DIGITAL OCULOMOTOR BIOMARKERS IN DEMENTIA

Kyriaki Mengoudi

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Computer Science
University College London
June 8, 2021

I, Kyriaki Mengoudi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Dementia is an umbrella term that covers a number of neurodegenerative syndromes featuring gradual disturbance of various cognitive functions that are severe enough to interfere with tasks of daily life. The diagnosis of dementia occurs frequently when pathological changes have been developing for years, symptoms of cognitive impairment are evident and the quality of life of the patients has already been deteriorated significantly. Although brain imaging and fluid biomarkers allow the monitoring of disease progression in vivo, they are expensive, invasive and not necessarily diagnostic in isolation. Recent studies suggest that eye-tracking technology is an innovative tool that holds promise for accelerating early detection of the disease, as well as, supporting the development of strategies that minimise impairment during every day activities. However, the optimal methods for quantitative evaluation of oculomotor behaviour during complex and naturalistic tasks in dementia have yet to be determined.

This thesis investigates the development of computational tools and techniques to analyse eye movements of dementia patients and healthy controls under naturalistic and less constrained scenarios to identify novel digital oculomotor biomarkers. Three key contributions are made. First, the evaluation of the role of environment during navigation in patients with typical Alzheimer disease and Posterior Cortical Atrophy compared to a control group using a combination of eye movement and egocentric video analysis. Secondly, the development of a novel method of extracting salient features directly from the raw eye-tracking data of a mixed sample of dementia patients during a novel instruction-less cognitive test to detect oculomotor biomarkers of dementia-related cognitive dysfunction. Third, the application of unsupervised anomaly detection techniques for visualisation of oculomotor anomalies during various cognitive tasks.

The work presented in this thesis furthers our understanding of dementia-related oculomotor dysfunction and gives future research direction for the development of computerised cognitive tests and ecological interventions.

Impact Statement

This thesis is intended to stimulate a paradigm-shift in attitudes toward the neuropsychology of dementia. As the prevalence of dementia continues to rise rapidly, there is a pressing need for greater insight into the nature and timing of the earliest subtle changes in cognition, and how these changes can best be measured. This study focuses on the investigation of oculomotor biomarkers using eye-tracking technology during navigation in a stimulated real-world environment and free-viewing of images as part of a novel instruction-less cognitive assessment. Data provided from the Computational PLatform for Assessment of Cognition In Dementia (C-PLACID) and Seing What They See (SWTS) programmes were analysed that aim to enhance the cognitive assessment of people with dementia and advance our understanding of the functional impact of dementia-related visual impairment respectively. The true impact of this study constitutes the empirically demonstrated potential of computational techniques and sensing systems to enable a change in approach toward addressing the current problems which impose a number of limitations.

The results in this study present evidence that firstly brief instruction-less eye-tracking tests can detect abnormal oculomotor biomarkers and secondly, representation learning techniques can extract more discriminative features than standard handcrafted eye-tracking metrics from an instruction-less eye-tracking cognitive test. These findings have implications for the patients, clinicians and researchers. They open the window to shorter, more personalised neuropsychological tests for patients with different educational and cultural backgrounds. They also offer the potential for remote testing, which can be of great value in studies seeking to screen or recruit large numbers of people. The novel eye-tracking cognitive measures demonstrate the importance of augmenting cognitive assessments with cognitively-relevant physiological data and might offer future potential as outcomes in clinical trials.

The results support that representation learning and anomaly detection techniques are suitable computational methods for neuropsychological problems that can be used by researchers for similar problems of interest. Additionally, these methods provide informative insights into individual abnormal cases that could be drivers of improvement and enrichment in neuropsychological testing and experiment design. Finally, the multi-modal approaches used in this thesis combining eye-movement with egocentric videos and motion capture data can be applied in future home-based interventions in dementia and other diseases in order to improve independence and quality of life of individuals and their caregivers.

The broader implications that derive from the attempt of this work to develop better cognitive tests, constitute the reduction of healthcare costs for dementia by improving

diagnosis, particularly in the earliest stages of disease when some established tests lack sensitivity and particularly for rarer types of dementia. Improved cognitive assessments will also play a critical role in validating and monitoring the effectiveness of novel therapies; secondary benefits of this contribution will include boosting the economy by expediting the process of bringing those treatments to market, and relieving suffering of individuals with dementia and their carers.

Acknowledgements

There are numerous people I would like to thank for supporting me during my PhD and sharing the journey with me.

Foremost, my supervisors. Prof. Daniel Alexander, for his expert guidance, open-minded spirit and inspiring enthusiasm for research, innovation and great ideas; Prof. Sebastian Crutch, for his creative thinking, bright ideas and support; Dr. Nicholas Firth, for helping me in the first years of my academic journey. Finally, my biggest thanks go to Dr. Keir Yong and Dr. Daniele Ravi who were always there for me and from whom I learnt so many things about science and life in general!

I would also like to thank the POND group, the VID group from the Dementia Research Centre and all the wonderful people from the Created Out of Mind project with whom I was delighted to collaborate. Also, many thanks to the CMIC family (especially my beloved friend Maura) and all the people that contributed to having such a warm and friendly working environment for the last four years. I am so grateful for all the friends I made within CMIC from all over the world!

Many thanks to my friends, flatmates and my family that were supportive and patient with me. I am also grateful for all my yoga and dance teachers, of course Despina and all inspiring artsy events I went to in London - this thesis couldn't be possible without them. And lastly, a big thanks to myself for all the hard work, resilience and courage. This PhD was an eye-opening experience for me that helped me grow as a researcher and most importantly as a person; hopefully more are to come in the future!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Dementia, Alzheimer's disease and subtypes

Dementia is a syndrome describing the progressive deterioration of cognitive and functional abilities. It is not a specific disorder or disease, but a description for a group of symptoms associated with the gradual disturbance of cognitive functions including memory, reasoning and judgement among others, that are severe enough to interfere with tasks of daily life, and not associated with loss of consciousness [88]. Deterioration in emotional control, behaviour and motivation is commonly observed along with cognitive impairment [139]. Dementia is an issue of enormous medical and socioeconomic significance particularly in societies with rapidly aging populations; 35.6 million people are currently living with dementia worldwide and this number is expected to double by 2030 and more than triple by 2050 [139].

The most prevalent dementia is typical Alzheimer's disease (tAD) and it is a chronic progressive neurodegenerative disorder that usually affects people over 65 years of age [5]. While characterised by gradual and progressive episodic memory impairment (the memory of autobiographical events), it is also linked with other cognitive impairments such as executive dysfunction (planning, self-control, focus), language and complex visual processing deficits [173, 116]. Amyloid protein deposits and intracellular neurofibrillary tangles are the hallmark pathological changes progressively invading the cerebral cortex, accompanied by global brain atrophy with particular burden on the temporal lobes and medial temporal structures [5]. Despite typical memory led AD being the most well recognised form of AD, atypical AD presentations exist characterised by predominant visual, language, behavioural/executive or motor presentations, with relatively spared memory.

Posterior Cortical Atrophy (PCA) is a usually young-onset (diagnosed in people under the age of 65) neurodegenerative syndrome characterised by a progressive decline primarily in visuoperceptual and visuospatial processing and dysfunction that depends on parietal occipital and occipitotemporal regions of the brain [41]. PCA is most often caused by Alzheimer's disease pathology (often being referred to as the visual variant of AD), although other underlying causes can include dementia with Lewy bodies, corticobasal degeneration or prion disease [41]. Common symptoms

include difficulties in tasks that require locating, interpreting and reaching items under guidance and other everyday activities such as reading, writing, spelling and driving [43]. It is distinguished from more common amnestic presentations by the fact that usually memory and language remain relatively preserved at least in earlier stages of the disease course [43].

Frontotemporal dementia (FTD) is the second most common young-onset dementia. It is clinically characterised by gradual changes in behaviour including social dysfunction, apathy, executive dysfunction in planning, organisation, problem solving and language difficulties. Memory, visual perception and spatial skills are usually relatively well preserved [152]. The main subtypes of FTD are the following distinct syndromes: the behavioural variant of FTD (bvFTD) and the language variants, semantic variant (svPPA) and progressive non-fluent variant of primary progressive aphasia (nfvPPA) [152]. In behavioural-variant FTD, brain atrophy is observed in symmetrical frontal and anterior temporal lobe regions and striking changes are reflected in behaviour and personality including apathy, lack of insight, reduced empathy and altered preference for food [152]. In semantic variants, the areas of the brain affected first are in the front of the left temporal lobe, dealing with verbal semantic memory which indicates reduced single-word comprehension and impaired object knowledge but preserved fluent speech abilities [100]. In contrast to svPPA, nfvPPA presents with a nonfluent expressive language disturbance characterised with phonological errors and agrammatism as the atrophy is present in the left frontal regions [152].

In recent years, logopenic variant primary progressive aphasia (lvPPA), another language related syndrome, has been described and is commonly associated with Alzheimer's disease characterised by difficulty finding words in spontaneous speech and repeating sentences and phrases. However, people living with lvPPA do not present impairment in understanding words, as in the case of svPPA. Preliminary studies suggest predominant left temporo-parietal involvement in this disorder [19].

The phenotypic variability across different forms of dementia is a well-known challenge in both clinical practice and research [178]. This heterogeneity has led to inconsistencies in diagnosis and poor support management. Greater understanding and awareness for the specific syndromes will enable accurate diagnosis, facilitate non-pharmacological disease modifying treatment trials and provide new insights into degenerative diseases.

1.1.1 Biomarkers and diagnosis

Standard diagnostic practice of dementia involves the evaluation of a person for whom there has been expressed some concern related to a change in cognitive function or behaviour. It starts with the assessment of the clinical history of the individual, followed by a neurological and cognitive examination and an interview with a relative [40]. Before concluding to a dementia diagnosis, different examinations are made to reject the potential existence of other physical or mental diseases that contribute to cognitive impairment (e.g. depression). A neuroradiological examination with Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is recently recommended to exclude conditions with similar clinical phenotype, mistaken for dementia,

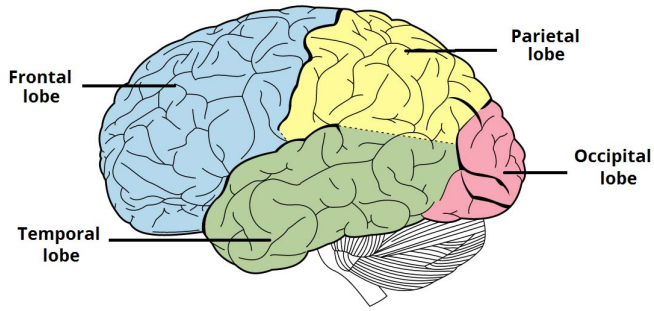

Figure 1.1: The celebral cortex of one hemisphere of a human brain is divided into four lobes: a. parietal: handles information from our senses about space, perception, size and integration of sensory and motor functions, b. temporal: deals with memory (including recognition of faces/objects) and language, c. occipital: deals with visual information, d. frontal: involved in movement, decision-making, problem-solving, and planning. Source: https://commons.wikimedia.org/wiki/File:Lobes_of_the_brain_NL.svg

but different brain structure causes such as subdural haematoma. Finally, if a deficit is detected in more than two cognitive domains at an extent that impairs functional activities and additionally, the symptoms progress over time, then a dementia diagnosis is defined [40].

The diagnosis of dementia occurs frequently when pathological changes have been developing for years, symptoms of cognitive impairment are evident and the functional independence of the patients has already been deteriorated significantly [75, 124]. Based on our current understanding, most subtypes of dementia develop over years, starting from an asymptomatic period where pathological changes accumulate in the absence of clinical manifestations, through subtle cognitive, behavioural or personality impairments and finally multiple cognitive domains are affected, as well as, noticeable decline appears in executive functions [40]. An early diagnosis of the disease is believed to offer the opportunity to reduce psychiatric symptoms associated with the disease (e.g. agitation), improve cognition through pharmacological (e.g. Donepezil) and non-pharmacological (e.g. cognitive stimulation therapy) interventions and help caregivers to carefully plan and adjust their lives [48]. Moreover, previous research supports that at least Alzheimer's disease pathological changes are present up to 20 years before its evident manifestations [11]. Presymptomatic diagnosis, therefore, might be feasible with the investigation of different sensitive and appropriate biomarkers related to the different types of dementia [11].

A biomarker is a "characteristic that can be objectively measured and evaluated as an indicator of normal biological or pathogenic processes or pharmacological responses to a therapeutic intervention" [2]. Reproducibility, availability and direct reflection of the disease process are considered to be some characteristics of the ideal biomarkers for dementia [2]. At present, the biomarkers that have been developed to monitor neuronal atrophy of the brain *in vivo* and are being used in clinic can be divided into imaging modalities and cerebrospinal fluid (CSF) measures. Blood-based or urine-based biomarkers are recommended although are not available for routine clinical use. However these methods have advanced our understanding of the dis-

ease, they are expensive, require exposure to radiation (PET imaging) and are invasive [105]. There is also considerable overlap in the clinical and pathological presentations in different forms of dementias, and indeed differences in clinical presentation in patients with the same underlying pathology (e.g. memory problems in typical AD vs visual problems in PCA). Some researchers, therefore, discourage diagnosis based solely on neuropathological criteria [151].

1.1.2 Cognitive Testing

Cognitive tests are used to measure changes in cognitive functioning due to neurodegenerative diseases [38]. The purpose of cognitive assessment is to eliminate the use of subjective and self-reported measures and rather provide an objective, quick to administer and minimally invasive evaluation of cognitive functions. Patients are typically assessed on a variety of tasks, each of which examines different cognitive domains such as attention, memory, language, and executive functions [10]. Behavioural observations and qualitative evaluation could also provide additional information for a patient's cognitive profile [10].

Cognitive scores are not meaningful in their raw form because performance variations are influenced by demographic factors [10]. A frame of reference is therefore required for impairment to be defined. Normative data are typically obtained from a large sample of cognitive healthy individuals appropriately stratified by demographics, reflecting healthy performance on a specific test. They are also adjusted for relevant demographic factors including age, education and sex [15]. Individuals performance is compared and contrasted to this reference group. These scores determine the cut-off points that define the level of performance associated with impairment. In clinical practice, impaired performance is conventionally defined as below the 5th or 1st percentile based on normative data [51].

The list of neuropsychological tests used in clinical and research settings is rich and diverse. The Folstein Mini-Mental State Examination (MMSE) is the most widely used test. The MMSE assessment consists of 11 –items that evaluate the cognitive function of the following domains: attention and orientation, memory, registration, recall, calculation, language and ability to draw a complex polygon [67]. The administration time is approximately seven minutes for a person with dementia and five minutes for cognitive normal individuals. Apart from MMSE, there are more than 40 tests for dementia including the Addenbrooke's Cognitive Examination Revised (ACE-R), the Mini-Cog test, the General Practitioner Assessment of Cognition (GPCOG) and the Montreal Cognitive Assessment (MoCA).

Due to the diverse properties of the batteries, different cognitive tests are used for dementia screening, staging or evaluating longitudinal change. Screening tools are used to support early diagnosis of the disease detecting subtle changes prior to clear impairment in daily functions and they are not intended to be diagnostic [40]. The Mini-Mental State Examination14 (MMSE), Montreal Cognitive Assessment (MoCA) and Addenbrooke Cognitive Examination (ACE) are commonly used for this purpose. The MoCA having an increased focus in multiple cognitive domains might outperform MMSE in detecting early cognitive changes [166]. On the other hand, MMSE and

ADAS-COG are more sensitive for determining the stage of the disease (e.g. mild or sever AD) and the longitudinal decline during the symptomatic phase.

1.2 Research Problem

Given the defining characteristics of most dementia syndromes are primarily cognitive in nature, assessment of a person's cognition is a vital component of both diagnostic services and research investigations, and is the most common outcome measure by which the effectiveness of potential pharmaceutical and non-pharmaceutical therapies is judged. Standardised paper-and-pencil cognitive assessment tools are a key component of the screening and diagnostic process, but have a number of limitations:

1. Psychometrics: Accurate assessments are long and associated with participant fatigue and stress, but brief tests often elicit floor and ceiling effects owing to a lack of dynamic range [122]. Particularly in longitudinal studies, variation in disease severity means that reduced variability in participant's scores is common. Time and patient effort is wasted completing items that are too easy or hard to contribute to ascertainment of an individual's exact level of functioning.

2. Administration: Practice effects and complex task demands mask longitudinal change and precise performance in certain cognitive domains. Practice effects in sequential assessments expressed in the form of reduced anxiety or improved performance due to test familiarity may hide evidence of cognitive decline or instability. Moreover, many tasks purport to measure one skill (e.g. Cogstate spatial problem solving) but are confounded by task demands that utilise other skills (e.g. complex verbal test instructions that aphasic participants fail to understand).

3. Quantification: Current routine tests fail to capture critical, sensitive aspects of task performance that alternative performance measures might be able to record (e.g. eye-tracking). For instance, many tasks capture accuracy data but miss other sensitive performance measures such as vocal reaction time and other measures reflecting higher-order cognition in dementia. In addition, cognitive profiles of individuals which characterise cognitive abilities across several cognitive domains and tasks are usually described qualitatively because test properties and normative samples differ across tasks. Therefore, there is need for quantitated scores across tasks that capture different aspects of performance and are not biased by assessor's subjectivity.

4. Statistical analysis: Cognition has traditionally been assessed using standardised test batteries administered to large numbers of people, providing information on average performance in broad domains (e.g. information-processing speed, memory, executive function). Two major drawbacks with this approach are: statistics based on accuracy or response latency are highly reductive, overlooking informative sources of variability (e.g. effort levels, response strategies). Individual differences are subsumed into descriptive summary statistics

(e.g. group means), which fail to account for heterogeneity in task performance. Moreover, performance across cognitive tasks is not independent. General factors (e.g. disease severity) and collateral deficits (e.g. language problems limiting performance on a verbal memory test) mean a multivariate rather than (mass) univariate approach is required.

5. Ecological validity: Certain domains (e.g. social cognition) and complex cognitive functions (e.g. navigation) are poorly assessed via traditional paper-and-pencil tests. Frontotemporal dementia patients exhibit profoundly abnormal behaviours in social settings (e.g. swearing; inappropriate comments; touching children; loss of empathy) but it is currently impossible to recreate those scenarios in the normal clinical or research environment. Static photo or picture-based tests of skills such as emotion recognition often lack ecological validity. Even more dynamic video-based tests (e.g. TASIT2) lack personal engagement and require advanced linguistic skills (e.g. accurate labelling of emotional states).

As the prevalence of these devastating conditions continues to rise rapidly, there is pressing need for greater insight into the nature and timing of the earliest subtle changes in cognition, and how these changes can best be measured. Recent studies suggest that eye-tracking-based cognitive assessment might ameliorate some of the existing problems as it enables a brief and quantitative evaluation of cognitive functions [128, 21, 138]. Eye-tracking provides fine-grained information regarding oculomotor information (pupil dilation and gaze) that provides additional information about the association between brain and behaviour and has been used to uncover eye movement abnormalities in different dementia syndromes [120]. It has also the potential to alleviate some problems of standard paper-and-pencil cognitive tests related to administration, quantification and ecological validity. Novel eye-tracking tests might be a window to a robust, natural and less complex and linguistically demanding evaluation of cognition.

Most previous studies in the context of dementia have used eye-tracking to look at basic oculomotor functions and checked if those have a relationship to disease [154, 66, 25]. A small number of studies has shown that eye-tracking metrics can be used as an outcome measure for evaluation of particular higher-order cognitive functions (e.g. memory, attention) [44, 145, 62, 138]. Although the eye-tracking measures from these studies capture critical aspects of task performance, the tests are still susceptible to the need for instructing patients on how to complete the tasks, which is prone to mistakes caused by misunderstandings, language difficulties or patients at the later stages of the disease. Additionally, although several investigations have explored dementia oculomotor biomarkers in controlled oculomotor tasks, more experiments are necessary under naturalistic scenarios. Greater understanding of eye movement abnormalities during Activities of Daily Living (ADL) might support the detection of early signs of dementia [11].

Identification of oculomotor biomarkers in dementia is still in its infancy. Apart from the need for more well-designed studies and cognitive tests for individuals with different disease severity and type that investigate more ecologically valid behaviour, the complexity of eye-tracking data constitutes a major challenge that should be

addressed. The transition from simple experimental tasks (e.g. anti-saccadic, fixation stability - for example, in continuously fixating a dot) to more complex ones requires the establishment of appropriate methods to analyse the dense and dynamically changing time series of eye movements. While complex computational methods are used for analysis of neuroimaging data in dementia research to detect changes in brain atrophy, only a few studies have attempted to apply similar methods to investigate cognitive changes manifested through eye-tracking datasets. So far the eye-tracking metrics used were solely based on the selection and intuition of neuropsychologists that were spending their time creating areas of interests one-by-one and visualising individual trials of experiments with hundred of trials to identify abnormal behaviour. Here I tackle the problem of identifying dementia oculomotor biomarkers by harnessing computational methods and artificial intelligent algorithms.

1.2.1 Problem Statement

My thesis will address the following problem:

- Identifying novel less-constrained and ecological valid tests designed to augment dementia cognitive assessment with oculomotor measurements has been investigated in a limited extent.

1.3 Thesis contributions

This thesis investigates the development of computational tools and techniques that enable the identification of novel digital oculomotor biomarkers of dementia under naturalistic and less constrained scenarios. With a multidisciplinary focus, it attempts to bridge the gap between computer science and neuropsychology developing algorithms that have an impact on our understanding of oculomotor abnormalities of dementia patients relative to controls.

Three contributions were made in this thesis:

1. **Investigating the effects of visual environment on navigation in Alzheimer's disease and Posterior Cortical Atrophy**

   This project attempts to address two core limitations of cognitive tests, namely, the lack of ecological validity and quantification of performance measures (described previously). Previous studies and design guidelines suggest that the physical environment may play a major role in mitigating dementia's functional impairment [113]. In this work, we combine eye movement and egocentric video analysis to investigate patients with PCA and tAD compared to a control group performing a real-world visual search task while navigating a controlled environment. The analysis of eye movement patterns in naturalistic settings is achieved through integrating gaze locations and scene information provided by egocentric videos. Computational attention modelling techniques with saliency maps of the

point of view (POV) frames used in [155] are combined with eye-tracking metrics and gait/orientation measures to investigate potential differences between groups in:

- the extent to which environmental features distinctive in colour or orientation predicted fixation position.
- the particularly salient environmental features within POV frames.
- the relationship between saliency at fixation and maximum saliency of POV frames and completion time of the tasks (general measure of functional performance).
- in individual cases by visualisation of trials including information for position in the room, orientation of the head and saliency measures.

Publications:

Yong, K.X., McCarthy, I.D., Poole, T., Ocal, D., Suzuki, A., Suzuki, T., Mengoudi, K., Papadosifos, N., Boampong, D., Tyler, N. and Frost, C., 2020. Effects of lighting variability on locomotion in posterior cortical atrophy. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 6(1), p.e12077.

Mengoudi, K., Firth, N.C., Suzuki, A., McCarthy, I., Suzuki, T., Ocal, D., Papadosifos, N.N., Tyler, N., Boampong, D., Alexander, D.C. and Crutch, S.J., 2018. P1?662: Effects of visual enviroment on fixation and gait parameters in Alzheimer's disease and posterior cortical atrophy. *Alzheimer's & Dementia*, 14(7S_Part_11), pp.P596-P596.

2. **Augmenting Dementia Cognitive Assessment with Instruction-less Eye-tracking Tests**

This project attempts to improve the administration and quantification of performance measures of cognitive tests, introducing novel instruction-less eye-tracking tests that reduce or eliminate task demands'. In this work, we introduce a novel way of detecting abnormal behaviour and automatically extracting salient features from a novel instruction-less eye-tracking cognitive test administered to well-characterised patients with a variety of dementia diagnoses and healthy controls. In more detail, the following contributions are made:

- We introduce a novel method for extracting features from instruction-less eye-tracking cognitive tests. Our approach is based on self-supervised representation learning where, by training initially a deep neural network to solve a pretext task using well-defined available labels (e.g. cognitive activity recognition in healthy individuals), the network encodes high-level semantic information which is useful for solving other problems of interest (e.g. dementia classification).
- Inspired by previous work in explainable AI, we use the Layer-wise Relevance Propagation (LRP) technique to describe our network's decisions in differentiating between the distinct cognitive activities.

- The extent to which eye-tracking features of dementia patients deviate from healthy behaviour is then explored, followed by a comparison between self-supervised and handcrafted representations on discriminating between participants with and without dementia.
- Our findings not only reveal novel self-supervised learning features that are more sensitive than handcrafted features in detecting performance differences between participants with and without dementia across a variety of tasks, but also validate that instruction-less eye-tracking tests can detect oculomotor biomarkers of dementia-related cognitive dysfunction.
- This work highlights the contribution of self-supervised representation learning techniques in biomedical applications where the small number of patients and the complexity of the setting can be a challenge using state-of-the-art feature extraction methods.

Publications:

Mengoudi, K., Ravi, D., Yong, K.X., Primativo, S., Pavisic, I.M., Brotherhood, E., Lu, K., Schott, J.M., Crutch, S.J. and Alexander, D.C., 2020. Augmenting Dementia Cognitive Assessment With Instruction-Less Eye-Tracking Tests. *IEEE journal of biomedical and health informatics*, 24(11), pp.3066-3075.

Mengoudi, K., Ravi, D., Yong, K.X., Primativo, S., Pavisic, I.M., Brotherhood, E., Lu, K., Schott, J.M., Crutch, S.J. and Alexander, D.C., 2020. Augmenting Dementia Cognitive Assessment With Instruction-Less Eye-Tracking Tests: A Machine Learning Approach for Detecting Abnormal Oculomotor Biomarkers. *Alzheimer's & Dementia.*

3. **Visualising oculomotor abnormalities based on unsupervised anomaly detection**

This project builds on the previous work on instruction-less tests and provides a data-driven way of detecting individual cases of abnormal trials during free-viewing of scenes for further clinical relevance and interpretation from neuropsychologists. The following contributions are made:

- We propose an unsupervised framework for anomaly detection in sequential data, based on representation learning using convolutional autoencoders. This method is well-suited to our problem of biomarker discovery when small number of patients data are only available since the models require only controls data to be trained.
- Selection and visualisation of abnormal trials based on the model's ranked anomaly scores for different dementia types and cognitive tasks.
- This work establishes a starting-point for getting further insights into eye movement abnormalities which are of greatest importance given the load of available data and the instruction-less nature of the tasks that render very difficult the prediction of anomalies even from experts.

1.3.1 Structure of this thesis

To identify ecological valid and less-constrained eye-tracking tests for dementia cognitive assessment, data from two novel neuropsychological experiments were explored in this thesis. The two batteries included participants (controls and dementia patients) navigating in a naturalistic setting and viewing images in a computer screen without any instructions given, respectively.

I firstly investigated in Chapter 3 oculomotor abnormalities during activities of daily living (i.e. navigation in a naturalistic setting using a mobile eye-tracker) which poses the challenge of integrating multi-modal datasets of egocentric videos and low frequency eye movement time series. To address the limitations of this work including the crude accuracy of the eye-tracker, a high frequency eye-tracker was used in the study presented in Chapter 4 under a more-constrained but ecological setting (viewing naturalistic images without following any instructions). Analytic approaches that identify properties of the complex eye-tracking time series that discriminate between the dementia and controls group were explored. Finally, in Chapter 5 I approached the task of identifying digital oculomotor biomarkers in instruction-less tests as an anomaly detection problem by defining impairment based on normative eye-tracking data in line with standard neuropsychological practices of defining abnormality.

The thesis has the following structure:

- Chapter 2 contains firstly background information about eye-tracking in dementia research in terms of evidence from existing biomarkers and data analysis techniques used. Then, as this thesis proposes the analysis of eye-tracking data as time series, the general processing pipeline used for biosignals is presented, followed by a more detailed description about the feature extraction, prediction and anomaly detection phases.
- Chapter 3 contains the first project on evaluation of clinical/environmental factors relating to function in naturalistic settings.
- Chapter 4 contains the second project on self-supervised representation learning of eye movement data from instruction-less cognitive tests.
- Chapter 5 contains the last project on visualisations of oculomotor abnormalities based on an unsupervised learning technique.
- Chapter 6 presents a summary of the work in this thesis, and proposes directions for further research.

Chapter 2

Background

Various eye-tracking approaches have been developed to study the oculomotor profile of people with neurodegenerative conditions. Among them there is a clear distinction between studies that characterise basic oculomotor function in people with dementia with no explicit reference to cognition (e.g. do people with AD have more saccadic intrusions than healthy age-matched controls) and those which use eye position as an outcome measure in a cognitive task designed to measure a particular ability. The aim of this thesis is to push the state of the art in the latter line of research in order to identify novel cognitive tests based on oculomotor measurements. To this aim, in this chapter, we firstly provide a brief overview of the basics of eye-tracking technology and a summary of the oculomotor deficits previously reported in Alzheimer's disease, frontotemporal dementia and posterior cortical atrophy. Secondly, we survey the methods previously used to extract eye-tracking metrics. These can be grouped into three categories: statistics over the pupil dilation or the gaze signal, machine learning methods and computational visual saliency models applied on static stimuli images and egocentric videos. Then, as this thesis investigates the development of computational tools to identify novel digital oculomotor biomarkers from eye-tracking data, we review eye-tracking studies that used machine learning techniques and then describe computational methods for time-series feature extraction in general and provide a comparison between classification and anomaly detection techniques. This chapter concludes with a description of the machine learning methods used in the thesis (including a machine learning glossary for neuropsychologists) and an overview of the identified gaps and limitations from the current state of the art.

2.1 Eye-tracking

One fundamental question that arises when one wants to interpret cognitive and emotional patterns is what measurable information is required to capture these changes. Manifestations of human behaviour can be measured using self-report, observed behaviour and physiology [185]. The first two methods are considered subjective (monitoring/observing questionnaire responses of subjects/patients) and they provide usually qualitative or quantitate data (e.g. ranks) with limited range of responses. Moni-

toring of subjects' physiological responses is considered a more objective route [71]. It is a more reliable neurophysiology-oriented approach based on rich quantitative data with reduced sources of bias. In clinical application in particular, physiology is of primary interest as some people living with dementia present difficulties in communication. Moreover, recent advancements in non-invasive technological sensors have also contributed to the increase of the user's comfort (e.g., reduced size sensors or wearable computers) and has assured the long-term physical contact with the subject (better quality real-time data) [71].

Eye-tracking is the process of measuring eye activity estimating the gaze location (where one looks) and the pupil size over time. It has both a physical (eye movement) and physiological (eye movement, pupil dilation) component as the latter may not be consciously controlled. To better understand the eye movement, the anatomy of the human eye is described here briefly (Figure 2.1). The cornea is a hard and transparent layer which forms the outer part of the eyeball. The visible parts of the eye are the sclera (the white part), the iris (coloured part) and the pupil which is in the centre of the iris and regulates the amount of light coming into the retina by changing its size [121]. The retina is responsible for the transformation of the visual stimuli to electric signals which pass to the visual cortex through the optic nerve [158]. Additionally, behind the pupil, the crystallines lens filter the input image by focusing the image on the retina. The fovea is a special region in the retina that processes high spatial resolution. Based on this information, the point of regard is calculated as the intersection of the axis defined as passing through the fovea and the Line of Sight (visual axis) with the closest object of the scene [121].

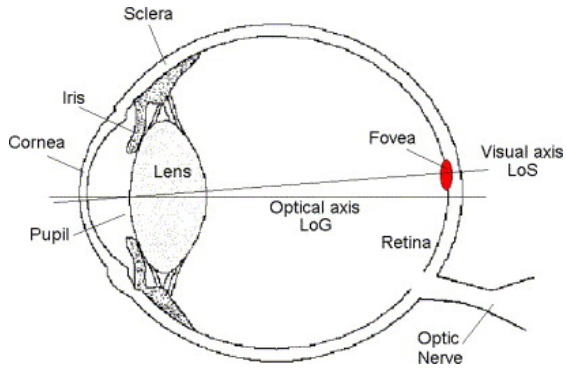

Figure 2.1: The anatomy of the human eye.Image produced with copyright permission from [121].

Eye movements are divided into two broad categories: a. stabilising movements that hold the image of an object on the retina and saccades that move the eye around the visual field bringing different objects to the fovea [158]. The first category consists of either movements in which the gaze is stable in one location called fixations or smooth pursuit and nystagmus in which the gaze is not stable (although the eyes look the same object) because of head or object motion. During fixations, the eyes are held fairly stable and they last between between 100-1000 ms, with the majority being between 200-500 ms depending on the task. The second category includes vergence movements which are involved in rotating the eyes in same or opposite directions and saccades that are rapid eye movements used to change the location of the fovea and

thus the position of the fixation.

In Figure 2.2 the time course of a saccadic movement is shown with the corresponding velocity and position profiles. Peak velocity and duration is calculated from the velocity profile, as the highest velocity during the saccadic movement and the time to complete the saccadic movement respectively. Saccadic amplitude is another commonly used measure that defines the size of the saccade (measured in degrees or mins) by computing the difference in gaze location before and after the initiation of the saccade [140]. The velocity of saccades rises to a maximum value which is approximately in the midpoint of the movement and then it drops until the new target location is reached. Peak accelerations can reach 40000 deg/s^2 [1] and peak velocities vary between 400 and 600 deg/s, depending on the amplitude of the saccade. The duration of saccades is influenced by the task and the distance covered [143]. Moreover, previous research has shown that during fixations information is taken in and during saccades new information is not obtained because the very rapid moving of the eyes allow only blur to be perceived [143].

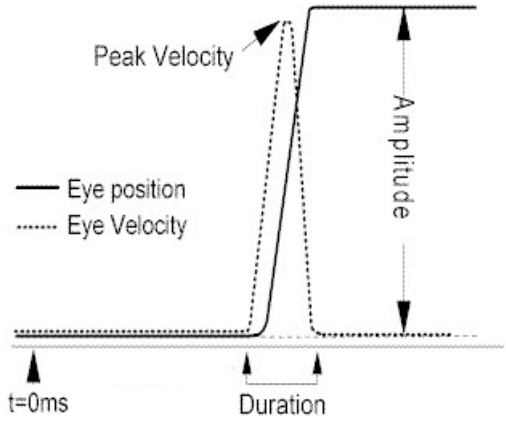

Figure 2.2: Characteristics of the velocity and position profile of a saccade between two fixations. Source: https://www.liverpool.ac.uk/~pcknox/teaching/Eymovs/params.htm.

Eye movements can be recorded with electrooculogram (EOG) signals measuring the corneo-retinal potential occurring between the back and the front of the human eye. During this method two electrodes are usually placed on either sides of the eye (left and right or above and below). Another more popular technique is the recording of eye movement with electronic devices called eye-trackers. The main advance in the last years in these technologies is the adoption of the video-based eye tracking technique in contrast to previously commonly used electrooculography, infrared reflection, search coil and dual purkinje tracking [121]. Video-based eye trackers have a camera (or cameras) which take a series of images of the eyes and an infrared illuminator (reflection of a fixed light source) next to the camera. The eye-tracking software uses image processing algorithms to identify in all the images taken by the cameras the centre of the pupil and centre of the corneal reflection (CR). Then, mathematical

[1] One degree of visual angle spans approximately 1 cm on a distance of 57 cm from the viewer's eye.

algorithms are used with the help of a calibration procedure (in which the participants look at predefined markers) to map the location of the centre of the pupil in images to screen coordinates in pixels [80]. There are both static and mobile eye-trackers that facilitate different purposes of research.

In practice, eye-tracking devices provide a 3-dimensional vector of x and y coordinates of gaze and pupil diameter size over the course of the experimental setting. The amount of the data produced depends on the sampling rate used by the sensor which can vary between 30 and 1000 Hz. Eye movement events such as fixations and saccades are identified from raw eye movement data by algorithms, which use information on spatial dispersion and temporal characteristics of raw data by the eye-tracking software [149].

2.2 Eye-tracking in Dementia

Recent technological advancement in hardware and software has enabled eye-tracking systems to collect behavioural information that accurately reflects the strategies people use to inspect visual stimuli and show preference over areas of interest [156]. As vision is one of the most dominant senses in everyday activities, eye-tracking technology provides fertile ground for different applications in a variety of research areas including neuroscience and psychology, psycholinguistics and healthcare, user experience and interaction, education, consumer research and marketing [95]. In dementia research, eye-tracking has been utilised to quantify cognitive functions and subsequently measure deviations from healthy cognitive profiles.

In this section, we provide the reader with a brief overview firstly of some cognitive functions that have been assessed with eye-tracking technology and secondly a summary of the oculomotor deficits in Alzheimer's disease, frontotemporal dementia and posterior cortical atrophy.

A cognitive domain that has been evaluated to an increasing extent with eye-tracking technology is executive functions which are cognitive processes associated with ones ability to initiate, inhibit and plan behaviour. The anti-saccade task is commonly used to evaluate this cognitive domain, where the subject is requested to suppress saccades towards a specific target and instead generate saccades in the opposite direction [78]. There are also suggestions that eye movements reveal different mnemonic processes that can be evaluated with tests such as the visual paired comparison task (VPC). This task involves firstly the presentation of an item by itself and then after a small delay, the previous presented item is presented side-by-side with a new item. The amount of time spent exploring each item is the measure of interest [44]. Additionally, different neuropsychological tests assessing language and social cognition have been recently adapted to be available for people with verbal and motor impairment [136]. Visual search tasks (active scan of a visual scene searching for a particular object among other objects) have been benefitted greatly from the use of eye-tracking technology, providing a continuous window on the allocation of attention [137]. The investigation of spatial navigation and the interaction of people with the surrounding environment has been also facilitated by eye-tracking technology, mainly

in laboratories, providing information about allocation of perceptual attention [93].

2.2.1 Alzheimer's Disease

Oculomotor testing in Alzheimer's disease reveals a range of eye movement abnormalities associated with impaired attentional processing, working memory, spatial disorientation and episodic memory. Saccadic intrusions during fixation are among the most common oculomotor features reported in tAD patients and are correlated with disease severity [90]. Saccadic intrusions are described as a pair of horizontal saccades, made in opposite directions that cause little change in eye position due to the corrective nature of the second saccade. These unwanted microsaccades support the presence of gaze-fixation instability in tAD. A number of studies also report increased latency for visually guided saccades (pro-saccades) [66, 25, 39]. This indicates a delay in the initiation of eye movement towards the presented target compared to controls. Moreover, in antisaccade tasks, patients with tAD have shown more antisaccade errors with fewer corrections than control groups [92, 174].

Apart from prossacade and antisaccade tasks that have been popular due to their simplicity, research studies have investigated more complex tasks such as reading, visual search and spatial orientation among others [63, 61, 155]. Findings support that mild Alzheimer's disease patients produce shorter outgoing saccades when reading sentences. A number of studies indicate that patients with tAD present longer reaction times and number of fixations in visual search and exploration compared to age-matched older adults [53, 169]. Lagun et al. [101] also found differences between healthy participants, mild cognitive impairment and AD patients in the memory recognition related VPC task using a machine learning method.

Notably, eye-tracking measures can offer additional information to augment cognitive assessment in dementia. Nevertheless, attentional dysfunction and disease severity may interfere with oculomotor control and patients cooperation to perform the task. A few studies have attempted to evaluate eye movement abnormalities on scenarios that allow naturalistic assessment. Davis and Ohman [47] investigated way-finding using VR and eye-tracking to assess whether salient cues make the environment more supportive for older adults with tAD. Eye movement during locomotion has been investigated with case studies by Suzuki et al. [164], but the area remains unexplored in AD with no studies providing quantitative findings at the group level.

2.2.2 Frontotemporal Dementia

Few eye-tracking findings have been reported for patients diagnosed with FTD. Individuals with FTD show impaired saccadic eye movements with increase pro-saccades latency and higher rates of antisaccadic errors [72]. Regarding complex cognitive tasks, Primativo et al. [138] developed a computerised version of the Brixton spatial anticipation task to evaluate executive functions in bvFTD patients. In this test, the participants must predict the location of a target which is following a specific pattern and thus it measures a person's ability to detect and follow a rule, as well, being cognitive flexible in new rules. Findings suggest that bvFTD patients produced

less correct and more incorrect anticipatory saccades compared with healthy controls and svPPA. Regarding language variants, Faria et al. [56] provided evidence using eye-tracking that patients with svPPA show uncertainty in matching names to objects. Finally, some studies evaluated pupillary responses as a biomarker of behaviour in language impaired dementias. Findings demonstrate that auditory salience (approaching 'looming' versus withdrawing sounds) differentially affected nfvPPA and svPPA patients; with looming sounds inducing greater pupil dilation in healthy controls and svPPA compared to nfvPPA [65].

2.2.3 Posterior Cortical Atrophy

In PCA, the first detailed assessment of oculomotor functions indicates that the most prominent oculomotor abnormalities were increased time to saccadic target fixation, increased first major saccade latency and decreased saccade amplitude. Also, the PCA patients show large saccadic intrusions, more saccades and lower pursuit gain in sinusoidal pursuit, but normal peak velocity of saccades [154]. Another line of research investigated scene perception in PCA using eye-tracking to evaluate the relationship between visual saliency (i.e., brightness and contrast in low-level stimulus features) and fixation location. Case studies from visual agnostic subjects reported contradictory results with some work demonstrating that saliency is a good predictor of fixation and some others indicating that top-down processes still have an effect in scene scanning [110, 68, 69]. In the group level, the only work so far that attempted to distinguish between top-down and bottom-up influences upon eye movement of patients with different types of dementia using eye-tracking has been conducted by Shakespeare et al. [155]. In this work, 7 PCA and 8 tAD patients undertook different search and non-search tasks when looking at images. Results suggest an increased tendency of individuals with PCA to fixate at salient locations compared to controls, however, we need to consider that the experiment just evaluated a narrow topic within visual search (vegetation within scenes). For further consideration of the role of visual saliency influencing gaze position and a novel investigation applied in a real-world setting during navigation of PCA and tAD patients, see section 2.3 and Chapter 3.

2.3 Computational Attention Modelling

One perspective in eye-tracking research has been the analysis of eye movement to estimate the focus of visual attention, or in other words, what attracts human attention. Some approaches suggest the analysis of eye-tracking data using computational visual saliency models applied on static images and egocentric videos [11]. In dementia research, limited studies have investigated the role of saliency and eye movements, and to the best of our knowledge there is no research using saliency maps on egocentric videos [68, 69, 155]. In this section, we will provide the reader with some definitions about visual attention and computational visual attention models.

2.3.0.1 Visual Attention and Eye Movement

Visual attention is directed using two information processing mechanisms. Bottom-up selection is a fast and stimulus-driven mechanism which involves shifting attention to conspicuous features based on colour, intensity, orientation and motion (e.g. red item against a field of green or the sudden movement that could be a predator) [35]. The other mechanism, top-down attentional selection, is slower and is guided by the observer's expectations, emotions and intentions such as biasing attention toward restaurants when we are hungry.

Visual attention is closely associated with eye movements which are considered as "a proxy for attention" since they constitute a way to get information about it. When we look at a scene or search for an object, we make saccades during which we do not obtain any information and fixations which determine the parts of the visual field which are consciously examined using higher order cortical brain functions [143]. Although attention is not always directed to the gaze location [130], eye movements are driven by both bottom-up and top-down attention [167, 176]. In other words, the location of fixations within a visual scene is not random; it is determined by low-level properties of the scene and high-level knowledge related to scene structures or items and task demands [155].

2.3.1 Computational Visual Saliency Models

Computational models have used to measure the likelihood of a location in the visual field to attract the attention of human observers. Saliency models "predict the probability distribution of the location of the eye fixations over the image, i.e. saliency map" [83]. Thus, given an image, a saliency map represents the extend to which the image regions are distinguiable from each other and the order in which the nervous system process them [18]. Each pixel of the image is represented by a scalar value that demonstrates its saliency.

Various models exist which can be grouped into those that model bottom-up attention and those that try to predict human fixations. It is widely accepted that bottom-up models are inadequate for modelling visual attention because of their lack of semantics features [83]. Object recognition methods have been incorporated in saliency models to include top-down attention and they improve prediction accuracy [28]. Recently models have been proposed that can automatically learn features for saliency prediction using deep neural networks pre-trained for object recognition in datasets including eye movements of people looking at images [83].

In this thesis, since there is previous evidence that individuals with cognitive impairments rely on bottom-up saliency with inefficient visual search predominantly attributed to posterior parietal damage, we are interested in how low-level visual properties guide attention and thus we focus on bottom-up models [155, 153]. The seminal works in this area have been conducted by Koch and Ullman [97] and Itti et al [86] and are based on the feature integration theory [167]. The second paper is one the most widely used for comparison purposes. In this model the visual features of an image are computed within three channels based on intensity, orientation and colour

using linear filtering at several spatial scales and calculating centre-background differences. The feature maps are combined into a single "conspicuity map" for each channel and then they are summed into one saliency map. In addition, an inhibition-of-return mechanism applies such that attention is not stuck in the most salient image location but rather is shifted to the next most salient point. An example of this model applied to three images of bars is displayed in Figure 2.3.

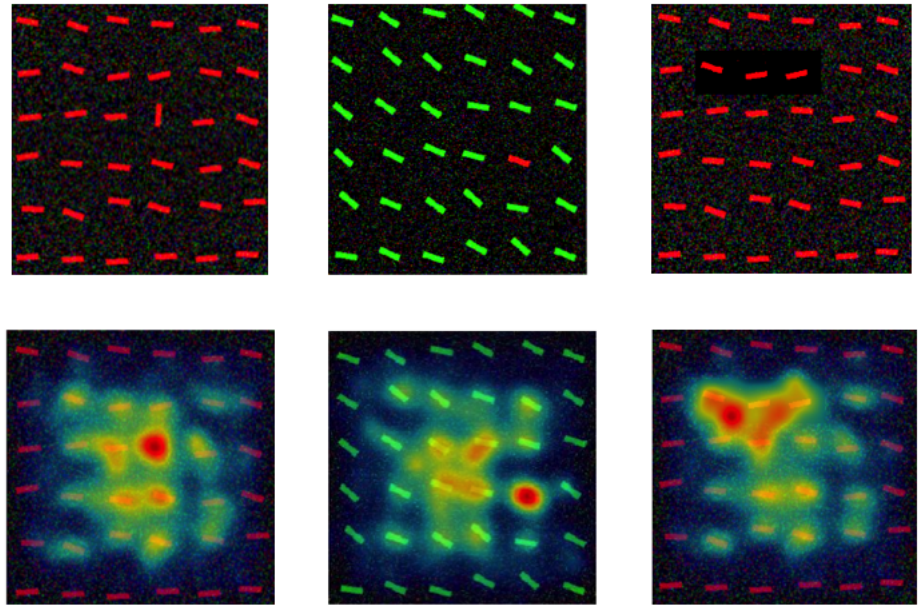

Figure 2.3: Itti et al [86] saliency maps overlapped with the original images that have regions that immediately pop-out based on orientation, colour and luminance contrast, from left to the right respectively. Regions coloured with orange are the most salient areas in the images based on the model.

2.3.1.1 Graph-Based Visual Saliency

Different bottom-up models exist (e.g. [20], [82]) but here we choose to use the Graph-Based Visual Saliency model [79] due to its improved prediction over other algorithms such as Itti et al [86] and its publicly available toolbox [79]. In addition, it was used by Shakespeare et al. [155] for the evaluation of low-level influences on scene perception in people with PCA and tAD and thus enables extension of these previous investigations involving computerised scenes to real world, naturalistic investigations.

The algorithm consists of three stages: feature maps are extracted from the image from which activation maps are generated that are normalised and finally combined to a single saliency map. In more detail, informative locations are extracted from the image forming feature maps via linear and nonlinear filters in order to convert pixels to regions. Orientation (e.g. based on Gabor filter), contrast based on luminance variance in a local neighbourhoods and luminance maps are examples typically used. Then for each feature map an activation map is computed in which the elements have

high values if they are considered unusual for its neighbourhood. In order to find "unusual" elements a dissimilarity metric is defined. Given a feature map M with (i, j) and (p, q) locations in M, the dissimilarity of these two elements is defined as

$$d((i,j) \parallel (p,q)) \equiv |\log \frac{M(i,j)}{M(p,q)}|.$$

The map is converted to a fully-connected graph with the edge weights of nodes (i, j) to (p, q) proportional to the dissimilarity of the two ends, as well as, to their relative position on the map:

$$w_1((i,j),(p,q)) \equiv d((i,j) \parallel (p,q)) \cdot F(i-p, j-p), \text{ where}$$
$$F(a,b) \equiv \exp -\frac{a^2+b^2}{2\sigma^2}.$$

Then a Markov chain is created on the previous graph. A Markov chain is a stochastic model which describes the transitions (changes) of a system from one state to another based on some probabilistic rules (transition probabilities). In the previous graph, the nodes are treated as states and the weights as transition probabilities normalising the weights to 1. Since transitions are more likely to nodes with higher weights, more mass will be concentrated at nodes dissimilar with their surrounding nodes. This results in an activation map which is normalised following a similar process and it ultimately creates the output saliency map.

2.4 Eye-tracking using Machine Learning

Recent eye-tracking applications have given rise to a number of studies incorporating eye-tracking technology with machine learning algorithms (see section 2.9) that demonstrate the potential of revealing information about human cognition, attention and learning using eye movements and computational approaches. The majority of research studies in this area are concerned with two phases; feature extraction and classification.

Eye movement feature extraction has been extensively studied and it concerns methods to summarise the large amount of spatiotemporal gaze recordings to understand cognitive processing [143]. Currently, the most commonly used way of summarising eye-tracking information is the computation of statistics over the pupil dilation and the gaze signal. The latter is converted to a sequential series of events predominantly consisting of fixations (eyes held stable) and saccades (rapid movements to change the position of fixations). Because of the high level of variability between and within-individuals, a single measure such as mean fixation duration per spatial unit (e.g. a word in a text) is not able alone to capture adequate characteristics of cognitive processes during complex tasks [143]. Thus, a set of features are calculated on carefully selected spatial areas of interest. Other methods for eye movement feature extraction include statistics and heatmaps over raw gaze data, similarity indices of scanpaths, as well as, the so-called n-grams features that encode information for

the direction and the amplitude of eye movements [81, 23]. In dementia research, the previous features take the form of abnormalities and are expressed in terms of latency, accuracy, stability and variability [132].

Regarding classification techniques, the support vector machine (SVM) classifier (see section 2.9) is the most commonly used method applied to a predefined set of features extracted from the raw eye-tracking data. It has demonstrated potential in a variety of eye-tracking applications including detection of cognitive and mood imbalances, learning disorders and performance prediction, among others. For instance, Lagun et al. [101] implemented a SVM model with eye-tracking metrics including novelty preference, fixation duration, refixations, saccade orientation, and pupillary diameter, which discriminates between controls and patients with mild cognitive impairment during a visual paired comparison task with 87% of accuracy, 97% of sensitivity, and 77% of specificity. Rello and Ballestero [144] utilised a SVM binary classifier for dyslexia detection during reading of texts with different font sizes using a sample size of 97 subjects (48 of which were diagnosed with dyslexia). Their model reaches 80.18% accuracy with features including a combination of text characteristics, eye-tracking metrics and subjects age. The feasibility of SVM in discriminating between low and high performance during problem solving from eye-tracking metrics is demonstrated in [54], where 87.5% accuracy is yielded from data of 14 participants completing puzzle games in a computer screen. Other studies that highlight the effectiveness of combining eye-tracking with SVM classifiers include detecting readers with low literacy skills [108], classification of the age of toddlers [46] and measuring learning attention in elearning [106].

The Random Forest (RF) algorithm has also attracted attention as it uses an ensemble of models (decision trees) with bootstrapping to make predictions [104]. In a comparative study of SVM and RF classifiers on detecting task load during complex mathematical problem solving, the latter produced higher accuracy (69.6% vs 56%)(using data from 48 participants on 10 tasks). Furthermore, a recent study on detection of personality traits by Berkovsky et al. [14] (while 21 subjects viewed image and video stimuli for 55 minutes), showed evidence for the superiority of naive Bayes classifier [146] (85.71% accuracy) over the state-of-art machine learning methods (including SVM and RF). The naive Bayes classifier is a probabilistic classifier with strong assumptions about the independence of input features.

The recent success of deep learning in a variety of applications (see section 2.9) has also encouraged researchers to investigate its effectiveness in eye-tracking data. Sims and Conati [157] developed an AI system for detecting user confusion episodes using data from 136 participants performing 5440 tasks while interacting with a visualisation tool for supporting decision making. They introduced a novel architecture combining Recurrent Neural Networks (RNN) [118] (the Gated Recurrent Units (GRU) variant of RNN trained on raw eye-tracking samples) and Convolutional Neural Networks (CNN) (see section 2.9.0.4) (trained on grayscale images of saccadic scanpaths) achieving an AUC score of 0.84 on confusion detection (22% improvement compared to a baseline RF model and 4% improvement compared to both baseline CNN and GRU networks when trained seperately). The strength of their approach is that it takes advantage of both temporal and visuospatial features of eye-tracking data.

Another study on identification of Autism Spectrum Disorder (ASD) using eye-tracking by Xia et al. [179] leveraged both CNN and visual saliency features to encode saccadic scanpaths of 74 children with and without ASD. Their experiment consisted of 82 images presented on a computer screen for 3 seconds each without particular instructions except to observe. By extracting features from each fixation using a pre-trained CNN on patches around the fixation location on the stimulus image, along with low and high level salient features and feeding them to an SVM classifier, they achieved a maximum classification accuracy of 94.28% in the diagnostic tests.

Apart from these studies that combined different aspects of eye-tracking data, other methods either used saliency-based approaches [46, 89] or focused on ways to encode scanpaths. In the latter case, eye movement scanpaths are either converted into images which are the input to 2D CNNs for classification tasks (e.g. classification of web user interfaces, nationalities of users, types of information presentation, relevance prediction [181, 180, 16]) or time series of x and y position of gaze during fixations using 1D CNN networks (e.g. age prediction from gaze [188]).

Another line of eye-tracking research using deep-learning is end-to-end classification for automatic detection of eye movement events (e.g. fixations, saccades). Zemblys et al. [187] implemented a RNN architecture using velocity of x and y coordinates of gaze and reported a classification performance equivalent of expert human coders. Furthermore, for the same purpose Goltz et al. [76] compared different neural network architectures including CNNs and sequential models (e.g. RNN) using as input the time series of velocity and acceleration of gaze. The authors found that small convolutional neural networks outperforms more complex architectures for eye movement event detection. Startsev et al. [162] implemented a sophisticated architecture capable of predicting smooth pursuits apart from fixations and saccades. The authors included not only unprocessed gaze coordinates but also the speed of gaze, its direction, and acceleration at different temporal scales to capture larger movement patterns. The network architecture consisted of a combination of one-dimensional convolutional neural network (1D-CNN) and bidirectional long short-term memory block (BLSTM) that outperformed state-of-the-art smooth pursuit detectors.

Lastly, another application combining eye-tracking and machine learning that reached the attention of researchers is the automatic prediction of gaze location from images captured by a camera (without any sophisticated eye-tracking hardware system). Krafka et al. [99] aiming to develop a software that works in mobile phones and tablets implemented a deep learning framework with convolutional layers that given the image of the face together with its location in the image and the image of the eyes accurately predicts the location of gaze.

To conclude, based on the reviewed literature it becomes clear that machine learning techniques have made a significant contribution to the advancement of eye-tracking research. Neural network architectures show particularly promising results on extracting features from eye-tracking data compared to less complex methods such as random forests and support vector machines. Although the most suitable architecture might be dependent on the task of interest, convolutional and variations of recurrent neural networks and their combinations demonstrate high performance in a variety of problems. Reviewing the studies above, a current trend becomes apparent towards

fusing features of the stimuli itself (e.g. saliency features) along with spatiotemporal features of the raw eye-tracking data (or variations e.g. velocity) and encoded versions of the reduced signal of scanpaths. Nevertheless, all the previous approaches focus on classification tasks. In this thesis, we first processed eye-tracking data extracting saliency measures of egocentric videos in Chapter 3 and then following the emerging research on deep learning, we implemented a variant of the existing CNN architectures with inputs including raw eye-tracking data of x and y coordinates of gaze and additionally pupil size. The innovation of our work is the application of these neural networks on a transfer learning setting in Chapter 4 and in an unsupervised learning setting in Chapter 5. Finally, in Chapter 4 we applied an explainable AI framework on eye-tracking data to investigate further the black box decisions of the networks.

2.5 Computational Eye-tracking Methods in Dementia Research

In dementia research, the eye-tracking metrics take the form of abnormalities and are expressed in terms of latency, accuracy, stability and variability [132]. However, the identification of a complete set of handcrafted features from cognitive tests sensitive to subtle task and participant-specific abnormalities is non-trivial and time-consuming. Additionally, these features are not generalisable to more complex stimuli because they rely on specific stimulus characteristics (e.g. regions of interest).

Most studies in dementia research use the extracted features as input to mixed-effects or generalised estimating equations statistical models to test univariate group differences and group by task condition interactions. An emerging line of work exploits machine learning methods mainly for automatic classification of groups directly from the features. Biondi et al. [17] developed a deep-learning framework for automated AD prediction based on eye movements during reading of predictable and unpredictable sentences and proverbs using features commonly extracted from eye-tracking data: mean and standard deviation of saccade amplitude, fixation duration and duration of the fixation on a single word for each sentence, counts of fixation, first fixations, re-fixations and single fixations on each sentence. Their autoencoder network predicts the probability of a trial belonging to an AD patient with 89.9% accuracy (see section 2.9 for more background on autoencoder networks).

A few other studies followed an alternative direction, developing machine learning models on raw eye-tracking data hypothesising that additional information, more informative than hand-crafted features, can be extracted from the raw signal. Primativo et al. [138] implemented a Bayesian model for dementia diagnosis which is first trained to predict gaze coordinates in controls in a spatial anticipation task. Then a dementia diagnosis prediction is made based on the magnitude of the error between the model predictions and the real values in patients (bvFTD and svPPA) compared to controls. This approach seems to achieve improved results in detecting cognitive deficits compared to state-of-art methods. Moreover, in [132] a hidden Markov model was used to extract feature vectors from a smooth pursuit task which achieved a 95% accuracy in discriminating Alzheimer's disease patients from controls. Nevertheless, it is explicitly

fitted to model data from smooth pursuit experiments (one is anticipating the location of a target over time) and possibly not generalisable in other less predictable tasks.

2.6 Modelling Eye-tracking as Time Series

A time series is a sequence of data points sampled from an underlying process over time. Time series data differ from other types of data because they are noisy, high dimensional and sometimes non-stationary. The property of non-stationarity implies that data characteristics such as mean, frequency and variance are changing over time. Time series are also highly time dependent which implies that in order to be modelled memory past inputs are required [102].

Eye-tracking data consists of a multivariate time series of x and y position of gaze and pupil size. As biosignals (signals that can be continually measured/monitored in living organisms), they are time series readings and thus can be modelled using techniques widely used in the literature for time series modelling [33].

In this section, we provide a general overview of the stages of data processing used for biosignals and sensor data. Several reviews describe the data processing of wearable sensors in health monitoring as a procedure of implementation of data mining tasks [161, 8]. Data mining is defined as the process during which algorithms are used for extracting patterns from data [60]. The common data mining architecture used is summarised in Figure 2.4 which was adapted from [8].

For both supervised (i.e., classification, regression) and unsupervised (i.e., clustering, association, summarisation) data mining tasks, the raw data are extracted from the sensors and are divided into a training set for model development and a testing set for model evaluation [8]. Data preprocessing is then implemented to maximise the signal-to-noise ratio and to remove artefacts and sensor errors. A challenging part of this step includes also data formatting, normalisation and synchronisation when numerous sensors are used for data collection [161]. Subsequently, the abstraction of raw data via feature extraction follows for the discovery of characteristics which are representative of the original data and then the most discriminative features are selected leading to dimensionality reduction of the input data. The feature extraction algorithms are typically hand-engineered by experts and recently deep learning methods are being applied which automate the process. Machine learning techniques and other models can then use these robust input features (considering expert knowledge and metadata) to perform tasks such as anomaly detection, prediction and decision making. Anomaly detection is used for retrieval of abnormal patterns in physiological data, prediction for the estimation of events that have not occurred yet and decision making for the evaluation of various patterns of the data which are meaningful for decisions.

Given the general framework for data mining, in the following sections, we provide more details about the feature extraction and the anomaly detection/prediction steps.

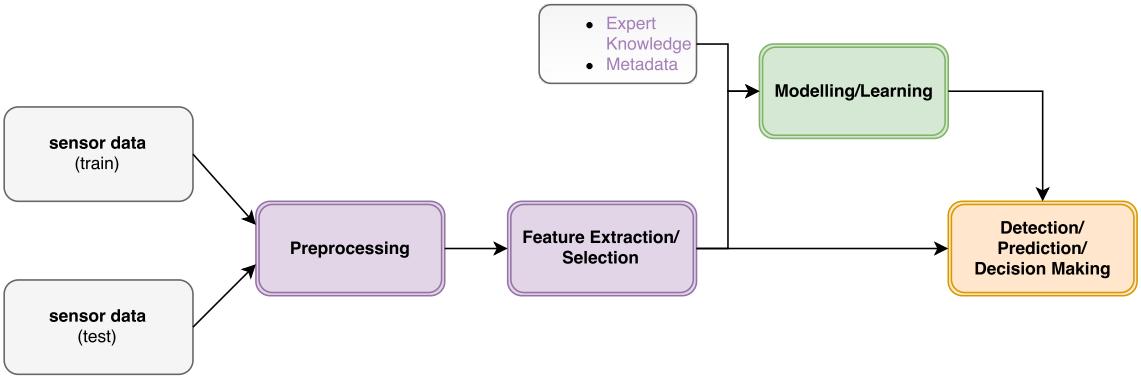

Figure 2.4: Generic architecture of the data mining approach for sensory data.

2.7 Feature extraction

The process of mining biosignals involves converting patterns in the data to features which constitute condensed representations preserving only salient information [34]. The set of features which is a reduced-dimensional representation of the observed patterns in the data is called feature vector. This dimensionality reduction reduces the computational cost of data processing and speeds up the model development [163]. The accuracy and generalisation of modelling methods such as classification also increase, as the input feature vector is in a form that is amenable to learning and it prevents overfitting [34, 52].

2.7.1 Hand-engineering Features

Hand-engineering of features is an approach that leverages ingenuity and expert knowledge to select a set of features which are appropriate and boost the perfomance of predictive models. The new features set is not generated by a machine but it is computed manually by human intervention [163]. Hand-engineering of temporal data often involves extracting features from the time and frequency domain.

The analysis in the time domain takes advantage of the temporal characteristic of the data and the observable trends in the signal. The extracted features include statistical parameters such as mean, median, variance and basic waveform characters such as peak counts and duration. The analysis in the frequency domain takes into account the periodic behaviour of time series. In this case, the frequency is the ordering dimension instead of time. Fourier and Wavelet transforms are frequently used tools for frequency domain feature extraction. The Fourier transform is based on the logic that any time series can be expressed as a number of sinusoidal waves that extend with equal amplitude to the range of the entire the time series. Common features derived from this transformation are spectral energy, power spectral density, low-pass filter and high/low frequency [8]. Wavelet transform is a time-frequency transformation which addresses the problem of non-stationarity by representing time series as fixed blocks called wavelets. More complex feature extractors inspired by signal processing have been also applied such as Legendre and Krawtchouk polynomials, approximate

entropy and parameters of regression models [114].

Although hand-engineering feature extraction techniques are beneficial for the data processing stages of complex time series, they present critical limitations and thus their use is cumbersome [114]. Firstly, the attribute selection depends solely on the creativity and knowledge of the expert. Secondly, these hand-crafted techniques make necessary the implementation of the feature selection stage which is computationally expensive, time-consuming with no guarantee of converging to an optimal features set. Therefore, automatic feature extraction approaches could help alleviate the limitations of hand-engineering techniques.

2.7.2 Feature Learning

Representation learning or feature learning is a set of techniques in machine learning that automatically discover salient features of the input data. A representation is considered to be good when it benefits supervised predictors when used as input. Representation learning is an important component of machine learning pipelines as the performance of any machine learning model is highly dependent on the input data. Traditionally, feature extraction methods have been developed that generate data representations by applying linear as well as nonlinear transformations to the input variables.

Linear methods Principal Component Analysis (PCA) is one of the most common linear feature extraction techniques that is used for transformation of the feature space. It transforms the original possibly correlated variables to an orthogonal set of Principal Components (PCs) which are linearly uncorrelated variables accounting for as much variability in the data as possible [163]. Linear Discriminant Analysis (LDA) is used for dimensionality reduction and PCA alike, the obtained features of LDA are linear combinations of the original data and not of constructed variables like principal components.

Nonlinear methods Some nonlinear feature extraction methods include kernel PCA, Restricted Boltzmann Machines (RBM) and manifold learning methods [31]. Kernel PCA is a non-linear extension of PCA which maps the original data vector into a feature space using a kernel function (see 2.9.0.2) and then linear PCA is performed. RBMs are undirected graphical models that consist of a two-layer neural network with one visible layer and one hidden layer; the outputs of the latter are the set of extracted features. The manifold learning methods attempt to provide a mapping from the high-dimensional space of the original features to a low-dimensional embedding. Multidimensional scaling, locally linear embedding and laplacian eigenmaps are techniques that fall under manifold learning. To conclude, although the above techniques constitute automatic techniques for feature extraction, they are usually applied to a set of features extracted a priori [114].

2.7.3 Deep Representation learning

There are various ways of learning representations, but here we focus on deep learning methods which are formed by combining multiple linear and non-linear transformations to the input aiming to learn abstract and ultimately more useful representations of the data. Deep learning is a subcategory of machine learning based on artificial neural networks (see section 2.9.0.3 for more details).

Representation learning [13] with neural networks can be either supervised or unsupervised:

- In supervised learning, given a dataset of n observations $X = \{x_i | i = 1, ..., n\}$, there is a set of output values $Y = \{y_i | i = 1, ..., n\}$ associated with the examples in X. The learning algorithm seeks to find a function $f(x; \theta)$, characterised by a set of parameters θ, that approximates Y as accurately as possible. The key idea is to not only learn the mapping between the inputs and outputs, but also the underlying structure of the data.
- In unsupervised learning, no labels are associated with the observations and an unsupervised learning algorithm seeks to discover features in a lower dimensional space than the original high-dimensional input that still capture some underlying patterns of the data.

One of the main challenges in machine learning is that we often have very large amounts of unlabelled training data and little labeled data. Therefore, supervised learning techniques trained on the labeled set often results in overfitting [77]. To address this problem, pretraining is a commonly used in deep learning to learn representations in a supervised way aiming to alleviate overfitting. There are two main pretraining methods: transfer learning (supervised) and self-supervised learning (unsupervised). The rationale behind both is that by training a network to solve a pretext task, it encodes high-level semantic representations that are useful for solving other tasks of interest that usually have little annotated data.

Transfer learning [129] is one of the most popular pretraning methods, which learns the parameters of a representation network by solving a supervised problem (source task) on a large-sized external dataset and then finetunes the parameters on the target data (available labeled examples) (Figure 2.5). This approach takes advantage of the abundance of source data and enables the network to learn powerful representations of the target task with resilience to overfitting. However, one disadvantage of transfer learning is that the representations might be biased to the source labels and not generalise well on the target task where the classes are different.

Apart from supervised learning techniques, unsupervised methods have also been used for feature learning since they need no labels to learn representations of the input. Various deep neural network structures have been explored such as restricted Boltzmann machines, deep belief networks and deep autoencoders; with the latter being the most flexible unsupervised neural networks. Autoencoders create a bottleneck or apply a restriction in the learnt representations and their training objective is the reconstruction of the original input from these representations [77]. They have

been used in several applications from dimensionality reduction for feature visualisation and denoising images, to detecting abnormal patterns in sequential data [31].

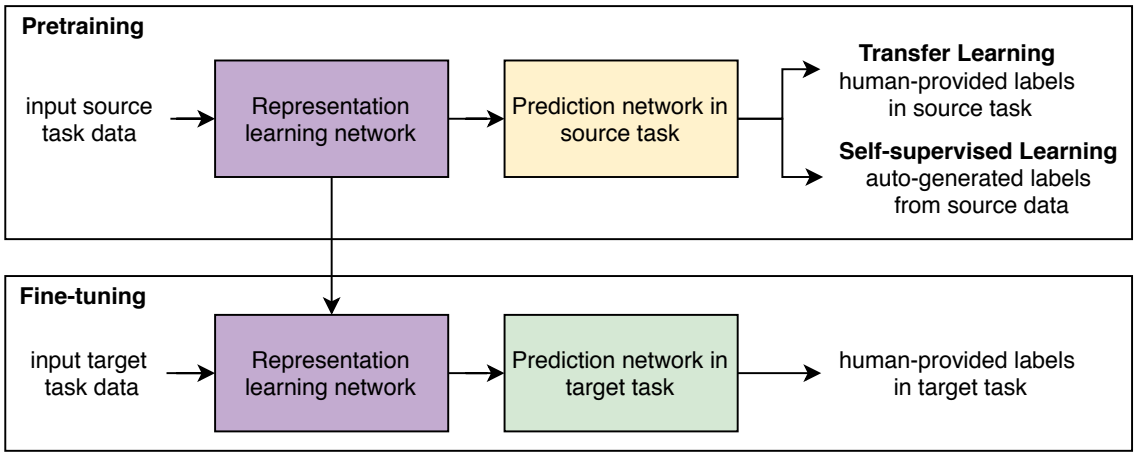

Figure 2.5: Workflow for transfer learning and self supervised learning.

A relatively new and promising subclass of unsupervised representation learning which has produced state-of-the-art visual representations in standard computer vision problems is self-supervised representation learning [125]. It attempts to alleviate the problems of previous approaches by unsupervised pretraining. This involves training the representation network in the source task only with the input data, as no labels are required. The output labels are constructed directly from the data, compared to transfer learning that human annotations are needed (Figure 2.5). Self-supervised learning models are not prone to be biased to the labels in the source task and this property might make them more generalisable. Therefore, this method uses information already present in the data as a supervision signal so that supervised learning techniques can be used.

The pretext tasks can be constructed using different mechanisms such as rotation prediction (the network recognises the geometric transformation applied to an image; [74]), inpainting (the network is trained to generate the contents of an arbitrary blank image region conditioned on its surroundings; [131]) and automatic colorisation (given a grayscale image, the network predicts a plausible colour version of the image; [189]).

2.8 Classification vs Anomaly Detection

Having discussed various methods for extracting useful information from a signal, in this section, we focus on the last stage of the data mining pipeline which concerns predictions. Predictions are closely linked with the existence or lack of labels. In standard neuropsychological practices we encounter two types of labels associated with a performance in a cognitive task: scores (defining the level of performance associated with impairment in a given scale) or a binary labels (impaired/normal). These two categories can be translated into two machine learning problems, namely, anomaly

detection and classification. The former deals with models that predict deviations from a normal distribution assigning anomaly scores to observations and the latter with assigning observations to specific classes. A comparison of classification and anomaly detection techniques are discussed in more detail here.

Table 2.1: Comparison of anomaly detection with classification in terms of properties of the dataset (Dataset), labels of the available classes (Class) and Category of representation learning (Category).

Method	Category	Dataset	Class
Anomaly Detection	Supervised Semi-supervised Unsupervised	Imbalanced	Binary
Classification	Supervised	Balanced Imbalanced	Binary Multi-class

Classification is a type of supervised learning which is used to classify observations into two or more classes. Its targets are always categorical and expected to be balanced so that all classes have almost equal importance.

Anomaly detection refers to the problem of discovering nonconforming patterns in the data such as anomalies, outliers and peculiarities. Anomalies are patterns in the data that deviate from normal behaviour [1]. There are different types of anomalies and can be classified into three categories [30]:

- Point anomalies refers to the case that an individual observation is anomalous with reference to the rest of the data.
- Contextual anomalies concern those data instances that are considered abnormal only in within a specific context [160].
- Collective anomalies are observed when a collection of related data instances is anomalous relative the entire data, although the individual data instances are not anomalies.

In anomaly detection problems, the labels associated with each data point denote whether that instance is normal or anomalous. Typically, getting labels for normal behaviour is easier than getting a labelled set of all possible anomalous behaviour. Therefore, based on the availability of the labels, anomaly detection techniques are grouped in three different types [30]:

- Supervised Anomaly Detection: Techniques trained in a supervised mode assuming availability of labels for both anomalous and normal observations. A

typical approach is to build a classification model for prediction of normal vs anomalous classes. However, there are major issues using this approach because firstly the anomalous instances are fewer compared to the normal ones in the training set (imbalanced classes) and secondly, obtaining accurate labels for the anomaly class is challenging.

- Semi-supervised Anomaly Detection: Techniques assuming that the training data has labels only for the normal class.
- Unsupervised Anomaly Detection: Techniques that do not require labels for training assuming that normal observations are far more than anomalous ones. Therefore, the model is assumed to be robust during training to a few anomalies.

The output of anomaly detection techniques are either scores or labels. Methods that produce scores, assign an anomaly score to each data instance in the test set analogous to the extend that the instance is considered anomalous. Therefore a list of ranked anomalies is provided by the algorithm and then either the top few anomalies are analysed or a cutoff threshold is used to select anomalies. On the other hand, methods that provide labels, assign a binary label (normal or anomalous) to each test instance.

Various models have been explored for anomaly detection, however in the last years Deep Anomaly Detection (DAD) has gained a lot of attention. DAD techniques solve the problem end-to-end; taking raw input data, learning hierarchical discriminative features and then providing anomaly scores. Generative models, autoencoders, sequence models, convolutional neural networks, word2vec models are some architectures that have been used successfully for anomaly detection [29]. A fundamental method for anomaly detection is using deep autoencoders. When these models are trained solely on normal data, they are not able to reconstruct previously not seen anomalous data. There samples that produce large reconstruction error are those predicted as outliers.

To conclude, anomaly detection and classification methods are two distinct machine learning problems. The key factors for differentiating them depend on the labeled classes and the imbalance of the dataset. In general, in binary problems where there are very little positive data instances and large number of negative, anomaly detection is recommended. If there are enough positive examples, then classification is preferred; whereas if they are many types of anomalies, anomaly detection might be advantageous.

2.9 Overview of Machine Learning Models

Here we briefly describe the machine learning methods used in this thesis. Firstly, we provide a glossary with common machine learning terms to provide a basis for non-technical readers. Then, more details for the support vector machine classifier and artificial neural networks are given for the interested readers. A definition of the support vector machine classifier is given which is used in Chapter 4 as the default method for classifying whether one has a dementia given various eye-tracking metrics. This is followed by a introduction to some basic neural network architectures such single-layer and multi-layer perceptrons which give the basis for the description of convolutional neural networks. Convolutional neural networks are used in both Chapters 4 and 5 to automatically extract features from the dense and dynamically changing eye movement time series. Lastly, the architecture of autoencoders is described which is the method used in Chapter 5 for detecting trial-level eye movement anomalies by defining the distribution of healthy eye-tracking data.

2.9.0.1 Machine Learning Glossary for Neuropsychologists

Terminology	Definition
Machine learning	Mathematical algorithms that have an ability to learn itself and predict future behaviour from data.
Neural Network	Mathematical algorithms inspired by the brain's architecture modelled to recognise patterns in data.
Deep learning	A branch of machine learning methods based on artificial neural networks.
Supervised learning	Training a model using a labeled dataset.
Unsupervised learning	Training a model to find patterns in unlabelled data.
Self-supervised representation learning	Training a model using information already present in the data in a supervised way without needing labels.
Convolutional Neural Network	Hierarchical models designed to process input data where a spatial or temporal relation exists (e.g. images, speech or physiological signals).
Autoencoder	A type of artificial neural networks that learns representations in an unsupervised manner, typically for dimensionality reduction.

Terminology	Definition
Latent space	A compressed representation of the data in which similar data are closer together.
Feature	Variables used as inputs to make predictions.
Feature vector	A list of features representing an input instance passed into the model.
Loss function	A method of evaluating how well an algorithms models the data.
Epoch	A full training pass over the entire dataset.
Ensemble	A merge of predictions from multiple models.
Optimiser	A specific implementation of the gradient descent algorithm which gradually adjusts the model's parameters to find the best combinations of weights.
Batch	The set of training examples used in one iteration of model training.
L1 regularisation	A regularisation technique that penalises the model's complexity.
Data augmentation	The process of using algorithms to increase the size of a collected dataset.
Learning rate	A scalar value used during model training in gradient descent.
SVM	A supervised machine learning model that maps input examples in a space so that examples of different classes are divided by a clear gap.
Kernel	A method of using a linear classifier to solve a non-linear problem applying non-linear functions.
F1-score	The weighted average of precision and recall.
AUC	An evaluation metric that considers different classification thresholds.

2.9.0.2 Support Vector Machine

Support Vector Machine [37] is an extension of Maximal Margin Classifier (MMC) and Support Vector Classifier (SVC) which are supervised learning models based on the concept of finding a hyperplane that best separates the different classes [87]. In MMC, the classification of observations is defined by a linear boundary which is a hyperplane that has the farthest minimum distance to the training observations or in other words a large margin. This approach was considered problematic for many applications, because it is very sensitive to a single observation changes (prone to overfit the training data). The SVC similarly classifies observations depending on whether they lie on the correct side of the margin for their class, allowing though some instances to be on the incorrect side of the margin or hyperplane. These observations lying directly on the margin or on the wrong side of the margin are called support vectors. The performance of this classifier depends on a parameter C which controls the tolerance of observations being on the wrong side of the margin; as C decreases, the margin narrows. A SVC can be represented as:

$$f(x) = \beta_0 + \sum_{i \in S} a_i \langle x, x_i \rangle$$

where S is the collection of indices of the support points, x is a new example, x_i is the i observation of the input data with n number of observations and the function $\langle\rangle$ represents the inner product.

The SVM classifier extends SVC by enlarging the feature space using kernels or in other words replacing the inner product form with a non-linear kernel function $K(x, x_i) = \langle \phi(x), \phi(x_i) \rangle$, where ϕ is a given function. This is equivalent to applying the function ϕ to the inputs and then learning a linear model in the new feature space with a computationally efficient way (ϕ is applied only to the support vectors). A popular kernel function is the radial basis function (RBF) that ensures that only nearby observations have an effect on the classification and it has the following form:

$$K(x, x_i) = e^{-\gamma \|x - x_i\|}$$

where γ is a positive constant that needs to be optimised.

Next, another model is described called Artificial Neural Network (ANN) that does not apply a linear model to a general function $\phi(x)$ to learn the non-linear mapping between inputs and outputs as seen in SVMs, but rather learns this function from training examples so that $y = \phi(a; \theta)^T w$ where θ, a are parameters and ϕ a hidden layer.

2.9.0.3 Artificial Neural Networks

An Artificial Neural Network (ANN) is a biologically-inspired computational model defined as a network of processing units called neurons [77]. Each neuron receives a number of inputs and calculates its output as follows:

$$y_j = f(\sum_{n=0}^{n-1} x_{ij} w_{ij} + \theta_j)$$

where $w_j = [w_{j0}, w_{j1}, \ldots, w_{jn-1}]$ are the connection weights, $x_j = [x_{j0}, x_{j1}, \ldots, x_{jn-1}]$ are the inputs, y_j is the output, θ_j is the bias term and $f(x)$ is the activation function of neuron j. The weights and bias are adjusted to yield different functions. The activation function is usually a logistic sigmoid or hyperbolic tangent, as their output ranges are bounded to the intervals $[0, 1]$ and $[-1, 1]$ respectively.

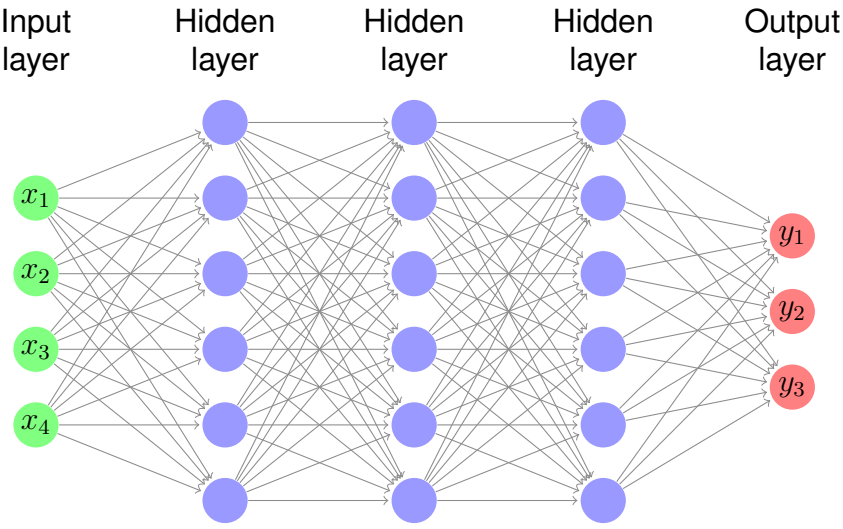

Figure 2.6: A multi-layer perceptron architecture with three hidden layers.

A single-layer perceptron (SLP) is the simplest ANN topology in which the inputs of the network are connected to neurons and the output of each neuron forms the outputs of the network. A multi-layer perceptron (MLP), a non-linear extension of SLP, is a neural network architecture where neurons are organised in stack hidden layers. The outputs of every neuron in one layer are connected only to every neuron in the next layer. In a multi-layer perceptron architecture, like the one presented in Figure 2.6, the outputs of the intermediate layers constitute abstract representation of the original input data.

The approximation of the mapping function between the inputs and outputs of the neural network is achieved using different training strategies. In this context, training refers to the process of selecting the weights, activation functions and topology of the the network. For this process, a cost function and a training algorithm is required to determine how good a given configuration is and to search the space of possible configurations respectively.

The cost function used in this thesis is the cross entropy or in other words the negative log-likelihood between the training data and the model's predictions. The objective therefore during the training phase is to minimise the dissimilarity between the data and model distribution.

For a network with C output neurons the categorical cross entropy is defined as following:

$$Loss = -\sum_{i=1}^{C} y_i log\hat{y}_i$$

where $\hat{y}_i$ is the i-th neuron's output in the model, y_i is the corresponding target value.

Given a fixed network topology and activation function, the back-propagation algorithm is used to configure the weights of the network [147]. This popular training algorithm optimises the cost function in an iterative way across a number of epochs (number of times the algorithm sees the entire data set) by adjusting the weights proportionally to the gradient of the cost function. In this thesis the Adam optimisation algorithm was used which is an adaptive learning rate algorithm in which the size of the update steps of the network weights during the optimisation process is changing as learning unfolds [94]. Typically a regulariser is also used to keep the weights low and thus avoid overfitting. Overfitting occurs when the model memorises the training data too well and it performs poorly on the test set.

Although the structure of MLPs enables the learning of complex non-linear functions, it presents certain limitations for time series data: *a.* it is computational expensive (many parameters) because of the interconnectivity of the units from different layers, *b.* same features in different time points in the input do not appear with the same representation in the output and *c.* subtle translations of the input change significantly the output [77].

2.9.0.4 Convolutional Neural Networks

To eliminate these shortcomings, another network architecture is developed which is called a Convolutional Neural Network (CNN) or Time-Delay Neural Network (TDNN) for 1-dimensional signals [77]. These networks have sparse interactions, are equivariant and invariant to translation. They have shown to be well-suited for pattern recognition in large input spaces with a spatial or temporal structure among the inputs. A common convolutional architecture for time series classification consists of a feed-forward network with convolution, pooling and fully connected layers (dense layers).

Convolutional layers contain a set of neurons that identify local patterns on a time window of the input time series (Figure 2.7). In contrast to MLP, in which the hidden units are fully-connected to the inputs, each neuron of a convolutional layer has trainable weights equal to the number of its inputs (equal to the patch size, called receptive field) and a trainable bias term. The output of each neuron is an activation function applied to the weighted sum of its inputs. This filter is applied to every patch of the signal (with same shared weights) or in other words, the weights of each neu-

ron convolve over the input signal assembling a feature map [127]. The number of feature maps, which are the outputs of the convolutional layers, are equal to the number of neurons in each layer. This operation makes feasible the removal of outliers, the filtering of data and the detection of patterns regardless of their location. Subsequently, the pooling technique aggregates features of the same feature map using average or maximum operations so that the features are presented in lower resolution. Lastly, the fully-connected layer establishes a weighted sum of all the outputs from the previous layer and it determines the output of the network [58]. Apart from the fully-convolutional layers described above, dilated convolutional layers have been applied for time series modelling (usually in cases where the input and output size of the model is equal) which enable an exponentially large receptive field by applying the convolution operation to samples d steps apart instead of consecutive samples of the input [7]. Expanding the receptive field of a convolutional network enables the network to accumulate information from very far into the past to make predictions [7].

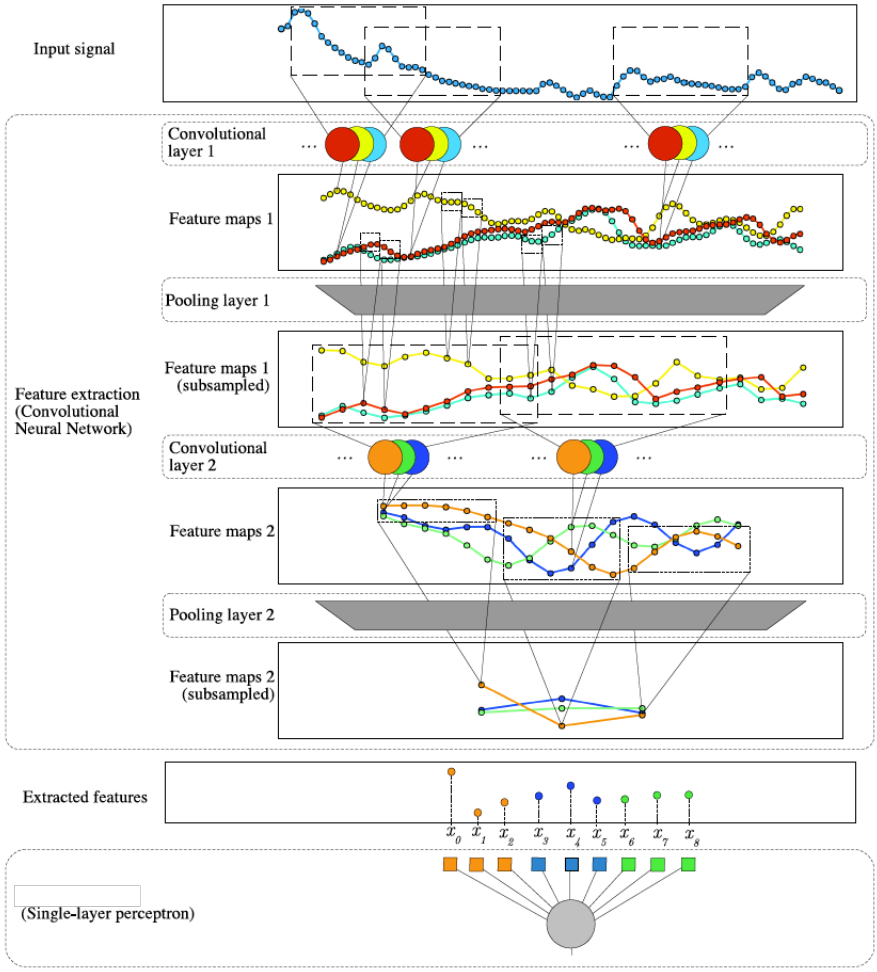

Figure 2.7: Example of structure of a deep CNN architecture modified from [114] (with copyright permission) which contains two convolutional blocks and a single-layer perceptron (SLP). The first convolutional layer has 3 neurons with patch size of 20 samples and an average pooling layer with window length of 3 samples. The second convolutional layer processes the features maps of the previous convolutional block with 3 neurons and patch length of 11, as well as an average-pooling layer of 6 samples. The final feature map of length 9 are the input of SLP.

2.9.0.5 Autoencoder

The autoencoder is a unsupervised neural network trained using unlabelled data that encodes these inputs in a small feature space and subsequently reconstructs them as precisely as possible (Figure 2.8) [77]. The encoder part of the network projects the original data into the feature space and the decoder performs the inverse operation. The training objective of the network is to recover as much information as possible from the reduced representation minimising the distance between the inputs and the outputs (reconstruction error). Restrictions can be imposed to produce interesting representations with models such as the sparse or denoising autoencoders. In the former, adding a sparsity penalty term in the loss function obtains representations with very few activated neurons. The latter adopts the loss function to minimise the error between the reconstruction and a corrupted noisy copy of the input attempting to repair corrupted data. Convolutional autoencoders are popular networks for non-static data that combine the bottle-tie architecture of autoencoders with the properties of the convolution operation.

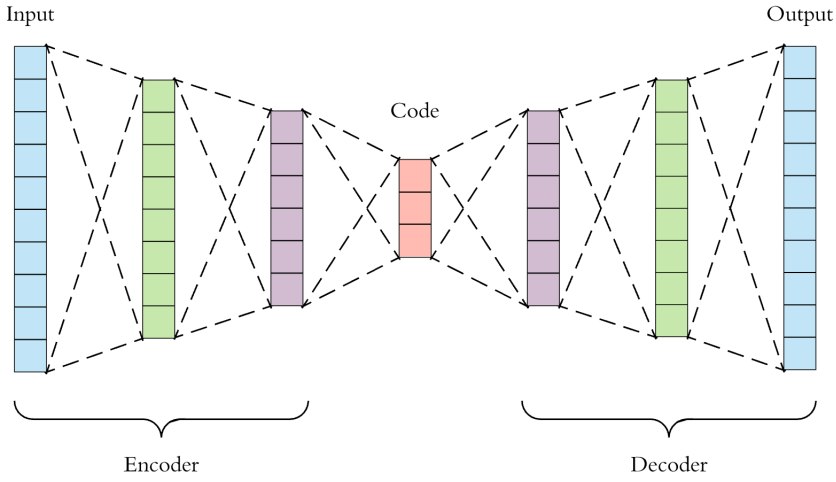

Figure 2.8: The bottleneck architecture of a basic autoencoder: an encoder and a decoder are linked by the encoding layer. Source: https://towardsdatascience.com/applied-deep-learning-part-3autoencoders-1c083af4d798.

2.10 Critical Assessment

To date, various eye-tracking approaches have been developed to study the oculomotor profile of people with dementia. However, they remain an emerging area of research, as those studies that have attempted to investigate eye-tracking as an outcome measure in cognitive tasks (deviating from the norm of identifying abnormalities in basic oculomotor functions) are very few and share a number of limitations. One strong assumption in these studies (e.g. [44, 145, 62, 138]) is that all dementia patients understand the instructions of the tests. However, mistakes caused by misunderstandings, language difficulties or patients at the later stages of the disease are known to mask disease signature. Moreover, previous studies do not consider that attentional dysfunction present in various types of dementia might interfere with oculomotor control and patients cooperation. Although there have been some attempts addressing this limitation with evaluation of oculomotor behaviour during naturalistic activities (e.g. [164]), they are restricted to specific dementia subtypes (AD), and are characterised by small sample sizes and inadequate quantification of the measurements. Apart from the limitations related to the methods used to acquire accurate oculomotor information, the analysis of the complex eye-tracking data is still a challenge for researchers. Some studies (e.g. [17]) rely on known biomarkers restricting in this way the potential of eye movement time series containing far richer relevant information. Some others (e.g. [132]) have attempted to find the discriminative boundary between the classification classes assuming a homogenous pattern of abnormality in the patient group (e.g. similar degrees of disease severity). Although, recent work by Primativo et al. [138] introduces the concept of normative eye-tracking data by defining abnormal behaviour based on deviations from a normative reference, it has results that are less interpretable compared to standard eye-tracking measures analyses. To conclude, a variety of eye-tracking batteries and data analytics have been investigated so far. However, augmenting cognitive assessment with novel outcome measures based on oculomotor metrics is still an emerging area of research, and will require further refinement and validation before it is translated into a useful clinical tool.

Chapter 3

Investigating the effects of visual environment on navigation in Alzheimer's disease and Posterior Cortical Atrophy

3.1 Introduction

Core characteristics of different forms of dementia create particular challenges to people's functional autonomy. Visual processing difficulties are under-recognised consequences of Alzheimer's disease (AD), the most common form of dementia. Such difficulties, combined with other cognitive deficits (planning, memory), affect people's perception and representation of the environment and underlie patients getting lost in both familiar and unfamiliar environments [32]. Previous studies and design guidelines suggest that the physical environment may play a major role in mitigating dementia's functional impairment [113]. However, limited quantitative investigation of effects of environmental conditions featuring well-characterised patients has prompted strong recommendations for further empirical research [64, 177]

Individuals with dementia have been proposed to rely more on salient visual landmarks that are prominent or conspicuous compared to other features in the environment for navigation [126]. Although intricately challenging to determined, "good" landmarks are considered those spatial features that either are architecturally differentiable (e.g in color, texture, size, shape; low-level features) or semantically salient (e.g. recognisable or idiosyncratic objects; high-level factors) [26]. As visual deficits in AD include symptoms of diminished spatial mapping, restricted window of spatial attention and inefficient visual search, visual environment might modulate functional abilities in individuals with AD [165, 111, 50]. However, the evidence supporting the benefits of environment in dementia is almost exclusively based on anecdote and observation rather than formal investigation, prompting strong recommendations for empirical research to evaluate the effect of environmental factors on supporting everyday patient functional abilities.

A neurodegenerative syndrome that might offer valuable insights into the role of the visual environment on functional abilities is Posterior Cortical Atrophy (PCA) [41]). PCA is characterised by a progressive decline in visuoperceptual and visuospatial processing ([41]) and particular pathological involvement of posterior parietal and occipito-temporal cortices. While most commonly caused by AD pathology, in contrast to typical Alzheimer's disease (tAD), PCA patients demonstrate relatively preserved episodic memory at least in early stages of the disease. PCA patients exhibit a range of complex and unusual visual deficits including excessive visual crowding, restrictions in the effective visual field and eye movement abnormalities [43, 155, 183]. Notably, clinical anecdote and empirical investigation emphasise how the expression of visual deficits is modulated by low-level environmental and stimulus conditions; with individuals with PCA presenting better perception of objects presented in isolation, with reduced clutter [183], small vs large objects [182], and better localisation of moving vs static objects [42]. Moreover, eye-tracking investigations of scene perception in PCA have noted the increased influence of low-level (visual saliency of parts of scenes) rather than top down factors (adapting gaze behaviour based on task demands) on fixation patterns.

Eye-tracking studies have demonstrated that individuals with PCA when viewing a scene that they find difficulty to recognise, present eye movements initially similar to those of controls and different only on later fixations [110]. This phenomenon was interpreted as a potential impairment of PCA patients in top-down control processing which is presumably activated after the initiation of bottom-up level mechanisms during scene perception [42]. Foulsham et al. [68] conducting a case study observed that the most bottom-up salient region in the scene was more likely to be fixated by the patient than by controls. Nevertheless, a following case study presented contradicting results supporting the idea that saliency is not always a good predictor of fixation in PCA [69]. Shakespeare et al. [155] attempted to extend the single-case observations in a quantitative group study evaluating how scanning patterns are influenced by task instructions and low-level visual properties in 7 patients with PCA, 8 patients with tAD, and 19 healthy age-matched controls. Participants viewed vegetation scenes under four task conditions (encoding, recognition, search and description). Interestingly, PCA patients presented significantly less consistent scanpaths than tAD patients and controls across tasks. The findings also suggest the influence of conspicuous, visually salient features of static scenes on fixation of PCA patients relative to controls irrespective of the viewing task and indicate no differences between AD and PCA or controls.

Although previous research demonstrates that the role of visual saliency in influencing gaze in static scenes, there is limited understanding about the extent to which these findings may generalise to everyday life in complex tasks such as navigation. During navigation, the saliency of the observed visual environment is complex, as it is dynamically changing based on the view of the world from the current position. It is also determined by the relationship between the observer, the referenced spatial features and the physical environment [26]. Therefore, advanced computational methods have particular promise to investigate visual perception of dementia patients in naturalistic settings. Categorical measures based on manually defined areas of interest need to be replaced by methods that provide a continuous window on the allocation

of attention considering dynamically changing spatiotemporal information. The integration of computational approaches with ecologically valid settings has potential to change drastically the way we understand egocentric navigation (compared to studies of static head viewing static scenes or virtual reality) as the role of proprioception (the ability to sense the orientation of the body in the environment) will be taken into account [73].

In this work, we combine eye movement and egocentric video analysis to investigate patients with PCA and tAD compared to a control group performing a real-world visual search task while navigating a controlled environment. The analysis of eye movement patterns in naturalistic settings is achieved through integrating gaze locations and scene information provided by egocentric video. Here computational attention modelling techniques with saliency maps of the point of view (POV) frames used in Shakespeare et al. [155] are combined with eye-tracking metrics and gait/orientation measures to investigate the following hypotheses:

- Both patient groups will exhibit greater functional impairment relative to controls.
- Group effects of POV frame saliency/saliency at fixation are expected to be higher in PCA, followed by tAD patients and controls (PCA>tAD>Control).
- A stronger relationship between saliency measure and functional performance is anticipated in both patient groups relative to controls.

3.2 Materials and Methods

3.2.1 Dataset

Eye-tracking data from a total of 10 PCA patients, 9 AD patients and 12 healthy controls with comparable age (mean $\pm$ SD: PCA:69.1 $\pm$ 7.6; tAD: 67.7 $\pm$ 7.5; Control: 68.4 $\pm$ 5.6) and relatively mild disease severity (MMSE/30 mean $\pm$ SD: PCA: 23.7 $\pm$ 5.6; tAD: 22.6 $\pm$ 4.8) were used in this study. The data were collected from a simulated domestic environment which was set up in Pedestrian Accessibility and Movement Environment Laboratory (PAMELA) at University College London. Patient groups fulfilled clinical criteria for PCA-pure ([115, 45]) and research criteria for probable AD respectively. Ethical approval was provided by the National Research Ethics Service Committee London Queen Square. All participants provided written informed consent.

3.2.1.1 Background Neuropsychology

A standard battery of neuropsychological tests was administered to PCA and tAD patients. Overall, PCA patients exhibited poorer performance relative to tAD patients on visual processing measures (visuoperceptual: related to recognition of form, pattern, colour; and visuospatial: related to spatial relationships between multiple objects/modalities, object localisation). In particular, PCA patients presented higher

visual deficits relative to tAD patients in the Fragmented letters and Dot Counting tasks from the Visual Object and Space Perception (VOSP) battery [171]. In the former task, participants were asked to identify visually degraded letters and in the later participants were asked to count the number of black dots presented as quickly as possible (from arrays of 5-9 black dots on white background). PCA patients exhibited poorer performance overall relative to tAD patients on a measure of visual search (Letter Cancellation) [175] in which participants were requested to mark as quickly as possible with a pencil the location of 19 targets (letter As) presented among distractors (letters B-E) in a grid on an A4 sheet. Overall, performance on a measure of recognition memory was comparable between patient groups [170].

3.2.2 Stimuli and Procedure

The experimental setting consisted of a room with two doors (positioned at 400cm from the starting position), a table and an entry corridor serving as the trial starting point. For each trial, one door was opened at 46 degrees to indicate the target door, with the other door serving as the distractor. Participants began each trial with their feet at the starting position, 0.4m before a blind restricting the view of the experimental setting. Participants were repeatedly requested (36 trials) to walk to the visible open door (target) under different environmental conditions:

- Door position (left/right)
- Lighting position (=target destination, =middle, =distractor)
- Table position (obstacle/no obstacle).

Trials were administered through a repeated-measures design ensuring an equal number of trials involving each of the following variables: target position (Left, Right), lighting position (Left, Middle, Right), clutter position (Left, Right). Lighting and clutter position variables were arranged in counterbalanced variants of a Latin square design (Lighting: 6 sets of 3 trials; Clutter: 2 sets of 6 trials). Variable combinations were assigned randomly to each participant to control for order effects. Mean ground illuminance was matched between lighting conditions (40lx). The setting was designed so as not to place requirements on spatial representation beyond the range of immediate perception; with both doors being visible from the starting point, the task could, in principle, be completed using only visual information available at the start of each trial. Overall, the experimental set up was built to assess low-level environmental effects and thus it is controlled and unfamiliar to participants. Different environmental conditions were featured to require participants to adapt behaviour to each trial and vary saliency of features between each trial so that sometimes these are more (Target) or less relevant (Distractor) to completing the task.

Participants' eye movements was recorded using SMI Eye-Tracking Glasses, a head-mounted binocular mobile hardware eye-tracker, at 30 Hz. Eye movement events (saccades, fixations, blinks) were automatically parsed using SMI software Begaze 3.6. The eye-tracker's ego-centric camera with a resolution of 1280 x 960 pixels recorded the user's surroundings at 24 fps. Each fixation event was matched to

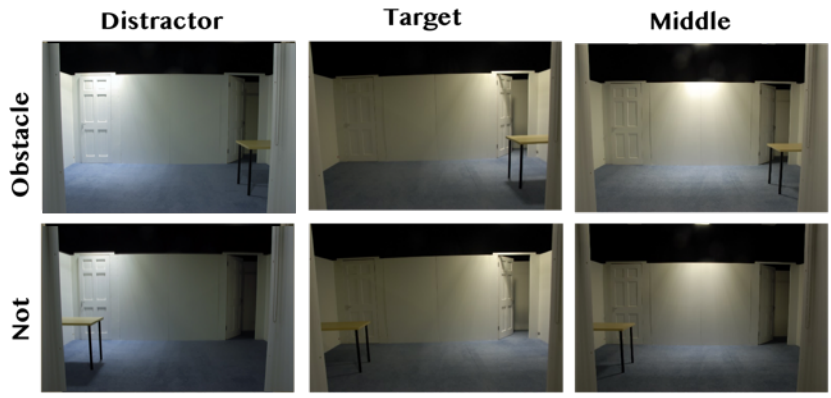

Figure 3.1: Room lighting (columns) and clutter conditions (rows).

the egocentric video's frames assuming a constant 24 Hz frame rate. The eye tracker was calibrated for each participant using 3 calibration points in the beginning of the experiment. The participants were asked to maintain fixation on a dot positioned at eye level on the blind before every trial.

3.2.3 Pre-processing and Analysis

3.2.3.1 Measures

Completion time is the primary measure of functional performance and is defined as the difference between the start of each trial and the time one reaches the target.

To quantify the effects of environmental features on visual attention, the following saliency measures of the egocentric video frames were computed:

1. The maximum normalised salience value (MaxS) of people's point of view (POV) frame during fixation periods is the difference between an area of peak visual saliency relative to the mean overall saliency of the frame.

2. The mean normalised saliency at fixation location (FixationMS) assesses the extent to which environmental features distinctive in colour or orientation predicted fixation position.

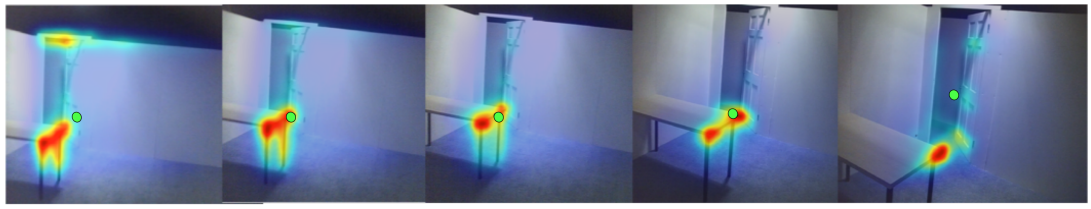

Figure 3.2: Saliency maps of consecutive frames from the perspective of a participant where red indicates salient regions and the green circle the fixation position in the course of a trial.

Observations with x, y coordinates outside of the stimulus dimensions (1280 x 960) and data points outside of the trial period were excluded. Fixations with duration less than 99 ms and more than 800 ms were also excluded from analysis based on typical ranges outlined by Munn et al. [123]. Fixation events with standard deviation of x and y coordinates more than 200 pixels were not considered reliable for the analysis and excluded. Additionally, observations which correspond to pupil diameter equal to zero in pixels or mm on either left or right eye were considered to indicate erroneous measurement and were excluded.

To calculate the saliency during fixation periods, the mapping between frames of the scene videos and the fixation coordinates provided by SMI software was used. Given that the gaze location of one fixation event corresponds to a time interval in the video, for each frame of this interval a low-level saliency map was calculated using the GBVS toolbox [79] with the default features (colour, intensity and orientation) (see section 2.3.1.1). The maps included the computed salience at each pixel of the image which ranges from 0 to 1. To obtain comparable results between different frames, the salience maps were normalised. Therefore, the normalised saliency at fixation was the mean normalised saliency of the gaze pixel across the frames during the fixation. The maximum normalised saliency value of each fixation was the average of the maximum normalised saliency across frames corresponding to each fixation time interval.

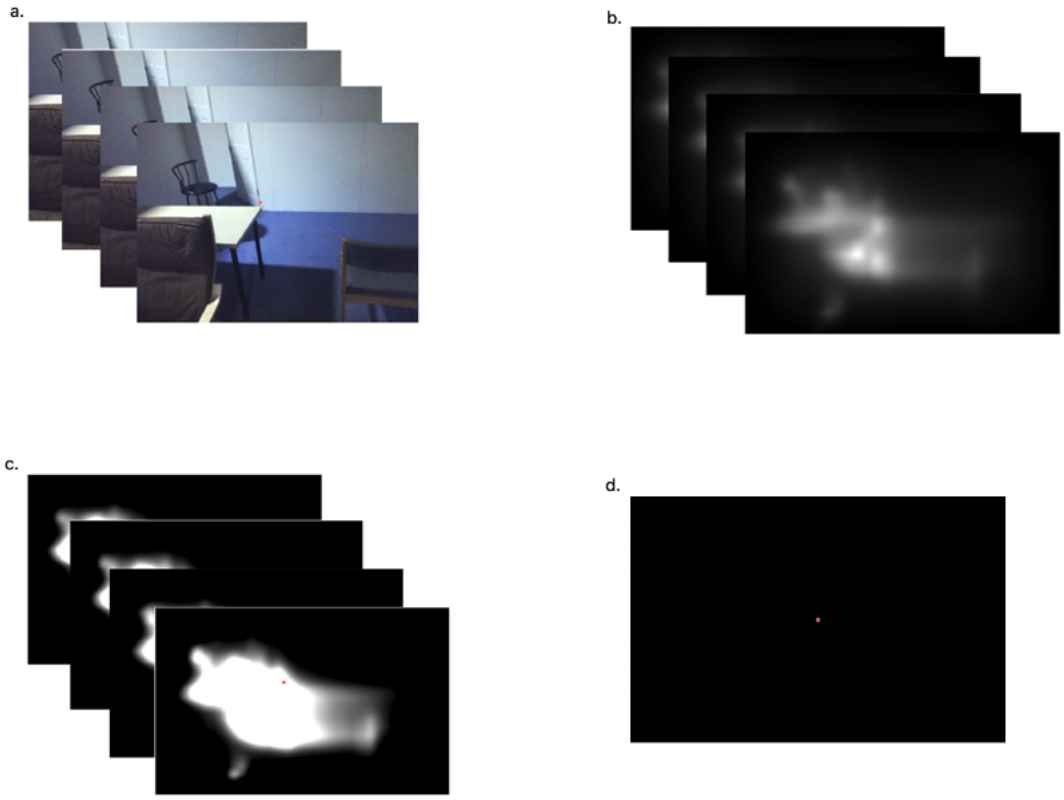

Figure 3.3: The process of computing the bottom-up saliency of a fixation event : (a) frames corresponding to fixation event overlapped with fixation location with red , (b) GBVS saliency map (the most salient areas are illustrated with lighter colour) for each frame, (c) the normalised salience maps, (d) the normalised saliency of the fixated pixel averaged over all frames (FixationMS).

The saliency measures (MaxS, FixationMS) averaged across each trial were used to test differences between groups, the effect of saliency on completion time and group by saliency interactions using Generalised Estimating Equation (GEE) model with independence correlation structure and robust standard errors to adjust for repeated measures for each subject [84]. In addition to group, GEE models included the following variables: group (PCA, tAD, Controls) and environmental condition (Lighting: L/M/R; Clutter:L/R; Door: L/R).

The same GEE model was used to test differences in completion time between groups. Tukey's multiple comparison tests were implemented to test differences between pairs of groups. All statistical analyses were conducted using R.

3.3 Results

Completion time Overall task performance was slower in both PCA (estimated mean completion time: 9.09 seconds; 95%CI [8.63, 9.55]) and tAD groups (7.10; 95% CI [6.76, 7.44]) relative to controls (5.46; 95% CI [5.30, 5.36]), and in the PCA relative to the tAD group. Between-group differences were statistically significant (all $p<0.005$).

Saliency features a) MaxS: The maximum normalised saliency of fixated frames was numerically higher in PCA (MaxS =6.65; p = 0.114) and tAD (MaxS = 6.49; p = 0.464) patients overall relative to controls (MaxS = 6.30). However, there were no statistically significant differences in maximum normalised saliency overall compared to controls or between patient groups (p=0.395).

b) FixationMS: Overall, although tAD patients fixated on less salient regions (FixationMS = 0.696) than controls (FixationMS = 0.826; p = 0.5620), this difference wasn't statistically significant. There was no evidence of a difference between PCA patients' tendency to fixate salient regions (FixationMS = 0.810) relative to controls (p = 0.9950) and AD (p =0.8090).

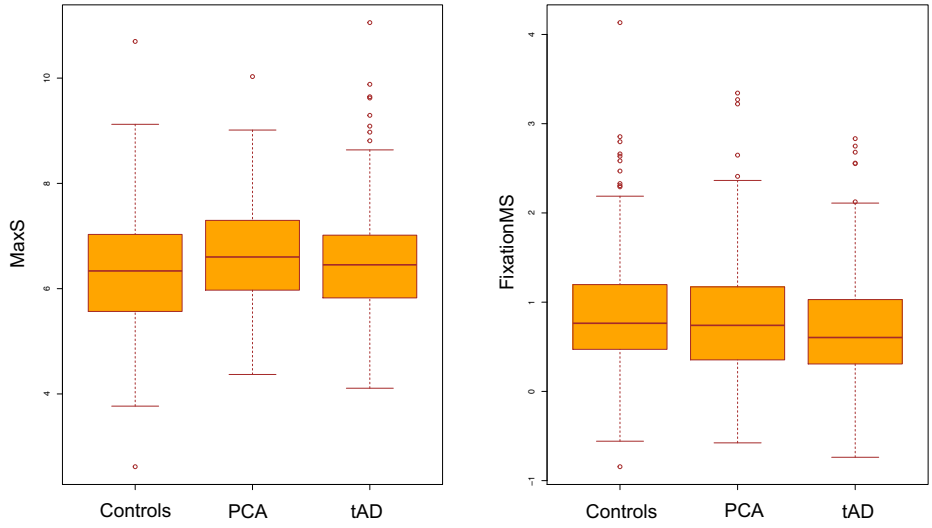

Figure 3.4: Boxplots maximum normalised saliency of fixated frames (MaxS) for each group (Controls, PCA and AD) in the left and mean normalised saliency at fixation (FixationMS) in the right side.

Completion time and saliency measures We compared the relationship between saliency and functional measures between patient and control groups. A Wald Chi-Squared test showed that there are no statistically significant effects of the interaction term maxS or FixationMS by group on completion time (Chisq = 2, df = 5.3 , p = 0.072; Chisq = 2, df = 2, p =0.36). This indicates that there are not expected to be any differences in the relationship between saliency and completion time by group.

To better understand the dynamic functional behaviour of participants, we attempted to further investigate these findings by including kinematic information, head orientation relative to the target door and displacement, as a third measure for visualisation of individual patient cases associated with slow completion times. Here, we picked two trials as examples of where patients had exhibited poor functional performance based on completion time and indirect walking paths. In Figure 4.3.4, we see a trial from a tAD patient when the light is a distractor and the table functions as clutter to his/her way to the target door. It is obvious from the diagram that during the period that the patient got lost (indicated with increased purple color) and started heading towards the close door (wrong target), there are no fixations measured. The same is the case for a PCA patient (Figure 3.6). Critical parts of the trials are missing, therefore, from our analysis. In addition, unexpectedly, in both trials we observe that the highest normalised saliency at fixation corresponds to a fixation on a small light close to the door (used to detect the end of each trial) and thus other areas (e.g. shadows on the floor due to lighting) don't appear to be as salient as expected.

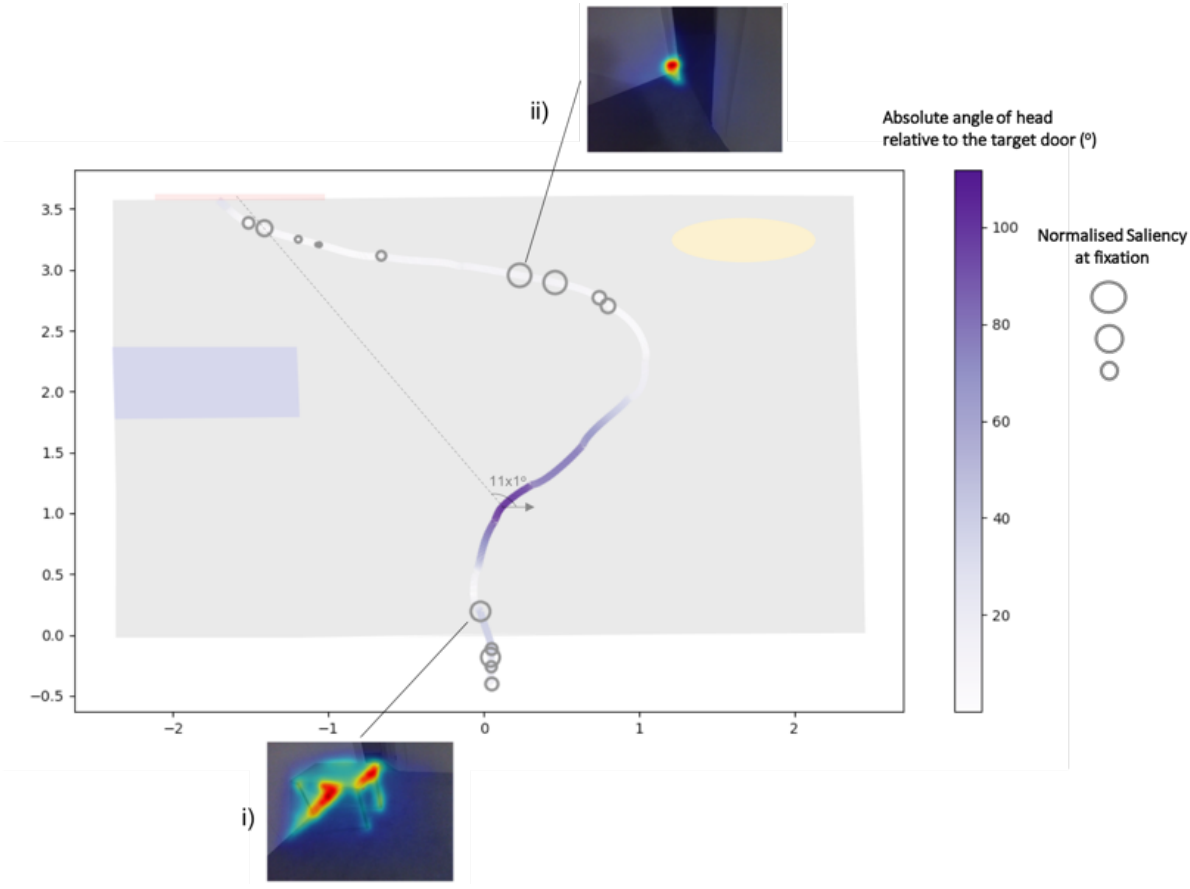

Figure 3.5: Trial from a tAD patient of light right (yellow), target left (pink) and table (lilac) overlapping with target. The plot depicts the location of the patient in the room coordinates with darker shades of purple indicating displacement values with higher absolute angle of head relative to the target door. Grey circles depict the fixation events during walking with bigger radius for higher salient values at fixations. Image i) corresponds to the last fixation before the point of maximum deviation of head orientation from the target and it depicts a saliency map with the most salient (75 percentile) parts of the image overlapped with the original image, ii) shows the frame with the maximum normalised saliency and highest normalised saliency at fixation.

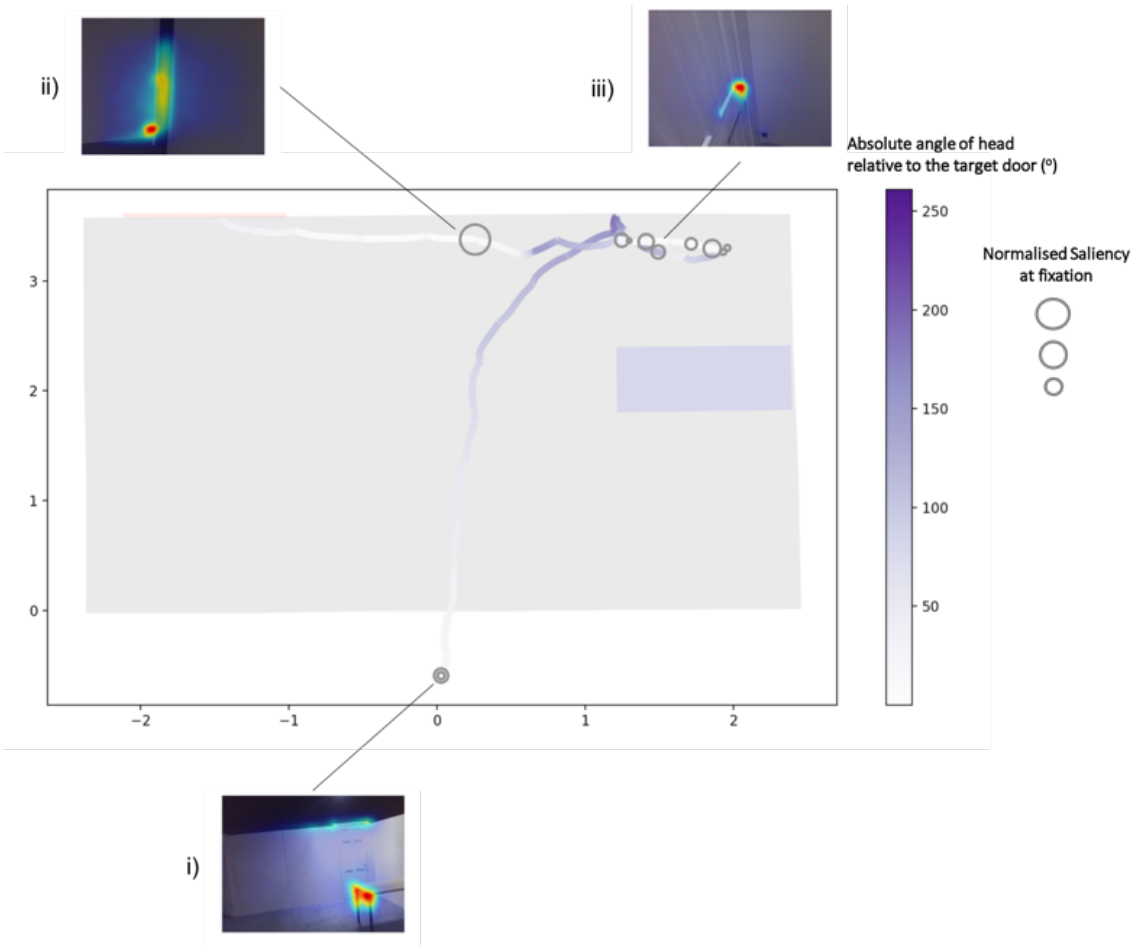

Figure 3.6: Trial from a PCA patient door left, light (3) and table not overlapping with target. Images i) corresponds to the second fixation during the trial, ii) maximum saliency at fixation during the trial and iii) highest maximum normalised saliency in POV.

3.4 Discussion

In this study, we investigated the effects of the saliency of environmental features on fixation dynamics in patients with PCA and tAD combining eye-tracking metrics with egocentric videos. Participants navigated to one of two destinations in a simplified real-world setting. The two patients' groups were matched in disease severity (mild) but differ in the clinical presentations of the disease to investigate the extent to which functional disabilities relate to visual and memory deficits in PCA and tAD respectively. To assess functional performance, a repeated-measures design was employed using completion time as a measure. Both patient groups were slower than controls, with PCA patients slower than both controls and tAD patients overall. There was not enough evidence to support patients' tendency to include salient environmental features within their egocentric frames or a strong relationship between saliency measures and function performance relative to controls.

To investigate the influences of environmental factors on slow performance, we evaluated two saliency measures for non-static scenes during recorded fixations. The first measure determined whether particularly salient environmental features were included within participants' egocentric frame. The second measure determined the saliency of features at fixation. In contrast to the static scene viewing task in Shakespeare et al. [155], in our real-world navigation setting there was no evidence of differences between PCA or tAD and controls.

These findings might show that bottom-up saliency is not the dominant factor that drives visual search in PCA and tAD during navigation in naturalistic settings, or demonstrate that the equipment and saliency measures used in this work are not able to capture saliency-related biomarkers. Another explanation might be that the patients are choosing or tending in directions not necessarily related to the room itself but to check, verify or gain confidence about things (e.g. leg position, floor flatness, walls when they are nearing a boundary) that healthy individuals don't even have to glance at/towards. This behaviour caused by patients anxiety during walking might have drag down the mean saliency at fixations. This is an inevitable downside of ecologically valid paradigms. As all glances in any direction will have some kind of environmental background, we can wrongly assume that these head and eye movements were directed to the visual scene in view, whereas there may have been other factors motivating this behaviour.

Previous findings confirm that the fixation-saliency changes with task and also the influence of top-down factors on viewing [70, 69]. Further analysis including a comparison with top-down models of attention would be needed to derive any concrete conclusions in our experiment. Mannan and colleagues, for instance, provide evidence of the dominant role of saliency when early viewing a scene and the divergence of attention to top-down processes as time progresses [110]. Our analysis aggregating the saliency measures over the course of each trial to a single might have masked the effects of this phenomenon. Alternatively, one hypothesis, for instance, might be that during visual search in navigation, the allocation of bottom-up attention might happen in different or even multiple time points or locations relative to the starting point for the different groups. However, the poor detection of fixations attributed to

fast movement of the head or body during walking (and particular when getting lost or frustrated) and the crude accuracy of mobile eye-tracking might not have permitted to extract precise measures for saliency analysis.

To better understand the dynamic functional behaviour of participants, we attempted to further investigate these findings by including kinematic information, head orientation relative to the target door and displacement, as a third measure for visualisation of individual patient cases associated with slow completion times. This visualisation gives insights about the entire time course of the trial and associates deviations in the walking path as well as in the head orientation relative to the target with characteristics (saliency) of the visual surroundings. From the case studies we investigated, it is interesting that the most salient visual feature (at least detected in the two example patient participants) is a part of the room that wasn't intended to provide a cue/salient feature (see Figure 4.3.4 (ii) and 3.6 (ii) where small light intended to detect the end of the trial is the most salient feature of the visual field).

To the best of our knowledge, this study represents the first empirical investigation of navigation in PCA using a high-volume multi-modal data to better understand which perceptual factors moderate the expression of functional impairment in real-world environments in tAD and PCA. However, this study has several limitations. The sampling frequency of the mobile eye-tracker was too low to accurately detect saccades and some fixations were detected to be out of the screen's resolution, rendering the investigation of the full spectrum of oculomotor patterns impossible and limited to a restricted number of fixations within trials. Therefore, caution is necessary with the selection of the mobile eye-tracking equipment in future investigations in naturalistic scenarios. Future directions might include the application of features used in this work (such as MaxS; maximum normalise saliency of frames) not only to eye-tracking devices but also cameras (e.g. SenseCam) to get insights related to the role of bottom-up saliency in point of view scenes. Additionally, measures such the angle of head relative to the objects in the environment (here limited to the target door), presented only in case observations in this study, could be implemented in the group level. This measure would be a surrogate of top-down saliency to measure the effect of distractors and targets in visual search during navigation, and it can also be combined with other modalities such as foot acceleration or displacement information.

Impaired way-finding and navigation abilities are disabling symptoms in dementia. Previous anecdotal evidence supports the benefits of salient visual cues on functional outcomes in AD. However, the relationship between function and the saliency of spatial features from the perspective of the observer in controlled naturalistic settings remains unclear. This study represents the first empirical investigation of navigation in PCA and tAD using eye-tracking data and egocentric videos to investigation perceptual biomarkers. It provides empirical directions for future ecological investigations in dementia with respect to the sensors and methods that can be used to extract relevant information from the signals.

Chapter 4

Augmenting Dementia Cognitive Assessment with Instruction-less Eye-tracking Tests

4.1 Introduction

Given the defining characteristics of most dementia syndromes are primarily cognitive in nature, assessment of a person's cognition is a vital component of both diagnostic services and research investigations, and is the most common outcome measure by which the effectiveness of potential pharmaceutical and non-pharmaceutical therapies is judged. Standardised paper-and-pencil cognitive assessment tools are a key component of the screening and diagnostic process, but have a number of limitations. Accurate assessments are long and associated with participant fatigue and stress, but brief tests often elicit floor and ceiling effects owing to a lack of dynamic range [122]. Literacy and education effects on cognitive scores due to the high linguistic demands of instructions, lack of reproducibility due to the assessor's subjectivity bias and ecological validity of some cognitive domains (e.g. social cognition) are further potential confounding factors [128].

Recent studies suggest that eye-tracking-based cognitive assessment might ameliorate some of the existing problems as it enables a brief and quantitative evaluation of cognitive functions [21, 128, 138]. Eye-tracking technology provides fine-grained information regarding oculomotor information (pupil dilation and gaze) and has been used to uncover eye movement abnormalities in different dementia syndromes [138, 155]. Previous studies explored its usability mainly for diagnostic purposes using it as a proxy to cognition during basic oculomotor functions (e.g. saccadic behaviour) and for evaluation of particular higher-order cognitive functions (e.g. memory, attention) [4, 120]. Recently Oyama et al. [128] used it as a communication tool during cognitive assessment to collect answers from patients with dementia and mild cognitive impairment that indicated their preference with their gaze while the tasks instructions were written on the screen. Although these tests capture critical aspects of task performance, they are still susceptible to the need for instructing patients on how to complete the tasks, which is prone to mistakes caused by misunderstandings,

language difficulties or patients at the later stages of the disease. Novel instruction-less tests might be a window to more natural, robust and ecologically valid cognitive evaluation.

Currently, the most commonly used way of summarising eye-tracking information is the computation of statistics over the pupil dilation and the gaze signal. The latter is converted to a sequential series of events predominantly consisting of fixations (eyes held stable), saccades (rapid movements to change the position of fixations) and blinks. Then, a set of eye movement event features are calculated on carefully selected spatial areas of interest. Other methods for eye movement analysis include statistics and heatmaps over raw gaze data, similarity indices of scanpaths, as well as, the so-called n-grams features that encode information for the direction and the amplitude of eye movements [81, 22, 110]. In dementia cognitive assessment, the previous features take the form of abnormalities and are expressed in terms of latency, accuracy, stability and variability [132]. However, the identification of a complete set of handcrafted features from cognitive tests sensitive to subtle task and participant-specific abnormalities is non-trivial and time-consuming. Additionally, these features are not generalisable to more complex stimuli because they rely on specific stimulus characteristics (e.g. regions of interest).

To overcome the limitations of handcrafted features, researchers have explored different computational approaches using unsupervised representation learning; by learning an embedding that captures some of the semantics of the input placing semantically similar inputs close together in the embedding space [12]. Self-supervised representation learning is a promising subclass of unsupervised representation learning which has produced state-of-the-art visual representations in standard computer vision problems [125]. This method uses information already present in the data as a supervision signal so that supervised learning techniques can be used. The rationale behind self-supervised learning is that by training a network to solve a pretext task, it encodes high-level semantic representations that are useful for solving other tasks of interest that usually have little annotated data. For sensors data, supervised representation learning with deep learning models has been shown to be competent in tasks including Human Activity Recognition (HAR) from wearable devices and detection of seizures or arrhythmia from electroencephalogram (EEG) and electrocardiogram (ECG), respectively [59, 141, 142]. However to our knowledge for eye-tracking data, supervised representation learning has only been used for detection of gaze events (e.g. fixations, saccades) from raw eye-tracking sequences and self-supervised representation learning has not been exploited [186].

In this work, we introduce a novel way of detecting abnormal behaviour and automatically extracting salient features from a novel instruction-less eye-tracking cognitive test administered to well-characterised patients with a variety of dementia diagnoses and healthy controls. We use the pretext task of identifying the particular cognitive task from which a particular eye-tracking sequence came. Labels are well-defined and known from this task and it supports self-supervised learning to identify salient features of eye-tracking sequences. Our results not only validate that instruction-less eye-tracking tests can detect dementia status but also reveal novel self-supervised learning features that are more sensitive than handcrafted features in detecting performance differences between participants with and without dementia across a variety

of tasks.

4.2 Materials

4.2.1 Datasets

Controls A: Eye-movement data from 432 healthy adults between 18 and 82 years were collected during a residency at the London Science Museum as part of the C-PLACID project. Thirty-one of these (mean age: 62.03 [SD: 7.79], 19 females [F], 12 males [M]) were over fifty years old, had proficient skills in English and reported no neurological conditions, visual impairment or dyslexia.

Controls B: Data from the Insight 46, a sub-study of the National Survey of Health and Development (NSHD) (British 1946 Birth Cohort) were also used for validation. 144 healthy individuals (67 F : 77 M) born in the same week in 1946 underwent the eye-tracking test and standard cognitive assessments at age 69-71 years. 121 of these individuals were cognitively healthy and amyloid negative based on Amyloid PET imaging.

Patients: Thirty patients with dementia (10 F : 20 M) participated in the study with mean age 68.9 years (SD : 9.16), of which 20 were less than 65 years of age at the time of their diagnosis. In terms of disease severity, their average MMSE score was 22.6 (SD: 6.68) and 18 of the patients had mild symptoms (based on correspondence with Clinical Dementia Rating scale; MMSE$>$20) [98]. These participants fulfilled standard clinical criteria for diagnosis of one of the following dementia subtypes: AD (6 subjects), bvFTD (7), lvPPA (5), nfvPPA (6) and svPPA (6).

Table 4.1: Demographic characteristics of the patients' group.

Group	Age	Gender (F:M)	MMSE
AD	74.33	3:3	17.33
bvFTD	64.14	1:6	24.85
lvPPA	65.60	1:4	22.0
nfvPPA	72.50	3:3	26.16
svPPA	66.66	2:4	22.16

4.2.2 Stimuli and Procedure

All participants (patients and Controls A & B) completed a free-viewing eye-tracking test in which 48 images were presented on a computer screen for 3 seconds each (for

a total of 192s) and their eye movements were recorded using a desk-mounted video-based eye-tracker (Eyelink 1000 Plus) at 1000 Hz. Participants were not given explicit task instructions; they were just asked to look at the screen. A chin rest was used to maintain a constant viewing distance of 80cm in all participants. Stimuli were selected to engage different cognitive functions:

1. **Scene exploration:** i) social interaction; 10 images with social (people present) and non-social context (people absent) (e.g. Figure 4.1 a.) , ii) missing items; 10 images, half complete and half incomplete (e.g. Figure 4.1 b.) and iii) social scenes; 8 images depicting either a garden or a kitchen scene where a person is present on one side of the screen and absent on the other (e.g. Figure 4.1 c.).

   The missing items task is a computerised instruction-less version of the Picture Completion subtest of the Wechsler Adult Intelligence Scale (WAIS) [172]. On this test, the participants are instructed to find the missing parts of the presented image. The test measures visual perception, in particular, visual recognition of essential details of objects and executive functioning [159]. Previous research has shown the efficacy of this test in discriminating between AD and other dementia subtypes [109, 107]. This task was included in the instruction-less battery to measure impairment in executive processes and also concentration/effort during the assessment.

   The social interaction and social scenes tasks were designed to evaluate decreased social interest which is a core diagnostic feature of bvFTD [134]. To date, the majority of research examining emotion and social perception has focused on pictures of basic facial expression. Hutchings et al. [85] found that bvFTD patients initiated more fixations to the eyes of emotional faces compared to controls. Patients with bvFTD are impaired in the Reading the Mind in the Eyes Test [9], in which participants are asked to label mental states based on visual information displayed by the eyes. Eye-tracking variants of this battery have also been developed [135]. Recently, Russel et al. [148] also created a novel instruction-less battery for identification of emotion recognition deficits in frontotemporal dementia which measures the extent by which participants match an emotion written on the screen with the corresponding photograph. Although previous research proved invaluable towards understanding social and emotional cognition, it fails to capture emotion and social skills as perceived in day-to-day life. Therefore, the social tasks in this battery attempt to explore high-order social cognitive functions including social apathy and disinhibition in more complex scenarios.

2. **Semantic processing:** 10 sentences, half of which were semantically congruent (e.g. “In the jungle there are many different animals.”) and half semantically incongruent (e.g. “She likes having a cup of injury in the morning.”), administered in pseudorandom order (e.g. Figure 4.1 d.). The sentences used in one of the batteries are displayed in Table 4.2. Half of the target words used have low frequency (based on bigram and celex frequencies).

   This paradigm was targeted towards the language-led dementia subtypes (semantic, logopenic and non-fluent variant primary progressive aphasia). Diffi-

culty in understanding sentences is a common symptom in many individuals with aphasia, especially in low frequency or complex sentences structures [49]. Previous studies have revealed increased ambivalence in gaze patterns in individuals with svPPA when looking at a target and semantically related foils compared to other variants of PPA and controls [57]. A recent study by Merck et al. [117] also reflected the semantic deficits in semantic dementia (compared to a control group) using an eye-tracking word-to-picture matching task, in which participants had to identify a target among semantically related items and distractors. Silent reading has been also evaluated using eye-tracking in individuals with aphasia [96, 49]. These studies demonstrated that structural, syntactic complexity and animated cues modulate online reading patterns in aphasia. Therefore, eye-tracking tasks seem particularly useful for discriminating individuals with svPPA and those with other variants of PPA. Here, the instruction-less semantic processing task included in this study combined stimuli from low frequency words and semantic incongruent sentences hypothesising higher overlap in gaze patterns between semantically congruent and incongruent sentences in svPPA compared to the other variants of PPA and controls.

3. **Recognition memory:** 10 pairs of images, one of which was seen previously in the social interaction task and the other which is a new image of equivalent style and complexity (e.g. Figure 4.1 d.).

   The social interaction task combined with the recognition memory task constitute a computerised version of the Visual Paired Comparison task for memory impairment detection [55]. During this task, in the familiarisation phase (social interaction) the subjects look at the pictures for a specific amount of time and then in the test phase, the subjects are presented with pictures of both old stimulus and novel stimulus (recognition memory). Previous eye-tracking studies by Crutcher et al. [44] and Lagun et al. [101] found that control subjects spend more time during the test phase looking at the novel stimulus which indicates that they have a memory of the old, now less interesting stimulus. The authors also demonstrated that this pattern is not evident in patients with tAD or mild cognitive impairment. Therefore, based on existing literature, this component of the instruction-less eye-tracking battery intends to detect memory impairment which we expect to be more prevalent on the tAD group.

There were four different versions of this test. V1-2 included all tasks (semantic processing, scene exploration, recognition memory) but had different stimulus sets. V3-4 included the same stimuli as V1-2 but excluded the semantic processing task. Controls A (Science Museum) participants were randomly assigned to one of the four versions. Controls B (Insight 46), Controls A elderly and patients were administered either version V1 or 2.

The eye-tracker was calibrated for each participant using 9 calibration points. Each trial was initiated by the experimenter and every trial was preceded by a centrally presented fixation point used as a drift correct stimulus. The fixation point also enabled a drift check, as the experiment only proceeded if the participants was looking at the drift target. Images were presented in a fixed random order within each task, and tasks were administered to all participants in the same order.

Figure 4.1: Example stimuli from the five cognitive tasks illustrated as they were presented on the computer screen sequentially (one image at a time) in the order administered: a. social interaction; image with people present (social), b. missing items; a chair with a missing part, c. social scenes; combination of pictures depicting two gardens scenes with a person present on the right, d. semantic processing; an example of a semantically incongruent sentence; ”She likes having a cup of injury in the morning”, e. recognition memory; the previously presented picture (from the social interaction task) on the right (see a.) is coupled with a new picture on the left.

sentences	target word	congruent
Some people questioned the equity of the action.	equity	true
She likes having a cup of injury in the morning.	injury	false
She had a major cinema on her leg after the crash.	cinema	false
A green candle was placed on the wooden table.	candle	true
If you ever feel in coffee run as fast as you can.	coffee	false
He will bring her a carafe of wine and flowers.	carafe	true
The most beautiful square is in that direction.	square	true
They went to the new bucket after a long time.	bucket	false
In the jungle there are many different animals.	jungle	true
Yesterday he used a danger to make sandcastles.	danger	false

Table 4.2: Sentences stimuli used in one the eye-tracking batteries.

4.3 Methodology

To mine the information of the eye-tracking time series of this instruction-less eye-tracking test, we implemented the following steps (Figure 4.2):

1. Cognitive activity recognition: Firstly, self-supervised representation learning (see section 2.7.3) was implemented in which condensed abstract representations of the input signal are learnt training a deep neural network on Cognitive Activity Recognition (CAR) based on healthy individual's data (Controls A). [section 4.3.2.1]

2. Feature relevance visualisation: Once the distribution of healthy behaviour was learnt, LRP was used to explain the networks' decisions in differentiating between cognitive activities and eventually to better understand the mechanisms

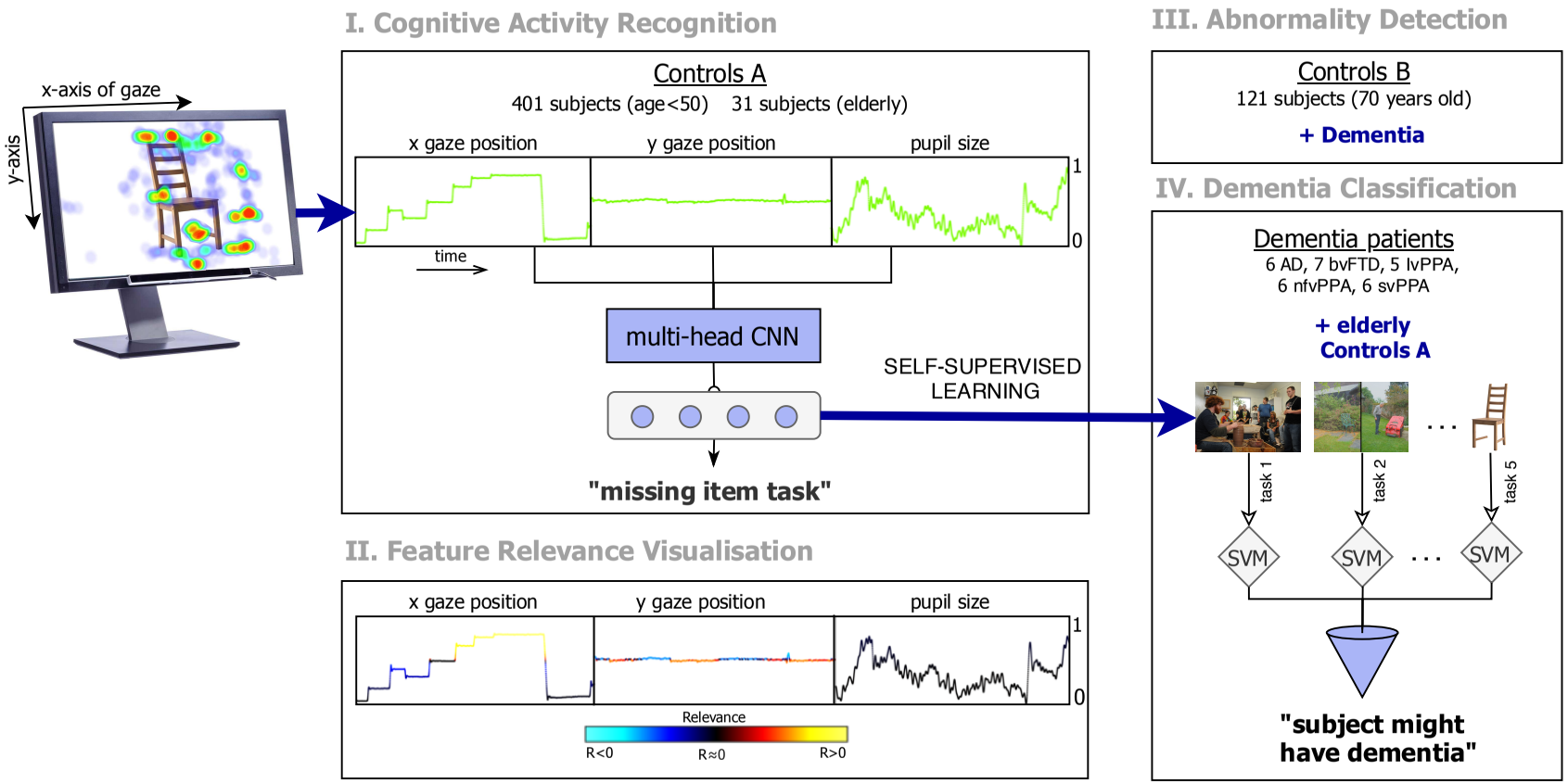

Figure 4.2: Outline of the Methodology: I. Two multi-head Convolutional Neural Networks (CNN), model A and B, were trained to identify the particular cognitive task from which a particular eye-tracking sequence came based on healthy individual's data (Controls A). II. Heatmaps based on Layer-wise Relevance Propagation technique applied on model A were visualised that show areas of the input that particularly contribute to a prediction of a cognitive task. III. Controls B and Dementia data were fed to model B for trial and subject-wise abnormality detection. IV. Model's B features learnt through self-supervised learning (by training initially a deep neural network to recognise distinct cognitive activities in healthy individuals) were transferred for subject-wise dementia classification using an support vector machine majority voting scheme.

underpinning healthy behaviour. [section 4.3.3]

3. Abnormality detection: Next, the extent to which eye-tracking features of dementia patients deviate from healthy behaviour was explored. [section 4.3.4]

4. Dementia classification: This was followed by a comparison between self-supervised and handcrafted representations on discriminating between participants with and without dementia. [section 4.3.5 for the comparison, section 4.3.1.1 for handcrafted features]

For the following analysis, the models' performance was evaluated in terms of $F1$ score which is the harmonic mean of precision and recall.

4.3.1 Data Processing

The EyeLink system recorded gaze position and pupil size in a monocular tracking mode providing 1000 samples per second. Gaze position reports the (x, y) coordinates of a subject's gaze on the display (resolution: 1920 x 1080) in actual display coordinates (pixels) with origin (0, 0) at the top left. Pupil size is reported as the pupil area measured in arbitrary units typically ranging between 100 to 10000 units. Raw samples, therefore, consist of three-time series of x, y coordinates of gaze and pupil size having a dimension of [sampling rate x trial duration].

Eye movement events were generated by the EyeLink tracker including fixations, saccades and blinks using standard velocity and acceleration thresholds. Saccades identified as containing blinks were considered blinks. Trials with total number of samples outside the screen's resolution or total blink duration more than 500ms were considered erroneous and were excluded from the analysis.

Gaze position signal was normalised to the display coordinates by dividing the gaze coordinates by the screen resolution. Missing values of gaze position were imputed with a constant zero value to avoid interpolation bias; as missing values might have a physical meaning indicating fatigue or cognitive load.

Processing of pupil size data involved discarding data before and after blinks and linear interpolation of missing values and lowpass Butterworth filter with cut-off frequency of 5 Hz. This cut-off frequency was found to be optimal for noise minimisation and signal restitution in our data. The baseline pupil size was measured as the average pupil size for a period of 300 ms immediately preceding each stimulus onset. This baseline value was selected because firstly it is a duration long enough to give a robust estimate which is longer than the average blink duration. Secondly, it is small enough to minimise the influence of pupil dilations from a previous trial since the inter-trial intervals in the battery are 1000 ms. Baseline corrected pupil diameters were computed by subtracting the baseline pupil size from the raw pupil size after stimulus onset.

4.3.1.1 Handcrafted Features

The following basic eye movement statistics were chosen to summarise the free-viewing tasks of the experiment:

saccade counts, total duration of saccades, median of the length of saccades (x-coordinate of gaze), number of progressive saccades (forwards), number of regressive saccades (backwards), fixation counts, mean/max/standard deviation, blinks counts, total duration of blinks and total duration of fixation duration, mean/std/min/max of peak velocity, visual angle, pupil size, pupil size during fixations, x and y coordinates of gaze.

Scanpath length, namely, the Euclidean distance of saccadic movements with respect to x and y position of gaze, was also selected as a measure of the overall functional performance of participants since it has been associated with higher fluid intelligence scores in healthy individuals [150].

Overall differences in eye-movement handcrafted features across all tasks between healthy controls and dementia patients were evaluated using a Generalised Estimating Equation (GEE) model with independence correlation structure and robust standard errors to adjust for repeated measures for each subject [103]. In addition to the group category (controls/dementia), the following variables were included in the GEE models: age, education, gender, task and task by group interactions.

4.3.2 Representation Learning Methodology

4.3.2.1 Cognitive Activity Recognition

Cognitive activity recognition from eye movements was used in this work as the pretext task and a deep neural network was trained in a trial-wise manner (rather than subject-wise) given the raw eye-tracking signals. CAR can be considered as a classification problem, where the inputs are time series and the outputs are the cognitive task one is being assessed on (semantic processing; scene exploration; recognition memory). In particular, a multi-head Convolutional Neural Network (CNN; see section 2.9.0.3) architecture was implemented which takes as an input the three eye-tracking time series, processes them separately by individual one-dimensional convolutional heads and extracts features specific to each time series [27]. This network's architecture processes the entire sequence at once generating a single feature map for each sample and then all the features maps are concatenated. In this way, the features extracted from each time series are kept separated which improves the interpretability of the model and captures better data of different natures and scales that are not correlated (e.g. gaze coordinates and pupil size). After the feature extraction stage, a global average pooling operation was applied which calculates the average output of each feature map and prepares the model for the final classification layer.

Defining the number of output classes is not straightforward, as the instruction-less nature of the test increases between-trial variability causing label ambiguity. Although the stimuli were designed to trigger specific reactions, it is not guaranteed that participants were performing in a similar/uniform way, especially on the missing item, social

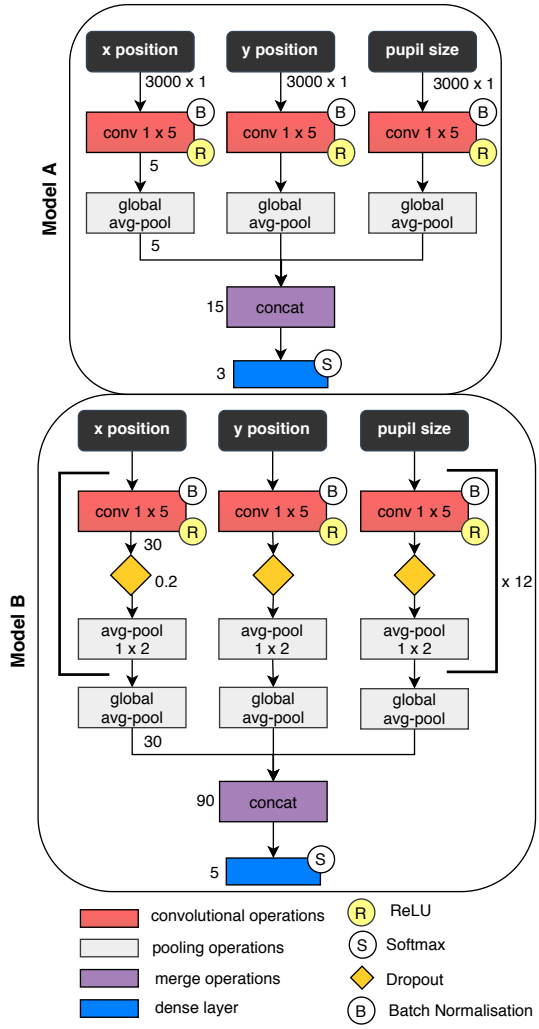

Figure 4.3: Model A and B Neural Networks Architectures for cognitive activity recognition with 3 (scene exploration, recognition memory and semantic processing task) and 5 (missing items, social scenes, social interaction, recognition memory and semantic processing task) output classes respectively.

scenes and interaction tasks which fall broadly into the scene exploration task.
The following two multi-class problems were investigated which differ on the number of output classes in the final classification layer:

i. Model A: Three-class problem (scene exploration, recognition memory and semantic processing task)

ii. Model B: Five-class problem (missing items, social scenes, social interaction, recognition memory and semantic processing task)

Figure 4.3 demonstrates the architectures of model A and B. Model A consists of a single 1-D convolutional layer (kernel size 5, stride equal to 1, no padding) with 5 features maps for each input signal followed by a batch normalisation layer and a ReLU activation function. Model B includes 12 blocks of the following architecture

in the order presented: 1-D convolutional layer (kernel size 5, stride equal to 1, no padding) with 30 features maps, batch normalisation layer, ReLU activation function, dropout layer ($p = 0.2$) and average pooling layer (pool size 2). In both cases, a global average pooling layer follows the feature extraction block and reduces the dimension to 15 and 90 features for model A and B respectively. These features are the input of a perceptron applied with a softmax activation function.

4.3.2.2 Training Details

Hyper-parameter selection and model comparisons were implemented within the following pipeline: data were split in train and test set under the constraint that trials of an individual appear in only one of the sets (leaving some participants out). 5-fold cross-validation was implemented on the train set for each combination of parameters selected using grid or random search. The set of parameters with the best 5-fold cross-validation score ($F1$ score) was selected and the model with weights from the best fold was evaluated on the test set.

The proposed framework was implemented and trained in KERAS. The network parameters were optimised by minimising the categorical cross-entropy loss function using gradient descent with Adam optimiser having learning rate of 0.001 and batch size of 50. Weights were randomly orthogonally initialised with the Glorot normal initialiser. The L1-norm weight regularisation was applied with regularisation rate of 10^{-5}. The maximum number of epochs was 50 and early stopping was implemented which stops the training process if the validation loss does not increase for 50 contiguous epochs. After training the model, only the weights of the epoch with the higher $F1$ score on the validation set were saved and used for the evaluation of the model.

4.3.2.3 Data Augmentation

Given the intrinsic within and between person variability and the limited amount of eye movement data, data augmentation can be used to prevent overfitting and improve the generalisability of the models [168]. Finding invariant properties of the data against certain transformations is the main idea behind the selection of the following four techniques implemented: shifting, jittering, scaling and cropping. For each gaze position time series, transformations were applied by randomly selecting two out of the four techniques.

Shifting involves generating samples by shifting x and y coordinate of gaze by a scalar (randomly sampled from the interval [-10, 100] for the semantic processing task and from [-100, 100] for all other tasks). In this way, we covered unexplored input space by accounting for variability of the movement while preserving the shape of it. Jittering is a way of simulating additive noise attributed to varying levels of gaze stability in individuals or noise associated with the sensor. Three seconds of Gaussian noise was generated with a standard deviation value sampled from a uniform distribution $U(0.05, 1)$. Scaling the input by multiplying the x and y coordinates of gaze with the same scaling factor attempts to change the magnitude of the signal and subsequently slightly the shape of the gaze scanpath. The scaling factor was sampled from

the normal distribution with a mean of one and standard deviation between 0.05 and 0.2 for the semantic processing task or 0.1 for the others. Cropping the input involves removing the first (or last) x time points (x in [5, 100]), shifting the signal x points in the time axis and consequently interpolating with zero values to keep the original dimension.

4.3.3 Feature Relevance Visualisation

LRP, proposed in [6], was applied to the best performing CAR model to better understand the mechanisms underpinning healthy behaviour during different cognitive activities. LRP attempts to explain the decisions of non-linear models such as deep neural networks. The goal of this technique is to quantify the contribution of each component of an input a to the prediction $f(a)$ made by a given decision function f. To this aim, LRP decomposes f attributing relevance scores R_i to all components i of a such that $f(a) = \sum_i R_i$. The algorithm starts from an output neuron j by defining $f(a) = R_j$ and it iterates over all the layers of the model backwards to the input attributing relevance messages to each neuron under the constraint that the total amount of relevance is conserved in each layer.
The relevance value being propagated from neuron j to its input i is proportional to each input i contribution to the activation of the neuron j and is defined as:

$$R_{i \leftarrow j} = \frac{z_{ij}}{z_j} R_j, \tag{4.1}$$

where where z_{ij} is the contribution of the input neuron i to the output neuron j and $z_j = \sum_i z_{ij}$. In this work a modification of this formula is used, the so-called ϵ-rule, which introduces a stabiliser $\epsilon > 0$ to the denominator of formula 4.1 to avoid possible unbounded values of $R_{i \rightarrow j}$ with small values of z_j. The relevance score R_i at input neuron i is then obtained by summing all incoming relevance values $R_{i \rightarrow j}$ from the output neurons to which i contributes to and is defined as:

$$R_i = \sum_j R_{i \leftarrow j}. \tag{4.2}$$

By replacing $R_{i \leftarrow j}$ with the above formula, it is obvious that a neuron is relevant if it contributes to neurons that are relevant themselves.

In the CAR classification setting, for input neurons i, $R_i \approx 0$ indicates inputs with no or little influence on the model's decision, $R_i > 0$ represents parts of the input that explain a specific class while $R_i < 0$ contradicts the prediction of that class.

Once the relevance values were computed, they were normalised to the interval $[-1, 1]$ by dividing with the maximum absolute relevance value of the entire input signal.

4.3.4 Abnormality Detection

Once the normal behaviour during this cognitive assessment was learnt, the extent to which the eye movement patterns of dementia patients deviate from it was investi-

gated. In particular, a question of interest is whether dementia patients passively look at the screen without following the implied instructions (e.g. reading when a sentence is presented on the screen) which is the expected activity from the controls. To investigate that, data from unseen elder controls (controls B) and dementia patients were fed into the pre-trained CAR neural network and the number of misclassified samples were estimated for each cognitive task and group. Since misclassifications might be attributed to behaviour patterns not seen in the training set, abnormality was defined in relation to an unseen elderly controls' dataset. A threshold that discriminates normal from abnormal cases for each cognitive task was created by calculating the average predicted probability of an elder control trial belonging to each cognitive task classes. A trial from the dementia group was considered abnormal if the model assigned it to a class with probability less than the threshold value of that specific class. Finally, a majority voting strategy was applied to determine abnormal participants of the dementia cohort using the median value of the abnormality scores of their trials.

4.3.5 Dementia Classification

Since the ultimate purpose of this instruction-less cognitive assessment is the detection of dementia related oculomotor biomarkers, we evaluated whether the representations learnt using the cognitive activity recognition task (i.e. pretext task) are useful for dementia classification (i.e. target task). If the representations learnt are general and not specific to the pretext task, then the target task is expected to perform well. To this aim, the data of the elderly healthy controls A and dementia patients were fed to the pre-trained CAR neural network and the outputs of the average global pooling layer were the features to be transferred for dementia classification. The performance of a support vector machine (SVM) classifier with these abstract features and handcrafted features was compared. The following procedure was applied to both feature sets.

The features were adjusted by controlling for potential confounding effects of gender, age and education levels before being fed to the classifier. A multivariate multiple linear regression model was fitted on the controls' data with the features as dependent variables and age, education and gender as independent variables. Subsequently, the residuals were calculated for the features matrix which measure unexplained variance presumably attributed to the task or to other individual characteristics. Using the model with the estimated coefficients the residuals were also calculated for the patients' features. The inputs, therefore, of the classifier were the residuals instead of the initially calculated features. In addition, all the features were standardised by removing the mean and scaling to unit variance inside the cross-validation procedure using the mean and variance of the train set.

For each cognitive task, we extracted abstract or handcrafted features for each trial and then made trial-wise predictions of dementia status based on all set of features, whether abstract or handcrafted. Five SVMs (with a radial basis function (RBF) and tuning parameters the kernel coefficient (γ) and the penalty parameter (C); see section 2.9.0.2) were fitted to the features of each task separately (missing items, social scenes, social interaction, recognition memory and semantic processing task)

[36]. This ensemble approach was preferred to a single global classifier, because we hypothesised that the task information would improve the predictions. Lastly, to obtain subject-specific from trial-wise predictions, a majority voting scheme (median operation) was applied to the five classifiers' predictions; twice for each subject (Figure 4.2). In more detail, for each cognitive task, the corresponding SVM made several predictions (votes) for all trials of each subject (e.g. for semantic processing 10 predictions for each subject). The final prediction for each subject's performance on a particular task was the one that received the most votes. Finally, the global output prediction of each subject was the one that received more than 3 votes (out of 5).

Nested cross-validation was implemented for the evaluation of the classifiers: data were split into a train set, within which parameters were selected with 5-fold cross-validation, and a test set, for evaluation. It was ensured in the process that the same participants appeared in the test sets for all five classifiers. This process was repeated 100 times and since the classifier's performance was evaluated in terms of $F1$ score, 100 $F1$ scores were obtained for each experiment.

A label permutation test was implemented as a baseline which determines the performance of the model when there is no relationship between the features and the output labels. The features of the best performing model were selected for this procedure. Comparisons between the performance of the model with different feature sets were made using Mann-Whitney U test and bootstrap confidence intervals with 1000 iterations were calculated.

4.4 Results

4.4.1 Cognitive Activity Recognition

Table 4.3 summarises the results of the CAR model, which classifies cognitive activity given eye-tracking data, in terms of 5-fold cross-validation and left-out test set performance with two combinations of output classes (3 or 5 output neurons) and four combinations of training datasets (with or without augmented data, controls B and combined controls A and B). The original dataset includes 15,996 trials and the augmented 18,793. Figure 4.4 shows that the distribution of the computed statistics (mean, variance, range) for x and y coordinates of gaze for original and augmented data are similar. In cropping, the range of the samples is higher because the minimum value is always zero, as the signal is interpolated. The best performance in terms of $F1$ score on the test set appears to be on the simplest model with three classes (scene exploration, memory and semantic processing) trained on the original dataset of healthy controls. Data augmentation and increasing the size of the training set (A and B) improves slightly the performance of CAR5 but not CAR3 model.

4.4.2 Feature Relevance Visualisation

Relevance maps were computed for the best performing model (CAR3) for both dominantly and not-dominantly firing output neurons as the latter can reveal interest-

Table 4.3: Performance scores of the multi-head CNN models on activity recognition with different multi-class and augmentation settings evaluating with 5-fold cross validation and the left-out test set.

Model	Control Dataset	Output classes	Augment	CV $F1$	Test $F1$
CAR3	A	3	False	0.955 (0.01)	0.967
CAR3_AUG	A	3	True	0.948 (0.014)	0.954
CAR3	B	3	False	0.941 (0.014)	0.926
CAR3	A + B	3	False	0.946 (0.012)	0.959
CAR5	A	5	False	0.841 (0.023)	0.821
CAR5_AUG	A	5	True	0.857 (0.019)	0.834
CAR5	B	5	False	0.854 (0.014)	0.859
CAR5	A + B	5	False	0.852 (0.01)	0.854

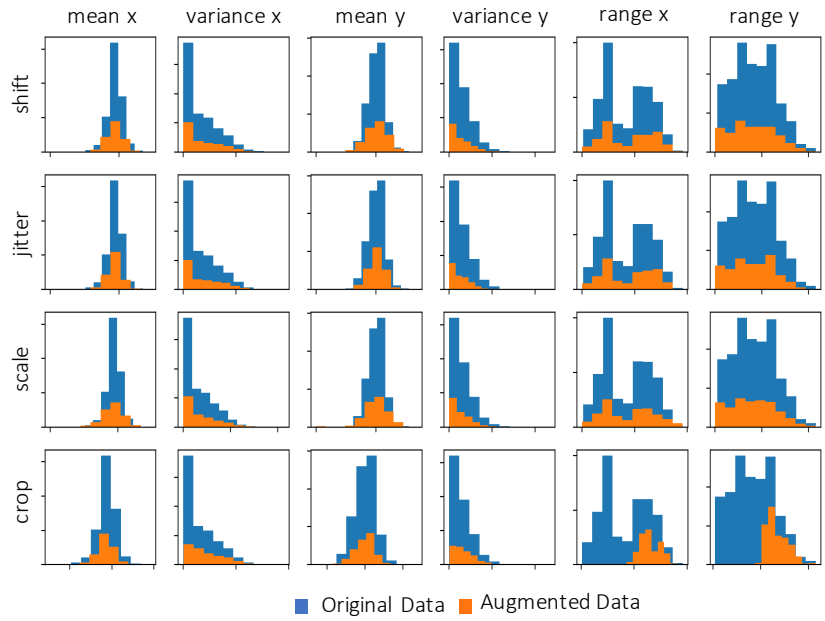

Figure 4.4: Histogram of statistics for original and augmented data. Statistics: mean, variance and range of the time series signal were computed for all samples for x and y coordinate of gaze.

ing information about the learnt strategy of the model e.g. why a certain class has not been picked for prediction. Figure 4.5 provides some insight on the methods the network uses to classify with high certainty eye-tracking trials belonging to the scene exploration (neuron 1), semantic processing (neuron 2) or memory recognition task (neuron 3). It shows the contribution of the input to the prediction of each class, or in other words, to the output of each neuron of the final layer of the model. Positive values of relevance in the not-dominantly firing neurons indicate parts of the input sharing properties with the dominant neuron. Negative values in the not-dominantly firing neurons indicate parts of the input that significantly oppose the properties of the dominant neuron.

Figure 4.5.a constitutes an example of a semantic processing trial of a healthy control which is correctly classified with 0.982 probability. The neural network attributes positive relevance to peaks, high values or gradually increasing stair-wise trends of the x coordinate of gaze. The network discriminates between scene exploration and semantic processing or recognition memory largely by these x position properties of the signal. This is reflected in the negative relevance of the same points in neuron 1 compared to neuron 2 and 3.

Figure 4.5.b displays an example of a memory recognition task (class 3) which is correctly classified by the network with probability 0.981. As seen in Figure 4.5.a previously, positive relevance values are attributed primarily in the x position of gaze in both neuron 2 and 3. Here to discriminate between memory recognition and semantic processing, the network looks at the y position of gaze and gives higher values of relevance to fixations (flat areas) with higher values (bigger jumps) of the y coordinate of gaze. Intuitively this means that it learns that in the semantic relative to the memory

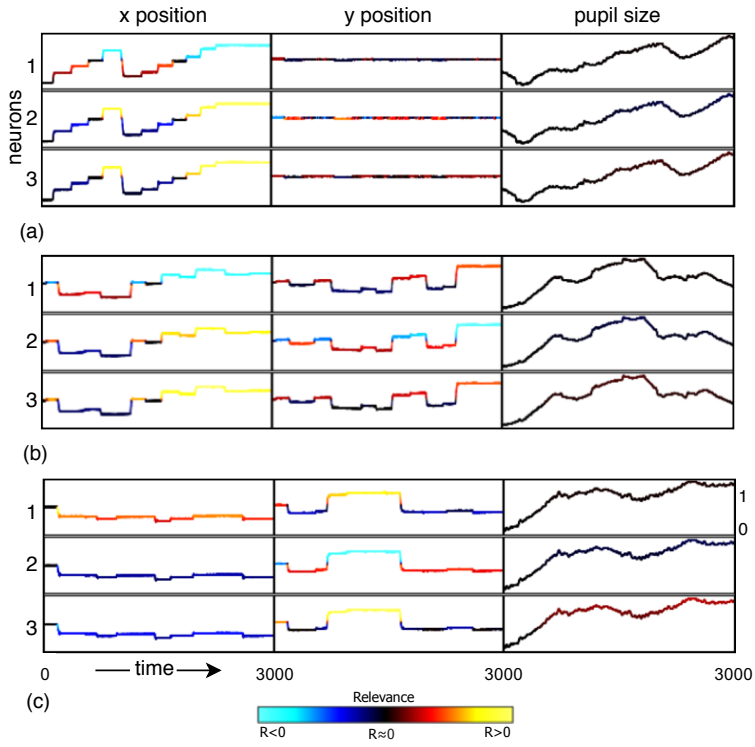

Figure 4.5: Relevance plots of Cognitive Activity Recognition (CAR) features discriminating between cognitive tasks in healthy controls. Features are presented from the best performing activity recognition model (CAR3). Three different input eye-tracking samples (a, b, c) are presented, each of a separate healthy control performing a reading (a) or episodic memory (b, c) task. Rows: Relevance maps with respect to the network's (class representing) output neurons (neuron 1: scene perception, neuron 2: reading, neuron 3: episodic memory). Columns: x, y coordinate of gaze and pupil size respectively. Warm hues (standing for $R > 0$) identify input components supporting the model prediction and cold hues (mapped from $R < 0$) pointing out evidence in the input considered as contradictory to the learned class by the model. (a). The model indicates that peaks, high values or gradually increasing stair-wise trends of x coordinate of gaze are associated with either the reading or episodic memory task. (b). These trends accompanied by relatively stable values of y position of gaze are attributed to the reading task, whereas long fixations with higher values (bigger jumps) of the y coordinate of gaze to the episodic memory task. (c) In samples where there are no jumps with respect to the horizontal axis of the screen, the network identifies big jumps in y position of gaze and variations in pupil size as features associated with the episodic memory task.

task the eyes stay relatively still with respect to the vertical axis of the screen while moving horizontally to read the sentence.

In the memory recognition task when the previously seen feature of x position of gaze is not apparent, i.e. there are no jumps with respect to the horizontal axis of the screen, the network looks for big jumps at the y position of gaze (Figure 4.5. c, probability = 0.99). Interestingly, since this property is shared between scene exploration and memory recognition task, the network classifies the trial as memory recognition

relying also on the pupil size signal.

4.4.3 Abnormality Detection

4.4.3.1 Handcrafted Features

Based on the results of the handcrafted features, overall dementia patients searched less extensively and scanned the stimuli significantly more slowly than controls with lower scanpath lengths (estimate = -276.56, SE = 97.09, z = 8.11, p = 0.00439) (Table 4.4). Mean x position of gaze was lower in dementia patients compared to controls (estimate = -22.895, SE = 11.220, z = 4.16, p = 0.0413). There was also a significant interaction between the effects of group and task ($p < 0.0001$); dementia patients showed a greater relative impairment relative to controls in the semantic processing task, looking at lower values of the x coordinate of gaze on average when the sentences appeared on the screen. The same patterns appear on median and max x coordinate of gaze (max: estimate = -37.466, SE = 15.999, z = 5.48, p =0.0191, median: estimate = -29.1, SE = 12,7, z = 5.25, p =0.022).

4.4.3.2 Self-supervised Learning Features

To investigate whether dementia patients passively look at the screen without following the implied instructions, the CAR5_AUG model trained on healthy behaviour was used. The percentage of misclassified trials when the controls B validation set vs dementia data were fed into the model were higher for the dementia patients for all the cognitive tasks: social scenes (Controls: 27.4% vs Dementia group: 37.3%), semantic processing (0.4% vs 2.7%), memory recognition (5.7% vs 8.8%), social interaction (22.4% vs 28.5%), missing items (13.6% vs 19.2%). The distribution of the predicted probabilities of a trial belonging to a task were statistically significantly different between controls B and dementia patients trials apart from the social interaction task (social interaction: z = 20067.5, p = 0.067, semantic processing: z = 24236, $p <$ 0.0001, missing items: z = 14576.5, p= 0.0084, social scenes: z = 13682, $p < 0.0001$, memory recognition: z = 17305, p = 0.0002).

In terms of the detection of abnormal participants, even in the absence of explicit task instructions, 13 out of 30 dementia patients were considered abnormal in the social scenes task (threshold p = 0.6851), 10 in the social interaction task (p = 0.71) and 4 in the missing items task (p = 0.808).

4.4.4 Dementia Classification

Figure 4.6 and Table 4.5 summarise the results of the model on the dementia classification task using handcrafted and deep learning features obtained from different variations of the CAR models presented above. Overall, the features from CAR5 present the best results capturing differences between the two groups (95% CI: [0.7870, 0.8241]). The handcrafted features [0.6175, 0.6723] show lower performance compared to CAR5_AUG [0.7522, 0.7944] (t = 2334 , $p < 0.0001$), CAR5 (t

Table 4.4: Wald statistic and p values for group category variable (dementia /controls) of the GEE models with outcome the handcrafted features and independent variables: age, education, gender and task.

Features	**Wald statistic**	**p-value**
fixation counts	0.28	0.59468
saccade counts	0.78	0.38
saccade median length	3.67	0.0553
progressive saccades counts	0.12	0.727
regressions counts	1.66	0.20
blink counts	2.32	0.1279
fixation duration sd	1.01	0.315
fixation duration mean	0.17	0.6773
fixation duration sum	0.06	0.813
saccade duration sum	3.18	0.07468
blink duration sum	1.67	0.19602
peak velocity mean	1.87	0.17124
peak velocity sd	0.02	0.902
peak velocity min	2.35	0.1257
peak velocity max	0.54	0.463
pupil during fixation mean	0.51	0.4743
pupil during fixation sd	0.06	0.81
pupil raw mean	0.0001	0.99
pupil raw sd	0.02	0.8845
pupil raw min	0.0001	0.96
pupil raw max	0.0002	0.98
pupil raw median	0.02	0.878
x mean	4.16	0.0413 *
x sd	2.86	0.091
x min	0.04	0.84
x max	5.48	0.01919 *
x median	5.25	0.022 *
y mean	2.09	0.148
y sd	0.09	0.7615
y min	1.31	0.25
y max	2.85	0.091
y median	1.30	0.255
scanpath length	8.11	0.00439 **

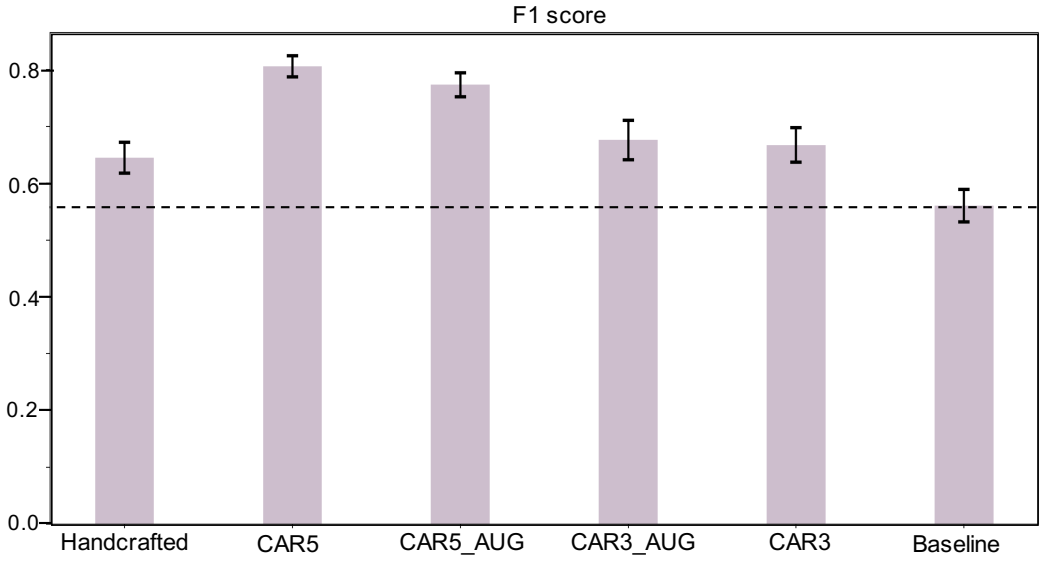

Figure 4.6: Performance of the SVM-ensemble model trained on the dementia classification task in terms of $F1$ score using different handcrafted and deep learning features. The self-supervised learning features were transferred from different variations of the CAR neural network trained on activity recognition. CAR3 and CAR3_AUG have three output neurons (scene exploration, recognition memory and semantic processing task) and are trained with and without data augmentation, respectively. CAR5 and CAR5_AUG have five output neurons (missing items, social scenes, social interaction, recognition memory and semantic processing task). For a baseline, the case where there is no relationship between the features and the output labels is considered. Bars represent 95% bootstrap confidence intervals.

= 1628, $p < 0.0001$) but not CAR3_AUG [0.6412, 0.71097] (t = 4371, p = 0.094) and CAR3 [0.6367, 0.6979] (t = 4287, p = 0.064), mainly as CAR3_AUG and CAR3 are performing better, but this improvement is not statistically significant. There was no evidence that data augmentation improved classification performance in the CAR5, (t = 4195.5, p = 0.013) nor in CAR3 problem (t = 4894.5, p = 0.398). Both handcrafted features and CAR5 differ significantly from the baseline case (t = 3479, $p < 0.001$, t = 1628, $p < 0.001$).

Table 4.5: Mean of 100 iterations of nested cross-validation metrics (TN: True Negative, FP: False Positive, FN: False Negative, TP: True Positive) of the SVM-ensemble model trained on the dementia classification task and tested on 6 patients and 6 controls.

Model	TN (%)	FP (%)	FN (%)	TP (%)
CAR3	37.08	12.91	16.33	33.66
CAR3_AUG	33.5	16.5	15.25	34.75
CAR5	33.5	16.5	5.16	44.8
CAR5_AUG	30.5	19.5	6.33	43.66
Baseline	11.08	38.91	12.91	37.08
Handcrafted	34	15.98	17.94	32.05

4.5 Discussion

A mixed sample of well-characterised dementia patients varying in disease severity participated in a free-viewing test which was designed to assess specific cognitive functions selectively impaired in different subtypes of dementia. In this study, we present evidence that firstly brief instruction-less eye-tracking tests can detect abnormal oculomotor biomarkers and secondly self-supervised representation learning techniques can extract more discriminative features from this instruction-less eye-tracking cognitive test that are more discriminative than standard handcrafted eye-tracking metrics.

To assess the overall functional performance of participants in the cognitive test, scanpath length and some relevant handcrafted features were computed. We found that dementia patients search less extensively and scan the stimuli significantly more slowly than controls. They also present a tendency to fixate towards the left side of the screen during sentence presentation compared to controls, which might indicate that either they are slower in reading or they are not reading the sentences. While these primary findings are unable to indicate the basis of such abnormal performance, such as whether this relates to a diminished ability to adapt eye behaviour in response to task demands [155], they demonstrate that features extracted from even this brief and instruction-less test may detect abnormal oculomotor biomarkers of dementia-related cognitive dysfunction.

With the aim to evaluate whether the dementia patients were performing the activities that were implied by the test, we first tried to understand the cognitive behaviour of the average healthy control. We trained a neural network on healthy controls to predict cognitive activity from eye movements and the network's decision strategies were analysed. All three input channels of the network, x and y coordinates of gaze and pupil size, contribute to the prediction of the cognitive task one is performing;

with the first two having a higher impact. The network's decisions seem to associate the combination of jumps towards high values of x position and stable y position or jumps towards y position with the sentence reading task (since one normally scans the screen from left to right when reading) and the recognition memory respectively. Interestingly, to discriminate between scene exploration and memory recognition, the networks seem to use information from pupil size which is consistent with previous evidence of pupil response being modulated during memory tasks [91].

When this network was used to classify cognitive activities of elderly controls and dementia patients, higher misclassification errors were observed in dementia patients than in controls indicating that dementia patients perform the distinct cognitive tasks differently than healthy participants. If slower performance in a task is associated with eye-tracking sequences with the same features to the average performance but shifted later in time, then we know that trials of slower participants are not misclassified by the network. This is because according to the equivariance to translation property of convolutional neural networks, two different input signals with the same feature presented in different locations in the input space produce the same output. Misclassified trials of dementia patients, therefore, might be attributed to cases where the mechanisms underpinning cognitive activities differ substantially to controls.

The cognitive activity recognition pretext task not only contributes to the detection of abnormal behaviour but also provides general condensed representations of the eye-tracking data useful for dementia classification from a variety of cognitive tasks. This is demonstrated by the ability of our framework to predict dementia status in this heterogeneous group with an $F1$ score between 0.7870 and 0.8241. This result was achieved with abstract features obtained from the most complex model (deep neural network) trained on the most difficult classification problem (activity recognition of five cognitive tasks) with 90 features in total. Although this model achieves a significantly lower performance on activity recognition than other less complex models, it learns richer representations of eye-tracking data that are more sensitive in detecting performance differences between participants with and without dementia. In addition, all sets of abstract features outperform standard handcrafted features, highlighting the added value of new feature extraction techniques for eye-tracking data from cognitive tests especially under the lack of instructions. These findings demonstrate the importance of self-supervised representation learning to healthcare applications in the absence of a large number of patients data. Future work would investigate further whether the improvement in performance using abstract features is attributed to a specific input channel (e.g. horizontal gaze movement and not vertical movement or pupil size, as indicated by the statistical analysis in Table 4.4) by fitting models separately for each channel.

To the best of our knowledge, this is the first application of deep learning for classifying and interpreting cognitive activity and dementia status from raw eye-tracking measurements. These methods were applied to a particularly complex dataset that included different versions of an instruction-less cognitive test with varying levels of stimulus complexity (abstract scene viewing versus simple sentence stimuli). Additionally, the test was also administered to clinically well-characterised patients, not only those with typical presentations, but a combination of rare dementia syndromes varying in disease severity. Our results show that self-supervised representation

learning methods hold promise for augmenting cognitive assessment with instruction-less eye-tracking tests to monitor patients at different stages of the disease in a brief, low-stress manner.

This study opens the door to a more ecologically valid assessment of natural cognitive behaviour in dementia. It develops an instruction-less paradigm for the assessment of multiple domains of cognition. This paradigm provides an alternative to lengthy batteries of individual tests that have different task demands, properties and instructions. However, this is not a clinic-ready screening test, as more work needs to be done to evaluate its validity. Future evaluation of patients in the early stages of dementia, with mild cognitive impairment or living at autosomal dominant genetic risk of a dementia, might determine whether this battery can be used for early detection of cognitive change/impairment. Next stages of development should include a comparison of test performance with results on established paper-and-pencil neuropsychological tests of each cognitive domain. The battery could also be expanded to incorporate additional cognitive domains (e.g. social cognition). Moreover, the current method is not able to show whether eye-tracking metrics are sensitive to the different dementia subtypes nor evaluate the effectiveness of the specific parts of the tests to the targeted groups (e.g. memory test for tAD patients). This can be potentially addressed in the future with the recruitment of larger within-subtype dementia cohorts. From a methodological perspective, although the current dementia classification methodology shows whether the features learnt in the pretext task are meaningful for dementia-related abnormality detection, it might not be the best approach for screening patients highlighting abnormalities in different subtypes and stages of the disease. The reason being is that it assumes a homogenous pattern of abnormality in the dementia group, which might not be true given the variability of eye movement behaviour between subjects. Anomaly detection based on detecting outliers given a distribution of normal behaviour might be a more appropriate tool here for future research.

To conclude, this work highlights the contribution of self-supervised representation learning techniques in medical applications where the small number of patients, the non-homogenous presentations of the disease and the complexity of the setting can be a challenge using state-of-the-art methods. It also demonstrates that the application of methods for interpreting artificial intelligence systems constitutes a window to better understand human cognitive functions. The proposed methodology of the unsupervised representation learning technique with the LRP interpretability framework presented above is applicable to different cognitive tests, instruction-less or not, under the only assumption that they include activities associated with distinct eye-movements.

Chapter 5

Oculomotor anomalies in instruction-less eye-tracking tests

5.1 Introduction

Defining impairment typically based on normative data is critical in clinical neuropsychology. Neuropsychologists and neurobehavior specialists are faced with the challenge of decoding a plethora of numerical and qualitative data, that to be meaningful must have a frame of reference. Normative data are typically obtained from a large sample of cognitive healthy individuals appropriately stratified by demographics, reflecting healthy performance on a specific test adjusted for relevant demographic factors. Individuals performance is compared and contrasted to this reference group [119]. Although such definitions are considered standard in neuropsychological assessments, they have not been used in more complex cognitive measures such as eye-tracking data [24].

Most eye-tracking studies in dementia research have focused on describing statistically significant differences in basic oculomotor features or features within areas of interest between different cognitive tasks or groups (e.g. patients versus controls). Although these approaches can reveal associations between cognitive function and oculomotor behaviour, this information can not be used for a diagnosis of a single new subject. Eye-tracking technology could be used to support diagnosis only if expert review by a highly specialised professional was available. Additionally, relying only on known biomarkers restrains the potential of eye movement time series containing far richer relevant information. A few studies have used machine learning methods to eliminate these problems; these methods identify discriminative patterns in the data between groups of subjects. For instance, Pavisic et al [132] and Biondi et al [17] applied a Markov model and a neural network, respectively, to eye-tracking data from a reading and smooth pursuit task that discriminate between Alzheimer's patients and controls. These classification approaches try to find the discriminative boundary between the classification classes assuming a homogenous pattern of abnormality in the patient group (e.g. similar degrees of disease severity) and well-balanced classes in terms of sample size; which might not be always valid.

Following an alternative approach, Primativo et al. [138] implemented a Bayesian model which is first trained to predict gaze coordinates in controls in a spatial anticipation task. Then a dementia diagnosis prediction is made based on the magnitude of the error between the model predictions and the real values in patients (bvFTD and svPPA) compared to controls. Although this approach is explicitly fitted to model data from a spatial anticipation task, it introduces the concept of defining abnormal behaviour based on deviations from a normative reference.

Anomaly detection involves identifying unexpected items or events in unseen data instances that deviate from the normal distribution (which is learnt during training). Deep learning-based algorithms have received a lot of attention recently and have been applied in various tasks ranging from video analysis to Internet of Things (IoT) systems. Anomaly detection plays a particularly prominent role in the healthcare domain; with studies demonstrating that malignant tumors can be inferred from anomalous MRI images and cardiac problems can be detected from anomalous electrocardiogram traces [29]. Anomaly detection has the potential of providing healthcare professionals with useful information to make clinical decisions as well as respond faster to adverse events. For sensor data, although both supervised and unsupervised deep learning models are competent for anomaly detection, the latter plays a more and more important role for applications including detection of seizures or arrhythmia from electroencephalogram and electrocardiogram, respectively [133, 184]. To the best of our knowledge, for eye-tracking data, unsupervised anomaly detection has not been explored yet.

In this work, we propose a data-driven way of detecting anomalous trials of dementia patients using healthy controls as a frame of reference during an instruction-less eye-tracking test (Chapter 4). Under this setting, we define as anomalous those trials that deviate from healthy controls normal oculomotor behaviour. Thus anomalies could be related to cognitive processing of the stimuli, as well as, eye-tracking issues unavoidably inherited to the dementia group such as motion artefacts (due to increased head motion), poor calibration and irregular blinks. Given the instruction-less nature of the battery, we also consider the potential lack of engagement or attention of the patients to the stimuli as an anomaly which might be a valuable information for the diagnostic procedure and the experimental design of cognitive tests. The main contribution of this work is a novel unsupervised anomaly detection framework for eye-tracking data based on representation learning using convolutional autoencoders. This framework could be used to identify disease biomarkers, assess the quality of recordings, as well as, the efficacy of the experimental stimuli towards a specific target group. This work establishes a starting-point for getting further insights into eye movement abnormalities which is of greatest importance given the load of available data and the instruction-less nature of the tasks that render very difficult the prediction of anomalies even from experts.

5.2 Methods

To detect abnormal trials of dementia patients, the distribution of normal eye-tracking data was first learned on data of healthy individuals using an autoencoder

neural network (see section 2.9.0.3) that models the variety of healthy performance. We then performed the following experiments:

i. We firstly investigated the ability of the model to capture the distribution of eye-tracking data by closely reconstructing the input signal. Thus, we assessed the performance of the model on trials of healthy individuals extracted from the test set.

ii. We then examined the accuracy of our approach in detecting eye-tracking data of dementia patients. We calculated anomaly scores on the trial and patient level and reported the results.

iii. The accuracy was also compared between the different dementia groups and cognitive tasks.

iv. We provided more details in individual trial abnormalities by visualising trials of patients and controls with high and low anomaly scores.

5.2.1 Data

The materials and data used in this chapter are the same as presented in section 4.2.

5.2.1.1 Data preprocessing

The EyeLink system recorded gaze position and pupil size in a monocular tracking mode providing 1000 samples per second. Gaze position reports the (x, y) coordinates of a subject's gaze on the display (resolution: 1920 x 1080) in actual display coordinates (pixels) with origin (0, 0) at the top left. Raw samples, therefore, consist of three-time series of x, y coordinates of gaze and pupil size having a dimension of[sampling rate x trial duration].

Given the aim of this chapter is the identification of oculomotor abnormalities, only the x and y coordinates of gaze were included for further analysis. Additionally, we used velocity per sampling interval instead of position signal since we were interested in global abnormalities rather than stimulus-specific abnormalities; thus we believe that the patterns of velocity over time would be more informative than the specific displacement of gaze [186].

The velocity of gaze signal was firstly normalised by subtracting the mean value of each trial and dividing by its corresponding standard deviation. Missing values of velocity of gaze were imputed prior to normalisation with a constant zero value to avoid interpolation bias. Trials with total number of missing samples more than 1500ms were considered erroneous and were excluded from the analysis. Additionally, trials with normalised signal values more or less than 13 and -13, respectively, were excluded from the analysis because we found empirically that it helps with the convergence of the algorithm as extreme values correspond to incidents where the participant looks out of the screen resolution. Finally, each time series is scaled to

the $[0, 1]$ interval by applying a min-max transformation; subtracting from each time point the minimum value that appears in the training set and dividing by the difference between the maximum and the minimum value of the training set.

5.2.2 Proposed Pipeline

5.2.2.1 Anomaly Detection Problem

This work is based on an unsupervised approach for anomaly detection and the conceptual framework builds on work similar to [3]. We train our proposed convolutional network architecture in an unsupervised manner; training the model on normal samples and testing on both normal and abnormal ones. The input dataset D is split into train $Dtrn$ and test set $Dsts$, where $Dtrn = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ and $y_i = 0$ corresponds to the normal class. The test set $Dtst = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$, where y_i in $[0, 1]$, consists of normal (Controls B) and abnormal classes (patients data), so that the number of data from normal classes in the test set equals the abnormal ones. Therefore, we hypothesise that all patients data are anomalous. Data from battery controls A and the rest of controls B data that were not included in the test set were used for training.

The training objective of the model is to capture the distribution of $Dtrn$ which will enable the model to learn features that are representative of normal eye-tracking data. The hypothesis here is that an anomaly score based on the training objective would generate minimal anomaly scores for normal samples (since they have the same distribution as the training samples), but higher scores for abnormal samples.

5.2.2.2 Pipeline

Figure 5.1 shows the bow-tie architecture of the proposed approach, which comprises an encoder and a decoder network. The encoder network captures the distribution of the input data by mapping the 2-dimensional time series x (dimension: 2 x 3000; 2 channels of x and y velocity of gaze for 3 seconds) into a lower-dimensional latent representation z (dimension: 300 x 1). The encoder passes the input through five Convolutional and Batch Normalisation layers and one linear activation layer (dense layer) as well as ReLU activation functions and generates the latent representation z which carries a unique representation of the input. The linear layers were added to the architecture, so that the dimension of the latent representation could stay the same and can be easily compared with different manifestations of the network (e.g. with a variational autoencoder architecture). The decoder upsamples the latent representation z back to dimension of the original input time series and produces an output x' which is the reconstruction of the input.

The model is trained by minimising a reconstruction loss, enforcing the output of the decoder to be similar to the original input signal. Since the values of the time series lie in the $[0, 1]$ interval, the binary cross entropy as a measure of uncertainty between the training data and the model distribution, is the loss function used:

$Loss = -\sum_n x_n log(x'_n) + (1 - x_n) log(1 - x'_n)$, where x is the input, x' is the output

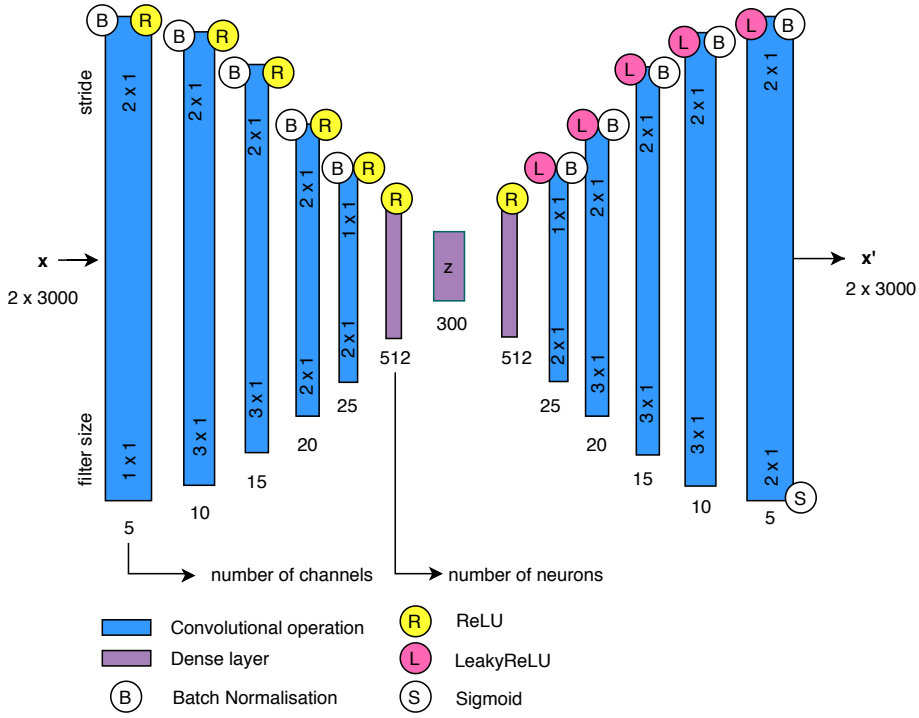

Figure 5.1: Details of the network architecture used.

of the network and n corresponds to each point of the input.

Implementation details The training objective is optimised via Adam optimiser with learning rate $lr = 10^{-3}$ and momentums $\beta_1 = 0.5, \beta_2 = 0.999$ and batch size 200. The model is trained for 20 epochs and we save the weights of the network during training that corresponds to the lowest reconstruction error in the test set.

5.2.2.3 Detection of anomalies

To find the anomalies during testing, the reconstruction error (anomaly score) for each sample in the test set is calculated. The anomaly scores for each test trial are then scaled, following the same procedure proposed in [3], within the probabilistic range of $[0, 1]$, by subtracting the calculated minimum anomaly score from the test set and dividing by the difference between the maximum and minimum anomaly score in the test set. The trial-level anomaly scores were converted to individual-level scores by calculating the average normalised anomaly score per individual.

The metric used for the the evaluation of model's performance is the area under the curve (AUC) of the receiver operating characteristic curve (ROC) which is a plot of true positive and false positive rates at various threshold values.

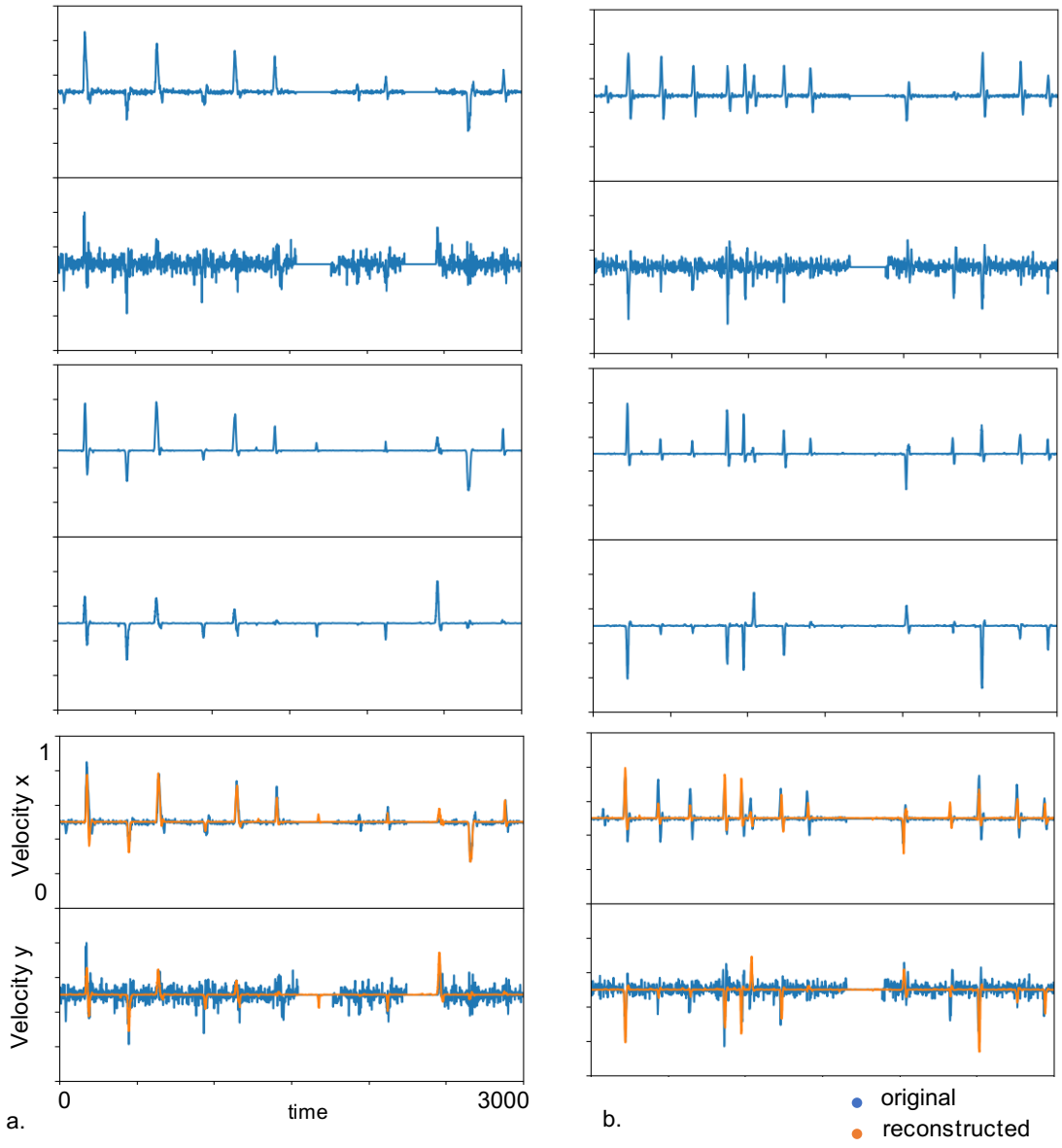

Figure 5.2: Example of two semantic processing trials (a, b) from two healthy participants in the test set. First block row: Real input time series of x and y velocity of gaze. Second block row: Corresponding eye-tracking data generated by the model. Third block row: Overlapped real and generated time series.

5.3 Results

Can the model generate realistic eye movement data? The model captures many properties of the signal and it generates a denoised version of the input time series (second block row in Figure 5.2). In Figure 5.2, the pairs of input and generated time series of the two case observations from healthy participants drawn from the test set suggest that the latent representation includes information about the location of saccades in time and the direction of saccades (e.g. negative and positive velocity), as well as the duration of fixations. The amplitude of saccades is also adequately

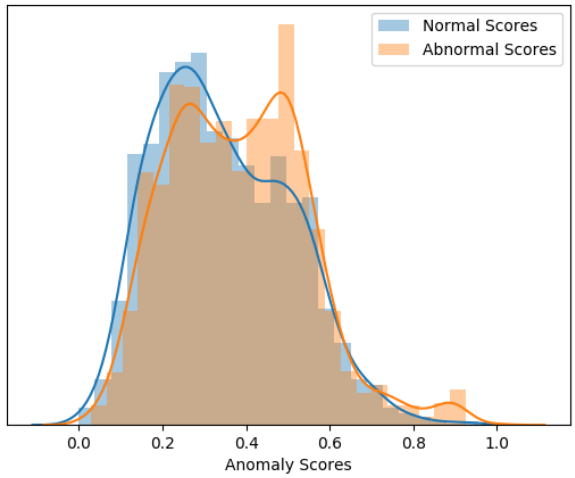

Figure 5.3: Distribution of the anomaly score evaluated on normal trials of the test set (blue), and on trials extracted from diseased cases (orange).

reconstructed. At least in these two trials the reconstruction precision is very high for velocity x and less for velocity y, as the latter has higher amount of noise in the signal. In some cases the residual values were higher between the real and generated signal, demonstrating the high variability and thus unexplained variance even in controls data. Overall, the model is able to capture the distribution of healthy eye-tracking data since it reconstructs quite precisely unseen time series trials from the test set.

Can the Model Detect Anomalies? The AUC score for trial and participant level anomaly detection based on the anomaly score described in Section 5.2.2.3 is 0.5650 and 0.652 respectively. In addition, the distribution of the anomaly scores of the normal and abnormal cases in the test set are presented in Figure 5.3. It is worthy mentioning that all trials from the dementia patients are considered abnormal and this might contribute to the little separability of the distributions of normal and abnormal samples.

Which dementia groups and cognitive tasks present the most abnormal eye movements? Figure 5.4 presents the AUC for the anomaly scores between the dementia and control group for the five different cognitive tasks of the battery. The semantic processing and social interaction task present the least abnormal cases (AUC: semantic processing; 0.4353, social interaction; 0.5297), whereas the discriminative power of the model is slightly higher for the missing items (0.6257), social scenes (0.6117) and recognition memory task (0.60091).

To further investigate oculomotor abnormalities within dementia subgroups, the anomaly scores were normalised for each cognitive task and AUC scores were computed to compare each dementia subgroup with the healthy controls group. The anomaly detection performance of the model on the trial and participant level, are summarised in Table 5.1 and 5.2. Based on our approach, the bvFTD group trials in the missing items, social scenes and recognition memory task, as well as the trials of the tAD group in the missing items task, seem to be the cases in which the most

Table 5.1: Trial level AUC scores on anomaly detection for each cognitive task and dementia group versus healthy controls.

Task	tAD	bvFTD	lvPPA	nfvPPA	svPPA
social interaction	0.533	0.599	0.493	0.533	0.484
missing items	0.670	0.711	0.589	0.596	0.565
social scenes	0.611	0.715	0.563	0.600	0.565
semantic processing	0.429	0.406	0.456	0.373	0.528
recognition memory	0.554	0.727	0.534	0.574	0.604

Table 5.2: Participant level AUC scores on anomaly detection for each cognitive task and dementia group versus healthy controls.

Task	tAD	bvFTD	lvPPA	nfvPPA	svPPA
social interaction	0.608	0.566	0.460	0.590	0.488
missing items	0.747	0.735	0.615	0.636	0.627
social scenes	0.687	0.752	0.587	0.655	0.597
semantic processing	0.359	0.389	0.439	0.276	0.524
recognition memory	0.590	0.804	0.513	0.607	0.636

abnormal eye movements are detected. Additionally, the AUC values are highest for svPPA patients on the semantic processing task. Combined with the bvFTD's having the highest AUC values for social scenes, there is a suggestion that this approach can yield cognitive profiling information which is consistent with established cognitive phenotypes of each dementia subtype.

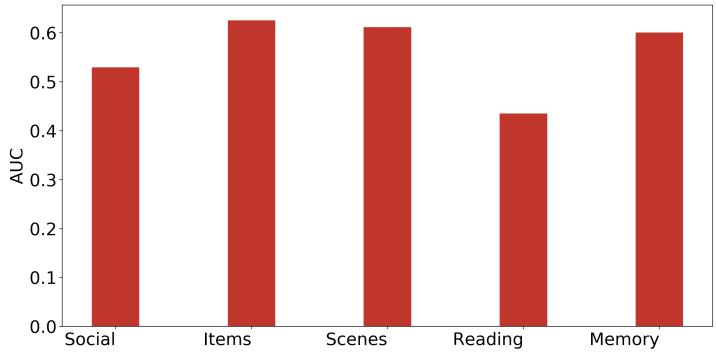

Figure 5.4: Barplots for the trial level AUC scores for each cognitive task in the computerised test: Social Interaction (social), Missing Items (items), Social Scenes (scenes), Semantic Processing (reading) and Recognition Memory (memory).

What do the most and least abnormal trials based on the model look like? To better understand the oculomotor abnormalities that our model is able to detect, we visualised individual trials from patients and controls for all five cognitive tasks of the test. Figure 5.5 shows the input data (normalised velocity of x and y coordinates of gaze) of patients with the highest anomaly score (relative to the data in the test set), controls with highest anomaly scores and controls with the lowest anomaly scores. The latter present velocity patterns with clear saccade (peaks) and fixation (flat areas) sequences with minimal noise. Compared to the least abnormal trials showing distinct and well defined saccades, the abnormal patients trials correspond to cases in which the patient does not engage or fixates in one location of the image (social interaction, social scenes), gets distracted and looks at something out of the stimuli (semantic processing), presents very noisy scanpaths (recognition memory) or there is a tracking error (saccades are not tracked) (missing items). The anomalies in the controls trials seem to be related to unusual saccadic patterns (social scenes, semantic processing), limited engagement suggestive of tracking error (social interaction, recognition memory) and noisy samples presumably associated with poor recording quality (missing items).

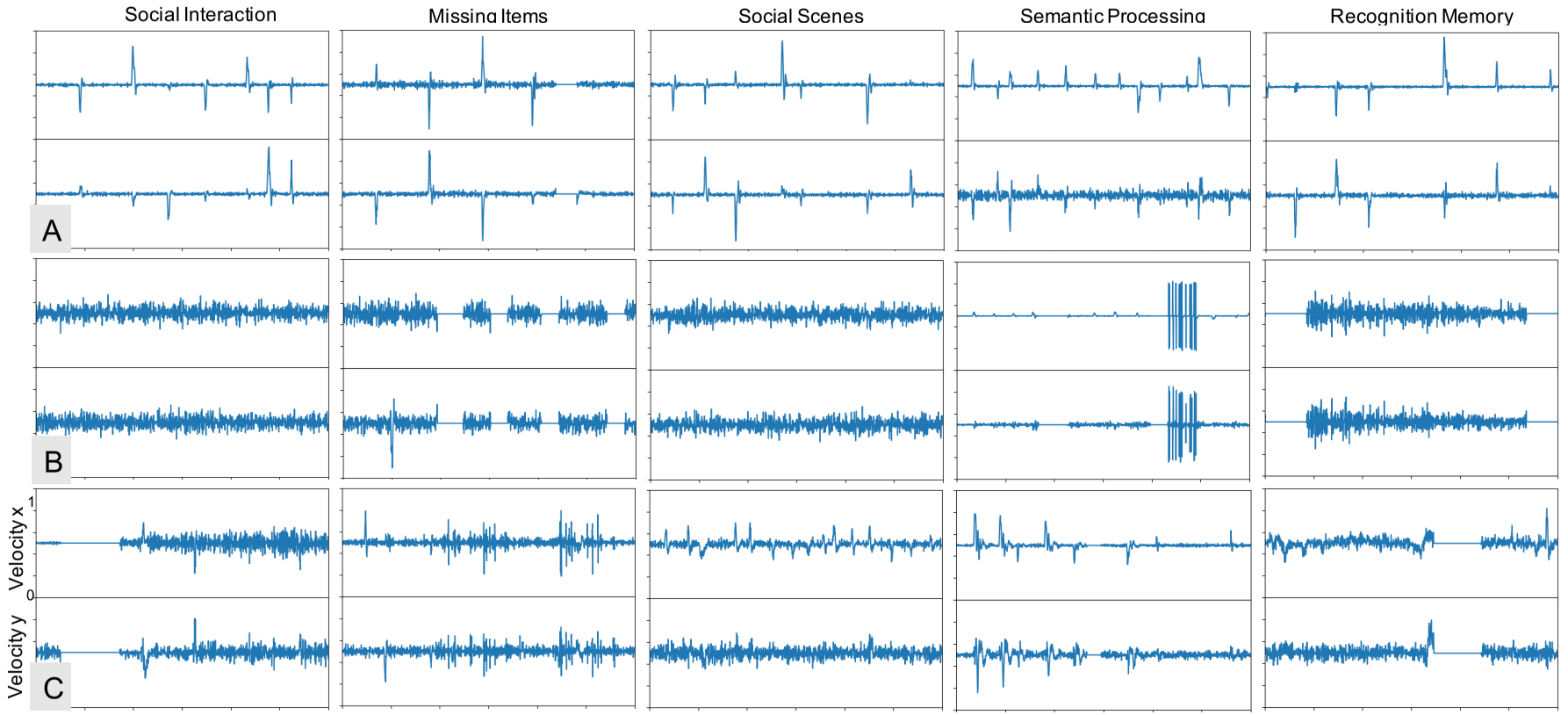

Figure 5.5: A: Healthy controls with the least abnormal scores: $(0, 0.038, 0.038, 0.355, 0.015)$. B: Dementia patients' trials with the highest anomaly scores (from left to right: nfvPPA $[0.906]$, tAD $[0.909]$, nfvPPA $[0.908]$, lvPPA $[1]$, tAD $[0.917]$). C: Healthy controls with the highest anomaly scores: $(0.743, 0.778, 0.687, 0.9558, 0.722)$.

5.4 Discussion

In this chapter we propose an application of a convolutional autoencoder neural network as a framework to assess deviations from normative behaviour based on eye-tracking data. We investigated whether we could use this framework to identify outliers in eye-tracking trials of dementia patients and healthy controls participating in a free-viewing test which was designed to assess particular cognitive functions that are impaired to varying extents in different subtypes of dementia.

We explored a data driven-way of detecting abnormal trials by creating a normative reference of the healthy eye-tracking appearance and comparing the patients' data to it. Normative eye-tracking data were created by a neural network with a convolutional autoeconder architecture which was trained to learn the distribution of normal eye-tracking data based on the data from healthy cases. Deviations of the patients and controls data held in the test set from the normal distribution, or in other words the anomaly scores, were estimated by calculating the reconstruction error of the trained model on the test data. The values were then scaled to the probabilistic interval [0, 1] to correspond to the probability of a trial being abnormal and the AUC score was computed.

Although the model seems to encode a rich representation of the original signal in the latent space, the AUC metric for the dementia versus controls case is 0.5650 and 0.652, for trial and participant level anomaly detection. This is not unexpected as we anticipate not all dementia trials to be abnormal and thus we are not necessarily interested in the model performance.

Our method demonstrates that fewer anomalies are present in the semantic processing task. This might be due to the specific type of anomalies that our approach is trained to detect. From the visualisations of individual trials and since our method is not task or trial specific, the anomalies detected seem to be global rather than local patterns, such as frozen gaze, error of the device or distractions. Since the semantic processing task involves participants voluntarily reading sentences, the participants might be more engaged to this task compared to the scene perception tasks. This is because the nature of the semantic processing task yields a much more predictable set of fixations and saccades (reading left to right along a fixed y value) compared to the other tasks (free viewing scenes).

Additionally, our method is particularly sensitive to the bvFTD group in the recognition memory and social scenes tasks, demonstrating that these tasks' stimuli modulate participants eye movements during scene perception. Patients with bvFTD are characterised by early behavioural disinhibition, apathy, loss of empathy and preservative compulsive behaviour. There is therefore a suggestion that their eye movements on these tasks reflect cognitive profiling information which is consistent with established cognitive phenotypes for bvFTD.

This work establishes a starting-point for getting further insights into eye movement abnormalities in dementia patients. This is of greatest importance because the load of available data and the instruction-less nature of the tasks render the identification of anomalies very difficult (even from experts). Future directions could involve

the investigation of different types of abnormalities which are trial or task specific by incorporating this information in the latent space. Comparisons of the autoencoder architecture with variational autoencoders or Generative Adversarial Networks for anomaly detection might also show whether richer representations can be encoded. Another interesting direction might be the application of clustering algorithms in the latent space features to identify potential clusters of anomalies. The purity of the clusters (dementia vs controls) might indicate whether detected anomalies are disease specific and not due to a device error. Finally, future work could involve the applications of our anomaly detection pipeline to other eye-tracking large datasets and also the development of new cognitive batteries driven by the findings presented here (e.g. stimuli which generate more predictable gaze pathways).

To conclude, this work presents an alternative approach for oculomotor biomarker discovery; instead of building models to identify properties of the data that discriminate between the dementia and controls group, we address the problem of measuring departures from a distribution of typical oculomotor patterns during instruction-less eye-tracking tests. The biggest challenge in this context compared to other medical problems (e.g. MRI tutor anomaly detection) is the lack of labels determining abnormal trials within patients and controls data. However, our visualisations of abnormal cases improve our understanding of the oculomotor anomalies in dementia. Our approach indicates that limited engagement and distractions during free-viewing might be markers of the disease that can be detected in this short instruction-less test.

Chapter 6

Conclusions

In this thesis I have presented my work on the application of machine learning methods for the detection of dementia oculomotor biomarkers from eye-tracking based assessments. In this chapter I will present a summary of the thesis and future research directions for the potential application of the techniques used here to similar neuropsychological problems and also suggestions for the further improvement of the techniques themselves.

6.1 Thesis Summary

The aim of this thesis was to investigate computational approaches for the analysis of eye-tracking data of dementia patients and healthy controls under naturalistic and less constrained scenarios to identify novel digital oculomotor biomarkers. Towards this goal, this thesis brings together expertise from the disciplines of dementia research, cognitive neuropsychology, biosensors and machine learning.

Early detection of the disease, before the quality of life of the patients has already been deteriorated significantly, could revolutionise the healthcare system; reducing healthcare costs and suffering of individuals with dementia and their carers. From the neuropsychological perspective, although the defining characteristics of most dementia syndromes are primarily cognitive and behavioural in nature, the current cognitive assessment tools do not capture adequately and precisely different aspects of cognitive function. Technological innovation might support substantial improvements in techniques and devices that analyse and acquire cognitive data. Given the recent success in the area of machine learning, this thesis attempts to identify novel cognitive tests based on oculomotor measurements using computational techniques suitable for neuropsychological problems.

To summarise, the computational principals and techniques that have been investigated in this thesis as suitable for neuropsychological problems are the following:

- Anomaly detection (Chapter 5)
- Deep neural networks (Chapter 4, 5)

- Data augmentation (Chapter 4)
- Unsupervised learning with auto-encoders (Chapter 5)
- Layer-wise Relevant Propagation for feature visualisation (Chapter 4)
- Self-supervised learning/ transfer learning (Chapter 4)
- Fusion of multi-modal data (videos with eye movement time series) (Chapter 3)

Chapter 3 investigates the potential of identifying oculomotor biomarkers during activities of daily living. In particular, the effect of saliency of environmental features on oculomotor dynamics in patients with posterior cortical atrophy and Alzheimer's disease is explored to investigate the extent to which physical environment mitigates dementia functional impairment. I combined eye movement and egocentric videos of participants performing a real-world visual search task navigating in a controlled environment. I extracted two saliency based features that could also be generalised in other naturalistic experiments. I found that although both patient groups were slower in reaching their target destinations, there is no evidence of a strong relationship between saliency and fixation or completion time. In comparison, the findings by Shakespeare et al. [155] suggest the influence of conspicuous, visually salient features of static scenes on fixation of PCA patients. To conclude, it is difficult to infer, from this study only, whether bottom-up saliency is the dominant factor that drives visual search during navigation in AD and PCA. A mobile eye-tracker with higher frequency is recommended to be used in future investigations to investigate the full spectrum of oculomotor patterns.

Chapter 4 explores the extent to which eye-tracking metrics capture dementia related oculomotor deficits during a novel instruction-less eye-tracking cognitive test. To address the limitations of the previous chapter, a high frequency (1000 Hz) eye-tracker was used in this study under a more-constrained but ecological setting (naturalistic images) with analytic approaches accommodating complex time series data. This chapter also introduces a novel method for extracting features from instruction-less eye-tracking cognitive tests based on self-supervised representation learning and a technique for interpreting the network's decisions in differentiating between the distinct cognitive activities.

I found that dementia patients search less extensively and scan the stimuli significantly more slowly than controls. They also present a tendency to fixate towards the left side of the screen during sentence presentation compared to controls, which might indicate that either are slower in reading or they are not reading the sentences. This study provides quantitative evidence that eye-tracking metrics reflect dementia-related oculomotor deficits during processing of complex visual stimuli even under the lack of any instructions given to the participants. I also found that self-supervised learning features are more sensitive than handcrafted features in detecting performance differences between participants with and without dementia across a variety of tasks. This work highlights the contribution of self-supervised representation learning techniques in biomedical applications where the small number of patients, the

non-homogenous presentations of the disease and the complexity of the setting can be a challenge using state-of-the-art feature extraction methods.

Chapter 5 proposes anomaly detection as a framework to assess deviations of patients oculomotor behaviour from healthy controls. This work presents an alternative approach for oculomotor biomarker discovery; instead of building models to identify properties of the data that discriminate between the dementia and controls group, I address the problem of measuring departures from a distribution of typical oculomotor patterns during instruction-less eye-tracking tests. The focus of this chapter is exploratory, providing visualisations and interpretations of abnormal patterns of eye movements of dementia patients and healthy controls. This work establishes a starting-point for getting further insights into eye movement abnormalities relative to normative data using cutting-edge computational techniques.

6.2 Future Directions

Standardised paper-and-pencil cognitive assessment tools are a key component of the screening and diagnostic process of dementia patients, but have a number of limitations. The work presenting in thesis introduces a more ecological valid assessment of natural cognitive behaviour in dementia. It develops an instruction-less paradigm for the assessment of multiple domains of cognition. This paradigm provides an alternative to lengthy batteries of individual tests that have different task demands, properties and instructions. There are a number of specific ways in which the research presented in this thesis could be taken forward to improve our knowledge on oculomotor biomarkers in dementia and potentially be translated in a clinical or standard research tool. These future advances can be broken down into a. improving data acquisition techniques, and b. data analysis methods, c. running large scale tests for validation and performance assessment and d. considering instruction-less tests alongside the battery of other clinical examinations.

6.2.1 Improving data acquisition techniques

The research presented in this thesis apart from improving analytical techniques suitable for detecting hidden patterns in large eye-tracking datasets, could also guide the future design of future cognitive tests involving eye-tracking. This could include suggestions about the selection of stimuli given a specific population but also the overall procedure of data collection during cognitive testing. For instance, as far as anomaly detection is concerned, the more predictable and invariant the pattern of normal behaviour elicited by a set of stimuli, the easier to detect gross anomalies. From the various stimuli used in this thesis, reading tasks evoking a specific left-to-right pattern of fixations and saccades seem to a stimuli recommended for future investigations since it generates more predictable patterns. Future work could focus on the appropriate selection of the duration of stimuli evaluating whether the abnormal behaviour occurs in the beginning of the trials and therefore avoiding gathering a lot

of data with limited discriminative power.

6.2.2 Improving data analysis methods

Integrating stimulus properties with physiological data

One particularly interesting possibility is a novel way of feature generation from neuropsychological tests integrating stimulus properties with physiological data. In particular, in the instruction-less eye-tracking battery, the response characteristics (x, y and pupil size over time) could be augmented with two time series: the bottom-up saliency and the semantics of gaze location over time. The bottom-up saliency time series could be computed in ways similar to those in Chapter 3; calculating saliency maps of the input image and then creating a time series of the value of saliency for each gaze location over time. The semantics of eye movement over time could be implemented by applying first a semantic segmentation algorithm to images (e.g. [190]) that labels each pixel of the stimulus images with a corresponding class. A categorical time series of objects that the eyes looked at could be generated in this way. Representation learning methods such the self-supervised learning or unsupervised learning algorithms used in this thesis could be applied to the oculomotor, saliency and semantics time series to extract features and then comparisons could be made to investigate the source of abnormal behaviour of dementia patients (bottom-up, top-down and oculomotor mechanisms).

Improving anomaly detection with clustering and generative modelling techniques

Another interesting possibility that presents itself from Chapter 5, is the continuation of the unsupervised representation learning of time series comparing different deep learning for anomaly detection frameworks including variational autoencoders and Generative Adversarial Networks (GAN) (e.g. [3, 133]) that have shown better performance in anomaly detection problems when applied in image datasets. These comparisons will shed light into which neural network architecture is the most suitable for learning the distribution of normal behaviour of eye movements when free-viewing at images. In particular, the skip-GANomaly model which employs an encoder-decoder convolutional neural network with skip connections and an adversarial training scheme that given the original input and the output of the autoencoder discriminates between the real and the fake input, has potential in our application. We believe that adapting the discriminator from a binary (fake vs real) to a multi-class classifier (discriminating between the different cognitive tasks activities for real data vs fake data) could help regularise better the network. Additionally, apart from detecting anomalies in patient groups, the anomaly detection framework could be used in control populations for data cleaning of eye-tracking data to improve data quality by removing for further analysis participants with unknown oculomotor problems or trials in which the sensor has tracking or calibration problems. Another potential research direction that comes from this line of work, is the application of clustering methods

on the latent space representations generated by the networks to identify patterns of abnormality potentially grouped by stimuli, task or dementia subtype.

Incorporating multi-modal data with bayesian optimisation techniques

Another future direction might be the potential use of bayesian optimisation in neuropsychological tasks and cognitively-relevant physiological data (e.g., pupil size, electrodermal activity, heart rate) to provide a novel way of assessing cognitive function at the individual level [112]. Bayesian optimisation is a sequential method for global parameter optimisation, using priors to turn objective functions into random functions. The space of random functions is then sampled, and the resulting acquisition function used to update the prior space and determine the next sampling point, eventually converging on the optimal point in space. Using this paradigm, highly multivariate multi-modal data can be analysed to examine questions such as: a) which biological variables optimally predict cognitive performance (mechanism identification), b) in which measures an individual differs most from the norm (personalised outlier detection), c) how real-time adaptations can be made to neuropsychological tests to optimally elicit specific cognitive functions (the Automated Neuropsychologist).

6.2.3 Running large scale tests for validation and performance assessment

Future investigation regarding whether the instruction-less eye-tracking battery is clinically ready and potential improvements to be done might be of interest. Next stages could involve the recruitment of a larger within-subtype dementia cohort which will facilitate the evaluation of potential correlation of eye-tracking metrics from specific cognitive tasks and the different dementia subtypes. Future evaluation of patients in the early stages of dementia, with mild cognitive impairment or preclinical AD, might determine whether this battery can be used for early detection of cognitive impairment from eye movements.

6.2.4 Comparing with other clinical examinations

Once the stimuli and the methods used to analyse the data are optimised and the instruction-less eye-tracking test is evaluated in large scale datasets, comparisons with other clinical batteries are required before translating it into a research or clinical tool. A comparison of this computerised test's performance with results on established paper-and-pencil neuropsychological tests and cognitive screening batteries (e.g. MMSE) will indicate the actual value of the cognitive battery.

Bibliography

[1] Charu C Aggarwal. An introduction to outlier analysis. In *Outlier analysis*, pages 1–34. Springer, 2017.

[2] R M Ahmed, R W Paterson, J D Warren, H Zetterberg, J T O'Brien, N C Fox, G M Halliday, and J M Schott. Biomarkers in dementia: Clinical utility and new directions. *Journal of Neurology, Neurosurgery and Psychiatry*, 85(12):1426–1434, 2014.

[3] Samet Akçay, Amir Atapour-Abarghouei, and Toby P Breckon. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[4] Tim J Anderson and Michael R Macaskill. Eye movements in patients with neurodegenerative disorders. *Nature Reviews Neurology*, 9(2):74–85, 2013.

[5] Alzheimer Association. 2015 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, 11(3):332–384, 2015.

[6] Sebastian Bach, Alexander Binder, Grégoire Montavon, and Frederick Klauschen. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PloS one*, 10(7):e0130140, 2015.

[7] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. 2018.

[8] Hadi Banaee, Mobyen Uddin Ahmed, and Amy Loutfi. Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges. *Sensors*, 13(12):17472–17500, 2013.

[9] Simon Baron-Cohen, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb. The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry*, 42(2):241–251, 2001.

[10] Maura Bellio, Neil P Oxtoby, Zuzana Walker, Susie Henley, Annemie Ribbens, Ann Blandford, Daniel C Alexander, and Keir X X Yong. Analyzing large Alzheimer's disease cognitive datasets: Considerations and challenges. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 12(1):e12135, 2020.

[11] Jessica Beltrán, Mireya S García-Vázquez, Jenny Benois-Pineau, Luis Miguel Gutierrez-Robledo, and Jean François Dartigues. Computational Techniques for Eye Movements Analysis towards Supporting Early Diagnosis of Alzheimer's Disease: A Review. *Computational and Mathematical Methods in Medicine*, 2018, 2018.

[12] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning : A Review and New Perspectives. 35(8):1798–1828, 2013.

[13] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[14] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. Detecting personality traits using eye-tracking data. *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–12, 2019.

[15] Lilah M Besser, Walter A Kukull, Merilee A Teylan, Eileen H Bigio, Nigel J Cairns, Julia K Kofler, Thomas J Montine, Julie A Schneider, and Peter T Nelson. The revised National Alzheimer's Coordinating Center's Neuropathology Form—available data and new analyses. *Journal of Neuropathology & Experimental Neurology*, 77(8):717–726, 2018.

[16] Nilavra Bhattacharya, Somnath Rakshit, Jacek Gwizdka, and Paul Kogut. Relevance Prediction from Eye-movements Using Semi-interpretable Convolutional Neural Networks. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 223–233, 2020.

[17] Juan Biondi, Gerardo Fernandez, Silvia Castro, and Osvaldo Agamennoni. Eye-Movement behavior identification for AD diagnosis. pages 1–11, 2017.

[18] Monika Biskupska. Bottom-up saliency maps – a review. *Elektronika: konstrukcje, technologie, zastosowania*, 54(7):53–57, 2013.

[19] Daniel Borja-Cacho and Jeffrey Matthews. The logopenic variant of primary progressive aphasia. *Current opinion in neurology*, 6(23):633–637, 2010.

[20] Neil D B Bruce, John K Tsotsos, and John K Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):5–5, 2009.

[21] A P A Bueno, J R Sato, and M Hornberger. Eye tracking – The overlooked method to measure cognition in neurodegeneration? *Neuropsychologia*, 133:107191, 2019.

[22] Andreas Bulling, Student Member, Jamie A Ward, Hans Gellersen, and Gerhard Tr. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE transactions on pattern analysis and machine intelligence*, 33(4):741–753, 2010.

[23] Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Tröster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2011.

[24] Robyn M. Busch and Jessica Smerz Chapin. Review of normative data for common screening measures used to evaluate cognitive functioning in elderly individuals. *Clinical Neuropsychologist*, 22(4):620–650, 2008.

[25] Frederick W. Bylsma, D. Xeno Rasmusson, George W. Rebok, Penelope M. Keyl, Larry Tune, and Jason Brandt. Changes in visual fixation and saccadic eye movements in Alzheimer's disease. *International Journal of Psychophysiology*, 19(1):33–40, 1995.

[26] David Caduff and Sabine Timpf. On the assessment of landmark salience for human navigation. *Cognitive Processing*, 9(4):249–267, 2008.

[27] Mikel Canizo, Isaac Triguero, Angel Conde, and Enrique Onieva. Multi-head CNN – RNN for multi-time series anomaly detection : An industrial case study. *Neurocomputing*, 363:246–260, 2019.

[28] Moran Cerf, Neural Systems, Jonathan Harel, Christof Koch, and Wolfgang Einh. Predicting human gaze using low-level saliency combined with face detection. *Nips2007*, pages 1–8, 2007.

[29] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[30] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[31] David Charte, Francisco Charte, María J. del Jesus, and Francisco Herrera. An analysis on the use of autoencoders for representation learning: Fundamentals, learning task case studies, explainability and challenges. *Neurocomputing*, 404:93–107, 2020.

[32] Yi Chen Chiu, Donna Algase, Ann Whall, Jersey Liang, Hsiu Chih Liu, Ker Neng Lin, and Pei Ning Wang. Getting lost: Directed attention and executive functions in early Alzheimer's disease patients. *Dementia and Geriatric Cognitive Disorders*, 17(3):174–180, 2004.

[33] Jongyoon Choi, Beena Ahmed, and Ricardo Gutierrez-Osuna. Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 16(2):279–286, 2012.

[34] Edward J Ciaccio, Steven M Dunn, and Metin Akay. Biosignal pattern recognition and interpretation systems. 2. Methods for feature extraction and selection. *IEEE Engineering in Medicine and Biology Magazine*, 12(4):106–113, 1993.

[35] Charles E Connor, Howard E Egeth, and Steven Yantis. Visual attention: Bottom-up versus top-down. *Current Biology*, 14(19):850–852, 2004.

[36] C Cortes and V Vapnik. Support-Vector Networks. 20:273–297, 1995.

[37] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[38] Paul K Crane, Emily Trittschuh, Shubhabrata Mukherjee, Andrew J Saykin, R Elizabeth Sanders, Eric B Larson, Susan M McCurry, Wayne McCormick, James D Bowen, Thomas Grabowski, and Others. Incidence of cognitively defined late-onset Alzheimer's dementia subgroups from a prospective cohort study. *Alzheimer's & Dementia*, 13(12):1307–1316, 2017.

[39] Trevor J. Crawford, Alex Devereaux, Steve Higham, and Claire Kelly. The disengagement of visual attention in Alzheimer's disease: A longitudinal eye-tracking study. *Frontiers in Aging Neuroscience*, 7(JUN):1–10, 2015.

[40] Sam T Creavin, Susanna Wisniewski, Anna H Noel-Storr, Clare M Trevelyan, Thomas Hampton, Dane Rayment, Victoria M Thom, Kirsty J E Nash, Hosam Elhamoui, Rowena Milligan, Anish S Patel, Demitra V Tsivos, Tracey Wing, Emma Phillips, Sophie M Kellman, Hannah L Shackleton, Georgina F Singleton, Bethany E Neale, Martha E Watton, and Sarah Cullum. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews*, (1), 2016.

[41] Sebastian J. Crutch, Manja Lehmann, Jonathan M. Schott, Gil D. Rabinovici, Martin N. Rossor, and Nick C. Fox. Posterior cortical atrophy. *The Lancet Neurology*, 11(2):170–178, 2012.

[42] Sebastian J Crutch, Jonathan M Schott, Gil D Rabinovici, Bradley F Boeve, Stefano F Cappa, Bradford C Dickerson, Bruno Dubois, Neill R Graff-Radford, Pierre Krolak-Salmon, Manja Lehmann, and Others. Shining a light on posterior cortical atrophy. *Alzheimer's & Dementia*, 9(4):463–465, 2013.

[43] Sebastian J Crutch, Keir X X Yong, and Timothy J Shakespeare. Looking but Not Seeing: Recent Perspectives on Posterior Cortical Atrophy. *Current Directions in Psychological Science*, 25(4):251–260, 2016.

[44] Michael D. Crutcher, Rose Calhoun-Haney, Cecelia M. Manzanares, James J. Lah, Allan I. Levey, and Stuart M. Zola. Eye tracking during a visual paired comparison task as a predictor of early dementia. *American Journal of Alzheimer's Disease and other Dementias*, 24(3):258–266, 2009.

[45] R. C. Petersen D. F. Tang-Wai, N. R. Graff-Radford, B. F. Boeve, D. W. Dickson, J. E. Parisi, R. Crook, R. J. Caselli, D. S. Knopman. Clinical , genetic , and neuropathologic. *Neurology*, 63(7):1168–1174, 2004.

[46] Kirsten A. Dalrymple, Ming Jiang, Qi Zhao, and Jed T. Elison. Machine learning accurately classifies age of toddlers based on eye tracking. *Scientific Reports*, 9(1):1–10, 2019.

[47] Rebecca Davis and Jennifer Ohman. Wayfinding in ageing and Alzheimer's disease within a virtual senior residence: study protocol. *Journal of Advanced Nursing*, 72(7):1677–1688, 2016.

[48] Marjolein E De Vugt and Frans R J Verhey. The impact of early dementia diagnosis and intervention on informal caregivers. *Progress in Neurobiology*, 110:54–62, 2013.

[49] Gayle DeDe and Denis Kelleher. Effects of animacy and sentence type on silent reading comprehension in aphasia: An eye-tracking study. *Journal of Neurolinguistics*, 57:100950, 2021.

[50] A R DeIpolyi, K P Rankin, Lennart Mucke, B L Miller, and M L Gorno-Tempini. Spatial cognition and the human navigation network in AD and MCI. *Neurology*, 69(10):986–997, 2007.

[51] Sergio Della Sala and Roberto Cubelli. Alleged" sonic attack" supported by poor neuropsychology. *Cortex; a journal devoted to the study of the nervous system and behavior*, 103:387–388, 2018.

[52] Pedro Domingos. A few useful things to know about machine learning, 2012.

[53] Michelle C. Dragan, Timothy K. Leonard, Andres M. Lozano, Mary Pat McAndrews, Karen Ng, Jennifer D. Ryan, David F. Tang-Wai, Jordana S. Wynn, and Kari L. Hoffman. Pupillary responses and memory-guided visual search reveal age-related and Alzheimer's-related memory decline. *Behavioural Brain Research*, 322:351–361, 2017.

[54] Shahram Eivazi and Roman Bednarik. Predicting Problem-solving Behavior and Performance Levels from Visual Attention Data. *Proc. of the Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI*, pages 9–16, 2011.

[55] Joseph F Fagan III. Memory in the infant. *Journal of experimental child psychology*, 9(2):217–226, 1970.

[56] Andreia V. Faria, David Race, Kevin Kim, and Argye E. Hillis. The eyes reveal uncertainty about object distinctions in semantic variant primary progressive aphasia. *Cortex*, 103:372–381, 2018.

[57] Andreia V Faria, David Race, Kevin Kim, and Argye E Hillis. The eyes reveal uncertainty about object distinctions in semantic variant primary progressive aphasia. *Cortex*, 103:372–381, 2018.

[58] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U. Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, 2018.

[59] Oliver Faust, Yuki Hagiwara, Tan Jen, Oh Shu, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals : A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, 2018.

[60] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54, 1996.

[61] Gerardo Fernández, Liliana R. Castro, Marcela Schumacher, and Osvaldo E. Agamennoni. Diagnosis of mild Alzheimer disease through the analysis of eye movements during reading. *Journal of Integrative Neuroscience*, 14(1):121–133, 2015.

[62] Gerardo Fernández, Facundo Manes, Nora P Rotstein, Oscar Colombo, Pablo Mandolesi, Luis E Politi, and Osvaldo Agamennoni. Lack of contextual-word predictability during reading in patients with mild Alzheimer disease. *Neuropsychologia*, 62(1):143–151, 2014.

[63] Gerardo Fernández, Marcela Schumacher, Liliana Castro, David Orozco, and Osvaldo Agamennoni. Patients with mild Alzheimer's disease produced shorter outgoing saccades when reading sentences. *Psychiatry Research*, 229(1-2):470–478, 2015.

[64] Richard Fleming and Nitin Purandare. Long-term care for people with dementia: Environmental design guidelines. *International Psychogeriatrics*, 22(7):1084–1095, 2010.

[65] Phillip D Fletcher, Jennifer M Nicholas, Timothy J Shakespeare, Laura E Downey, Hannah L Golden, Jennifer L Agustus, Camilla N Clark, Catherine J Mummery, Jonathan M Schott, Sebastian J Crutch, and Jason D Warren. Dementias show differential physiological responses to salient sounds. *Frontiers in Behavioral Neuroscience*, 9, 2015.

[66] William A. Fletcher and James A. Sharpe. Saccadic eye movement dysfunction in Alzheimer's disease. *Annals of Neurology*, 20(4):464–471, 1986.

[67] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

[68] Tom Foulsham, Jason J S Barton, Alan Kingstone, Richard Dewhurst, and Geoffrey Underwood. Fixation and saliency during search of natural scenes: The case of visual agnosia. *Neuropsychologia*, 47(8-9):1994–2003, 2009.

[69] Tom Foulsham, Jason J S Barton, Alan Kingstone, Richard Dewhurst, and Geoffrey Underwood. Modeling eye movements in visual agnosia with a saliency map approach: Bottom-up guidance or top-down strategy? *Neural Networks*, 24(6):665–677, 2011.

[70] Tom Foulsham and Geoffrey Underwood. How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, 36(8):1123–1138, 2007.

[71] Christos A Frantzidis, Charalampos Bratsas, Manousos A Klados, Evdokimos Konstantinidis, Chrysa D Lithari, Ana B Vivas, Christos L Papadelis, Eleni Kaldoudi, Costas Pappas, and Panagiotis D Bamidis. On the classification of emotional biosignals evoked while viewing affective pictures: An integrated data-mining-based approach for healthcare applications. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):309–318, 2010.

[72] Siobhan Garbutt, Alisa Matlin, Joanna Hellmuth, Ana K. Schenk, Julene K. Johnson, Howard Rosen, David Dean, Joel Kramer, John Neuhaus, Bruce L. Miller, Stephen G. Lisberger, and Adam L. Boxer. Oculomotor function in frontotemporal lobar degeneration, related disorders and Alzheimer's disease. *Brain*, 131(5):1268–1281, 2008.

[73] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[74] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.

[75] Caroline Giulioli and Helene Amieva. Epidemiology of Cognitive Aging in the Oldest Old. *Revista de investigacion clinica; organo del Hospital de Enfermedades de la Nutricion*, 68(1):33–39, 2016.

[76] Jonas Goltz, Michael Grossberg, and Ronak Etemadpour. Exploring simple neural network architectures for eye movement classification. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–5, 2019.

[77] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[78] P. E. Hallett. Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18(10):1279–1296, 1978.

[79] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.

[80] J. Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer. *Eye tracking: a comprehensive guide to methods and measures*. 2011.

[81] Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience*, 12:105, 2018.

[82] Xiaodi Hou and Liqing Zhang. Dynamic visual attention: Searching for coding length increments. *Advances in neural information processing systems*, 21(800):681–688, 2008.

[83] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. *Proceedings of the IEEE International Conference on Computer Vision*, pages 262–270, 2016.

[84] Alan E Hubbard, Jennifer Ahern, Nancy L Fleischer, Mark der Laan, Sheri A Satariano, Nicholas Jewell, Tim Bruckner, and William A Satariano. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, pages 467–474, 2010.

[85] Rosalind Hutchings, Romina Palermo, Jason Bruggemann, John R Hodges, Olivier Piguet, and Fiona Kumfor. Looking but not seeing: Increased eye fixations in behavioural-variant frontotemporal dementia. *Cortex*, 103:71–81, 2018.

[86] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[87] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[88] Kurt A. Jellinger. Should the word 'dementia' be forgotten? *Journal of Cellular and Molecular Medicine*, 14(10):2415–2416, 2010.

[89] Ming Jiang and Qi Zhao. Learning visual attention to identify people with autism spectrum disorder. In *Proceedings of the ieee international conference on computer vision*, pages 3267–3276, 2017.

[90] A. Jones, R. P. Friedland, B. Koss, L. Stark, and B. A. Thompkins-Ober. Saccadic intrusions in Alzheimer-type dementia. *Journal of Neurology*, 229(3):189–194, 1983.

[91] Alexandros Kafkas and Daniela Montaldi. The pupillary response discriminates between subjective and objective familiarity and novelty. 52:1305–1316, 2015.

[92] Liam D. Kaufman, Jay Pratt, Brian Levine, and Sandra E. Black. Executive deficits detected in mild Alzheimer's disease using the antisaccade task. *Brain and Behavior*, 2(1):15–21, 2012.

[93] Peter Kiefer, Ioannis Giannopoulos, Martin Raubal, and Andrew Duchowski. Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition and Computation*, 17(1-2):1–19, 2017.

[94] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[95] Ahmad F. Klaib, Nawaf O. Alsrehin, Wasen Y. Melhem, Haneen O. Bashtawi, and Aws A. Magableh. Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. *Expert Systems with Applications*, 166(September 2020):114037, 2021.

[96] Jessica Knilans and Gayle DeDe. Online sentence reading in people with aphasia: Evidence from eye tracking. *American journal of speech-language pathology*, 24(4):S961—-S973, 2015.

[97] Christof Koch and Shimon Ullman. Shifts in selective visual attention:Towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227, 1985.

[98] Katja Komossa, Timo Grimmer, Janine Diehl, and Alexander Kurz. Mapping Scores Onto Stages : Mini-Mental State Examination and Clinical Dementia Rating. *American Journal of Geriatric Psychiatry*, 14(2):139–144, 2006.

[99] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, 2016.

[100] Fiona Kumfor and Olivier Piguet. Disturbance of emotion processing in frontotemporal dementia: A synthesis of cognitive and neuroimaging findings. *Neuropsychology Review*, 22(3):280–297, 2012.

[101] Dmitry Lagun, Cecelia Manzanares, Stuart M Zola, Elizabeth A Buffalo, and Eugene Agichtein. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of Neuroscience Methods*, 201(1):196–203, 2011.

[102] Martin Längkvist, Lars Karlsson, and Amy Loutfi. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42(1):11–24, 2014.

[103] B Y Kung-yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[104] Andy Liaw and M Wiener. Classification and Regression by randomForest. *R news*, 2(December):18–22, 2002.

[105] Jeremiah K H Lim, Qiao Xin Li, Zheng He, Algis J Vingrys, Vickie H Y Wong, Nicolas Currier, Jamie Mullen, Bang V Bui, and Christine T O Nguyen. The eye as a biomarker for Alzheimer's disease. *Frontiers in Neuroscience*, 10:536, 2016.

[106] Chien Hung Liu, Po Yin Chang, and Chun Yuan Huang. Using eye-tracking and support vector machine to measure learning attention in elearning. In *Applied Mechanics and Materials*, volume 311, pages 9–14. Trans Tech Publ, 2013.

[107] David W Loring, K J Meador, Roderick K Mahurin, and John W Largen. Neuropsychological performance in dementia of the Alzheimer type and multi-infarct dementia. *Archives of Clinical Neuropsychology*, 1(4):335–340, 1986.

[108] Ya Lou, Yanping Liu, Johanna K. Kaakinen, and Xingshan Li. Using support vector machines to identify literacy skills: Evidence from eye movements. *Behavior Research Methods*, 49(3):887–895, 2017.

[109] Florence Mahieux, Gilles Fénelon, Antoine Flahault, Marie-José Manifacier, Denyse Michelet, and François Boller. Neuropsychological prediction of dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 64(2):178–183, 1998.

[110] Sabira K Mannan, Kennard Christopher, and Masud Husain. The role of visual salience in directing eye movements in visual object agnosia. *Current biology*, 19(6):R247–R248, 2009.

[111] Mark Mapstone and Sandra Weintraub. Closing the window of spatial attention: Effects on navigational cue use in Alzheimer's disease. In *Vision in Alzheimer's Disease*, volume 34, pages 290–304. Karger Publishers, 2004.

[112] Roman Marchant, Fabio Ramos, Scott Sanner, and Others. Sequential Bayesian optimisation for spatial-temporal monitoring. In *UAI*, pages 553–562, 2014.

[113] Gesine Marquardt. Wayfinding for people with dementia: A review of the role of architectural design. *Health Environments Research and Design Journal*, 4(2):75–90, 2011.

[114] HéctorP Martínez, Yoshua Bengio, and Georgios Yannakakis. Learning deep physiological models of affect. *IEEE Computational Intelligence Magazine*, 8(2):20–33, 2013.

[115] Mario F. Mendez, Mehdi Ghajarania, and Kent M. Perryman. Posterior cortical atrophy: Clinical characteristics and differences compared to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 14(1):33–40, 2002.

[116] Janine D Mendola, Alice Cronin-Golomb, Suzanne Corkin, and John H Growdon. Prevalence of visual deficits in Alzheimer's disease. *Optometry and Vision Science*, 72(3):155–167, 1995.

[117] Catherine Merck, Audrey Noël, Eric Jamet, Maxime Robert, Anne Salmon, Serge Belliard, and Solene Kalenine. Overreliance on thematic knowledge in semantic dementia: Evidence from an eye-tracking paradigm. *Neuropsychology*, 34(3):331, 2020.

[118] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černock\'y, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.

[119] Maura Mitrushina, Kyle B Boone, Jill Razani, and Louis F D'Elia. *Handbook of normative data for neuropsychological assessment*. Oxford University Press, 2005.

[120] Robert J Molitor, Philip C Ko, and Brandon A Ally. Eye Movements in Alzheimer ' s Disease. *Journal of Alzheimer's Disease*, 44(1):1–12, 2015.

[121] Carlos H. Morimoto and Marcio R.M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[122] Martha Clare Morris, Denis A Evans, Liesi E Hebert, and Julia L Bienias. Methodological issues in the study of cognitive decline. American journal of epidemiology,. *American journal of epidemiology,*, 149(9):789–793, 1999.

[123] Susan M Munn and Jeff B Pelz. 3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 181–188, 2008.

[124] Agneta Nordberg, Juha O. Rinne, Ahmadul Kadir, and Bengt Lngström. The use of PET in Alzheimer disease. *Nature Reviews Neurology*, 6(2):78–87, 2010.

[125] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation Learning by Learning to Count. *Proceedings of the IEEE International Conference on Computer Vision*, pages 5898–5906, 2017.

[126] Mary O'Malley, Anthea Innes, and Jan M. Wiener. Decreasing spatial disorientation in care-home settings: How psychology can guide the development of dementia friendly design guidelines. *Dementia*, 16(3):315–328, 2017.

[127] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 2016.

[128] Akane Oyama, Shuko Takeda, Yuki Ito, Tsuneo Nakajima, Yoichi Takami, Yasushi Takeya, Taiichi Katayama, Hiromi Rakugi, and Ryuichi Morishita. Novel Method for Rapid Assessment of Cognitive Impairment Using High- Performance Eye-Tracking Technology. *Scientific Reports*, 9(1):1–9, 2019.

[129] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[130] Derrick Parkhurst, Klinton Law, and Ernst Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123, 2002.

[131] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[132] Ivanna M. Pavisic, Nicholas C. Firth, Samuel Parsons, David Martinez Rego, Timothy J. Shakespeare, Keir X.X. Yong, Catherine F. Slattery, Ross W. Paterson, Alexander J.M. Foulkes, Kirsty Macpherson, Amelia M. Carton, Daniel C. Alexander, John Shawe-Taylor, Nick C. Fox, Jonathan M. Schott, Sebastian J. Crutch, and Silvia Primativo. Eyetracking metrics in young onset alzheimer's

disease: A Window into cognitive visual functions. *Frontiers in Neurology*, 8:1–16, 2017.

[133] Joao Pereira and Margarida Silveira. Learning representations from healthcare time series data for unsupervised anomaly detection. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 1–7. IEEE, 2019.

[134] Richard Perry and Bruce Miller. Behavior and treatment in frontotemporal dementia. *Neurology*, 56(11), 2001.

[135] Barbara Poletti, Laura Carelli, Federica Solca, Annalisa Lafronza, Elisa Pedroli, Andrea Faini, Nicola Ticozzi, Andrea Ciammola, Paolo Meriggi, Pietro Cipresso, and Others. An eye-tracker controlled cognitive battery: overcoming verbal-motor limitations in ALS. *Journal of neurology*, 264(6):1136–1145, 2017.

[136] Barbara Poletti, Laura Carelli, Federica Solca, Annalisa Lafronza, Elisa Pedroli, Andrea Faini, Stefano Zago, Nicola Ticozzi, Andrea Ciammola, Claudia Morelli, Paolo Meriggi, Pietro Cipresso, Dorothée Lulé, Albert C. Ludolph, Giuseppe Riva, and Vincenzo Silani. An eye-tracking controlled neuropsychological battery for cognitive assessment in neurological diseases. *Neurological Sciences*, 38(4):595–603, 2017.

[137] Stefan Pollmann. *Spatial Learning and Attention Guidance*. Springer, 2020.

[138] Silvia Primativo, Camilla Clark, Keir X X Yong, Nicholas C Firth, Jennifer Nicholas, Daniel Alexander, Jason D Warren, Jonathan D Rohrer, and Sebastian J Crutch. Eyetracking metrics reveal impaired spatial anticipation in behavioural variant frontotemporal dementia. *Neuropsychologia*, 106:328–340, 2017.

[139] Martin Prince and Jim Jackson. World Alzheimer Report 2009. *Alzheimer's Disease International*, pages 1–96, 2009.

[140] Fitzpatrick D Purves D, Augustine GJ. *Types of Eye Movements and Their Functions*. Sunderland (MA): Sinauer Associates, 2nd editio edition, 2001.

[141] Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang Zhong Yang. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, 2017.

[142] Daniele Ravi, Charence Wong, Benny Lo, and Guang-zhong Yang. Deep Learning for Human Activity Recognition : A Resource Efficient Implementation on Low-Power Devices. *IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, pages 71–76, 2016.

[143] K Rayner. Eye movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 124(3):372–422, 1998.

[144] Luz Rello and Miguel Ballesteros. Detecting readers with dyslexia using machine learning with eye tracking measures. *W4A 2015 - 12th Web for All Conference*, 2015.

[145] Jenny Richmond, Paula Sowerby, Michael Colombo, and Harlene Hayne. The effect of familiarization time, retention interval, and context change on adults' performance in the visual paired-comparison task. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 44(2):146–155, 2004.

[146] Irina Rish and Others. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[147] David E Rumelhart, Richard Durbin, Richard Golden, and Yves Chauvin. Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications*, pages 1–34, 1995.

[148] Lucy L Russell, Caroline V Greaves, Rhian S Convery, Jennifer Nicholas, Jason D Warren, Diego Kaski, and Jonathan D Rohrer. Novel instructionless eye tracking tasks identify emotion recognition deficits in frontotemporal dementia. *Alzheimer's research & therapy*, 13(1):1–11, 2021.

[149] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*, pages 71–78, 2000.

[150] Bahman Abdi Sargezeh, Ahmad Ayatollahi, and Mohammad Reza. Investigation of eye movement pattern parameters of individuals with different fluid intelligence. *Experimental Brain Research*, 237(1):15–28, 2019.

[151] Philip Scheltens and Kenneth Rockwood. How golden is the gold standard of neuropathology in dementia? *Alzheimer's and Dementia*, 7(4):486–489, 2011.

[152] Harro Seelaar, Jonathan D Rohrer, Yolande A L Pijnenburg, Nick C Fox, and John C Van Swieten. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: A review. *Journal of Neurology, Neurosurgery and Psychiatry*, 82(5):476–486, 2011.

[153] Sarah C. Seligman and Tania Giovannetti. The Potential Utility of Eye Movements in the Detection and Characterization of Everyday Functional Difficulties in Mild Cognitive Impairment. *Neuropsychology Review*, 25(2):199–215, 2015.

[154] Timothy J. Shakespeare, Diego Kaski, Keir X X Yong, Ross W. Paterson, Catherine F. Slattery, Natalie S. Ryan, Jonathan M. Schott, and Sebastian J. Crutch. Abnormalities of fixation, saccade and pursuit in posterior cortical atrophy. *Brain*, 138(7):1976–1991, 2015.

[155] Timothy J Shakespeare, Yoni Pertzov, Keir X X Yong, Jennifer Nicholas, and Sebastian J Crutch. Reduced modulation of scanpaths in response to task demands in posterior cortical atrophy. *Neuropsychologia*, 68:190–200, 2015.

[156] Mina Shojaeizadeh, Soussan Djamasbi, Randy C. Paffenroth, and Andrew C. Trapp. Detecting task demand via an eye tracking machine learning system. *Decision Support Systems*, 116(June 2018):91–101, 2019.

[157] Shane D Sims and Cristina Conati. A neural architecture for detecting user confusion in eye-tracking data. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 15–23, 2020.

[158] Hari Singh and Jaswinder Singh. Human Eye Tracking and Related Issues: A Review. *International Journal of Scientific and Research Publications*, 2(1):2250–3153, 2012.

[159] Ryan E Solomon, Kyle Brauer Boone, Deborah Miora, Sherry Skidmore, Maria Cottingham, Tara Victor, Elizabeth Ziegler, and Michelle Zeller. Use of the WAIS-III picture completion subtest as an embedded measure of response bias. *The Clinical Neuropsychologist*, 24(7):1243–1256, 2010.

[160] Xiuyao Song, Mingxi Wu, Christopher Jermaine, and Sanjay Ranka. Conditional anomaly detection. *IEEE Transactions on knowledge and Data Engineering*, 19(5):631–645, 2007.

[161] Paul Sowden and Paul Barrett. Psychophysiological Methods. *Research Methods in Psychology*, 8:146–159, 2006.

[162] Mikhail Startsev, Ioannis Agtzidis, and Michael Dorr. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, 51(2):556–572, 2019.

[163] Akara Supratak, Chao Wu, Hao Dong, Kai Sun, and Yike Guo B. *Survey on feature extraction and applications of biosignals*. Springer, 2016.

[164] Tatsuto Suzuki, Keir Yong, Biao Yang, Amelia Carton, Ian McCarthy, Nikolaos Papadosifos, Derrick Boampong, Catherine Holloway, Nick Tyler, and Sebastian Crutch. Locomotion and eye behaviour under controlled environment in individuals with Alzheimer's disease. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 6594–6597. IEEE, 2015.

[165] A. Tales, S. R. Butler, J. Fossey, I. D. Gilchrist, R. W. Jones, and T. Troscianko. Visual search in Alzheimer's disease: A deficiency in processing conjunctions of features. *Neuropsychologia*, 40(12):1849–1857, 2002.

[166] Robyn L. Tate and Nasreddine. Montreal Cognitive Assessment (MoCA). *A Compendium of Tests, Scales and Questionnaires*, 27(1):161–164, 2020.

[167] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[168] Terry T Um, Franz M J Pfister, Ludwig München, Daniel Pichler, Satoshi Endo, Muriel Lang, Urban Fietzek, and Dana Kulić. Data Augmentation of Wearable Sensor Data for Parkinson' s Disease Monitoring using Convolutional Neural Networks. *arXiv preprint arXiv:1706.00527*, 2017.

[169] Vanessa Vallejo, Dario Cazzoli, Luca Rampa, Giuseppe A. Zito, Flurin Feuerstein, Nicole Gruber, René M. Müri, Urs P. Mosimann, and Tobias Nef. Effects of Alzheimer's disease on visual target detection: A "peripheral bias". *Frontiers in Aging Neuroscience*, 8:200, 2016.

[170] Elizabeth K Warrington. *The Camden memory tests manual*, volume 1. Psychology Press, 1996.

[171] Elizabeth K Warrington and Merle James. The visual object and space perception battery. 1991.

[172] David Wechsler. Wechsler adult intelligence scale–. 1955.

[173] Salmon D P Weintraub S, Wicklund AH. The Neuropsychological Profile of Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 4(2):a006171, 2012.

[174] Thomas D W Wilcockson, Diako Mardanbegi, Baiqiang Xia, Simon Taylor, Pete Sawyer, Hans W Gellersen, Ira Leroi, Rebecca Killick, and Trevor J Crawford. Abnormalities of saccadic eye movements in dementia due to Alzheimer's disease and mild cognitive impairment. *Aging (Albany NY)*, 11(15):5389, 2019.

[175] J R Willison and E K Warrington. Cognitive retardation in a patient with preservation of psychomotor speed. *Behavioural neurology*, 5(2):113–116, 1992.

[176] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):419–433, 1989.

[177] R. Woodbridge, M. P. Sullivan, E. Harding, S. Crutch, K. J. Gilhooly, M. L.M. Gilhooly, A. McIntyre, and L. Wilson. Use of the physical environment to support everyday activities for people with dementia: A systematic review. *Dementia*, 17(5):533–572, 2018.

[178] Yu Tzu Wu, Linda Clare, John V. Hindle, Sharon M. Nelis, Anthony Martyr, and Fiona E. Matthews. Dementia subtype and living well: results from the Improving the experience of Dementia and Enhancing Active Life (IDEAL) study. *BMC medicine*, 16(1):140, 2018.

[179] Chen Xia, Kexin Chen, Kuan Li, and Hongxia Li. Identification of Autism Spectrum Disorder via an Eye-Tracking Based Representation Learning Model. In *2020 7th International Conference on Bioinformatics Research and Applications*, pages 59–65, 2020.

[180] Yuehan Yin, Yahya Alqahtani, Jinjuan Heidi Feng, Joyram Chakraborty, and Michael P McGuire. Classification of Eye Tracking Data in Visual Information Processing Tasks Using Convolutional Neural Networks and Feature Engineering. *SN Computer Science*, 2(2):1–26, 2021.

[181] Yuehan Yin, Chunghao Juan, Joyram Chakraborty, and Michael P McGuire. Classification of eye tracking data using a convolutional neural network. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 530–535. IEEE, 2018.

[182] Keir X X Yong, Timothy J Shakespeare, Dave Cash, Susie M D Henley, Jason D Warren, and Sebastian J Crutch. (Con) text-specific effects of visual dysfunction on reading in posterior cortical atrophy. *cortex*, 57:92–106, 2014.

[183] Keir X.X. Yong, Timothy J. Shakespeare, Dave Cash, Susie M.D. Henley, Jennifer M. Nicholas, Gerard R. Ridgway, Hannah L. Golden, Elizabeth K. Warrington, Amelia M. Carton, Diego Kaski, Jonathan M. Schott, Jason D. Warren, and Sebastian J. Crutch. Prominent effects and neural correlates of visual crowding in a neurodegenerative disease population. *Brain*, 137(12):3284–3299, 2014.

[184] Sungmin You, Baek Hwan Cho, Soonhyun Yook, Joo Young Kim, Young-Min Shon, Dae-Won Seo, and In Young Kim. Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network. *Computer Methods and Programs in Biomedicine*, page 105472, 2020.

[185] Zachary Infantolino; Gregory A. Miller. Psychophysiological Methods in Neuroscience. *Introduction to Psychology: The Full Noba Collection*, pages 51–64, 2014.

[186] Raimondas Zemblys, Diederick C Niehorster, and Kenneth Holmqvist. gazeNet : End-to-end eye-movement event detection with deep neural networks. *Behavior Research Methods*, 51(2):840–864, 2019.

[187] Raimondas Zemblys, Diederick C Niehorster, and Kenneth Holmqvist. gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior research methods*, 51(2):840–864, 2019.

[188] A Tianyi Zhang and B Olivier Le Meur. How Old Do You Look? Inferring Your Age From Your Gaze. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2660–2664. IEEE, 2018.

[189] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[190] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision*, 127(3):302–321, 2019.