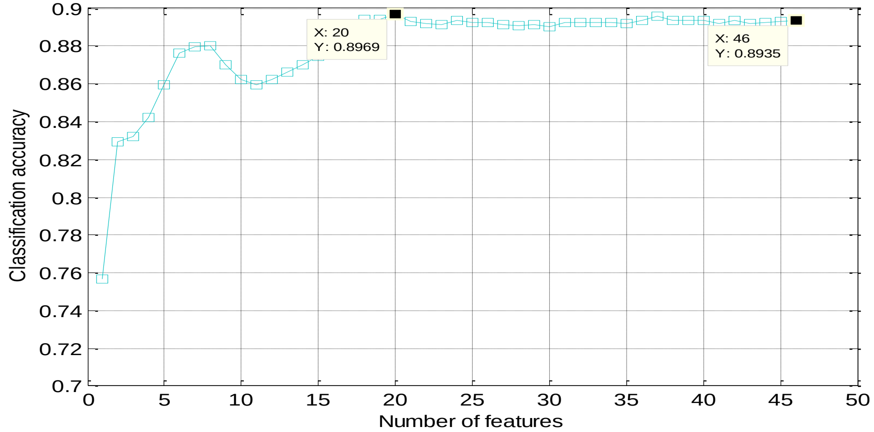
No	Feature	Normative
1	Count of numbers within area 1.	12
2	Count of numbers within area 2.	0
3	Count of numbers within area 3.	0
4	Count of numbers within quadrant 1.	(2,3,and 4)
5	Count of numbers within quadrant 2.	(2,3,and 4)
6	Count of numbers within quadrant 3.	(2, and 3)
7	Count of numbers within quadrant 4.	(2,3,and 4)
8	Minimum size of the numbers mm2.	[1-10]
9	Maximum size of the numbers mm2.	[15-150]
10	Ratio between the maximum number size and minimum	[2-20.8]
11	Count of numbers outside the contour.	0
12	Minimum angle between numbers.	[16-30]
13	Maximum angles between numbers.	[31-44]
14	Count of numbers whose rotation is over 25 degree.	0
15	Count of numbers left out from the drawing.	0
16	Count of duplicated numbers.	0
17	Sequential numbers are written following 12 (13, 14,	0
18	Numbers not in sequence.	0
19	Numbers 3 and 11 not present.	0
20	Arabic only numbers used.	1
21	Direction of written numbers.	1
22	Self-correction of numbers.	0
23	Minute hand is present.	1
24	Hour hand is present.	1
25	More than two hands are drawn.	0
26	Self-correction of hands.	0
27	Time is correct.	1
28	Time is indicated by writing minute number next to 3 or	0
29	Straight line is used between the two numbers.	0
30	Displacement of hour hand or mark from the target	0
31	Displacement of minute hand or mark from the target	0
32	Hands connected with target number.	0
33	Arrows on hands.	3
34	Displacement of arrows less than 4mm.	(-1,0, and 1)
35	Arrows are pointing in the wrong direction.	0
36	Presence of superfluous.	0
37	Hands are joint or within 12 mm.	1
38	Position of minute hand.	1
39	Position of hour hand.	3
40	Angle between clock hands.	[98-137]
41	Ratio between hands.	[0.71-2.77]
42	Presence of stem of clock hands (near to the center) is	0
43	Time is written across the clock.	0
44	Time is written outside the clock.	0
45	Picture of a human face is drawn on clock.	0
46	Presence of written words.	0
47	Distance between the position of hands intersection	[0-9]

Appendix D

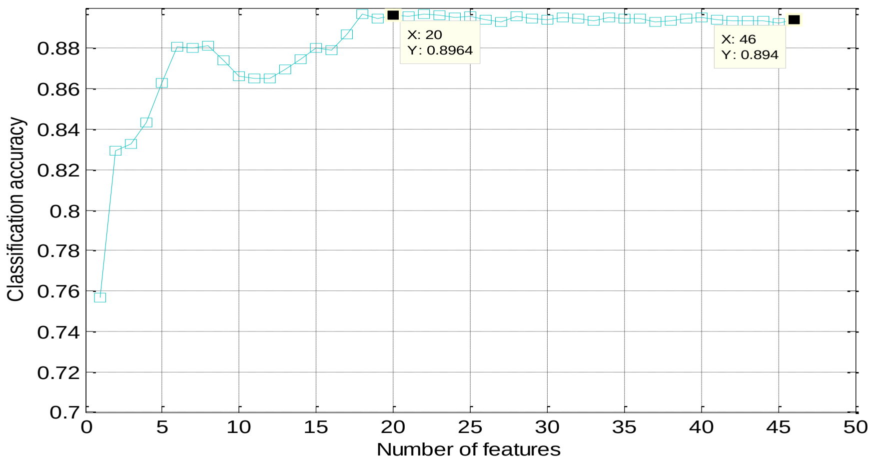
The Comparative study of Cascade Classifier



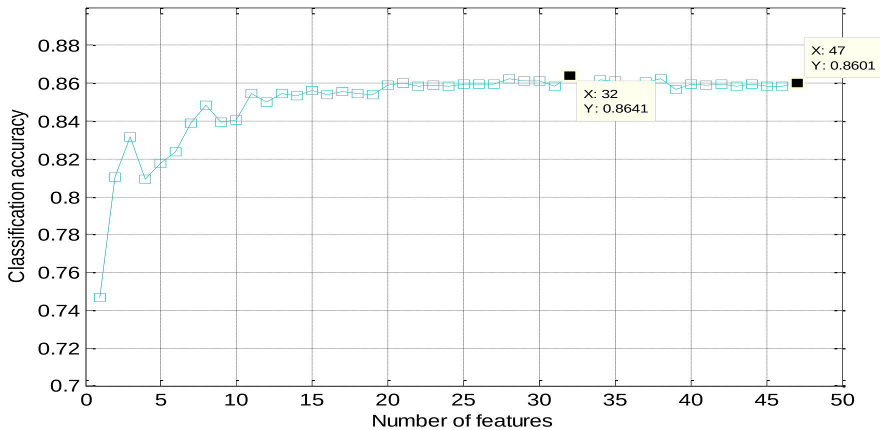
a- RF classifier, No. of sub-trees= 100, with MDL method



b- RF classifier, No. of sub-trees= 200, with MDL method

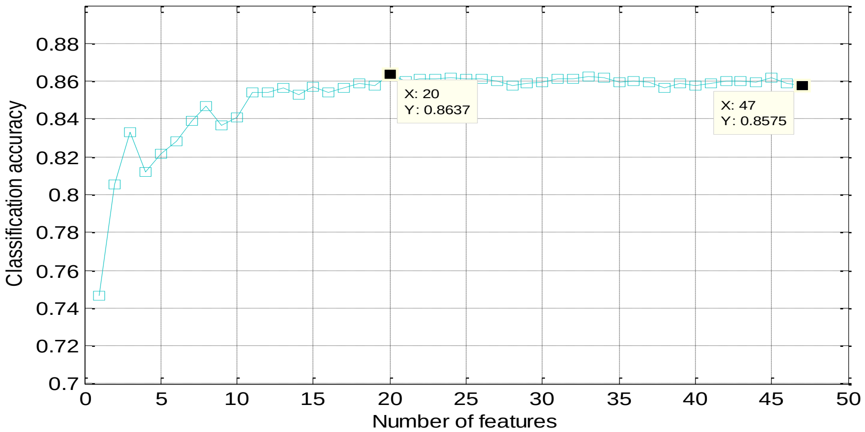


c- RF classifier, No. of sub-trees= 500, with MDL method

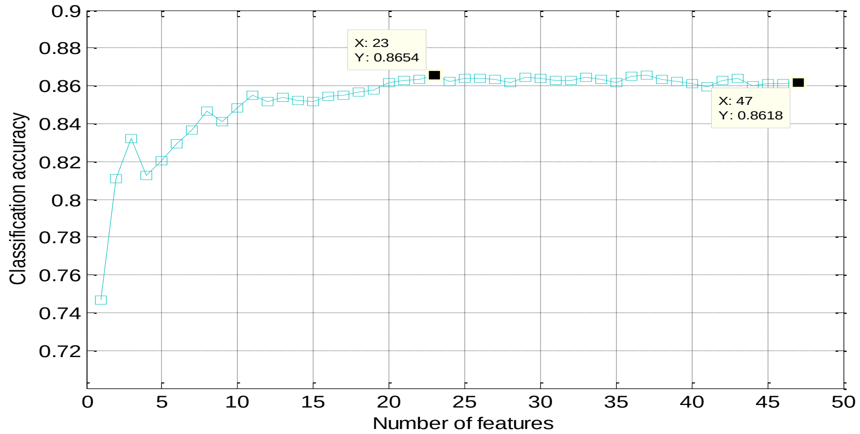


d- RF classifier, No. of sub-trees= 100, with EWD method

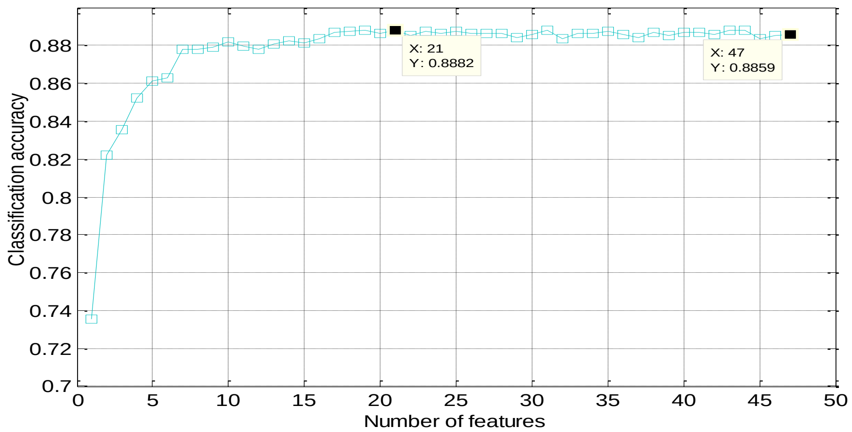
Figure D1: The classification accuracy of classifier 1 in the cascade structure.



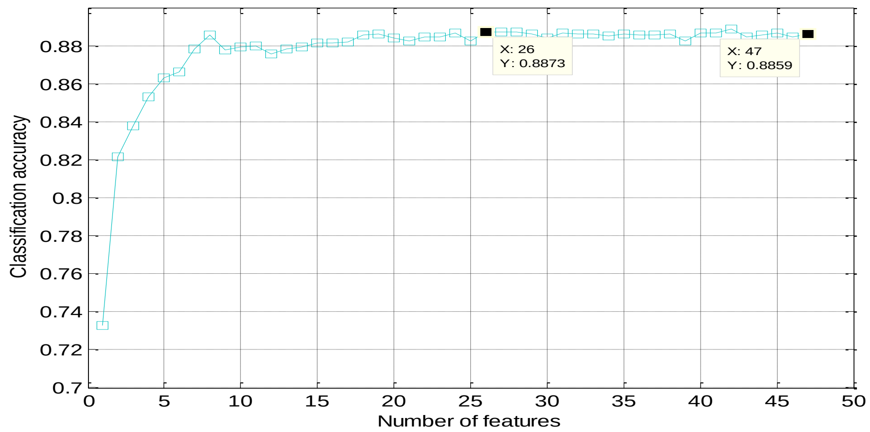
e- RF classifier, No. of sub-trees= 200, with EWD method



f- RF classifier, No. of sub-trees= 500, with EWD method

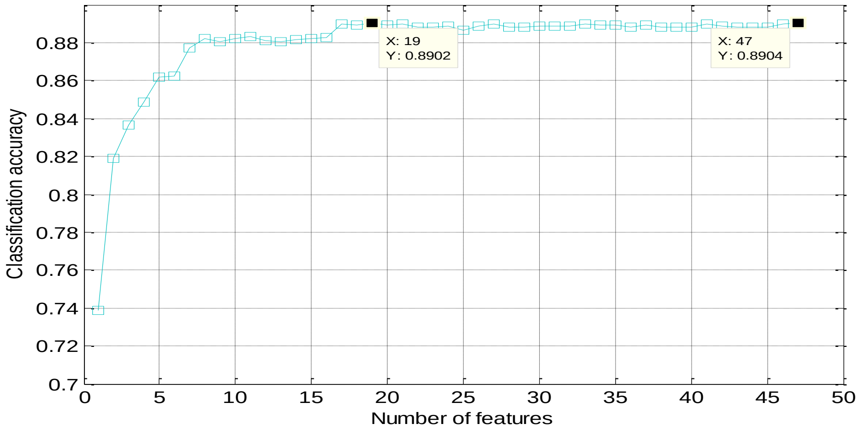


g- RF classifier, No. of sub-trees= 100, with EFD method

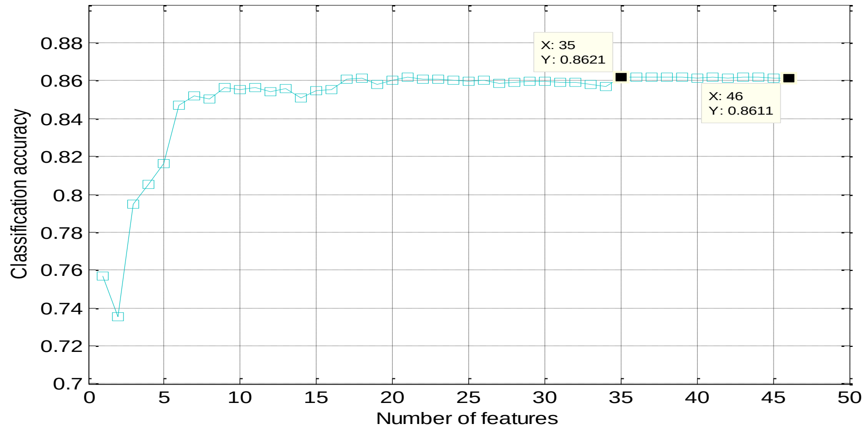


h- RF classifier, No. of sub-trees= 200, with EFD method

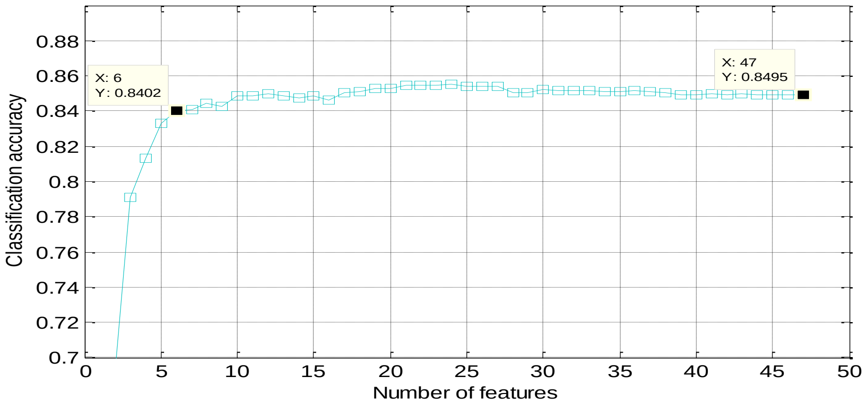
Figure D1: The classification accuracy of classifier 1 in the cascade structure.



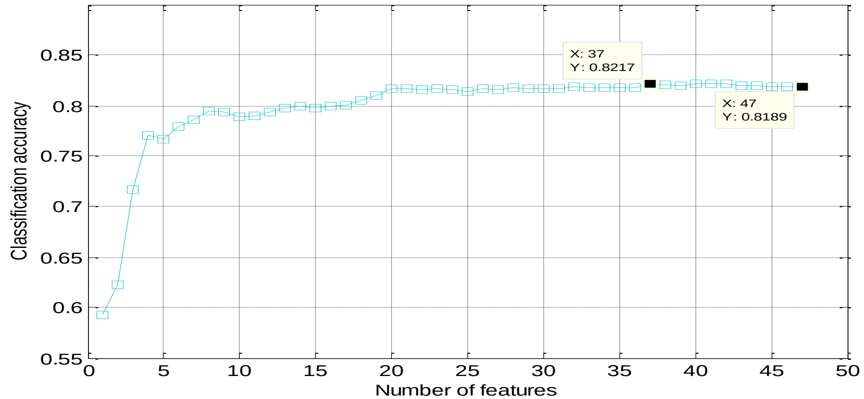
i- RF classifier, No. of sub-trees= 500, with EFD method



j- SVM classifier, with MDL method

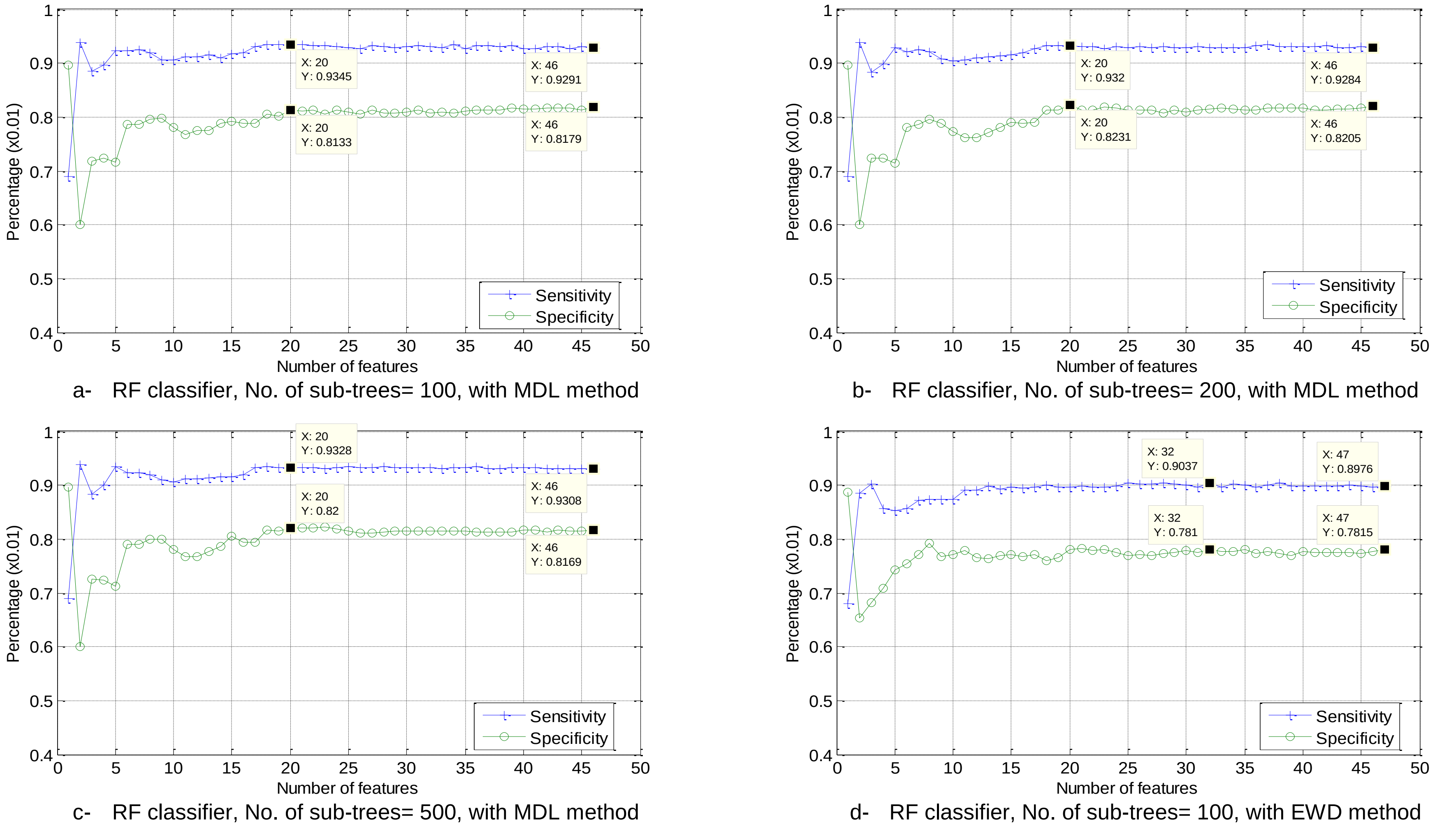


k- SVM classifier, with EFD method



l- SVM classifier, with EWD method

Figure D1: The classification accuracy of classifier 1 in the cascade structure.



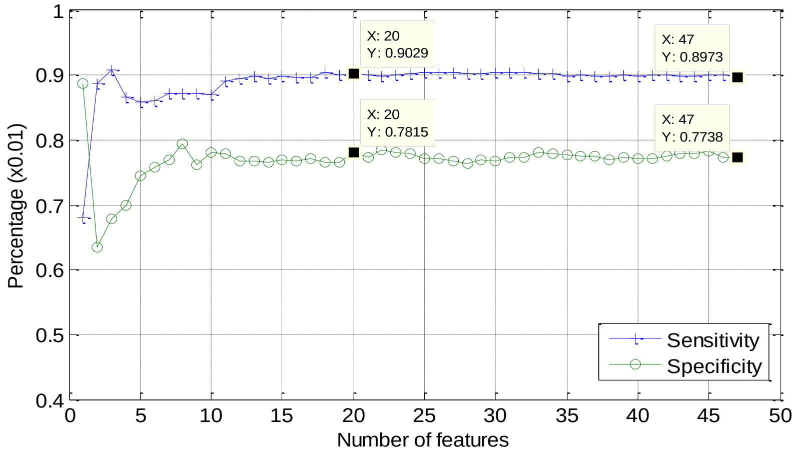
a- RF classifier, No. of sub-trees= 100, with MDL method

b- RF classifier, No. of sub-trees= 200, with MDL method

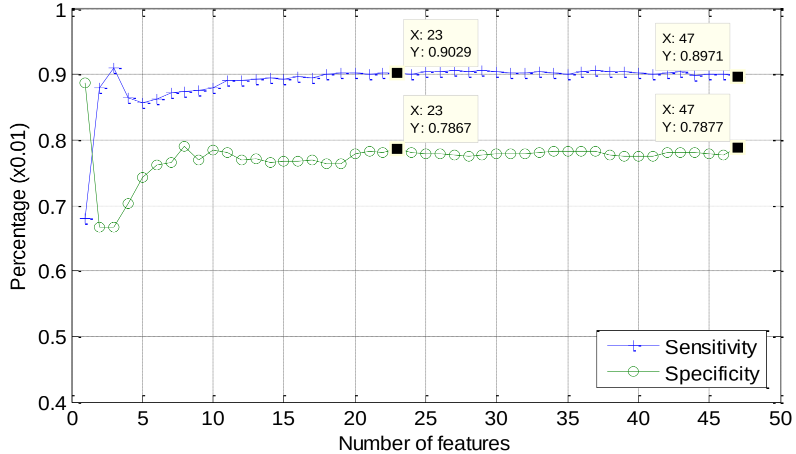
c- RF classifier, No. of sub-trees= 500, with MDL method

d- RF classifier, No. of sub-trees= 100, with EWD method

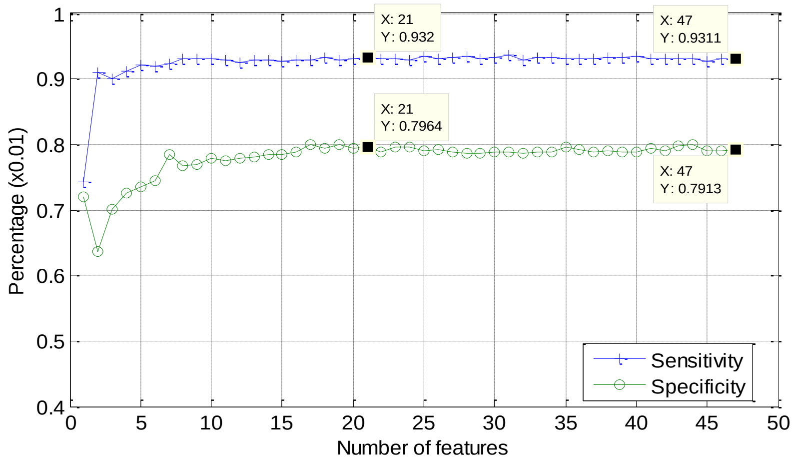
Figure D2: The sensitivity and specificity of classifier 1 in the cascade structure.



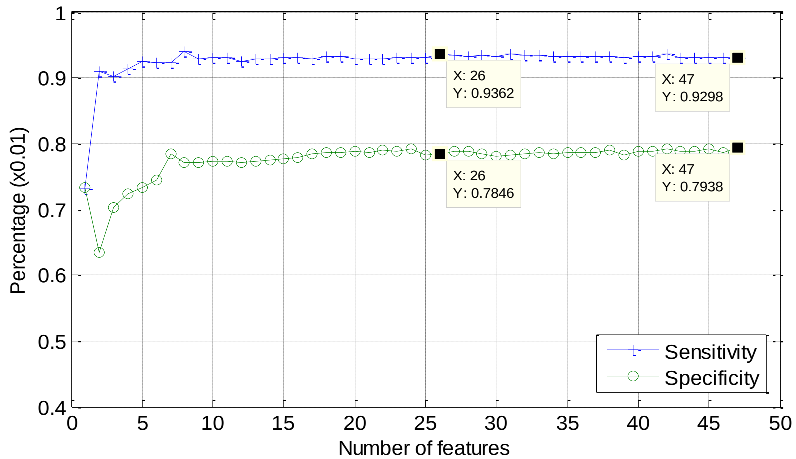
e- RF classifier, No. of sub-trees= 200, with EWD method



f- RF classifier, No. of sub-trees= 500, with EWD method

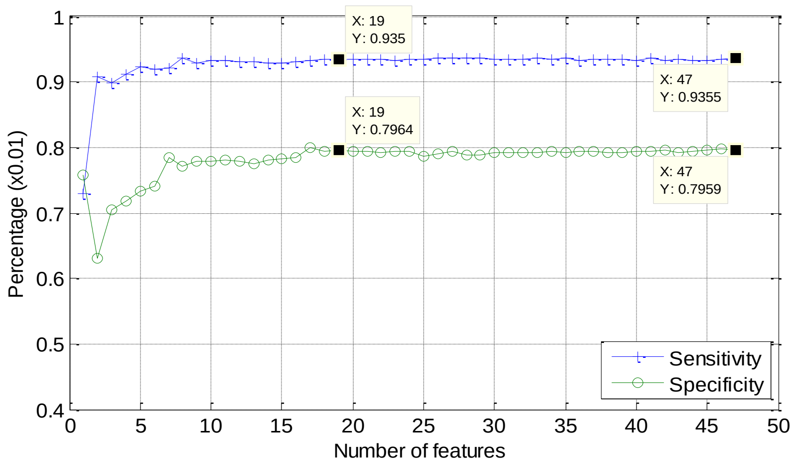


g- RF classifier, No. of sub-trees= 100, with EFD method

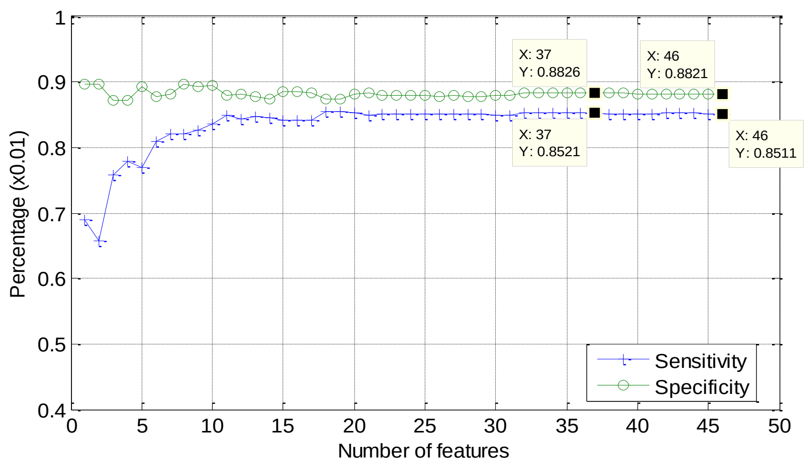


h- RF classifier, No. of sub-trees= 200, with EFD method

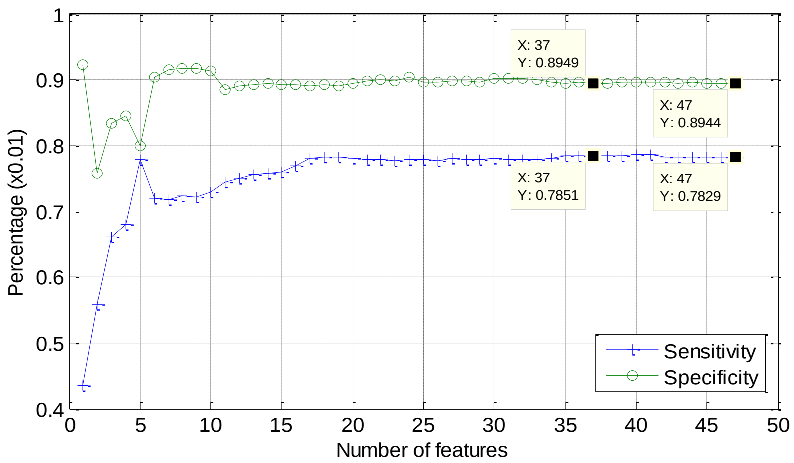
Figure D2: The sensitivity and specificity of classifier 1 in the cascade structure.



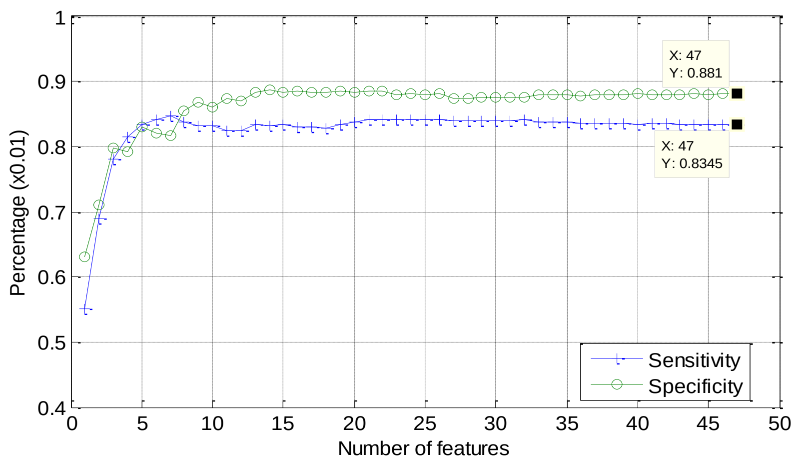
i- RF classifier, No. of sub-trees= 500, with EFD method



j- SVM classifier, with MDL method

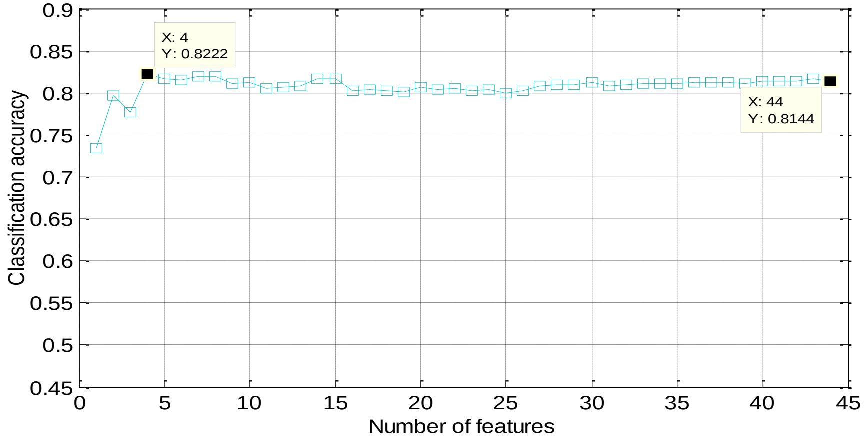


k- SVM classifier, with EWD method

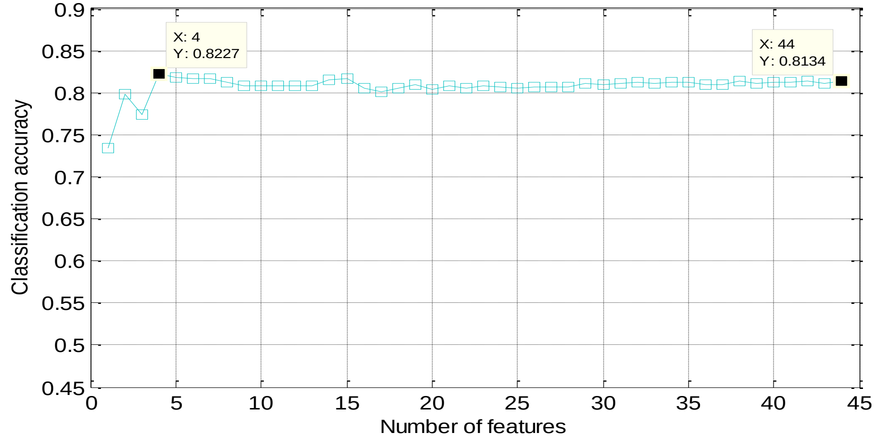


l- SVM classifier, with EFD method

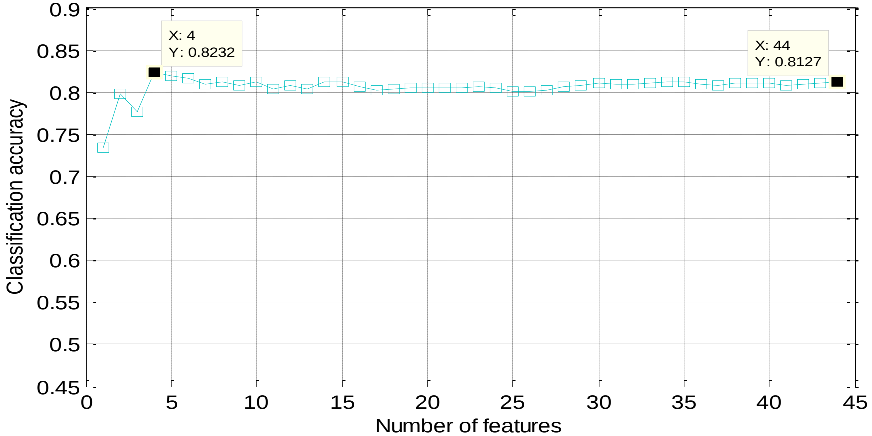
Figure D2: The sensitivity and specificity of classifier 1 in the cascade structure.



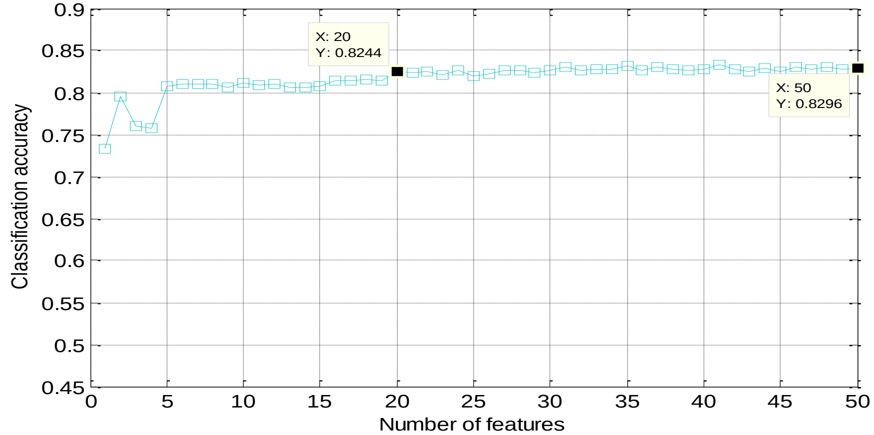
a- RF classifier, No. of sub-trees= 100, with MDL method



b- RF classifier, No. of sub-trees= 200, with MDL method

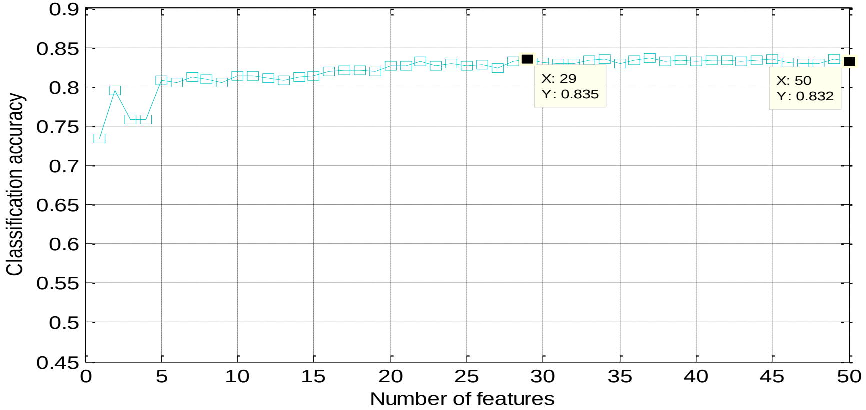


c- RF classifier, No. of sub-trees= 500, with MDL method

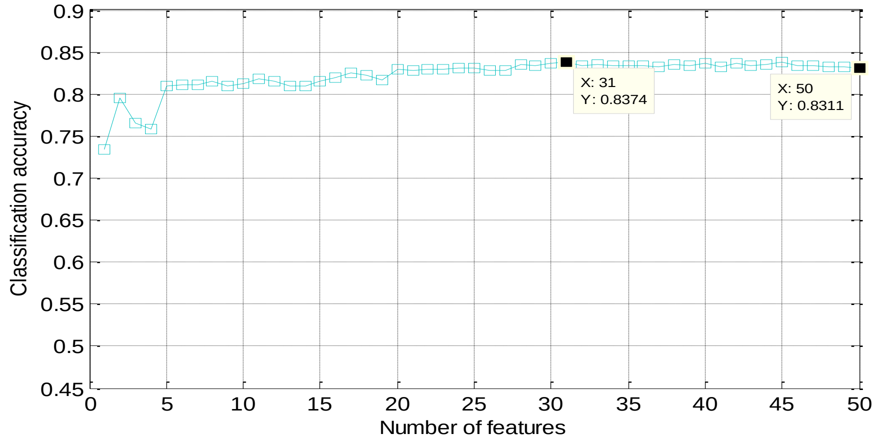


d- RF classifier, No. of sub-trees= 100, with EWD method

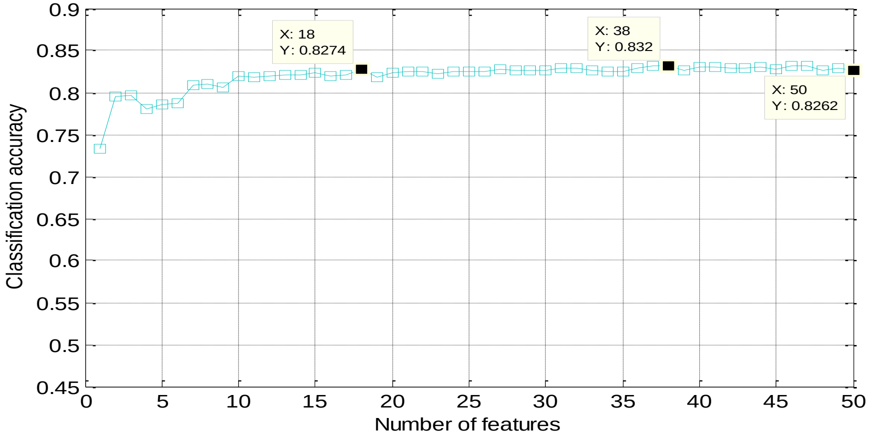
Figure D3: The classification accuracy of classifier 2 in the cascade structure



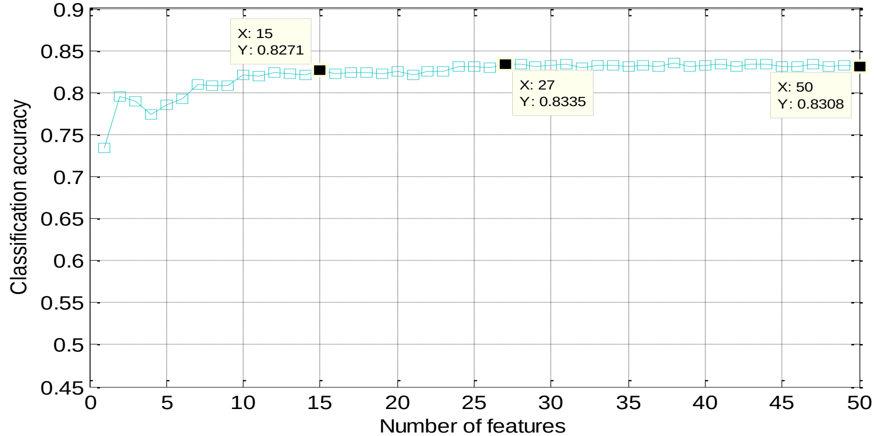
e- RF classifier, No. of sub-trees= 200, with EWD method



f- RF classifier, No. of sub-trees= 500, with EWD method

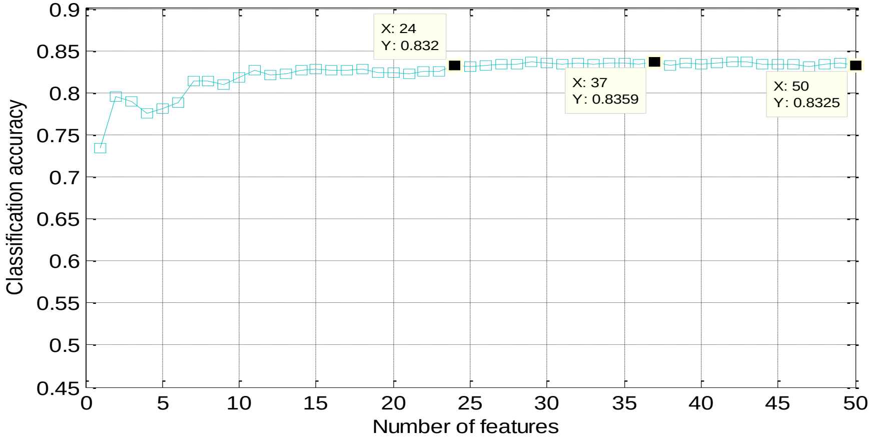


g- RF classifier, No. of sub-trees= 100, with EFD method

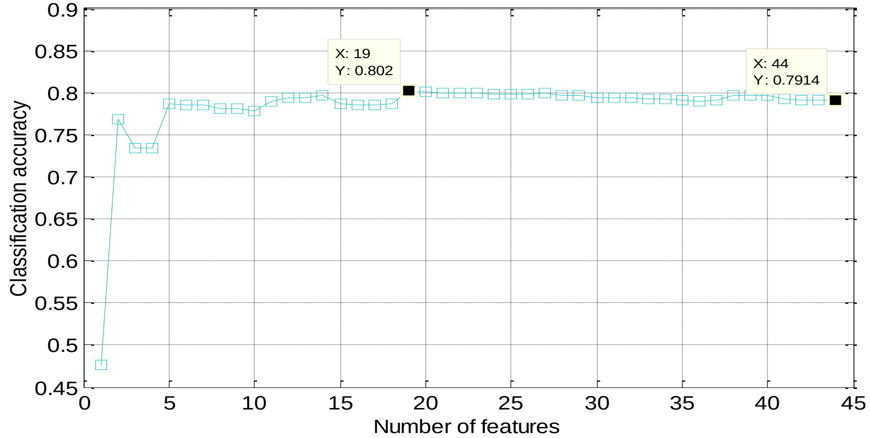


h- RF classifier, No. of sub-trees= 200, with EFD method

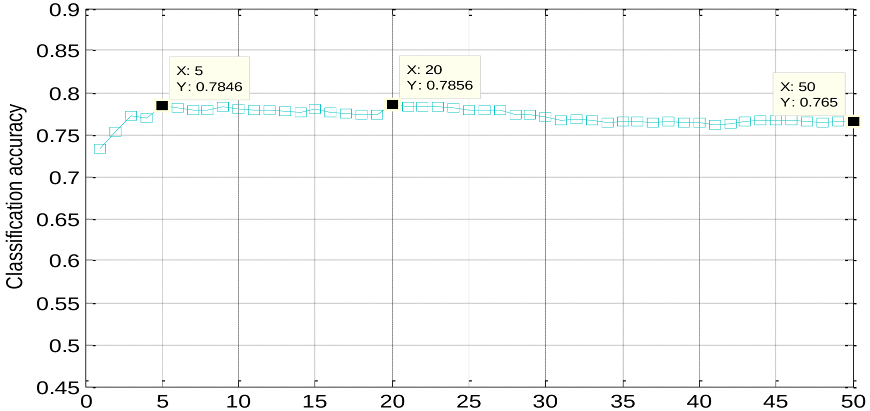
Figure D3: The classification accuracy of classifier 2 in the cascade structure



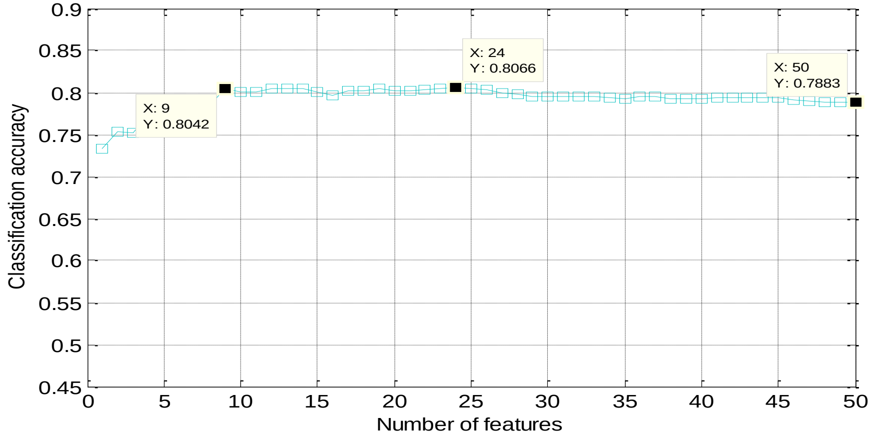
i- RF classifier, No. of sub-trees= 500, with EFD method



j- SVM classifier, with MDL method

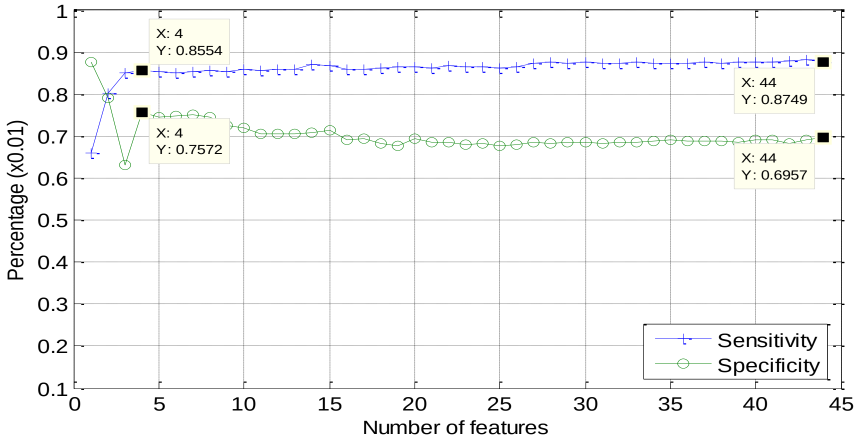


k- SVM classifier, with EWD method

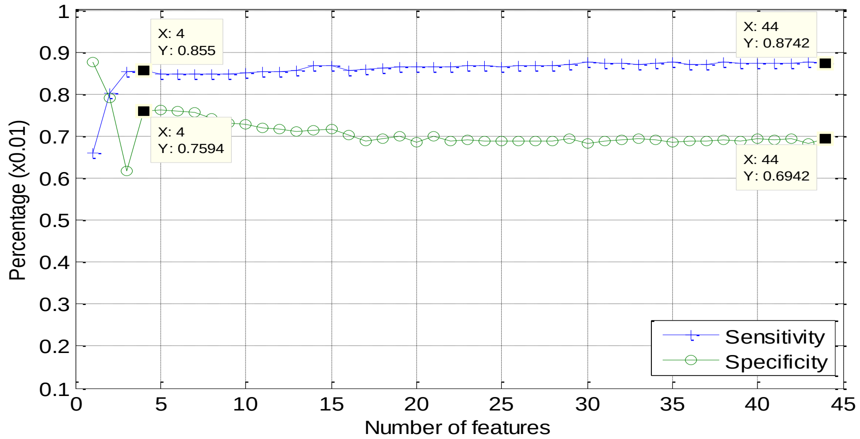


l- SVM classifier, with EFD method

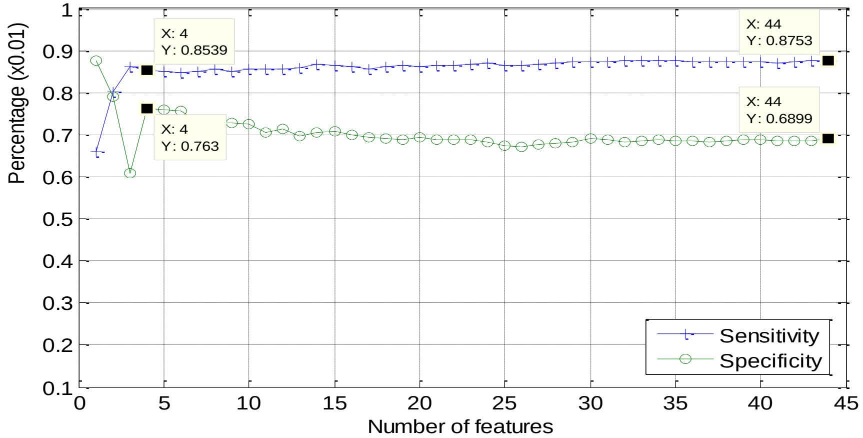
Figure D3: The classification accuracy of classifier 2 in the cascade structure



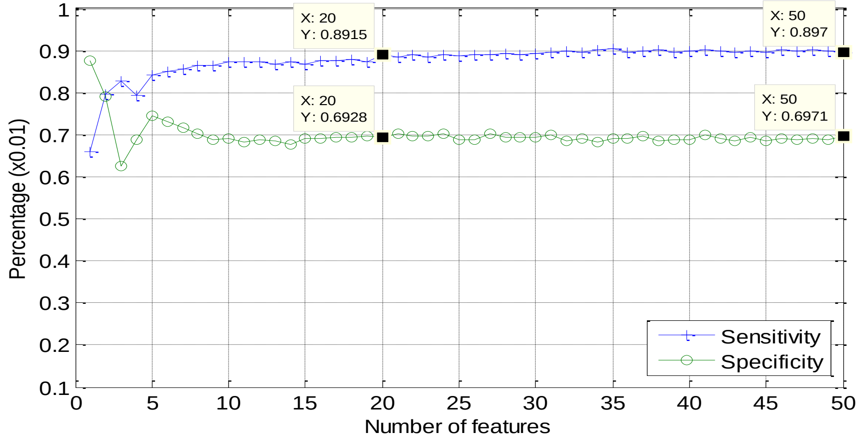
a- RF classifier, of sub-trees= 100, with MDL method



b- RF classifier, of sub-trees= 200, with MDL method

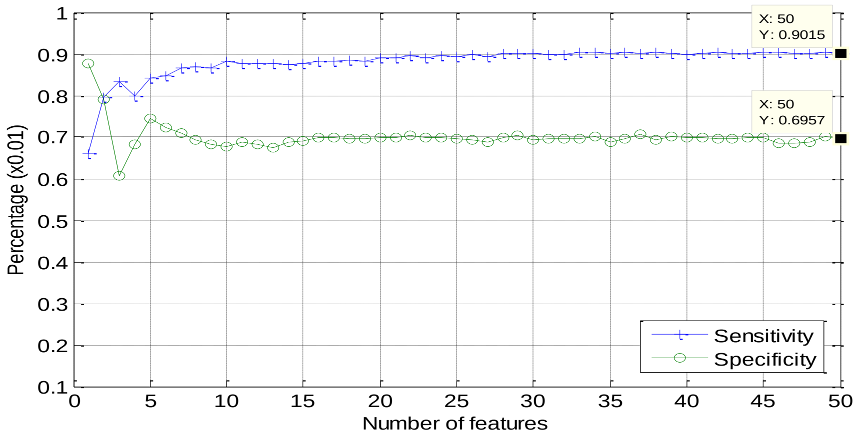


c- RF classifier, of sub-trees= 500, with MDL method

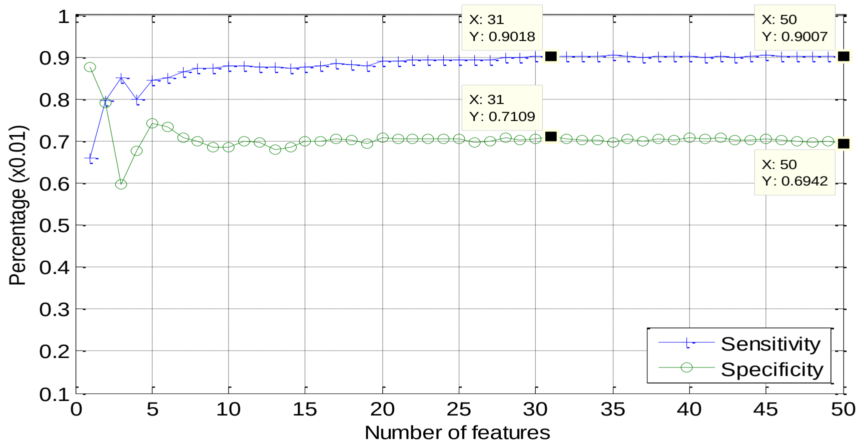


d- RF classifier, of sub-trees= 100, with EWDmethod

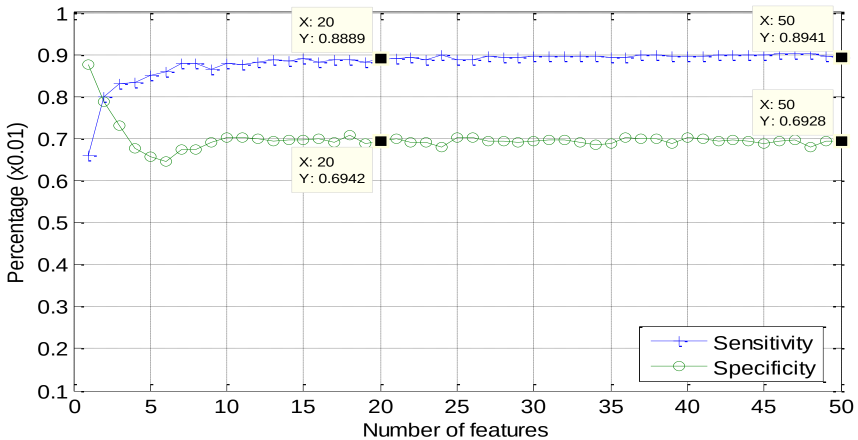
Figure D4: The sensitivity and specificity of classifier 2 in the cascade structure



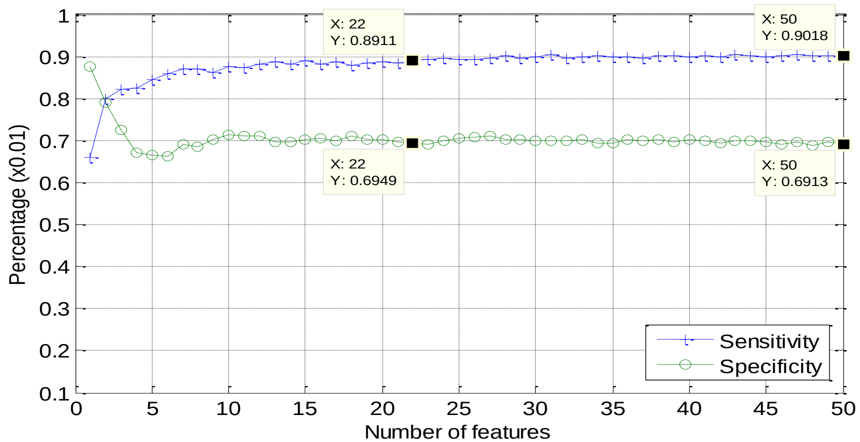
e- RF classifier, of sub-trees= 200, with EWD method



f- RF classifier, of sub-trees= 500, with EWD method

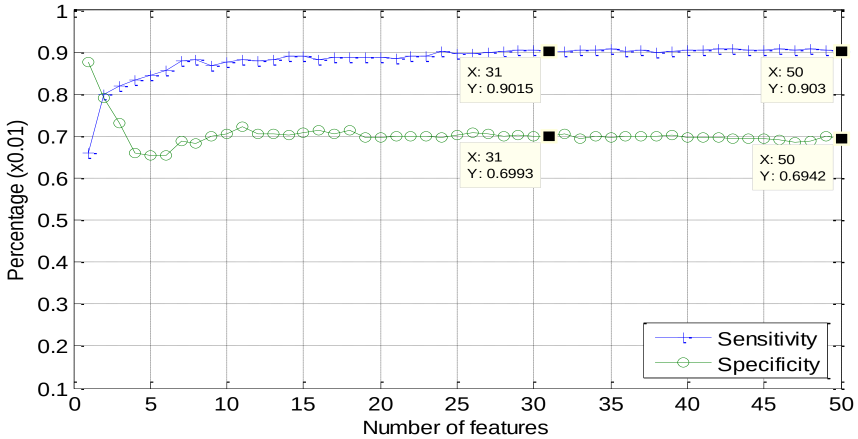


g- RF classifier, of sub-trees= 100, with EFD method

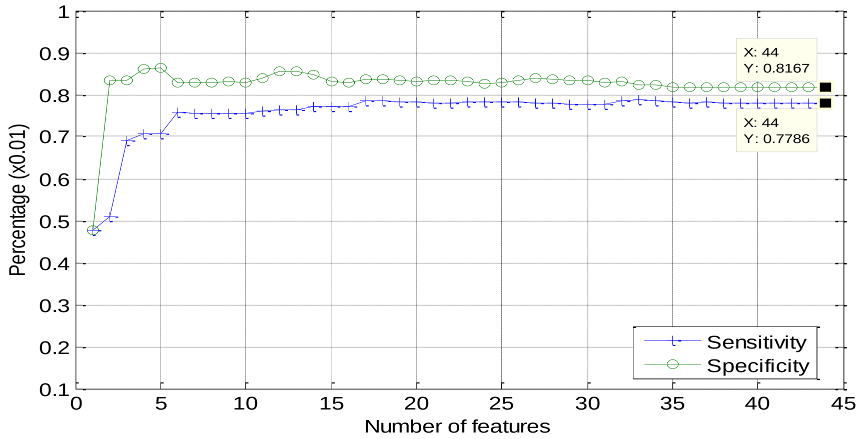


h- RF classifier, of sub-trees= 200, with EFDmethod

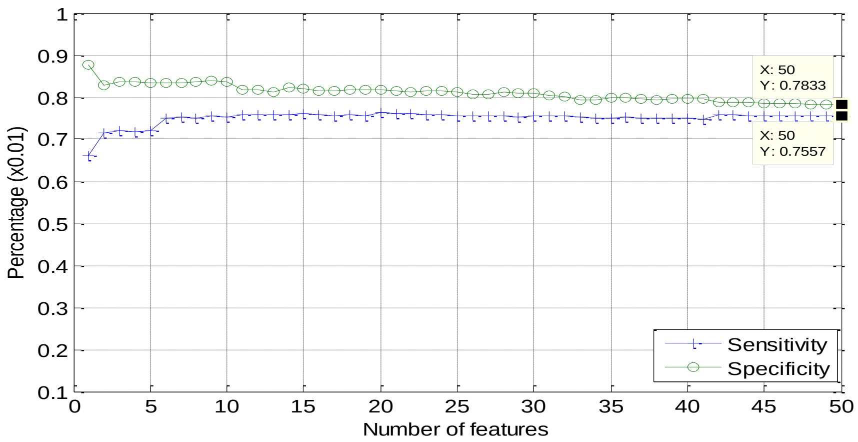
Figure D4: The sensitivity and specificity of classifier 2 in the cascade structure



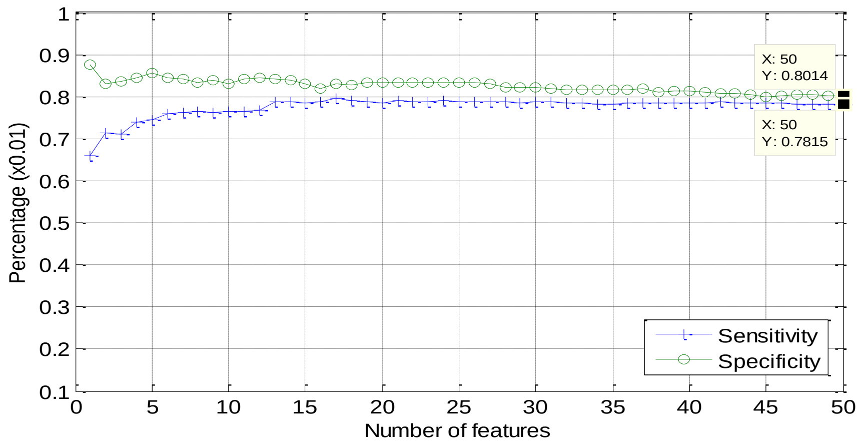
i- RF classifier, of sub-trees= 500, with EFD method



j- SVM classifier, with MDL method

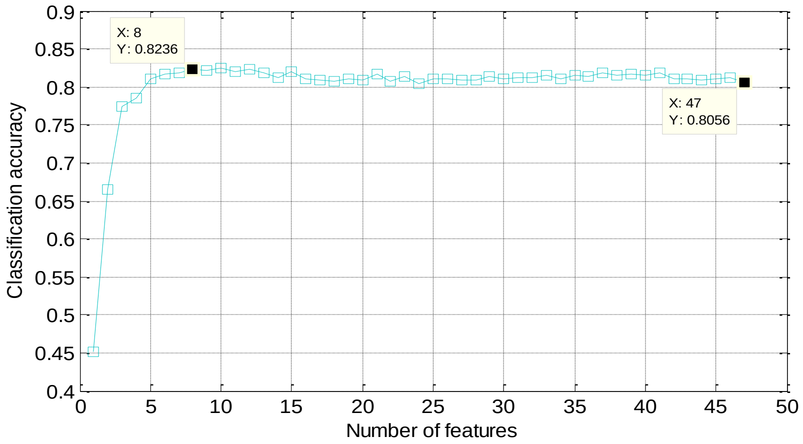


k- SVM classifier, with EWD method

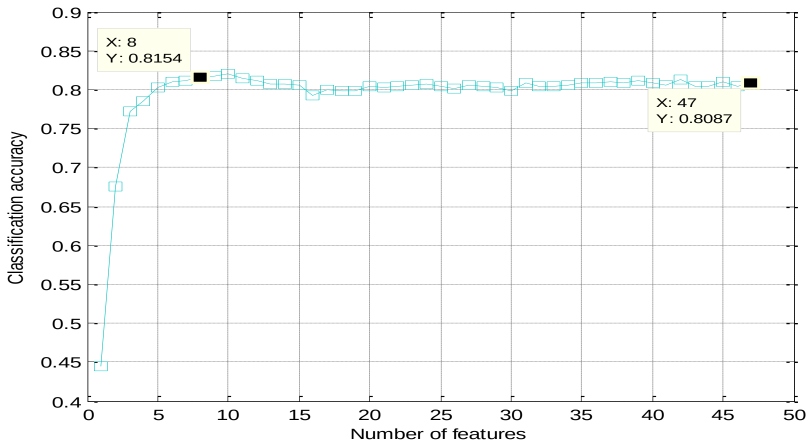


l- SVM classifier, with EWD method

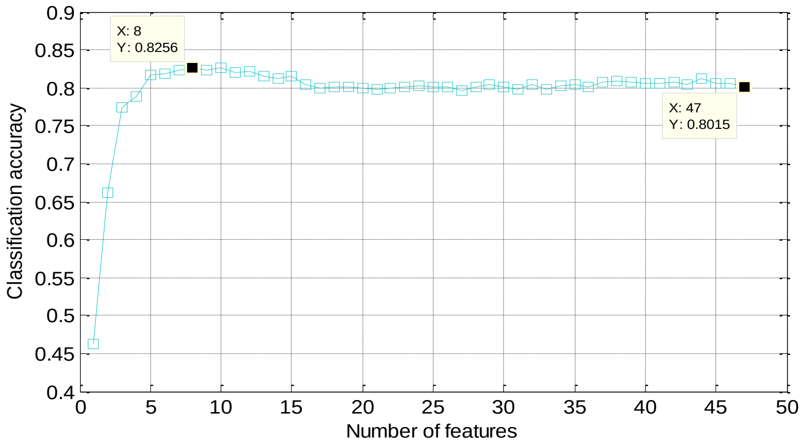
Figure D4: The sensitivity and specificity of classifier 2 in the cascade structure



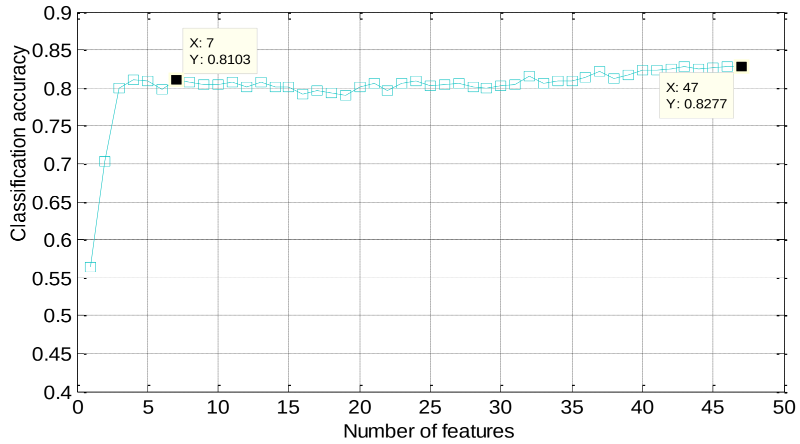
a- RF classifier, No. of sub-trees= 100, with EWD method



b- RF classifier, No. of sub-trees= 200, with EWD method

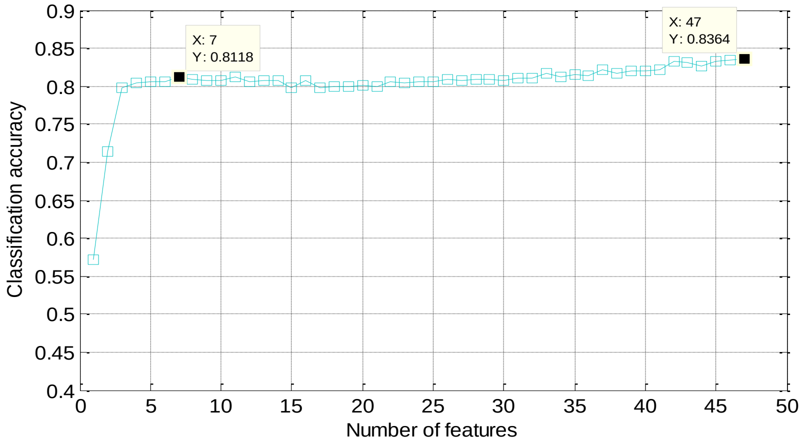


c- RF classifier, No. of sub-trees= 500, with EWD method

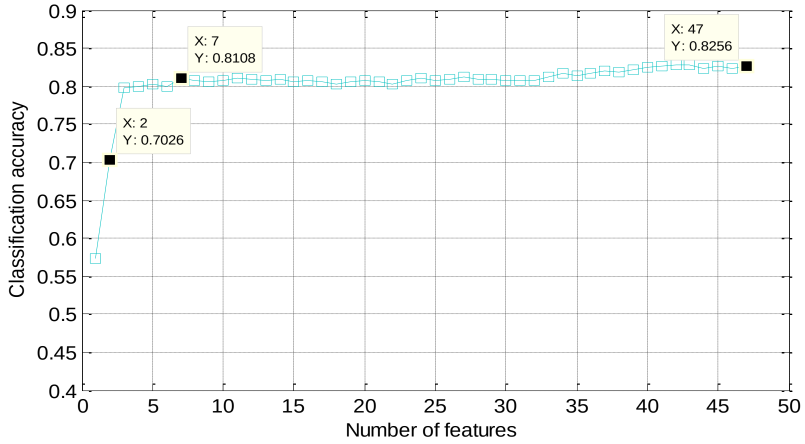


d- RF classifier, No. of sub-trees= 100, with EFD method

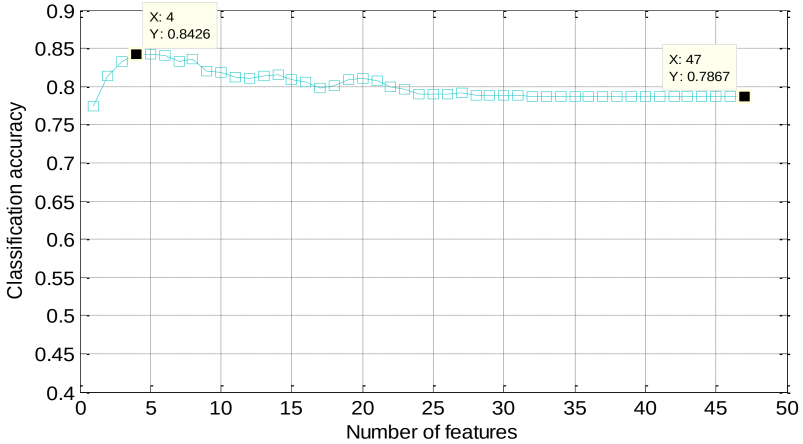
Figure D5: The classification accuracy of classifier 3 in the cascade structure



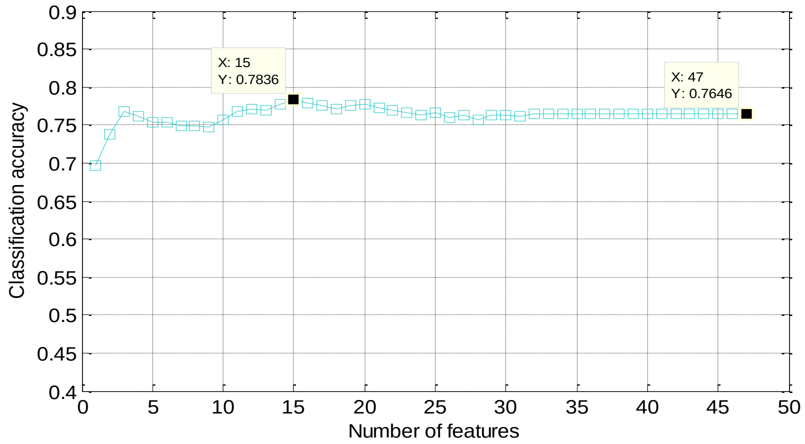
e- RF classifier, No. of sub-trees= 200, with EFD method



f- RF classifier, No. of sub-trees= 500, with EFD method

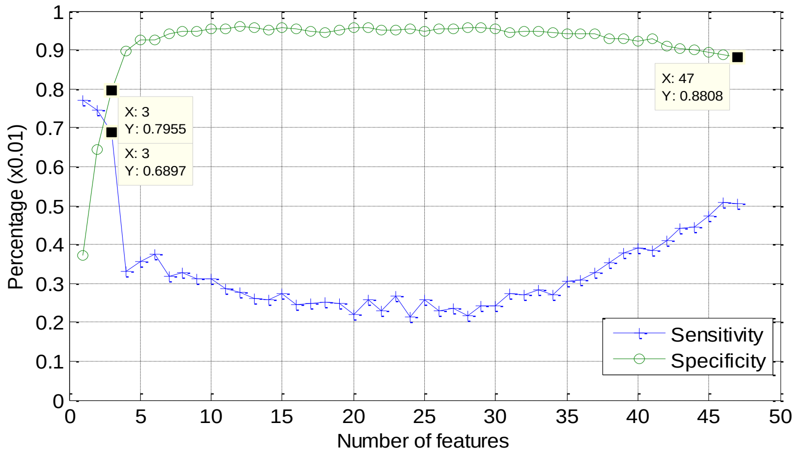


g- SVM classifier, with EWD method

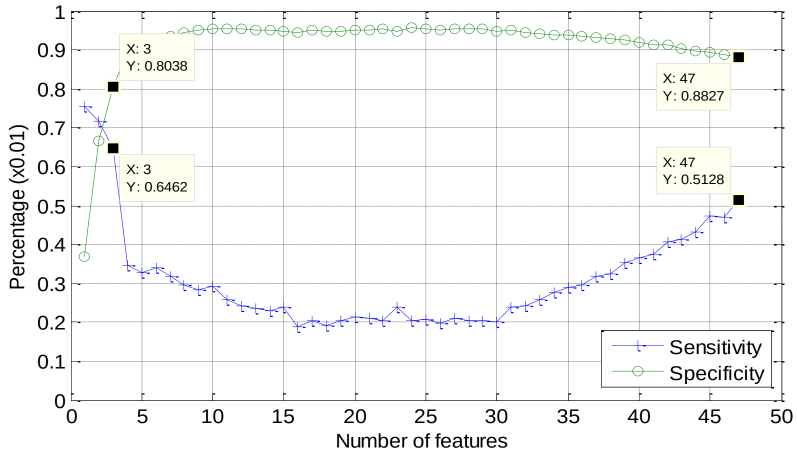


h- SVM classifier, with EFD method

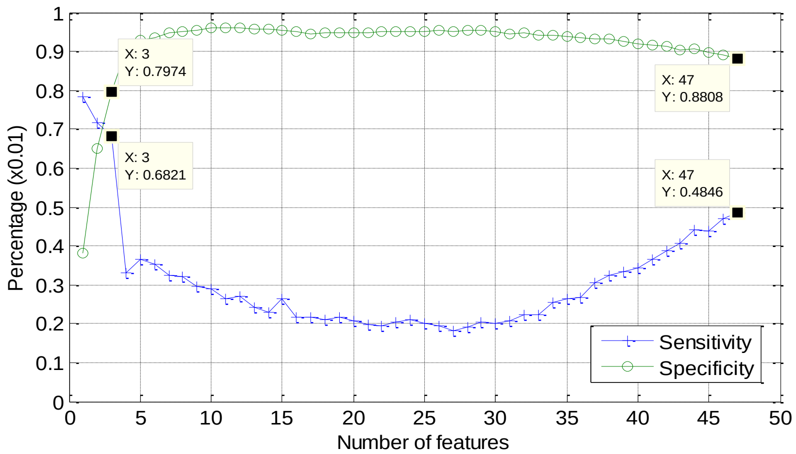
Figure D5: The classification accuracy of classifier 3 in the cascade structure



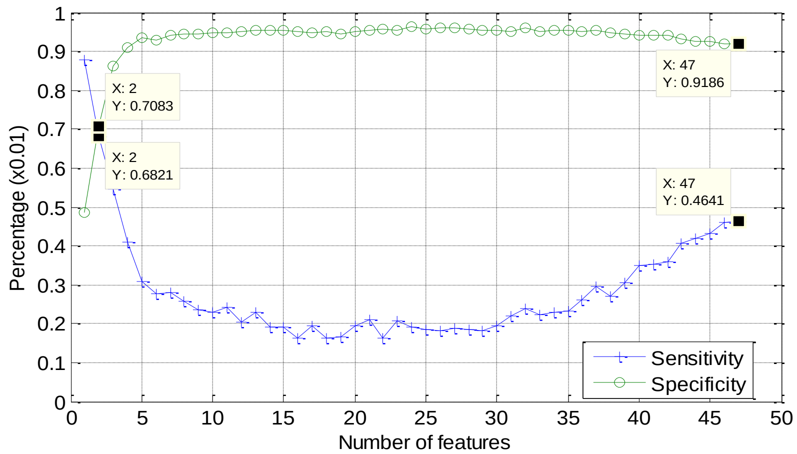
a- RF classifier, No. of sub-trees= 100, with EWD method



b- RF classifier, No. of sub-trees= 200, with EWD method

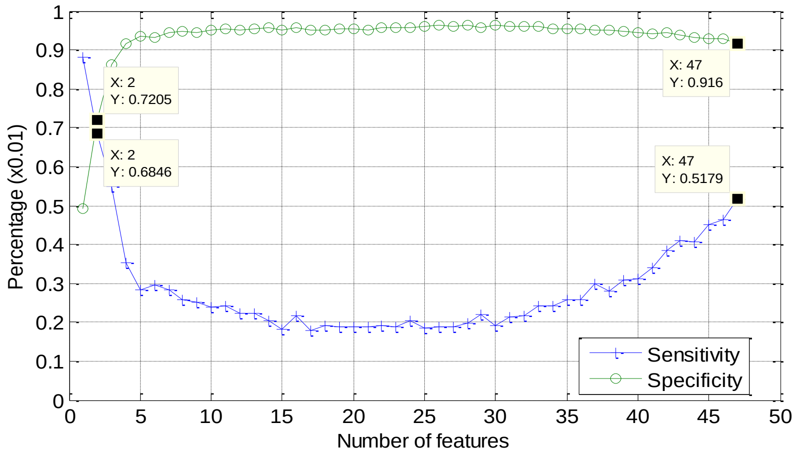


c- RF classifier, No. of sub-trees= 500, with EWD method

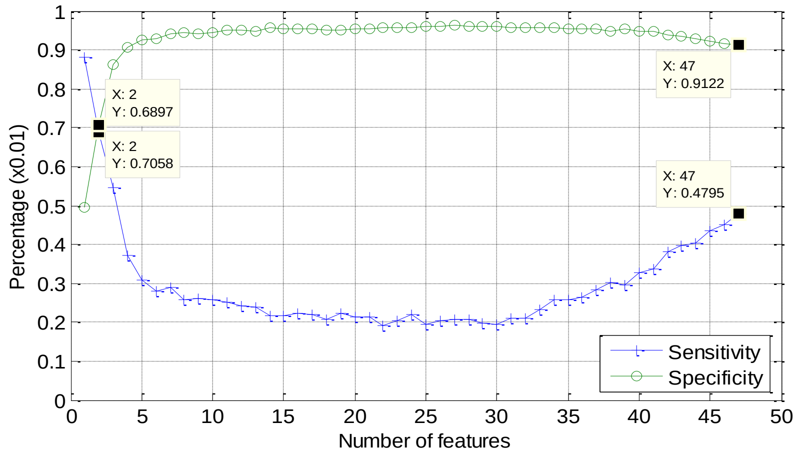


d- RF classifier, No. of sub-trees= 100, with EFD method

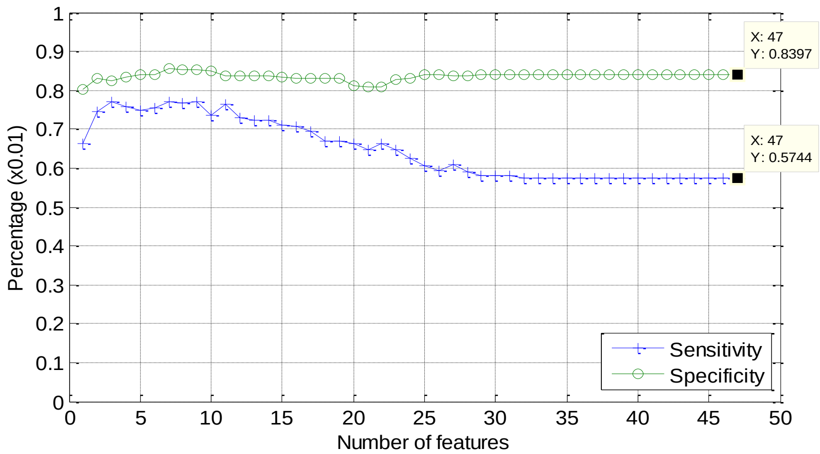
Figure D6: The sensitivity and specificity of classifier 3 in the cascade structure



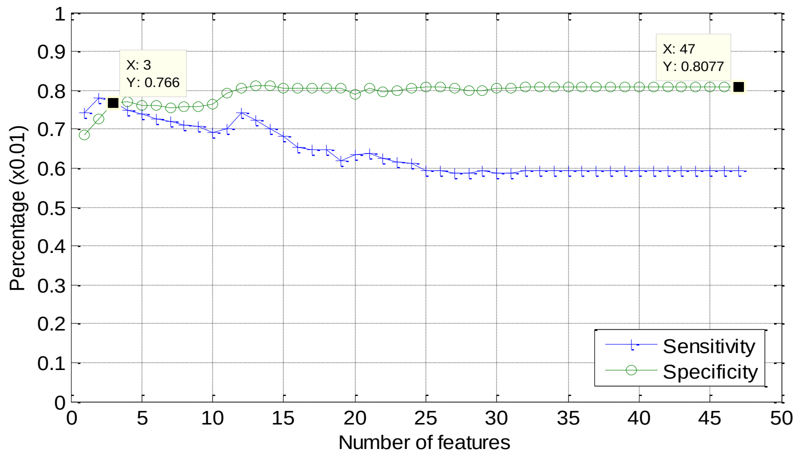
e- RF classifier, No. of sub-trees= 200, with EFD method



f- RF classifier, No. of sub-trees= 500, with EFD method



g- SVM classifier, with EWD method



h- SVM classifier, with EFD method

Figure D6: The sensitivity and specificity of classifier 3 in the cascade structure