Automatic Interpretation of Clock Drawings for Computerised Assessment of Dementia

Zainab Harbi

Cardiff University

School of Engineering

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

Doctor of Philosophy

September 2017

All praise is due to Allah,the most gracious, the most merciful.

Peace and blessings be upon our prophet Mohammed and upon his family and companions

******

To My Family

Declaration and Statemnets

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ……………………………………………… (candidate) Date …………………

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of ……… (insert MCh, MD, MPhil, PhD etc, as appropriate)

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

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

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

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

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

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

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

Abstract

The clock drawing test (CDT) is a standard neurological test for detection of cognitive impairment. A computerised version of the test has potential to improve test accessibility and accuracy. CDT sketch interpretation is one of the first stages in the analysis of the computerised test. It produces a set of recognised digits and symbols together with their positions on the clock face. Subsequently, these are used in the test scoring. This is a challenging problem because the average CDT taker has a high likelihood of cognitive impairment, and writing is one of the first functional activities to be affected. Current interpretation systems perform less well on this kind of data due to its unintelligibility. In this thesis, a novel automatic interpretation system for CDT sketch is proposed and developed. The proposed interpretation system and all the related algorithms developed in this thesis are evaluated using a CDT data set collected for this study. This data consist of two sets, the first set consisting of 65 drawings made by healthy people, and the second consisting of 100 drawings reproduced from drawings of dementia patients.

This thesis has four main contributions. The first is a conceptual model of the proposed CDT sketch interpretation system based on integrating prior knowledge of the expected CDT sketch structure and human reasoning into the drawing interpretation system. The second is a novel CDT sketch segmentation algorithm based on supervised machine learning and a new set of temporal and spatial features automatically extracted from the CDT data. The evaluation of the proposed method shows that it outperforms the current state-of-the-art method for CDT drawing segmentation. The third contribution is a new

handwritten digit recognition algorithm based on a set of static and dynamic features extracted from handwritten data. The algorithm combines two classifiers, fuzzy *k*-nearest neighbour's classifier with a Convolutional Neural Network (CNN), which take advantage both of static and dynamic data representation. The proposed digit recognition algorithm is shown to outperform each classifier individually in terms of recognition accuracy.

The final contribution of this study is the probabilistic Situational Bayesian Network (SBN), which is a new hierarchical probabilistic model for addressing the problem of fusing diverse data sources, such as CDT sketches created by healthy volunteers and dementia patients, in a probabilistic Bayesian network. The evaluation of the proposed SBN-based CDT sketch interpretation system on CDT data shows highly promising results, with 100% recognition accuracy for heathy CDT drawings and 97.15% for dementia data.

To conclude, the proposed automatic CDT sketch interpretation system shows high accuracy in terms of recognising different sketch objects and thus paves the way for further research in dementia and clinical computer-assisted diagnosis of dementia.

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, Dr. Yulia Hicks, and Prof. Rossitza Setchi not only for their constant assistance, excellent guidance, and thoughtful advices, but also for their constant support and encouragement. I would also like to thanks my co-supervisor Prof. Antony Bayer from Llandaough hospital for his valuable advice.

I would like also to thank my colleagues in Knowledge Engineering Systems (KES) group, Dr. Bennaser, Dr. Romli, Dr. Anuar, Dr. Saadi, and Mr. O'Regan for stimulating discussion and their kind support.

I would like also to gratefully acknowledge establishment of martyrs, republic of Iraq-presidency of ministers for my PhD sponsorship.

Finally, I would like to express my sincere gratitude and appreciation to my late father, and my beloved family: my mother, sisters, husband, for their endless love, support and prayers. Thanks also to my beloved kids Mohammed and Zahraa.

Zainab Harbi

September 2017

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CDT	Clock Drawing Test
AI	Artificial Intelligence
ML	Machine Learning
SVM	Support Vector Machine
kNN	*k*-Nearest Neighbours
DNN	Deep Neural Networks
CNN	Convolutional Neural Network
BN	Bayesian Network
SBN	Situational Bayesian Network
MLP	Multi-Layer Perceptron
MNIST	Modified National Institute of Standards and Technology
UCI	University of California, Irvine
MatConvNet	Matlab Convolutional Network
RBF	Radial Basis Function
CPT	Conditional Probabilistic Table
SSBN	Structured Situational Bayesian Network
USBN	Unstructured Situational Bayesian Network

List of Publications

Publications as a main author:

- Harbi, Z., Hicks, Y., Setchi, R. and Bayer, T., 2015. “Segmentation of Clock Drawings Based on Spatial and Temporal Features”, 19th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Singapore, pp.1640-1648.

- Harbi, Z., Hicks, Y. and Setchi, R., 2016. “Clock Drawings Test Digit Recognition Using Static and Dynamic Features”, 20th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, York, United Kingdom, pp.1221-1230.

- Harbi, Z., Hicks, Y. and Setchi, R., 2017. “Clock Drawings Test Interpretation System”, 21 International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Marseille, France, pp.1641-1650.

The author has also contributed to other publications in related work. These publications were not a part of this thesis, these are:

- Shigemori, T., Harbi, Z., Kawanaka, H., Hicks, Y. and Setchi, R., 2015. “Feature Extraction Method for Clock Drawing Test”, 19th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Singapore, pp.1707-1714.

- Shigemori, T., Harbi, Z., Kawanaka, H., Hicks, Y. and Setchi, R., 2015. “A Study on Feature Extraction Method for Clock Drawing Images Using Weighted Direction Index Histogram”, 31st Fuzzy System Symposium, Chofu, September 2-4, 2015.

Chapter: 1

Introduction

1.1 Motivation

The Clock Drawing Test (CDT) has been used for decades as a screening tool to differentiate normal individuals from those with cognitive impairment. The CDT assesses the cognitive domains of comprehension, planning, visuospatial ability, visual memory, motor programming and execution, abstraction, concentration and response inhibition. The major benefit of the CDT is to provide a concrete visual reference of the patient's cognitive dysfunction associated with neurological disorders such as Alzheimer's and Parkinson's diseases and other dementia types.

At present, a CDT is usually administrated in a hospital environment by a clinician. Patients are asked to draw a clock by using a pencil on a sheet of paper. The most common instructions are: 'Draw a clock face in a pre-drawn circle and place the numbers on it. Then set the time to a specific time [usually five minutes to three].' After that, the clinicians would spend hours analysing and scoring the tests using one of a number of subjective scoring systems that have a set of analytical rules. These systems have different degrees of complexity, which can range from simple binary rating to more complex qualitative and quantitative systems that capture a wide variety of errors in the drawn clock. This sort of assessment can often result in the assignment of a specific ‘score’ to the patient’s output, based on which conclusion about the patient’s status can be made.

The main drawbacks of these scoring systems is their subjectivity and the time required for evaluation. In addition, not a single scoring system has been universally accepted as the most effective. Moreover, the inter-rater reliability in scoring can undermine effective objective assessment. Finally, these assessments almost always relate to the analysis of the end-product of the patient activity (i.e., the finished test output, seen as a complete entity, which is the drawn clock); they rarely attempt to provide an indicator of the underlying execution process. Investigating behavioural dysfunction during the test’s execution can provide important extra information about the impairments attributable to various neurological conditions.

Recently, a number of computerised systems for administering the CDT have been proposed, such as online data capture using smart pens or graphic tablets. Moving towards a system of online data capture and automated assessment can bring significant solutions to the shortcomings noted above. In addition to eliminating the problem of subjective scoring, online data capture is able to acquire additional dynamic data, whereby sequential and constructional information about the drawing process is collected and stored.

When administrating the test using a digital pen, touch screen tablet or digitiser, the test is recorded as a sequence of time-stamped drawing pixels. These input devices enable the capture of new information regarding the completion and thinking time, the planning strategies and the pressure and inclination of the pen on the paper (or the stylus on the tablet surface). In previous research, such data was shown to be significant when assessing the mental state of the patients. Supplementing the computerised CDT with online data capture offers an advantage for diagnosing impairments related to various neurological disorders.

With computer-based CDT, the problem of automatic assessment has been highlighted. The assessment is based on identifying abnormalities in the drawings, which may include poor number positioning, omission of numbers, incorrect sequencing, missing clock hands and the presence of irrelevant writing. An automatic CDT sketch interpretation system is required so that the assessment is effective and reliable. The system can be challenged by unexpected input from a patient with a cognitive disorder. The limitations of the current state of the art make it difficult for an automated system to interpret clock drawings in a manner that matches human cognition. For example, the labelling of each of the sketched objects such as clock hours, hands or other drawing artefacts is still considered a challenging task.

Automatic sketch interpretation has received considerable attention in recent years as a means of communicating between humans and machines. Sketch-based communication is very attractive to people, as drawing by hand provides an easy way to input information. Sketch interpretation is a special case of image interpretation in which two general approaches can be identified. There is the classical recognition approach—in which object recognition depends solely on the object's features—and knowledge-based where additional features relating to the context is used in the recognition process. Acknowledging the inherent trade-off between drawing freedom and interpretation accuracy, a number of successful approaches in limited interpretation domains have been developed, which benefit from exploiting the contextual information. A CDT interpretation system could further utilise the advantages of prior knowledge, along with possible interpretations from the classical recognition systems, to improve the accuracy of recognition.

However, interpreting a CDT sketch produces new challenges regarding the abnormalities of the test context. As this test is designed for assessing cognitive impairment, a poorly-drawn object is to be expected. The main challenges in any CDT interpretation system are: first, segmenting the CDT sketches into objects; second, recognising these objects; and finally, deriving inferences from prior knowledge in order to improve and complete the CDT drawing interpretation. All these challenges have been addressed in this research and a solution is proposed in each area.

In order to recognise objects in a CDT sketch, accurate segmentation of the sketch is required. Segmentation is an important step in any recognition system, since it directly influences the recognition accuracy. A drawing of a clock is a special case of handwriting, but in terms of segmentation, it is much more difficult. This difficulty is due to the fact that in a CDT, the writing does not proceed in one direction, but changes direction arbitrarily. Moreover, in cognitive impairment cases, most objects' aspects and spaces between elements are not preserved. Thus, the traditional handwriting segmentation algorithm would not be applicable in this area, and so a new algorithm is required. The supervised learning-based segmentation method is a new approach, and its application in a free handwriting platform would be attractive to resolving the CDT segmentation problem. Furthermore, a suitable segmentation system would consider the advantage of acquired temporal information recorded by the online data capture of computerised CDT system in addition to the conventional spatial features.

In this research, most of the sketched objects are numbers, so the focus is on handwriting recognition. Recognising the handwriting of people with cognitive impairment is challenging, since their writing skills are often affected. Handwriting recognition is a large area in which many algorithms for both offline and online handwriting have been developed. Previous studies in computerised CDT have highlighted the handwriting recognition problem, but since their focus was on data capturing, they used existing algorithms. These algorithms relied on the shape or position of the objects which, in the case of CDT drawings from people with moderate or severe dementia, could lead to misclassifications due to the unusual positions or shapes of the clock digits. Hence, a new algorithm is required to tackle the problem of recognising CDT sketch objects. Offline handwriting recognition systems focus on the static image features, while online systems consider dynamic features related to the process of writing. Combining multiple classifiers, with each one an expert on a particular feature, has been a successful means of handwriting recognition and could improve the accuracy of CDT object recognition.

Integrating a state-of-the-art recognition system with domain-specific reasoning would improve the interpretation system's performance. The more a dementia patient deviates from the expected baseline, the more challenging becomes CDT sketch interpretation. Patients can sketch elements in the wrong position, repeat numbers or omit them completely. They can also add components that are neither hands nor numerals. Thus, a flexible and comprehensive reasoning system is required. In terms of image interpretation, there are two general approaches: the rule-based or expert system; and the probabilistic-based. Both approaches have their advantages and disadvantages, but within their application area, there has been little work investigating them. Moreover, the unstructured nature of abnormal CDT sketches (i.e. those produced by severe dementia patients) can introduce new challenges to applying these approaches. Thus, both avenues have been investigated in this research to highlight their functioning in this area.

1.2 Aims and Objectives

The scope of this project is a system for interpreting CDT sketches. The project aims to develop an automatic CDT sketch interpretation system as a step towards a computerised CDT. This system is envisaged to assist in the detection of dementia symptoms at the point of care. This will be achieved by developing an intelligent interpretation system based on artificial intelligence (AI) and machine learning techniques.

The specific objectives necessary to achieve the aim are identified as:

1. Creation of a conceptual model for an automatic CDT sketch interpretation system that can assist in developing the computerised CDT system for automatic assessment and early diagnosis of dementia.

2. Development of a novel algorithm to segment the CDT sketches into a number of objects, based on spatial and temporal features.

3. Development of a new handwriting recognition system that can recognise bad handwriting that is expected within the problem context.

4. Development of a knowledge-based reasoning approach for CDT sketch interpretation that can integrate domain knowledge with reasoning to improve the interpretation process. Two approaches for reasoning to be considered are the rule-based and probabilistic-based approaches.

1.3 Thesis Outline

This thesis is organised into the following structure:

- Chapter 1 provides an introduction to the work.

- Chapter 2 presents a review of the background information surrounding CDT and the computerised system, and discusses related work in the field. It also reviews the relevant literature related to image interpretation, image segmentation, object recognition, knowledge representation and reasoning systems.

- Chapter 3 proposes a conceptual model for the CDT sketch interpretation system. It also provides an overview of the process of collecting clock sketch data and the computerised CDT.

- Chapter 4 introduces a novel segmentation algorithm for segmenting the CDT sketches.

- Chapter 5 proposes a new recognition system, along with its validation by using Pendigits, a benchmarking data set.

- Chapter 6 introduces the knowledge rule-based approach for automatic CDT sketch interpretation.

- Chapter 7 introduces the probabilistic approach for automatic CDT sketch interpretation and compares it to the rule-based approach.
- Chapter 8 highlights the contributions, and conclusions of this thesis, and proposes further work.

Chapter: 2

Literature Review

This chapter reviews previous research related to the work presented in this thesis. In particular, research in the areas of manual and computerised CDT assessment is reviewed in Section 2.2. Related areas of automatic sketch interpretation and machine learning methods are considered in the context of automatic CDT interpretation in Sections 2.3 and 2.4. Research on handwriting segmentation and recognition is reviewed in Sections 2.5 and 2.6. Finally, methods for knowledge representation, reasoning systems and situation assessment are reviewed in Sections 2.7–2.9. Section 2.10 summarises the chapter.

2.1 Clock Drawing Test

The clock drawing test has been widely employed as a tool for the assessment of cognitive abilities and neurological disorders over the past 30 years (Shulman, Shedletsky and Silver, 1986; Rouleau *et al.*, 1992; Juby, Tench and Baker, 2002; Pinto and Peters, 2009; Spenciere, Alves and Charchat-fichman, 2017). It is used as a stand-alone test or as a part of general screening for cognitive changes such as: the Mini-Cog screening (Borson *et al.*, 1999); the Cambridge Cognitive Examination (CAMCOG) (Schmand *et al.*, 2000); and the Montreal Cognitive Assessment (MoCA) (Charbonneau, Whitehead and Collin, 2005). 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 is widely employed as a follow-up instrument because it can be easily and quickly applied, is well accepted by patients and is relatively independent of language, education and culture (Shulman, 2000).

During the test, the individual is asked to draw a clock face by using a pencil on a provided sheet of paper, then draw the hands to a specific time. The task of drawing a clock seems a simple one; however, to a neurologist it is a complex task, one that requires many neurological functions involving various brain regions. The CDT assesses the cognitive domains of comprehension, planning, visuospatial ability, visual memory, motor programming and execution, abstraction, concentration and response inhibition (Freedman *et al.*, 1994). The CDT has been shown to be a useful screening tool to differentiate normal individuals from those with cognitive impairments such as Alzheimer's and other dementia conditions (Pinto and Peters, 2009; Mazancova *et al.*, 2016; Vyhnálek *et al.*, 2016). Figure 2-1 shows an example of clocks drawn by patients with different types of dementia.

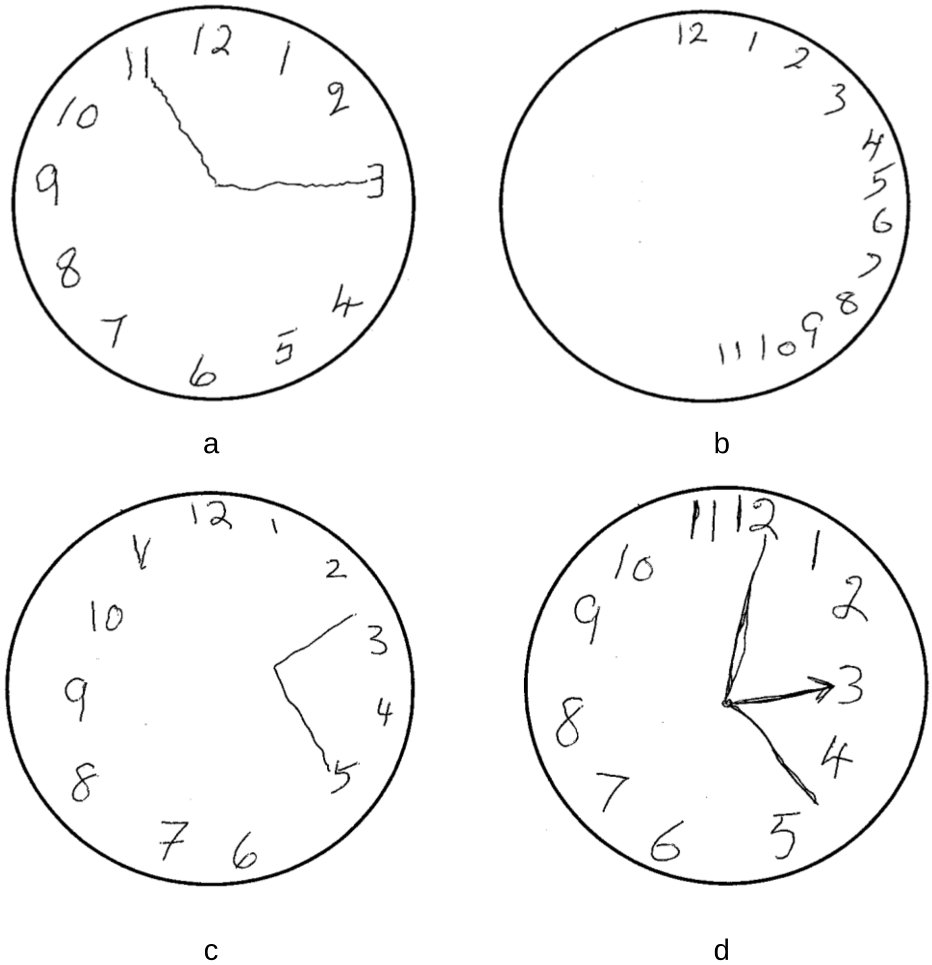

Figure 2-1: Examples of clocks drawn by patients at Llandough Hospital, Cardiff, UK: (a) Normal elderly; (b) Alzheimer's disease; (c) Mild dementia; (d) Vascular dementia.

There are two ways in which individuals may be asked to perform the test. They could be given either 1) a verbal command to draw the clock and set the hands to a specific time; or 2) instructions to copy a given template. There is a difference between these methods in terms of the cognitive function required to understand the instructions and perform the test (Freedman *et al.*, 1994). The verbal command is often further categorised into three different forms: 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 which 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 with the numbers and time; and 3) the examiner CDT, where the individual is given a clock drawing with all the numbers written on it, then asked to set the hands to a specific time. The free-hand CDT has been criticised for the influence of the clock's contour; for example, if the circle is drawn too small or is distorted, then the rest of the drawing will also be impaired (Pinto and Peters, 2009).

The time setting on the clock can also be varied in a CDT. The time settings most sensitive to neurocognitive dysfunction, and hence the most widely used, are '10 past 11', '20 past 8', and '5 to 3'. For these 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.

Currently, the clinical administration of the CDT in a hospital environment is performed by medical practitioners such as neuropsychologists, neurologists and clinical nurses (Shulman and Feinstein, 2005; Grober *et al.*, 2008). The clinicians score the tests manually using various subjective scoring systems that have a set of analytical rules. There are many scoring systems, which have been developed to assess clock drawings and to diagnose cognitive impairments from the drawings (Shulman, Shedletsky and Silver, 1986; Mendez, Ala and Underwood, 1992; Rouleau *et al.*, 1992; Tuokko *et al.*, 2000; Jouk, 2007; Parsey and Schmitter-Edgecombe, 2011; Mendes-santos *et al.*, 2015; Ricci *et al.*, 2016). Each of these systems places an emphasis on a subset of cognitive functions. They have a different range of complexities, from a simple binary rating to more complex qualitative and quantitative assessments (Ismail, Rajji and Shulman, 2010). The most recent literature in this area can be found in (Spenciere, Alves and Charchat-fichman, 2017)). During the development of the scoring systems, most of the researchers proposed their own methods to assess errors in the drawing, which could be qualitative and/or quantitative. The methods were usually proposed based on the personal experiences of the researchers using the CDTs, which explains why most of the scoring systems employ different lists of assessed errors (Shulman, Shedletsky and Silver, 1986; Mendez, Ala and Underwood, 1992).

While there are many well-regarded systems for manually scoring CDTs, they have their drawbacks, some of which are listed below:

1. Many of the scoring systems can be complex in design and require very detailed assessment rules. While the simplest ones can undermine effective objective assessment, the complex methods are far too labour-intensive and time consuming for routine use (Ricci *et al.*, 2016).

2. The scoring systems often rely on the clinician's subjective assessment (Price *et al.*, 2011). For instance, one current scoring system calls for judging whether 'the clock face must be a circle with only minor distortion acceptable'. Without providing a quantitative definition, this can lead to variability in scoring and analysis.

3. A lack of consensus on which criteria produce the best results (Spenciere, Alves and Charchat-fichman, 2017). Because of the inter-rater reliability in scoring, no system has been universally accepted as the most effective.

4. Although visual analysis can determine both qualitative and quantitative assessment of performance, these are almost always related to the analysis of the end-product of the test, which is the drawn clock without any indication of the underlying execution process. There is

temporal/constructional (also called 'dynamic') information inherently embedded in the test response (Potter *et al.*, 2000). This dynamic information can characterise a behavioural dysfunction and provide a much richer source of valuable diagnostic data on various neurological conditions (Werner *et al.*, 2006; Heinik *et al.*, 2010; Rosenblum *et al.*, 2013).

The above limitations may reduce the robustness of the CDT test and its usefulness as a cognitive impairment-screening tool. In recent research, computerised systems for administering the CDT have been proposed (Heinik *et al.*, 2010; Kim, 2013). Using automatic assessment has the potential to bring significant advantages in dealing with the issues listed above, as will be discussed in the next section.

2.2 Computer-Based CDT

Since the advent of the personal computer, a significant amount of effort has been invested in developing a computerised tool to perform neurological assessment (Wild *et al.*, 2008). Some of these assessment tools adapted standardised tests in a new way in order to benefit from the computer's abilities in terms of administration and analysis of existing tests. Other tools presented an entirely new computer test to assess cognitive function. Examples include the Automated Neuropsychological Assessment Matrix (ANAM) (Rice *et al.*, 2011), the Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment (CANS-MCI) (Tornatore *et al.*, 2005), the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Robbins and

Sahakian, 1988) and the Computerised Neuro-psychological Test Battery (CNTB) (Culter *et al.*, 1993).

Recently, a number of computerised systems have been developed to analyse and diagnose CDT drawings (Heinik *et al.*, 2010; Kim, 2013; Bennasar, 2014; Davis *et al.*, 2014). The research in this area has generally fallen into two groups. The first group focuses on developing a computer system for analysing and assessing the output of the paper-based CDT (the drawn clock), while the second administers the CDT using a digital pen, a digitiser or tablet computer, with these devices capturing the test onto an online database.

In recent research (Bennasar, 2014), supervised classification algorithm SVM and Random Forest (RF) with a set of features extracted from the clock images were used to score the CDT drawings automatically. The system was able to classify CDT drawings into several classes, including healthy and several kinds of dementia with an accuracy of 89.5%, which is comparable to that of medical practitioners. The author also identified new CDT features important for such a classification. However, most of the scoring features are extracted manually from the drawn clocks and only features extracted from the clock images (static features) are employed.

In another approach in computerised CDT, the system is used to administer the test in addition to automatically assessing it. The first studies in this area (Heinik *et al.*, 2010) used a graphical tablet to capture the CDT output. A set of extracted kinematic features were examined for their importance in the diagnosis of mild Major Depressive Disorder (MDD). The focus group included 20 patients and 20 healthy individuals. The study explored 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 was 81.1%, with the most important factors being the writing pressure. Although the importance of dynamic features was highlighted, this study did not include the static features in which the original test was proposed.

Another study (Kim, 2013) employed a tablet computer for recording the drawing process so that clinicians could examine the planning strategy of the patients. The study introduced air time as a new feature, which was defined as the time during which the individual stopped drawing before resuming. The usability of the system was examined, with great acceptability reported from both subjects and practitioners. The work focused primarily on the user interface and did not specify the accuracy of the system in differentiating between normal and abnormal cases. However, the problem of labelling clock elements such as hands and numbers was identified.

A research neuropsychologist group at MIT (Davis *et al.*, 2014) has been administering the CDT by using a digitising pen (DP-201, Anoto, Inc.) which works by recording its position on the page on temporal bases. More than 200 static and dynamic features have been introduced and used with machine learning in order to classify various dementia conditions. The challenge of automatically understanding the drawings (that is, determining the right label for every object) was also noted (Song *et al.*, 2016). More recently, the CDT has been administered by a Wacom critique 13HD pen display (Shi *et al.*, 2016), and new 3D orientation dynamic features were introduced for diagnosis.

Computing technologies have many advantages over traditional paper-based test, as have been shown in many studies (Brown and Brown, 2009; Dougherty *et al.*, 2010; Leposavić, Leposavić and Šaula-Marojević, 2010). First, they can support the administration process through automatic data collection. Second, they can decrease the subjectivity of test examiners' interpretations because of the support of automatic analysis, which can increase test standardisation. Third, they are able to acquire new behavioural data by capturing the execution process (Fairhurst *et al.*, 2008; Liang *et al.*, 2010). Dynamic data could not be captured or measured using paper tests, and now they provide a much richer source of potentially valuable diagnostic data. These systems capture both the result of the drawing and the behaviour that produced it: every pause, hesitation and time spent simply holding the pen versus writing. Such information helps in assessing the cognitive functions of the patients even when the test result appears superficially normal. Information such as how long the participant spent in the test could spotlight a cognitive problem with normal drawings (Guha, Kim and Do, 2010). Moreover, the test can be administered by a GP or even a nurse without the need for specialist medical practitioners.

To conclude, the computerised system could bring a number of advantages in terms of increased diagnostic accuracy and simplification the administration process. However, in order to take advantage of these opportunities, an automatic system for labelling different parts of the CDT sketches is required.

2.3 Sketch Interpretation

Image interpretation is the process of giving meaning to an image by identifying and labelling its objects (Kumar and Desai, 1996). For example, after segmenting an image—that is, partitioning it into multiple segments—each segment is interpreted as being a road, river, building, trees, vehicle, etc. The human visual system is able to describe the image content instantly and without effort. However, for artificial intelligence, image interpretation remains a fundamental challenge (Fleuret *et al.*, 2011). Much of the early successes in machine image interpretation have been made in constrained environments, e.g., remote sensing (Forestier *et al.*, 2012; Huang *et al.*, 2017), text recognition (Aurangzeb *et al.*, 2010) and medical image analysis (Tadeusiewicz and Ogiela, 2004). Image interpretation in unconstrained environments is still largely an open problem (Sonka, Hlavac and Boyle, 1999). Sketches are a special case of images which offer the advantage of being able to adapt computer vision methodology and algorithms (Jorge and Samavati, 2011). For example, one system might represent the sketch as a bitmapped image, then perform image interpretation (Hse and Newton, 2004; Kara and Stahovich, 2005). In addition to these advantages, sketches can be recorded online, with the temporal information being useful in interpretation as reported by other authors (Arandjelović and Sezgin, 2011; Bresler, Průša and Hlaváč, 2016). Nonetheless, the unconstrained nature of sketches produces its own challenges for current image interpretation algorithms (Jorge and Samavati, 2011).

In image interpretation, two main approaches exist: bottom-up and top-down (Behnke, 2003). The bottom-up approach consists of a sequence of steps that transform an image from one representation into another. These steps generally are pre-processing, segmentation, feature extraction and recognition. Image interpretation systems that follow this approach are called classical recognition-based systems (Kopparapu and Desai, 2002). It does not have a hierarchical decision scheme of elimination and acceptance. One popular method belonging to this approach is labelling an image using a statistical classifier based on information at a local level only (Shih, 2010). However, no matter how good the classifier technique is, or how much training data is available, the classifier's performance will be limited if there is ambiguity in an object's appearance. While certain post-processing procedures are used to smooth out the outputs of a local classifier, they are usually problem-specific and based on heuristics (Campbell, Mackeown and Thomas, 1997). In CDT sketches, badly drawn objects are expected since the writing skills of people with cognitive impairment are often affected (Rosenblum *et al.*, 2013). Thus, the classical recognition-based approach will be challenged with the impaired object's appearance.

The top-down approach to image analysis works in the opposite direction. It starts from objects' models, generates a set of hypotheses in which additional information related to the objects is considered—for example, their position. The hypotheses are matched with the features extracted from the image in order to accept or reject them. This method is successful if good models of image objects are available and one ensures that the correct hypothesis is among the first ones (Behnke, 2003).

Sketch interpretation research with a focus on specific domains has been gaining interest in recent years (Olsen *et al.*, 2009; Jorge and Samavati, 2011). Popular application domains include electrical circuits (Guihuan Feng, Christian Viard-Gaudin, 2009), mathematical equations (Laviola and Zeleznik, 2004) and digital logical circuits (Johnston and Alvarado, 2013). Domain knowledge has been used in most of the above systems in the form of constraints to ease the recognition of objects. However, this was limited to the type of objects expected in a sketch without utilising other knowledge, such as the expected positions of the objects or their relationships with other objects. This kind of information is used extensively by people to solve ambiguity when objects have a similar appearance (Hudelot, Atif and Bloch, 2008).

For computerised CDTs, previously developed sketch interpretation approaches were a classical recognition-based system, in which all objects are segmented, and then a classifier employed for labelling them. In early research (Cho, Kim and Do, 2010) on computerised CDT interpretation, only the shape of the drawing strokes was used to recognise the clock numbers, which could lead to low recognition accuracy when the numbers were badly drawn. A similar approach was used by Shi *et al.* (2016). Kim *et al.* (2010) used Microsoft's recognition engine for recognising clock numbers, and suggested further post-processing by considering the context. That is, the recognised characters were converted into appropriate numeric digits. As Kim stated, the context is a clock and so no letters were expected. However, irrelevant handwriting is expected from dementia patients (Freedman *et al.*, 1994). In early research (Song *et al.*, 2016), the context of the CDT sketch objects was also taken into account. However, this was limited to recognising clock numbers only, which was similar to all previous proposed CDT sketch interpretation systems that did not include hands or irrelevant writing.

Previous research in image and sketch interpretation focused on top-down or bottom-up approaches, but no research explored the advantages of integrating the two approaches. Yet human vision employs such an integration during the task of image interpretation. A person first produces a set of possible interpretations depending on the object's appearance. Then the important properties of the objects are compared with a set of expectations that is derived from prior knowledge. Following this process, the nominated hypotheses are further eliminated or accepted. Thus the domain of prior knowledge is employed for final visual inference (Ogiela, 2008). Using such a reasoning process, a person can make interpretations even in difficult cases. This integration has been successfully applied in object detection (Hoiem, Efros and Hebert, 2008; Hedau, Hoiem and Forsyth, 2010; Bao, Sun and Savarese, 2011; Hsiao and Hebert, 2012) and interpreting video-based traffic scenes (Geiger *et al.*, 2014).

Computer vision supports a wide variety of competing paradigms in image interpretation. One approach is supervised machine learning (Kotsiantis, 2007). This approach accounts for many of the successful methods in computer vision such as segmentation and recognition (Wu, 2013). These two processes are essential in the conceptual model of CDT sketch interpretation. More details of supervised machine learning and some of their classifiers are presented in the next section.

2.4 Supervised Machine Learning

Machine Learning (ML) is a set of algorithms that give the computer its ability to perform a specific task by learning from data. The most significant application of ML is predictive data mining, which is about extracting patterns and data classification (Somvanshi and Chavan, 2016). These algorithms learn by using training data, in which every instance in the data set is represented by the same set of features. These features could be binary, continuous or categorical. If labels are given, then the learning is called supervised, in contrast to unsupervised learning, where data set is unlabelled (Jain, Murty and Flynn, 1999).

The process of ML begins with data collection and processing (Kotsiantis, Zaharakis and Pintelas, 2006). During this step, issues such as missing values, discretisation and noise removal are resolved by employing pre-processing algorithms. The next step is feature extraction to identify the most relevant features. Then classification, and the final step is the evaluation, in which there are various approaches to perform, all of which are based on dividing the available data into training and testing sets. The only difference is how the division between these data sets is made. Examples of these validation procedures are cross-validation, ‘leave one out’ and ‘third to one’.

Supervised classification is one of the tasks most frequently carried out by so-called Intelligent Systems (Kotsiantis, Zaharakis and Pintelas, 2006). In the following subsection a relevant classifier of this type will be introduced.

Support Vector Machine (SVM)

SVM are discriminative classifier that use a separating hyper-plane to distinguish between the classes (Cortes and Vapnik, 1995). The selected hyper-plane is the one that maximises the margins between the classes. The rationale behind searching for the hyperplane with a large margin is that the largest margin hyperplane should be the one that is more resistant to noise (Ertekin, 2009).

An 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 projection of initial data to a feature space with a higher dimension using a kernel function. In this new space, the data are considered as linearly separable. A linear kernel has been recorded to provide the best performance in many applications (Ben-Hur and Weston, 2010).

In addition to the wide use of SVM in recognition systems (Burges, 1998), they have been employed in many handwriting segmentation systems in order to make segmentation point decisions. SVM has been applied successfully for segmentation in text handwriting (Sun *et al.*, 2004), online overlaid handwriting (Lv *et al.*, 2013) and online Japanese handwriting (Fukushima and Nakagawa, 2000).

The k-Nearest Neighbour Classifier

The k-nearest neighbour *kNN* classifier is one of the oldest and simplest methods used for classification (Cover and Hart, 1967). Despite its simplicity, the *kNN* gives competitive results and has the advantage of learning from even a small set of examples (Bay, 1998). The label of the observation is predicted based on which class is more common among the k^{th} closest point to that observed in the labelled sample. The best choice of *k* depends upon the data. Generally, larger values of *k* reduce the effect of noise on the classification. Several different distance measurements for classification have been used, but the most commonly used function is Euclidean distance (Fosseng, 2013; Zunjarrao, 2017). One difficulty with *kNN* is that, once an input vector is assigned to a class, there is no indication of the strength of its ‘membership’ in that class. This problem has been addressed by developing fuzzy *kNN*s in which fuzzy set theory is incorporated into the *kNN* rule (Keller, Michael and Givens, 1985).

kNN has been widely employed and a good performance has been reported in online handwriting recognition (Castro-Bleda *et al.*, 2009; Li, Zhang and Su, 2012; Zunjarrao, 2017) and online symbol sketch recognition (Ouyang and Davis, 2009).

Multilayer Perceptron

The Multilayer perceptron (MLP) is a feed-forward neural network consisting of multiple mutually interconnected layers of neurons (Hornik, Stinchcombe and White, 1989). The layers are stacked one onto each other. Every neuron in one layer is connected to every neuron in the following layer. The ability to learn is the key concept of neural networks. The aim of the learning process is to find the optimal parameters of the network for solving the given task. Learning is carried out on the training set by feeding the training data through the network. It is an iterative process, where the outputs produced on each input from the training set are analysed and the network is repeatedly being adjusted to produce better results. The network is considered to be trained after reaching the target performance on the training data. Learning is performed in multiple epochs. After each epoch the error is validated on the validation data set. Once this error starts to increase, the learning process stop at this point.

Deep Neural Networks

Neural networks are systems inspired by parallel distributed processing in the brain (Zurada, 1992). They help to group unlabelled data according to similarities among the example inputs, and they classify data when they are trained on a labelled data set. A Deep Neural Network (DNN) is the name used for ‘stacked neural networks’; that is, networks composed of several layers.

One of the most widely-used types of deep network is the deep Convolutional Neural Network (CNN). This network uses a special architecture that is particularly well-adapted to classifying images (Ciresan, Meier and Schmidhuber, 2012). It consists of a succession of convolutional and max pooling layers, which are general, hierarchical feature extractors that map raw pixel intensities of the input image into a feature vector. Several subsequent fully connected layers classify these features.

Each convolutional layer performs a 2D convolution of its M^{n-1} input maps with a filter of size $K_x^{\,n}$ x $K_y^{\,n}$. The resulting activations of the M^{n} output maps are given by the sum of the M^{n-1} convolutional responses, which are passed through a nonlinear activation function. Thus, in the convolutional step of the CNN, the input image is convolved with filter to obtain the convolutional response map.

Max-pooling layer creates slight position invariance over larger local regions and down-samples the input image. The output of a max-pooling layer is given by the maximum activation over non-overlapping rectangular regions of size K_x x K_y. This represent a reduction in dimensionality of the convolutional responses and confer a small degree of translational invariance into the model.

Kernel sizes of convolutional filters and max-pooling rectangles are chosen such that either the output maps of the last convolutional layer are down-sampled to 1 pixel per map, or a fully connected layer combines the outputs of the last convolutional layer into a 1D feature vector. The last layer is always a fully connected layer with one output unit per class in the recognition task.

With the aid of GPU-based computers, CNN has won many image recognition competitions (Schmidhuber, 2015). In addition, CNN reported impressive results in recognising handwritten digits (Ciresan, Meier and Schmidhuber, 2012).

2.5 Segmentation

Segmentation is the process of extracting individual objects from a given image. It is a crucial step in further image analysis and object recognition. Segmentation used for text-based images aims at retrieving written characters from the entire image (Lu and Shridhar, 1996; Mehul *et al.*, 2014). Unlike printed documents, segmentation of unconstrained handwritten documents has remained a key problem in character recognition (Stafylakis *et al.*, 2008; Papavassiliou *et al.*, 2010; Dave, 2015). Since the main objects in CDT sketches are handwritten digits, the focus in this section is on handwriting segmentation.

Handwriting segmentation can be generally classified as segmentation-recognition and recognition-based (Casey and Lecolinet, 1996). In the former, segmentation is performed separately before recognition, and the result of the segmentation is the isolated characters. In the latter, segmentation and recognition happen simultaneously. This method produces good results, but is computationally expensive (Kokkinos and Maragos, 2009).

Segmentation can be done both offline and online (Plamondon, Srihari and S., 2000). In the offline approach, the completed writing is available as an image, and segmentation algorithms analyse the image's spatial characteristics. In the online case, the two-dimensional coordinates of the writing are stored in the order of strokes made by the writer as a function of time, so both the temporal and spatial information is available for the segmentation.

Recently, several state-of-the-art segmentation methods have been tested on a common database used by (Liu *et al.*, 2003; Ribas *et al.*, 2013; Stamatopoulos *et al.*, 2013). Eleven offline algorithms were evaluated and some of them showed high segmentation accuracy. Most of the algorithms use statistical information on spacing, together with horizontal and vertical projection, in order to segment the written document. These methods can work well with normal images, but would fail when applied to drawings of clocks. In clock drawings, the writing does not proceed from one side to another, but changes direction—especially when the drawings are made by people with cognitive impairment. Figure 2-2 shows examples of clock images that produce segmentation challenges using a traditional segmentation method.

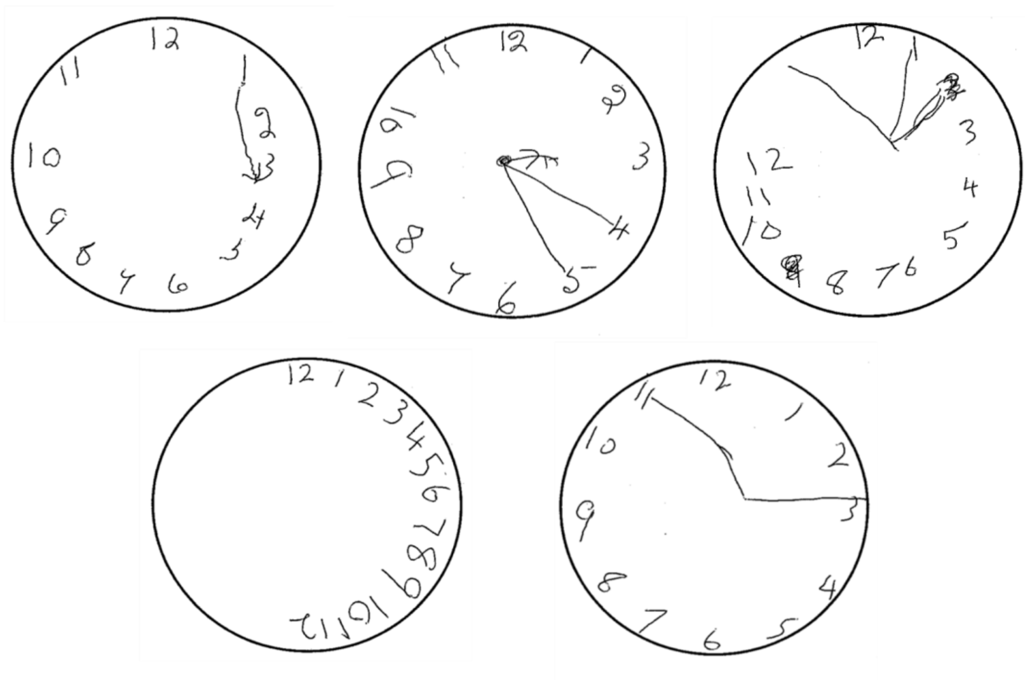

Figure 2-2: Examples of segmentation challenges in CDT sketches.

Most sketch recognition systems override the segmentation step by employing stroke-based recognition. The sketch is recorded online as a set of strokes that are identified as a set of points from the pen on paper until lifting it. These strokes are recognised as lines, curves or other basic subcomponents of diagrams (Alvarado and Davis, 2005). This kind of segmentation was developed to work with shapes rather than handwriting (Alvarado, 2011). Other sketch recognition systems place restrictions on the user's writing style in order to make recognition easier and to eliminate the segmentation problem. For example, users must press a recognition button after each symbol or pause notably between symbols (Cohen *et al.*, 1997; Hse and Newton, 2005). These approaches are inapplicable for CDT sketch segmentation, because the test needs to be unrestricted.

In CDT, sketch segmentation systems (Kim, 2013) used a connected component algorithm to segment the clock elements. The connected component algorithm (Gonzalez, Woods and Eddins, 2004) has been widely used in image segmentation. It segments binary images by detecting connected components between pixels. This method has some limitations in the segmentation of clock drawings. For example, when elements are connected, as in the first clock in Figure 2-2 where the number 3 is connected to the clock hand, it will lead to an under-segmentation error (Cardoso and Corte-Real, 2005) (i.e., when the segmented object contains more than one character). In the case when a number consists of two fragments (as the number 5 in the second clock in Figure 2-2), it will lead to an over-segmentation error (when the segmented object is not a complete character). In (Shi *et al.*, 1997), one proposed solution to the problem of fragmented characters uses vertical blurring of the original images before segmentation. However, this solution is limited to short gaps (no more than two pixels), which is impractical in clock drawings, where subjects may leave a large gap between a character's parts.

In the recent Think project (Davis *et al.*, 2014), the clock drawings were segmented by first separating the clock hands based on their position near the centre of the clock. Then the *k*-means algorithm was applied to cluster the remaining elements into 12 groups using the time when they were drawn and the distance from the centre. This proposed method works well with clock drawings produced by people with no cognitive impairment. However, as reported by the authors, in drawings by impaired people, assistance was required to achieve accurate segmentation results.

The supervised classification algorithms such as SVMs have been used successfully in segmentation. In (Sun *et al.*, 2004), an SVM with a set of geometrical features was used to classify inter- and intra-word gaps in the segmentation of handwritten text. An SVM was also used to successfully segment the online overlaid handwriting (Lv *et al.*, 2013). High segmentation accuracy has been reported in online Japanese handwriting segmentation in (Furukawa *et al.*, 2006; Bilan and Masaki, 2008) using a set of spatial and temporal stroke-based features. However, these systems are specifically designed for handwritten document segmentation and the employed features are specified for this handwriting style. Developing new spatio-temporal-based segmentation for CDT sketches would be valuable.

2.6 **Handwriting Recognition**

Handwriting recognition systems can be divided into two categories: online and offline. In an online system, a sequence of time-stamped coordinates representing the movement of the pen tip is transformed into meaningful text. By contrast, in an offline recognition system, only the scanned image of the text is available (Plamondon, Srihari and S., 2000). Over the last few decades, a large number of algorithms have been proposed in both categories; for a detailed review the reader is referred to the broader survey in (Plamondon, Srihari and S., 2000; Arica and Yarman-vural, 2001; Tanaka, Iwayama and Akiyama, 2004).

Recently, CNNs (LeCun *et al.*, 1998) showed state-of-the-art performance in various domains such as speech recognition and visual object detection and recognition (Schmidhuber, 2015). They consist of multiple alternating convolutional and pooling layers. Their structure of successive alternating layers is designed to learn progressively higher-level features, where the last layer produces the classification results. CNNs were applied to recognise offline-handwritten digits with an error of just 0.23%, as reported on the MNIST database of handwritten digits (Ciresan, Meier and Schmidhuber, 2012). This is comparable to human performance.

With the recent development of smartphones and digital tablets, interest in online handwriting recognition has increased. Several new algorithms have been proposed (Namboodiri and Jain, 2004; Castro-Bleda *et al.*, 2009; Deborah, Elijah and Olaosebikan, 2012; Tagougui, Kherallah and Alimi, 2013; Keysers *et al.*, 2017), with good results reported. Most of these algorithms use a set of features such as the normalised first and second derivatives of the *x,y* coordinates, the distance and angles between pairs of points, the curvature, the start and end point positions and whether the pen is on and off (if there are multiple strokes). These features could be called ‘dynamic’ since they are derived from the process of writing. Online recognition methods have been shown to be more robust than offline algorithms against variations in input shape. A significant recognition accuracy of 99.4%. was reported in (Castro-Bleda *et al.*, 2009) for online handwritten digit recognition. The *kNN* classifier that uses approximate Dynamic Time Warping (DTW) (Berndt and Clifford, 1994) with a histogram of directional features is employed. However DTW is a computationally expensive process. More recently, *kNN* and backpropagation also employed in Lipi toolkit (Zunjarrao, 2017),an online character recognition engine, yet the reported result was only 90%.

So far, research efforts have focused on getting the best recognition accuracy with any one of the approaches listed above (static images, offline and online). Relatively little effort has been spent to explore how various recognition methods can be used as individual sources of information and combined to boost the recognition accuracy. Specifically, how can online recognition methods be combined with offline recognition methods.

Combining multiple classifier systems is an approach proposed in (Liwicki and Bunke, 2007), where a system based on hidden Markov models was used with digit representations. The result was derived using a voting strategy. The results showed a 2% improvement over the best individual recogniser. Other methods (Alimoglu and Alpaydin, 2001) achieved an improvement of 1.2% by combining the results of two multi-layer perceptron (MLP) classifiers, one trained on dynamic *x,y* coordinates and the other on the static images of digits.

In the area of computerised CDT research, several approaches to digit recognition have been employed. The first was by (Cho, Kim and Do, 2010), in which the system recognises the unique shape of each number by comparing the curved and straight lines and the writing sequences of strokes. The system has been tested on 20 clocks drawn by healthy participants and achieved

97.67% recognition accuracy. Although good accuracy was achieved, the developed system was only tested on a small number of healthy people. With bad handwriting resulting from cognitive impairment, the patient does not follow the standardised digit writing form of curves and straight lines.

The second approach (Davis *et al.*, 2014; Souillard-mandar *et al.*, 2016) adapted a symbol recognition algorithm originally developed by Ouyang (Ouyang and Davis, 2009) for electrical circuit diagrams. It uses *k*-nearest neighbour as a classifier with five feature images. Four of the feature images describe the stroke from the horizontal, vertical and diagonal point of view while the fifth represents the stroke endpoint. The reported result was over 96% recognition accuracy after the algorithm was trained and tested on clocks from healthy individuals. More recently, probabilistic Conditional Random Field (CRF)(Lafferty, McCallum and Pereira, 2001) has been used for classifying clock numbers (Song *et al.*, 2016). In addition to the numbers' visual shapes, features related to context information are included, including their position within the clock and the preceding and succeeding numbers' visual shapes. Using this contextual information, the classifier accuracy improved from 88.96% to 99.3%. However, there was no indication of classifier performance in cases when the numbers were not in their expected position nor in sequence with the previous and next numbers. Such cases would affect the classification results. More than fifty percent of the dementia data set involved in the current study are neither in their correct position nor in their expected sequence.

Convolutional neural networks achieved promising results in offline handwriting (LeCun *et al.*, 1998), but their application is limited due to the huge amount of data required for training to avoid overfitting (Srivastava *et al.*, 2014).

Nonetheless, their performance can be improved by using them in combination with other classifiers, which are not highly dependent on the size of the training data set. An example of such a classifier is *kNN*. In addition, the classification features employed with CNNs are static, such as the image of the digit. Dynamic features that describe the sequence of writing would be useful when the digits are visually unclear. This combination will contain information about ‘what’ and ‘how’, which both support the recognition process.

2.7 Knowledge Representation and Ontology

Knowledge representation as a branch of AI relates to the formalisation of knowledge and its processing within machines (Grimm, Hitzler and Abecker, 2007). Knowledge represented in a machine-interpretable form enables computer systems to make decisions based on reasoning about particular domains, similar to humans. Knowledge representation has been extensively used in image interpretation (Monique, 2002).

Ontology is a popular knowledge representation technology in information science. Ontologies have become popular as computational artefacts capable of providing systems with a conceptual model of a particular domain of interest. Ontologies encode a partial view of the world, with respect to a given domain, by defining a set of concepts and their relational structure, which can be used to describe and reason about a domain. Their robustness as valid contexts of knowledge representation have been proven by many researchers (Gu *et al.*, 2004; Grimm, Hitzler and Abecker, 2007).

Ontological modelling of domain knowledge has been applied in many real-world applications, such as medicine (Luciano *et al.*, 2011) and decision making

(Kornyshova and Rebecca, 2010; Bastinos and Krisper, 2013). It has also been widely used in image interpretation (Durand *et al.*, 2007; Hudelot, Atif and Bloch, 2008; Bannour and Hudelot, 2011; Huang *et al.*, 2017).

In addition to the context representation model, appropriate reasoning mechanisms are required to exploit the available context information and add intelligence to the systems. The classical reasoners that have been proposed on top of the ontology context model are ontological reasoning (Description Logic) and rule-based reasoning (such as First Order Logic) (Bikakis and Antoniou, 2010). However, most real-world domains contain uncertain knowledge and vague information. This problem has been discussed in the literature, and several approaches have been proposed to extend these ontologies to deal with these challenges. These approaches include combining ontologies with either fuzzy logic (Bobillo and Straccia, 2016) or with a probabilistic Bayesian network (Nikolopoulos *et al.*, 2011; Aguilar, 2016). Using a fuzzy rule or Bayesian inference would enable the system to represent the procedural knowledge (Van and Rene, 1999), which is purpose-specific knowledge related to the problem of sketch interpretation.

2.8 Reasoning Systems

Reasoning is a process of deriving a new statement (conclusion) from other known statements (premises) of knowledge. Almost any practical artificial intelligence requires dealing with uncertainty. Interpretation of CDT sketches is a good example. Most CDT sketch elements are noisy. This noise is present due to irrelevant writing and badly written digits and clock hands. The popular approaches used for dealing with uncertainty are fuzzy logic and Bayesian networks. Both logic and probability fit well together with human reasoning (Khemlani and Goodwin, 2015).

Although there are many studies describing the theoretical differences between them (Dubois and Prade, 1993), there is little research on the advantages of using one of these approaches over the other (Jeet and Dhir, 2012), and what little exists is usually constrained to a specific context. One of these studies compared these two approaches in the context of medical diagnostics (Onisko, Lucas and Druzdzel, 2001). The authors stated that BN is the best approach when there are missing values in the data. Another study examined the advantage of each approach when the available data were limited (Chen, 2000). The author claims that in such cases a rule-based method can be better. Both fuzzy logic and BN are attractive for CDT sketch interpretation and they have been successfully applied in image interpretation (Dipartimento *et al.*, 1991; Kopparapu and Desai, 2002; Luo, Savakis and Singhal, 2005; Berka, Athanasiadis and Avrithis, 2006; Nikolopoulos *et al.*, 2011). However, the uncertain nature of CDT sketches provides challenges to both approaches.

2.8.1 Rule-based Reasoning

Rule-based reasoning (Buchanan and Shortliffe, 1984) is one of the oldest and most popular reasoning paradigms used in AI. The rules are simple, flexible, formal and expressive, and they facilitate high-level context abstraction and integration with ontology models (Bikakis and Antoniou, 2010). The originality of rule-based approach modelling lies in the interpretability of the models. Rule-based systems, which are also called knowledge-based or expert systems, are the most popular approaches for image interpretation (Dipartimento *et al.*, 1991; Berka, Athanasiadis and Avrithis, 2006; Forestier *et al.*, 2012). The designer of this system has to develop rules for interpretation based on the features of the object under consideration and other constraints. For example, there could be adjacency constraints, as the road can be adjacent to a sidewalk but need not be adjacent to the sky. One of the drawbacks of this approach is its strong dependence on domain knowledge and the rigidity of the rules. CDT sketches can contain many uncertainties caused by badly drawn and incorrectly placed elements. Such problems can be overcome to an extent, by incorporating uncertainty factors. Fuzzy logic provides a systematic mechanism to utilise uncertain and imprecise information (Zadeh, 1992).

The concept of fuzzy logic was first introduced by L. A. Zadeh (1965) as a mathematical tool to deal with uncertainty. In this theory, uncertainty has more to do with vague definitions of criteria than randomness (Dubois and Prade, 1993). To date, fuzzy rule-based reasoning has been deployed in an enormous number of engineering and science areas, e.g. in bioinformatics (Zhou *et al.*, 2012), data mining (Ishibuchi, Nakashima and Nii, 2006), finance (Boyacioglu and Avci, 2010) and robotics (Bai, Zhuang and Roth, 2005). Recently, fuzzy rules have been employed with image interpretation (Hudelot, Atif and Bloch, 2008; Jabari and Zhang, 2013).

2.8.2 **Bayesian Networks**

Until the mid-1980s, applying a probabilistic approach to reasoning under uncertainty was considered impractical. This was mostly due to the problem of computing the joint probability distribution of a large number of random variables involved in the reasoning (Pomerol, 1997). This barrier has been swept away by (Pearl, 1988) and others (Lauritzen and Spiegelhalter, 1988; Neapolitan, 1990). They developed algorithms able to propagate probabilities in the case of a dependency, in which a node is supposed to be independent of its ancestors conditionally to its parents. This field is known as a Bayesian or probabilistic network.

A Bayesian network (BN) is a graphical representation of a joint probability distribution (Neapolitan, 1990). The name comes from the fact that most theories about these networks are based on Bayesian probability. BNs were applied in many diverse domains, including but not limited to medicine (Barbini, Manzi and Barbini, 2013), biology (Needham *et al.*, 2007), natural language processing (Rindflesch and Fiszman, 2003) and forecasting (Zuo and Kita, 2012). Most of these applications were concerned with probabilistic decision making to manage uncertainty. In the area of image interpretation, many researchers noted the advantages of BNs in which prior knowledge about visual scenes is combined with image features to infer the most probable interpretation of the image. For instance (Velikova *et al.*, 2013) applied a BN to the interpretation of mammogram images in an attempt to classify these images into cancer cases. Features extracted directly from the images and others computed from a set of classifiers (based on pixel- or region-based features) were combined. The learnt BN structures supported many of the expert-originated relationships but also revealed some novel relationships between the mammographic features. Another approach by (Luo, Savakis and Singhal, 2005) incorporated semantic features related to context and low-level features (such as colour, texture and shape) into a BN for improving the performance of classifying indoor and outdoor images. Although good results were achieved by these studies, the areas of application were images that have fixed structures. This is in contrast to drawn sketches, which have relatively large variations due to their free-hand nature (Li, Song and Gong, 2013). The same object can be drawn with wide varieties of detail/abstraction depending on the drawer. This problem has been highlighted in research with sketch interpretation using Bayesian networks (Jorge and Samavati, 2011). To provide more flexibility, these authors used dynamically-constricted BNs with shape descriptors. However, this application was restricted to recognising shapes rather than handwritten text as the authors stated. In addition, these shapes had to be representable by a set of components based on a time series. Moreover, this application was tasked with interpreting diagrams that have constrained content, such as circuit diagrams and family trees that were drawn by cognitively healthy people. But since the CDT is a cognitive test, the constraints of the sketch are inherently illogical.

BNs do have several advantages: 1) A Bayesian probabilistic approach is robust enough to manage the uncertainty as stated by the literature review. 2) It offers researchers a flexible and consistent framework for incorporating context information. 3) It is powerful in modelling the qualitative relationships among beliefs.

When comparing BNs against rule-based reasoning approaches, the rule-based system can be highly unclear about the underlying dependence and independence of beliefs. The same reasoning beliefs could be used in both systems, but in BNs they are embedded in a probability distribution function; thus they do not appear as rigid parameters as in the rule-based system. These advantages give the incentive to employing BNs in CDT sketch interpretation as a probabilistic reasoning inference.

However, the main challenge in a Bayesian network is determining the network structure (Margaritis *et al.*, 2003), especially when multiple data belong to the same context but from different sources. Such variation in data sources would affect the dependencies of variables within a BN structure. This problem was identified in the CDT sketch data, since some are drawn by heathy people and others by dementia patients. These data are characterised by a highly unstructured and messy nature, one in which employing features related to context would be a challenge. For example, some objects are not in their proper positions and others are missing. Thus, a more flexible BN structure is needed to overcome this problem. Considering different situations in BNs, and constructing different BN structures based on these situations, could provide more flexibility to BN modelling in CDT sketch interpretation. However, with current BNs, such properties are not available.

2.9 Situation Assessment

Situation assessment is a process of identifying the relations between objects in the environment. The purpose of situation assessment is to increase awareness of the current operational situation (Laxhammar, 2007). It was originally developed to model human decision making during evaluation of complex dynamic systems (Endsley, 1995). In such systems, the environment changes continuously, yet a person can arrive at a suitable decision in a brief time. Situation assessment is also combined with data fusion (Roy, 2001; Dong, Berti-Equille and Srivastava, 2013), that is, when multiple and sometimes conflicting sources of data are present. The developed systems in this area are often related to control systems and sensor fusion (Hanson *et al.*, 2004). In general, rule-based is the most popular approach employed in situation assessment (Rohitha *et al.*, 2007; Kokar, Matheus and Baclawski, 2009);

however, there have been attempts to use Bayesian networks (Wright *et al.*, 2002; Li, Cao and Tian, 2015).

In CDT sketch interpretation, similar processes of situation assessment could be supported, since the drawn clock under interpretation could have different situations. For instance, when the clock is normal and structured, all expected elements are present and in their correct position. And when they are not, the human interpreter is aware of these situations. Considering this knowledge and this situation evaluation process in CDT sketch interpretation, and more specifically in the reasoning process within a BN, would bring advantages to the overall system.

2.10 Summary

This chapter has provided the background of a CDT and sketch interpretation. It has also presented a review of the techniques relevant to the system proposed for solving the CDT sketch interpretation problem. The main findings of the literature review are as follows:

1. While the advantage of a computerised CDT is clear, the problem of automating the scoring of CDT sketches is unresolved. Previous research has shown the importance of correct CDT sketch interpretation in terms of correctly labelling all elements in the sketch, along with collecting information about any missing or repeated numbers, numbers in the wrong position and so on . An automatic CDT sketch interpretation system is required so that the assessment would be effective in diagnosing the abnormality of cognitive functions. However, a very limited number of studies have addressed this issue.

2. While a variety of approaches for image and sketch interpretation have been proposed in the literature, CDT sketch interpretation is still a challenge due to the badly sketched objects related to the cognitive impairment of the patients taking the test. Reasoning by using domain knowledge combined with an automatic object recognition system can improve the performance of the interpretation systems. To achieve interpretation robustness and efficiency, a hybrid system that combines bottom-up and top-down interpretation algorithms is needed. The most promising interpretations would be generated first (bottom-up) by using classical recognition systems, and then actively seek out context knowledge for further verification (top-down).

3. Segmentation of sketches, that is, separating them into semantically meaningful objects, is a crucial step for further object recognition. However, the segmentation of unconstrained handwriting such as found in CDT sketches is still an open research problem that has not been resolved. Supervised machine learning has been successfully applied to online handwriting segmentation; however, further improvements could be made by utilising the temporal information recorded by the computerised CDT system in addition to the spatial information.

4. Handwriting recognition is still an unsolved problem when there is totally unconstrained input. Classifier fusion is considered the optimal approach for improving the classification rates of the individual classifiers. The strengths of one could complement the weaknesses of the other. Online and offline handwriting recognition could be used as

individual sources of information, each preserving different aspects of valuable information, and combined to boost recognition accuracy.

5. Incorporating domain knowledge in sketch interpretation is valuable, and can further improve the interpretation system's performance. Ontologies as artefacts of knowledge representation enable the computer system to reason about possible interpretations of sketched objects. In addition to ontologies, a reasoning system is required.

6. Rule-based reasoning is the most popular approach applied to image interpretation. Moreover, it is the most compatible reasoner with ontology. However, when dealing with uncertainties in CDT sketch interpretation, fuzzy logic has the advantage.

7. The Bayesian network is a probabilistic reasoning approach, which brings the advantages of modelling the rational dependencies between beliefs. However, the inflexibility of modelling the network structure in cases of healthy and dementia data would limit its usefulness within a CDT sketch interpretation system.

8. Situation assessment is an artificial intelligence approach that models human abilities to reason in various situations. Employing this process for CDT sketch interpretation could overcome the obstacle of modelling heathy and dementia sketches in one model within Bayesian networks.

The next chapter introduces the conceptual model of the proposed CDT sketch interpretation system and explains the CDT data capturing approach.

Chapter: 3

Conceptual Model of a CDT Sketch Interpretation System

The main objective of the computerised CDT is to develop an automated assessment of cognitive impairment. The automated assessment would bring great advantages as has been shown in the literature review. It would reduce the labour needed for scoring and could improve the consistency of scoring and reduce its subjectivity. More importantly, an online approach to CDT delivery also offers the potential to extract a much richer set of features that could not be obtained through conventional paper-based testing. The new features would relate specifically to the characteristics of the drawing execution pattern. In previous works (Rosenblum and Livneh-Zirinski, 2008; Heinik *et al.*, 2010; Rosenblum *et al.*, 2013; Davis *et al.*, 2014), such dynamic features have been shown to be useful in diagnostic classification and as additional individual performance metrics.

There are four main issues that must be addressed in order to define the computerised CDT as an effective diagnostic tool. These are: 1) a robust data acquisition strategy; 2) an effective system for interpreting the sketched clock; 3) a strategy for analysing data and specifying diagnostic features; and 4) a strategy for feedback of the clinical diagnostic indicators. In this chapter and within the scope of this research the first two issues are addressed. These represent the fundamental steps for further research in terms of data analysis and feedback generation for an effective automatic cognitive-impairment diagnosis system.

The main contribution of this chapter is the conceptual model for a CDT sketch interpretation system. What is proposed is a hybrid system that integrates contextual knowledge such as sketch structure with an object's visual appearance. This integration representation is encoded in an inference engine and used to facilitate the interpretation of CDT sketches. The details of the CDT capturing system are also presented, along with data-collecting details and participants' information.

This chapter is organised as follows: The data capture process employed in this research is explained in Section 3.1; Section 3.2 presents the collected data; Section 3.3 introduces the conceptual model of the proposed CDT sketch interpretation system; and finally, Section 3.4 summarises the chapter.

3.1 Data Capture

The key to the computerised CDT is the capture of online information about the patient's drawing activity, so that subsequent automated data analysis can yield an appropriate diagnosis. A typical infrastructure for capturing these data is a standard computer-linked graphics tablet. In this scenario, the conventional paper test sheet can be placed over the tablet and positioned in front of the patient on a table. The use of a cordless inking pen allows the patient to perform the test directly onto the paper and receive a familiar visual feedback from the drawing process. There is evidence that, because the patient is using an apparatus identical to that used when taking the conventional test (i.e., a normal-looking pen and a standard sheet of paper), test equivalence is maintained (Potter *et al.*, 2000). At the same time, positional information about the pen is stored sequentially for later analysis. In addition, most graphics-capture systems allow the collection of other pen-dependent data—e.g., data relating to the pressure applied during the writing process and the tilt of the pen as it moves across the surface. Thus, the experimental setup is made to parallel almost exactly the prevailing conditions when the manual test is taken. The interface remains familiar while permitting the seamless capture and analysis of relevant data.

In this research, data is captured using a WACOM Intuos Pro digitising tablet, which has a recording area of approximately 32.5 cm by 20.3 cm. The used pen is a Wacom Intuos wireless inking pen that is compatible with the tablet, and has a pressure-sensitive tip (pressure: 2,048 levels). Its shape and size are similar to regular pens, offering an experience that is no different from the normal paper and pen test.

The participant is given an electronic inking pen and asked to draw a clock on A4 paper affixed to the surface of the digitiser. The clock circle is pre-drawn on the paper (the pre-drawn circle is the general CDT form used in clinical tests), so the participant has to add only the numbers and the hands. The clock hands have to display a specific time, which in this case is five minutes to three. The test instruction and time setting are the same as used in the clinic during the assessment process. Figure 3-1 shows the data captured using the Wacom digitiser tablet.

Figure 3-1: Data capturing using a Wacom digitiser tablet.

As the participant writes on the tablet's surface, the tablet's digitiser reports the sequence of time-stamped coordinate pairs (x,y) (with 200 Hz sampling rate) as well as information about pen pressure and tilt. Figure 3-2 shows an example of the collected online data. This information is transferred directly via a wireless connection to the computer. The computer interfaces with the digitiser using software developed by the author during this research. The developed interface is shown in Figure 3-3; the interface is designed to be friendly. This makes the online CDT test easy to use, and thus it could be conducted by any health care worker without the need for technicians or medical practitioners. The collected data are stored on the computer and analysis is performed offline.

X	Y	Pressure	X_tilt	Y_tilt	Time(msec)
19483	13533	571	1440	570	37220
19481	13503	567	1440	570	37228
19480	13479	541	1440	570	37235
19479	13460	501	1440	570	37243
19479	13448	449	1430	560	37250
19479	13443	381	1430	560	37258
19479	13443	285	1430	560	37265
19479	13446	178	1430	560	37273
19479	13465	50	1430	560	37280
19479	13465	0	1430	560	37288
19472	13516	0	1430	560	37296
19466	13545	0	1430	560	37303
19464	13575	0	1430	560	37311
19456	13601	0	1430	560	37318

Figure 3-2: CDT-captured online data.

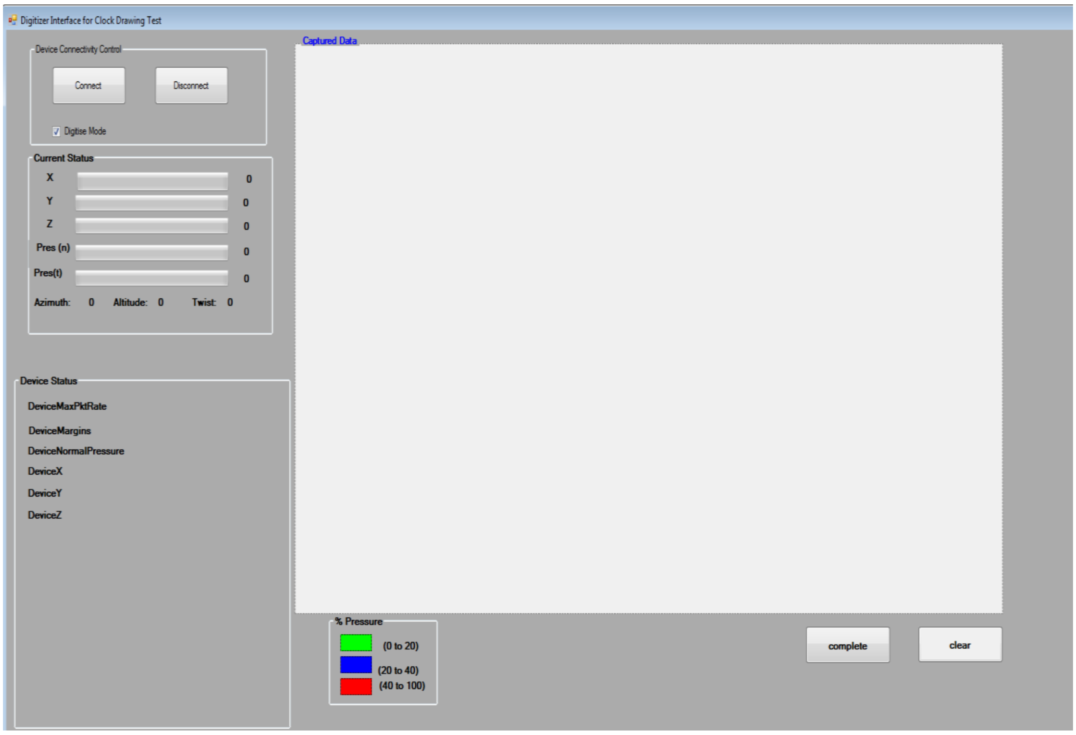

Figure 3-3: CDT digitiser interface.

3.2 Data Sets

In this research, two data sets have been collected: the first consists of 65 drawings made by healthy people and the second is 100 images reproduced from the authentic drawings of dementia patients. The data set was collected under the ethical approval of Cardiff University. All participants were given a participant information sheet and they signed a consent form to participate in this study.

Data Set 1

During data collection, 65 volunteers aged from 25 to 87 years participated. The group included 15 individuals who were older than 60. The participants comprised 32 females and 33 males, and their educational attainments ranged between basic and college graduate. The participants were asked to draw a clock on a paper laid on the surface of the digitiser as shown in Figure 3-1. Each person drew one clock, so 65 clock images were collected. These drawings will be referenced as ‘normal drawings’ in this research.

Data Set 2

The second data set came from dementia patients’ drawings. Because of the difficulty in obtaining data from cognitive impairment patients, and because this study was concerned with the interpretation system rather than patient diagnosis, 100 drawings were reproduced by an authorised person onto the digitiser from the original drawings of dementia patients. These patients were diagnosed with dementia during their examination at Llandough Hospital in Cardiff, UK. More than 100 drawings of different types of dementia were obtained: 37 drawings were reproduced from the drawings of patients diagnosed with mild cognitive impairment, 55 came from Alzheimer's patients, and 8 drawings were from vascular dementia patients. These drawings are referred to in this research as 'abnormal drawings'. Examples of abnormal data are shown in Figure 3-4.

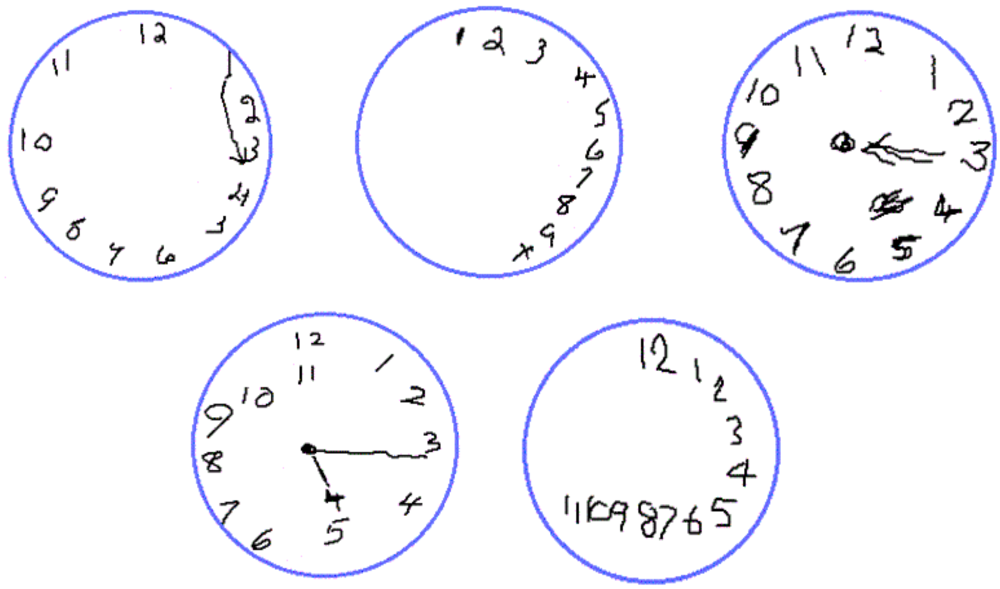

Figure 3-4: Examples of clock drawings produced by people diagnosed with dementia.

3.3 CDT Sketch Interpretation System

3.3.1 CDT Sketch Interpretation Definition

CDT sketch interpretation will be developed to recognise the objects hand-sketched by the user during a cognitive impairment-screening test. The user is requested to draw a clock by placing the clock's numbers in their expected places and setting the clock's hands to a predetermined time. Sketch interpretation is a crucial step in computerised CDT, as it enables the appropriate measurement of clock features, for example the occurrence of all numbers, the hands showing the correct time, positioning numbers in their correct places, etc. Humans are able to interpret CDT sketches almost instantly and reliably even with badly sketched objects. However, computer vision systems work well only under strictly controlled conditions in a limited domain (Geman *et al.*, 2015).

Since the performance of the human eye far exceeds that of current computer vision, it may prove useful to follow the pattern of the human visual inference system when designing a computer vision system. During sketch interpretation, a person generates a set of possible interpretation hypotheses relying on the object's visual appearance and then prior knowledge is employed for final inference. Figure 3-5 shows an example of contextual effects in human reasoning. Here, the context of an ambiguous letter decides whether it is to be interpreted as 'H' or 'A'. The chosen letter is always the one that completes a word. So the first step is to generate a hypothesis for interpretation guided by the letter's visual appearance (either H or A). Then the word context is used to make the final decision.

Figure 3-5: Contextual effect on handwriting recognition.

Following the human reasoning process to resolving uncertainty in difficult cases, the proposed CDT interpretation system generates a set of possible hypotheses relying on the classical object recognition process (i.e. segmentation, feature extraction and classification). The final decision is conducted by the inference engine, in which other information derived from the CDT sketch is incorporated. This is a flexible way to incorporate context information. When the interpretation cannot be decided locally (due to visually impaired cases), the decision is deferred until further evidence arrives from the context.

The proposed system is a hybrid that integrates two approaches in image interpretation: bottom-up, which is feature extraction and recognition, and the top-down strategy, which is hypothesis generation and adding features related to the context.

3.3.2 Conceptual Model

The conceptual model of the CDT sketch interpretation system is shown in Figure 3-6. The system consists of five main components: segmentation, feature extraction, recognition, knowledge representation and the inference engine.

Segmentation

The first step in the interpretation system is segmentation, which involves defining a set of possible objects. The CDT data was recorded when the clocks were drawn on the tablet surface as a set of time-stamped *x,y* coordinates. These points were segmented into a set of objects using a new segmentation algorithm proposed in this research.

A new set of spatial and temporal features automatically extracted from the CDT drawings is proposed. Consequently, a supervised machine learning approach is employed to segment the CDT drawings into their elements, such as number and clock hands, on the basis of extracted features. This new segmentation

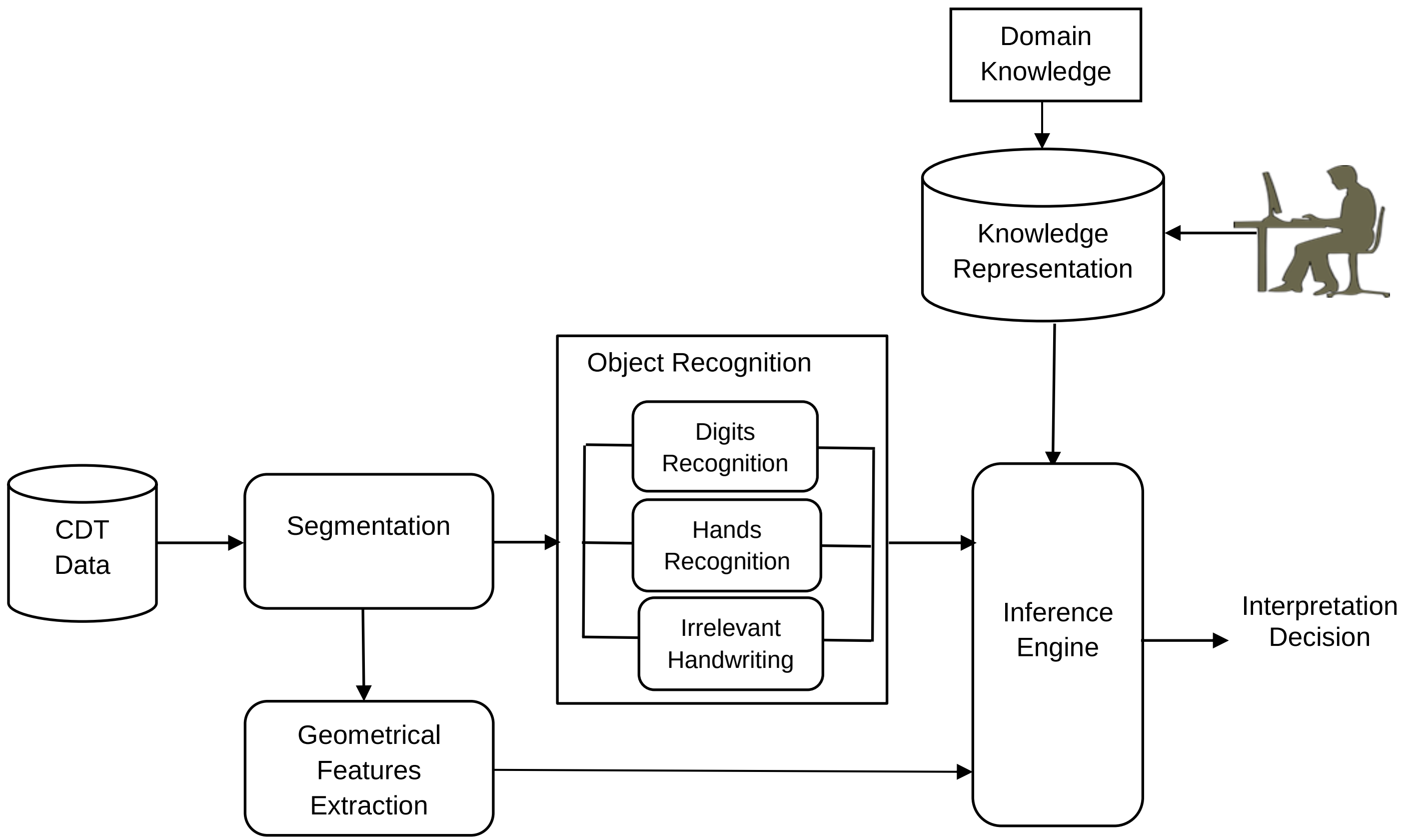

Figure 3-6: Conceptual model of CDT interpretation system.

system is detailed in Chapter 4. The work developed in this area provides a new segmentation system in the domain of unconstrained handwriting segmentation.

Geometrical Features Extraction

After normalisation, a set of geometrical features are extracted for each segmented object. These features are selected in relation to their importance in the interpretation process according to the CDT domain knowledge (features used in human visual inference during CDT sketch interpretation). Some of these features are related to the object itself, and others are related to the object's position within the clock and its relation to the other objects.

Object Recognition

The next step is the recognition process, in which a set of possible classifications for the individual elements according to their visual appearance is generated. After the CDT sketch is segmented into a set of elements, the elements are separated into hands and digits based on their position from the clock's centre. Thus, the recognition process consists of several stages. First, the digit recogniser classifies each potential digit. Second, the arrow recogniser differentiates between hands and non-hands. The non-hands are further classified as irrelevant writing. These two recognisers produce a set of probabilities for each individual element of the clock sketch, which will be used in the inference engine. In this research, a new system for numeral handwriting recognition in the CDT is proposed. The system is based on two complementary sources of data, namely static and dynamic features extracted from handwritten data. The proposed recognition system can take advantage of both static and dynamic data. The details of the proposed recognition system can be found in Chapter 5.

Knowledge Representation

The vision systems require prior knowledge of the images to be interpreted. Knowledge plays a prominent role in image understanding, which can be classified into two types: declarative and procedural knowledge (Van and Rene, 1999). Declarative knowledge is used to describe certain abstract concepts, such as shape and objects, relationships between objects, and components of the objects in an image. Procedural, also termed as functional knowledge, addresses the problem of selecting and applying some purpose-specific knowledge to problems related to image interpretation, and to guide the inferencing process.

In the CDT interpretation system, the declarative knowledge is represented by the clock. The clock is a simple object, but in order for the CDT sketch interpretation to be a general approach, one that could be employed in a similar problem domain (that is, the uncertainty of interpretation), the declarative knowledge of the clock is represented in this research. Moreover, representing this knowledge is essential for better understanding of the problem domain. In this thesis Ontology has been employed for modelling the prior context knowledge, which is the clock; the details related to the clock's ontology are presented in Chapter 6.

The functional knowledge in the CDT interpretation system is related to the knowledge used to guide the interpretation inference process, so it is highly related to the used inference engine. If, for example, a rule-based inference engine were employed, this knowledge would be represented by a set of rules. An example of this procedure is detailed in Chapter 6.

Inference Engine

The inference engine is the main component of the CDT interpretation system, in which the final interpretation is derived. In this research, two inference engines have been employed: rule-based and probabilistic-based. The rule-based is the traditional engine employed in image interpretation, and has the advantage of direct implementation and simplicity. However to deal with uncertainty in CDT sketch interpretation, fuzzy logic is proposed in this research. The details of this approach are described in Chapter 6.

By contrast, the probabilistic Bayesian approach brings the advantage of better modelling of the rational dependencies between reasoning elements. However, the problem arises of modelling the normal and abnormal sketches simultaneously, which reduces the applicability of this approach. Further improvement is proposed in this study by developing a new Bayesian network model called the Situational Bayesian Network (SBN). A new hierarchical probabilistic framework, one that integrates situational assessment, a process well known in decision making with Bayesian networks, is proposed. Details of this model are presented in Chapter 7. The developed work in this area provides a new flexible modelling approach for Bayesian networks.

3.4 Summary

This chapter presented the computerised CDT data capturing system that was used in this research, and described the CDT data that were collected and subsequently employed in this study. The main contribution of this chapter is the proposed conceptual model of the automatic CDT sketch interpretation system. The new hybrid interpretation system that integrates contextual knowledge with human reasoning can enhance the performance of CDT sketch interpretation, and consequently assist in developing an online CDT system. The next chapter discusses the first part in the interpretation system, which is the segmentation, and compares the proposed segmentation system performance to the traditional approach.

Chapter: 4

Segmentation of CDT Sketches Based on Spatial and Temporal Features

A drawing of a clock is a special case of handwriting as the clocks numbers are usually its main components. In such sketches, the writing does not proceed in one direction, but can change its direction arbitrarily. Moreover, the distance between objects and their dimensions are not necessarily consistent for example, in cases of incorrect object positioning related to the cognitive impairment of the person being tested. Thus the standard handwriting segmentation algorithms, which rely on the width and height of segment pattern and the horizontal gap between segments, would not be applicable to CDT sketch.

In this chapter, a new segmentation algorithm is proposed, in which the advantage of using the dynamic CDT data acquired by a graphics tablet is used to improve the segmentation process. A new set of temporal and spatial features are extracted for this purpose from the data. Consequently, a SVM classifier is employed to segment the CDT sketches into their elements, such as numbers and clock hands, on the basis of the extracted features.

The rest of this chapter is organized as follows. Section 4.1 describes the proposed segmentation algorithm. Section 4.2 describes the data set and the experimental setup. Section 4.3 presents the results and discussion, and finally, Section 4.4 summarises this chapter.

4.1 Proposed Segmentation System

The proposed segmentation system (Figure 4-1) consists of three parts: 1) data capture and pre-processing; 2) feature extraction; and 3) classification. In the first part, a digitizer is used to collect the clock drawing data from the participant. The data is then pre-processed to remove irrelevant information. The next step is feature extraction, wherein a set of proposed temporal and spatial features is extracted for each stroke. A stroke is defined as a sequence of points starting from the point where the pen touches the paper to the point where it is lifted from the paper. Whenever a 'pen-off' (i.e. the pen lifts between strokes) is detected, the last stroke is treated as a new candidate segment. Lastly, using these extracted features, each stroke is judged by a two-class classifier (SVM with a linear kernel) to decide whether it is a new segment or not. In the following subsections, the components of the segmentation system are explained in detail.

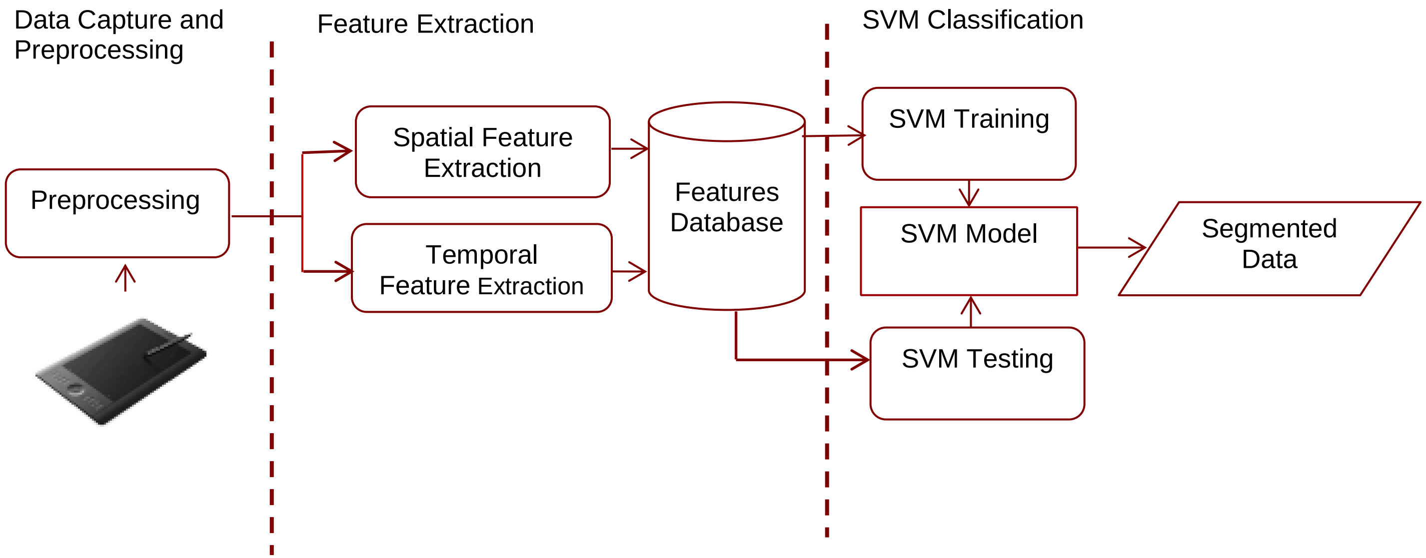

Figure 4-1: Proposed segmentation system

4.1.1 **Data Preprocessing**

As the participant writes on the tablet's surface, the tablet's digitiser reports the sequence of time-stamped coordinate pairs (x, y) as well as information about pen pressure and tilt. This information is transferred directly via wireless connection to the computer. The collected data is stored on the computer and analysis is performed offline. All the collected data details in previous chapter are considered for further segmentation.

The pen-down and pen-up signals are used to determine where a stroke starts and ends, with the overall handwriting data represented as a sequence of strokes. These strokes are further pre-processed to remove noise. This noise occurs when the participants touch the digitiser surface using the pen without writing. In such cases a stroke, characterised by a small size (that contains a small number of points) is recorded. From data observation, these strokes are no more than 10 points. These small strokes are removed from the data, so it could not affect on other strokes segmentation.

4.1.2 **Feature Extraction**

As the clock is being drawn, digitised data is captured as a temporal sequence of strokes. Let us represent this set of strokes as S. Then $S = \{S_1, S_2, \cdot\cdot\cdot, S_n\}$, where n represents the number of strokes in the data set, and $S_i = \{p_1, p_2, \cdot\cdot\cdot, p_m\}$, where m represents the number of points in the stroke. For each stroke S_i in the data set, the system extracts 15 spatial and temporal features (Table 4-1).

Table 4-1: List of stroke features.

	Features
1-4	Stroke width, height, length and size
5	Centroid distance between the current stroke and the succeeding stroke
6	Minimum distance between the current stroke and the succeeding stroke
7	Distance between the end point of the current stroke and the starting point of the next stroke (which represents the off stroke distance)
8-9	Vertical and horizontal alignment between the current stroke and the succeeding stroke
10	Distance between the centroid of the stroke and the centre of the clock circle
11	Angle between the centroid of the stroke and the centre of the clock circle within the x-axis
12	Connection with the succeeding stroke
13	Overlap with the succeeding stroke
14	Completion time of the stroke
15	Difference in time between finishing the current stroke and starting the succeeding stroke

The stroke width and height represent the width and height of the bounding box surrounding the stroke (as shown in Figure 4-2(a)). Stroke length (S_{ilen}) represents the total path length of the pen trajectory from the start point of the stroke until the last one, which can be calculated using the following equation:

$$S_{i_{len}} = \sum_{j=1}^{m} dist(p_j, p_{j+1}), \qquad \textbf{(4-1)}$$

where m is the number of points in the stroke, and $dist$ represent the Euclidean distance between each point and its successor. Such distances in a correctly

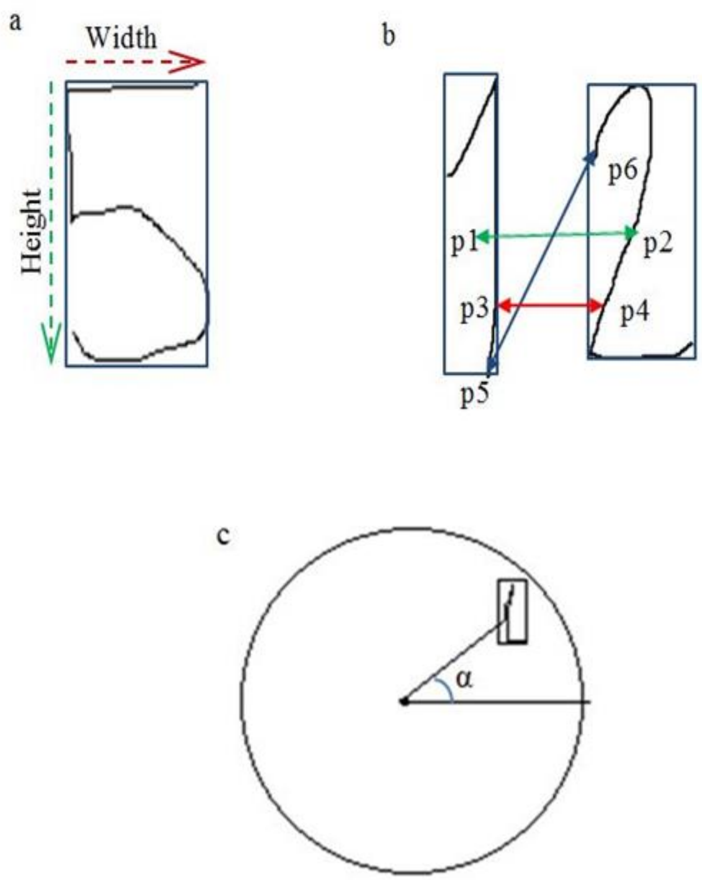

Figure 4-2. (a) Width and height of stroke; (b) Distances between strokes: (p1-p2) centroid, (p3-p4) minimum, and (p5-p6) end to start distance; (c) Angle between the stroke and the centre.

segmented character have relatively small values. Using this property, three complementing measurements of the Euclidean distance between strokes are defined: centroid distance, minimum distance, and the end-to-start point distance (i.e. the distance between the end point of the previous stroke and the start point of the current stroke). The centroid distance can be measured as the distance between the centre of the bounding box of the current stroke and the centre of the bounding box of the succeeding stroke (i.e. the stroke that follows the current stroke in time). The minimum distance can be calculated by measuring the shortest distance between the current stroke and the next one in the sequence. End-to-start distance represents the distance of the off stroke, i.e. the pen's lift distance between strokes. Figure4-2 (b) illustrates these distance measurements.

Clock drawings are considered to be similar to unconstrained writing, with the writer being able to write in any direction. Thus the horizontal alignment feature, which has been used successfully in standard handwriting segmentation, is not sufficient. The proposed segmentation system includes a vertical alignment feature. Horizontal alignment between successive strokes is measured as the difference in the x- axis coordinate of the right point of the bounding box of the first stroke and the left point of the bounding box of the second stroke, while the vertical alignment is the difference in the y-axis between them. Other important features in clock drawing segmentation are the distance between a stroke and the centre of the clock, and the angle between the stroke and the centre of the clock (Figure 4-2(c)). The strokes, which have close values of these features and are consecutive in time are likely to belong to the same character.

All spatial features are normalised with respect to average character size to overcome the problem of different character sizes among different writers, which may affect the classifier operation. However, it is advisable to retain the character size information for further cognitive assessment, since the letter size may be indicative of cognitive impairment. The average character size is estimated by measuring the length of the longer side of the bounding box for each stroke, sorting the lengths of the strokes in descending order (excluding the two clock hands), and finding the average of the first half of the sorted strokes.

The temporal information provided by the graphics tablet in the proposed segmentation system is also taken into account and appropriate features are extracted. For example, the ordering of the strokes in the sequence is considered important in character segmentation as well as in handwriting recognition (Furukawa *et al.*, 2006). The time taken to finish the stroke and the time between strokes are also important for the segmentation process (Table 4-1).

After features calculation for each stroke, SVM model are trained by setting positive and negative examples. SVM classify each stroke into segmentation or non-segmentation point.

4.1.3 SVM Classification

SVM (Cortes and Vapnik, 1995) has been widely used in recent years as an alternative to other popular classification methods such as neural networks, with good results shown in various applications. SVM has also been used successfully in handwriting segmentation (Furukawa *et al.*, 2006; Bilan and Masaki, 2008; Lv *et al.*, 2013), thus making it a good choice of a classifier in this research. The SVM is trained using a training pattern of strokes features with the target value of segmentation points set to 1 and that of non-segmentation point to -1.

Given a training set of instance of labelled pairs $(x_i, y_i), i = 1 \ldots., N,$ where $x_i \in X = R^d$ stand for the features vector with dimension *d* of training pattern i, and $y_i \in Y = \{1, -1\}$ is an associated class label of traning pattern i, N is number of training pattern.

The key idea of SVM is to learn the parameters of the hyperplane that has maximum margin to classify two classes on training set. To find the hyperplane $wx_i + b = 0$, SVM require the solution to the following optimization problem:

$$
\begin{aligned}
min: \quad & \frac{1}{2} w^T w + \mathrm{C} \sum_{i=1}^{N} \xi_i \\
s.t.: \; & y_i \left(w^T x_i + b\right) \geq 1 - \xi_i, \;\; \xi_i \geq 0
\end{aligned}
\qquad (4\text{-}2)
$$

where w is the weight vector, b the bias, and $\mathrm{C} > 0$ is the penalty parameter of the error term. ξ_i is the learning error of a training pattern i. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. Then, the feature vectors are mapped into an alternative space choosing kernel function. Examples of these kernel are linear, polynomial, radial bias function and sigmoid. In this thesis, linear kernel function is employed, since it produced the best result.

For evaluating the performance of the proposed system the data has been divided into training and testing sets using 5-fold cross validation. At the decided segmentation points, the adjacent primitive segments merged to form candidate object patterns.

4.2 Data Set and Experiment Setup

The proposed system has been tested on the two data sets detailed in chapter 3, which is healthy and dementia in which it denoted as normal and abnormal subsequently. Feature extraction and classification was implemented using the Matlab language. The dataset was divided into training and testing sets using the 5-fold cross-validation method, four parts used for training and one for testing. All the features listed in Table 4-1 were used, with the exception of completion time and time between writing. These features were excluded because they showed no effect on the result: time of writing depends on the writer's speed, which varies between individuals. Moreover, some user such as in dementia cases stop for long time before resuming writing. The drawings were converted into two sets of strokes, which were classified as segmentation or non-segmentation point as described in previous section .

4.3 Results and Discussion

The performance of the proposed segmentation system was evaluated in both normal and abnormal drawings. In order to compare the proposed methods with other methods, the connected component algorithm (Gonzalez, Woods and Eddins, 2004) was employed on the two data sets. This method has been used before in (Kim, 2013) for the clock drawing segmentation problem. The results show that there is a significant improvement when using the proposed strategy, in comparison with the connected component algorithm.

Table 4-2 shows the average of ten segmentation test results. The proposed system achieved 99.5% segmentation accuracy for the normal drawing data set and 96.1% for the abnormal ones. More than 4% improvement in segmentation accuracy was reported in both normal and abnormal cases.

An experiment for training and testing the classifier with a different data set size was conducted to examine the effect of the data set size on classification accuracy. 1200 strokes were obtained from 65 normal drawings and 1800 strokes from 100 abnormal ones. The higher SVM classifier accuracy can be achieved by increasing numbers of strokes, as shown in Figure 4-3.

Table 4-2: Segmentation accuracy of normal and abnormal drawings.

	Connected Component Algorithm	Proposed Segmentation System
Normal	95.6%	99.5%
Abnormal	92.27%	96.1%

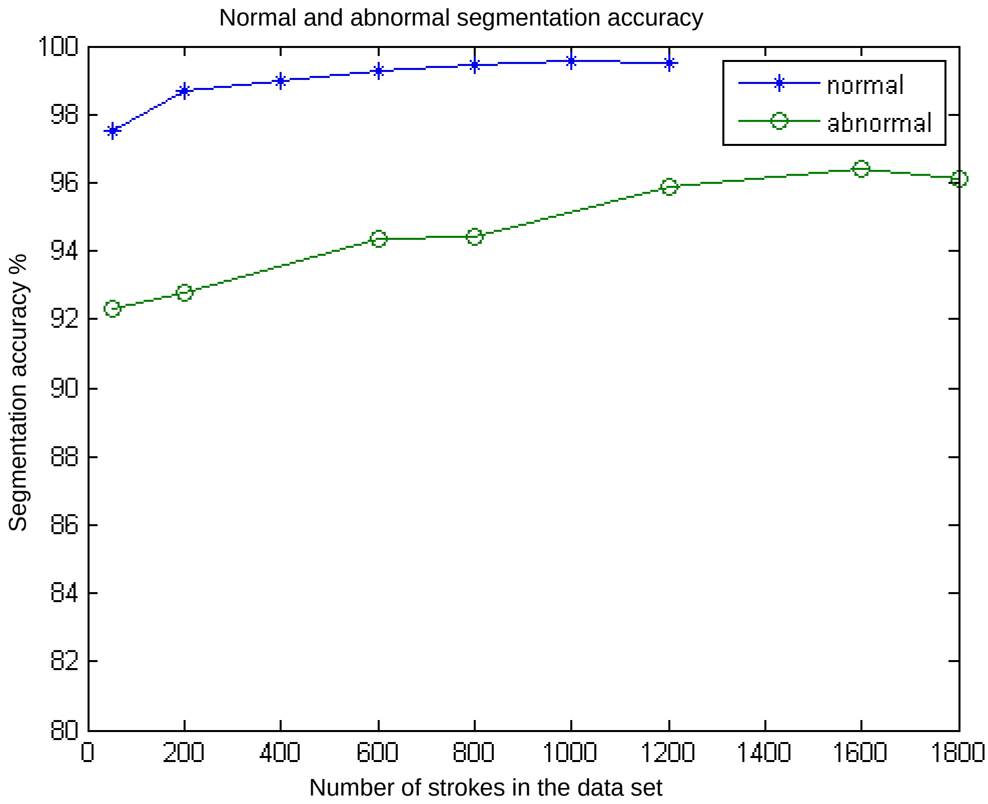

Figure 4-3: Segmentation accuracy versus size of the data set for both normal and abnormal drawings.

Two examples of segmented data produced by the connected component algorithm and the proposed method are illustrated in Figure 4-4. In these drawings each square represents a segmented object. The connected component algorithm failed to detect all segment properly, as can seen in Figure 4-4 (b). For instance, the second part of the number 11 is connected to the hand arrow. Another example of incorrect segmentation is where the hand and arrows are not connected. The same problem has been detected in the second example of Figure 4- 4 (b), where the number 5 is broken into two parts. These problems can be solved by the proposed system as shown in Figure 4-4 (c). These drawings are segmented correctly, because other features such as the sequence of writing, the size of the stroke, and the distance between strokes have been considered by the classifier. Analysis of segmentation error produced by the proposed system shows that most segmentation errors are related to clock hand detection, something a future work should consider in order to achieve a higher accuracy.

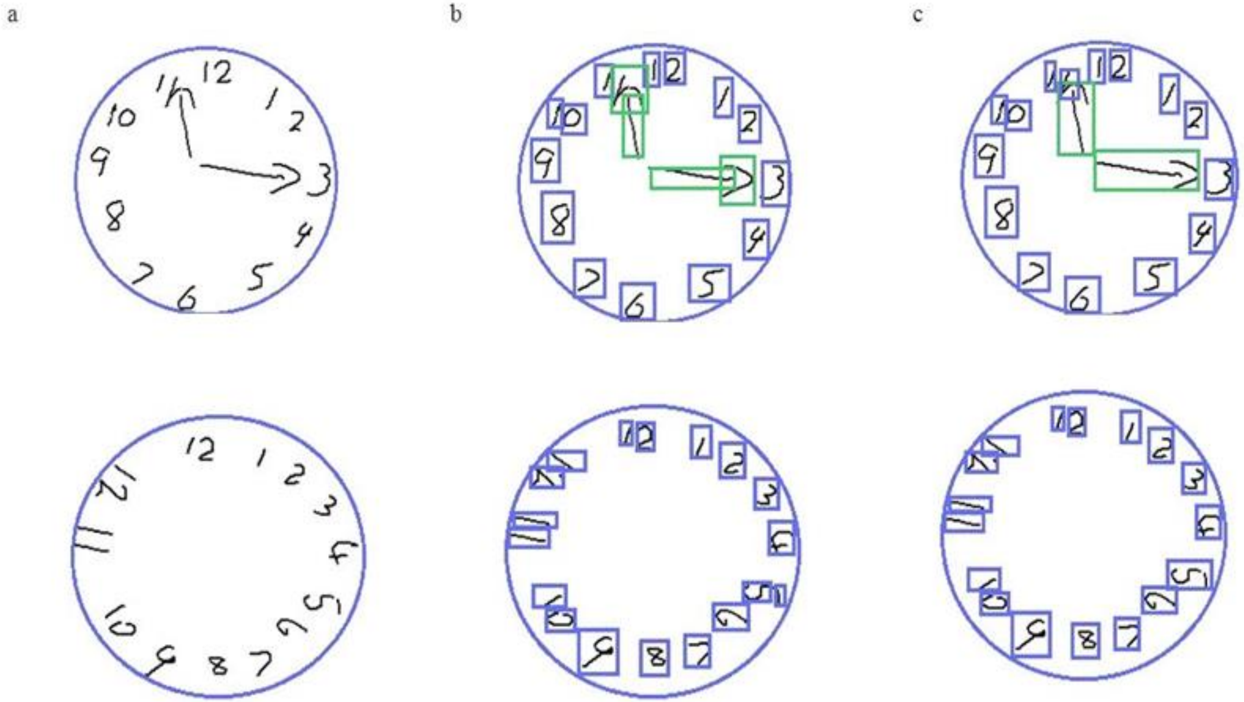

Figure 4-4. (a) Original image; (b) Segmentation using a connected component algorithm;

(c) Segmentation using the proposed algorithm.

4.4 Summary

A new approach for object segmentation in CDT drawings was proposed in this chapter. Conventional handwriting segmentation methods cannot be used in the case of clock drawings where the writing does not follow a standard format of direction and spacing. Current algorithms that rely on the width and height of segmented patterns and horizontal gaps between segments are not applicable to CDT drawings, especially those made by dementia patients. By using SVM as a classifier with a combination of temporal and spatial features, the proposed method has achieved a high segmentation accuracy of 99.5% for normal drawings and 96.1% for abnormal drawings. The system shows promising results, even for abnormal drawings. The segmentation process is the first step towards handwriting recognition in which a new recognition system is proposed and presented in the next chapter.

Chapter: 5

CDT Object Recognition Using Static and Dynamic Features

After the CDT sketch is segmented into a set of elements, the segmented elements are separated into hands and digits based on their position from the clock's centre. A digit recogniser processes the off-centre objects, that is, the objects close to the boundary of the clock circle, and the objects in the middle of the clock face are fed to an arrow recogniser. Object recognition is a critical step in the CDT sketch interpretation system, since it is the main source of possible objects interpretation. Therefore, a reliable handwriting recognition system needs to be applied. However, recognising the handwriting of people with cognitive impairment is challenging, since their writing skills are often affected by cognitive impairment (Rosenblum *et al.*, 2013). In the past, many algorithms have been developed for handwriting recognition (Tappert, Suen and Wakahara, 1990). However, there is no actual evaluation of such algorithms using data sets collected from the elderly or persons with cognitive impairment.

In this chapter, a new system for numeral handwriting recognition is proposed. The system is based on two complementary sources of data, namely static and dynamic features extracted from handwritten data. The main novelty of this chapter is a new handwriting digit recognition system, in which two classifiers are combined: fuzzy *k*NN for dynamic stroke-based features and CNN for static image based features, which can take advantage of both classifiers with static and dynamic data representation.

This chapter is structured as follows: Section 5.1 describes the proposed digit recognition system, along with other related classifiers; Section 5.2 presents the clock hands recognition system; Experiment setting and data set are presented in Section 5.3; Comparative results and analysis are shown in Sections 5.4; and finally, Section 5.5 summarises the chapter.

5.1 The Proposed Handwriting Digit Recognition System

This section presents the proposed digit handwriting recognition system (Figure 5-1) and its components in detail. The proposed system consists of a preprocessing and normalisation component, static and dynamic feature classification components and the unit for combining output of these components. In the following subsections, the components of the system are explained in detail.

5.1.1 Preprocessing and Normalisation

The data were captured when the clock is drawn as a set of (x,y) coordinates. These points were segmented into a set of objects using the segmentation algorithm previously proposed in previous chapter. The segmented object was further separated into hands and digits based on their position from the clock's centre. The coordinate sequence received from the tablet was normalised to eliminate differences due to sampling, scale, and translation. This improve the robustness of the recognition system, ensuring that all the digits are centred and scaled appropriately.

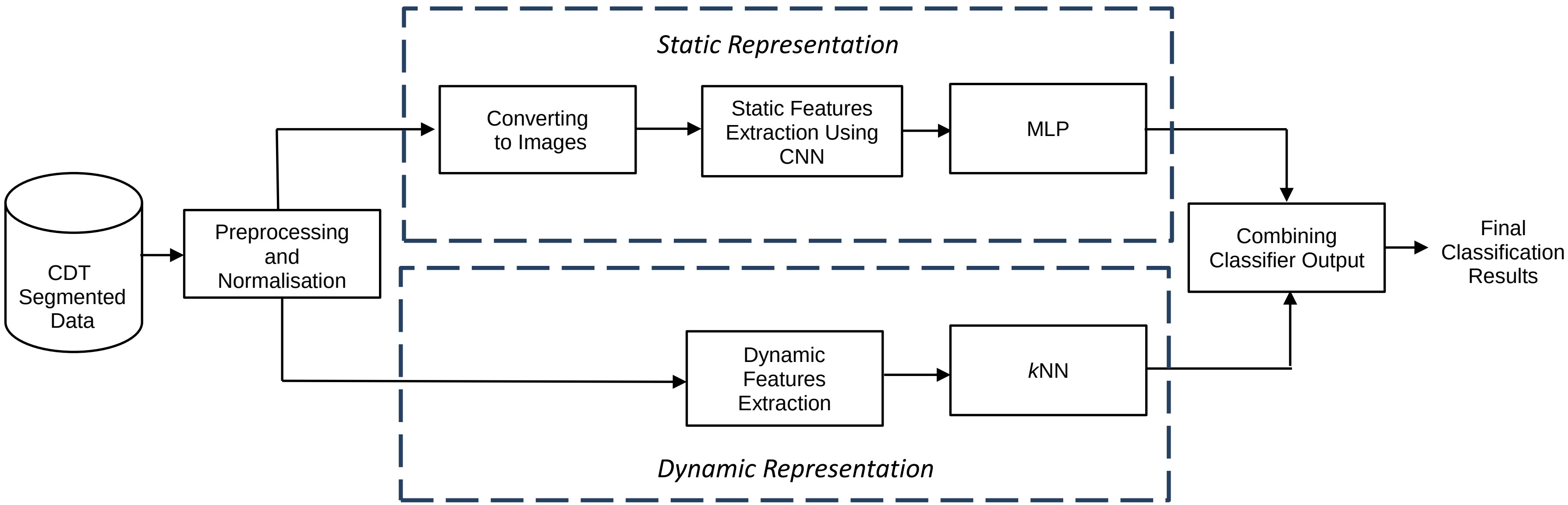

Figure 5-1: Proposed handwriting digit recognition system.

The input digit size depends on the user writing, and its coordinates depends on where the item placed on the tablet surface. All digits are transformed so that they have the same bounding box dimensions while preserving their aspect ratio. Depending on the aspect ratio, the normalised image is centred in the plane with one dimension filled. Assume the standard plane is square and the side length is denoted by l . Denote the width and height of the input pattern i as w_i and h_i, and the width and height of the corresponding normalised one as w'_i and h'_i. The normalised pattern filled one dimension by $\max(w'_i, h'_i) = l$, that is to keep the aspect ratio unchanged. The linear mapping is shown in equation (5-1), where $\propto$ and β are parameter computed by equation (5-2).

$$x_i' = \propto x_i\ , \qquad y_i' = \beta\ y_i \tag{5-1}$$

$$\alpha = \frac{w'_i}{w_i}, \qquad \beta = \frac{h'_i}{h_i} \tag{5-2}$$

where x, y are the origin points, and x', y' are the normalised point. The normalised dimension used in this study is 100x100 following the same approach in Pendigits data set.

Different writers write with different speeds and the online stroke are typically sampled at a constant temporal frequency, thus the distance between neighbouring points in the pen trajectory varies based on the speed of the pen. In addition, more samples could be found in the corner or regions of high curvature, where the pen is typically slower. In order to eliminate these variations Bresenham's line algorithm (Bresenham, 1965) is used. It is widely employed to create bitmaps from normalised on-line handwritten data (Alimoglu and Alpaydin, 2001). Given two endpoints, Bresenham's algorithm determines subsequent points from the first to the second point. This algorithm is iterated through each pairs of sequential point samples within each pen stroke in the data set.

5.1.2 Static Feature Classification

The static feature classification starts with converting the preprocessed and normalised data into grey scale images. In order to obtain the images from sequences of x-y coordinates, these coordinates are further down sampled and mapped to a 20x20 pixels box to store each image. Furthermore, the images are smoothed using a Gaussian-smoothing function to increase tolerance to local shift and distortion. The Gaussian filter used is 3x3 pixel with 0.75 uniform standard deviation. Finally, the images were centred in a 28x28 image by computing the centre of mass of the pixels, and translating the image to position this point at the centre of the 28x28 field.

After the digits data are converted into images, the next step is feature extraction and classification. CNNs are well known for their invariance to distortion and simple geometrical transformations such as translation, scaling and rotation (Ciresan, Meier and Schmidhuber, 2012). This feature makes a CNN an appropriate candidate for tackling the problem of CDT handwritten digit. However CNN success stories notwithstanding, there are some limitations to their application, such as the large quantity of training data, which is required in order to avoid overfitting the CNN to the training data set. Since the available CDT data set is relatively small, one can consider using CNN as features extractor by employing the advantage that a CNN network converges from high-resolution information to reduced but highly informative space recognition. First CNN was trained on a data set that was close to the CDT data set, which is Pendigits online digits data set and this network more augmented by adding a data from our data set. Then the pre-trained network used as a feature extractor to the CDT data sets. Investigating the physical meaning of these features is outside of the scope of this work. These features are fed to MLP classifier. The MLP with backpropagation training is the standard algorithm for any supervised learning pattern recognition process, and it can fit well in the CNN architecture. Moreover the MLP's output is considered as posteriori class probabilities, with useful properties (e.g. positivity, summing to one), providing an efficient framework for a classifier combination (Pinto and Hermansky, 2010).

The CNN architecture used in this research is LeNet (Ciresan, Meier and Schmidhuber, 2012), which is a deep convolutional neural network known to work well on MNIST handwritten digit classification tasks. LeNet architecture used in this work is the default one using MatConvNet toolbox (Chatfield *et al.*, 2014) which consist of eight layers. The first one is a convolutional layer with a filter bank of 20 single-channel filters of 5 × 5 size. The second one is a max pooling layer. Third is another convolutional layer with 50 different filters of 5 × 5 followed by another max pooling layer. The fifth layer contains a filter bank of 500 filters of 4 × 4 pixels. The sixth layer contains a rectifier linear unit followed by another convolutional layer of 10 filters with one single pixel and at the end; the eighth layer applies the softmaxloss operation. The CNN network is trained first and after training, it is used as feature extractor by replacing the last two layers with a MLP. The MLP used in this work is a simple two-layer perceptron with a logistic sigmoid activation function.

5.1.3 Dynamic Feature Classification

The first step in dynamic feature classification is dynamic features extraction. The used dynamic features are the normalised x-y coordinates. After a preprocessing and normalisation step, the data set consists of variable numbers of sequence points. In order to have constant- length feature vectors, the data was spatially resampled into sequence of points regularly spaced in arc length. Following the same approach of Pendigts data set, we used 8 points per digit.

Using *k*NN in online handwritten digit recognition demonstrated impressive results even when considering only simple directional features and a small data set (Castro-Bleda *et al.*, 2009). In this work, a fuzzy *k*NN (Keller, Michael and Givens, 1985) algorithm is used where the algorithm assigns label probabilities to a sample rather than assigning the sample to a particular class. The inclusion of fuzzy set theory into these classifiers deals with imprecision when defining the classes, which is caused by the large variability of the samples belonging to the same class. Consequently, it improves the results. The following relationship assigns class labels to the sample as a function of the sample's distance from its *k*NN training samples:

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} \left(\frac{1}{\|x - x_j\|}\right)^{\frac{2}{m-1}}}{\sum_{j=1}^{k} \left(\frac{1}{\|x - x_j\|}\right)^{\frac{2}{m-1}}} \tag{5-3}$$

where $u_i(x)$ is the membership probability of the test sample x to class i, and m is the 'fuzzifier' which is a fuzzy strength parameter that determines how the distance is weighted when calculating each neighbour's contribution to the membership value. The variable k is the number of nearest neighbours; u_{ij} is the membership value of the j-th neighbour to the i-th class, which can be defined by giving them complete membership in their own class and no membership in all other classes. This is because the prototypes should naturally be assigned complete membership in the class that they represent. As seen from equation(5-3), the assigned memberships of x are influenced by the inverse of the distances from the nearest neighbours and their class memberships, this inverse distance serves to give more weigh to vector's membership if it is closer to, and less if it is further from, the vector under consideration. To calculate the distance there are many distance algorithms. Dynamic time warping is widely used with time series data, but since each digit represents by eight points only with a fair distance between them, Euclidean distance is most appropriate and less computational complexity and time consuming.

5.1.4 Combining Classifiers Outputs

One effective approach to improve the performance of handwriting recognition is to combine multiple classifiers (Xu *et al.*, 1992). Following this approach, a combination of two classifiers is used in order to obtain better digit recognition accuracy. CNN and *k*NN are built as individual classifiers for recognising offline and online patterns respectively. Data processed by each classifier are very different: dynamic representation (online) contains spatial and temporal information (stroke coordinates and order), while static (offline) representation consists of the image of a digit.

The advantage of the CNN classifier is that it automatically extracts the salient features of the input image. The features are largely invariant to the shift and shape distortions of the input characters. This invariance occurs because CNN adopts a weight-sharing technique on one feature map. CNNs are efficient at learning invariant features from the offline patterns, but do not always produce optimal classification results, particularly when there are small data sets or unbalanced training data. Conversely, *k*NNs, with their distance measuring, cannot learn complicated invariance. However, they do produce good decisions when considering the sequencing of points in an online pattern, which can be achieved with a small number of patterns.

Overwriting is one of the problems that can cause misclassifications to the dynamic classifier, as the stroke point sequencing will be changed dramatically while the final shape of writing will be the same. In this case, the advantage will be for the static classifier. In other cases, the shape of the written digits may be overly distorted but the same sequencing information is preserved, which will give credit to the dynamic classifier.

In the proposed combination system, the CNN is trained with normalised images and used as features extractor, while the *k*NN classifier is trained with the normalised (x,y) coordinates that represent the dynamic data. The two classification results are then processed by a combination scheme, and this scheme generates a ranked list of predictions for the input image. Classifier combination techniques operate on the outputs of individual classifiers, while a function or a rule combines the classifier scores in a predetermined manner. The formula is defined as follows:

$$P\,(c_i|S) = f\{P(c_i\,|C, P(c_i|K)\} \qquad i = 1..m \tag{5-4}$$

where $P(c_i\,|C)$ is a posterior probability for one class (i), computed from the CNN model; $P(c_i|K)$represents a probability for the same class (i) given by the *k*NN model; $P(c_i|S)$ is the combination probability for the class (i); and f represents the function applied to the classifier probability results. The average is used in this chapter experiment as it generated better results than other combination methods such as maximum, product and weighted sum. Finally, a ranked list of candidates is obtained with a decreasing order of probabilities after the combination process. The top candidate is then chosen as the predicted class for the input pattern.

To assess the accuracy of the combination CNN and *k*NN model, as well as of separate classifiers, they are applied to the Pendigits data set and the clock drawing data set.

5.2 Clock Hands Recognition

The CDT objects recognition problem is separated into two parts, digits and hands recognition since both have different aspects. Digits are handwriting while clock's hands mostly represented by arrows are related to shape. In CDT sketch interpretation system, objects in the middle of the clock face are forwarded to the arrow recogniser. Where they classify as hands or non-hands. The non- hands are further classified as irrelevant writing, that is, any object in the centre of the clock face, which is not a digit or an arrow.

All the previous research study in CDT sketch interpretation system (Kim, 2013; Davis *et al.*, 2014; Shi *et al.*, 2016) ignored hand recognition step as revealed in the literature review. All the objects in the middle of the clock were considered as clock hands. However, irrelevant handwriting are almost expected in dementia drawings. Moreover, full sketch interpretation means recognising all other elements (Freedman *et al.*, 1994). Thus, clock hands recognition is considered in this study.

Clock hands are normally represented by arrows. An arrow could be defined as a symbol consisting of two parts: shaft and head. The head determines the orientation of the arrow. However, arrow's appearance within the clocks drawings can be changing arbitrarily according to users drawing styles. They can have various shapes, length, heads, and direction. Examples of these arrows in CDT dataset are shown Figure (5-2). Therefore, it is a challenging task to detect arrows with ordinary classifiers based on symbol appearance. Moreover, most users repeat the ink layers several times while drawing the hands, for instance the first arrow in dementia data clocks examples in Figure (5-2(b)). This form of overwriting and variation make the dynamic features relate to the (x,y) coordinates do not provide discriminative features for recognition system, specifically with abnormal data set. Therefore, only static classifier that is proposed for digit recognition in Section 5.1 is used for clock hands recognition.

Recently CNN has been achieved high performance in sketched object recognition (Sarvadevabhatla and Babu, 2015; Yu *et al.*, 2015). Moreover it is featured as invariant to distortion and simple geometrical transformation (Ciresan, Meier and Schmidhuber, 2012), give the motivation to employ CNN in the task of CDT sketch hands recognition. Similar steps that are applied in preprocessing and digits static features

Figure 5-2: Examples of clock hands from CDT data set (a) healthy, (b) dementia

classification detailed in previous section are employed with clock hands recognition. Clock hands are normalised, preprocessed, and converted into images. Then the extracted

features by CNN (LeNet architecture) are fed to MLP classifier, but this time for binary classification, hand or non-hands. Details of the data sets, experiments and results are described in the next sections.

5.3 Data Sets and Experiments Set up

Two different sets of online isolated handwritten digits were used in the set of implemented experiments: Pendigits and CDT digits. Pendigits (Alimoglu and Alpaydin, 1997) is an online data set available from the UCI Machine Learning Repository (Lichman, 2013). This resource contains handwritten instances of 10 digits from several writers. 7,494 glyphs from 30 writers are used as a training set and 3,498 glyphs from 14 different writers are used as test data.

Each digit is represented by eight successive pen points in a two-dimensional coordinates system. The second data set is the CDT digits.

Two representations of handwritten digit samples are used in the proposed system: the static representation, where each digit is represented by a fixed size 28 × 28 pixels grayscale image with each pixel value ∈ {0,... , 255}., and the dynamic representation where each digit is represented by a sequence of (x_t,y_t) coordinates in time. Pendigits is an already normalised data set in dynamic representation; the only pre-processing applied on them is converting them into grayscale images for static representation.

All the clock hands in the CDT data sets are employed in the hand recognition experiment part, in which 130 arrows relate to healthy CDT data set, and 162 relate to dementia data set are detected. Number of irrelevant handwriting objects in dementia data set are relatively small, in which only 22 object detected. Since the classification purpose is to detect arrow or non-arrow, 270 digits randomly selected from CDT digits data set are combined with the irrelevant handwriting to have equalised numbers of patterns for classification.

Data pre-processing and classification models were all implemented using MATLAB. The Pendigits data set was divided into training and testing using the original settings. Following the same approach, the CDT data set was divided into fixed 70% training and 30% testing sets for fair comparison between the proposed classification systems.

5.4 Results and Discussion

5.4.1 Digits Recognition

This section presents the performance evaluation and comparison of the proposed digit classification system. To evaluate which method would be more accurate—the individual classifiers or the combination system—a number of experiments were conducted on a public Pendigits data set as well as CDT digits, 2438 digits are extracted from both normal and abnormal drawings. All CDT data introduced in chapter 3 are used in the experimental part of this chapter.

5.4.1.1 CNN and Static Representation

In order to train a CNN, a large amount of data is required. To overcome this problem the CNN used as a feature extractor as explained in Section 5-1. LeNet was first trained on Pendigits, which is an online handwritten digits data set. Next, the pretrained network was used as a feature extractor for digits that were extracted from the CDT data set. These features were fed to the MLP classifier with a simple structure consisting of one hidden layer with 200 neuron and one output layer. An implementation from the MatConvNet MATLAB library (Chatfield *et al.*, 2014) was used. The model was trained using a stochastic gradient descent with a batch size of 100 samples and 0.001 learning rate. The network was trained for 100 epochs on an NVIDIA GeForce 970 GX 4 GB GPU. In addition, data augmentation was applied to the data sets. In particular, the data set were increased by a factor of ten by rotating the image through ten different angles [-25,-20,-15,-10,-5, 5, 10, 15, 20, 25]. Rotation is the most suitable augmentation, which can be applied to the clock drawing digits, since most people try to write with some rotation according to the circle of the clocks.

Table 5-1 shows the recognition accuracy of CNN as a classifier and using it as a feature extractor by replacing the last fully-connected layer with MLP. The first row of the table represents the result of training and testing the network from scratch on the data set. This experiment is applicable only to the Pendigits data; the CDT data set is not large enough for such experiments. The result for MLP shows the case when the MLP was trained directly on the row images rather than using the features map extracted by CNN.

Table 5-1: Recognition accuracy of Pendigits, normal and abnormal digits with MLP and CNN.

	Pendigits	Normal digits	Abnormal digits
CNN	97.2%	-	-
MLP	94%	89%	84%
CNN + MLP	98%	97.3%	93%

It is clear from the recognition accuracy that using CNN as a feature extractor outperforms the MLP when trained on the row images. CNN is a considerable feature learner, even in the case of a small data set such as CDT data set digits. Moreover, the difference in accuracy between normal and abnormal digits indicates that there is a remarkable effect in recognition algorithm performance when considering digits drawn by healthy people and others with cognitive impairment.

5.4.1.2 *k*NN and Dynamic Representation

In this set of experiments, the accuracy of *k*NN using the dynamic representation of data (i.e. a set of temporal x,y-coordinates) were compared with other machine learning classifiers: LIB SVM, Naïve Bayes and RBF Network. A detailed description of these learning algorithms can be found in

(Mitchell, 1979). Weka data mining software (Hall *et al.*, 2009)was used for the implementation of the classifier with the same setting parameters. The training and testing data size is the same as used for all previous experiments. There is no data augmentation here. As shown in Table 5-2, the experiments indicate that *k*NN outperforms other classifiers for different k values.

Table 5-2: Recognition accuracy of Pendigits, normal and abnormal digits with dynamic representation and *k*NN, MLP, Lib SVM, Naïve Bayes and RBF Network.

	Pendigits	Normal digits	Abnormal digits
*k*NN (*k*=1)	97.68%	95.2%	92.5%
*k*NN (*k*=3)	97.8%	96.7%	92.6%
*k*NN (*k*=5)	97.7%	95.5%	92.7%
MLP	94.5%	90.1%	89.1%
Lib SVM	96.9%	93.8%	91.4.%
Naïve Bayes	89%	87.5%	85.2%
RBF Network	95.6%	92.3%	88.3%

The performance of SVM is very is close to the RBF network; however the SVM is affected by the small data set. As in the case of the CDT data set with normal digits, the size of the data is smaller than in other cases.

All the classifier results show differences in recognition performance between normal and abnormal digits. *k*NN with three nearest neighbours has achieved a slightly better result than other values, so this setting will be used in the combination classifier.

5.4.1.3 Classifier Combination

In this section, two classifiers model are combined: CNN as features extractor with MLP and *k*NN. Each classifier was trained individually on the same training data; however, the training data were in two different representations. Dynamic representation used the set of sequential (x,y) points for the *k*NN, while static representation took the form of images in the case of CNN. In testing, the patterns were presented to both classifiers simultaneously.

The output probability of both classifiers was combined using the average, which was given the best result in this chapter. Finally, a ranked list of candidates was obtained with a decreasing order of probabilities. The top candidate was then chosen as the predicted class for the input pattern. By combining two pieces of knowledge, the accuracy was increased considerably: 99% for the Pendigits data, 98% for normal digits and 94.5% for abnormal cases. A significant improvement can be reported, especially with abnormal digits: about 2%. This was the most challenging task (see Table 5-3).

Table 5-3: Recognition accuracy of Pendigits, normal and abnormal digits for the combination system

	Pendigits	Normal digits	Abnormal digits
CNN-MLP	98%	97.3%	93%
KNN	97.8%	96.7%	92.6%
CNN-MLP + KNN	99%	98%	94.5%

The advantage comes from the diversity of the classifiers' strengths on different input patterns. Moreover most of the abnormal digits are overwritten, that confused the *k*NN classifier, while the CNN has a higher certainty of it is classification result, and this can improve the final classification results.

In another hands some ab-normal digits are badly distorted but still the sequencing information preserved, In such cases the *k*NN has some confidence of the classification results and that improves the final classification results subsequently. However, the small size of CDT normal digit data set has an impact on recognition accuracy, in comparison between normal digits and the Pendigits data set. In addition, the context is different: writing in a round clock is different from writing in a linear document. The examples in Figure 5-3 show that many of missed digits could be considered difficult even for the human.

Figure 5-3: Examples of incorrectly classified digits by the proposed handwriting digit recognition system. First line shows examples from pendigits data set, second line shows examples from normal digits while third one are examples from abnormal digits. The label is corresponding truth->predicted.

5.4.2 Clock hands Recognition

The results of clock hands object recognition are presented in this section; Table 5-4 shows the recognition accuracy of recognising clock hands in CDT dataset for both normal and abnormal data sets. It also shows the result of recognising the irrelevant handwriting objects, which is related to abnormal drawings, since no such kind of object were detected in normal drawings. The result for MLP shows the case when the MLP was trained directly on the row images. Following the same experiment in digits static classification, MLP in the second part are trained and tested on the features map extracted by CNN.

Table 5-4: Recognition accuracy of normal and abnormal clock hands and irrelevant handwriting

	Normal clock hands	Abnormal clock hands	Irrelevant handwritings
MLP	95.5%	85%	68%
CNN-MLP	97.7%	90.1%	77.2%

The obtained results show impressive improving when using CNN as a feature extractor for both normal and abnormal clock hands recognition. However, the improving in normal drawings was only 2% compared to more than 5% in abnormal data, due to the fact that most normal data are systematic. That is normal clock hands have discriminated appearance compared to other clock's objects. Most misrecognised cases in normal data relate to the fact of similarity between number '1' and the clock hands represented by arrows. Since all arrow's images are processed to be similar size and are rotated, the differences between this number and an arrow become unclear. In these cases, incorporating further information regarding object position and size could further improve the recognition accuracy and can resolve that conflict.

The reported results in this section support other researchers finding of effectiveness and relative compactness of CNN deep features (Chatfield *et al.*, 2014; Sarvadevabhatla and Babu, 2015). In addition, this research considers the first study that use CNN in recognising a sketched object that drawn by cognitive impairment patients.

5.5 Summary

This chapter investigated the unconstrained objects recognition problem in CDT sketch interpretation system. Clock digits and hands are the main part in the

CDT sketches, hence two recogniser are developed. Different data representations and classification techniques were developed. In addition, a new combination of classifiers is proposed by combining *k*NN and CNN in recognising clocks digits. The combination system has the advantages of both classifiers and static and dynamic data representations. In the evaluation section, the proposed system is shown to be more effective than using each representation alone. The proposed system's ability to recognise digit in both normal and abnormal drawings is evaluated on two data sets. The first was a publically available online Pendigits data set and the second consisted of digits extracted from the CDT data set (both healthy and dementia data). Due to the lack of discriminative abilities in dynamic features of the clock hands in CDT sketches, only static CNN classifier was employed in recognising the clock hands. Although a considerable results reported in this chapter, including context information such as the object position in the sketch and its relation with other objects would further improve the recognition accuracy of the CDT sketch objects.

Object recognition is the source of possible interpretation hypotheses for CDT sketch interpretation, and it is employed in the CDT sketch interpretation system in the next chapters.

Chapter: 6

Rule Based Approach for CDT Sketch Interpretation System

Most classical automatic object recognition systems use information related to the visual or geometrical appearance of the object for classification. Whilst this step is essential in sketch interpretation, incorporating additional features related to the context of the object and its relation to the other objects can effectively enhance the performance of the recognition system and reduce ambiguities in the interpretation. This chapter describes a newly proposed rule based CDT sketch interpretation system, which represents the prior knowledge of the CDT structure by using ontology and integrating human reasoning through a fuzzy inference engine. This integration combines multiple sources of information concerning the sketch structure and the visual appearance of the sketched objects and deals with the interpretation uncertainty inherent to CDT sketches.

The main contribution of this chapter is combining ontology and fuzzy logic to allow the fusion of the object recogniser probabilities based on the visual appearance of the objects and the context embedded in the domain knowledge ontology.

The rest of this chapter is organised as follows: Section 6.1 describes the proposed system; Section 6.2 introduces the proposed system evaluation and discussion; and finally, this chapter summery is presented in Section 6.3.

6.1 The Proposed System

The proposed interpretation system is shown in Figure 6-1. Ontology-based knowledge representation is enhanced by fuzzy reasoning to benefit from CDT sketch domain knowledge and to deal with uncertainty in the interpretation process. The system generates a set of possible interpretation hypotheses relying on the classical object recognition process as explained in chapter 3.

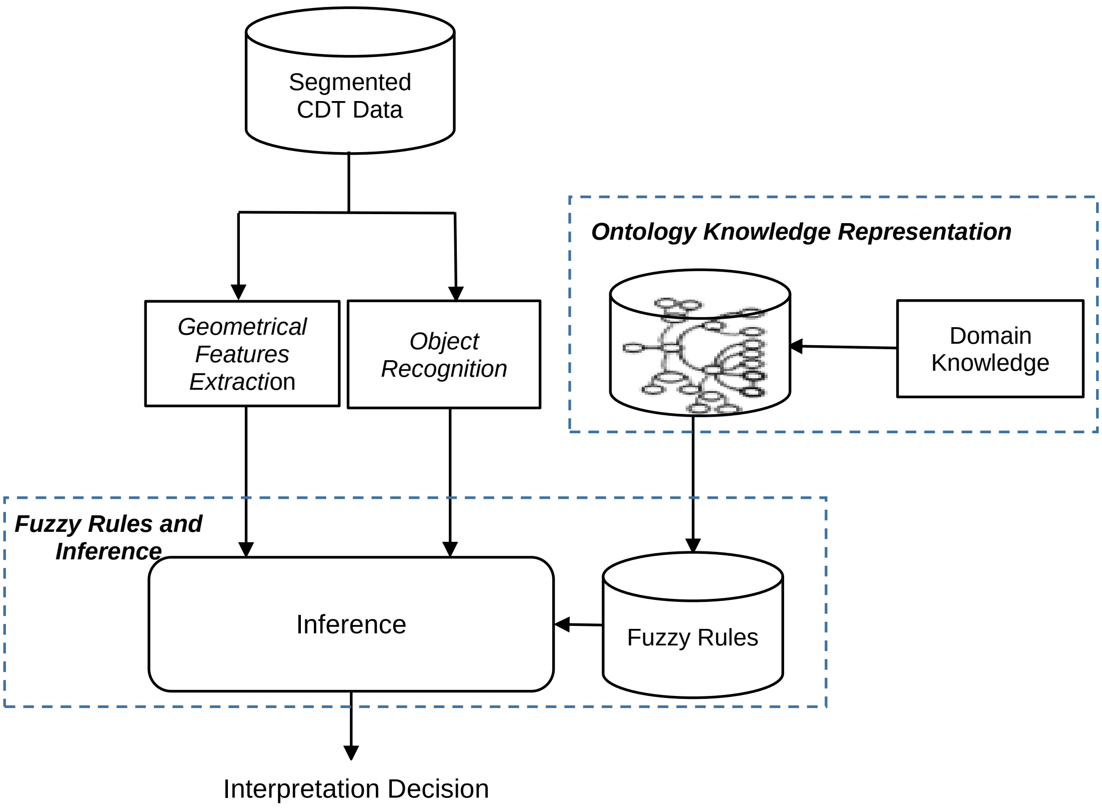

Figure 6-1: Rule-based CDT interpretation system

The final decision incorporates other information derived from the CDT sketch domain that is CDT ontology representation and the fuzzy inference engine. In the following subsections, the components of the system are explained in detail.

6.1.1 Geometrical Features Extraction

CDT data was segmented into a set of objects using a segmentation algorithm proposed in chapter 4. A set of geometrical features were extracted for each segmented object. These features were selected in relation to their importance in the interpretation process according to the domain knowledge represented in the CDT sketch ontology. Two group of features could be identified: local and global. The local features are related to the object itself, such as the object's height and width, while global features are related to the object position within the clock and its relative position to other objects. These features will be used in the reasoning process, as explained in the following subsections.

6.1.2 Object Recognition

The first stage in the human understanding of a given scene is defining a set of possible classes for each object based on the object's shape. After the CDT sketch is segmented into a set of elements, the segmented elements are separated into hands and digits based on their position from the clock's centre. After this step, each object is forwarded to specific recognisers. These recognisers are digit and hand recogniser detailed in previous chapter. A digit recogniser processes the off-centre objects, that is, the objects close to the boundary of the clock circle and its task is classifying each potential digit. Hand recogniser deals with the objects in the middle of the clock face. It differentiates between hands and non-hands. The non-hands are further classified as irrelevant writing.

A CNN is employed in the task of recognising digits and clock hands based on their visual appearance. CNNs have been proven as a very successful framework for object recognition and are well known as a powerful visual features extractor. Further information for the detailed algorithm and implementation can be found in the previous chapter. These recognisers produce a set of probabilities for each individual elements of the clock sketch, which will be used in the inference engine as a predicted level (*PreLev*).

6.1.3 Ontology Knowledge Representation

This section discusses how knowledge about the clock structure is represented and utilised in the reasoning process. Since no existing ontology describing clock drawings was available, a new ontology was developed to capture the domain concepts, such as clock numbers and clock hands. Any other objects found in the clock were considered irrelevant. Figure 6.2 shows the domain-specific ontology model for the CDT sketch. It represents all the elements contained in the clock sketch and involved in the CDT sketch interpretation process. The ontology defines the basic concepts of a clock. The clock contains a set of objects, with the numbers and hands being instances of the objects, which share similar attributes, such as size and location. Numbers from ‘1’ to ‘12’ are instances of numbers, and the relation between them is the sequence represented in ‘after’ and ‘before’ relations.

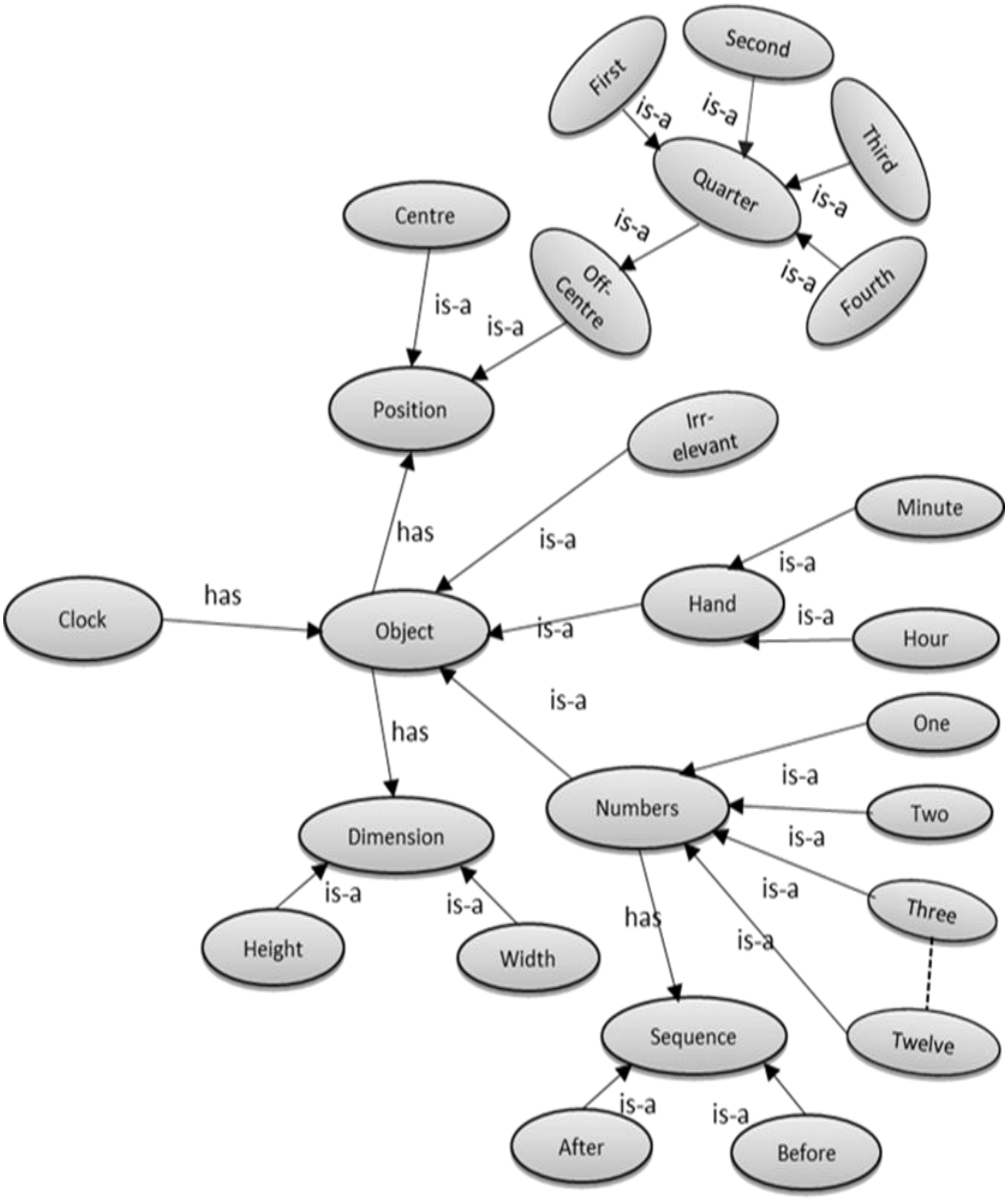

Figure 6-2: Clock Ontology.

All the objects in the clock sketch have specific constraints associated with them; for instance, each number has a specific location within the clock face, and the clock hands should be located in the centre. The clock circle is divided into four quarters rather than twelve parts to give more flexibility to the numbers position. Moreover, the twelve numbers of the clock face should be placed in increasing numerical order starting from number 1and ending with number 12, which is represented in the 'after' and 'before' relations in the ontology.

This information is used by people when interpreting the drawings of the clock. Even with the worst clock sketches with severely distorted visual appearance, a human can arrive at reasonable sketch interpretation. To transfer this knowledge to the machine and to enable the reasoning process, a set of rules is associated with each element defined in the clock ontology, as explained in detail in the next section.

6.1.4 Fuzzy Rules and the Inference Engine

In this chapter, knowledge is introduced in the form of fuzzy logic if–then rules for the concepts and properties defined in the ontology. Since the domain problem is a sketch of a clock drawing, relying on human knowledge for generating these rules is considered an adequate approach. Meanwhile, in a more complex context for which defining a set of rules would be a challenging task, an appropriate learning method could be utilised to derive the rules from a training data set (Roubos, Setnes and Abonyi, 2003).

The variables used in the fuzzy rules in the proposed systems are listed and explained below. The clock digit and hand recognisers' probabilistic output predicted level $Prelev$ was used as a measure of certainty that the object belonged to a particular class and was calculated for all classes. $PreLev$ was further fuzzified to three levels, high H, medium M and low L according to the predefined thresholds from the data observation before its inclusion in the set of rules. Another element included in the fuzzy rules was the position (Pos) of the element, which can take on either a True or False value depending on whether the digit is in the quarter of the clock face that it should be in. The information about whether a digit appeared in the correct sequence with the previous and following digit was included in two variables of the fuzzy rules: *SeqAft* and *SeqBef*. In total, 24 inference rules were generated using the above variables and their possible values. The set of rules for clock numbers is presented in Table 6-1.

Table 6-1: Fuzzy rules to be used by the interpretation system inference engine for clock numbers (Ac for accepted and Rej for rejected).

Rule #	Pre Lev	Pos	Seq Aft	Seq Bef	Dec	Rule #	Pre Lev	Pos	Seq Aft	Seq Bef	Dec
1	L	F	F	F	Rej	13	M	T	F	F	Rej
2	L	F	F	T	Rej	14	M	T	F	T	Ac
3	L	F	T	F	Rej	15	M	T	T	F	Ac
4	L	F	T	T	Ac	16	M	T	T	T	Ac
5	L	T	F	F	Rej	17	H	F	F	F	Rej
6	L	T	F	T	Ac	18	H	F	F	T	Ac
7	L	T	T	F	Ac	19	H	F	T	F	Ac
8	L	T	T	T	Ac	20	H	F	T	T	Ac
9	M	F	F	F	Rej	21	H	T	F	F	Ac
10	M	F	F	T	Rej	22	H	T	F	T	Ac
11	M	F	T	F	Rej	23	H	T	T	F	Ac
12	M	F	T	T	Ac	24	H	T	T	T	Ac

For the clock hands, an alternative set of rules was generated to reflect some changes in the fuzzy rules variables, as shown in Table 6-2. The element used in this set of rules, in addition to the *PreLev*, was the *Size*, which was fuzzified to take one of two values – large *Lr* or small *S*.

Table 6-2: Fuzzy rules to be used by the interpretation system inference engine for clock hands (Ac for accepted and Rej for rejected**).**

Rule #	Pre Lev	Size	Dec	Rule #	Pre Lev	Size	Dec
1	L	S	Rej	4	M	Lr	Ac
2	L	Lr	Rej	5	H	S	Ac
3	M	S	Rej	6	H	Lr	Ac

The implementation of fuzzy inference approach in various applications commonly involves two inference models Mamdani (Mamdani, 1974) and Takagi-Sugeno inference model (Sugeno, 1985). Both models employ slightly different approaches in the output aggregation process. Since the intended outputs are constant in a binary form (i.e. accept or reject the hypothesis), that is constant, the Sugeno inference is very suited to this study. Sugeno inference has the form of:

If input 1 = x and input 2 = y, then output is z (linear or constant).

$$z = ax + by + c \tag{6.1}$$

For a zero-order Sugeno model, the output level z is a constant ($a=b=0$).

To conclude, in the proposed CDT interpretation system, the visual appearance was used to produce possible hypotheses for interpreting the clock sketch elements, as described earlier in this section. These hypotheses were further processed in the inference engine to be accepted or rejected according to the predefined fuzzy rules. The entire hypotheses were processed sequentially, and the decision was made in favour of the first accepted hypothesis.

6.2 **Evaluation and Discussion**

This section presents the performance evaluation of the proposed system. All the experiments were implemented using Matlab language. The evaluation analysis was conducted using two data sets, normal and abnormal, All the CDT sketch data set detailed in chapter 3 are employed in this experiments. To evaluate the performance of the proposed system, two experiments were conducted. The baseline experiment evaluated the performance of the sketch interpretation based solely on visual appearance recognition, when the recognisers' hypotheses with highest probabilities were accepted. The second experiment used the proposed interpretation system. The interpretation performance was measured as the ratio of the number of objects identified correctly in each sketch to the total number of objects present in the sketch, averaged over all sketches.

The proposed system significantly outperformed the baseline system, as shown in Table 6-3, with 98.5% of all objects correctly identified in the first set of sketches and 94.8% of objects correctly identified in the second set of sketches. Up to 2.3% improvement was recorded for abnormal sketches, showing the effectiveness of the proposed system in using both prior knowledge of the sketch structure and human reasoning in the interpretation process.

Table 6-3: Recognition accuracy of normal and abnormal sketch objects using the baseline and the proposed systems.

	Recognition Based System	Proposed System
Normal	97.2%	98.5%
Abnormal	92.5%	94.8%

Several examples of the CDT drawings labelled by the proposed system are shown in Figure 6-3. These examples demonstrate the advantage of the proposed system when digits are badly drawn or misplaced.

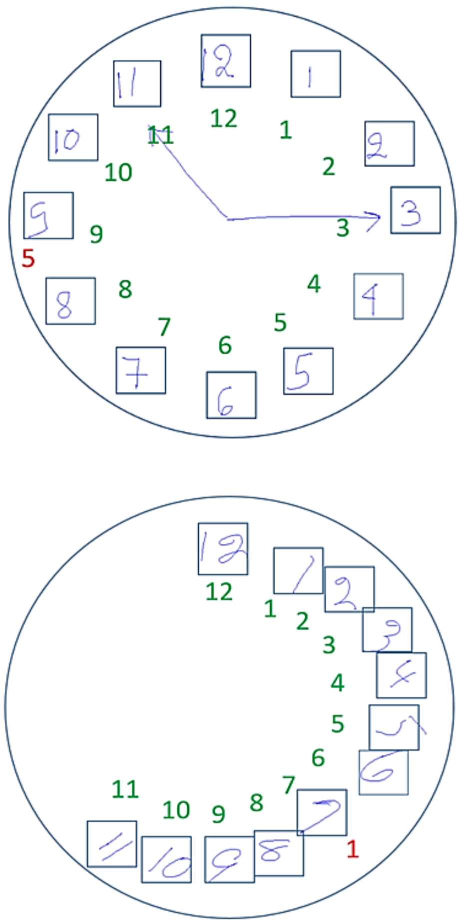

Figure 6-3: Examples of CDT interpretation system labelling: (a) example from normal data and (b) example from abnormal data.

The first CDT sketch was drawn by a healthy volunteer. Here, the digit recogniser labelled the clock number 9 as 5 using the highest probability score of 0.585 for number 5 versus 0.415 for number 9. However, the hypothesis that this was number 5 was rejected by the interpretation system, as the number was recognised with a medium level of confidence and it was in an incorrect position and in an incorrect sequence with both numbers before and after. The second hypothesis that this was number 9 was considered, as the first one was rejected. According to the inference rule number 15, this hypothesis was accepted, as the number was recognised with a medium level of confidence and it was in a correct position and in a correct sequence with both numbers before and after. The second CDT sketch was drawn by a person with dementia, and although the clock numbers were misplaced, the interpretation system was able to correct the labelling of number 7, which was misrecognised as number 1 initially. After the first hypothesis was rejected, the second hypothesis that it was number 7 was considered. The recogniser confidence was medium, and the number was in a correct sequence with both numbers before and after. According to rule number 12, this hypothesis was accepted.

Further analysis of the successful and unsuccessful sketch interpretations was conducted to understand how the system could be improved further. The analysis showed that the causes of incorrect interpretations of sketch objects were often different for the CDT sketches made by healthy people as opposed to the sketches made by people with dementia. For example, in some sketches made by healthy people, the sketch objects were misclassified due to the misrecognised neighbouring objects (i.e., when both the previous and next-in-sequence objects were classified incorrectly). According to the fuzzy rules, the predicted object class is accepted if it is in sequence with one of its neighbours. To resolve this conflict, one can consider giving more weight to the position of the objects in comparison to the sequencing information. In general, the object positions were more important than the sequencing information in the sketches made by healthy people, as they tended to place the objects in the correct places even if the writing was not legible. The situation for the CDT sketches made by dementia patients was the opposite, as they often placed objects in the wrong positions. The sketched objects were not in their correct positions for more than 25% of the sketches made by the patients with dementia. A possible solution to this problem would be to have a different set of rules for the sketches made by healthy people and people with dementia, if there were a way to distinguish between them automatically.

6.3 Summary

This chapter investigated the knowledge fuzzy rule based reasoning approach for sketch interpretation problem for online CDTs. It was shown that modelling the structure of the domain knowledge using ontology and human reasoning using fuzzy logic improved the accuracy of the interpretation system. The preliminary performance analysis showed the effectiveness of the proposed system in correcting the elements that were misrecognised by the standard recognition approach. The proposed fusion of information leads to improvement in objects' labelling even in cases containing abnormalities.

In analysing the results, it was found that generating a set of rules that works equally well for sketches made by healthy people and by those with dementia is a challenge, since there are many unusual sketches in the latter case. In addition to that, further improvement could be achieved by consider the statistical dependencies between variables and the interpretation decision for both sketches that made by healthy people and subjects with dementia. Thus a probabilistic based approach for CDT sketch interpretation will be considered in the next chapter.

Chapter: 7

Probabilistic Approach for CDT Sketch Interpretation

The primary appeal of probability theory is the ability to express useful qualitative relationships among beliefs and to process these relationships in a way that yields intuitively plausible conclusions. In the previous chapter, the limitation of the rule-based approach to capture this kind of relationship was highlighted. For example, in a healthy CDT sketch, the object's positions within the clock can have a high impact on interpreting these objects. However, within an unstructured clock drawn by a dementia patient, this information would have less impact or may be misleading. Thus different data sources, even in the same context, produce different rational dependencies. In such cases, the dependencies and the conditional probabilities between variables should be constructed dynamically to reflect these different situations. The Situational Bayesian Network (SBN) model proposed in this chapter offers a solution by providing more flexibility for the Bayesian network to provide a different BN structure based on the situation under consideration.

The main contribution of this chapter is the SBN, a hierarchal Bayesian model proposed to solve the problem of situation dependencies in CDT data. The first higher-level layer of the proposed model is a situation assessment layer, which could be regarded as a management layer responsible for guiding the flow of information and the activation process of the next layer. The second layer is a set of BNs in which each represents a specific situation, as shown in Figure 7-2 on page 108.

This chapter is organised as follows: The proposed SBN model is introduced in Section 7.1. In Section 7.2 the proposed model is applied to CDT sketch interpretation. The evaluation and discussion of the results is detailed in Section 7.3. Finally, Section 7.4 summarises the chapter.

7.1 Situational Bayesian Network

The proposed SBN model is described in this section. Bayesian networks and situation assessment backgrounds are introduced first and then the proposed SBN model is detailed.

7.1.1 Bayesian Network

A BN allows subjective interpretation of a probability: i.e., the probability of an event is the degree to which you believe that the event is true. Bayesian inference is based on Bayes' theorem, which allows inference about the parameter given the observed data. This formalism offers a natural way to represent the uncertainties in decision-making.

Bayes' theorem expresses the posterior probability $p(H|E)$ of the hypothesis after observing the evidence in terms of the prior probability $p(H)$, probability of the evidence $p(E)$ and the conditional probability of the evidence given the hypothesis $p(E|H)$.

Formally, Bayes' theorem is stated as:

$$p(H|E) = \frac{P(H,E)}{p(E)} = \frac{p(E|H)p(H)}{p(E)}\ , P(E) \neq 0 \qquad (7.1)$$

The prior $p(H)$ represents the prior belief about the hypothesis before observing any evidence. The probability of observing the evidence given the hypothesis $p(E/H)$ is called the likelihood function. The numerator $p(E)$ is called the marginal probability of evidence (E).

The structure of the Bayesian network consists of:

- a set of nodes, one per variable
- a directed, acyclic graph (links represent the direct influences)
- a conditional distribution for each node given its parents: $P(X_i \mid Parents(X_i))$

In the simplest case, conditional distribution, represented as a Conditional Probability Table (CPT), gives the distribution over X for each combination of parent values, which quantifies how much a node depends on its parents. When nodes have no parents, the CPT then represents the prior probabilities, which are obtained from the distribution of X over data.

The variables gathered from the problem domain are directly mapped onto nodes in a BN model (Constantinou, Fenton and Neil, 2016). Once all the variables of a problem domain are identified, the next stage is to describe the relationships between variables. One commonly-used approach is the causal relationship analysis (Constantinou, Fenton and Neil, 2016) by the domain experts. In addition, the dependency structure of a BN can be learnt from data (Margaritis *et al.*, 2003). Finally, the CPT parameter is defined for each node. The CPT quantifies the strength of causal relations and could be estimated subjectively using an expert's opinion. Moreover, it could be learnt statistically from the data using frequency counting. This would be applicable if there is a complete data set, meaning that observations or measurements are available on all nodes in the network (Husmeier, 2005).

Inference with BN is a large topic, since the main purpose of building a BN is to use it for inference (Guo and Hsu, 2002), that is computing the answer for particular queries about domain using a set of evidence. A number of different inference algorithms have been proposed, but in general all can be categorised into two types: exact and approximate inference. For an in-depth survey of the methods, see the paper by Guo and Hsu (Guo and Hsu, 2002).

There are well-established approaches to structure and parameter learning of a BN from a random sample of independent and identically distributed observations. However, most observations are cluster correlated. Examples include people belonging to the same family, children attending the same class or people of the same age. Learning the structure and parameters of a Bayesian network from such observations, yet ignoring their clustered correlations, typically increases the rate of false positive associations (Bae *et al.*, 2016). Data from different sources have a set of values that often provide conflicting features for the reasoning system (Dong, Berti-Equille and Srivastava, 2013). When there are several data sources, modelling the data for each individual source can solve the problem of merging conflicting data.

7.1.2 Situation Assessment

Situation assessment is a key component of any decision-making process. It is about the perception of the elements in the decision environment (Endsley, 1995). The common approach for situation assessment is the fusion of data, from different sources with different attributes, in order to draw inferences about the current situation (Bladon, Hall and Wright, 2002). A human has the ability to make a decision based on a situation. Generally, the first step conducted in the human decision process is analysing the situation. Based on the analysis of the underlying situation and the observed environmental condition, the final decision is made. Figure 7-1 depicts the modelling approach for the human decision process based on situation assessment.

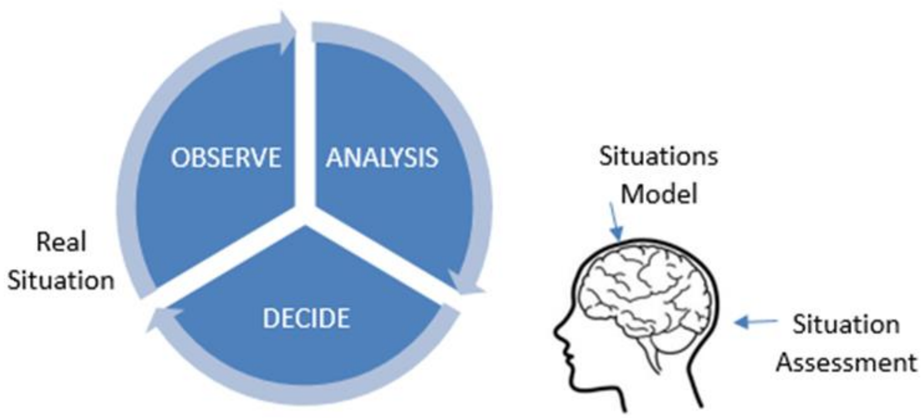

Figure 7-1: Decision process and situation analysis adapted from Bladon *et al.* (2002).

The human's ability to learn from different situations and to make a consequent decision is characterised by building a model in the mind for each situation. Transferring such a human decision process to a BN can overcome the problem of situation-related data modelling, by building a model for each situation. Then the activation of a specific model is based on the process of situations assessment.

7.1.3 SBN Model

In this section, an SBN model is proposed in order to overcome the problem of situational dependencies between data and their source. This model integrates the situation assessment approach with the BN for decision making. The proposed model is shown in Figure 7-2. The first layer in the SBN architecture is where the situations of interest are assessed. Then every identified situation is modelled by a simple BN in the subsequent layers.

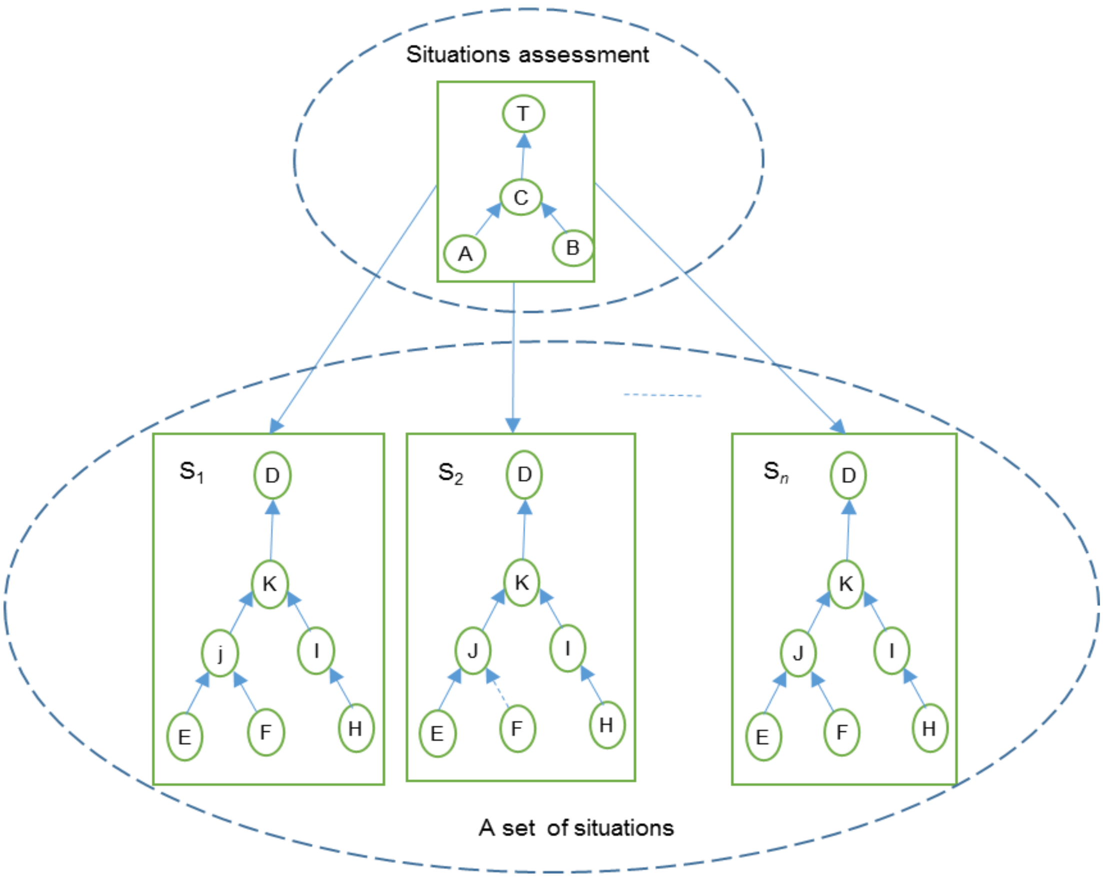

Figure 7-2: SBN Model.

The situation assessment layer is designed to evaluate the particular situation based on specific features. An important step is to determine which features are involved in the situation. Utilising human knowledge is straightforward. If there are *n* possible situations to be observed from the data, then the expected situations could be represented as a set of $S = \{S_1, S_2 ... S_n\}$. After the situation assessment is finished in the top layer, the situation's nature is determined and this will be forwarded to the next layer. Only that situation corresponding to the BN will be active in the next layer and it will be responsible for making the decision. Structure and parameter learning in the SBN is performed separately for each BN using the data corresponding to each situation or using expert knowledge. Since only one active BN will be working at a time, then, based on the situation assessment, the SBN inference will be a direct process using the appropriate BN inference algorithm.

Although each environment could be represented by different situations, this model has been developed for a limited number of situations. A larger number would increase the number of situation-specific BNs, which would increase the computational complexity and the data required in learning each BN.

7.2 CDT Sketch Interpretation Based on SBN

CDT sketch interpretation describes the process of deriving a label for each object within a clock sketch. In many applications of image interpretation, it is not sufficient to detect and classify objects based on their appearance alone. Instead, extensive prior knowledge about the scene descriptor and the spatial objects' configurations are utilised to infer the possible interpretations. Prior knowledge can also be useful to improve the results of purely appearance-based object recognition methods by ruling out unlikely detections and focussing on objects that are likely to occur but are not detected.

Recently, the advancement in probabilistic graphical models, such as BN, provided a more flexible yet consistent framework for incorporating context information in the process of sketch interpretation. The probabilistic approach models the interactions between these components using parameterised probability families by treating the image components as random variables. The dependency of each label on its context is represented by the dependency between these random variables, which can be presented on a graph. Moreover, BNs provide a clear way to map contextual constraints from the scene onto the computation of the visual interpretation by combining known causal dependencies with estimated statistical knowledge. These properties of probabilistic methods have encouraged the development of probabilistic-based models for CDT sketch interpretation in this study, which is intended to model prior knowledge about possible objects using probability theory.

The main architecture of an SBN for a CDT sketch interpretation system consists of two main components: the situational assessment layer, responsible for evaluating the situation under consideration; and the situation specific BNs. Both are detailed in the next subsections.

7.2.1 Situational Assessment Layer

Each situation in CDT sketch interpretation is represented by a simple BN in the SBN model. The dependencies between variable and CPT parameter are different from situation to situation. From observation of the CDT data, there are two situations that can be identified. These are structure and non-structure situations, which effect the dependencies between variables within a BN. A situational assessment layer based on these discriminative situations is identified. The flowchart in Figure 7-3 shows the situation assessment layer for the SBN that interprets the CDT sketch.

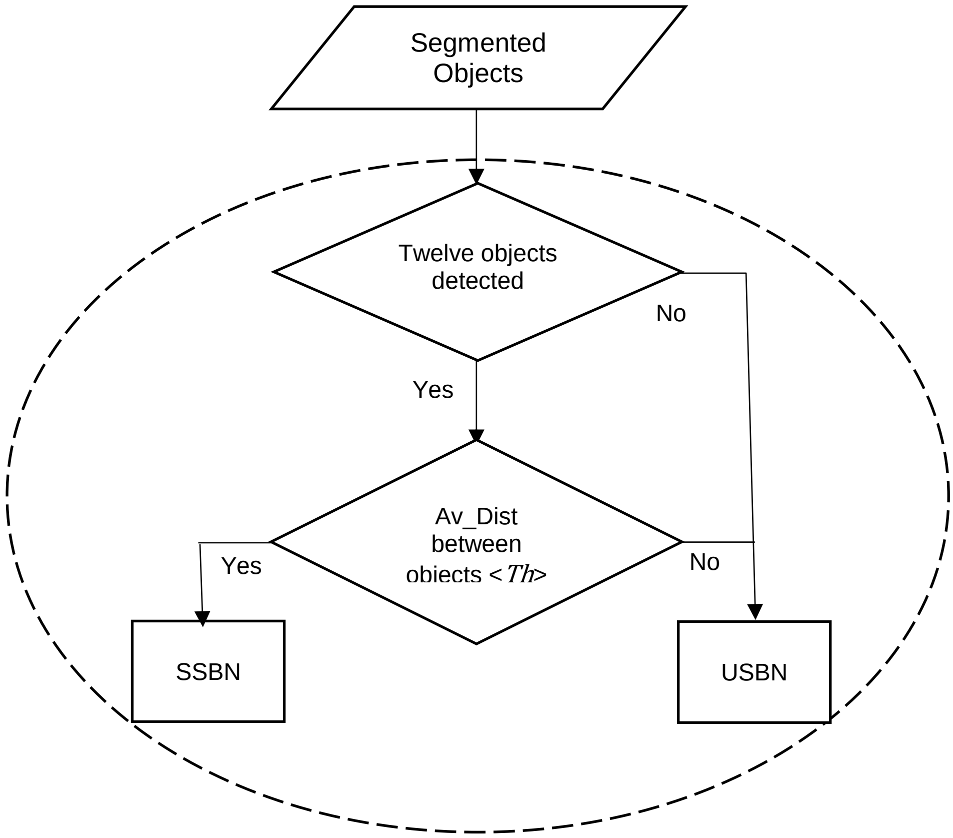

Figure 7-3: Situation assessment layer for SBN-based CDT interpretation.

The structured situation could be specified when there are twelve objects representing the numbers, and they are distributed uniformly on the circle of the clock. The average distance between these objects should be no more than a specific threshold *Th,* in order to be uniformly distributed. This threshold value could be determined empirically from the data. The unstructured situation, however, could be determined when there are fewer than twelve objects or they are not distributed uniformly. The occurrence of all twelve numbers and the uniform spacing between them is a good rule for structuring a clock as they are the first thing to be perceived by a human.

Since there are two distinct situations in a CDT that could be identified directly from the data, the situation layer could be implemented using the rules of ‘if then else’. Each situation is represented by a BN. These BNs are named the Structured Situation Bayesian Network (SSBN) and the Unstructured Situation Bayesian Network (USBN).

7.2.2 Bayesian Network Structure

The network structure for the SSBN and USBN are similar, apart from their dependencies on position. Because of the unstructured nature of a USBN, these position dependencies are eliminated. The network structure for interpreting CDT sketch objects is shown in Figure 7-4.

Identifying the main components of this structure is based on the clock ontology developed in the previous chapter. Every hypothesis generated by the recognisers is considered for verification. The digit and arrow recognisers do different tasks. The digit recogniser identifies which class the numbers belong to, while the arrow recogniser identifies whether the object is an arrow or not. More details of these recognisers are found in Chapter 5. In addition to the recogniser predictive level, other important information to include are object sequencing, position and dimension.

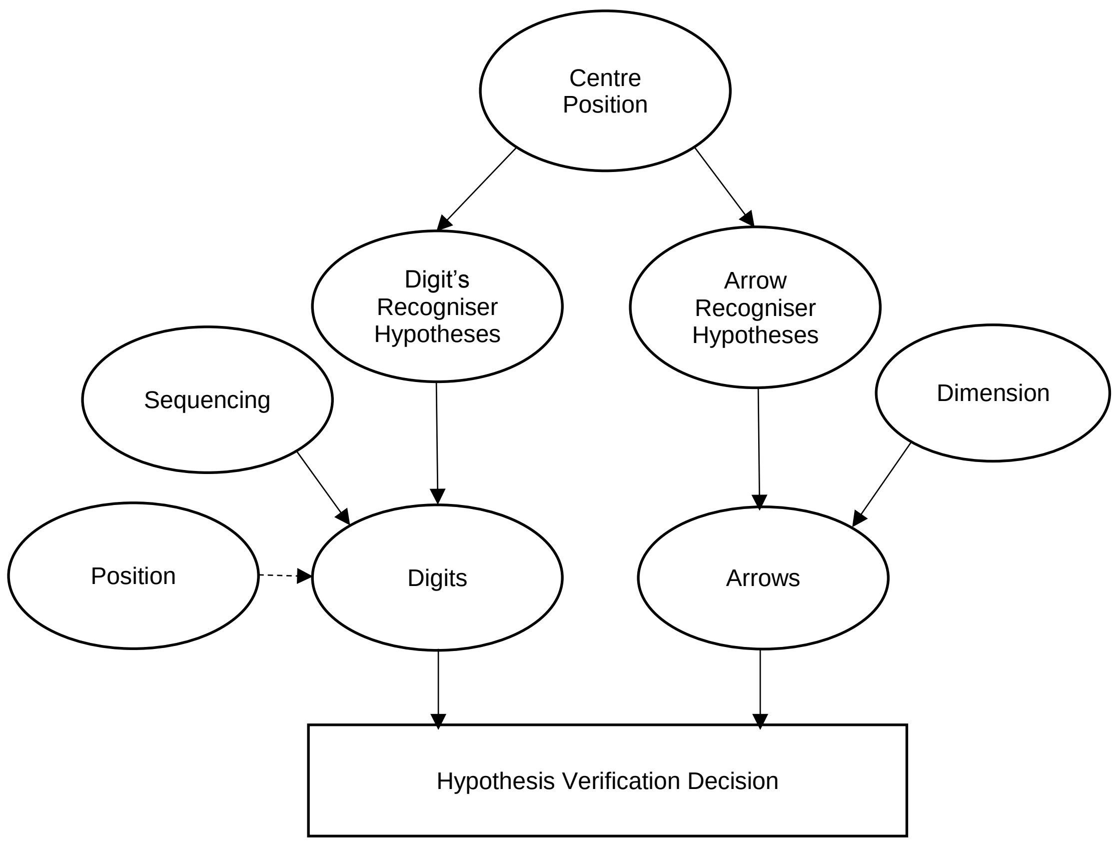

Figure 7-4: BN structure for SBN-based CDT interpretation.

Since there are several variables in the problem under consideration, building a BN structure using a human expert is the simplest way. The dependencies between variables depend on the situation under consideration. The dependencies of position are noted in Figure 7-4 to show that this dependency is removed in the case of USBN.

When the data set is complete, the CPT value is learnt from the data, as it is trivial to learn the parameters for a given structure from a complete data set. The observed frequencies are optimal with respect to the maximum likelihood estimation (Cooper and Herskovits, 1992). Moreover, the CPT values of both BNs are identified from the data that were originally divided into two cases according to the situation identified previously.

The recogniser generates a set of hypothesis based on the visual appearance of the object, and each hypothesis is verified in the BN, which is identified by the situation assessment layer. Then a decision is made as a probability, which reflects the acceptance of the hypothesis with a degree of uncertainty. All recogniser hypotheses are considered for verification and the one with the highest probability of acceptance is considered as the final interpretation of the object. Both the digit and hand recogniser are defined in Chapter 5, in which a deep neural network was involved for this purpose.

7.3 Evaluation and Discussion

This section discusses the results and performance evaluation of the proposed SBN model for a CDT sketch interpretation system. The database used in this chapter is the same data used in the previous chapters which is healthy and dementia dataset. The data set was randomly split into two subsets, one used for training and the other for testing. BNs were implemented using the BayesNet Matlab library (Murphy, 2001). Two experimental cases were conducted with each repeated ten times and the average is considered as the final results. The first experiment evaluated the efficiency of SBNs against BNs, while the second compared the SBN CDT sketch interpretation system with the classical recognition and rule-based approaches presented in previous chapters. Recognition accuracy was used as a performance indicator, which reflects the interpretation system's ability to correctly label the clock sketch's elements. The results and discussion of these two experiments are detailed in the following subsections.

7.3.1 Situational Bayesian Network

The detailed experimental results using both the healthy and dementia data sets are shown in Table 7-1. In order to show the effects of different data sources in the structure and parameter learning of the BNs, two different BNs were included and their results compared with the SBN model.

Table 7-1: CDT interpretation accuracy using BN trained only on normal data, BN trained on all data, and SBN.

	BN (normal based)	BN (all data)	SBN
Normal	100%	97.7%	100%
Abnormal	87.2%	91.85%	97.15
Overall	92.3%	94.1%	98.3%.

In the first BN, only the data obtained from healthy volunteers were used in training. In the second BN, the data from dementia patients' sketches were included in addition to heathy data.

It is clear from the results that including or excluding healthy or dementia data in a BN parameter affects the final performance of the interpretation system. When using only healthy data to construct the first BN, 100% recognition was achieved, as well as when the system was tested on data from the same healthy source. This means the system is able to correctly label all the elements of the clocks. In these data, there was no irrelevant handwriting, and all objects were present and placed within their expected positions. The incorrect labelling of the classical recogniser is amended using information gathered from object neighbours and their positions. Using the same model for interpreting sketches from dementia patients can degrade the efficiency of the interpretation system, in which an accuracy of 87.2% was achieved. This result is expected, as dementia data naturally have abnormal cases and the BNs had never learnt from them. More than 50% of these sketches are abnormal from a general consideration, such as occurrence of numbers and their position. In the second BN, all clock sketches sourced from healthy and demented people were considered for the structuring and learning of the BN model. The recognition accuracy of the interpretation system improved to 91.85% when tested with dementia data. However, a decline of 2.3% was recorded in interpreting healthy sketches using this model in comparison with the previous model, where dementia data was not included.

The final part of this section compares the proposed SBN with the normal BN. Using the SBN, the interpretation system’s performance improved: there was 100% accuracy in labelling all objects in the clocks sketched by healthy people. Moreover, a 97.15% recognition accuracy was reported in interpreting other sketched objects related to dementia cases, with more than 5% improvement in performance over the previous model (that is, BNs reading heathy and dementia data). This major improvement shows the effectiveness of the proposed SBN model in dealing with data that are within the same context but originating from different sources, by modelling each situation individually. The system performance is associated with the situation assessment layer performance. However, with good structuring of the situational layer in the proposed system, all data successfully associated with it is intended BN. That is the SSBN and USBN, as detailed in Section 7.2.

Representing a single model to deal with dementia and healthy sketches is a challenge, one which reduces the applicability of BNs in such problems. The proposed SBN model enhances the applicability of BN in this area, by enabling flexible network structuring.

7.3.2 Probabilistic CDT Sketch Interpretation

A comparative study between different approaches for CDT sketches interpretation is described in this section to show the effectiveness of the probabilistic-based SBN for this task. The system performance is compared to other interpretation approaches conducted in previous chapters. Table 7-2 shows the detailed recognition accuracy of the interpretation system in terms of recognising clock elements—digits, hands and irrelevant handwriting—for both healthy and dementia data. The comparative analysis is based on using the classical recognition, the fuzzy rule and the probabilistic-based SBN approaches.

The overall recognition accuracy for the proposed SBN is 98.3%. Using the rule-based approach, 96.2% was recorded. Both approaches used the domain knowledge to improve the recognition accuracy of the recogniser but in different forms of representation. Meanwhile, in the classical recognition-based method, the recogniser made a decision based only on the visual appearance of the object; 94.45% accuracy was achieved in recognising overall CDT sketched elements.

Table 7-2: CDT interpretation accuracy using recognition-, rule-based and SBN approaches.

Data Set	Objects	#objects	% Accuracy Recognition Based	Rule Based	SBN
Over all Data	Digits	2438	94.7	96.6	98.3
	Hands	292	93.4	94.5	97.94
	Irrelevant	22	77.2	81.8	100
	Overall	2752	**94.45**	**96.2**	**98.3**
Normal Data	Digits	975	97.3	98.5	100
	Hands	130	97.7	98.4	100
	Irrelevant	_	_	_	_
	Overall	1105	97.2	98.46	100
Abnormal Data	Digits	1463	93	95.4	97.2
	Hands	162	90.1	91.3	96.2
	Irrelevant	22	77.2	81.8	100
	Overall	1647	92.5	94.8	97.15

In order to determine the significance of the improvement achieved due to the proposed probabilistic approach over the classical recognition approach, the experiments were conducted ten times and the results were subjected to statistical analysis. The results of the statistical t-test showed the P-value to be 1.6E-13 rejecting the null hypothesis, which means that the improvement is unlikely to happen by chance.

The proposed SBN interpretation system achieved 100% accuracy for labelling the CDT sketch elements in healthy data. The SBN interpretation system correctly interpreted all elements that were visually hard to recognise by the classifier and achieved equivalent performance in deciding the correct labelling for both digits and hands. However, the performance was less in the dementia data, with reported results of 97.15%. But this result is considered a success due to the impaired degree of these sketches, both in terms of appearance and unstructured writing. In the dementia data a great improvement was reported in recognising clock hands—more than 3% greater than the rule-based system and 4% larger than recognition-based. Moreover, the system successfully labelled all irrelevant writing in the dementia data in which only 22 objects were noticed.

Two examples of CDT drawings that were successfully labelled by the proposed system are shown in Figure 7-5. These examples demonstrate the advantage of the proposed system over the fuzzy rule-based system. In the first CDT sketch, the digit recogniser visually recognised the number 5 with a high level of confidence. Using the rule-based system such a hypothesis was rejected, since the number 5 is not supposed to be in this position and it is not in sequence with previous and subsequent digits. A similar problem was detected in the second example, with the numbers 2, 4, 6 and the two 8s. Other clock numbers in this example were recognised, even when they are not in their correct position, because their sequencing is preserved. Since both sketches are unstructured, these limitations can be overcome by using the SBN interpretation system that greatly depends on the recogniser by associating more weight to its prediction, which could be considered the best assessor in such situations.

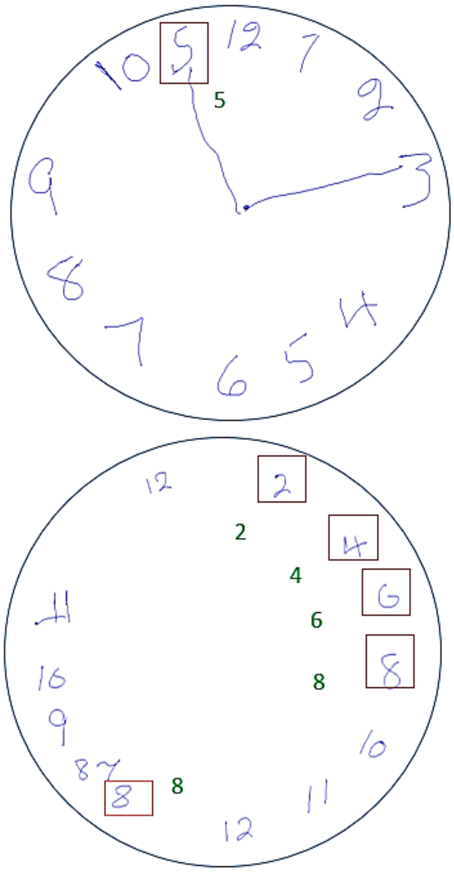

Figure 7-5: Successful CDT interpretations from dementia data using the proposed SBN.

Further analysis of the unsuccessful sketch interpretation cases were conducted to better understand system performance. The least successful cases came from unstructured data, where the SBN interpretation system highly relied on the recogniser's assessment, as stated before. If the recogniser misclassifies sketch elements with a high predictability, the possibility of correction by the interpreter is very low. Examples of such cases are given in Figure 7-6. Sketch elements that are visually distorted and not in their correct position or sequence could not be handled by the interpretation system. These cases would be difficult even for human interpreters.

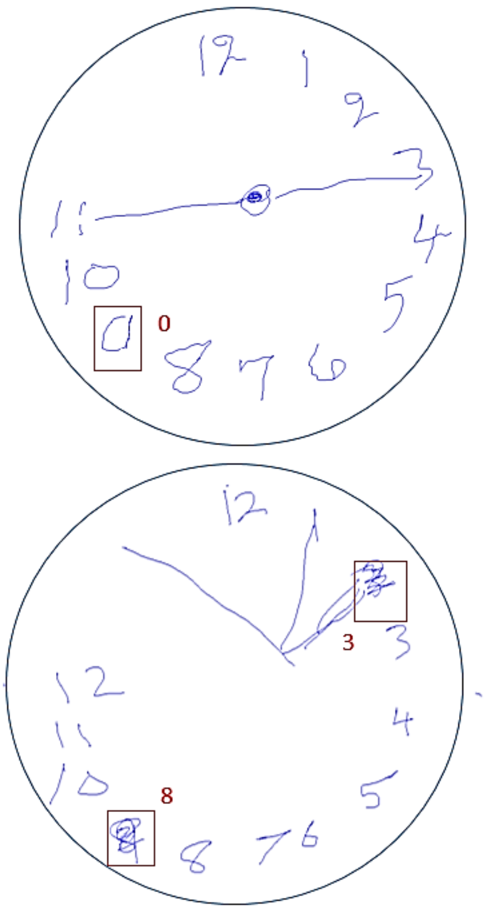

Figure 7-6: Unsuccessful CDT interpretations from dementia data using the proposed SBN.

Representing knowledge in the form of a ruleset does not provide flexibility to the decision system, since it decides strictly by the rules. These rules at the best of times cannot reflect the rational dependencies between the reasoning variables. Each variable in the reasoning stage should have some effect on the final decision. That would reflect the human's everyday conditional reasoning in which each piece of evidence is given weight. Moreover, fine-tuning a rule-based system to comprise a data set is elaborating (Onisko, Lucas and Druzdzel, 2001) and this task could be challenging if these data are related to two different situations. BNs resolve this problem by representing domain knowledge and decision making in the form of conditional dependencies between reasoning elements in a probabilistic form. However, the dependencies between BN learning and data is an obstacle if the data relate to different sources. The proposed SBN provides flexibility to the BN to deal with such cases by modelling each situation individually.

To conclude, the probabilistic SBN approach for reasoning showed improvement over classical recognition- and rule-based approaches. Involving the theory of situation assessment in decision making in modelling BN is highly promising.

7.4 Summary

This chapter presented a novel SBN model for interpreting CDT sketches. A probabilistic-based approach was proposed for modelling human reasoning in interpreting these drawings. The object's visual appearance forms the first decision hypothesis for the system, and then reasoning is conducted using domain knowledge in a probabilistic form to arrive at a final decision. The recogniser result from Chapter 5 is used as a hypothesis generator, in which a set of possible interpretations is suggested with a different level of certainty. By using a mixed database in which some follow a regular structure and others do not, one can influence the learning and structuring of a BN and limit its applicability in this area. SBN is a proposed framework that solves this problem by constructing a different situation-based model according to a situation assessment approach. In this way, the SBN overcomes the inflexibility of BNs to adapt to the environment being modelled, since it has a fixed structure and fixed conditional probability for each node. In addition, the presented approach contributes to the efficient application of probability theory for the interpretation of images depicting complex scenes, such as sketches by dementia patients. The results show the novelty of the proposed model in comparison with the approaches outlined in previous chapters.

Moreover, while SBN is proposed for interpreting CDT sketches, it is applicable in other domains where learning a single model for data is not convenient due to the irregularity and fusion of different situation data within the same context.

Chapter: 8

Conclusion and Future Work

This chapter concludes the work presented in this thesis. Section 8.1 lists the main contributions of this research. Section 8.2 provides the conclusions. Directions for future work are discussed in Sections 8.3.

8.1 Contributions

The contributions of this study are as follows:

- A conceptual model of an automatic CDT sketch interpretation system is proposed and described. The model is designed to support the development of a computerised CDT for automatic assessment of neurological cognitive impairment. The system integrates the low level features, such as object appearance, with high level features extracted from contextual knowledge, such as sketch structure in a reasoning platform.

- A new segmentation algorithm is developed, one which utilises the advantages of employing online temporal features along with the traditional spatial features extracted from the sketched images.

- A new algorithm for recognising handwritten digits is developed. The algorithm is based on two complementary sources of data, namely static and dynamic information extracted from handwritten data. The proposed system advances the state of the art in the field of recognising totally unconstrained numerals that are handwritten online.

- Two approaches are proposed based on different methodologies for modelling human reasoning during the task of CDT sketch interpretation. These are fuzzy rules and BN-based approaches enriched with ontology. They allow for the managing of uncertainty related to the expert knowledge in a particular domain. These approaches constitute an adequate framework for knowledge representation and reasoning that can be used as a basis for decision making, predictions and other related tasks.

- A new framework for a BN is proposed, termed an SBN. The SBN is a hierarchal model that is based on integrating situation assessment processes within the BN architecture. This framework increases the flexibility of a BN for dealing with the fusion of data from different sources within the same context. Whilst the system was developed for CDT sketches, the idea of integrating situational assessment with BNs is universal and could be considered for other domains where combining different situational data might cause a conflict during the BN's structure and parameter learning.

8.2 Conclusions

The main aim of this study was to develop an automatic CDT sketch interpretation system to support the development of a computerised CDT for neurological cognitive impairment assessment.

The first objective was to propose a conceptual model for a CDT sketch interpretation system. The conceptual model was designed to incorporate context knowledge along with an object's visual appearance. The integration was implemented using a hybrid interpretation strategy. The bottom-up strategy, which entails the object's feature extraction and recognition, and the top-down strategy, which is hypothesis generation and feature expansion, were combined to exploit the advantages of both approaches. The model provides an overview of the employed processing steps, such as segmentation, feature extraction, recognition and knowledge representation. The information derived from these different forms of representation was integrated using an inference engine to derive the final interpretation of CDT sketches.

The next objective was to develop a segmentation algorithm to address the problem of unconstrained handwriting segmentation. Current algorithms that rely on the width and height of segmented patterns and horizontal gaps between segments are not applicable to this kind of handwriting, as exemplified by CDT sketches. This thesis proposes a new set of temporal and spatial features that are automatically extracted from the CDT data. Consequently, a supervised machine learning classifier is proposed to segment the CDT drawings into their elements, such as numbers and clock hands, based on the extracted features. The developed algorithm was evaluated on CDT data sets; drawings made by both healthy people and dementia patients were considered. The experimental results showed a significant improvement over a competing approach employed by other researchers within the context of CDT sketch segmentation.

The next objective was to develop a new system to deal with recognition of unintelligible handwriting. The problem was divided into two parts, based on the main objects in the CDT: that is, digit recognition and clock hand recognition. A new combination of classifiers was proposed in this area, with CNN representing the static image features, and *kNN* representing dynamic features extracted from the online sequence of stroke points. Classifiers using these representations make errors on different patterns, which suggests that a suitable combination of the two would lead to higher accuracy. This combination of the classifiers was applied to digit recognition. The proposed digit recognition algorithm was tested on two sets of data: 1) Pendigits online handwritten digits; and 2) digits from the CDT data set collected for this study. The test on both data sets showed that the proposed combination system outperformed each classifier individually in terms of recognition accuracy, especially when assessing the handwriting of people with dementia.

Due to the fact that the clock hands (normally represented by arrows) have a high level of deformation, only static image representation was used for recognising these objects. Recently, CNNs have emerged as a powerful framework for representing and recognising features from a variety of image domains. However, this is the first study that explored applying CNNs to recognition of objects sketched by cognitively impaired persons. The experiment results showed that a CNN is a good feature learner and has the potential to improve performance in related problems of recognising unintelligible handwriting associated with medical conditions.

The final objective was to develop a knowledge-based reasoning system for CDT sketch interpretation. This involved two subtasks: representing knowledge and creating an inference engine. An Ontology was proposed to represent the declarative knowledge represented by the clock concept, and then two alternative approaches were developed in reasoning—namely, rule-based and probabilistic-based approaches. In the first approach, Ontology was enriched with fuzzy rules to deal with uncertainty. The preliminary performance analysis showed the effectiveness of the proposed system in correcting the elements that were misrecognised by the standard recognition approach. However, after analysing the results, it was revealed that further improvement could be achieved by considering the statistical dependencies between beliefs and by dealing separately with normal and abnormal sketches. This problem was resolved by proposing a new situational probabilistic network SBN model for CDT sketch interpretation.

Data from different sources can have a set of features containing conflicting values for the reasoning system, which is the case with normal and abnormal sketches in a CDT dataset. In this thesis, SBN was proposed as a hierarchal model that allows fusing this kind of data by constructing a model for each situation. On top of these models, a situation assessment layer was added. This incorporating of a situational assessment approach from a decision making discipline within the BN architecture provided a mechanism for managing the flow of information within its hierarchal architecture. The evaluation results of this proposed system demonstrated the advantages of SBN in CDT sketch interpretation in comparison with classical recognition-based and rule-based approaches. A significant improvement of recognition accuracy was reported with 100% for normal sketches and 97.15% for abnormal sketches. The system showed promising results, even for abnormal sketches that present a challenge for human interpreters.

The conducted study has demonstrated that the developed approach is feasible for CDT sketch interpretation; however, some limitations have been identified. The main limitation of this study is the lack of standard public data sets of CDT sketches, a lack which may be attributed to the newness of the topic and the privacy of medical data. This limitation prevented the proposed approaches from being compared in more detail to the works of other researchers in the field of CDT sketch interpretation. Another identified limitation is the relatively small size of the data set involved in this study. That was due to lack of access to dementia patients for ethical reasons. Although the main application of this study is not medical diagnosis, a larger data set could further improve the system's training, especially in abnormal cases.

8.3 Future Work

The techniques proposed in this thesis have raised a number of interesting possibilities for future work. Some of the more prominent research directions are noted here.

- Although the performance figures suggest that the system works well for interpreting CDT sketches by both healthy and impaired users, there is a need to improve the interpreter's performance—not only in its recognition accuracy, but also its speed. This factor is especially important in pen-based or touch screen applications, in which it is desirable to have the results appear in real time.

- Object segmentation plays an important role in the CDT sketch interpretation system. In the proposed segmentation method, however, some authentic segmentation points are misclassified as non-segmentation points and vice versa. Perhaps the most important direction for improvement is to explore alternative approaches by considering feedback from the recognition system, that is, segmentation and recognition work simultaneously. Human segmentation of objects is clearly guided in part by understanding the drawings.

- Another research path which deserves to be followed is finding a new representation format that combines both static and dynamic features in sketched object recognition using a CNN. Recently, a CNN was employed for sketched object recognition; however, dynamic features were not considered. Such features account for unique sketch characteristics and are shown to be important for the recognition process, especially when they are combined with complementary static features as revealed by this research.

- In this thesis, the focus was on CDT sketch interpretation. However, the developed interpretation model could be applied to other domains that require context modelling such as sketched-based image retrieval and video retrieval, which could be interesting avenues for future work.

References

Aguilar, J. (2016) ‘Autonomic decision making based on bayesian networks and ontologies’, *International Joint Conference on Neural Networks (/JCNN)*, pp. 3825–3832.

Alimoglu, F. and Alpaydin, E. (1997) ‘Combining Multiple Representations and Classifiers for Pen-based Handwritten Digit Recognition’, in *Proceedings of the International Conference on Document Analysis and Recognition,ICDAR*, pp. 637–640.

Alimoglu, F. and Alpaydin, E. (2001) ‘Combining Multiple Representations for Pen-based Handwritten Digit Recognition’, *Turk J Elec Engin*, 9(1), pp. 1–12.

Alvarado, C. (2011) ‘Multi-domain Hierarchical Free-Sketch Recognition Using GraphicalModels’, in *Sketch-based Interfaces and Modeling*, pp. 1–402.

Alvarado, C. and Davis, R. (2005) ‘Dynamically constructed Bayes nets for multi-domain sketch understanding’, *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1407–1412.

Arandjelović, R. and Sezgin, T. M. (2011) ‘Sketch recognition by fusion of temporal and image-based features’, *Pattern Recognition*, 44, pp. 1225–1234.

Arica, N. and Yarman-vural, F. T. (2001) ‘An Overview of Character Recognition Focused on Off-Line Handwriting’, *IEEE Transaction on System, Man, and Cybernetics-Part C*, 31(2), pp. 216–233.

Aurangzeb, K. *et al.* (2010) ‘A Review of Machine Learning Algorithms for Text-Documents Classification’, *journal of Advances in information technology*, 1(1).

Bae, H. *et al.* (2016) *Learning Bayesian Networks from Correlated Data*. Nature Publishing Group.

Bai, Y., Zhuang, H. and Roth, Z. S. (2005) ‘Fuzzy logic control to suppress noises and coupling effects in a laser tracking system’, *IEEE Transactions on*

Control Systems Technology, 13(1), pp. 113–121.

Bannour, H. and Hudelot, C. (2011) 'Towards ontologies for image interpretation and annotation', *Proceedings - International Workshop on Content-Based Multimedia Indexing*, pp. 211–216.

Bao, S. Y., Sun, M. and Savarese, S. (2011) 'Toward coherent object detection and scene layout understanding', *Image and Vision Computing*. Elsevier B.V., 29(9), pp. 569–579.

Barbini, E., Manzi, P. and Barbini, P. (2013) 'Bayesian Approach in Medicine and Health Management', in *Current Topics in Public Health 2.* InTech, pp. 17–35.

Bastinos, A. Š. and Krisper, M. (2013) 'Multi-criteria decision making in ontologies', *Information Sciences*. Elsevier Inc., 222, pp. 593–610.

Bay, S. D. (1998) 'Nearest Neighbor Classification from Multiple Feature Subsets', *Intelligent Data Analysis*, 3, pp. 191–209.

Behnke, S. (2003) *Hierarchical Neural Networks for Image Interpretation*. Springer Science & Business Media.

Ben-Hur, A. and Weston, J. (2010) 'A user's guide to support vector machines', in *Data Mining Techniques for the Life Sciences*, pp. 223–239.

Bennasar, M. (2014) *Clinical Decision Support System for Early Detection and Diagnosis of Dementia, PhD thesis*. Cardiff University, UK.

Berka, P., Athanasiadis, T. and Avrithis, Y. (2006) 'Rule-based reasoning for semantic image segmentation and interpretation', in *the 1st International conference on Semantic and Digital Media Technologies*, pp. 39–40.

Berndt, D. and Clifford, J. (1994) 'Using dynamic time warping to find patterns in time series', *Workshop on Knowledge Knowledge Discovery in Databases*, 398, pp. 359–370.

Bikakis, A. and Antoniou, G. (2010) 'Rule-Based Contextual Reasoning in Ambient Intelligence', in *Proc. International Symp. on Semantic Web Rules*, pp.

74–88.

Bilan, Z. and Masaki, N. (2008) ‘Segmentation of on-line freely written Japanese text using SVM for improving text recognition’, *IEICE Transactions on Information and Systems*, E91–D(1), pp. 105–113.

Bladon, P., Hall, R. J. and Wright, W. A. (2002) ‘Situation Assessment using Graphical Models’, in *Proceedings of the Fifth International Conference on Information Fusion , IEEE*, pp. 886–893.

Bobillo, F. and Straccia, U. (2016) ‘The fuzzy ontology reasoner fuzzyDL’, *Knowledge-Based Systems*, 95, pp. 12–34.

Borson, S. O. O. *et al.* (1999) ‘THE MINI- COG :A COGNITIVE “VITAL SIGNS” MEASURE FOR DEMENTIA SCREENING IN MULTI - LINGUAL ELDERLY’, *Intenational Journal Of geriatric Psychiatry*, (15), pp. 910–916.

Boyacioglu, M. A. and Avci, D. (2010) ‘An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange’, *Expert Systems with Applications*. Elsevier Ltd, 37(12), pp. 7908–7912.

Bresenham, J. E. (1965) ‘Algorithm for computer control of a digital plotter’, *IBM Systems Journal*, 4(1), pp. 25–30.

Bresler, M., Průša, D. and Hlaváč, V. (2016) ‘Online recognition of sketched arrow-connected diagrams’, *International Journal on Document Analysis and Recognition*, 19(3), pp. 253–267.

Brown, J. and Brown, L. A. (2009) ‘Self administered cognitive screening test (TYM) for detection of Alzheimer’s disease: cross sectional study’, *BMJ 2009;338:b2030*.

Buchanan, B. and Shortliffe, E. (1984) *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Reading, MA: Addison-Wesley.

Burges, C. J. (1998) ‘A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery*, 2(2), pp. 121–167.

Campbell, N. W., Mackeown, W. P. J. and Thomas, B. T. (1997) 'Interpreting Image Databases by Region Classi cation', *Pattern recognition*, 30(4), pp. 555–563.

Cardoso, J. S. and Corte-Real, L. (2005) 'Toward a generic evaluation of image segmentation.', *IEEE transactions on image processing*, 14(11), pp. 1773–1782.

Casey, R. G. and Lecolinet, E. (1996) 'A survey of Methods and Strategies in Character Segmentation', *IEEE Transaction on Pattern Analysis and machine Intelligence*, 18(7), pp. 690–706.

Castro-Bleda, M. J. *et al.* (2009) 'Improving a DTW-Based Recognition Engine for On-line Handwritten Characters by Using MLPs', in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 1260–1264.

Charbonneau, S., Whitehead, V. and Collin, I. (2005) 'The Montreal Cognitive Assessment , MoCA : A Brief Screening', *Journal of the American Geriatrics Society*, 53(4), pp. 695–699.

Chatfield, K. *et al.* (2014) 'Return of the Devil in the Details: Delving Deep into Convolutional Nets', in *BMCV*.

Chen, S. Q. (2000) *Comparing Probabilistic and Fuzzy Set Approaches for Design in the Presence of Uncertainty, PhD thesis*. Virginia Polytechnic Institute and State University, Virginia, USA.

Cho, Y. S., Kim, H. and Do, E. Y.-L. (2010) 'Curved-Straight Lines-Analysis (CSLA) Algorithm for Handwritten Digit Recognition Enhancement', in *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, pp. 150–154.

Ciresan, D., Meier, U. and Schmidhuber, J. (2012) 'Multi-column Deep Neural Networks for Image Classification', in *CVPR*.

Cohen, P. R. *et al.* (1997) 'QuickSet: multimodal interaction for distributed applications', in *Proceedings of the 5th ACM International Multimedia Conference*, pp. 31–40.

Constantinou, A. C., Fenton, N. and Neil, M. (2016) 'Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved', *Expert Systems with Applications*. Elsevier Ltd, 56, pp. 197–208.

Cooper, G. F. and Herskovits, E. (1992) 'A bayesian method for the induction of probabilistic networks from data', *Machine Learning*, 9(4), pp. 309–347.

Cortes, C. and Vapnik, V. (1995) 'Support Vector Networks', *Machine Learning*, 20(3), pp. 273–297.

Cover, T. and Hart, P. (1967) 'Nearest neighbor pattern classification', *IEEE Transactions on Information Theory*, 13(1), pp. 21–27.

Culter, N. R. *et al.* (1993) 'the Use of the Computerized Neuropsychological Test Battery (Cntb) in an Efficacy and Safety Trial of Bmy-21,502 in Alzheimers-Disease', *Annals of the New York Academy of Sciences*, 695(1), pp. 332–336.

Dave, N. (2015) 'Segmentation Methods for Hand Written Character Recognition', *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(4), pp. 155–164.

Davis, R. *et al.* (2014) 'Think : Inferring Cognitive Status from Subtle Behaviors', in *Proceedings of the Twenty-Sixth Conference on Innovative Applications of Artificial Intelligence*, pp. 2898–2905.

Deborah, F. O., Elijah, O. O. and Olaosebikan, A. F. (2012) 'Development of a Feature Extraction Technique for Online Character Recognition System', *Innovative Systems Design and Engineering*, 3(3), pp. 10–23.

Dipartimento, M. *et al.* (1991) 'An APL Rule-Based System Architecture For Image Interpretation Strategies', *In ACM SIGAPL APL Quote Quad*, 21(4), pp. 51–61.

Dong, X. L., Berti-Equille, L. and Srivastava, D. (2013) ‘Data fusion: Resolving conflicts from multiple sources’, in *Handbook of Data Quality*. Springer Berlin Heidelberg, pp. 293–318.

Dougherty, J. H. *et al.* (2010) ‘The computerized self test (CST): an interactive, internet accessible cognitive screening test for dementia.’, *Journal of Alzheimer’s Disease*, 20(1), pp. 185–95.

Dubois, D. and Prade, H. (1993) ‘Fuzzy sets and probability: misunderstandings, bridges and gaps’, in *Second IEEE International Conference on Fuzzy Systems*, pp. 1059–1068.

Durand, N. *et al.* (2007) ‘Ontology-based object recognition for remote sensing image interpretation’, in *19th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2007*, pp. 472–479.

Endsley, M. (1995) ‘Toward a theory of situation awareness in dynamic systems’, *The Journal of the Human Factors and Ergonomics Society,* 37(1), pp. 32–64.

Ertekin, S. (2009) *Learning in Extreme Conditions: Online and Active Learning with Massive, Imbalanced and Noisy Data, PhD thesis*. The Pennsylvania State University.

Fairhurst, M. C. *et al.* (2008) ‘Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling’, *Behavior Research Methods*, 40(1), pp. 290–303.

Fleuret, F. *et al.* (2011) ‘Comparing machines and humans on a visual categorization test.’, in *Proceedings of the National Academy of Sciences of the United States of America*, pp. 17621–5.

Forestier, G. *et al.* (2012) ‘Knowledge-based region labeling for remote sensing image interpretation’, *Computers, Environment and Urban Systems*. Elsevier Ltd, 36(5), pp. 470–480.

Fosseng, S. (2013) *Learning Distance Functions in k-Nearest Neighbors, MSc thesis*. Norwegian University of Science and Technology.

Freedman, M. *et al.* (1994) *Clock Drawing A Neurological Analysis*. New York: Oxford University Press.

Fukushima, T. and Nakagawa, M. (2000) ‘Online writing-box-free recognition of handwritten Japanese text considering character size variations’, in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*.

Furukawa, N. *et al.* (2006) ‘Online Character Segmentation Method for Unconstrained Handwriting Strings Using Off-stroke Features’, in *Tenth International Workshop on Frontiers in Handwriting Recognition*.

Geiger, A. *et al.* (2014) ‘3D traffic scene understanding from movable platforms’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5), pp. 1012–1025.

Geman, D. *et al.* (2015) ‘Visual Turing test for computer vision systems.’, *Proceedings of the National Academy of Sciences*, 112(12), pp. 3618–23.

Gonzalez, R., Woods, R. and Eddins, S. (2004) *Digital Image Processing Using MATLAB*, *Digital Image Processing Using Matlab*. New Jersey: Prentice Hall.

Grimm, S., Hitzler, P. and Abecker, A. (2007) ‘Knowledge Representation and Ontologies Logic, Ontologies and Semantic Web Languages’, *Semantic Web Services*, pp. 51–105.

Grober, E. *et al.* (2008) ‘Primary care screen for early dementia’, *Journal of the American Geriatrics Society*, 56(2), pp. 206–213.

Gu, T. *et al.* (2004) ‘An Ontology-based Context Model in Intelligent Environments’, in *Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference*, pp. 270–275.

Guha, A., Kim, H. and Do, E. Y. (2010) ‘Automated Clock Drawing Test through Machine Learning and Geometric Analysis’, in *Proceedings of the International Conference on Distributed Multimedia Systems, DMS*, pp. 311–314.

Guihuan Feng, Christian Viard-Gaudin, Z. S. (2009) ‘On-line hand-drawn electric circuit diagram recognition using 2D dynamic programming’, *Pattern*

Recognition, 42, pp. 3215–3223.

Guo, H. and Hsu, W. (2002) ‘A survey of algorithms for real-time Bayesian network inference’, in *AAA/KDD/UAI-2002 Joint Workshop on Real-Time Decision Support and Diagnosis System*.

Hall, M. *et al.* (2009) ‘The WEKA Data Mining Software: An Update’, *SIGKDD Explorations*, 11(1), pp. 10–18.

Hanson, M. L. *et al.* (2004) ‘Automated Data Fusion and Situation Assessment in Space Systems’, *International Journal on Artificial Intelligence Tools*, 13(1), pp. 255–272.

Hedau, V., Hoiem, D. and Forsyth, D. (2010) ‘Thinking inside the box: Using appearance models and context based on room geometry’, *Lecture Notes in Computer Science*, 6316 LNCS(PART 6), pp. 224–237.

Heinik, J. *et al.* (2010) ‘Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study.’, *International psychogeriatrics*, 22(3), pp. 479–88.

Hoiem, D., Efros, A. A. and Hebert, M. (2008) ‘Putting objects in perspective’, *International Journal of Computer Vision*, 80(1), pp. 3–15.

Hornik, K., Stinchcombe, M. and White, H. (1989) ‘Multilayer feedforward networks are universal approximators’, *Neural Networks*, 2(5), pp. 359–366.

Hse, H. H. and Newton, A. R. (2005) ‘Recognition and beautification of multi-stroke symbols in digital ink’, *Computers and Graphics (Pergamon)*, 29(4), pp. 533–546.

Hse, H. and Newton, A. R. (2004) ‘Sketched symbol recognition using zernike moments’, in *Proceedings - International Conference on Pattern Recognition*, pp. 367–370.

Hsiao, E. and Hebert, M. (2012) ‘Occlusion reasoning for object detection under arbitrary viewpoint.pdf’, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 36(9), pp. 3146–3153.

Huang, H. *et al.* (2017) 'Ontology-Guided Image Interpretation for GEOBIA of High Spatial Resolution Remote Sense Imagery: A Coastal Area Case Study', *ISPRS International Journal of Geo-Information*, 6(4), pp. 1–17.

Hudelot, C., Atif, J. and Bloch, I. (2008) 'Fuzzy spatial relation ontology for image interpretation', *Fuzzy Sets and Systems*, 159(15), pp. 1929–1951.

Husmeier, D. (2005) 'Introduction to Learning Bayesian Networks from Data', in *Probabilistic modeling in bioinformatics and medical informatics*. Springer London, pp. 17–57.

Ishibuchi, H., Nakashima, T. and Nii, M. (2006) *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*. Springer-Verlag Berlin Heidelberg.

Ismail, Z., Rajji, T. K. and Shulman, K. I. (2010) 'Brief cognitive screening instruments : an update', *International Journal of Geriatric Psychiatry*, 25, pp. 111–120.

Jabari, S. and Zhang, Y. (2013) 'Very high resolution satellite image classification using fuzzy rule-based systems', *Algorithms*, 6(4), pp. 762–781.

Jain, A. K., Murty, M. N. and Flynn, P. J. (1999) 'Data clustering: a review', *ACM Computing Surveys*, 31(3), pp. 264–323.

Jeet, K. and Dhir, R. (2012) 'Bayesian and Fuzzy Approach to Assess and Predict the Maintainability of Software: A Comparative Study', *ISRN Software Engineering*, 2012, pp. 1–10.

Johnston, D. and Alvarado, C. (2013) *Sketch Recognition of Digital Logical Circuits*. California University, USA.

Jorge, J. and Samavati, F. (2011) *Sketch-based Interfaces and Modeling*, London: Springer Science and Business Media.

Jouk, A. (2007) *A Reduced Scoring System for the Clock-Drawing Test Using a Population-Based Sample, MSc thesis*. University of Victoria,USA.

Juby, A., Tench, S. and Baker, V. (2002) 'The value of clock drawing in identifying executive cognitive dysfunction in people with a normal Mini-Mental State Examination score.', *CMAJ : Canadian Medical Association journal CMAJ*, 167(8), pp. 859–64.

Kara, L. B. and Stahovich, T. F. (2005) 'An image-based, trainable symbol recognizer for hand-drawn sketches', *Computers and Graphics (Pergamon)*, 29(4), pp. 501–517.

Keller, J. M., Michael, G. R. and Givens, J. james A. (1985) 'A fuzzy K-Neirest Neighbor Algorithm', *IEEE Transactions on Systems, Man and Cybernetics*, SMC-15(4), pp. 580–589.

Keysers, D. *et al.* (2017) 'Multi-Language Online Handwriting Recognition', *IEEE Transaction on Pattren Analysis and Machine Intelligence*, 39(6), pp. 1180–1194.

Khemlani, S. S. and Goodwin, G. P. (2015) 'Logic , probability , and human reasoning', *Trends in Cognitive Sciences*, 19(4), pp. 201–203.

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

Kokar, M. M., Matheus, C. J. and Baclawski, K. (2009) 'Ontology-based situation awareness', *Information Fusion*. Elsevier B.V., 10(1), pp. 83–98.

Kokkinos, I. and Maragos, P. (2009) 'Synergy between Object Recognition and Image Segmentation Using the Expectation-Maximization Algorithm', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8), pp. 1486–1501.

Kopparapu, S. K. and Desai, U. B. (2002) *Bayesian Approach to Image Interpretation*. Springer Science & Business Media.

Kornyshova, E. and Rebecca, D. (2010) 'Decision-making ontology for information system engineering', *Lecture Notes in Computer Science*, 6412, pp. 104–117.

Kotsiantis, S. B. (2007) ‘Supervised Machine Learning: A Review of Classification Techniques’, *Informatica*, 31, pp. 249–268.

Kotsiantis, S. B., Zaharakis, I. D. and Pintelas, P. E. (2006) ‘Machine learning: A review of classification and combining techniques’, *Artificial Intelligence Review*, 26(3), pp. 159–190.

Kumar, V. F. and Desai, U. B. (1996) ‘Image Interpretation Using Bayesian Networks’, *IEEE transactions on pattern analysis and machine intelligence*, 18(1), pp. 74–77.

Lafferty, J., McCallum, A. and Pereira, F. C. N. (2001) ‘Conditional random fields: Probabilistic models for segmenting and labeling sequence data’, *ICML ’01 Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988) ‘Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems’, *Journal of the Royal Statistical Society*, 50(2), pp. 157–224.

Laviola, J. J. and Zeleznik, R. C. (2004) ‘MathPad 2 : A System for the Creation and Exploration of Mathematical Sketches’, *ACM Transactions on Graphics*, 23(3), pp. 432–440.

Laxhammar, R. (2007) *Artificial Intelligence for Situation Assessment, MSc thesis*. Royal Institute of Technology.

LeCun, Y. *et al.* (1998) ‘Gradient-Based Learning Applied to Document Recognition’, in *Proceedings of the IEEE*, pp. 2278–2324.

Leposavić, I., Leposavić, L. and Šaula-Marojević, B. (2010) ‘Neuropsychological assessment: Computerized batteries or standard tests’, *Psychiatria Danubina*, 22(2), pp. 149–152.

Li, C., Cao, M. and Tian, L. (2015) ‘Situation Assessment Approach Based on a Hierarchic Multi-timescale Bayesian Network’, in *2nd International Conference on Information Science and Control Engineering*, pp. 911–915.

Li, L., Zhang, L. L. and Su, J. F. (2012) ‘Handwritten character recognition via direction string and nearest neighbor matching’, *Journal of China Universities of Posts and Telecommunications*, 19(SUPPL. 2), p. 160–165,196.

Li, Y., Song, Y.-Z. and Gong, S. (2013) ‘Sketch Recognition by Ensemble Matching of Structured Features’, in *BMCV*.

Liang, Y. *et al.* (2010) ‘A learning model for the automated assessment of hand-drawn images for visuo-spatial neglect rehabilitation.’, *IEEE transactions on neural systems and rehabilitation engineering*, 18(5), pp. 560–70.

Lichman, M. (2013) *UCI Machine Learning Repository*, *University of California, Irvine, School of Information and Computer Sciences*. Available at: http://archive.ics.uci.edu/ml.

Liu, C. L. *et al.* (2003) ‘Handwritten digit recognition: Benchmarking of state-of-the-art techniques’, *Pattern Recognition*, 36(10), pp. 2271–2285.

Liwicki, M. and Bunke, H. (2007) ‘Combining On-line and Off-line Systems for Handwriting Recognition’, in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 372–376.

Lu, Y. I. and Shridhar, M. (1996) ‘Character Segmentation in Handwriting words-An Overview’, *Pattren Recognition*, 29(1), pp. 77–96.

Luciano, J. S. *et al.* (2011) ‘The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside.’, *Journal of biomedical semantics*, 2(2), pp. 1–21.

Luo, J., Savakis, A. E. and Singhal, A. (2005) ‘A Bayesian network-based framework for semantic image understanding’, *Pattern Recognition 38*, 38(6), pp. 919–934.

Lv, Y. F. *et al.* (2013) ‘Learning-based candidate segmentation scoring for real-time recognition of online overlaid chinese handwriting’, in *Proceedings of the 12 International Conference on Document Analysis and Recognition, ICDAR*, pp. 74–78.

Mamdani, E. H. (1974) ‘Application of fuzzy algorithms for control of simple dynamic plant’, in *Proceedings of the Institution of Electrical Engineers IEE*, p. 1585. doi: 10.1049/piee.1974.0328.

Margaritis, D. *et al.* (2003) *Learning Bayesian Network Model Structure from Data, PhD thesis*, *Learning*. Carnegie Mellon University.

Mazancova, A. F. *et al.* (2016) ‘The Reliability of Clock Drawing Test Scoring Systems Modeled on the Normative Data in Healthy Aging and Nonamnestic Mild Cognitive Impairment’, *Assessment SAGE*, pp. 1–13.

Mehul, G. *et al.* (2014) ‘Text-Based Image Segmentation Methodology’, in *2nd International Conference on Innovations in Automation and Mechatronics Engineering, ICIAME*. Elsevier B.V., pp. 465–472.

Mendes-santos, L. C. *et al.* (2015) ‘Specific algorithm method of scoring the Clock Drawing Test applied in cognitively normal elderly’, *Dement Neuropsychol*, 9(2), pp. 128–135.

Mendez, M. F., Ala, T. and Underwood, K. L. (1992) ‘Development of Scoring Criteria for the Clock Drawing Task in Alzheimer’s Disease’, *Journal of the American Geriatrics Society*, 11(40), pp. 1095–1099.

Mitchell, T. (1979) *Machine learning*, *IJCAI International Joint Conference on Artificial Intelligence*. McGraw Hill.

Monique, T. (2002) ‘Knoweldge-based Techniques for Image Processing and for Image Understanding’, *Journal de Physique*, 4(12), pp. 189–235.

Murphy, K. P. (2001) ‘The Bayes Net Toolbox for Matlab’, *Computing Science and Statistics*, 33(2), pp. 1024–1034.

Namboodiri, A. M. and Jain, A. K. (2004) ‘Online Handwritten Script Recognition’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), pp. 124–130.

Neapolitan, R. E. (1990) *Probabilistic reasoning in expert systems: theory and algorithms*. John Wiley & Sons, Inc. New York, NY, USA.

Needham, C. J. *et al.* (2007) ‘A primer on learning in Bayesian networks for computational biology’, *PLoS Computational Biology*, 3(8), pp. 1409–1416.

Nikolopoulos, S. *et al.* (2011) ‘Evidence-driven image interpretation by combining implicit and explicit knowledge in a bayesian network’, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(5), pp. 1366–1381.

Ogiela, L. (2008) ‘Modelling of cognitive processes for computer image interpretation’, in *Proceedings - EMS 2008, European Modelling Symposium, 2nd UKSim European Symposium on Computer Modelling and Simulation*, pp. 209–213.

Olsen, L. *et al.* (2009) ‘Sketch-based modeling: A survey’, *Computers and Graphics*, 33(1), pp. 85–103.

Onisko, A., Lucas, P. and Druzdzel, M. J. (2001) ‘Comparison of Rule-Based and Bayesian Network Approaches in Medical Diagnostic Systems’, in *8th Conference on Artificial Intelligence in Medicine in Europe, AIME*, pp. 283–292.

Ouyang, T. Y. and Davis, R. (2009) ‘A Visual Approach to Sketched Symbol Recognition’, in *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 1463–1468.

Papavassiliou, V. *et al.* (2010) ‘Handwritten document image segmentation into text lines and words’, *Pattern Recognition*. Elsevier, 43(1), pp. 369–377.

Parsey, C. M. and Schmitter-Edgecombe, M. (2011) ‘Quantitative and qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: evaluation of a modified scoring system.’, *Journal of geriatric psychiatry and neurology*, 24(2), pp. 108–18.

Pearl, J. (1988) *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, california: Morgan Kaufmann.

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(3), pp. 201–213.

Pinto, J. and Hermansky, H. (2010) 'Analysis of MLP Based Hierarchical Phoneme Posterior Probability Estimator', *IEEE Audio, Speech and Language Processing*, 19(2), pp. 225–241.

Plamondon, R., Srihari, S. N. and S., S. (2000) 'On-Line and Off-Line Handwriting Recognition : A Comprehensive Survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), pp. 63–84.

Pomerol, J.-C. (1997) 'Artificial intelligence and human decision making', *European Journal of Operational Research*, 2217(96), pp. 1–28.

Potter, J. *et al.* (2000) 'Computer recording of standard tests of visual neglect in stroke patients', *Clinical Rehabilitation*, 14(4), pp. 441–446.

Price, C. C. *et al.* (2011) 'Clock drawing in the montreal cognitive assessment: Recommendations for dementia assessment', *Dementia and Geriatric Cognitive Disorders*, 31, pp. 179–187.

Ribas, F. C. *et al.* (2013) 'Handwritten digit segmentation: A comparative study', *International Journal on Document Analysis and Recognition*, 16, pp. 127–137.

Ricci, M. *et al.* (2016) 'The clock drawing test as a screening tool in mild cognitive impairment and very mild dementia: a new brief method of scoring and normative data in the elderly', *Neurological Sciences*, 37(6), pp. 867–873.

Rice, V. J. *et al.* (2011) 'Automated Neuropsychological Assessment Metrics (ANAM) Traumatic Brain Injury (TBI): Human Factors Assessment', *Army Research Laboratory*, (July), pp. 1–42.

Rindflesch, T. C. and Fiszman, M. (2003) 'The interaction of domain knowledge and linguistic structure in natural language processing: Interpreting hypernymic propositions in biomedical text', *Journal of Biomedical Informatics*, 36(6), pp. 462–477.

Robbins, T. W. and Sahakian, B. J. (1988) 'Computerized Psychological Assessment: A Practitionerś Guide', *Principles and Practice of Geriatric Psychiatry*, (33), pp. 147–151.

Rohitha, K. K. *et al.* (2007) ‘Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections’, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(6), pp. 1446–1459.

Rosenblum, S. *et al.* (2013) ‘Handwriting as an objective tool for Parkinson’s disease diagnosis.’, *Journal of neurology*, 260(9), pp. 2357–61.

Rosenblum, S. and Livneh-Zirinski, M. (2008) ‘Handwriting process and product characteristics of children diagnosed with developmental coordination disorder’, *Human Movement Science*, 27(2), pp. 200–214.

Roubos, J. A., Setnes, M. and Abonyi, J. (2003) ‘Learning fuzzy classification rules from labeled data’, *Information Sciences*, 150(1–2), pp. 77–93.

Rouleau, I. *et al.* (1992) ‘Quantitative and qualitative analyses of clock drawings in Alzheimer’s and Huntington’s disease’, *Brain and Cognition*, 18(1), pp. 70–87.

Roy, J. (2001) ‘From Data Fusion to Situation Analysis’, in *Proceedings of the 4th International Conference on Information Fusion*, pp. 7–10.

Sarvadevabhatla, R. K. and Babu, R. V. (2015) ‘Freehand Sketch Recognition Using Deep Features’, *arXiv preprint arXiv:1502.00254*, pp. 1–6.

Schmand, B. *et al.* (2000) ‘Early detection of Alzheimer ’ s disease using the Cambridge Cognitive Examination (CAMCOG)’, *Psychological Medicine*, 30(3), pp. 619–627.

Schmidhuber, J. (2015) ‘Deep learning in neural networks: An overview’, *Neural Networks*, 61, pp. 85–117.

Shi, D. *et al.* (2016) ‘Integrating Digital Pen Devices with Traditional Medical Screening Tool for Cognitive Assessment’, in *Information Science and Technology (ICIST), Sixth International Conference on IEEE.*, pp. 42–48.

Shi, Z. *et al.* (1997) ‘A system for segmentation and recognition of totally unconstrained handwritten numeral strings’, in *Proceedings of the Fourth*

International Conference on Document Analysis and Recognition ICDAR, pp. 455–458.

Shih, F. H. (2010) *Image Processing and Pattern Recognition Fundemental and techniques*. a John Willey & Son Publication,Inc., Hoboken, New Jersey.

Shulman, K. I. (2000) 'CLOCK -DRAWING : IS IT THE IDEAL COGNITIVE SCREENING TEST ?', *International Journal of Geriatric Psychiatry*, 15, pp. 548–561.

Shulman, K. I. and Feinstein, A. (2005) 'Quick Cognitive Screening for Clinicians', *International Psychogeriatrics*, 17(1), pp. 129–130.

Shulman, K. I., Shedletsky, R. and Silver, I. L. (1986) 'The challenge of time: Clock-drawing and cognitive function in the elderly', *International Journal of Geriatric Psychiatry*, 1(2), pp. 135–140.

Somvanshi, M. and Chavan, P. (2016) 'A review of machine learning techniques using decision tree and support vector machine', in *International Conference on Computing Communication Control and automation (ICCUBEA)*, pp. 1–7.

Song, Y. *et al.* (2016) 'Balancing Appearance and Context in Sketch Interpretation', *arXiv preprint arXiv:1604.07429.*, pp. 2625–2632.

Sonka, M., Hlavac, V. and Boyle, R. (1999) 'Image processing, analysis, and machine vision second edition', (January).

Souillard-mandar, W. *et al.* (2016) 'Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test', *Machine learning*, 102(3), pp. 393–441.

Spenciere, B., Alves, H. and Charchat-fichman, H. (2017) 'Scoring systems for the Clock Drawing Test a historical review', *Dement Neuropsychol*, 11(1), pp. 6–14.

Srivastava, N. *et al.* (2014) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *Journal of Machine Learning Research*, 15, pp. 1929–1958.

Stafylakis, T. *et al.* (2008) ‘Robust text-line and word segmentation for handwritten documents images’, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 3393–3396.

Stamatopoulos, N. *et al.* (2013) ‘ICDAR 2013 handwriting segmentation contest’, in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 1402–1406.

Sugeno, M. (1985) *Industrial applications of fuzzy control, Elsevier Science Pub. Co.* Elsevier Science Pub.

Sun, Y. *et al.* (2004) ‘Segmenting handwritten text using supervised classification techniques’, in *Proc. of IEEE International Joint Conference on Neural Networks*, pp. 657–662.

Tadeusiewicz, R. and Ogiela, M. R. (2004) *Medical Image Understanding Technology*. Springer, Heidelberg.

Tagougui, N., Kherallah, M. and Alimi, A. M. (2013) ‘Online Arabic handwriting recognition: A survey’, *International Journal on Document Analysis and Recognition*, 16, pp. 209–226.

Tanaka, V. H., Iwayama, V. N. and Akiyama, V. K. (2004) ‘Online Handwriting Recognition Technology and Its Applications’, *Fujitsu Science Technology Journal*, 40(1), pp. 170–178.

Tappert, C. C., Suen, C. Y. and Wakahara, T. (1990) ‘The state of the art in online handwriting recognition’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(8), pp. 787–808.

Tornatore, J. B. *et al.* (2005) ‘Self-Administered Screening for Mild Cognitive Impairment: Initial Validation of a Computerized Test Battery’, *Biophysical Chemistry*, 257(5), pp. 2432–2437.

Tuokko, H. *et al.* (2000) ‘A comparison of alternative approaches to the scoring of clock drawing.’, *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 15(2), pp. 137–48.

Van, T. B. T. and Rene, H. (1999) ‘Procedural and Declarative Knowledge’, *Sage Publications*, 9(5), pp. 605–624.

Velikova, M. *et al.* (2013) ‘Artificial Intelligence in Medicine On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks’, *Artificial Intelligence In Medicine*. Elsevier B.V., 57(1), pp. 73–86.

Vyhnálek, M. *et al.* (2016) ‘Clock drawing test in screening for Alzheimer ’ s dementia and mild cognitive impairment in clinical practice’, *International journal of Geriatric Psychiatry*.

Werner, P. *et al.* (2006) ‘Mild Alzheimer ’ s Disease and Mild Cognitive Impairment’, *Journal of Gerontology: Psychological Science*, 61B(4), pp. 228–236.

Wild, K. *et al.* (2008) ‘The status of computerised cognitive testing in aging: A systematic review’, *Alzheimers Dementia*, 4(6), pp. 428–437.

Wright, E. *et al.* (2002) ‘Multi-entity Bayesian networks for situation assessment’, in *Proceedings of the 5th International Conference on Information Fusion*, pp. 804–811.

Wu, Y. (2013) ‘An Introduction to Computer Vision What Is This Area About ?’, *EECS 432-Advanced Computer Vision Notes Series*, pp. 1–7.

Xu, L. *et al.* (1992) ‘Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition’, *IEEE Transactions on Systems, Man and Cybernetics*, 22(3), pp. 418–435.

Yu, Q. *et al.* (2015) ‘Sketch-a-Net that Beats Humans’, in *British Machine Vision Conference (BMVC)*, pp. 1–11.

Zadeh, L. (1992) ‘Knowledge Presentation In Fuzzy Logic’, in *An Introduction to Fuzzy Logic Applications in Intelligent Systems*. New York: Springer Science and Business Media.

Zhou, S. M. *et al.* (2012) ‘Constructing Compact Takagi-Sugeno Rule Systems: Identification of Complex Interactions in Epidemiological Data’, *PLoS ONE*, 7(12).

Zunjarrao, T. (2017) ‘Recognition of Online Handwritten Characters Using LIPI Toolkit’, *International Journal of Advance Research, Ideas and Innovations in Technology.*, 3, pp. 689–692.

Zuo, Y. and Kita, E. (2012) ‘Stock price forecast using Bayesian network’, *Expert Systems with Applications*. Elsevier Ltd, 39(8), pp. 6729–6737.

Zurada, J. M. (1992) *Introduction to Artificial Neural Systems*. St. Paul, USA: West Publishing.