DEVELOPMENT OF A NOVEL VIRTUAL ENVIRONMENT FOR ASSESSING COGNITIVE FUNCTION

Design, Development and Evaluation of a Novel Virtual Environment
to Investigate Cognitive Function and Discriminate between
Mild Cognitive Impairment and Healthy Elderly

Syadiah Nor WAN SHAMSUDDIN

Submitted for the Degree of

Doctor of Philosophy

School of Computing, Informatics and Media

University of Bradford

2012

Abstract

Alzheimer's disease (AD) is neurodegenerative disorder that causes memory loss and cognitive dysfunction. It affects one in five people over the age of 80 and is distressing for both sufferers and their families. A transitional stage between normal ageing and dementia including AD is termed a mild cognitive impairment (MCI). Recent studies have shown that people with MCI may convert to AD over time although not all MCI cases progress to AD. Much research is now focussing on early detection of AD and diagnosing an MCI that will progress to AD to allow prompt treatment and disease management before the neurons degenerate to a stage beyond repair. Hence, the ability to obtain a method of identifying MCI is of great importance.

Virtual reality plays an important role in healthcare and offers opportunities for detection of MCI. There are various studies that have focused on detection of early AD using virtual environments, although results remain limited. One significant drawback of these studies has been their limited capacity to incorporate levels of difficulty to challenge users' capability. Furthermore, at best, these studies have only been able to discriminate between early AD and healthy elderly with about 80% of overall accuracy.

As a result, a novel virtual simulation called Virtual Reality for Early Detection of Alzheimer's Disease (VREAD) was developed. VREAD is a quick, easy and friendly tool that aims to investigate cognitive functioning in a group of healthy elderly participants and those with MCI. It focuses on the task of following a route, since Topographical Disorientation (TD) is common in AD. An investigation was set up with two cohorts: non-elderly and elderly participants. The findings with regard to the non-elderly are important as they represent a first step towards implementation with elderly people. The results with elderly participants indicate that this simulation based assessment could provide a method for the detection of MCI since significant correlations between the virtual simulation and existing neuropsychological tests were found. In addition, the results proved that VREAD is comparable with well-known neuropsychological tests, such as Cambridge Neuropsychological Automated Test Battery, Paired Associate Learning (CANTAB PAL) and Graded Naming Test (GNT). Furthermore, analysis through the use of machine learning techniques with regard to the prediction of MCI also obtained encouraging results. This novel simulation was able to predict with about 90% overall accuracy using weighting function proposed to discriminate between MCI and healthy elderly.

Keywords :

Virtual Environment, Spatial Memory, Topographical Disorientation, Mild Cognitive Impairment.

Acknowledgements

First, I would like to extend my utmost gratitude to my main supervisor, Prof Hassan Ugail for his guidance, attention and support. My deepest and sincere appreciation goes to my co-supervisor, my mentor, Dr Valerie Lesk, for her encouragement, guidance, support and understanding throughout this study. I feel privileged having two knowledgeable and dedicated supervisors for guiding me without fail, assisting and pointing me in the right direction.

I would not have been able to complete a PhD without the support from my family and friends. My sincere gratitude goes to my beloved husband, Mokhairi Makhtar, for his love, support, encouragement and always standing by me. To my beautiful children; Syahmi, Syafiq, Syazwan and Syasya, thank you for your love, understanding, patience and inspiration on this wonderful journey. My foremost gratitude also goes to my parents for their support, prayers and encouragement.

Thank you to all my wonderful friends, in Malaysia and Bradford, UK especially and staff who have been supporting me for their contributions to many moments of insight, inspiration, laughter, and support throughout my completion of this study. My special appreciation goes to Elizabeth Walters for her helped in collecting samples, commitment, support and inputs. To Helen Jones, thank you for proofreading my thesis. To staff of Psychology Division, University of Bradford, thank you very much. Finally, I would like to thank all the volunteer participants who gave their time and effort for this research and made this thesis possible. Last but not least, thank you to Ministry of Higher Education, Malaysia and University Sultan Zainal Abidin, Malaysia (UNisZa) for the financial supports received and University of Bradford for giving me the opportunity to continue my study.

All of my thanks go to God for enabling me to accomplish this and giving me the support I need on a daily basis.

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

Parts of the original work proposed in this thesis have appeared in the publications as listed in publications section.

Publications

- Lesk, V., Wan Shamsuddin, S. N., Walters, E. & Ugail, H. Mac 2012. Using A Virtual Environment To Assess Cognition In The Elderly. *Alzheimer's & Dementia: Journal of the Alzheimer's Association.* (Submitted).

- Wan Shamsuddin, S. N., Ugail, H. & Lesk, V. 2011. Development and Usability Evaluation of Virtual Environment for Early Diagnosis of Dementia, *Proceedings of the Second International Conference on Visual Informatics: Sustaining Research and Innovations, 9 – 11 November, Malaysia: Springer.* 13-22.

- Wan Shamsuddin, S. N., Walters, E., Ugail, H. & Lesk, V. 2011. Evaluation of Users Acceptance in Virtual Environment for Early Diagnosis of Dementia. *Paper presented at 2nd World Conference on Information Technology.* (WCIT), November 23-26. Antalya Turkey.

- Wan Shamsuddin, S. N., Lesk, V. & Ugail, H. 2011. Virtual Environment Design Guidelines for Elderly People in Early Detection of Dementia. *Proceedings of the International Conference of Computer and Information Science 2011, November 14-16, Paris, France.* 751-755.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
ADDIE	Analysis, Design, Development, Implementation and Evaluation
CANTAB PAL	Cambridge Neuropsychological Automated Test Battery, Paired Associate Learning
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GNT	Graded Naming Test
KDD	Knowledge Discovery Dataset
MCI	Mild Cognitive Impairment
MMSE	Mini Mental State Examination
TD	Topographical Disorientation
TN	True Negative
TNR	True Negative rate
TP	True Positive
TPR	True Positive rate
VE	Virtual Environment
VR	Virtual Technology
VREAD	Virtual Reality for Early Detection of Alzheimer's Disease
WEKA	Waikato Environment for Knowledge Analysis

Chapter 1

Introduction

The first chapter provides readers with an overview of the research and addresses the significance of the contribution made by this research. In addition, the research aims and objectives, as well as its motivations, are presented. Furthermore, the research methodology section explains the procedures utilized and emphasizes ethical considerations. This chapter concludes with an outline of the thesis and provides a summary.

This thesis explores the efficacy of a newly developed virtual environment as a novel test of spatial memory for investigating cognition in the elderly. The overarching aim is to use this test as a method for the early diagnosis of Alzheimer's disease (AD). The specific design criteria for the development of this simulation - which is called Virtual Reality for Early Detection of Alzheimer's Disease or VREAD - are essential for motivating elderly people to take part in this simulation. Throughout the thesis, data collected from virtual simulations and statistical analysis are described. Further analysis using data mining techniques for the prediction of mild cognitive impairment (MCI) and identifying the healthy elderly also gave positive results. Finally, the results for the usability of users' acceptance and satisfaction of the test among the non-elderly and elderly will be discussed.

1.1 Background

There are currently 820,000 people living with dementia in the United Kingdom alone (Alzheimer's Association, 2011). The National Health Service uses the term dementia to describe the symptoms that occur when the brain is affected by a specific disease and conditions. However, an effective treatment or cure for dementia remains elusive and its incidence will continue to increase (Singh-Manoux and

Kivimäki, 2010). Alzheimer's disease (AD) is the most common type of dementia (Hettinga et al., 2009) and the most well-known symptom of AD is memory loss (Hettinga et al., 2009).

MCI is widely believed to be a transitional stage between normal ageing and dementia (Petersen, 2011). Dubois et al. (2007) have described in their paper that imaging studies of MCI patients reveal an elevated volumetric decline in the entorhinal cortex of the hippocampus. The entorhinal cortex is a major gateway that communicates information to and from the hippocampus. MCI, along with images from magnetic resonance imaging scans of healthy elderly and MCI patients will be discussed further in Chapter 2. According to Petersen (2011), MCI has an estimated prevalence of between 10% to 20% of individuals over the age of 65 in population-based studies. Several studies have shown that some cases of MCI progress to AD, reaching as high as 15% (Decarli, 2003). A study by Busse et al. (2003) reported that 23% to 47% of MCI converted to dementia over 2.6 years. In up to six years, approximately 80% of the 220 patients in Mayo Alzheimer's Disease Research Center will have converted to dementia (Petersen, 2004). Another study by Blackwell et al. (2004) showed that psychological tests could detect cognitive dysfunction characteristics among subjects.

Although studies have recognized the subsequent stage of AD, some studies show that not all people diagnosed as MCI will go on to develop AD (Petersen et al., 1999). In a follow-up study, Larrieu et al. (2002) demonstrated that more than 40% of her subjects reverted to normal within two to three years. Another study by Ganguli et al. (2004) revealed that 33% to 55.6% of their subjects had reverted to normal.

The ability to obtain a method of identifying MCI is therefore of great importance. It is also essential that techniques are established to discriminate between those with an MCI that will progress to a dementia and those who will not. Early diagnosis is necessary before the neurons degenerate to a stage beyond repair. One importance of early diagnosis is to allow people with MCI to plan ahead and make important decisions about their future. Early diagnosis allows prompt treatment before patients suffer potential cognitive problems (Weimer and Sager, 2009; Holt et al., 2009). Another difficulty is that diagnosis often occurs at a late stage, when structural damage to the hippocampus and surrounding areas has already reached too great an extent (Pantel et al., 2004).

Various studies have focused on the detection of early AD and this area is progressively gaining more attention. In recent studies, neuroimaging, neuropsychological tests and laboratory tests have been carried out to establish a diagnosis of AD (Liu et al., 2011). Neuropsychological tests can provide evidence of memory impairment (Maruta et al., 2011; Karageorgiou et al., 2011). On the other hand, structural MRI can offer evidence of atrophy in the structures related to AD (Frisoni et al., 2010). However, there is no standard protocol in the early diagnosis of AD (Mistur et al., 2009), which is still difficult to establish (Liu et al., 2011). The Mini Mental State Examination (MMSE), a type of neuropsychological test that is widely used, has previously been criticized for its inability to detect mild changes in cognition (De Jager et al., 2002; Pasquier, 1999).

1.1.1 Topographical Memory and AD

Topographical disorientation (TD) is the reduced ability to find one's way and navigate in familiar and unfamiliar environments towards a specific destination (Barrash, 1998). TD is very common in AD and dementia (Cherrier et al., 2001). As the disease progresses, AD patients are likely to become lost even in familiar places and to require help from others to locate their target destination. In order to enable this population to continue participating in social and leisure activities in their environment, issues with orientation, navigation and well-being need to be identified (Blackman et al., 2003). Cherrier et al. (2001) investigated performance measurement of TD in AD patients. In their study they demonstrated that AD patients show poor performance compared to control groups on route learning tests due to the characteristics of poor spatial orientation or spatial reasoning (Cherrier et al., 2001)

Clear navigational deficits were seen in this patient cohort, leading to sufferers getting lost when out and about, wandering and taking the wrong direction on car journeys (Silverstein et al., 2006). This has potentially devastating consequences and causes significant worry for the carers and families involved. Few well-established tests for assessing navigational deficits early on in the disease exist. Duffy (2009) suggests that visual information processing is impaired in AD and a number of studies have shown that both AD and MCI populations have problems in terms of visual motion processing, which is thought to underlie the navigational problems (Mapstone et al., 2003; Tetewsky and Duffy, 1999; Duffy, 2009; O'brien et al., 2001; Dubinsky et al., 2000; Uc et al., 2004). The testing of navigational deficits using 'real-world' methods is both difficult and time-consuming. As discrimination between healthy elderly and AD is problematic, more tests are needed to detect or screen between these subgroups.

1.2 Motivation

In recent years, computer technology has evolved and the use of virtual environments has increased to simulate real-world tasks (Tippett et al., 2009). Virtual Reality (VR) has been recognized as a technology that provides an effective and motivating way to help in many different areas. VR is a technology with many applications, from the arts to healthcare. In the latter area it has been used in fields such as phobias with regard to spiders (Garcia-Palacios et al., 2002), fear of flying (Banos et al., 2002), acrophobia (Coelho et al., 2009), fear of public speaking (Slater et al., 2006) and motor rehabilitation (Holden, 2005). VR provides valuable daily life scenarios with familiar elements for patients, thus extending the range of cognitive rehabilitation applications (Marques et al., 2008). With VR technology, a virtual environment lets the user freely explore a 3D space.

A virtual environment is advantageous because the experimenter has control over the stimuli and complexity of the environment, allowing him or her to observe the patient in detail. Similarly, a computerized system allows for the accurate recording of data, including the number of errors made, without any reliance on subjective assessments and note taking on the part of the researcher.

Previously, Werner et al. (2009) developed a Virtual Supermarket to examine the feasibility and the validity of a virtual action planning supermarket (VAP-S) for the diagnosis of patients with MCI. Another study by Pengas et al. (2010) also used a virtual environment to discriminate between early AD patients and controls. Tippett et al. (2009) examined behavioural performance across different populations, using a virtual environment. A recent study by Cushman et al. (2008) investigated the efficacy of a virtual reality environment to assess navigational performance in people with early AD and MCI, along with young and old normal controls. Chapter 2 provides a further review of virtual environments for early detection of MCI.

Despite the many research efforts addressing the use of virtual environments for the early detection of dementia, these examples are limited to navigation tasks that would challenge users' motivation. In addition, the studies lack specific criteria, such as the design and development of a virtual environment for people with dementia, which need to be considered. Furthermore, the previous studies have only been able to discriminate between early AD and healthy elderly with about 80% of overall accuracy.

1.3 Research Aims and Objectives

The aim of this research is to develop a new virtual environment for spatial memory with the purpose of discriminating between participants with MCI and the healthy elderly. The aim can be achieved by following these objectives:-

1. To develop a novel virtual environment as a test for spatial memory to assess cognitive functions.
2. To implement and evaluate a newly developed virtual environment for non-elderly and elderly people.

3. To diagnose MCI at an early stage with data collected from the virtual simulation using machine learning techniques.
4. To propose design criteria specifically for elderly people to help with the virtual environment process.
5. To compare performance on the virtual simulation with existing neuropsychological tests used in AD research.

1.4 Research Methodology

The aims and objectives of this research can be accomplished by implementing a structured research methodology. This research was carried out by following the research methodology depicted in Figure 1.1, which shows that this study begins with a statement of the aims and objectives. A thorough literature review relating to the domain was conducted, leading to the production of design criteria for the development of a virtual environment for elderly people with AD (Wan Shamsuddin et al., 2011b). The primary goal of these specific criteria is to promote accessibility because the effectiveness of the technology relies on the ability of research participants to use it.

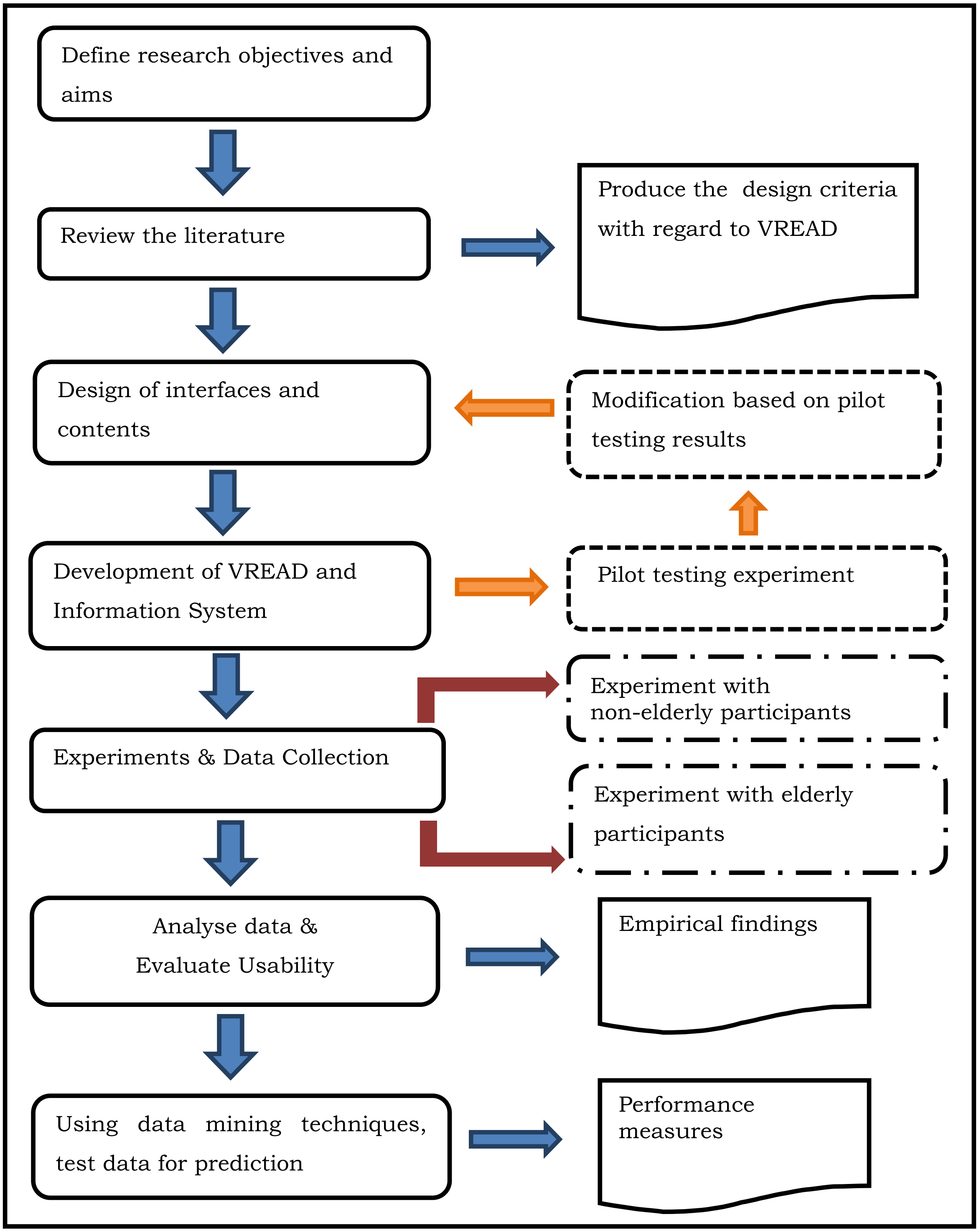


Figure 1.1: Research methodology for this research study

The next phase involves the design of interfaces and contents. A storyboard of interfaces and contents was first produced to generate some ideas for the simulation. Then, VREAD was developed on the basis of the proposed design critera. Once the simulation was complete, it underwent pilot testing for experiment and evaluation. The results from the pilot test revealed the need for some modifications to the simulation, which are discussed in Chapter 3.

The focus then moved to the VREAD implementation in terms of experiments and data collection with non-elderly and elderly participants. Chapter 4 describes in detail the experiments that were conducted and the methods of data collection. The experimental setup and methods used for the investigation of cognition in spatial memory are explained, along with the detailed procedures. Two experiments involving 23 non-elderly and 31 elderly participants were conducted.

The next phase involved the analysis of collected data as well as an evaluation of its usability. Once the experiments were completed, the data was analysed to find any significant correlations between attributes and established neuropsychological tests. This aspect is discussed in Chapter 4. The evaluation of usability, in terms of user acceptance and satisfaction, was undertaken in conjunction with non-elderly and elderly participants. The findings with regard to non-elderly participants comprise an important first step, prior to the implementation with elderly people. This allowed (i) the verification of the feasibility of the environment in a group with no brain impairment, (ii) assessment of the time taken to run the test and (iii) the acquisition of normative data for comparison with elderly participants. It also facilitated assessment of the memory mechanisms utilized in VREAD, which are harder to assess in a group of participants with cognitive impairment or even healthy ageing, as neurones degenerate even with healthy ageing.

Further analysis of the data was conducted through the knowledge discovery process in order to predict differences between MCI and healthy elderly people, as discussed in Chapter 6. Five different algorithms, including Neural Networks, Decision Trees, K-Nearest Neighbour, Ensemble and Bayesian Networks, were used in this experiment to provide the highest performance measures in terms of prediction. Using WEKA package software, a comparative study of different selections and classifications was undertaken. In addition, in Chapter 6, the Weighting functions were highlighted to improve performance measures in terms of prediction using data mining techniques.

1.5 Research Contributions

The seven individual research contributions of the thesis are as follows:-

1. A novel VREAD simulation to discriminate between healthy elderly and MCI patients (Chapter 3).
2. The design criteria for the development of the VR simulation for spatial memory (Chapter 3).
3. The production of a dataset, which can be used for comparison by other researchers in the future (Chapter 4).
4. The accuracy rate for discrimination as a result of data mining techniques could be used as a benchmark by other researchers in future (Chapter 6).
5. Empirical findings in this study could be usefully explored in further research (Chapter 4, 5 and 6).
6. A new measurement to test ways of assessing performance in spatial memory (Chapter 3).
7. A methodology to discriminate between MCI patients and the healthy elderly using the VREAD simulation (Chapter 1).

1.6 Ethical Consideration

The participants involved in this research are non-elderly and elderly people. For non-elderly participants, a total of 23 participants consisting of 10 males and 13 females below the age of 65 were recruited for this experiment. The participants were students or staff of the University of Bradford who were following different courses.

In terms of the elderly participants, the group consisted of 31 individuals of whom 10 were males and 21 females over the age of 65. All elderly participants were recruited from the University of Bradford Division of Psychology healthy elderly participant pool. All participants had no existing neurological disorder or existing memory impairment. Furthermore, none of the participants had an uncorrectable visual impairment. The study was approved by the University of Bradford Ethics Committee. The participants were asked to sign a consent form prior to the experimental procedure (see Appendix A).

Sample sizes are important because it affects the results of the research. Choosing a small sample size may not give accurate results and in contrast, choosing a large sample size may be wasteful and sometimes impossible to complete.

A statistical significant approach is appropriate for the calculation of sample sizes (Whitley and Ball, 2002). The term significant with p-value of less than 0.05 is used merely as a way of indicating the chances are at least 95 out of 100 that the findings obtained from the sample regardless of the size (Gupta et al., 2012). Furthermore, the sample size in this study is assumed to be sufficient as proposed by Roscoe (1975) and were also used for the studies by Hox, 1998 and Kasiulevičius et al., 2006

For this research study, there was a significant correlation between VREAD level 4 scores on both PAL six pattern stage and GNT with $p<0.02$ respectively, and the results do discriminate between the two subgroups. Thus, it indicates the sample size for this research study is sufficient.

1.7 Thesis Organisation

This thesis is divided into seven chapters.

Chapter 2 is about related work in the form of a literature review of relevant studies and fundamental knowledge of the subject matter.

Definitions and some examples of the techniques used in the study are also discussed in this chapter.

Chapter 3 discusses the proposed design criteria for the development of the VREAD simulation. In addition, it also discusses the step-by-step approach of the development process, with an emphasis on hardware and software.

In **Chapter 4**, a thorough analysis of the data collected was undertaken. The aim was to determine whether VREAD could be used as a novel test of spatial memory and its correlation with established neuropsychological tests of cognitive function is explained. The interpretations of the significant findings are reviewed here.

Chapter 5 explains in detail the levels of acceptance and satisfaction of the tests by non-elderly and elderly participants.

Chapter 6 demonstrates further analysis through the use of data mining techniques with regard to the prediction of MCI and healthy elderly. This chapter also discusses the significance of the findings of the study.

Lastly, **Chapter 7** concludes the thesis and summarises how the research method used have led to the research objectives being achieved. The chapter also identifies the significant contribution of this research to this knowledge domain. In order to improve research in the future, some ideas for further work are proposed.

1.8 Summary

This chapter explains the overall framework of the research in order to offer readers a quick overview. Background information is provided by an explanation of the aims and objectives of the research and the research methodology is presented. Details of each phase of the methodology are explained in subsequent chapters. Finally, significant contributions and an outline of the thesis are given.

Chapter 2

Literature Review

2.1 Introduction

The purpose of this chapter is to give the reader an overview of the knowledge relevant to this thesis. It contains a literature review of virtual reality and its application to healthcare. Furthermore, comparison and summarisation of other related works are briefly explained. The chapter starts with an overview of Alzheimer's disease, MCI, spatial memory and topographical disorientation. The chapter concludes with knowledge discovery in databases and machine learning algorithms used in the experimental work.

2.2 Alzheimer's Disease

Alzheimer's disease is a neurological disorder that destroys memory and thinking skills. The Alzheimer's Association (2011) defines AD as a deficit in the formation of new memories, difficulty remembering events and names, apathy and depression. AD is the most common form of dementia, which is associated with a decline in mental illness (Small and Cappai, 2006). Currently, AD affects one in five people over the age of 80 (Alzheimer's Association, 2011). AD is named after Dr. Alois Alzheimer. In 1906, Alois Alzheimer examined the brain tissue of a woman who had died of a mental illness that had attacked nerves, brains cells and neurotransmitters.

AD not only affects the person himself, but also their family, friends and carers. These patients will need their families and carers to support their daily living. A person with AD may also suffer from mood changes and depression (Starkstein et al., 2008). Indeed, behavioural symptoms of AD are a major challenge and distressing for patients and carers. Although intensive efforts are being made and research is being carried out to find effective treatments, there is currently no cure for Alzheimer's disease (Jack Jr et al., 2008; Korczyn, 2012). It is therefore necessary to get early diagnosis for this disease, not only to allow prompt treatment but also for the future plans of both patients and families.

The early symptoms of AD are as below (Alzheimer's Association, 2011):-

i. Difficulty performing familiar tasks
ii. Getting lost in familiar places
iii. Language problems
iv. Memory Loss
v. Personality changes
vi. Misplacing things
vii. Poor judgement
viii. Withdrawing from social contact

A number of processes have been identified as a diagnosis of AD, as below (Alzheimer's Association, 2011):-

i. Personal History

   The doctor may take a medical history, ask about the ability to do daily activities and gather information about changes in behaviour and personality.

ii. Memory Examination

The examination may include problem solving, attention, counting and language.

iii. Laboratory examination

The doctors may carry out medical tests, such as blood and urine tests.

iv. Brain Scans

Scans, such as computed tomography and magnetic resonance imaging, may be performed.

2.3 Mild Cognitive Impairment

The transitional zone between normal ageing and dementia is defined as mild cognitive impairment (Morris, 2006; Petersen et al., 1999). Morris (2012), in his paper, reported that 99.8% of 17,535 individuals from the National Alzheimer's Coordinating Centre are currently diagnosed with very mild Alzheimer's disease and the large majority, 92.7% of those diagnosed with mild AD could be reclassified as having MCI.

Table 2.1: Characteristic of mild cognitive impairment

MCI Categories	**Definition**		**Progression**
Amnestic	Isolated Memory Disorder	→	• Alzheimer's disease
Multiple Domain Slightly Impaired	Mild impairment in a multiple domain of cognition with or without memory impairment domain	→	• Alzheimer's disease • Vascular dementia • Normal ageing
Single Non Memory Impaired	Single impairment in a single non memory.	→	• Fronto-temporal dementia • Lewy body dementia • Vascular dementia • Primary progressive aphasia • Parkinson's disease • Alzheimer's disease

The criteria for amnestic MCI, as defined by (Petersen et al., 2001), include the possession of (1) a memory complaint (2) objective memory impairment (3) problems with typical activities of daily living (4) some normal general cognitive functions and (5) the full diagnostic criteria for dementia not being met. Different types of MCI can be seen in Table 2.2, as adapted from (Petersen et al., 2001). As seen in Table 2.1, Amnestic MCI and Multiple Domain MCI are likely to be progressive to AD. Amnestic MCI patients have a relatively higher risk of progressing to Alzheimer's disease (Gauthier et al., 2006; Weniger et al., 2011).

2.3.1 Magnetic Resonance Imaging and MCI

Magnetic Resonance Imaging has been found to be a useful tool for analysing brain anatomy in developing MCI (Van De Pol et al., 2007). Petersen (2011) reported that the increased ventricular size and increased medial temporal lobe atrophy at the basal scan are more predictive of progression to AD. Medial temporal lobe atrophy is a diagnostic marker for AD (Petersen et al., 2000). MCI patients have, on imaging scans, presented with an elevated volumetric decline in the entorhinal cortex of the hippocampus in the medial temporal lobe, an area responsible for spatial memory (Dubois et al., 2007). Imaging techniques and lesion studies have shown that this deterioration occurs early in the disease pathway (Petrella et al., 2003) and, therefore, tests sensitive to spatial memory and route learning should show deficiencies in performance if this area is affected.

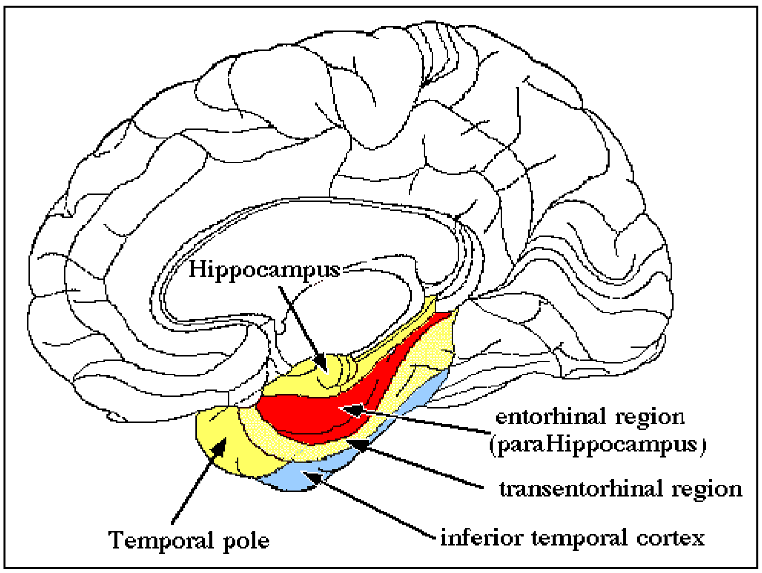


Figure 2.1: A diagram showing the different structures of the medial temporal lobe and related brain regions
(Fuchs and Ogilvie, 2012)

Figure 2.1 illustrates the different structures of the medial temporal lobe and related brain regions (Fuchs and Ogilvie, 2012). Figure 2.2 illustrates MRI scans from patients in the categories of (A) normal elderly, (B) MCI patient and (C) AD patient (Petersen, 2011).

Images A, B and C from Figure 2.2 have an arrow which highlights the hippocampal formation and area of atrophy. Regions of the brain such as hippocampus and parts of the cerebral cortex shrink severely as the disease progress.

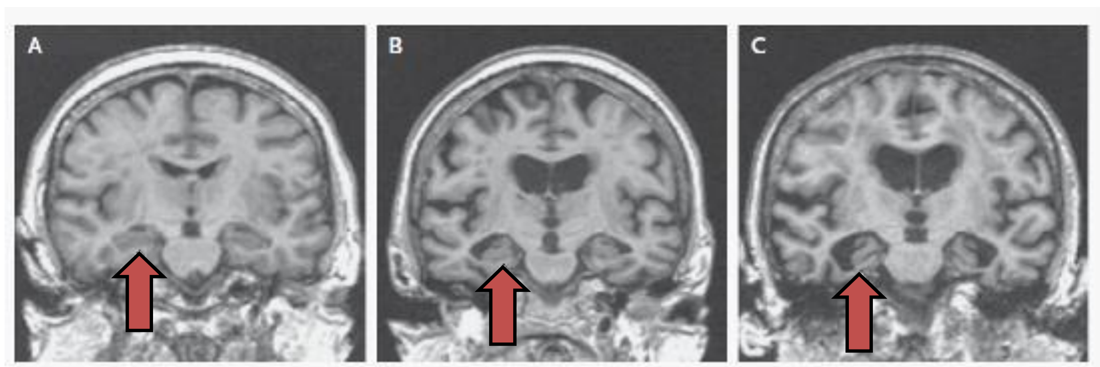


Figure 2.2: MRI scan images: (A) normal elderly (B) MCI patient (C) AD patient (Petersen, 2011)

In order to have a better understanding of cognition in the elderly, identifying the early stages of MCI is important. A better understanding of MCI can provide better treatment. Furthermore, early identification of MCI could provide an opportunity for intervention that might delay cognitive and functional impairment.

2.4 Virtual Reality

The rapid growth in virtual reality technology has proved effective in various applications. There are many definitions of the term 'virtual reality'. One can be described as '*a computer-generated immersive environment with which users have real-time, multisensoral interactions'* (Burdea et al., 1995).

Virtual environments present a unified workspace, allowing more or less complete functionality without requiring all the functions to be located in the same physical space (Moline, 1997). Virtual environments can be defined as interactive, with virtual image displays enhanced by special processing and by non-visual display modalities, such as auditory and haptic, to convince users that they are immersed in a synthetic space (Ellis, 1994). Another definition is the presence of environments simulated by computer that can be experienced (Sherman and Craig, 2003). A virtual environment can be quite effective in immersing the user within a realistic 3D world and in providing a high level of sensory richness (Bowman et al., 2003). With VR technology, a virtual environment lets the user freely explore the 3D space.

VR has applied in many applications, including gaming (Piekarski and Thomas, 2002), entertainment (Cheok et al., 2009), education (Virvou and Katsionis, 2008) and healthcare (Laver et al., 2012). At present, this technology can be said to offer specific attributes or ingredients well-matched to the needs of many mental health and rehabilitation targets (Rizzo et al., 1998).

2.4.1 Types of Virtual Reality Systems

VR systems can be divided into three main categories: non-immersive, semi-immersive and fully-immersive systems. These can be described as follows:-

- Non-immersive VR systems

Non-immersive systems are also called desktop systems because the virtual environment is viewed through a standard high-resolution monitor. Although this is the least immersive implementation of VR techniques, there are many advantages in using these systems. The main advantage of this type of VR is that it does not require the highest level of graphic performance. Additionally, no specific hardware is needed and it can be implemented in any PC system. In conclusion, relatively inexpensive VR systems can be implemented and used for many applications.

- Semi-immersive VR systems

A semi-immersive VR system is viewed mainly through a wide-field projection system. The system attempts to give users a feeling of being at least slightly immersed in a virtual environment and therefore provides a better sense of presence than non-immersive environments. This type of system requires a high-performance graphics computing system and can be integrated with a large-screen monitor, a large-screen projector system or multiple television projection systems.

- Fully-immersive VR systems

This category of system is a computer-generated environment and provides an immersive experience by blocking out the real world. Cave Automatic Virtual Environment (CAVE) is an example of a fully-immersive VR system, consisting of a cube-shaped room in which the walls are rear-projection screens. Users need to wear special glasses inside the CAVE. With these glasses, people using the CAVE can actually see objects floating in the air and can walk around them, gaining a realistic view of what the object looks like.

Table 2.2: Types of VR System (Mujber et al., 2004)

Types of VR systems			
VR system	Non-immersive	Semi-immersive	Fully-immersive
Input devices	Mouse, keyboards, joysticks and trackballs	Joystick, space balls and data gloves	Gloves and voice commands
Output devices	Standard high-resolution monitor	Large screen monitor, large screen projector system, and multiple television projection systems	Head mounted display (HMD), CAVE
Resolution	High	High	Low–medium
Sense of immersion	Non-low	Medium–high	High
Interaction	Low	Medium	High
Price	Lowest cost	Expensive	Very expensive

A summary of these types of VR System is shown in Table 2.2, which compares the three types of VR System in terms of input & output devices, resolution, sense of immersion, interaction and price.

The study by Krijn et al. (2004) investigated patients with acrophobia and comparison to the degree of presence between CAVE and HMD systems. There were 37 patients in their study that were randomly assigned to either CAVE or HMD. Their results showed no differences in effectiveness between the two visualization systems. Although, sense of presence was significantly higher in the CAVE system, this did not produce any differences in outcomes to the treatment.

Recently, Santos et al. (2009) carried out on an experiment to evaluate and compare a traditional desktop setup with VR system using HMD. Their experiment recruited 42 participants to navigate in a complex virtual environment using these two different systems. The results showed that although users were satisfied with the HMD system and found it more interactive, the performance was better using the desktop setup.

Another study by Kim et al. (2012), also investigated the effects of three different types of virtual environment technologies on emotion and task performance. They concluded from their findings that different VE systems are suitable in different scientific purposes where CAVE systems may be effective for relaxation treatment and an

HMD system may be useful for treatment of anxiety disorders in psychotherapy.

Thus for this research, as this is a preliminary study, a prototype was developed using the desktop system which focused only on discriminating between healthy elderly and MCI using a virtual environment regardless of the sense of presence and immersion. This is due to the expensive cost of other different virtual systems and in addition; the use of the desktop system is more suitable to be place in any hospital for a quick screening. As a conclusion, at this stage having a high-end system is not necessary but it can be part of the future work.

2.5 Virtual Reality in Healthcare

Virtual Reality has now emerged as a useful tool in many areas of therapy and rehabilitation. It has played an important role in healthcare domains, including psychology (Schultheis and Rizzo, 2001) and neuropsychology (Rizzo et al., 2004). Rizzo et al. have identified twelve assets of virtual reality applications in neuropsychology as follows (Rizzo et al., 2004):-

i. *The capacity to systematically deliver and control dynamic, interactive 3D stimuli within an immersive environment that would be difficult to present using other means.*

ii. *The capacity to create more ecologically valid assessments and rehabilitation scenarios.*

iii. *The delivery of immediate performance feedback in a variety of forms and sensory modalities.*

iv. *The provision of "cueing" stimuli or visualization tactics designed to help guide successful performance to support an error-free learning approach.*

v. *The capacity for complete performance capture and the availability of a more naturalistic/intuitive performance record for review and analysis.*

vi. *The capacity to pause the assessment, treatment and training for discussion and/or integration of other methods.*

vii. *The design of safe testing and training environments that minimize the risks due to errors.*

viii. *The capacity to improve availability of assessment and rehabilitation for persons with sensorimotor impairment via the use of adapted interface devices and tailored sensory modality presentations built into VE scenario design.*

ix. *The introduction of "gaming" features into VR rehabilitation scenarios as a means of enhancing motivation.*

x. *The integration of virtual human representations (avatars) for systematic applications addressing social interaction.*

xi. *The potential availability of low-cost libraries of VEs that could be easily accessed by professionals.*

xii. *The option of self-guided independent testing and training by clients when deemed appropriate.*

With this promising tool, virtual reality has been used most effectively in the treatment of specific phobias: acrophobia, arachnophobia, post-traumatic stress disorder.

- *Acrophobia*

Jang's group has designed software for the treatment of the fear of heights. The software has a virtual scenario, including an open lift surrounded by props, which allows the patient to feel a sense of heights. Their research has proved that this VR environment is effective and realistic for overcoming acrophobia for two reasons. Firstly, by comparing the results of questionnaires before and after treatment and secondly by noting that subjects' that that the VE seemed to evoke more fearful feelings than the real situation (Jang et al., 2002).

- *Arachnophobia*

Arachnophobia, also known as fear of spiders and other arachnids, also uses virtual reality as a treatment. Software was developed for this phobia. Twenty-three participants took part in a study. The results shows that 83% of patients in the VR treatment group showed a clinically significant improvement and no patients dropped out. This study shows that VR exposure can be effective in the treatment of phobias (Garcia-Palacios et al., 2002).

- *Post-traumatic stress disorder*

An application was developed for the treatment of post-traumatic stress disorder and the results showed some preliminary promise for this treatment (Gerardi et al., 2008).

VR is becoming a useful tool for assessing participants on a wide range of tasks, particularly in the healthcare domain with a specific emphasis on navigational difficulties (Rizzo and Kim, 2005). It is a promising tool in healthcare and offers opportunities for the early detection of AD and MCI. Thus, this research was designed to investigate whether a new VR environment could be used as a novel test of spatial memory for investigating cognition in the elderly and to find out how it correlates with established neuropsychological tests of cognitive function. The overarching aim is to use this test as a method for the early diagnosis of AD in the future.

2.6 Comparison of Related Works Using Virtual Environment for Early Detection of Alzheimer's Disease

Various studies have focused on early AD and MCI and this area is progressively gaining more attention. Previously, Werner et al., (2009) developed a Virtual Supermarket to examine the feasibility and validity of a virtual action planning supermarket (VAP-S) for the diagnosis of patients with MCI (Werner et al., 2009). The main task of this experiment is to purchase seven items from a list and to pay at the counter. The results of their study showed that the use of VAP-S resulted 71.7% of overall accuracy.

Another study by Pengas et.al (2010) also uses a virtual environment with the aim of discriminating between early AD patients and controls. The topographical memory task involves three activities that include the Virtual Route Learning Test (VRLT). Participants need to navigate each of four routes of increasing complexity with 6 trials. One weaknesses of the study is that, if users make a mistake while remembering the route, the trial is then terminated. The results illustrated of over 80% overall accuracy to discriminate early AD patients (Pengas et al., 2010).

Recently, Tippet et al. (2009) examined behavioural performance across different populations using a virtual environment. The participants viewed the route passively 3 times for Path A and once for path B. Then they were required to replicate the given path. Although the study was unable to discriminate between MCI and normal controls, the data can be used for comparison with other work (Tippett et al., 2009).

Cushman et al., (2008) investigated the efficacy of using a virtual reality environment to assess navigational performance in people with early AD and MCI, along with young and old normal controls. Their results were clear: correlations were seen between the real-world navigational deficits and those from the virtual reality environment in all groups, highlighting the potential efficacy of using virtual reality as a test for dementia. None of the measures, however, predicted the group to which the individual belonged (Cushman et al., 2008).

Table 2.3: Comparison of other related works

	Related Works			
	Werner et. al. (2009)	Pengas et.al. (2010)	Tippett et.al. (2009)	Cushman et. al. (2008)
Virtual Simulation able to Discriminate	Yes , accuracy of 71.7%	Yes, accuracy of 80%	No	No
Types of VE Tasks	Purchasing	Navigating	Navigating	Navigating
VE Performance Measurement	Scores are based on twelve correct actions for the task.	Score = total number of errors incurred over all four routes.	VE navigation index = VR memory index + VR movement index	Score of total right points.
VE Level of Difficulties	*No*	Yes, 4	Yes, 2	No
VE Real time data collection	Yes	No	No	Yes
Comparison with Neuropsychological test	Yes	Yes	Yes	Yes

Table 2.3 illustrates the comparison with other related works. From Table 2.3, it can be explained that only the work of Werner and that of Pengas have been able to discriminate between the two subgroups with accuracy of 71.7% and 80% respectively. As illustrated, there are navigation tasks and one purchasing task presented by Werner. As discussed earlier, TD is very common in AD and sufferers are likely to become lost even in familiar places. Thus, testing of their ability to remember a route is most important. In addition, it is interesting to see different scores calculated to produce the performance measures. Furthermore, only the work of Pengas and that of Tippet incorporate levels of difficulty to motivate participants to walk through the virtual simulation to complete the tasks. Apart from that, real time data collection only can be seen in the work of Werner and that of Cushman, in which this real time data collection makes the findings more accurate.

Despite the widespread efforts of research into the use of virtual environments for detection of early AD and MCI, there are still a number of limitations. From a comparison with other related works mentioned, there are some drawbacks that can be addressed. One limitation of the previous works is the difficulty in finding a navigation task that would challenge the user’s motivation. In addition, the highest accuracy to be able to discriminate MCI and healthy elderly using virtual simulation was only over 80%. For these reasons, a new virtual simulation was developed to improve on the existing virtual simulation used for early detection of early MCI. Details of VREAD design, implementation and results of discrimination will be discussed in the following chapters.

2.7 Spatial Memory

Spatial memory involves the ability to encode, store and retrieve information about spatial locations, configurations or routes (Kesselsa et al., 2001). These cognitive functions enable us to remember the location of objects, to find our way around our environment and to recognize familiar places (Moscovitch et al., 2005).

2.7.1 Spatial Knowledge

A theoretical framework has been proposed to describe and explain the process of knowledge development over time in a new environment (Siegel and White, 1975) Generally, there are three categories of spatial knowledge: survey knowledge, procedural knowledge and landmark knowledge (Werner et al., 1997).

The first category is landmark knowledge. Landmarks are defined as the predominant objects in the environment (Charito, 1997). Landmarks are unique objects in fixed routes or locations. For example, a person will decide which path to follow or direction to turn in order to get to the target destination by looking at the specific landmark.

Route knowledge consists of the procedural knowledge required to traverse the distance between origins and destinations successfully (Golledge, 1991). This category of knowledge is usually gained by direct experience of the routes through navigation.

Survey knowledge is analogous to a map of the environment, except that it does not encode a typical map's top-down or bird's-eye-view perspective (Vinson, 1999). This knowledge involves an understanding of the place and can be developed by reading maps.

2.7.2 Topographical Disorientation

Topographical disorientation is the inability to orient and navigate in the environment (Iaria et al., 2009). Topographical disorientation is usually the expression of defects in a variety of cognitive processes, including memory, attention, spatial skills or visual perceptual skills (Aguirre and D'esposito, 1999).

Aguirre and D’Espositi (1999) proposed four categories of topographical disorientation. These four types are egocentric disorientation, heading disorientation, landmark agnosia and anterograde disorientation.

The first type, described as egocentric disorientation, is difficulty in representing the location of objects with respect to self, despite the ability to recognize visible landmarks. These patients also have difficulties in imagining or describing routes.

The second type, landmark agnosia, is characterized by the inability to recognize salient environmental features, such as landmarks for orientation. These landmarks could include the patient’s own house or famous landmarks, although the patient may compensate for their deficit by using small details, such as door number, mailbox or lamppost as cues.

The third category, heading disorientation, is the inability to derive directional information from landmarks that people recognize although the ability to recognize familiar buildings and homes may be retained.

The last category, anterograde disorientation, is the inability to learn to navigate a new environment. These patients are able to recognize previously experienced locations but unable to learn new landmarks.

This is a common manifestation in AD that can occur in the early stages of the disease, particularly in less familiar settings (Passini et al., 1995). The ability to find one's way involves knowing the procedural components of the route to the destination, which may require topographical knowledge about the surroundings (Aguirre and D'esposito, 1999). It is not unusual for Alzheimer's disease patients to get lost in unfamiliar places in the early stages of the disease and, in the later stages, they may become lost or disoriented in familiar places, such as their home or neighbourhood (Cherrier et al., 2001).

Current studies have shown that the allocentric component of spatial memory can be a predictor of AD (Iachini et al., 2009). As visuospatial deficits are manifested in early-stage AD, it is important to test spatial memory to investigate cognition in the elderly. The following chapters discuss the newly developed virtual simulation to investigate cognitive function in the elderly.

2.8 Knowledge Discovery in Databases

Knowledge Discovery in Databases offers many interesting challenges from large amounts of data, providing valuable insight into identifying novel and useful characteristics that may exist in the data. A definition by Fayyad et al. (1996) of KDD refers to "*the nontrivial process of identifying valid, novel, potentially useful and ultimate understandable patterns in data*". Data mining, as defined by (Fayyad et al., 1996), is *"a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations produce a particular enumeration of patterns over data.*" In other words, KDD is an overall process of discovering useful knowledge from data while data mining is a particular step in KDD.

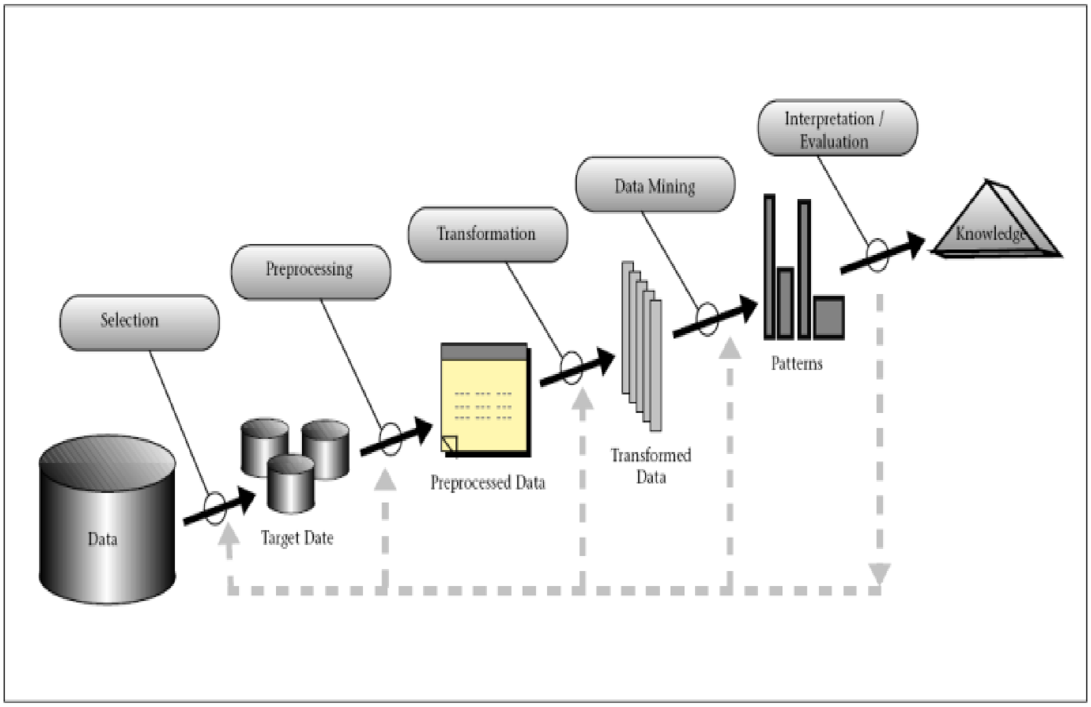


Figure 2.3: An overview of the KDD process (Fayyad et al., 1996).

The steps that compose the KDD process include Selection, Pre-processing, Transformation, Data Mining, Interpretation and Evaluation (see Figure 2.3). The process starts by determining the KDD goals and developing an understanding of the application domain. The next phase involves selecting and creating a target data set on which discovery is to be performed. This is followed by data cleansing and pre-processing. In this phase, basic operations are conducted on data, including the removal of noise and the handling of missing values. Data transformation is the fourth step of the process, in which data are transformed into a form suitable for analysis. The next phase is data mining, which involves data mining algorithms to discover hidden patterns. In the evaluation stage results are evaluated and analysed. In addition, mine patterns are interpreted in line with the goals defined in the first step. The final stage uses the knowledge directly or consolidates the discovery knowledge.

2.8.1 Performance Measures for Machine Learning Algorithms

The performance measures of a machine learning algorithm are commonly evaluated by a confusion matrix (Chawla et al., 2011). Kohavi and Provost (1998) defined a confusion matrix as a table that contains information about actual and predicted classes produced by the data mining tools. Table 2.4 presents a confusion matrix for binary classification. True Positive (TP) is the number of correct predictions for positive output; False positive (FP) is the number of incorrect predictions for the negative output; False Negative (FN) is the number of incorrect prediction for the positive output and True Negative (TN) is the number of correct predictions for the negative output.

Table 2.4: Confusion matrix for binary classification

		Predicted	
		Positive Yes	Negative No
Actual	Positive Yes	**TP**	**FN**
	Negative No	**FP**	**TN**

The confusion matrix is represented by a matrix in which each row represents the instances in an actual class, while each column represents a predicted class. TP, TN, FP and FN are the four different possible outcomes of a single prediction for a binary case with classes consisting of yes or no. The table also shows the accuracy of the classifier as a percentage of correct classifications.

The performance measures can be calculated as follows (Fawcett, 2004; Kohavi and Provost, 1998).

- The *true positive rat*e (*TPR*) is the proportion of positive cases that were correctly identified.

- The *false positive rate* (*FPR*) is the proportion of negative cases that were incorrectly classified as positive.

- The *true negative rate* (*TNR*) is defined as the proportion of negative cases that were classified correctly.
- The *false negative rate* (*FNR*) is the proportion of positives cases that were incorrectly classified as negative.
- The *accuracy* is the total number of predictions that were correct for all classes.

$$TPR = \frac{TP}{TP + FN} \tag{2.1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2.2}$$

$$TNR = \frac{TN}{TN + FP} \tag{2.3}$$

$$FNR = \frac{FN}{FN + TP} \tag{2.4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5}$$

Moreover, there are other measures that can analyses the performance such as presicion, recall and Receiver Operator Characteristic(ROC). ROC curves determine how the number of correctly classified positive examples varies with the number of incorrectly classified examples. Recall is similar to TPR, whereas precision measures the fraction of examples classified as positive that are truly positive. From the analysis carried out in this research, the results were sufficient to discriminate between the two subgroups. Therefore, the other measures are not important for this case study.

One of the objectives of this research was to discriminate between healthy elderly and MCI. With this objective in mind, a machine learning algorithm could contribute to the effort of predicting accurate results to discriminate between healthy elderly and MCI. Furthermore, machine learning techniques and algorithms used in the prediction will be explained in Chapter 6.

2.8.2 Waikato Environment for Knowledge Analysis

Waikato Environment for Knowledge Analysis or WEKA was developed at the University of Waikato in New Zealand (Holmes et al., 1994). It is a comprehensive collection of machine learning algorithms for data mining tasks and can be either applied directly or using the Java code (Bouckaert et al., 2010). WEKA contains tools for data pre-processing, classification, regression, clustering rules and visualization.

As shown in Figure 2.4, this is the screenshot for applying classifiers in WEKA. Classifier can be selected at this stage to make prediction between healthy elderly and MCI patients.

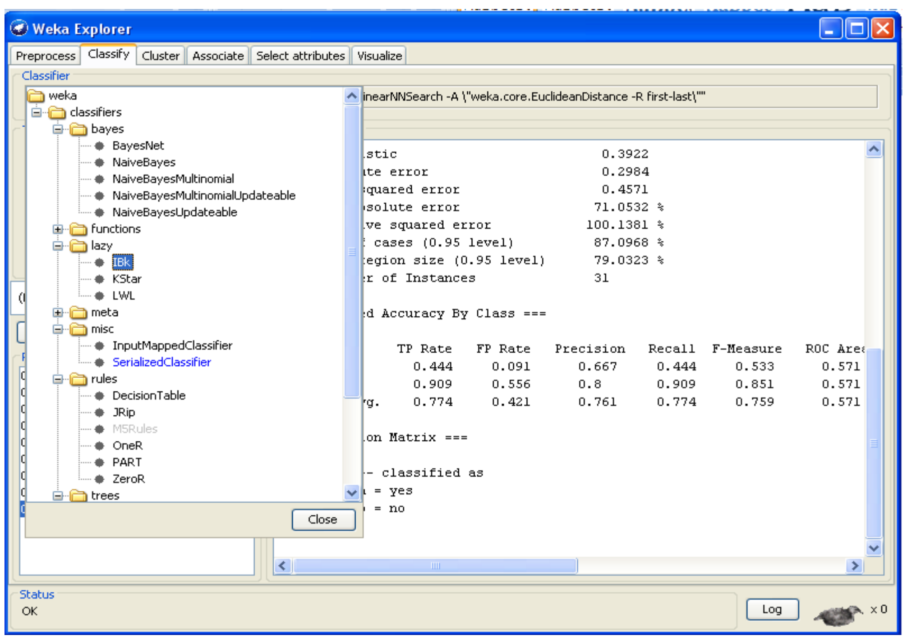


Figure 2.4: A screenshot of using selected classifier in WEKA

The WEKA platform was selected for the implementation of the selected algorithms because it is an open source software which is user-friendly. The tools are able to make predictions using the algorithms provided. Details of the functions used and the results from this research will be explained in Chapter 6.

2.9 Summary

This chapter has presented a literature review of the fundamental knowledge required for this research. Discussions about virtual reality are presented with VR technology in healthcare. In addition, an overview of AD and MCI is discussed. Subsequently, this chapter explains the comparison with other related works and ends with a discussion of knowledge discovery in databases and WEKA software.

Chapter 3

Virtual Environment: Design and Development of VREAD

3.1 Introduction

This chapter presents the procedural steps undertaken in the design and development phase of the virtual simulation. Firstly, the proposed design criteria, which are essential to the development of testing spatial memory in the elderly, are explained. In addition, this chapter presents an Analysis, Design, Development, Implementation and Evaluation (ADDIE) model, a step-by-step and systematic approach used in the design and development of this simulation. Furthermore, this chapter describes the initial design and development of a novel virtual environment simulation, with the aim of being able to discriminate between MCI and healthy elderly.

3.2 Proposed Design Criteria

Elderly people who grew up without the benefit of computers may feel that computers are not meant for their use or are not relevant to them. The technological revolution has developed after the early years of this generation (Leonardi et al., 2008). Research by Dickinson et al. (2005) concluded that elderly people find it difficult to use a standard interface compared to young people.

It is of great importance that a design for the elderly should follow specific criteria to take their physical age impairment into account (Mahmud et al., 2010). Numerous studies have highlighted the design criteria for elderly users. Ijsselsteijn et al. (2007) identify several game design opportunities for the creation of digital games for elderly users, as depicted in Table 3.1. Another study, by Flores et al. (2008), lists gaming design criteria for the elderly from their comprehensive reading, as shown in Table 3.1. Yusoff et al. (2009) propose game attributes to meet users' requirements on the basis of psychological theories, such as cognitive, constructivist, behaviourist and neuroscience perspectives, for serious game design.

Taking this into consideration and for the full potential of technology to be useful to these people, their needs and requirements must be considered in the design of the virtual environment. This research focuses on the early detection of MCI, requiring participants to walk through a virtual environment and complete tasks, necessitating a design that motivates them to participate. Thus, for this study, several criteria have been selected and embedded in the virtual simulation specifically for the testing of spatial memory in elderly populations. The findings demonstrate that it is feasible to work in virtual environments with elderly people by implementing the design criteria (Wan Shamsuddin et al., 2011b).

Table 3.1: Design criteria used for prototype development for elderly.

Research Study	**Design Criteria**
Ijsselsteijn et al. (2007)	• Content that is beneficial to the elderly. • Learn the necessary skills. • Easy computer tasks. • Simple interaction design. • Functional limitations • Feedback
Flores et al. (2008)	• Appropriate game challenging to motivate users. • Simple interface and controls. • Multiple levels of difficulty. • Sensitive to decreased sensory acuity. • Sensitive to slower response. • Creation of new learning.

VREAD was developed based on the proposed design criteria as follows:-

i. Levels of Difficulty.

ii. Different Maps.

iii. Linking of Psychological Theories.

iv. Minimal use of Landmarks.

v. Data Collection.

vi. Collision Detection.

vii. Training Module.

viii. Simple and Structured Interface.

ix. Scenarios.

The following section briefly explains each of these criteria.

A. Levels of Difficulty

One important consideration when designing this simulation was to maximise the enjoyment and motivation of users. The ease of learning in the initial stage and ability to master the other stages should be considered when designing successful computer games (Bekker et al., 2005). Levels of increasing difficulty in most games motivate the players and challenge them to adapt their specific skills (Rego et al., 2010).

In this simulation, there were five levels of difficulty ranging from easy to complex. Level 1 was considered to be the easiest, with each level getting progressively harder and requiring more complex navigation. The level of difficulty was defined on the basis of junctions, paths and target destinations for each of the routes that users had to navigate in the testing phase. Additionally, extra destinations were added as a ‘dummy’ to make the path more complex and difficult. These extra dummies also have underlying grids where data can be collected if users go through the dummy destinations. This allowed the research to investigate the users’ task performance at each level. Table 3.2 summarises the level of difficulty of the junctions and turns the users had to make in order to get to the destinations.

Table 3.2: Summary of the junctions and turns in the simulation.

Levels	Junctions and Turns
Level 1	1
Level 2	2
Level 3	3
Level 4	4
Level 5	5

B. Different Maps

Different versions of maps are routinely used in experimental psychology. In this simulation, different maps were implemented for each module: VR Park and VR Games. Both maps were scenarios set in a virtual park with other objects, such as trees, plants, buildings, benches and street lamps. VR Park had the target destination in a park layout which included a playground, picnic area, art gallery, park squares and mini garden. For the second module, VR Games had the target destination in an outdoor games setting that included giant chess, giant board games, lawn bowls, mini golf and a picnic area.

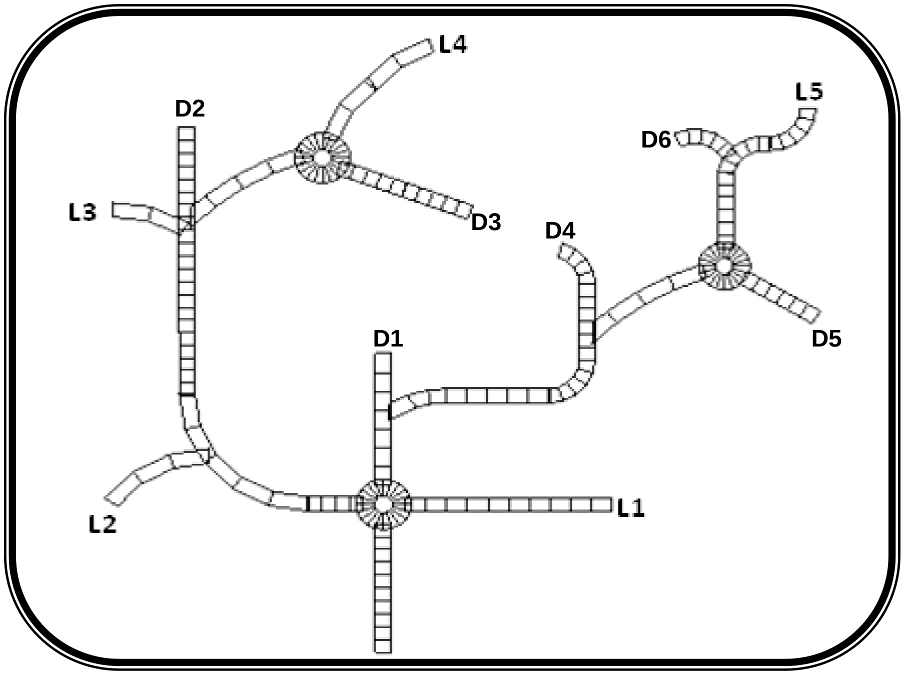


Figure 3.1(a)

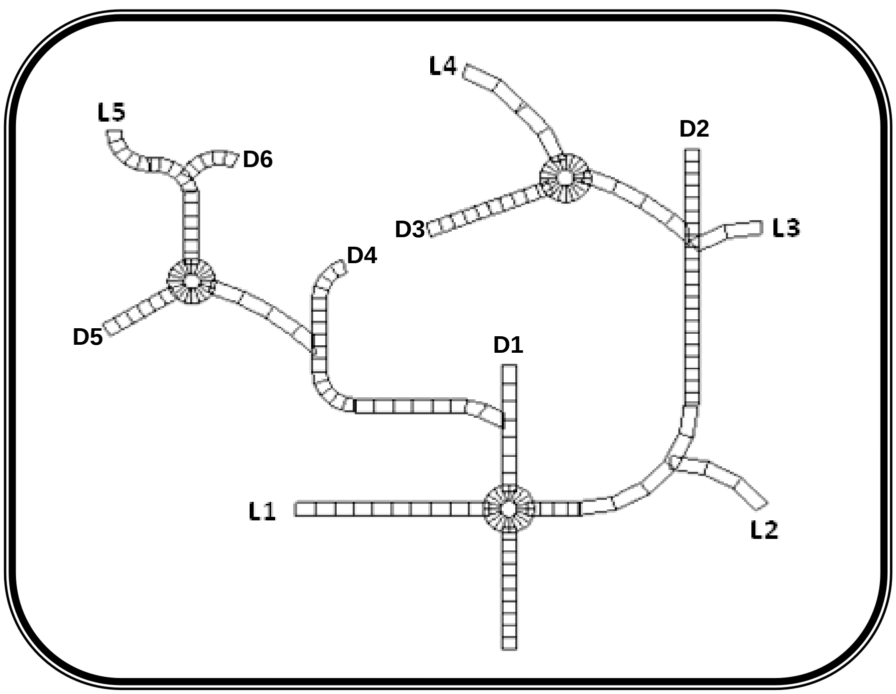


Figure 3.1(b)

Figure 3.1: (a) Path in VR Park (b) Path in VR Games.

Both modules had the same number of levels and landmarks. Moreover, VR Park's path was transposed to VR Games' path. Figure 3.1 presents two different path styles to test the efficiency and effectiveness of the experiments. L1, L2, L3, L4 and L5 are the destinations of five levels and D1, D2, D3, D4 and D5 are the dummy destinations for VR Park and VR Games.

C. Linking of Psychological Theories

This research embeds psychological theories into the design and development of the simulation. The theories embedded were behaviourism, cognitivism, constructivism and error based learning. Zaibon and Shiratuddin (2010), summarise the learning theories as follows:

1. Behaviourism

- The concept of repetition in a game and receiving rewards.

2. Cognitivism

- Integrating learning and game experience to build intrinsic motivation.

- Game experience that integrates learning and play through a discovery process.

3. Constructivism

- Offers game challenges to solve problems.

Errorless learning is one of the successful techniques used in rehabilitating people with memory disorders, especially those with a more severe impairment (Tailby and Haslam, 2003). Errorless learning can be defined as presenting information across acquisition trials in such a way that the learner is prevented from making errors, learning exclusively by repeated exposure to correct information (Lloyd et al., 2009).

With regard to the theories above, the simulation successfully embedded all of them in the design and implementation, as summarised in Table 3.3.

Table 3.3: Summary of psychological theories and their Implementation.

Theories	Implementation of VREAD
Behaviourism	• Training module in understanding the movement e.g: using the cursor keys. • Overall scores of the performance as rewards.
Cognitivism	• Simple and structured interface and contents. • Game information to existing knowledge.
Constructivism	• Different types of modules and scenarios. • Five different levels from easy to complex.
Errorless Learning	Repetition of the exercise of reaching the target destination three times by following the red lines attached to the path.

D. Minimal Use of Landmarks

Landmarks are useful when finding directions and communicating about space (Weissensteiner and Winter, 2004). They can be divided into two major categories: geometric and object landmarks. Geometric landmarks are the salient features of the environmental construction, whilst object landmarks are the salient objects within the environment that can be easily distinguished from the background, such as a street lamps or buildings (Lin et al., 2008).

Although landmarks are an integral aspect of navigation, in this experiment to investigate spatial memory through navigation minimal use of landmarks was required. Table 3.4 shows landmarks used in the applications for all levels. There were no maps in this simulation to help users; signboards were hidden and could only be seen in the training phase. This was to assess users' spatial cognition in memorising a route with limited landmarks.

In Table 3.4 the limited landmarks are summarised. The design of the modelling objects, such as trees, benches, fences, junctions and turns were identical in each category. Although there are number of landmarks in the simulation, they still cannot be considered to be cues to memorize the route.

Table 3.4: Landmarks used in the simulation for all levels.

Landmarks	Numbers of Landmarks
Statue	1
Trees	94
Benches	26
Fences	Completely around the path
Junctions and turns	18
Dummy destination	6
Buildings	39

E. Data Collection

There are path squares, which consist of underlying grids unknown to users. These path squares have certain values that will be collected and subsequently used to calculate the scores. During the testing phase, data were collected automatically in real time. These data comprise the five attributes: correct path, incorrect path, correct sequences, incorrect sequences and time.

1. Path

During testing, users were allowed to move freely to the target destination. All movements were captured and recorded for data analysis.

2. Path Sequences and Path Squares

Users were required to reach the target destination using the correct path sequence. The data from path-tracking would show correct sequences, incorrect sequences, correct squares and incorrect squares. Path squares comprised an underlying grid unseen by participants. While users moved along the path, the unseen grid was tracked and recorded. This was to ensure all movements were captured for analysis.

3. Timing

During testing, the amount of time taken to complete the journey from the starting point to the target destination was recorded.

4. Score

Scores were calculated based on path sequences and path squares. The scores show the percentage performance of the users. The Route Test Score (*RTS*) was proposed to calculate the total score (see Equation 3.1), which includes the values of the correct and incorrect sequences as follows:

$$RTS_i = \left(\frac{cs_i}{is_i + es_i} \right) * 100 \qquad 3.1$$

RTS is a Route Test Score where;

l is a level,

cs is correct sequence recorded for the particular level,

is is incorrect sequence recorded for the particular level,

es is exact sequence for each level.

Table 3.5: Sample calculation based on Route Test Score.

User ID	Correct Sequence	Incorrect Sequence	Exact Sequence	Calculation of (RTS)	RTS Score
101	15	0	15	15/(15+0)*100	100%
102	15	3	15	15/(15+3)*100	83%
103	14	2	15	14/(15+2)*100	82%
104	14	0	15	14/(15+0)*100	93%

Table 3.5 shows sample calculations based on the Route Test Score. The three main attributes were first extracted from the data collections: correct sequence, incorrect sequence and exact sequence. Based on Table 3.5, the calculation column shows how RTS has been calculated using equation 3.1, with the value of RTS in the last column.

F. Collision detection

Collision detection in virtual environments is very important since it determines how close you can get to the 3D object. It will automatically stop the movement before collision with an object. In this simulation, users were allowed to move only on the path without going onto the grass or elsewhere. The main reason for this collision detection was to avoid disorientation of the users. Boundaries were programmed in this simulation to prevent users navigating beyond the path or route given.

G. Training module

A training module is essential, especially when the study involves elderly people who may be less proficient at computer use than younger people. In this simulation, a training module entitled VR Practice was introduced to users to give them the opportunity to practice with the arrow keys on the keyboard. The four arrow keys of up, down, left and right were used to navigate through the virtual environment. This practice would help them to use the simulation easily. In addition, this module would help users become familiar with the virtual environment.

H. Simple and Structured Interface

The simplicity of this virtual environment is essential when used to study spatial memory for navigation. A structured and simple interface is important in understanding the interactive process so that the user can fully concentrate on the task (Wallace et al., 2010). Simple and relevant graphics should be used in the design, whilst decorative animation and pictures, wallpaper patterns and flashing text should be avoided (Hawthorn, 2000). Furthermore, icons such as exit or home buttons should also be familiar to the users (Heim, 2008). The font sizes of a site should be 14 point or more and have very clear headings due to a general decline in visual acuity among elderly persons (Hartley, 1994).

I. Scenarios

The outdoor environment seems to offer an opportunity for elderly people to be physically active, to have contact with nature and to meet with friends and neighbours (Sugiyama and Thompson, 2006). For that reason, VREAD was designed in the outdoor environment. As an alternative to a real life environment, due to health and safety issues, a virtual environment provides one of the best ways of investigating spatial memory. The scenarios chosen for this simulation were outdoor environments, in the park or an open environment with trees surrounded by tall buildings and blue sky. The scenario was created on account of its similarity to a real life environment.

3.3 Design and Development of the Simulation

All of the design criteria that have been discussed were successfully applied to the virtual simulation. The following sections discuss the integration of the design criteria in the design and development of VREAD. The ADDIE model was adopted as the flow guideline for the development process. Although many different generic models could be adapted to meet specific user needs, ADDIE is a commonly used method that can be effective in almost every circumstance (Shen-Tzay et al., 2005). Each phase of the prototype development, including Analysis, Design, Development, Implementation and Evaluation, makes its own contribution towards the main objective of this research.

The five development phases are summarised in Figure 3.2. The analysis phase was the first and the most fundamental phase in this model. During analysis, goals and objectives were identified; hardware, software and user requirements were documented along with consideration of the timeline or constraints for this development.

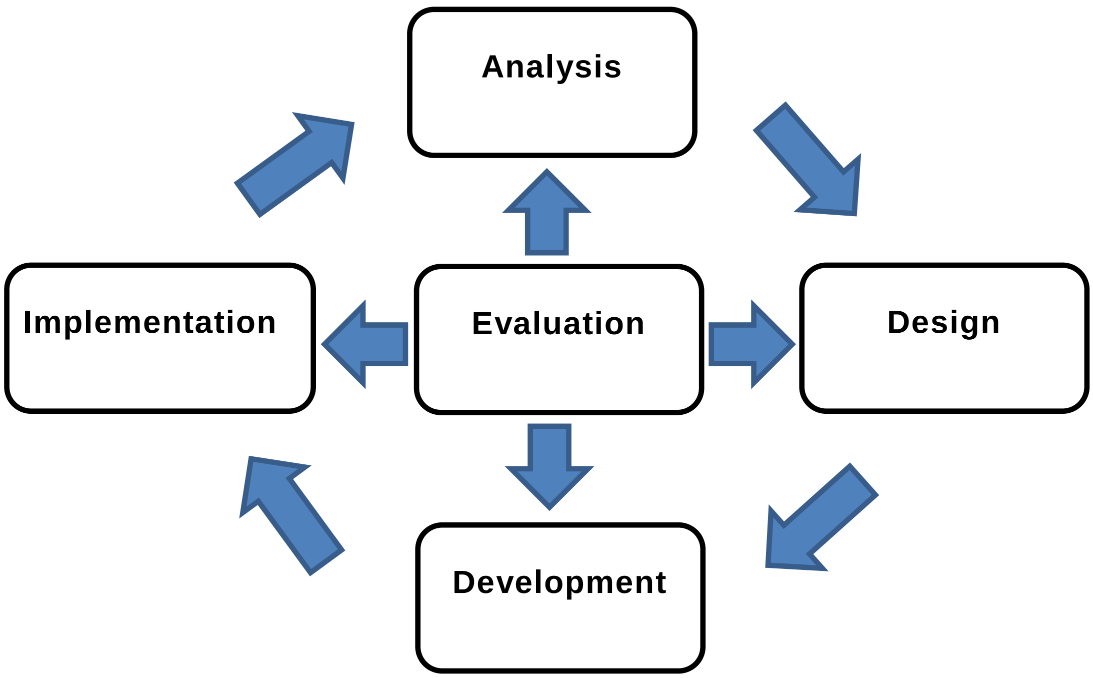


Figure 3.2: ADDIE model for design and development methodology (Gustafson and Branch, 2002).

In the design phase, the complete design of the prototype, including detailed storyboards with contents, was carried out. During this phase, information was compiled from the analysis phase then organised into an outline of instructions that became a blueprint for the development phase.

During the development phase, the design specifications were produced. All aspects of the design interface from the previous phase were utilized to develop the working prototype. This phase also involved development through the programming mentioned in the design phase.

The implementation phase involved the deployment of the prototype. This was the phase in which the prototype was actually administered to the target users. The delivery environment should be prepared for implementation of this prototype.

The evaluation phase was considered to be the most crucial in this methodology since it determines and evaluates the final findings based on the objectives identified earlier. This phase decides whether the prototype is effective and satisfies the objectives. It also considers feedback from users.

Further explanation for the development of VREAD will be discussed in the following sections.

3.4 Hardware and Software Requirements

This section discusses the hardware and software needed for the development of VREAD

A. Hardware

The development of this virtual environment required a high specification of hardware. In this development, the hardware consisted of a Pentium-based computer with NVIDIA graphic card of 512MB memory, 4MB RAM and an Intel Core Duo processor. Standard input and output devices, such as keyboard, mouse and computer monitor were used for this experiment.

B. Software

Software requirements for the development of this simulation were Maya Autodesk, Virtools 5.0, Visual Basic and 3DVIA Players. Modelling was performed using Maya to create models and scenes (see Figure 3.3). Subsequently, the scenes were exported to Virtools, a 3D authoring tool which handled all the programming, including interactivity, setting and configuration (see Figure 3.4). Virtools is a complete development and deployment platform with an innovative approach to interactive 3D content creation. It integrates rich interaction block and custom scripting. To run the VR simulation 3DVIA Players needed to be installed (see Figure 3.5). For analysis and representation of the results, an information system was developed using Visual Basic (see Figure 3.6).

Other than the requirements mentioned above, there are other various options of development tool and hardware for the development of VREAD. For instance, Quest3D is also a development tool for creating real-time applications and was used in many commercial games, but the cost is expensive.

Another type of development tool is Unity which is a game development that allows games to be created for different platforms such as MAC, PC or IPhone (Petridis et al, 2010). Therefore, although these game engines have their own advantages, the goal is to be able to create a virtual environment that satisfies user needs.

On the other hand, there are other alternatives for hardware. A 3D display could replace a traditional monitor to provide better visualisation. A desktop system may also be a substitute with a more complex system such as HMD. These alternatives may give a higher sense of presence and immersion but there is study that shows performance was better in using the desktop system rather than HMD (Santos et al, 2009).

Below is a list of the software used in this development process:-

i. Maya: Creates scenarios and 3D modelling objects

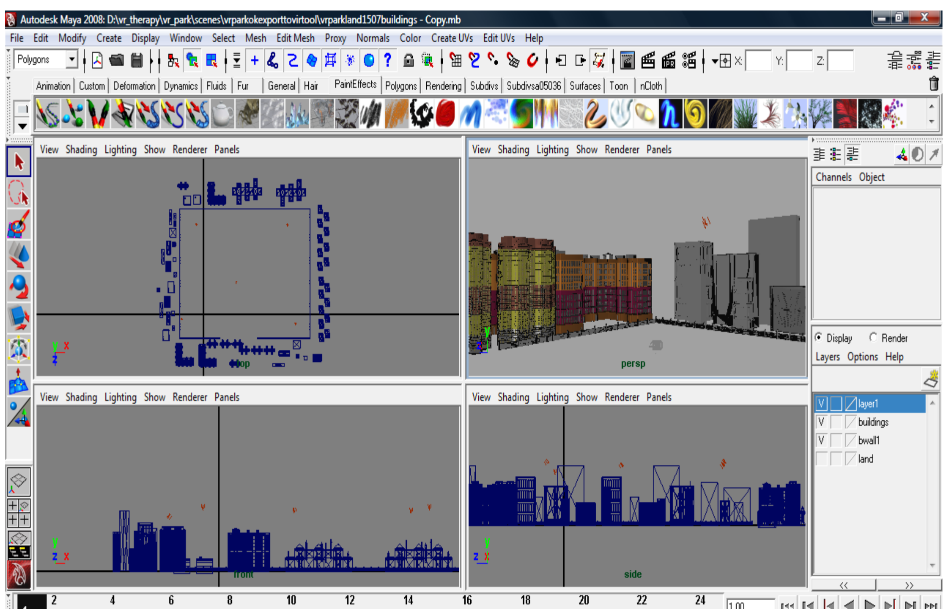


Figure 3.3: A screenshot of 3D Modelling being created in Maya

ii. Virtools: Simulation engines development and integration of 3D modelling objects.

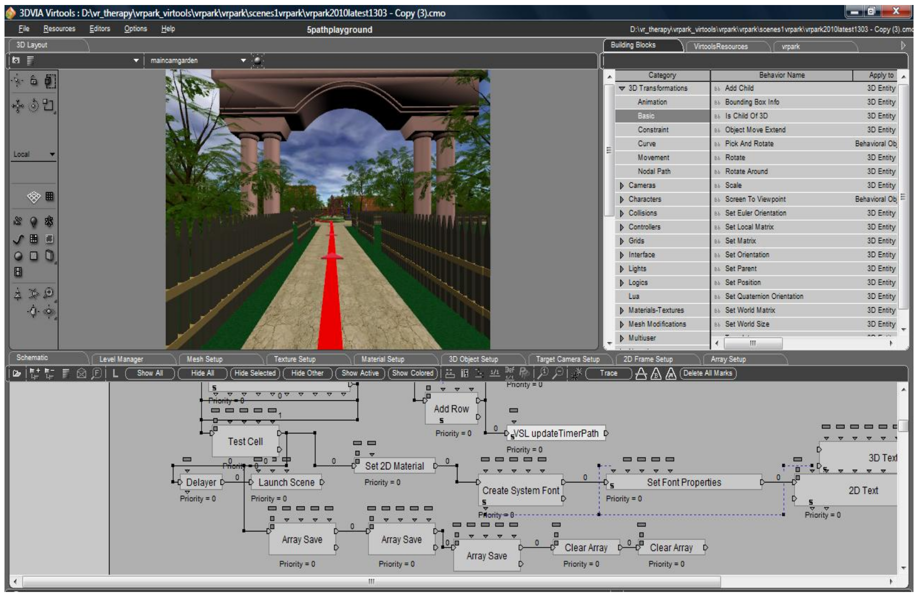


Figure 3.4(a)

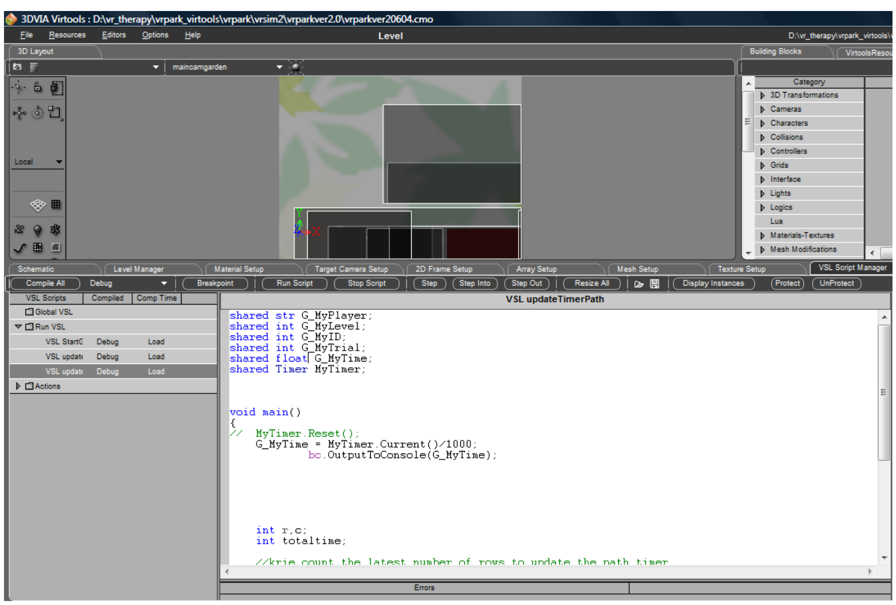


Figure 3.4(b)

Figure 3.4: (a)(b)Screenshots of Programming 3D objects in Virtools.

iii. 3DVIA Player: A browser plugin to run VREAD.

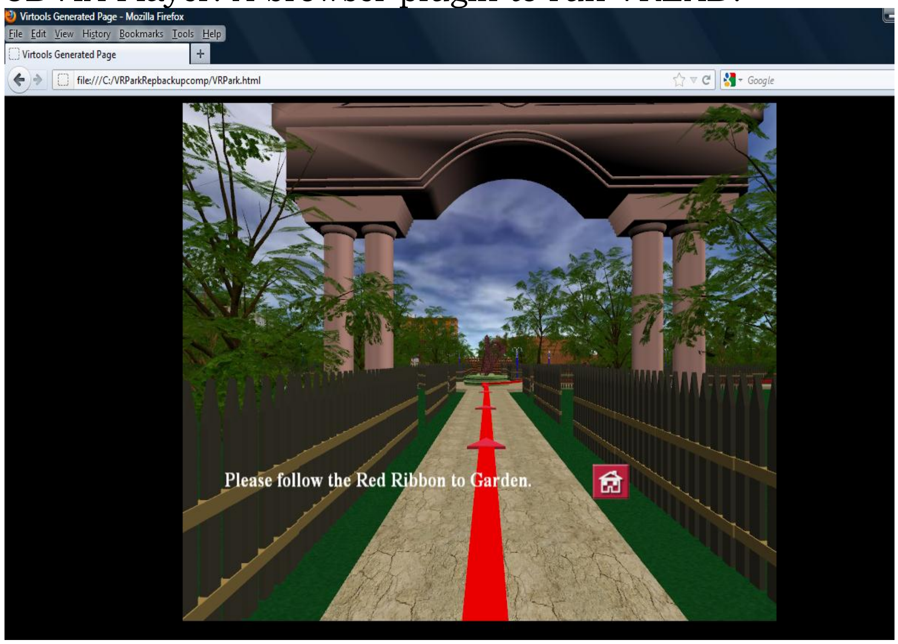


Figure 3.5: A screenshot of 3DVIA Player Plugin to run VREAD.

iv. Visual Basic: Developing and programming the information system

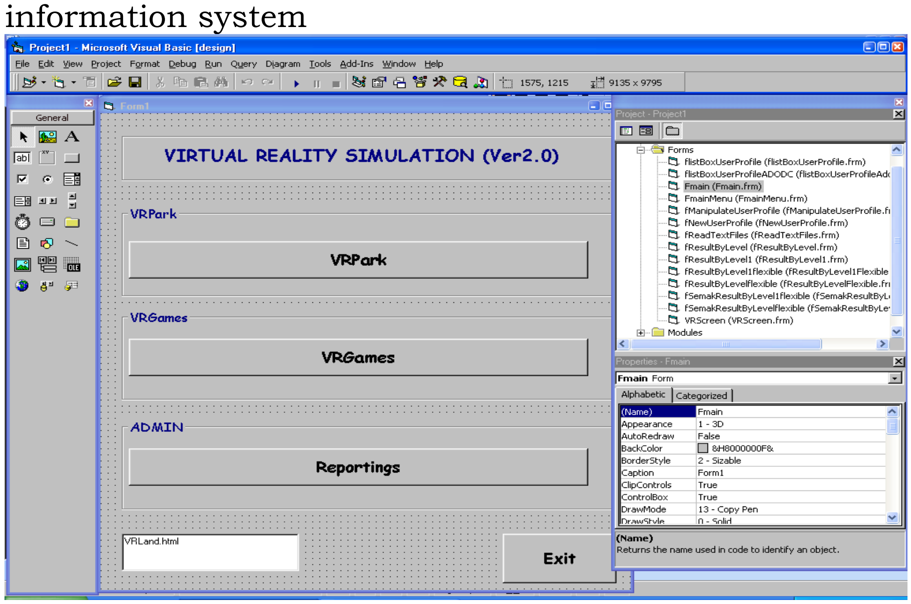


Figure 3.6: A screenshot of programming the information system.

3.5 Virtual Environment

A virtual environment was developed with the aim of identifying cases of MCI. The virtual environment was developed to contain scenery of a park with high buildings scattered around the surrounding area. The proposed virtual simulation offered a friendly environment to users. VREAD consisted of three modules: VR Practice, VR Park and VR Games (see Figure 3.7).

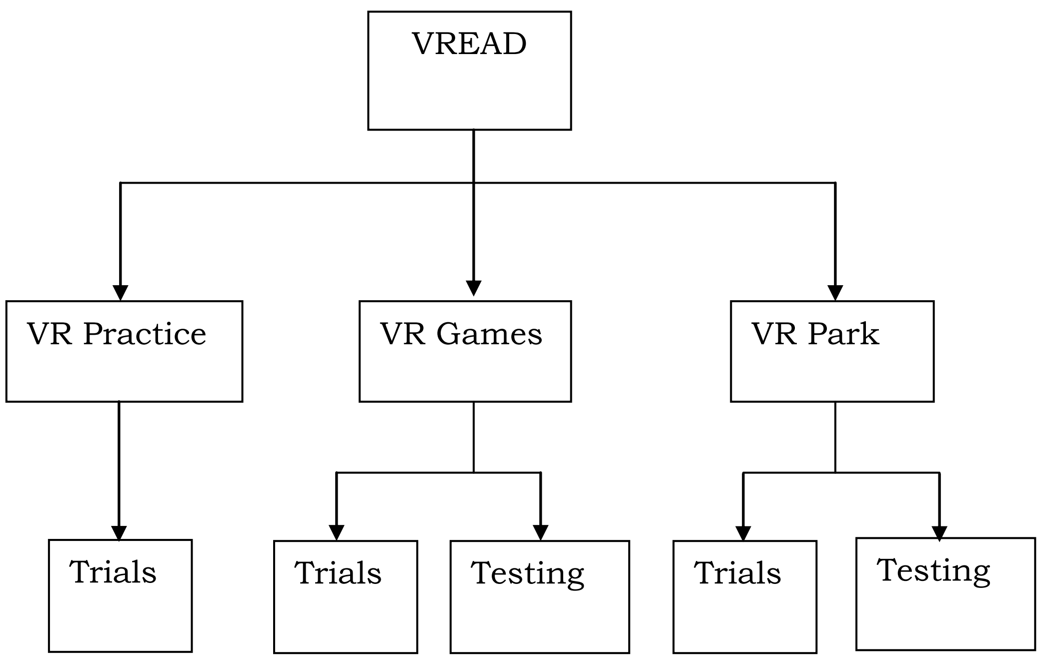


Figure 3.7: VREAD Modules.

3.5.1 VR Practice

VR Practice is a training module. In this module, users were trained to use the four cursor keys with the red lines attached to the path. Moreover, users had the opportunity to become familiar with the virtual environment. The participants were also being tested indirectly to ensure that they understand the various function keys and interface of the system. In addition, there were five levels of difficulties for VR Park and VR Games where level 1(easiest level) was to ensure that the participants understand the use of the arrow keys and interface in navigating the virtual environment. Figure 3.8 illustrates a screenshot of the training module.

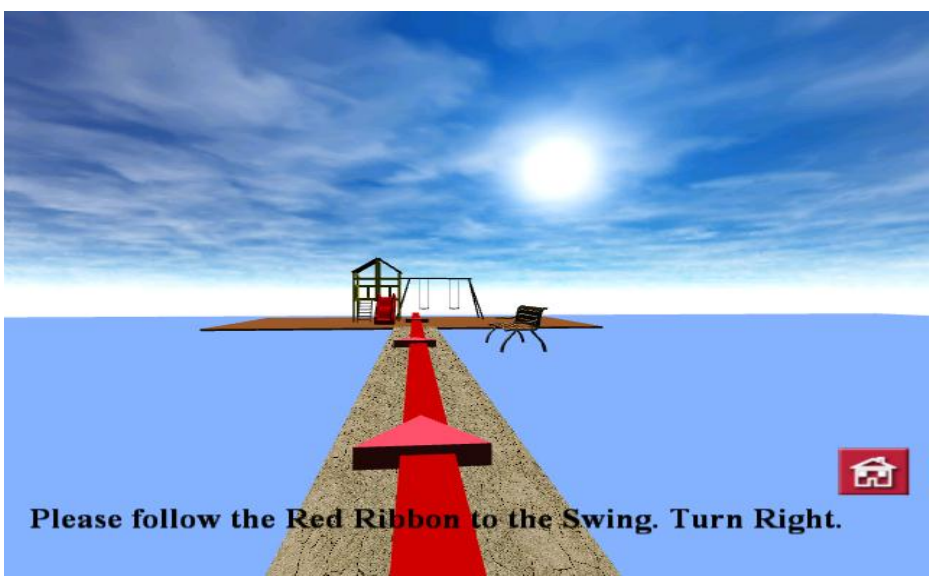


Figure 3.8: A screenshot of the Training Module (VR Practice).

3.5.2 VR Park

The scenario settings for this module were within a park. There were five specific target destinations, including a playground, art gallery, garden, rest area and picnic area. Users experienced a walk through a park in the city where there were tall buildings to be seen, trees all around and various other features (see Figure 3.9).

Figure 3.9(a)

Figure 3.9(b)

Figure 3.9: (a) Screenshot of VR Park environment: Garden;
(b) Screenshot of VR Park environment: Gallery.

3.5.3 VR Games

In this module, the location was surrounded by tall buildings and houses. Inside the park, there were five different and specific types of games available, such as giant chess, giant board games, lawn bowls and mini golf (see Figure 3.10).

Figure 3.10(a)

Figure 3.10(b)

Figure 3.10: (a) A screenshot of VR Games environment: Mini Golf; (b) A screenshot of VR Games environment: Giant Chess.

3.6 Procedures of VREAD

Figure 3.11 represents the procedures performed in VREAD. Firstly, users' information was recorded in the system. A unique user ID was given to each user.

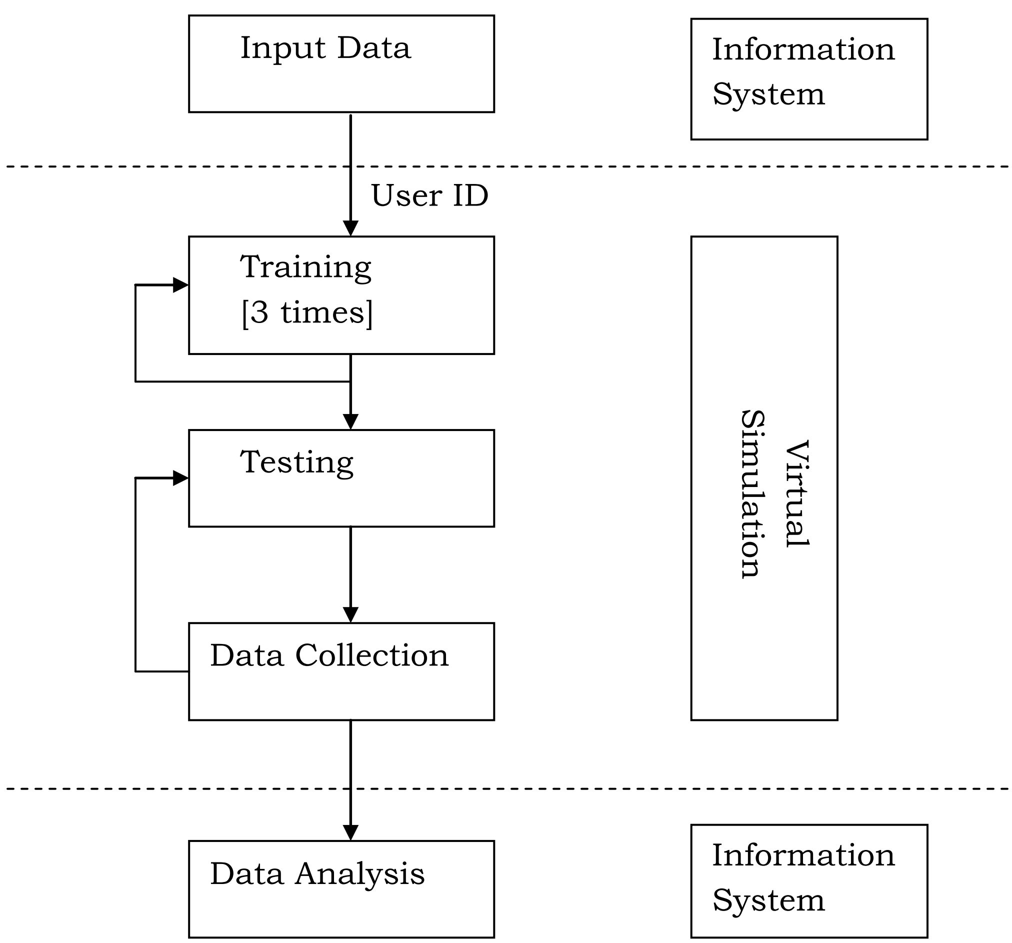


Figure 3.11: Procedures of VREAD.

Next, in the virtual simulation, either VR Park or VR Games was introduced to users. The users were allowed to repeat the exercise of reaching the target destination three times by following the red lines attached to the path. Following this, they were tested on their ability to memorise the given path without the red lines. All data were collected and recorded during this phase. Lastly, all data were exported to the information system for analysis and for scores to be calculated.

3.7 Information System

The purpose of the development of the information system was to input users' information. Moreover, the other main purpose of this information system was the import and export of data from the virtual simulation, which were then to be calculated and analysed. These data were also exported to Microsoft Excel for subsequent export to the WEKA software for prediction. Further discussion on the use of WEKA software will be discussed in Chapter 6.

Figure 3.12 shows the diagram of the process in the information system. Inputs for the information system comprised information from users. In addition, data were also imported via a text file from the virtual simulation, which were then processed and the outputs were the results of the scores calculated and other attributes that could be visualised in a meaningful way. Moreover, this information system could present reports via Microsoft Word format and could also produce a file that could be saved in Microsoft Excel and directly imported to the WEKA software for the prediction process.

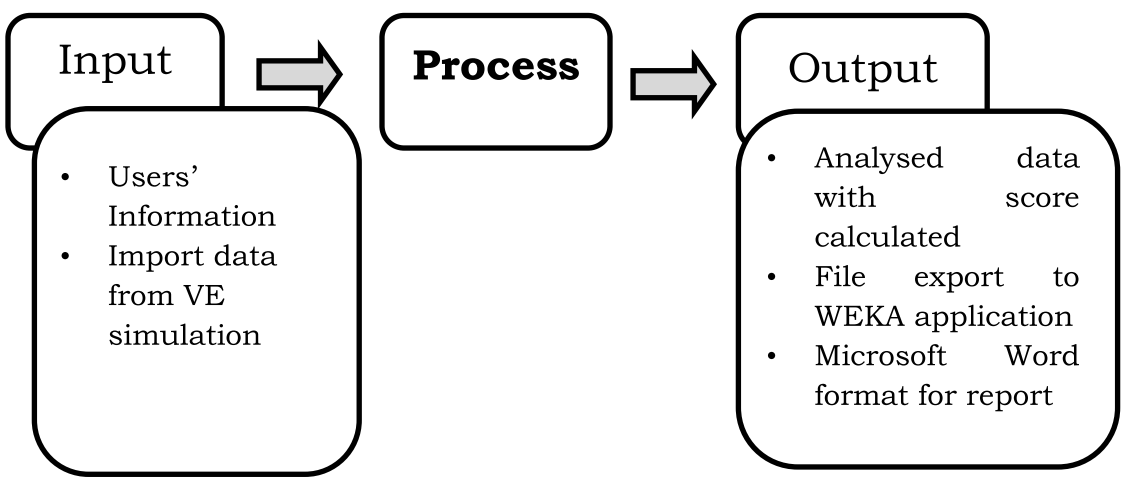


Figure 3.12: Diagram of the Information System process.

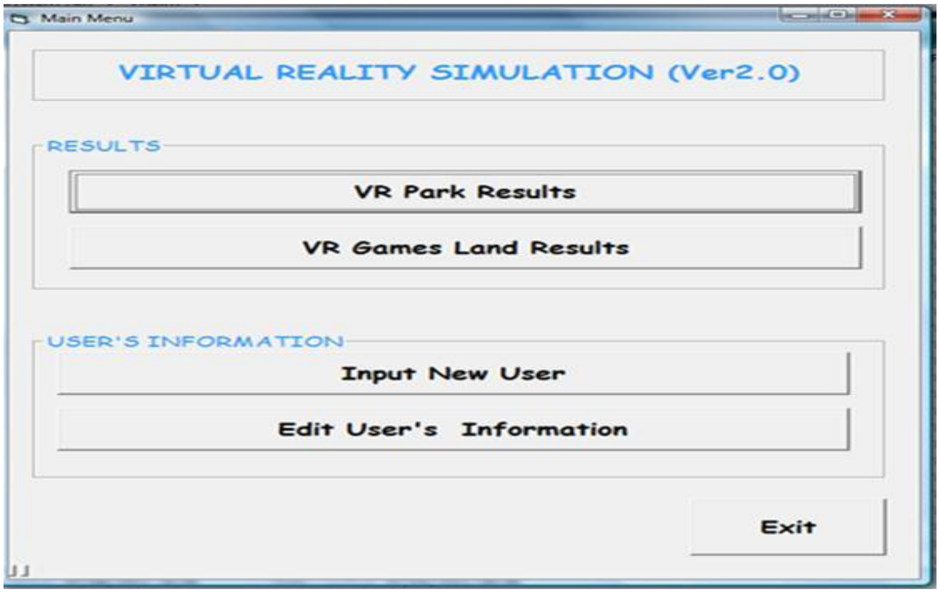


Figure 3.13(a)

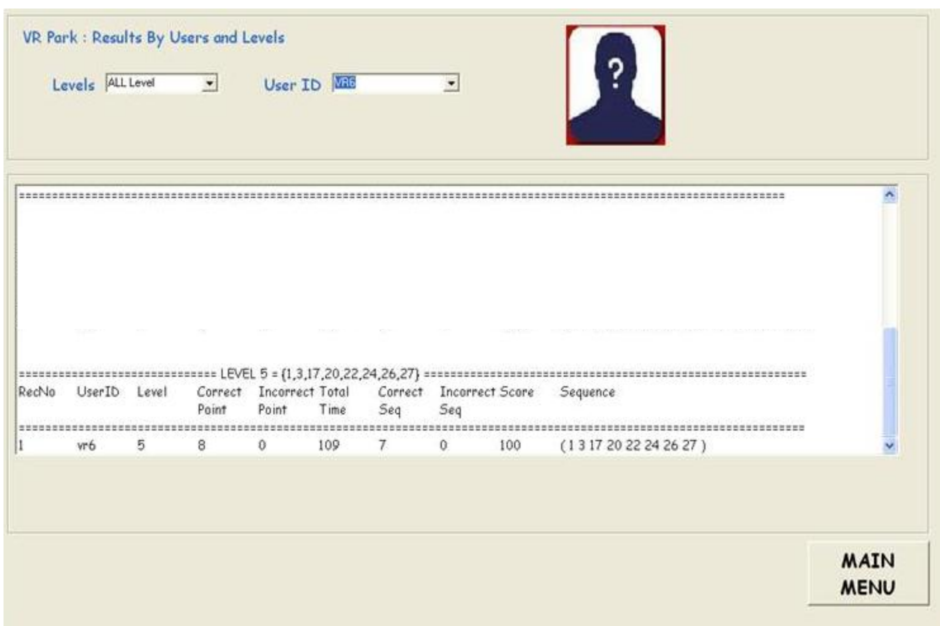


Figure 3.13(b)

Figure 3.13: (a) A screenshot of interface for menu;

(b) A screenshot of interface for results.

Figure 3.13 demonstrates screenshots of the information system. Figure 3.13(a) illustrates the screenshot of the interface for Menu. Figure 3.13(b) illustrates the screenshot of the interface that shows results or data that have been calculated and analysed.

3.8 Summary

This chapter has placed emphasis on the design and development of this virtual simulation. The design criteria presented here are essential to the development of a virtual environment for the elderly with MCI in order to test their spatial memory. It has been developed by adopting the ADDIE model as the main guideline for the entire development process.

Chapter 4

VREAD Simulation and Implementation: Experiments and Results

4.1 Introduction

This chapter describes the experimental setup and methods used for the investigation into whether a new VR environment could be used as a novel test of spatial memory to investigate cognition in the elderly. It details the procedure of a technical experiment as a feasibility test for the virtual simulation. Two types of experiments are discussed: experimentation with non-elderly participants and experiments with elderly participants. Analyses of the results are explained in detail. The results of this study indicate that a virtual simulation based assessment for spatial memory could be a possible method for the early detection of MCI since significant correlation with existing measures implies reliability.

4.2 The Experiments of VREAD with Elderly Participants

This section investigates our newly developed virtual environment test to see if it can discriminate between healthy elderly and people with MCI. Furthermore, the study also compares score on the VR assessments with existing neuropsychological tests.

4.2.1 Participants

The group study consisted of 31 elderly participants. All controls possessed no existing neurological disorder or existing memory impairment. Furthermore, none of the participants had an uncorrectable visual impairment. All participants were recruited from the University of Bradford Division of Psychology over-65 Participant Pool. This is a database of participants who had volunteered their details in order to take part in psychology research. They were recruited via local community centres and groups. The study was approved by the Board of Ethics, University of Bradford and the participants were included in the study after signing a consent form.

4.2.2 VE Experimental Apparatus: Virtual Reality for Early Detection for Alzheimer's Disease (VREAD)

The experimental environment operated on a standard PC and simulated two fully various textured scenes. VREAD consists of three modules: VR Practice, VR Park and VR Games, as discussed in Chapter 3 earlier.

4.2.3 Procedures

The experimental procedure was as follows. Those interested in participating in the experiment were given an information sheet to sign and were then asked to sign a consent form. They also had the chance to ask any further questions if they so wished.

The participants were then given a training and familiarisation session through the VR Practice module. At the beginning, the controls of the keyboard were explained to participants. Then they were instructed how to navigate in the VR simulation and how to use the tools in the application to perform navigation tasks. They were allowed to practice navigating using the cursor keys to move around the virtual environment until they were satisfied, using the VR Practice module.

Afterwards, each participant was able to see one of the two modules: VR Park or VR Games. The participants were instructed to repeat the exercise of reaching the target destination three times by following the red lines attached to the path. Following this, the participants were tested on their ability to recollect the given path without the use of the red line. All data were collected and recorded during this phase. This was repeated from level one to level five, with each level becoming progressively harder, with more complex routes and junctions.

Following completion of all five levels of the assessment, the participants completed a brief usability questionnaire before commencing a battery of neuropsychological assessments, comprising of the Mini-Mental State Exam (MMSE), word recall, the CANTAB Paired Associates Learning (PAL) and the Graded Naming Test (GNT). The purpose of these tests was to compare the prediction results using those approved tests with the proposed virtual simulation. After completion, participants were given a usability questionnaire and thanked for their support. No clinical interpretation of the experiments or questionnaire was given to the participants. Lastly, all data were exported to the information system to be analysed.

4.2.4 Neuropsychological Tests

Below are neuropsychological tests used in this research:-

1. *Graded Naming Test (GNT)*

The GNT was developed to provide an instrument with the potential to detect mild word retrieval difficulties and to monitor improvement or deterioration of naming abilities (Warrington and Mckenna, 1983). Participants had to identify an object shown on the screen. There were 30 different drawings, which became progressively more difficult.

2. CANTAB Paired Associate Learning (CANTAB PAL) test

CANTAB is the Cambridge Neuropsychological Automated Test Battery and the PAL is an assessment of visuo-spatial associative memory and requires the participant to learn an association between a visual stimulus and a spatial location (Égerházi et al., 2007). The test was delivered via a touch screen computer and took approximately 10 minutes to administer. The participant was presented with a screen showing six white boxes which opened up in a random order. Then, these participants were asked to touch the correct boxes where the pattern was located, with two, then three, followed by six different patterns. The number of errors made at the six pattern stage is of importance. This level is highly sensitive to MCI and early (Blackwell et al., 2004). In their paper, Blackwell et al. (2004) demonstrated that the number of errors made at the six pattern stage of the PAL in combination with the score on another well-established CANTAB test, the Graded Naming Test (GNT), can be used to predict a later diagnosis of probable AD with excellent accuracy. They created an algorithm using scores on these tests to identify converters to AD with 100% accuracy within a 32 month period for a sample of 40 patients.

3. Mini Mental State Examination (MMSE)

The MMSE is a quick test of general cognition, assessing numerous domains, such as orientation, short term memory, recall, language, attention, copying and verbal understanding (Folstein et al., 1975). The patient receives a score out of a possible 30 (See Appendix B). A score of 24 or above signifies no impairment is present, 18 -23 represents a mild impairment and 0-17 signifies a severe impairment in cognitive function. The test is particularly poor at detecting mild cases of impairment, which is problematic for patients who score on the borderline. Baseline scores in the MMSE can vary immensely; Blackwell et al. (2004) found that, in a group of AD convertors, baseline scores varied from 23 and 28.

The NICE (National Institute for Clinical Excellence) guidelines currently use the MMSE score to determine whether a patient receives pharmacological treatment. Currently, these guidelines state that drugs should be given only to those individuals with a moderate form of dementia. This means that a patient with mild impairment will not be treated until they continue to decline further and there are concerns as to whether or not this is ethical. Additionally, patients who receive drug treatment will have their medication removed when they decline past a score of ten, even if they find the treatment beneficial. Furthermore, the accuracy and reliability of the MMSE can be affected by mood, time of day, education and age (Jacova et al., 2007). The MMSE is perhaps best used as a method to exclude dementia rather than aiming to diagnose the degree of impairment the patient presents with.

4. Word Recall

Word recall is an assessment of short-term memory and typically involves reading the patient a list of common nouns and then asking them to repeat back as many as possible. Scores were out of 15, with one point awarded per correct word recalled.

5. Questionnaires

Participants were requested to complete a questionnaire on VREAD feedback. The VREAD feedback questionnaires were designed to capture information about user acceptance and satisfaction of the virtual environment (See Appendix C). The questionnaire assessed the following: whether the participants learned controls quickly, moved easily around the environment, were given clear instructions, found the system controls effective and the freedom of movement allowed. The results of this questionnaire are discussed in Chapter 5.

4.2.5 Results

Initial analysis of the data showed that, from a total of 31 participants who undertook the experiments, there was a mean age of 70.6 years and a mean length of time spent in education of 16.6 years. Results from the virtual simulation assessment tests were analysed using correlational, univariate and multivariate analysis of variance. Table 4.1 displays the mean and standard deviations for all neuropsychological tests. In addition, Table 4.2 demonstrates the mean and standard deviation obtained by each user for all levels in VREAD. Interestingly, there was a significant correlation between VREAD level 4 scores and scores on the neuropsychological tests, but not for other levels, as depicted in Table 4.2.

Table 4.1: The mean and standard deviations for all neuropsychological tests.

Test	N	MIN	MAX	Mean	S.D
MMSE	31	26	30	28.84	1.21
Word Recall	31	3	10	5.55	1.84
PAL 6 Errors	31	0	49	10.13	11.14
GNT	30	19	30	25.9	3.14

Table 4.2: The mean and standard deviation obtained by each user for all levels in VREAD

Level	N	MIN (%)	MAX (%)	Mean (%)	S.D.
1	31	60	100	97.90	8.34
2	31	6.41	100	93.00	23.42
3	31	41.67	100	95.06	12.98
4	31	19.35	100	94.60	16.92
5	31	56	100	89.80	13.56

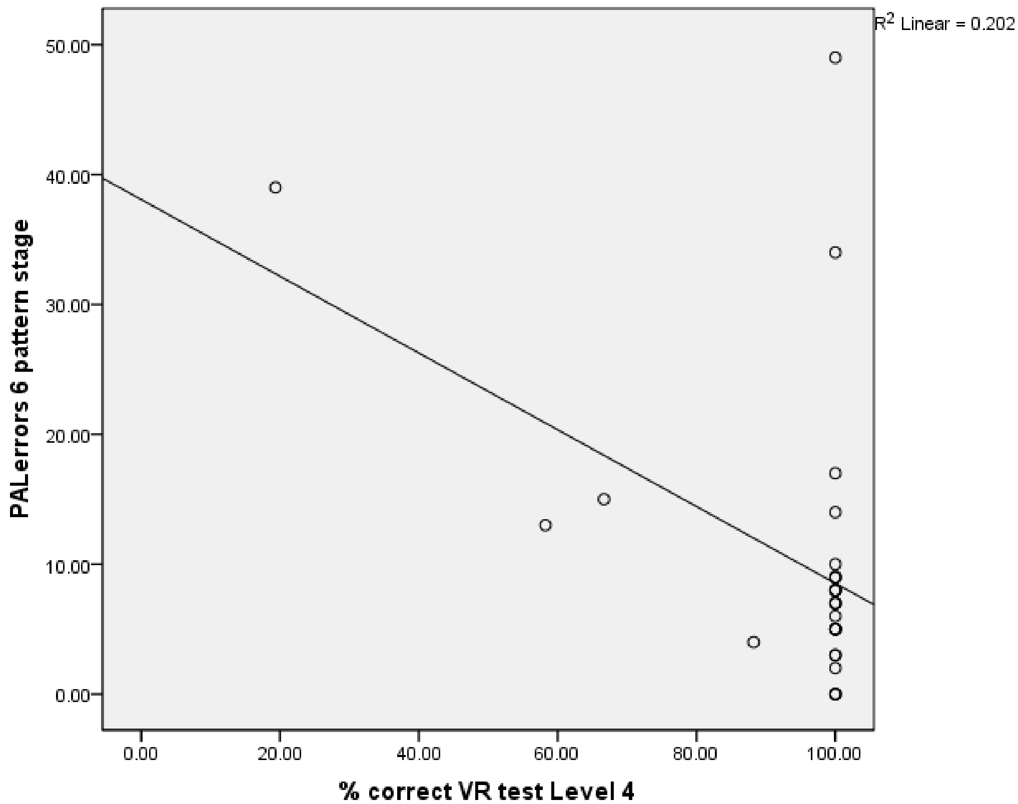


Figure 4.1: Graph showing correlation between scores on the 6 pattern stage of the CANTAB Pal and VREAD level 4.

There was a significant positive correlation between VREAD level 4 scores on both PAL six pattern stage and GNT, r=-0.45, p<0.02 and r=0.43, p<0.02 respectively. Figure 4.1 illustrates that, as the numbers of errors on the PAL 6 decreased, the performance on VREAD level 4 increased. From Figure 4.2, it can be noted that, as the numbers of correctly identified pictures on the GNT increased, VREAD level 4 also increased. Furthermore, from the results obtained, there was a trend towards significance for the word recall test at p=0.069.

A univariate analysis indicated no significant effect of age, gender and education on the scores on VREAD level 4. No significance was observed for the MMSE score. No other level significantly correlated with any of the tests.

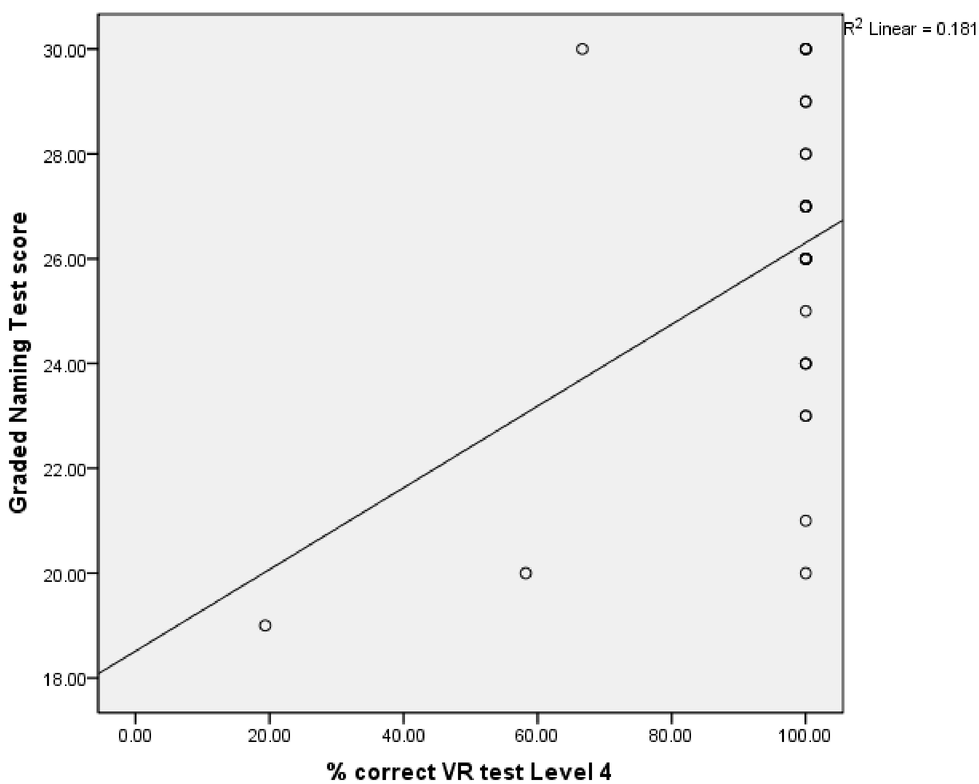


Figure 4.2: A graph that shows correlation between scores on the GNT and VREAD level 4.

4.2.5.1 Discrimination between Groups

From the results obtained, participants were diagnosed as either healthy elderly or MCI from the number of errors made on the PAL 6 pattern stage combined with the total score for MMSE and using the Blackwell algorithm. An independent sample t-test was used to analyse discrimination between the two groups. The result indicates a significant difference ($t(29)=3.17$, $p<0.01$), showing that VREAD level 4 can discriminate between the two groups. Figure 4.3 below shows that the mean score for VREAD level 4 was lower in the MCI group compared with the healthy group.

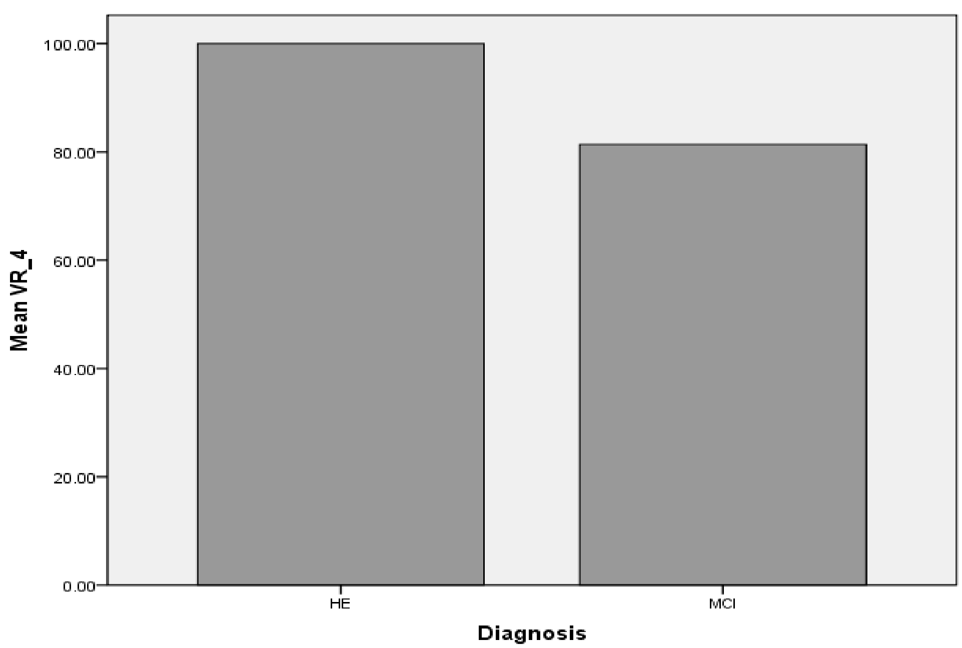


Figure 4.3: Graph showing the mean score for VREAD level 4.

4.2.6 Discussion

The main objective of this part of the study was to investigate whether VREAD could be used as a novel test of spatial memory as an aspect of cognitive function in the elderly. Furthermore, it was of interest to discover how this test correlates with existing neuropsychological tests. From the experiments conducted, each participant carried out all 5 levels in the module of VREAD and the results demonstrated that VREAD level 4 score significantly correlated with existing neuropsychological tests, namely the CANTAB PAL 6 pattern error score and the GNT, in the investigation of cognition in the elderly. It remains unclear why VREAD level 4 was the level which correlated with the scores on the neuropsychological tests. Each level became progressively harder, based on the number of turns and number of objects, such as trees, benches, street lights, fences and buildings. The overarching aim is to develop a simple test with one level; this is the next stage of our research.

Further analysis showed that there was also a trend toward significance for the word recall task. It is apparent from the results that there was no significant difference between VREAD level 4 and the MMSE score, which is important since the MMSE is not sensitive to mild degrees of impairment (Chertkow et al., 2007). Additionally, there was no association with the MMSE, which had previously been criticised for its inability to detect mild changes in cognition (Pasquier, 1999; De Jager et al., 2002). Jacova et al. (2007) have highlighted the fact that the MMSE remains insensitive to a mild degree of impairment. As navigational difficulties occur early in the disease pathway it would therefore seem unlikely that the MMSE would detect such impairments and this is reflected in the results of the current study.

The CANTAB PAL 6 pattern error score and the GNT, combined with age, were used in a study by (Blackwell et al., 2004). The three factors were put into a regression analysis and the algorithm was able to predict, with an accuracy of 100%, which participants converted to AD and those who remained in the category of questionable dementia over a 32 month period. The VREAD level 4 score correlated with the two tests and there was no correlation with age, suggesting that it could be used as a more straightforward method of testing cognition and cognitive decline.

An important requirement of neuropsychological tests should be their ability to correctly identify a person's clinical group. Scores obtained on VREAD level 4 of the VR assessment were significantly worse in those participants with MCI in comparison with the healthy group, confirming that this novel assessment can discriminate reliably between the two groups. This has implications for the use of this particular test as a potential screening device for the detection of MCI or early AD. It is clear from the results of this iteration that, as a number of participants scored 100%, more subtle impairment in participants could be obscured. The results do discriminate between groups, so it seems unlikely that this is a simple ceiling effect, but future studies will be making the virtual simulation test harder to avoid this.

In some previous investigations, researchers have raised concerns about the possible side effects of using a virtual environment in which some participants have reported feelings of disorientation or nausea during the task (Moffat et al., 2001). This was not problematic in the current study and there were no reports of illness, headaches or nausea, as assessed by the usability questionnaire. Participants were informed in the invitation letter that the study was unable to recruit those with uncorrectable visual impairments.

4.3 The Experiments of VREAD with Non-Elderly Participants

The purpose of this investigation was to determine the reliability and effectiveness of VREAD in a group of non-elderly participants. This constituted a first step prior to its implementation with elderly participants. Furthermore, the purpose is to investigate whether VREAD tests general memory function or pure spatial memory in non-elderly participants. Another rationale for testing among non-elderly participants was to get feedback regarding user acceptance and satisfaction for subsequent modification. Further explanations of usability in terms of user acceptance and satisfaction are in the next chapter.

The participants were students and staff of the University of Bradford from different departments. A total of 23 non-elderly participants were recruited for this experiment. Ethical approval was obtained from the University of Bradford Ethics Committee. All participants possessed no existing neurological disorder or existing memory impairment. Moreover, none of the participants had an uncorrectable visual impairment.

4.3.1 VE Experimental Apparatus: VREAD

The experimental environment operated on a standard PC and simulated two fully textured scenes. Virtual Reality for Early Detection of Alzheimer's Disease, VREAD, consisted of three modules: VR Practice, VR Park and VR Games.

4.3.2 Procedures

The experimental procedure was the same as discussed in Section 4.2.3.

4.3.3 Neuropsychological Tests

1. Word recall

Word recall is an assessment of short term memory. Sixty basic words were shown through the computer screen. After seeing the words on screen, participants were to write down as many words as they could in any particular order. Overall the score was out of 60, with one point per correct word recalled.

The neuropsychological test such as MMSE, CANTAB PAL and GNT were not used for the non-elderly cohort as these are standardised neuropsychological tests specifically designed for assessing cognition in the elderly.

4.3.4 Results and Discussion

The aim of this experiment was to examine, in non-elderly participants, whether this test was examining spatial memory alone or if it tested memory processing overall.

Table 4.3: The mean of the reaction time for all levels and mean scores of the word list

VREAD Reaction Time					**Word List**
L1	**L2**	**L3**	**L4**	**L5**	
22.39	26.83	39.83	41.74	17.00	22.39

Table 4.3 presents the mean score obtained for participants in each level of the VREAD score and the word list score. A statistical analysis was undertaken to investigate spatial memory between reaction times on VREAD level 4 and word recall test scores, as depicted in Table 4.3. The data was submitted to a multiple regression analysis using SPSS to see if the reaction times of any of the five levels could predict scores on the word recall test. None of the predictor variables was significant, suggesting VREAD was not looking at general memory function but was central to the spatial domain. As a result, it was important to ensure that this (the spatial hippocampus) was being tested in the elderly, rather than spatial memory processing, thus adding to the novelty of the virtual environment.

4.4 Summary

This chapter has described the two main experiments to examine whether VREAD is a reliable means of investigating cognitive function in a group of elderly participants and discriminating between MCI and healthy groups. The findings have demonstrated positive results since significant correlations have been found with existing neuropsychological measures, specifically the CANTAB PAL and the GNT for the detection of mild degrees of impairment. This suggests reliability for the early detection of AD. Results from non-elderly have verified that it is a test specific to spatial memory.

Chapter 5

Usability Evaluations of User Acceptance and Satisfaction

5.1 Introduction

The objective of this chapter is to evaluate the usability of the developed simulation, with a focus on user acceptance and satisfaction. Firstly, the reasons why ease of use is important are explained and then the methods used in the usability evaluation are discussed. In addition, the questionnaire and the selection of participants are also explained. This chapter then discusses the three usability experiments: (i) the pilot study, (ii) the study on non-elderly participants and (iii) the study on elderly participants. Following the discussion, this chapter also presents a descriptive analysis of the data obtained and discusses the results for each experiment.

5.2 Usability Evaluation

Usability evaluation is one of the most important quality characteristics to ensure that systems meet both design specification and user requirements. Various definitions for usability exist. A definition from ISO 9241 (1998), states usability as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*. One well-known definition has been produced by (Nielsen, 1993), who associates it with five parameters:-

i. *Learnability: it is easy to learn and get work done quickly with the system.*
ii. *Efficiency: a high level of productivity is possible once the user has learnt the system.*

iii. *Memorability: The casual user is able to use the system after not having used it for period of time, without having to learn everything over again.*

iv. *Error: The system should have a low error rate. Also, if users do make errors, they should easily recover from them.*

v. *Satisfaction: Users are subjectively satisfied by using the system. In other words, they like using the system.*

Preece et al.(2002) defined usability goals which include effectiveness, safety, utility, learnability and memorability.

- Effectiveness explains how good the system is in achieving the tasks.
- Efficiency describes how good the system is in helping and supporting users in their tasks.
- Safety refers to the protection of users from undefined situations such as when the user could not click on certain buttons.
- Learnability explains how the system was easy to use and learn.
- Memorability describes remembering the procedure in using the system once learned.

- Utility refers to how the system provides the right functionality or support to all kind of interactions to users.

Another study by Fatouhi-Ghazvini et al.(2007) embedded other usability factors in the games development. These included control, feedback and sense of danger. All these elements are important because the element of feedback describes whether the task has been completed successfully, and the sense of danger refers to the traps or surprises that users should be ready for. Additionally, the element of control refers to how learners can control the game and learn from it.

Usability testing is critical to the success of any prototype (Olsen et al., 2011; Stoddard et al., 2006; Carvalho, 2001). Clearly, a usability evaluation is important because it can identify the efficacy of the virtual simulation, as well as any problems. Many studies have been conducted on the usability testing of prototypes.

Kizony et al. (2006) conducted usability testing on a virtual rehabilitation system for the elderly, looking at users' sense of presence, the perceived difficulty of the task and any discomfort that users may have felt. Their results were encouraging in terms of enjoyment and usability. Another study, conducted by Wilson et al. (2007) on a stroke rehabilitation system, emphasized the importance of usability testing for understanding the capabilities, limitation and needs of users. The results of this study contributed to their future development of the system. Another relevant study, by Fitzgerald et al. (2008), carried out a usability evaluation of a virtual rehabilitation system. In their study, the usability testing was to identify errors or problems and obtain user suggestions. The results of their study provided them with information to guide them in the development of system alterations.

There are numerous existing methods for usability evaluation but the outcome should provide some reports on the effectiveness, efficiency or user satisfaction. Some of the evaluation methods include cognitive walkthrough, formative evaluation, heuristic evaluation, summative evaluation (Bowman et al., 2002), questionnaire and formal experiment (Andrews, 2008). Table 5.1 summarises some of the evaluation methods mentioned.

Table 5.1: Summarisation of Evaluation Methods

Evaluation Methods	**Description**
1. **Cognitive Walkthrough**	It requires several expert evaluators. Check-list approach to learnability and ease of learning
2. Formative Evaluation	Identifies problems before being implemented.
3. Summative Evaluation	Assesses the usability of a prototype after it has been developed.
4. Heuristic Evaluation	Expert evaluates, examines and judges the interface using check-list provided.
5.Questionnaire	Set of questions given to users after involvement in the experiment. The questionnaires require rating of several issues on a series of scales.
6. Formal Experiment	Samples of users are recruited to do the tasks. The data collected will be analysed.

5.3 Methodology of Usability Evaluation for the Research

The usability evaluation of this simulation was broken into two different evaluation stages and certain modifications were undertaken. A pilot test was conducted first, using user based evaluation from five experimental users. The findings and feedback of this pilot study were then analysed and changes made accordingly. Next, the experiment was carried out again with non-elderly participants before it was eventually used on elderly people. All of these experiments were based on user evaluation.

There were five stages involved in the methodology for this evaluation (see Figure 5.1).The first stage was analysis of usability evaluation, including planning and identification of the method to be used. For this research, usability evaluation of VREAD was done through formal experiment and questionnaires. The next stage involved setting the materials. During this stage, one of the main tasks was the recruitment of participants. Other tasks included setting the test environment and preparing the questionnaire. Next, the experiment was conducted. For this stage, the main task was to organise the experiment itself. Ethical issues for participants were also considered at this point by explaining and giving participants the information sheet, instruction sheet and consent forms. The next stage of the process was to provide the usability questionnaire for participants to fill in. Lastly, the final stage was to analyse data collected from the questionnaire and from the experiment.

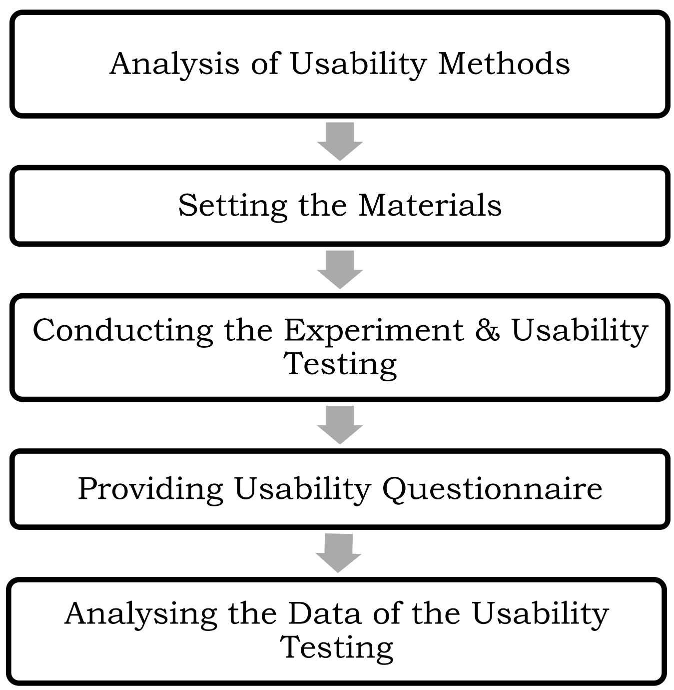


Figure 5.1: Methodology of usability evaluation.

5.4 Pilot Study

A pilot study was undertaken first in order to ensure effectiveness, efficiency and user satisfaction of this simulation before implementation. The participants recruited for pilot testing were experimental users who were experts in the use of computers, with experience of playing virtual games. This was to enable these participants to compare this simulation to the ones they had experience of. A pilot test was conducted with five healthy participants to investigate the following:-

- System errors.
- User satisfaction.
- Clear instructions.
- Whether the prototypes are easy to use and function correctly.
- Whether the data are tracked and recorded correctly.

Table 5.2: Results obtained from questionnaire.

	Female Age : 29	Female Age : 35	Male Age : 19	Female Age : 34	Male Age : 33
1. Learned controls quickly	7	4	6	7	7
2. Moved easily	6	5	6	6	6
3.Clear instructions	4	4	6	6	7
4. System control	5	5	6	6	6
5. Preferred more freedom	3	3	2	2	2
6. Looking around easily	6	5	6	6	7

After the experiment, these users were given a questionnaire. The questionnaire focused on control, movement, instructions and ease of use. This 7-point Likert scale has a value of 7-Strongly Agree, 6-Agree, 5-Slightly Agree, 4-Undecided, 3-Slightly Disagree, 2-Disagree and 1-Strongly Agree (Carifio and Perla, 2007). The results of the questionnaire can be seen in Table 5.2. From Table 5.2 it can be noted that the female test participants obtained a low 'learned control quickly' score of 4, while the others mainly scored between 6 and 7. They were all able to learn the controls fairly quickly, which is represented by scores of between 5 and 6. Only two participants were not very clear about the instructions given, scoring 4, while the others scored between 5 and 7. They liked using the system controls, as represented by the scores [5 to 6]. They all preferred the degree of freedom of movement, with some restricted and limited access areas, and were less happy if given more freedom of movement. The scores ranged from 2 to 3. All the participants also agreed that it was easy to look around, as represented by the scores [5 to 7].

Usability evaluation for this pilot test identified issues concerning the systems and provided suggestions for improvements in the virtual simulation. Following the pilot testing, some modifications were undertaken on the basis of the recommendations. The modifications were that some of the 3D models were reconstructed to give a more 'real' environment with level 3 and level 4 paths had to be adjusted to be more suited to each level. Furthermore, some instructions were amended and some programming errors were fixed so that the correct scores were displayed on the system.

5.5 Non-Elderly Participants

After the modifications resulting from the pilot testing, the second phase of the experiment was carried out with non-elderly participants.

A total of 23 participants, comprising 10 males and 13 females aged below 65 years, were recruited for this experiment. The participants were students or staff of the University of Bradford with different specialisations. All controls possessed no existing neurological disorder or existing memory impairment. Furthermore, none of the participants had an uncorrectable visual impairment. The study was approved by the Board of Ethics, University of Bradford. The participants were included in the study after signing a consent form.

This experiment produced results that measured usability. Table 5.3 provides results from the questionnaire. From the results, it can be observed that learning to move through the application using the cursor keys on the keyboard obtained an average score of 4.57. A possible explanation for this was that users were asked to use VR Practice first before going through the testing modules. The users were required to move through the application without any help from researchers, using the control keys on the keyboard.

Table 5.3: Findings from the questionnaire.

Five-Point Likert Scale : 1-Completely agree; 2-Disagree; 3-Undecided; 4-Agree; 5-Strongly agree	
Categories	Average out of 5
1. Learned to move quickly through application	4.57
2. Moved easily	4.48
3. Clear instructions	4.83
4. Control using arrow keys	4.17
5. Feel disoriented	1.26
6. Looking around easily	4.22
7. Preferred more freedom	3.09
8. Fast speed of movement	1.61
9. Comfortable using application	4.74
10.Overall Satisfaction	4.52

The general purpose of this research was to identify sufferers of MCI; therefore instructions had to be simple to remember. From the results, it shows that an average score of 4.48 was given for agreeing that clear instructions were given throughout this application. It was found that an average score of 4.17 was given for participants being able to use the control keys for navigation and interaction without any problem. Disorientation is one of the main health and safety implications arising from the use of VR. The overall response to this question was very positive as participants did not feel disoriented while using the application, which had a score of 1.26 on average.

Participants agreed that it was easy to look around in the application, giving an average score of 4.22. An average score of 3.09 was given for the question whether they would prefer to have more freedom. This may be because users were limited to walking on the path without going on the grass and other places. The reason for limited movement was to avoid the problems of users feeling disoriented or getting lost while using the application. The average score was 1.61 in response to the speed of movement, which was a positive result. For both questions of overall satisfaction and comfort whilst using the application, the overall response was positive, receiving average scores of 4.74 and 4.52 for each response.

This evaluation focused on user satisfaction for non-elderly participants, looking at the complexity and effectiveness of the prototype prior to the implementation of the experiment with elderly individuals and those with a mild cognitive impairment. The findings demonstrated that the application was reliable and suitable for use by elderly participants.

5.6 Elderly Participants

The aim of this experiment was to investigate whether this simulation could be used as a novel test for the early detection of MCI in a group of over 65 year olds. The group study consisted of elderly participants. All controls possessed no existing neurological disorder or existing clinically diagnosed memory impairment. Furthermore, none of the participants had an uncorrectable visual impairment. All participants were recruited from the University of Bradford Psychology Department Participant Pool. This is a database of participants who volunteered their details in order to take part in psychology research. They were recruited via local community centres and groups. The study was approved by the Board of Ethics, University of Bradford. The participants were included in the study after signing a consent form.

Table 5.4: Feedback results from questionnaires.

Five-Point Likert Scale : 1-Completely agree; 2-Disagree; 3-Undecided; 4-Agree; 5-Strongly agree	
Categories	**Average Out of 5**
1. Learned to move quickly through application	4.4
2. Moved easily	4.3
3. Clear instructions	4.8
4. Control using arrow keys	4.1
5. Feel disoriented	2.0
6. Looking around easily	4.0
7. Preferred more freedom	2.7
8. Fast speed of movement	1.9
9. Comfortable using application	4.6

Usability evaluation of the prototype was vital to ensure that the systems met user requirements and the system specification. User evaluation study with the participants was done to evaluate the usability of our prototype system. Following the experiment, a usability evaluation questionnaire was completed by participants. A usability evaluation was then performed on the questionnaire given to participants. The questionnaire was based on participants' responses spanning five levels. The responses are summarised in Table 5.4.

Below are the results by categories from the questionnaire:-

i. Speed and Movement

The results indicated that evaluation of movement returned responses with high scores. Learning how to move quickly through the application scored 4.4 out of 5 on average and the ability to move easily through the application scored an average of 4.3 out of 5. From the results, the participants seemed to disagree with the suggestion that movement was too fast since this response scored an average of 1.9 out of 5. For the response concerning the use of the arrow keys, an average score of 4.1 out of 5 was given. Furthermore, an average of 2.7 out of 5 was given for the responses concerning freedom of movement.

ii. Instructions

The participants' responses about clear instructions were positive, scoring an average of 4.8 out of 5.

iii. Disorientation

The findings for the feeling of disorientation showed an average score of 1.7 out of 5. This positive result shows that the participants were not confused while using the virtual simulation. Overall, participants' responses were 4.6 for feeling comfortable using the virtual simulation and an average of 4.8 out of 5 for overall satisfaction. The findings demonstrated that the virtual simulation was reliable and suitable for use by elderly participants.

iv. Feedback Comments

When participants answered the usability questionnaire, they gave some feedback or comments regarding the application. Table 5.5 shows the positive and negative comments made by elderly participants towards the application. All the comments and suggestions will be carried forward for consideration in future work.

Table 5.5: Feedback comments reported during evaluation.

No	***Participants Comments***
1.	It was fun.
2.	Interesting to play.
3.	Not able to know where I was on the path, no point of reference, so took away my concentration to where I was going.
4.	Two practices would have been sufficient. Boredom and lack of concentration sets in with 3rd practice.
5.	Trying to keep on the ribbon distracted me from remembering the route. In normal circumstances, scenery helps.
6.	Not adept at using cursor keys, would have preferred a mouse.
7.	Knowledge of keyboard is a big advantage. Also knowledge of driving.
8.	I thought it was fun but not being computer literate, the movement round the circles was a bit difficult.
9.	Concentrating on ribbon, couldn't look around.
10.	Cornering was hard, difficult for left handers.

The findings are important as a step prior to implementation of the application for elderly people with MCI. The findings demonstrate that it is feasible to work in a virtual environment with elderly people and the results also indicated a high level of usability in all categories evaluated.

5.7 The Comparison Study of Age between Non-elderly and Elderly Participants

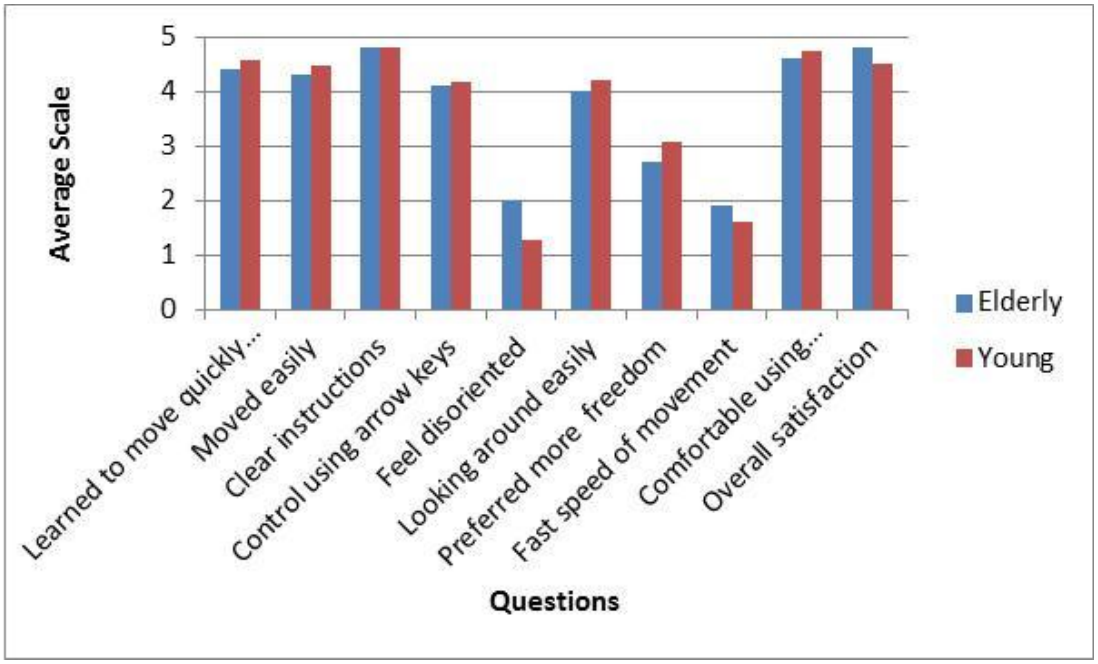


Figure 5.2: Age comparison between non-elderly and elderly participants

The comparison of usability evaluations between the two age groups can be found in Figure 5.2. It is noticeable from Figure 5.2. that a negligible difference can be seen between the two age groups. From the findings, it had been expected that young participants would prefer more freedom although the results were positive. A possible explanation for this might be that young participants are exposed to and have more experience of online or virtual games.

Another finding that showed little difference was the feeling of disorientation although the results were encouraging.

5.8 Summary

This chapter has described the three experiments which focused on usability evaluation in the research. In this chapter, usability evaluation of user acceptance and satisfaction were determined by young and elderly participants. The results from the pilot test provided essential insight into usability issues, allowing subsequent modifications and amendments to the simulation and system to be made. The findings have demonstrated that it is feasible to work in a virtual environment with elderly people and the results have also indicated a high level of usability in all categories evaluated.

Chapter 6

VREAD Prediction: Experiments and Results

6.1 Introduction

Results collected from the VREAD experiment of the whole cohort of elderly people together can be further analysed in order to predict the subgroups of those with MCI and those healthy elderly. VREAD can be considered as a reliable test to discriminate between MCI and healthy elderly using data mining techniques applied to the results for prediction. The aim of this chapter is to achieve the highest performance measure of accuracy in predicting VREAD dataset using data mining techniques. A range of data mining techniques are used in this study, including Neural Network, Decision Trees, K-Nearest Neighbour, Ensemble and Bayesian algorithm. In this chapter, VREAD Knowledge Discovery Databases (KDD) processes are discussed, followed by a comparative study of different feature selections and classifiers. In addition, the proposed weighting function to improve performance measures is also explained.

6.2 VREAD Knowledge Discovery Process

The KDD comprise a vast domain with the purpose of extracting hidden knowledge from the collection of data. Data mining is one of the processes in KDD that uses machine learning algorithms to extract patterns from the data. In recent years, KDD has been used in many applications including, medicine (Lavrač and Zupan, 2010), financial applications (Kovalerchuk and Vityaev, 2010), and in customer relations management (Thearling, 2010).

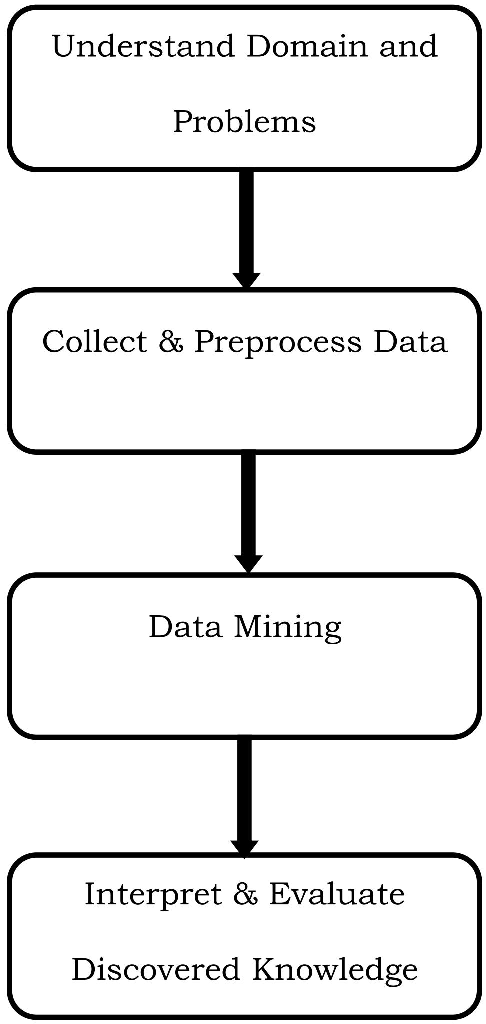


Figure 6.1: VREAD KDD process

For this research, KDD is used as a method to discriminate between healthy elderly and MCI. The VREAD KKD process can be summarised as in Figure 6.1, based on Fayyad et al. (1996) KDD Process.

6.2.1 Understand Domain and Problems

The most important process in VREAD KDD is to understand the problems of the research. The problem formulated was to predict MCI and healthy elderly.

6.2.2 Collect and Preprocess Data

This phase includes data collection, data cleaning and data selection. Firstly, data was collected automatically from VREAD. These data have six main attributes including: correct path, incorrect path, correct sequences, incorrect sequences, overall score and time. The sample group consisted of 31 elderly participants. The next step involved data cleaning, which checks missing data and derives new attributes.

6.2.3 Mining VREAD Dataset

Data mining is a process of extracting knowledge or patterns from a collection of data. It is based on pattern recognition techniques and statistical techniques. Data mining software is used as an analytical tool for analysing data. WEKA is a data mining tool that offers a comprehensive collection of machine learning algorithms and data pre-processing tools (Hall et al., 2009). The WEKA package can be included in Java sources and runs in various platforms. For the end users, WEKA software is easy to use since it has a graphical user interface and the algorithms can either be applied directly to a dataset or they can be called from your own Java code.

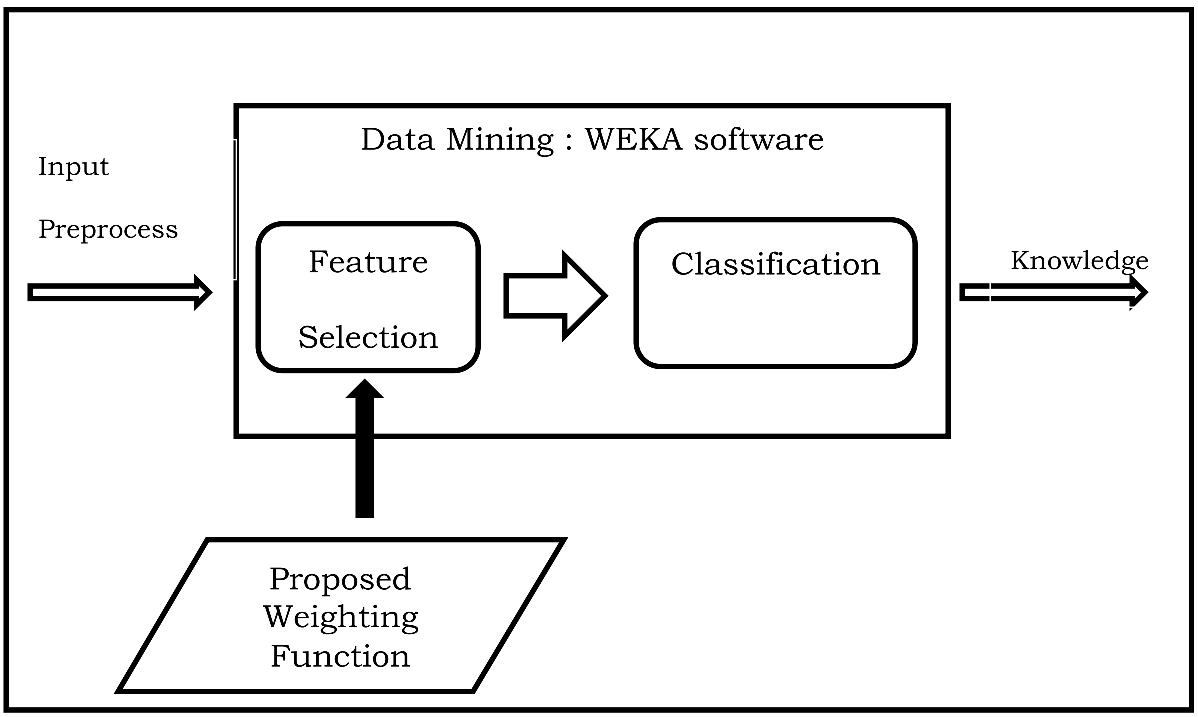


Figure 6.2: Data mining process for this study.

For this study, the WEKA package was used for prediction and to discriminate between MCI and healthy elderly people. All of the functions inherited from the WEKA package were very reliable. Functions that were used in this research were the select attribute (feature selection) and classification (see Figure 6.2).

Figure 6.2 illustrates the process of data mining using the WEKA software. First, for the pre-processing, a dataset was input to the WEKA. Table 6.1 describes the VREAD dataset used for this study, 31 instances and no missing values found. Following this, there is a process in the WEKA software which involves feature selection and classification. Another process involved in this stage is weighting, whereby performance in selecting attributes is improved. This will be discussed in a later section.

Table 6.1: VREAD dataset description

Dataset	**Number of Instances**	**Missing Value**
VREAD Dataset	31	No

i. Feature Selection

The first phase is feature selection, which is a process that chooses a subset of attributes relevant to the whole dataset analysis. In addition, it will improve the results by getting higher classification accuracy. The feature selection process will increase the accuracy of the classification by reducing irrelevant attributes in the dataset (Luukka, 2011). The functions used for the attribute evaluators (feature selection) within this research are as follows (Hall et al, 2009):

- Cfs Subset Evaluator

  It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.

- Info Gain Attribute Evaluator

  It evaluates the worth of an attribute by measuring the information gain with respect to the class.

- Gain Ratio Attribute Evaluator

  It evaluates the worth of an attribute by measuring the gain ratio with respect to the class.

By using the feature selection process, it can reduce noise and insignificant attributes in the training dataset and therefore improve the classification accuracy. For this study, feature selection in the WEKA package was used to determine the important attributes in predicting elderly people with MCI. Table 6.2 shows the results obtained from the preliminary analysis using the WEKA package with different attributes and evaluator methods. From Table 6.2, it can be summarised that attribute VREAD level 4 score or VR4 and other attributes related to VR4 were always selected as the top attributes. This result may be explained by the fact that VREAD level 4 score is significantly correlated with CANTAB PAL 6 Errors and GNT, as discussed earlier in Chapter 4. These selected attributes are used to discriminate between healthy elderly and MCI, as discussed in the next section.

Table 6.2: Results of selected attributes using feature selection provided by the WEKA package

Feature Selection	
Attributes Evaluator and Search Method	Selected Attributes
Cfs Subset Evaluator and Best First	• VR4 • VR4IncorrectPoint • VR4IncorrectSeq • VR5CorrectSeq • VR5IncorrectSeq
Gain Ratio Attribute Evaluation and Ranker	• VR4IncorrectPoint • VR4IncorrectSeq • VR4 • VR1IncorrectSeq • VR1
Info Gain Attribute Evaluator and Ranker	• VR4IncorrectPoint • VR4IncorrectSeq • VR4 • VR5IncorrectSeq • VR1IncorrectSeq

ii. Classification

The next phase in a data mining process is classification. Classification is a data mining function that accurately predicts group membership and data instances. For the machine learning algorithms, the classifier functions that apply within this thesis were as follows:-

• IBK (K-nearest neighbours classifier)

K-nearest neighbours classifier. Can select appropriate value of K based on cross-validation. Can also do distance weighting.

• J48 (Decision trees)

Class for generating a pruned or un-pruned C4.5 decision tree.

• Naïve Bayes (Naive Bayesian)

Class for a Naive Bayes classifier, using estimator classes. Numeric estimator precision values are chosen based on analysis of the training data.

• Bagging (Ensemble)

Class for a Bagging classifier to reduce variance. Can do classification and regression, depending on the base learner.

• Multilayer Perceptron (Neural Network)

A classifier that uses backpropagation to classify instances. This network can be built by hand, created by an algorithm or both.

These different classifiers from different categories are applied so that a comparison can be made, which therefore enables identification of the most accurate classification algorithm.

6.3 Comparative Study of the Data Mining Technique to the VREAD Dataset

In this section, investigation of the use of different data mining techniques to get the best results of performance measure in prediction is explained. From the statistical analysis discussed in

Chapter 4, it was found that MMSE did not correlate with any of VREAD scores. These findings are in agreement with (Jacova et al., 2007; Money et al., 2009; Zadikoff et al., 2008), who have emphasized that MMSE was found not sensitive to mild degree impairment. Therefore, to verify the findings as stated above, this study compares prediction using VREAD dataset but with two different class distributions. One experiment used VREAD dataset with class distribution of 9 classified as MCI out of 31 participants, which was diagnosed by MMSE, CANTAB PAL 6 pattern error and Algorithm Blackwell. Another experiment used VREAD dataset with class distribution of 8 classified as MCI out of 31 participants, which was diagnosed by CANTAB PAL 6 pattern error and Algorithm Blackwell.

A range of data mining techniques, including decision trees, K-nearest neighbour classifier, Bagging, Multilayer Perceptron and Naïve Bayes were selected from the WEKA package. Table 6.3, 6.4 and 6.5 summarise the results achieved using the WEKA package when trained using 10-fold cross-validation. In Tables 6.3, 6.4 and 6.5, there are three performance measures presented, which include Accuracy, TPR (True Positive Rate) and TNR (True Negative Rate), along with feature selection and classifiers. These three performance measures were chosen because these criteria are the most important measures in verifying the performance of classifiers' algorithms for MCI prediction.

Accuracy is the total of correct observations for each class. In addition, TPR is the correct observation for class Yes. In this study, TPR is the rate of correct classifications for the MCI class. In contrast, TNR is the correct observation for class No and for this study, is the correct classification for the Non-MCI patients.

Table 6.3 illustrates the results of the three performance measures obtained from the WEKA package using the CFs Subset Evaluator as Feature Selection and by using the same dataset but with different class distribution. It can be seen in Table 6.3 (a) that Bagging Classifier gives the highest accuracy, which is 0.84 and the highest TPR by the Naïve Bayes and IBK Classifiers, which was only 0.5. Furthermore, the highest TNR was 1, obtained by the Bagging Classifier. From the results depicted in Table 6.3 (b), it can be demonstrated that Naïve Bayes Classifier gives the highest measure for accuracy of 0.81 and TPR of 0.56. In addition, the highest TNR was also 1, obtained by J48, Bagging and Multilayer Perceptron Classifiers.

Table 6.3: Results of Accuracy, TPR and TNR obtained from the WEKA Package using CFs Subset Evaluator as feature selection with (a) using dataset with 8 MCI diagnosed (b) using dataset with 9 MCI diagnosed.

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.81	0.50	0.91
J48	0.81	0.38	0.96
IBk	0.77	0.50	0.87
Bagging	0.84	0.38	1.00
Multilayer Perceptron	0.81	0.38	0.96

Table 6.3(a)

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.84	0.56	0.96
J48	0.81	0.33	1.00
IBk	0.74	0.22	0.96
Bagging	0.74	0.11	1.00
Multilayer Perceptron	0.71	0.11	1.00

Table 6.3(b)

Table 6.4: Results of Accuracy, TPR and TNR obtained from the WEKA Package using Gain Ratio Attribute Evaluator as feature selection with (a) using dataset with 8 MCI diagnosed (b) using dataset with 9 MCI diagnosed.

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.87	0.50	1.00
J48	0.84	0.38	1.00
IBk	0.84	0.38	1.00
Bagging	0.81	0.25	1.00
Multilayer Perceptron	0.81	0.25	1.00

Table 6.4(a)

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.84	0.44	1.00
J48	0.81	0.33	1.00
IBk	0.81	0.33	1.00
Bagging	0.77	0.22	1.00
Multilayer Perceptron	0.77	0.22	1.00

Table 6.4(b)

As shown in Table 6.4(a), Naïve Bayes Classifier technique is shown to have the greatest accuracy, which was 0.87 and 0.50 for TPR. Naïve Bayes Classifier also shows the greatest accuracy, which was 0.84 and 0.44 for TPR for Table 6.4 (b). TNR for all classifiers obtained outstanding results of 1, as can be seen in Table 6.4

The other results obtained from the WEKA package using the Info Gain Attribute Evaluator as Feature Selection can be seen in Table 6.5 (a). Table 6.5(a) shows that the Naïve Bayes and Bagging Classifier have the greatest accuracy, which was 0.84. Moreover, the Naïve Bayes and IBk Classifiers presented the highest TPR of 0.5, and the Bagging Classifier illustrates the highest TNR of 1. As seen in Table 6.5 (b), Naïve Bayes and J48 Classifier achieved the greatest accuracy of 0.81. Also, for this result, the highest TNR was 0.44 by Naïve Bayes and IBk Classifiers. The highest TNR obtained was 1 by J48 and Bagging Classifiers.

Table 6.5: Results of Accuracy, TPR and TNR obtained from the WEKA Package using Info Gain Attribute Evaluator as feature selection with (a) using dataset with 8 MCI diagnosed (b) using dataset with 9 MCI diagnosed.

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.84	0.50	0.96
J48	0.81	0.38	0.96
IBk	0.81	0.50	0.91
Bagging	0.84	0.38	1.00
Multilayer Perceptron	0.74	0.25	0.91

Table 6.5(a)

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	0.81	0.44	0.96
J48	0.81	0.33	1.00
IBk	0.77	0.44	0.90
Bagging	0.74	0.11	1.00
Multilayer Perceptron	0.74	0.22	0.96

Table 6.5(b)

As a conclusion, it can be seen from Tables 6.3, 6.4 and 6.5 that all techniques achieved an accuracy of between 0.74 and 0.87. Additionally, the findings illustrate that the use of Gain Ratio Attribute Evaluator as feature selection shows the highest performance measure of 87% accuracy. Furthermore, from the comparison in study presented, it is shown that the dataset with a class distribution of 8 MCI classified obtained a higher result. In other words, the findings proved that the use of MMSE is not highly sensitive to the detection of MCI. As a result, the following sections will use only those dataset with a class distribution of 8 MCI out of 31 participants.

Early results generated by these methods demonstrated that there is a need for more sophisticated methods of pre-processing data. The weighting has been proposed to improve performance accuracy, TPR and FNR, which will be discussed in the next section.

6.4 Proposed Weighting Function to Improve Performance Measures

Weighting can be used in pre-processing to assign important features with higher weights than others and to assign less important features with lower weights. These weights can be assigned on the basis of domain knowledge of relative features or they can be determined automatically.

The results can be simplified by giving a weight to each score obtained for each level. The objective in giving the score is to give more weight to certain variables. In this study, the weights were assigned to the score for each level.

In Equation 6.1, the user can use the weights of the score for each level by embedding them to calculate a new score as follows:

$$Score_{Weight} = (w_1 * VR1) + (w_2 * VR2) + (w_3 * VR3) + (w_4 * VR4) + (w_5 * VR5) \quad (6.1)$$

$Score_{Weight}$ is a Score for a VREAD Simulation where;

VR1, VR2, VR3, VR4 and *VR5 are the scores for each level,*

w_1, w_2, w_3, w_4 and w_5 are the weights for *VR1, VR2, VR3, VR4* and *VR5* respectively.

The values of w_1, w_2, w_3, w_4 and w_5 are between 0 and 1 and the sum of ($w_1 + w_2 + w_3 + w_4 + w_5$) *is 1.*

Table 6.6: Example of the VREAD score for each level with User ID

User ID	**VR1**	**VR2**	**VR3**	**VR4**	**VR5**
1	100	100	100	90	100
2	100	100	100	100	90
3	100	100	100	90	90

Consider Table 6.6 as an example of the score for each level in the VREAD simulation. The $Score_{Weight}$ can be calculated for VREAD dataset by applying Equation 6.1. Let us say that the VREAD level 4 score and VREAD level 5 scores are the most significant attributes to determine the MCI. Thus, the weight for w_4 is 0.5 and w_5 is also 0.5. The rest of the weights are 0. Thus, for User ID 1, the $Score_{Weight}$ = ((0*100)+(0*100)+(0*100)+(0.5*90)+(0.5*100)) = 95. Table 6.7 shows the $Score_{Weight}$ calculated using Equation 6.1 where different values have been given to w_1, w_2, w_3, w_4 and w_5. The comparison between different values of weight from Table 6.7 will be discussed in the following section.

Table 6.7: $Score_{Weight}$ given different values for w_1, w_2 , w_3, w_4 and w_5.

UserID	$\mathbf{Score_{WeightA}}$	$\mathbf{Score_{WeightB}}$	$\mathbf{Score_{WeightC}}$
	w_1=0.0, w_2=0.0, w_3=0.0, w_4=0.7, w_5=0.3	w_1=0.0, w_2=0.0, w_3=0.0, w_4=0.5, w_5=0.5	w_1=0.2, w_2=0.2, w_3=0.2, w_4=0.2, w_5=0.2
1	93	95	98
2	97	95	98
3	84	80	92

6.5 Comparison Results of Different Weighting Values to VREAD Dataset

For this study, the weights are assigned to the important features based on statistical analysis, as discussed earlier in Chapter 4. From the statistical analysis discussed in Chapter 4, the variable VREAD level 4 scores seem to have significant correlation followed by VREAD level 5 scores. Therefore, these two variables are considered as the most important variables and the weightings are assigned to these variables.

Based on the discussion above, different weighting was implemented to VREAD datasets, as seen in Table 6.8. The results indicate that $Score_{WeightA}$ shows the most significant results in improving performance measures, where weighting of 0.7 was assigned to w_4 and weighting of 0.3 was assigned to w_5. A justification w_4 was assigned as the uppermost weighting to VREAD Level 4 scores because of its significant results, followed by w_5, since the significant impact of VREAD level 4 scores was greater than VREAD level 5 scores, as discussed earlier in Chapter 4.

Table 6.8: $Score_{Weight}$ given different values for w_1, w_2, w_3, w_4 and w_5.

Classifiers	$Score_{WeightA}$			$Score_{WeightB}$			$Score_{WeightC}$		
	w_1=0.0, w_2=0.0, w_3=0.0, w_4=0.7, w_5=0.3			w_1=0.0, w_2=0.0, w_3=0.0, w_4=0.5, w_5=0.5			w_1=0.2, w_2=0.2, w_3=0.2, w_4=0.2, w_5=0.2		
	Acc	TPR	TNR	Acc	TPR	TNR	Acc	TPR	TNR
Naïve Bayes	**0.90**	**0.75**	**0.96**	0.87	0.63	0.96	0.77	0.38	0.91
MultiPerceptron	0.87	0.63	0.96	0.87	0.63	0.96	0.74	0.25	0.91
Bagging	0.77	0.38	0.91	0.81	0.50	0.91	0.68	0.00	0.91
IBk	0.81	0.50	0.91	0.84	0.63	0.91	0.68	0.25	0.83
J48	0.87	0.50	1.00	0.87	0.63	0.96	0.74	0.25	0.91

The detailed explanations of the implementation of $Score_{WeightA}$ in improving performance are discussed in the next section.

6.6 The Implementation of Weight Functions to Improve Performance Measures

With reference to Table 6.8, this section details the implementation of $Score_{WeightA}$ in improving performance measures. The same five classifiers, including Naïve Bayes, Bagging, IBK, J48 and Multilayer Perceptron, were used for this experiment.

In this section, the detailed results for the Naïve Bayes algorithm are presented. The extracted outputs from the Naïve Bayes results, consisting of ‘detailed accuracy by class’ and ‘confusion matrix’, can be found in Figure 6.3. The overall accuracy for the Naïve Bayes algorithm was 0.90, while the TPR was 0.75 and the TNR was 0.96. From the confusion matrix, 6 out of 8 MCI were correctly classified, and 22 out of 23 non-MCI were correctly classified.

```
=== Run information ===
Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     VRdataelderly21022012lattributes42-weka.filters.unsupervised.attribute.Remove-R1-6,8-42
Instances:    31
Attributes:   2
              weightage
              Class
Test mode:    10-fold cross-validation
=== Summary ===
Correctly Classified Instances          28               90.3226 %
Incorrectly Classified Instances         3                9.6774 %
Kappa statistic                          0.7365
Mean absolute error                      0.1452
Root mean squared error                  0.3066
Relative absolute error                 36.8491 %
Root relative squared error             69.3847 %
Coverage of cases (0.95 level)          90.3226 %
Mean rel. region size (0.95 level)      72.5806 %
Total Number of Instances               31
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
                 0.75      0.043     0.857      0.75      0.8         0.769      yes
                 0.957     0.25      0.917      0.957     0.936       0.769      no
Weighted Avg.    0.903     0.197     0.901      0.903     0.901       0.769
=== Confusion Matrix ===
  a   b   <-- classified as
  6   2 |  a = yes
  1  22 |  b = no
```

Figure 6.3: Extracted output results using the Naïve Bayes Algorithm

The detailed results for the J48 algorithm are presented in Figure 6.4. The extracted outputs from the J48 results, which consist of 'detailed accuracy by class' and 'confusion matrix', can be found in Figure 6.4. The overall accuracy for the J48 algorithm was 0.88, while the TPR was 0.5 and the TNR was 1. From the confusion matrix, 4 out of 8 MCI were correctly classified, and all 23 non-MCI were correctly classified.

```
=== Run information ===
Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     VRdataelderly21022012lattributes42-
weka.filters.unsupervised.attribute.Remove-R1-6,8-42
Instances:    31
Attributes:   2
              weightage
              Class
Test mode:    10-fold cross-validation
=== Summary ===
Correctly Classified Instances         27               87.0968 %
Incorrectly Classified Instances        4               12.9032 %
Kappa statistic                         0.5974
Mean absolute error                     0.202
Root mean squared error                 0.3395
Relative absolute error                51.2616 %
Root relative squared error            76.8293 %
Coverage of cases (0.95 level)        100      %
Mean rel. region size (0.95 level)     93.5484 %
Total Number of Instances              31
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.5       0          1          0.5     0.667       0.617     yes
                 1         0.5        0.852      1       0.92        0.617     no
Weighted Avg.    0.871     0.371      0.89       0.871   0.855       0.617
=== Confusion Matrix ===
  a   b   <-- classified as
  4   4 |  a = yes
  0  23 |  b = no
```

Figure 6.4: Extracted output results using the J48 Algorithm

The detailed results for the IBk algorithm are presented in Figure 6.5. The extracted outputs from the IBk results, which consist of 'detailed accuracy by class' and 'confusion matrix', can be found in Figure 6.5. The overall accuracy for the IBk algorithm was 0.81, while the TPR was 0.5 and the TNR was 0.91. From the confusion matrix, 4 out of 8 MCI were correctly classified, and 21 out of 23 non-MCI were correctly classified.

```
=== Run information ===
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R
first-last\""
Instances:    31
Attributes:   2
              weightage
              Class
Test mode:    10-fold cross-validation
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances          25               80.6452 %
Incorrectly Classified Instances         6               19.3548 %
Kappa statistic                          0.4497
Mean absolute error                      0.2475
Root mean squared error                  0.4281
Relative absolute error                 62.7907 %
Root relative squared error             96.8826 %
Coverage of cases (0.95 level)          87.0968 %
Mean rel. region size (0.95 level)
79.0323 %
Total Number of Instances               31
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
               0.5       0.087     0.667       0.5     0.571       0.609     yes
               0.913     0.5       0.84        0.913   0.875       0.609     no
Weighted Avg.    0.806     0.393     0.795     0.806     0.797     0.609
=== Confusion Matrix ===
  a   b   <-- classified as
  4   4 |   a = yes
  2 21 |   b = no
```

Figure 6.5: Extracted output results using the IBk Algorithm

The detailed results for the Bagging algorithm are presented in Figure 6.6. The extracted outputs from the Bagging results, which consist of ‘detailed accuracy by class’ and ‘confusion matrix’, can be found in Figure 6.6. The overall accuracy for the Bagging algorithm was 0.77, while the TPR was 0.38 and the TNR was 0.91. From the confusion matrix, 3 out of 8 MCI were correctly classified, and 21 out of 23 non-MCI were correctly classified.

```
=== Run information ===

Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W
weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Relation:     VRdataelderly21022012lattributes42-
weka.filters.unsupervised.attribute.Remove-R1-6,8-42
Instances:    31
Attributes:   2
              weightage
              Class
=== Summary ===
Correctly Classified Instances          24               77.4194 %
Incorrectly Classified Instances         7               22.5806 %
Kappa statistic                          0.3282
Mean absolute error                      0.2816
Root mean squared error                  0.3783
Relative absolute error                 71.4563 %
Root relative squared error             85.6132 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances               31
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
               0.375     0.087     0.6         0.375   0.462       0.647     yes
               0.913     0.625     0.808       0.913   0.857       0.647     no
Weighted Avg.    0.774     0.486     0.754     0.774     0.755     0.647
=== Confusion Matrix ===
  a  b   <-- classified as
  3  5 |  a = yes
  2 21 |  b = no
```

Figure 6.6: Extracted output results using the Bagging Algorithm

The detailed results for the Multilayer Perceptron algorithm are presented in Figure 6.7. The extracted outputs from the Multilayer Perceptron results, which consist of 'detailed accuracy by class' and 'confusion matrix', can be found in Figure 6.7. The overall accuracy for the algorithm was 0.88, while the TPR was 0.63 and the TNR was 0.96. From the confusion matrix, 5 out of 8 MCI were correctly classified, and 22 out of 23 non-MCI were correctly classified.

```
=== Run information ===
Scheme:       weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V
0 -S 0 -E 20 -H a
Relation:     VRdataelderly21022012lattributes42-
weka.filters.unsupervised.attribute.Remove-R1-6,8-42
Instances:    31
Attributes:   2
=== Summary ===
Correctly Classified Instances         27               87.0968 %
Incorrectly Classified Instances        4               12.9032 %
Kappa statistic                         0.6331
Mean absolute error                     0.209
Root mean squared error                 0.3279
Relative absolute error                53.0252 %
Root relative squared error            74.2113 %
Coverage of cases (0.95 level)        100      %
Mean rel. region size (0.95 level)     95.1613 %
Total Number of Instances              31
=== Detailed Accuracy By Class ===
               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.625     0.043      0.833     0.625     0.714      0.755    yes
                 0.957     0.375      0.88      0.957     0.917      0.755    no
Weighted Avg.    0.871     0.289      0.868     0.871     0.864      0.755
=== Confusion Matrix ===
  a   b   <-- classified as
  5   3 |   a = yes
  1 22 |   b = no
```

Figure 6.7: Extracted output results using the Multilayer Perceptron Algorithm

From the detailed results, as depicted from Figures 1.4 to 1.7, a summary of the results is shown in Table 6.9. In order to identify the best performance, accuracy is the most important feature, followed by TPR and TNR. Accuracy is the most important feature because the aim of this study is to discriminate between MCI and healthy elderly people. Therefore, higher accuracy indicates that the classifier could predict MCI correctly. The results of this study show that Naïve Bayes Classifiers obtained the highest results, with an accuracy of 0.90, followed by TPR with 0.75 and TNR with 0.96 accuracy (see Table 6.9).

Table 6.9: Summary of results obtained from the WEKA package

Classifiers	**Accuracy**	**True Positive Rate**	**True Negative Rate**
Naïve Bayes	**0.90**	**0.75**	**0.96**
J48	0.88	0.50	1.00
IBk	0.81	0.50	0.91
Bagging	0.77	0.38	0.91
Multilayer Perceptron	0.88	0.63	0.96

In addition, assigning weighting to the most important variables improved performance measures. The results also show that the TPR obtained was 0.75 compared to 0.50, without using weighting. Furthermore, from the results, it can be seen that accuracy improved to 0.90, compared to 0.87 without using weighting. Naïve Bayes successfully out-performed other methods in finding the highest accuracy. This may due to its technique that performs well with small sample data (Jing et al., 2005). Further investigations may be needed to look at the nature of the data itself. In conclusion, the findings of accuracy demonstrate the highest performance measure obtained compared to other works.

6.7 Summary

This chapter has described the VREAD KDD processes in detail. Data mining techniques were used in the prediction of MCI and healthy elderly people. A comparison study between different feature selections and classifiers were also presented. In addition, the use of the weighting function to improve performance measures was also discussed; it demonstrates promising results. The findings of this study did show the accuracy and TPR significantly increase in prediction and in discrimination of between MCI and healthy elderly people compared to other related work discussed earlier in Chapter 2.

Chapter 7

Conclusions and Future Work

7.1 Introduction

This chapter summarised the research activities within this thesis, which has been concerned with the design, development, evaluation and data analysis of a newly developed virtual simulation named VREAD to test spatial memory among elderly. The following section emphasized the main contributions that have been made. The limitations of the work are discussed in the other sections. This chapter concludes with recommendations for future work as a way of developing the research.

7.2 Review of This Research

One of the biggest fears in the process of ageing is the fear of cognitive decline, which may lead to AD, the most common cause of dementia (Hettinga et al., 2009) The issue in AD today is to identify the disease before neurons degenerate to a stage beyond repair. Therefore, early detection of MCI is really important: it may ensure early treatment and intervention. In addition, it can have a significant impact in improving quality of life. Indeed, the early detection of cognitive decline is critical for preventing the progression to AD.

Consequently, investigation of MCI and early AD has increased substantially, even though there is currently no cure for AD (Singh-Manoux and Kivimäki, 2010; Korczyn, 2012; Pasinetti and Ho, 2010). Recently, some studies have reported that MCI will progressively develop to AD (Decarli, 2003; Busse et al., 2003; Blackwell et al., 2004; Petersen, 2011), but other studies have shown that many MCI patients will remain stable and recover (Larrieu et al., 2002; Ganguli et al., 2004). Since not everyone with MCI goes on to develop AD, it is a challenge to be able to detect early AD or MCI. Existing methods for the early detection of MCI require clinical assessment of medical history, neuropsychological tests, laboratory examination and clinical judgment (Alzheimer's Association, 2011). In spite of this, early diagnosis is still difficult to make (Liu et al., 2011; Wee et al., 2011). MMSE, one of the neuropsychological tests that has been widely used, is not sensitive to MCI (De Jager et al., 2002; Schmand et al., 2000).

Recently, new tools, such as neuroimaging scans, have been used for early detection of MCI (Frisoni et al., 2010). But, before using these expensive tools, it is necessary to detect or screen for early AD or MCI with tests that are sensitive to cognitive decline. It is therefore important to find methods and techniques that are able to discriminate between healthy elderly and MCI patients.

The development of VR technology has brought major progress to the study of spatial navigation in a virtual environment (Moffat et al., 2001). The virtual environment provides an interesting opportunity for the evaluation of topographical disorientation by providing a representation of an interactive environment. Recent studies by Cushman et al. (2008); Werner et al. (2009); Tippett et al. (2009) and Pengas et al. (2010) has shown opportunities of using virtual environment for early detection of early AD and MCI. Despite the efforts of these researchers to explore the use of the virtual environment in the early detection of MCI, there is scope to improve on them. One weakness of these research examples is their limited capacity to develop navigation tasks that would challenge users' motivation. A further limitation of these studies is the work of Werner and that of Pegas have been able to discriminate between the two subgroups with accuracy of only 71.7% and 80% respectively.

Motivated by the promise offered by VR technology and the importance of early detection of MCI, the virtual simulation was developed. VREAD was developed with the specific objective of discriminating between healthy elderly and MCI and has been able to overcome the limitations of previous work in the field. The virtual simulation was designed on the basis of design criteria that have been discussed earlier. The design was unique, with special features such as the learning repetition of following a red line, the use of five different levels which progressively become harder, automatic data collection and two different modules in a virtual park, which is a suitable setting for elderly. None of the participants commented that they were unable to use the apparatus and there were no reports of illness or nausea.

At the beginning of the research, the aim and objectives were presented. The aims were achieved by meeting the objectives as stated in Section 1.4, Chapter 1. The objectives of this research were:

1. To develop a novel virtual environment as a test for spatial memory to assess cognitive functions.
2. To implement and evaluate a newly developed virtual environment among non-elderly and elderly people.
3. To diagnose MCI at an early stage with data collected from the virtual simulation using machine learning techniques.
4. To propose design criteria specifically for elderly people to help with the virtual environment process.
5. To compare performance on the virtual simulation with existing neuropsychological tests used in AD research.

7.3 Original Contribution of the Research

This section discusses the original contributions in detail and how the aims and objectives were fulfilled. The seven individual research contributions of the thesis were as follows:-

1. A novel VREAD simulation to discriminate between healthy elderly and MCI patients.

   The first contribution was the development of a new virtual simulation known as VREAD to assess cognitive function in the elderly (Wan Shamsuddin et al., 2011a). The design and development, as well as the functions and utilities offered by VREAD, were discussed in Chapter 3.

2. The design criteria for the development of the VR simulation for spatial memory.

   The specific design needs to be considered because the effectiveness of the technology relies on the ability of the end user to use it. The primary goal of these design criteria is to promote accessibility. Based on these design criteria, a virtual simulation was developed and evaluated (see Chapter 3). The results on usability, with regard to acceptance and satisfaction by participants, indicate that these criteria are reliable and useful for use with elderly people (Wan Shamsuddin et al., 2011c).

3. The production of a dataset, which can be used for comparison by other researchers in the future.

   The data collected automatically from VREAD and neuropsychological tests can be used by other researchers. The results show a significant correlation for VREAD level 4 and scores on the neuropsychological tests. There was a significant positive correlation between VREAD level 4 score on both PAL six pattern stage and GNT, $r=-0.45$, $p<0.02$ and $r=0.43$, $p<0.02$ respectively. The neuropsychological tests are important because they should be able to identify a person's clinical group correctly. The scores obtained for VREAD level 4 were significantly worse in those participants with MCI in comparison with the healthy group, as discussed earlier. The results indicate a significant difference ($t(29)=3.17$, $p<0.01$) showing that VREAD level 4 is a novel assessment which can discriminate reliably between the two groups. Moreover, there was no association with the MMSE which had previously been criticised for its inability to detect mild changes in cognition (Pasquier, 1999; De Jager et al., 2002).

4. The accuracy rate for discrimination, as a result of using data mining techniques, could be used as a benchmark by other researchers in future.

   Machine learning algorithms were applied to make predictions of healthy elderly and MCI. The promising results obtained show that VREAD could predict MCI correctly with 0.90 accuracy, 0.75 TPR and 0.96 TNR.

5. Empirical findings in this study could be usefully explored in further research.

   Data collected from VREAD automatically or from questionnaires were analysed and investigated. The encouraging results using SPSS for the early detection of early AD were explained in Chapter 4. The positive results obtained for user acceptance and satisfaction among non-elderly and elderly participants were discussed in Chapter 5. Further findings regarding the use of machine learning algorithms are also explained in Chapter 6.

6. A new measurement to test ways of assessing performance in spatial memory

   The raw dataset produced was automatically collected during VREAD testing. To visualize data, a new performance measurement to test wayfinding performance in spatial memory was proposed. The Route Test Score was proposed to calculate the total score, which includes the correct sequences, exact sequences and incorrect sequences values.

7. A methodology to discriminate between MCI patients and the healthy elderly using the VREAD simulation.

   The methodology proposed to discriminate between MCI patients and the healthy elderly using the VREAD simulation, as proposed in Chapter 1, could be implemented in other similar domains.

7.4 Research Limitations

The contributions and the promising results presented previously show that the research successfully contributes by discriminating between healthy elderly and MCI. However, the research has some limitations that should be acknowledged.

- The sample sizes were small. Further study needs to be done with a larger population to give better results to the findings.

- The virtual simulation was developed using Virtools software. Although Virtools is a VR development tools that enables easy creation, there is still a limitation: it uses an online 3DVIA Player to deploy the virtual simulation. Thus, it is difficult to run the experiments if internet connection is not available.

- In this research, a standard computer was used in the implementation and investigation of the newly developed virtual simulation. This gives users a feeling of being the least immersed in a virtual environment. In addition, the use of a keyboard seemed to restrain participants' movement, although overall experiments were successful.

- Follow up study of these participants is important and necessary to examine their cognitive decline. This would enable results obtained in this study to be compared with real decline in the follow-up session. By comparing the results, it can be determined whether VREAD is reliable in discriminating between healthy elderly and MCI.

7.5 Recommendation for Future Work

The current research has some shortcomings and limitations, listed above. These problems need to be addressed in future work as highlighted below:-

- Further modifications are required to make the virtual environment more immersive with other VR technology of either with head mounted display or Cave Automatic Virtual Environment. A 3D display may be an option if the costs of other systems are too expensive. An investigation needs to be consider whether the 3D display may help users navigate the 3D environment.Although the present results are important and significant, future research should strive to obtain more participants. A larger sample may reveal insights undiscovered by this research.

- In future, the data collection may be expanded to include users' behaviour, such as idle time and number of pauses. Furthermore, users' reactions may be recorded for future investigation. This may include facial expression and body gesture. By considering these factors in the future, a highly accurate result may be produced for predicting early AD.

- The overarching aim of this study was to use a simple and rapid test that is also sensitive to mild degree impairment. The VREAD level 4 score that significantly correlates with other neuropsychological tests, compared to other levels, may be used as a novel path to discriminate between healthy elderly and MCI. Thus, it is essential to have only one level that could discriminate between these two subgroups; this could be the next stage of this research.

- Machine learning techniques with more sophisticated algorithm to discriminate between healthy elderly and MCI can be one of the future works.

- This simulation could also be considered to be migrated to an online environment with cloud infrastructure that enables users to analyses themselves in their home environment. In addition, it could minimize cost and could be assessed anytime at their own convenience.

7.6 Summary

This chapter emphasizes and highlights the significant contributions of this research. The success of this research relies on the aims and objectives outlined being carried out. The listed contributions have shown that these aims and objectives were fulfilled. In addition, the structured research design proposed was implemented to accomplish the goals.

This research has overcome many of the limitations of current AD tests using a virtual environment. As a result, VREAD, a newly developed virtual simulation, could be used as a possible method for detection of a cognitive impairment. The results of VREAD provide an effective way to discriminate between healthy elderly and MCI patients (Lesk et al., 2012). The significant correlations were found between VREAD level 4 score, the CANTAB PAL and the GNT. The findings using machine learning algorithms also provide preliminary evidence that VREAD was sensitive to cognitive decline in the elderly with 90% accuracy. Thus, further evaluations with a larger cohort are necessary to confirm the results obtained and may provide new insight into the early detection of MCI. The results of this systematic study will serve as guidelines for clinical use and further research in the area.

References

ISO 9241. 1998. Ergonomic Requirements for Office Work with Visual Display Terminals(Part 2): Guidance on Usability [Online]. Available at: http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf (Accessed: February 16, 2012).

Aguirre, G. K. &D'esposito, M. 1999. Topographical Disorientation: A Synthesis and Taxonomy. *Brain,* 122, 1613-1628.

Andrews, K. 2008. *Evaluation comes in many guises*. Workshop on Beyond time and errors: novel evaluation methods for Information Visualization, CHI 2008. April 5, 2008. Florence, Italy. Available at: http://www.dis.uniroma1.it/beliv08/pospap/andrews.pdf. (Assessed : 3 March 2012).

Association, A. S. 2011. 2011 Alzheimer's Disease Facts and Figures. *Alzheimer's and Dementia,* 7, 208- 244.

Banos, R. M., Botella, C., Perpiñá, C., Alcañiz, M., Lozano, J. A., Osma, J. & Gallardo, M. 2002. Virtual Reality Treatment of Flying Phobia. *Information Technology in Biomedicine, IEEE Transactions on,* 6, 206-212.

Barrash, J. 1998. A Historical Review of Topographical Disorientation and Its Neuroanatomical Correlates. *Journal of Clinical and Experimental Neuropsychology,* 20, 807-827.

Bekker, M., Barendregt, W., Crombeen, S. & Biesheuvel, M. 2005. Evaluating Usability and Challenge During Initial and

Extended Use of Children's Computer Games. *People and Computers XVIII—Design for Life*, 331-345.

Blackman, T., Mitchell, L., Burton, E., Jenks, M., Parsons, M., Raman, S. & Williams, K. 2003. The Accessibility of Public Spaces for People with Dementia: A New Priority for The'open City'. *Disability & Society,* 18, 357-371.

Blackwell, A., Sahakian, B., Vesey, R., Semple, J., Robbins, T. & Hodges, J. 2004. Detecting Dementia: Novel Neuropsychological Markers of Preclinical Alzheimer's Disease. *Dementia and geriatric cognitive disorders,* 17, 42-48.

Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. 2010. WEKA-Experiences with a Java Open-Source Project. *The Journal of Machine Learning Research, 11*, 2533-2541.

Bowman, D. A., Gabbard, J. L. & Hix, D. 2002. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators& Virtual Environments,* 11, 404-424.

Bowman, D. A., North, C., Chen, J., Polys, N. F., Pyla, P. S. & Yilmaz, U. 2003. Information-Rich Virtual Environments: Theory, Tools, and Research Agenda. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. 2003 New York, USA. ACM, 81-90.

Burdea, G., Gomez, D., Langrana, N., Rokros, E. & Richard, P. 1995. Virtual Reality Graphics Simulation with Force Feedback.

International Journal in Computer Simulation, ABLEX Publishing, 5, 287-303.

Busse, A., Bischkopf, J., Riedel-Heller, S. G. & Angermeyer, M. C. 2003. Mild Cognitive Impairment: Prevalence and Incidence According to Different Diagnostic Criteria Results of the Leipzig Longitudinal Study of the Aged (Leila75+). *The British Journal of Psychiatry,* 182, 449-454.

Carvalho, A. 2001. Usability Testing of Educational Software: Methods, Techniques and Evaluators. Proceeding of the *3rd Simpósio Internacional de Informática Educativa*, 139-148.

Carifio, J. & Perla, R. J. 2007. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends About Likert Scales and Likert Response Formats and Their Antidotes. *Journal of Social Sciences,* 3**,** 106-116.

Charito, D. Designing Space in Virtual Environments for Aiding Wayfinding Behaviour *Proceedings of the 4th UK Virtual Reality Special Interest Group Conference*, 1997 Brunel University.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. 2011. SMOTE: Synthetic Minority over-Sampling Technique. *Journal of Artificial Intelligence Research.* 321-357.

Cheok, A. D., Haller, M., Fernando, O. N. N. & Wijesena, J. P. 2009. Mixed Reality Entertainment and Art. *The International Journal of Virtual Reality,* 8, 83-90.

Cherrier, M., Mendez, M. & Perryman, K. 2001. Route Learning Performance in Alzheimer Disease Patients. *Cognitive and Behavioral Neurology,* 14, 159-168.

Chertkow, H., Nasreddine, Z., Joanette, Y., Drolet, V., Kirk, J., Massoud, F., Belleville, S. & Bergman, H. 2007. Mild Cognitive Impairment and Cognitive Impairment, No Dementia: Part a, Concept and Diagnosis. *Alzheimer's and Dementia,* 3, 266-282.

Coelho, C. M., Waters, A. M., Hine, T. J. & Wallis, G. 2009. The Use of Virtual Reality in Acrophobia Research and Treatment. *Journal of Anxiety Disorders,* 23, 563-574.

Cushman, L. A., Stein, K. & Duffy, C. J. 2008. Detecting Navigational Deficits in Cognitive Aging and Alzheimer Disease Using Virtual Reality. *Neurology,* 71, 888-895.

De Jager, C., Milwain, E. & Budge, M. 2002. Early Detection of Isolated Memory Deficits in the Elderly: The Need for More Sensitive Neuropsychological Tests. *Psychological Medicine,* 32, 483-491.

Decarli, C. 2003. Mild Cognitive Impairment: Prevalence, Prognosis, Aetiology, and Treatment. *The Lancet Neurology,* 2, 15-21.

Dickinson, A., Newell, A. F., Smith, M. J. & Hill, R. L. 2005. Introducing the Internet to the over-60s: Developing an Email System for Older Novice Computer Users. *Interacting with Computers,* 17, 621-642.

Dubinsky, R. M., Stein, A. C. & Lyons, K. 2000. Practice Parameter: Risk of Driving and Alzheimer’s Disease: An Evidence-Based Review. *Neurology,* 54, 2205-2211.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S. & Jicha, G. 2007. Research Criteria for the Diagnosis of Alzheimer's Disease: Revising the Nincds-Adrda Criteria. *The Lancet Neurology,* 6, 734-746.

Duffy, C. J. 2009. Visual Motion Processing in Aging and Alzheimer's Disease. *Annals of the New York Academy of Sciences,* 1170, 736-744.

Égerházi, A., Berecz, R., Bartók, E. & Degrell, I. 2007. Automated Neuropsychological Test Battery (Cantab) in Mild Cognitive Impairment and in Alzheimer's Disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 31, 746-751.

Ellis, S. R. 1994. What Are Virtual Environments? *Computer Graphics and Applications, IEEE,* 14, 17-22.

Fawcett, T. 2004. Roc Graphs: Notes and Practical Considerations for Researchers. *Machine Learning,* 31, 1-38.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. *AI magazine,* 17, 37.

Fitzgerald, D., Kelly, D., Ward, T., Markham, C. & Caulfield, B. 2008. Usability Evaluation of E-Motion: A Virtual Rehabilitation System Designed to Demonstrate, Instruct and Monitor a

Therapeutic Exercise Programme.*Proceedings of Virtual Rehabilitation 2008.* August 25th-27th 2008. Vancouver, Canada. IEEE, 144-149.

Flores, E., Tobon, G., Cavallaro, E., Cavallaro, F. I., Perry, J. C. & Keller, T.*2008.* Improving Patient Motivation in Game Development for Motor Deficit Rehabilitation. *Proceedings of the 2008 International Conference in Advances on Computer Entertainment Technology, 2008*. ACM, 381-384.

Folstein, M. F., Folstein, S. E. & McHugh, P. R. 1975. Mini-Mental State: A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research.* 12, 189-198.

Fotouhi-Ghazvini, F., Earnshaw, R., Robison, D. & Excell, P. 2009. The Mobo City: A Mobile Game Package for Technical Language Learning. *International Journal of Interactive Mobile Technologies.* 3**,** 19-24.

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. 2010. The Clinical Use of Structural Mri in Alzheimer Disease. *Nature Reviews Neurology,* 6, 67-77.

Fuchs, A. B. & Ogilvie, S. 2012. *Psychlopedia Project* [Online]. Available:http://www.psychit.com.au/Psychlopedia/article.asp?id=1 [Accessed April 23 2012].

Ganguli, M., Dodge, H. H., Shen, C. & Dekosky, S. T. 2004. Mild Cognitive Impairment, Amnestic Type. *Neurology,* 63, 115-121.

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. & Botella, C. 2002. Virtual Reality in the Treatment of Spider Phobia: A Controlled Study. *Behaviour Research and Therapy,* 40, 983-993.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D. & Chertkow, H. 2006. Mild Cognitive Impairment. *The Lancet,* 367, 1262-1270.

Gerardi, M., Rothbaum, B. O., Ressler, K., Heekin, M. & Rizzo, A. 2008. Virtual Reality Exposure Therapy Using a Virtual Iraq: Case Report. *Journal of Traumatic Stress,* 21, 209-213.

Golledge, R. G. 1991. Cognition of Physical and Built Environments. In T. Gärling& G. Evans (eds). *Environment, Cognition, and Action: An Integrated Approach*. New York. Oxford University Press, 35-62.

Gupta, V. K., Gupta, S. G. & Gupta, P. 2012. Understanding P Value. *Indian Journal of Medical Specialities,* 3**,** 77-81.

Gustafson, K. L. & Branch, R. M. 2002. What Is Instructional Design. *Trends and Issues in Instructional Design and Technology*, 16-25.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2009. The Weka Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter,* 11, 10-18.

Hartley, J. 1994. Designing Instructional Text for Older Readers: A Literature Review. *British Journal of Educational Technology,* 25, 172-188.

Hawthorn, D. 2000. Possible Implications of Aging for Interface Designers. *Interacting with Computers,* 12, 507-528.

Heim, S. G. 2008. *The Resonant Interface: HCI Foundations for Interaction Design,* USA, Addison Wesley.

Hettinga, M., De Boer, J., Goldberg, E. & Moelaert, F. 2009. Navigation for People with Mild Dementia. *Studies in Health Technology and Informatics,* 150, 428.

Holden, M. K. 2005. Virtual Environments for Motor Rehabilitation: Review. *CyberPsychology & Behavior,* 8, 187-211.

Holmes, G., Donkin, A., & Witten, I.H. 1994. Weka: A Machine Learning Workbench. *Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems*, Brisbane, Australia, 357-361.

Holt, J., Stiltner, L. & Wallace, R. 2009. Do Patients at High Risk of Alzheimer's Disease Benefit from Early Treatment? *The Journal of Family Practice,* 58, 320-322.

Hox, J. J. 1998. Multilevel Modeling: When and Why. In R. M. Balderjahn I. & M. Schader (Eds.), *Classification, Data Analysis and Data Highways*, New York, Springer Verlag,147–154.

Iachini, I., Iavarone, A., Senese, V. P., Ruotolo, F. & Ruggiero, G. 2009. Visuospatial Memory in Healthy Elderly, AD and MCI: A Review. *Current Aging Science,* 2**,** 43-59.

Iaria, G., Bogod, N., Fox, C. J. & Barton, J. J. S. 2009. Developmental Topographical Disorientation: Case One. *Neuropsychologia,* 47, 30-40.

Ijsselsteijn, W., Nap, H. H., De Kort, Y. & Poels, K. 2007. Digital Game Design for Elderly Users. *Proceedings of the Conference on Future Play 2007.* New York, USA. ACM, 17-22.

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J. & Ward, C. 2008. The Alzheimer's Disease Neuroimaging Initiative (Adni): Mri Methods. *Journal of Magnetic Resonance Imaging,* 27, 685-691.

Jacova, C., Kertesz, A., Blair, M., Fisk, J. D. & Feldman, H. H. 2007. Neuropsychological Testing and Assessment for Dementia. *Alzheimer's and Dementia,* 3, 299-317.

Jang, D. P., Ku, J. H., Choi, Y. H., Wiederhold, B. K., Nam, S. W., Kim, I. Y. & Kim, S. I. 2002. The Development of Virtual Reality Therapy (Vrt) System for the Treatment of Acrophobia and Therapeutic Case. *Information Technology in Biomedicine, IEEE Transactions on,* 6, 213-217.

Jing, Y., Pavlović, V. & Rehg, J. M. 2005 Efficient Discriminative Learning of Bayesian Network Classifier Via Boosted Augmented

Naive Bayes. *Proceedings of the 22nd International Conference on Machine Learning*, USA, ACM, 369-376.

Karageorgiou, E., Schulz, S. C., Gollub, R. L., Andreasen, N. C., Ho, B. C., Lauriello, J., Calhoun, V. D., Bockholt, H. J., Sponheim, S. R. & Georgopoulos, A. P. 2011. Neuropsychological Testing and Structural Magnetic Resonance Imaging as Diagnostic Biomarkers Early in the Course of Schizophrenia and Related Psychoses. *Neuroinformatics,* 9, 321-333.

Kasiulevičius, V., Šapoka, V. & Filipavičiūtė, R. 2006. Sample Size Calculation in Epidemiological Studies. *Gerantologija,* 7, 225-231.

Kesselsa, R. P. C., Haana, E. H. F. D., Kappelleb, L. J. & Postmaa, A. 2001. Varieties of Human Spatial Memory: A Meta-Analysis on the Effects of Hippocampal Lesions. *Brain Research Reviews,* 35, 295-303.

Kim, K., Rosenthal, M. Z., Zielinski, D. & Brady, R. 2012. Comparison of Desktop, Head Mounted Display, and Six Wall Fully Immersive Systems Using a Stressful Task. *Proceedings of Virtual Reality*, USA, IEEE, 143-144.Kizony, R., Weiss, P., Shahar, M. & Rand, D. 2006. Theragame: A Home Based Virtual Reality Rehabilitation System. *International Journal on Disability and Human Development,* 5, 265.

Kohavi, R. & Foster, P. 1998. Glossary of terms, *Machine Learning,* 30,271-274.

Korczyn, A. D. 2012. Why Have We Failed to Cure Alzheimer's Disease? *Journal of Alzheimer's Disease*. 29,275-82.

Kovalerchuk, B. and Vityaev, E. 2010. Data Mining for Financial Applications, In Maimon, Oded, Rokach, Lior (Eds), *Data mining and Knowledge Discovery*. (2nd Ed). New York. Springer, 1154-1169.

Krijn, M., Emmelkamp, P. M. G., Biemond, R., De Wilde De Ligny, C., Schuemie, M. J. & Van Der Mast, C. a. P. G. 2004. Treatment of Acrophobia in Virtual Reality: The Role of Immersion and Presence. *Behaviour Research and Therapy,* 42**,** 229-239.

Larrieu, S., Letenneur, L., Orgogozo, J., Fabrigoule, C., Amieva, H., Le Carret, N., Barberger–Gateau, P. & Dartigues, J. 2002. Incidence and Outcome of Mild Cognitive Impairment in a Population-Based Prospective Cohort. *Neurology,* 59, 1594-1599.

Laver, K., George, S., Thomas, S., Deutsch, J. E. & Crotty, M. 2012. Virtual Reality for Stroke Rehabilitation [Online]. Available at:http://www.update-software.com/BCP/WileyPDF/EN/CD008349.pdf (Accessed: December 5 2011).

Lavrac,N. and Zupan,B. 2010. Data Mining in Medicine, In Maimon, Oded, Rokach, Lior , (Eds). *Data Mining and Knowledge Discovery*. (2nd Ed.). New York, Springer, 1111-1136.

Leonardi, C., Mennecozzi, C., Pianesi, F. & Zancanaro, M. 2008. Supporting Older Adults Social Network: The Design of E-Inclusion Communication Services. *Gerontechnology,* 7, 153.

Lesk, V., Wan Shamsuddin, S., Walters, E. & Ugail, H. 2012. Using a Virtual Environment to Assess Cognition in the Elderly. *Alzheimer's & Dementia: Journal of the Alzheimer's Association.* (Submitted).

Lin, W. J., Yu, B., Huang, T. Y., Lin, C. T., Hung, D. L., Tzeng, O. J. L. & Chang, E. 2008. Human Spatial Navigation in Virtual Environment: The Contributions of Different Types of Landmarks. *Symposium on Digital Life Technologies: Human-Centric Smart Living Technology,.* Tainan, Taiwan.

Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L. O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P. & Tsolaki, M. 2011. Combination Analysis of Neuropsychological Tests and Structural MRI Measures in Differentiating AD, MCI and Control Groups—the Addneuromed Study. *Neurobiology of Aging,* 32**,** 1198-1206.

Lloyd, J., Riley, G. A. & Powell, T. E. 2009. Errorless Learning of Novel Routes through a Virtual Town in People with Acquired Brain Injury. *Neuropsychological Rehabilitation,* 19, 98-109.

Luukka, P. 2011. Feature Selection Using Fuzzy Entropy Measures with Similarity Classifier. *Expert Systems with Applications,* 38, 4600-4607.

Mahmud, N., Vogt, J., Luyten, K., Slegers, K., Van Den Bergh, J. & Coninx, K. 2010. Dazed and Confused Considered Normal: An Approach to Create Interactive Systems for People with Dementia. *Human-Centred Software Engineering*, 119-134.

Mapstone, M., Steffenella, T. M. & Duffy, C. J. 2003. A Visuospatial Variant of Mild Cognitive Impairment. *Neurology,* 60, 802-808.

Marques, A., Queirós, C. & Rocha, N. 2008. Virtual Reality and Neuropsychology: A Cognitive Rehabilitation Approach for People with Psychiatric Disabilities. *Proceedings of 7th International Conference on Disability Virtual Reality and Associated Technologies*. Maia, Protugal, 39-46.

Maruta, C., Guerreiro, M., De Mendonca, A., Hort, J. & Scheltens, P. 2011. The Use of Neuropsychological Tests across Europe: The Need for a Consensus in the Use of Assessment Tools for Dementia. *European Journal of Neurology,* 18, 279-285.

Mistur, R., Mosconi, L., De Santi, S., Guzman, M., Li, Y., Tsui, W. & De Leon, M. J. 2009. Current Challenges for the Early Detection of Alzheimer's Disease: Brain Imaging and Csf Studies. *Journal of Clinical Neurology Seoul, Korea.* 5, 153.

Moffat, S. D., Zonderman, A. B. & Resnick, S. M. 2001. Age Differences in Spatial Memory in a Virtual Environment Navigation Task. *Neurobiology of Aging,* 22, 787-796.

Moline, J. 1997. Virtual Reality for Health Care: A Survey. In G. Riva (Ed), *Virtual Reality in Neuro-psycho-physiology: Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation,* IOS Press.

Money, A. G., Fernando, S., Lines, L. & Elliman, A. D. 2009. Developing and Evaluating Web-Based Assistive Technologies for Older Adults. *Gerontechnology,* 8, 165-177.

Morris, J. C. 2006. Mild Cognitive Impairment Is Early-Stage Alzheimer Disease: Time to Revise Diagnostic Criteria. *Archives of Neurology,* 63, 15.

Moscovitch, M., Westmacott, R., Gilboa, A., Addis, D. R., Rosenbaum, R. S., Viskontas, I., Priselac, S., Svoboda, E., Ziegler, M. & Black, S. 2005. Hippocampal Complex Contribution to Retention and Retrieval of Recent and Remote Episodic and Semantic Memories: Evidence from Behavioral and Neuroimaging Studies of Healthy and Brain-Damaged People. In Ohta N, MacLeod CM, Uttl B (Eds). *Dynamic Cognitive Processes*. Springer-Verlag, 333-380.

Mujber, T., Szecsi, T. & Hashmi, M. 2004. Virtual Reality Applications in Manufacturing Process Simulation. *Journal of Materials Processing Technology,* 155, 1834-1838.

Nielsen, J. 1993. *Usability Engineering*,San Diego, CA, Morgan Kaufmann.

O'brien, H. L., Tetewsky, S. J., Avery, L. M., Cushman, L. A., Makous, W. & Duffy, C. J. 2001. Visual Mechanisms of Spatial Disorientation in Alzheimer's Disease. *Cerebral Cortex,* 11, 1083-1092.

Olsen, T., Procci, K. & Bowers, C. 2011. Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness. *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, 625-634.

Pantel, J., Schönknecht, P., Essig, M. &Schröder, J. 2004. Distribution of Cerebral Atrophy Assessed by Magnetic Resonance Imaging Reflects Patterns of Neuropsychological Deficits in Alzheimer's Dementia. *Neuroscience letters,* 361, 17-20.

Pasinetti, G. M. & Ho, L. 2010. Role of Grape Seed Polyphenols in Alzheimer's Disease Neuropathology. *Nutr Diet Suppl,* 2, 97-103.

Pasquier, F. 1999. Early Diagnosis of Dementia: Neuropsychology. *Journal of neurology,* 246, 6-15.

Passini, R., Rainville, C., Marchand, N. & Joanette, Y. 1995. Wayding in Dementia of the Alzheimer Type: Planning Abilities. *Journal of Clinical and Experimental Neuropsychology,* 17, 820-832.

Pengas, G., Patterson, K., Arnold, R. J., Bird, C. M., Burgess, N. & Nestor, P. J. 2010. Lost and Found: Bespoke Memory Testing for Alzheimer's Disease and Semantic Dementia. *Journal of Alzheimer's Disease,* 21, 1347-1365.

Petersen, R. C. 2004. Mild Cognitive Impairment as a Diagnostic Entity. *Journal of Internal Medicine,* 256, 183-194.

Petersen, R. C. 2011. Mild Cognitive Impairment. *Journal of Medicine,* 364, 2227-2234.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L. & Winblad, B. 2001. Current Concepts in Mild Cognitive Impairment. *Archives of Neurology,* 58, 1985-1992.

Petersen, R. C., Jack Jr, C., Xu, Y. C., Waring, S., O'brien, P., Smith, G., Ivnik, R., Tangalos, E., Boeve, B. F. &Kokmen, E. 2000. Memory and Mri-Based Hippocampal Volumes in Aging and Ad. *Neurology,* 54, 581-581.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G. & Kokmen, E. 1999. Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology,* 56, 303.

Petrella, J. R., Coleman, R. E. & Doraiswamy, P. M. 2003. Neuroimaging and Early Diagnosis of Alzheimer Disease: A Look to the Future1. *Radiology,* 226, 315-336.

Petridis, P., Dunwell, I., De Freitas, S. & Panzoli, D. 2010. An Engine Selection Methodology for High Fidelity Serious Games. *Proceedings of the 2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, IEEE, 27-34.Piekarski, W. & Thomas, B. 2002. Arquake: The Outdoor Augmented Reality Gaming System. *Communications of the ACM,* 45, 36-38.

Preece, J., Rogers, Y. & Sharp, H. 2002. *Interaction Design: Beyond Human-Computer Interaction.* New York, John Wiley & Son.

Rego, P., Moreira, P. M. & Reis, L. P. 2010. Serious Games for Rehabilitation: A Survey and a Classification Towards a Taxonomy. *Proceedings of 5th Iberian Conference on Information Systems and Technologies 2010.* Santiago de Compostela, IEEE, 1-6.

Rizzo, A. A. & Kim, G. 2005. A Swot Analysis of the Field of Virtual Rehabilitation and Therapy. *Presence: Teleoperators and Virtual Environments,* 14, 1-28.

Rizzo, A. A., Schultheis, M., Kerns, K. A. & Mateer, C. 2004. Analysis of Assets for Virtual Reality Applications in Neuropsychology. *Neuropsychological Rehabilitation,* 14, 207-239.

Rizzo, A. A., Wiederhold, M. & Buckwalter, J. G. 1998. Basic Issues in the Use of Virtual Environments for Mental Health Applications. *Studies in Health Technology and Informatics*, 21-42.

Roscoe, J. T. 1975. *Fundamental Research Statistics for the Behavioral Sciences*, Holt, Rinehart and Winston New York.

Sousa Santos, B., Dias, P., Pimentel, A., Baggerman, J. W., Ferreira, C., Silva, S. & Madeira, J. 2009. Head-Mounted Display Versus Desktop for 3d Navigation in Virtual Reality: A User Study. *Multimedia Tools and Applications,* 41**,** 161-181.

Schmand, B., Walstra, G., Lindeboom, J., Teunisse, S. & Jonker, C. 2000. Early Detection of Alzheimer's Disease Using the Cambridge Cognitive Examination (Camcog). *Psychological Medicine,* 30, 619-627.

Schultheis, M. T. & Rizzo, A. A. 2001. The Application of Virtual Reality Technology in Rehabilitation. *Rehabilitation Psychology,* 46, 296.

Shen-Tzay Huang, Yi-Pei Cho, & Yu-Jen Lin. 2005. ADDIE instructional design and cognitive apprenticeship for project-based software engineering education in MIS, Proceedings of *Software Engineering Conference, 2005. APSEC '05.* 15-17 Dec. 2005. 8.

Sherman, W. R. & Craig, A. B. 2003. *Understanding Virtual Reality: Interface, Application, and Design.*San Francisco, CA, Morgan Kaufmann Publishers.

Siegel, A. W. & White, S. H. 1975. The Development of Spatial Representations of Large-Scale Environments. *Advances in Child Development and Behavior,* 10, 9-55.

Silverstein, N. M., Flaherty, G. & Tobin, T. S. 2006. *Dementia and Wandering Behavior: Concern for the Lost Elder*, Springer Publishing Company.

Singh-Manoux, A. & Kivimäki, M. 2010. The Importance of Cognitive Ageing for Understanding Dementia. *Age,* 32, 509-512.

Slater, M., Pertaub, D. P., Barker, C. & Clark, D. M. 2006. An Experimental Study on Fear of Public Speaking Using a Virtual Environment. *CyberPsychology & Behavior,* 9, 627-633.

Small, D. H. & Cappai, R. 2006. Alois Alzheimer and Alzheimer's Disease: A Centennial Perspective. *Journal of Neurochemistry,* 99, 708-710.

Starkstein, S. E., Mizrahi, R. & Power, B. D. 2008. Depression in Alzheimer's Disease: Phenomenology, Clinical Correlates and Treatment. *International Review of Psychiatry,* 20, 382-388.

Stoddard, J. L., Augustson, E. M. & Mabry, P. L. 2006. The Importance of Usability Testing in the Development of an Internet-Based Smoking Cessation Treatment Resource. *Nicotine & Tobacco Research,* 8, S87-S93.

Sugiyama, T. & Thompson, C. W. 2006. Environmental Support for Outdoor Activities and Older People's Quality of Life. *Journal of Housing for the Elderly,* 19, 167-185.

Tailby, R. & Haslam, C. 2003. An Investigation of Errorless Learning in Memory-Impaired Patients: Improving the Technique and Clarifying Theory. *Neuropsychologia,* 41, 1230-1240.

Tetewsky, S. J. & Duffy, C. J. 1999. Visual Loss and Getting Lost in Alzheimer's Disease. *Neurology,* 52, 958-958.

Thearling, K. 2010. Data Mining for CRM. In Maimon, Oded, Rokach, Lior , (Eds). *Data Mining and Knowledge Discovery. (2nd Ed.). New York, Springer***,** 1181-1188.

Tippett, W. J., Lee, J. H., Zakzanis, K. K., Black, S. E., Mraz, R. & Graham, S. J. 2009. Visually Navigating a Virtual World with Real-World Impairments: A Study of Visually and Spatially Guided Performance in Individuals with Mild Cognitive Impairments. *Journal of Clinical and Experimental Neuropsychology,* 31, 447-454.

Uc, E., Rizzo, M., Anderson, S., Shi, Q. & Dawson, J. 2004. Driver Route-Following and Safety Errors in Early Alzheimer Disease. *Neurology,* 63, 832-837.

Van De Pol, L. A., Korf, E. S. C., Van Der Flier, W. M., Brashear, H. R., Fox, N. C., Barkhof, F. & Scheltens, P. 2007. Magnetic Resonance Imaging Predictors of Cognition in Mild Cognitive Impairment. *Archives of Neurology,* 64, 1023.

Vinson, N. G. 1999. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: the CHI is the Limit*, 15-20 May 1999. Pittsburgh, Pennsylvania.278-285.

Virvou, M. & Katsionis, G. 2008. On the Usability and Likeability of Virtual Reality Games for Education: The Case of VR-Engage. *Computers & Education,* 50, 154-178.

Wallace, J., Mulvenna, M. D., Martin, S., Stephens, S. & Burns, W. 2010. ICT Interface Design for Ageing People and People with Dementia. *Supporting People with Dementia Using Pervasive Health Technologies***,** 165-188.

Wan Shamsuddin, S., Ugail, H. & Lesk, V. 2011a. Development and Usability Evaluation of Virtual Environment for Early Diagnosis of Dementia. *Proceedings of the Second International Conference on Visual Informatics: Sustaining Research and Innovations, 9 – 11 November, Malaysia: Springer.* 13-22.

Wan Shamsuddin, S. N., Lesk, V. & Ugail, H. 2011b. Virtual Environment Design Guidelines for Elderly People in Early Detection of Dementia. *Proceedings of the International Conference of Computer and Information Science*, November 14-16 2011, Paris, France. 751-755.

Wan Shamsuddin, S. N., Walters, E., Ugail, H. & Lesk, V. 2011c. Evaluation of Users Acceptance in Virtual Environment for Early Diagnosis of Dementia. *Paper presented at 2nd World Conference on Information Technology.* (WCIT), November 23-26. Antalya Turkey.

Warrington, E. & Mckenna, P. 1983. *The Graded Naming Test*,Windsor, NFER-Nelson.

Wee, C. Y., Yap, P. T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L. & Shen, D. 2011. Identification of Mci Individuals Using Structural and Functional Connectivity Networks. *NeuroImage.*59,2045-2056.

Weimer, D. L. & Sager, M. A. 2009. Early Identification and Treatment of Alzheimer's Disease: Social and Fiscal Outcomes. *Alzheimer's and Dementia,* 5, 215-226.

Weissensteiner, E. & Winter, S. 2004. Landmarks in the Communication of Route Directions. *Geographic Information Science*, 313-326.

Weniger, G., Ruhleder, M., Lange, C., Wolf, S. & Irle, E. 2011. Egocentric and Allocentric Memory as Assessed by Virtual Reality in Individuals with Amnestic Mild Cognitive Impairment. *Neuropsychologia,* 49, 518-527.

Werner, P., Rabinowitz, S., Klinger, E., Korczyn, A. D. & Josman, N. 2009. Use of the Virtual Action Planning Supermarket for the Diagnosis of Mild Cognitive Impairment. *Dementia and Geriatric Cognitive Disorders,* 27, 301-309.

Werner, S., Krieg-Bruckner, B., Mallot, H. A., Schweizer, K. & Freksa. Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation. In Jarke, M.; Pasedach, K.; Pohl, K. (Eds.), *Informatik '97*, 1997. Berlin, Springer. 41-50.

Whitley, E. & Ball, J. 2002. Statistics Review 4: Sample Size Calculations. *Critical Care,* 6**,** 335-341.

Wilson, S., Davies, R., Stone, T., Hammerton, J., Ware, P., Mawson, S., Harris, N., Eccleston, C., Zheng, H. & Black, N. 2007. Developing a Telemonitoring System for Stroke Rehabilitation. *Contemporary Ergonomics,* 2007, 505.

Yusoff, Amri, Crowder, Richard, Gilbert, Lester and Wills, Gary (2009) A conceptual framework for serious games. *Proceeding of the Ninth IEEE International Conference on Advanced Learning Technologies*, 15 - 17 Jul 2009.Riga, Latvia. 21-23.

Zadikoff, C., Fox, S. H., Tang-Wai, D. F., Thomsen, T., De Bie, R., Wadia, P., Miyasaki, J., Duff-Canning, S., Lang, A. E. &Marras, C. 2008. A Comparison of the Mini Mental State Exam to the Montreal Cognitive Assessment in Identifying Cognitive Deficits in Parkinson's Disease. *Movement Disorders,* 23, 297-299.

Zaibon, S. B. & Shiratuddin, N. 2010. Adapting Learning Theories in Mobile Game-Based Learning Development. *Proceedings of the Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL)*, 12-16 April 2010, Kaohsiung,Taiwan. 124-128.

Appendix A

Consent Form

Division of Psychology
E Floor, Richmond
Building
University of Bradford
Richmond Road
Bradford
Researchers : **Dr Valerie Lesk, Prof. Hassan Ugail, Syadiah Shamsuddin and Elizabeth Walters**

Testing spatial memory using a virtual reality environment

Thank you for considering to take part in this research study. I would be grateful if you would read through the following questions and indicate your response to each of them. The purpose of this is to ensure that you are fully aware of the purpose of the research and that you are willing to take part.

I have been informed about the purpose of the study and have had the opportunity to ask questions about it if I wished **YES/NO**

I understand that I can withdraw from the study at any stage, without giving a reason and that my data will not be included in the research **YES/NO**

I understand that I am free to choose not to answer a question without giving a reason why **YES/NO**

I understand that any information I provide will be kept confidential and only the researchers listed above will have access to the information. **YES/NO**

I give my consent to take part in the research.
Participant
Signed ……………………………………….
NAME IN BLOCK LETTERS ……………………………………….
Date ……………………………………….

Researcher
Signed ……………………………………….
NAME IN BLOCK LETTERS ……………………………………….
Date ……………………………………….

Information Sheet

Division of Psychology
E Floor, Richmond Building
University of Bradford
Richmond Road
Bradford.

Researchers: Dr Valerie Lesk, Prof. Hassan Ugail, Syadiah Shamsuddin and Elizabeth Walters

Testing spatial memory using a virtual reality environment

Thank you for taking part in this study.

You will shortly be given a computerised experiment testing your spatial memory. You will need to follow a red ribbon seen on the computer screen in order to reach your destination. This will be done three times. **You will be required to try to remember the route as you follow the red ribbon as the fourth time you will need to find your destination without the help of the ribbon**. There are 5 levels of increasing difficulty.

After you have done this task, there will be a short break when we will ask you to complete two simple questionnaires. The researcher will then give you a number of neuropsychological tests investigating your general cognitive function.

The computerised experiment and the cognitive tests are considered fun and enjoyable.

All the information that is collected from you during this research will be kept secure and any identifying material, such as your name will be removed in order to ensure your anonymity.

If you are happy to proceed, you will be given this information sheet to keep and asked to sign a consent form. You are still free to withdraw at any time without giving a reason.

Thank you for taking the time to read this information sheet and taking part in this research.

Dr Valerie Lesk and Prof. Hassan Ugail
E-mail: v.lesk@bradford.ac.uk
Tel: 01274 236832

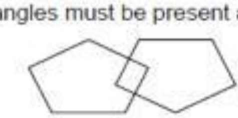
Appendix B

Mini Mental State Examination

Mini-Mental State Examination (MMSE)

Patient's Name: ______________________________ Date: ____________

Instructions: ***Ask the questions in the order listed. Score one point for each correct response within each question or activity.***

Maximum Score	Patient's Score	Questions
5		"What is the year? Season? Date? Day of the week? Month?"
5		"Where are we now: State? County? Town/city? Hospital? Floor?"
3		The examiner names three unrelated objects clearly and slowly, then asks the patient to name all three of them. The patient's response is used for scoring. The examiner repeats them until patient learns all of them, if possible. Number of trials: ___________
5		"I would like you to count backward from 100 by sevens." (93, 86, 79, 72, 65, …) Stop after five answers. Alternative: "Spell WORLD backwards." (D-L-R-O-W)
3		"Earlier I told you the names of three things. Can you tell me what those were?"
2		Show the patient two simple objects, such as a wristwatch and a pencil, and ask the patient to name them.
1		"Repeat the phrase: 'No ifs, ands, or buts.'"
3		"Take the paper in your right hand, fold it in half, and put it on the floor." (The examiner gives the patient a piece of blank paper.)
1		"Please read this and do what it says." (Written instruction is "Close your eyes.")
1		"Make up and write a sentence about anything." (This sentence must contain a noun and a verb.)
1		"Please copy this picture." (The examiner gives the patient a blank piece of paper and asks him/her to draw the symbol below. All 10 angles must be present and two must intersect.) 
30		TOTAL

Appendix C

Usability Questionnaire

Personal Details.

Age	:
Gender	:
Previous Virtual Reality Experience	: Yes ☐ No ☐

Please indicate with a number from 1 to 5 whether you agree with the following statements.

1 = Completely disagree 5 = Completely agree

1. I had quickly learned how to move through the virtual application.						
Completely disagree	1	2	3	4	5	Completely agree

2. I found it easy to move myself through the virtual application.						
Completely disagree	1	2	3	4	5	Completely agree
4.The Instructions for the tasks are clear.						
Completely disagree	1	2	3	4	5	Completely agree
5. I liked using the arrow keys to move around the application.						
Completely disagree	1	2	3	4	5	Completely agree
6. I often feel disoriented in using the application.						
Completely disagree	1	2	3	4	5	Completely agree
7. It was easy for me to look around in the virtual application.						
Completely disagree	1	2	3	4	5	Completely agree

8. I would have preferred to have more freedom of movement.						
Completely disagree	1	2	3	4	5	Completely agree
9. The speed of movement was too fast.						
Completely disagree	1	2	3	4	5	Completely agree
10. Overall, I feel comfortable using the application.						
Completely disagree	1	2	3	4	5	Completely agree
11. Overall, I am satisfied with the experience of using the application.						
Completely disagree	1	2	3	4	5	Completely agree

Any other comments ?

____

____

____

____

____

____

__________________________________**THANK YOU.**