Systematic Review

Jomar Alcantara
Department of Computer Science
School of Engineering and Applied Sciences
Aston University

Peter Sawyer
Department of Computer Science
School of Engineering and Applied Sciences
Aston University

George Vogiatzis
Department of Computer Science
School of Engineering and Applied Sciences
Aston University

April 18, 2019

Abstract

This is the paper's abstract ...

1 Introduction

Dementia has been identified as one of those fast growing difficulties facing the world. A recent report suggests that in 2015 there were 46 million people with a diagnosis of dementia and that number is expected to hit 131.5 million by 2050 [?]. The report also states that the worldwide cost of dementia in 2018 is estimated to be in the region of one trillion US dollars.

In 2009, the Department of Health designed it's National Dementia strategy and as part of this made early diagnosis and support one of it's key priorities [?]. A lot of work has gone into trying to find ways of improving the early diagnosis of Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) with research focused on two areas - identifying biological markers and analyzing the cognitive decline of those who are suspected to have the disease [?]. As described above, the numbers of those suffering from AD and MCI are going to increase as the population ages [?] and thus it is important that we utilize technology wherever possible to aid clinicians in the detection of MCI and AD. At the present time diagnosis is typically conducted at memory clinics by trained clinicians [?]. I theorize that we may be able to enable an earlier diagnosis of those with MCI and AD using samples of spontaneous speech, natural language processing (NLP) and machine learning (ML).

There is a large body of research that looks at the decline in language in those with MCI and AD [?, ?]. However there is conflicting evidence in these studies about which declining language factors are associated of MCI and AD [?, ?]. Research therefore should look at these features in more detail and a clarification of this currently disorganised picture should go some way to helping researchers further understand the disease and it's progression. Another area of focus for research of this nature is the process of collecting appropriate language samples. Whilst collecting samples of language is comparatively unintrusive, researchers recognise that these samples require a rich sample of language that potentially cannot be generated by tasks such as the picture description task. Therefore, it would be useful to explore whether spontaneous discourse such a semi-structured interview, has the ability to put pressure on both the cognitive and linguistic systems in the same way as traditional cognitive tests such that it might be able to distinguish between healthy controls, those with MCI and those with AD. There is some evidence to support this. Berisha et al [?], has shown through a longitudinal language analysis of spontaneous speech that there are marked differences in this process between those who would go on to have a diagnosis of AD and a healthy control.

The potential impact of this research in this area is immense. Research has shown that early diagnosis of people with AD or MCI improves sufferers quality of life and can, in some cases, slow the progress of the disease. Early diagnosis can increase the number of research opportunities for understanding the early stages of dementia and how the disease progresses so that more research can be conducted which may, in the future, lead to new treatments

and other interventions.

The aim of my research is to find less burdensome ways of detecting dementia without the use of invasive procedures (taking bloods, or using medical equipment such as MRI's and EEG's) and without resorting to to time-consuming and expensive psychological tests. There is a lot of research into the analysis of language as a bio-marker for MCI and Early Dementia. Given that sampling a person's language is relatively effortless, my research looks at whether we can find bio-markers of MCI and Early Dementia in natural language.

Concerns: Language and Memory are quite naturally intertwined and it would be difficult to test one without some reliance on the other. I'm not going to control for memory problems as a potential confound, but does this weaken the research? How do I defend this?

Introduction to the problem of dementia in the context of the wider world including quality of life and financial implications. Exploration of dementia as a syndrome rather than a disease, and a look at the different variants of dementia. A look at the rationale behind research into the early diagnosis of dementia as well as a brief look at what has been done in the area (wide context, so pharmacological and psychological).

The question to be addressed in this systematic review is how has the field of machine learning and natural language processing addresses language deterioration in the diagnosis of Mild Cognitive Impairment and Early Alzheimer's Disease.

The remainder of this article is organized as follows. Section 2 gives an account of the process of this Systematic Review. Our new and exciting results are described in Section 3. We discuss the results and implications in Section 4. Finally, Section 5 gives the conclusions.

2   Methodology

A systematic literature review (SLR) describes a process which aims to identify, evaluate and interpret the research and literature in a given area. They are designed to provide a complete and exhaustive summary of the current evidence relevant to an identified research question. SLR's conduct a thorough search of all literature following a pre-defined protocol that specifies

focused research questions, identifies criteria for the selection of studies and assessment of their quality, and forms to execute the data extraction and synthesis of results.

Common motivations for conducting a SLR are:

1. to summary all the evidence about a topic.

2. find gaps in the research.

3. to provide a ground for a fundament to new research.

4. and to examine how the current research supports a hypothesis.

Performing a SLR comprises the following steps:

1. (i) identify the need for performing the SLR.

2. (ii) formulate research questions.

3. (iii) execute a comprehensive search and selection of primary studies.

4. (iv) assess the quality and extract data from the studies.

5. (v) interpret the results.

6. (vi) report the SLR.

The main research question this SLR aims to address is: "How has the field of machine learning and natural language processing addresses language deterioration in the diagnosis of Mild Cognitive Impairment and Early Alzheimer's Disease?". This main question was decomposed further into five research questions:

1. RQ1: Which ML and MS techniques are being used in the dementia and comorbidities research?

2. RQ2: What data characteristics (variables, determinants and indicators) are being considered when applying the ML or/and MS techniques (physiological, demographic/social, genetics, lifestyle etc)?

3. RQ3: What are the goals of the studies that employ ML or MS techniques for prognosis of dementia and comorbidities?

4. RQ4: How is data censoring being handled in the studies?

5. RQ5: Do the studies focus on individuals or populations?

The present paper builds upon these questions and additionally presents the results of the other two additional research questions. Further, the key terms related to comorbidities were included in the search string to ensure that relevant studies about ML or MS for the prognosis of a disease, where dementia is considered a comorbidity to that disease would also be retrieved from the database searches, even when the term dementia was not mentioned in the paper's title or abstract.

2.1 Search strategy

To address the research questions, a search string was defined using the PICO approach, which decomposes the main question into four parts: population, intervention, comparison and outcome. The comparison component was discarded because the SLR was mainly concerned with a characterization. For each of the remaining components, keywords were derived and their rationale can be represented as follows:

1. Population: Studies that present research on dementia and comorbidities. Dementia's keywords were selected from the "Systematized Nomenclature of Medicine–Clinical Terms" and selected by A4. Comorbidities' keywords were extracted from the Marengoni et al. SLR in this topic.

2. Intervention: ML or MS techniques. The ML keywords were selected from the branch "Machine Learning Approaches" of the "2012 ACM Computing Classification System". The MS keywords were selected by A2.

3. Outcome: Prognosis on dementia and comorbidies. The prognosis keywords were provided by A4.

The automated searches were performed in the Pubmed, Web of Science and Scopus databases. Table 1 shows the search string used for the Pubmed automated search, but note that this search string was adapted to each of the other databases' search context.

(dementia OR MCI OR Mild Cognitive Impairment OR Alzheimer's OR Mild Neurocognitive Disorder OR AD) AND TOPIC: (machine learning OR Data Mining OR Decision Support System OR NLP OR Natural Language Processing) AND TOPIC: (prognosis OR prognostic estimate OR predictor OR prediction OR model OR patterns OR diagnosis OR diagnostic OR forecasting OR projection OR Deep Language Model OR Deep Neural Network) AND TOPIC: (classification OR regression OR kernel OR support vector machines OR Gaussian Process OR Bayesian Network OR Factor Analysis OR Deep Learning OR Neural Networks OR Maximum Likelihood OR Principal Component Analysis OR Markov OR Linear Model OR Mixture Model OR Perceptron Algorithm OR Logical Learning OR relational learning OR Supervised Learning OR Unsupervised Learning OR clustering OR Decision Tree) AND TOPIC: (Language OR Cognitive OR Speech OR Conversation OR Connected Speech OR Picture Description OR Discourse Analysis OR Verbal Fluency)

Searched - 4th April 2019 - Generated 1257 Articles

Table 1: Search Terms for Web of Science database

(dementia OR MCI OR Mild Cognitive Impairment OR Alzheimer's OR Mild Neurocognitive Disorder OR AD) AND TOPIC: (machine learning OR Data Mining OR Decision Support System OR NLP OR Natural Language Processing) AND TOPIC: (prognosis OR prognostic estimate OR predictor OR prediction OR model OR patterns OR diagnosis OR diagnostic OR forecasting OR projection OR Deep Language Model OR Deep Neural Network) AND TOPIC: (classification OR regression OR kernel OR support vector machines OR Gaussian Process OR Bayesian Network OR Factor Analysis OR Deep Learning OR Neural Networks OR Maximum Likelihood OR Principal Component Analysis OR Markov OR Linear Model OR Mixture Model OR Perceptron Algorithm OR Logical Learning OR relational learning OR Supervised Learning OR Unsupervised Learning OR clustering OR Decision Tree) AND TOPIC: (Language OR Cognitive OR Speech OR Conversation OR Connected Speech OR Picture Description OR Discourse Analysis OR Verbal Fluency)

Searched - 4th April 2019 - Generated 1257 Articles

Table 2: Search Terms for Scopus database

(dementia OR MCI OR Mild Cognitive Impairment OR Alzheimer's OR Mild Neurocognitive Disorder OR AD) AND TOPIC: (machine learning OR Data Mining OR Decision Support System OR NLP OR Natural Language Processing) AND TOPIC: (prognosis OR prognostic estimate OR predictor OR prediction OR model OR patterns OR diagnosis OR diagnostic OR forecasting OR projection OR Deep Language Model OR Deep Neural Network) AND TOPIC: (classification OR regression OR kernel OR support vector machines OR Gaussian Process OR Bayesian Network OR Factor Analysis OR Deep Learning OR Neural Networks OR Maximum Likelihood OR Principal Component Analysis OR Markov OR Linear Model OR Mixture Model OR Perceptron Algorithm OR Logical Learning OR relational learning OR Supervised Learning OR Unsupervised Learning OR clustering OR Decision Tree) AND TOPIC: (Language OR Cognitive OR Speech OR Conversation OR Connected Speech OR Picture Description OR Discourse Analysis OR Verbal Fluency)

Searched - 4th April 2019 - Generated 1257 Articles

Table 3: Search Terms for ProQuest database

(dementia OR MCI OR Mild Cognitive Impairment OR Alzheimer's OR Mild Neurocognitive Disorder OR AD) AND TOPIC: (machine learning OR Data Mining OR Decision Support System OR NLP OR Natural Language Processing) AND TOPIC: (prognosis OR prognostic estimate OR predictor OR prediction OR model OR patterns OR diagnosis OR diagnostic OR forecasting OR projection OR Deep Language Model OR Deep Neural Network) AND TOPIC: (classification OR regression OR kernel OR support vector machines OR Gaussian Process OR Bayesian Network OR Factor Analysis OR Deep Learning OR Neural Networks OR Maximum Likelihood OR Principal Component Analysis OR Markov OR Linear Model OR Mixture Model OR Perceptron Algorithm OR Logical Learning OR relational learning OR Supervised Learning OR Unsupervised Learning OR clustering OR Decision Tree) AND TOPIC: (Language OR Cognitive OR Speech OR Conversation OR Connected Speech OR Picture Description OR Discourse Analysis OR Verbal Fluency)
Searched - 4th April 2019 - Generated 1257 Articles

Table 4: Search Terms for IEEE Xplore database

Scopus	Web of Science	PubMed	IEEE Xplore
1002	991	376	230

Table 5: Totals from each database

Stage	Papers remaining
Start	2599

Table 6: Totals at each stage

2.2 Study selection

A total of 2599 papers were identified through the searches conducted above. Papers that were duplicated were removed and Papers were then evaluated by the JA based on the inclusion and exclusion criteria defined previously (see Table 6). Where JA could not reach a decision, PS and GV were consulted and majority vote on inclusion was conducted. Each paper was initially presented with just the title and the abstract.

After the evaluations, 37 papers were selected. Then a one-iteration backward snowballing was carried out looking for possible additional studies. The 1199 new identified studies were assessed analogously as the previous ones, resulting in 41 new selected papers. Throughout the whole selection process, A3 and A4 acted in conflict resolution in the case where A1 and A2 couldn't reach an agreement.

In total, 78 papers were selected to be fully read and assessed regarding its eligibility. The ones that successfully passed the established criteria previously defined in the protocol, had their relevant data extracted. In order to minimize the chance of selecting studies with bias evidence, a quality assessment questionnaire was used. This questionnaire was adapted from Kitchenham's guidelines and can be found in the SLR protocol. If the grading attributed to a paper fell below 8 points (out of a total of 12), it would be rejected for quality reasons. The 8-point threshold was decided in the research group discussions involving all the authors. In this phase, a paper could also be rejected due to inclusion and exclusion criteria because the selection process adopted an inclusive approach. This means that during the reading of the titles and abstracts, in the case where the information provided was incomplete or too general it was selected to be fully read in the posterior phase. A common example is the case when the data analysis technique specified in the abstract was merely "classification", so it was not possible to know if any machine learning occurred.

As in the study selection, a quality assessment evaluation round was performed beforehand to ensure consistency in the evaluations. A1, A2 and A5 participated in this task. In total, 37 studies composed the final set of

Inclusion Criteria	Exclusion Criteria
Be a primary study in English; AND address research on dementia and comorbidities; AND address at least one ML or MS technique; AND address a prognosis related to dementia and comorbidities.	Be a secondary or tertiary study; OR be written in another language other than English; OR do not address a research on dementia and comorbidities; OR do not address at least one ML or MS technique; OR do not address a prognosis related to dementia and comorbidities.

Table 7: Inclusion and Exclusion Criteria

included primary studies and had their relevant data extracted, 7 papers were rejected due quality reasons, and 34 papers were rejected due to failing the inclusion and exclusion criteria. One reason for the high number of the latter was the decision to exclude the papers that used solely statistical methods as data analysis tech- niques to build the prognostic models. The selected studies were also assessed for the risk of cumulative evidence bias. This was done by checking, in the case of the same research group with different studies in the final set of included primary studies, if it was justified having both studies (i.e different samples).

A search of the literature was conducted using ProQuest (PsychArticles), SCOPUS, Web of Science. The following results were found (Table 1). All papers were then reviewed for relevance by reading the abstract and full text where appropriate and a shortlist was compiled. An additional search through references of shortlisted papers was also conducted and any papers who upon further review appeared relevant were added to the shortlist. Papers were included where researchers used machine learning to classify participants as MCI, AD or Healthy using language. We excluded any papers that focused on other forms of dementia or cognitive impairment, as well as any papers in which the language being analysed was not English.

2.3   Data collection

For the data collection, a base extraction form was defined in the protocol, but later in the study it was evolved based on the research group discussions. Table 3 lists and defines the collected variables.

Variable	Definition
Conditions Studied	For which dementia disorder is the study deriving a prognosis.
Database used in the study	Name and origin of the data source used to derive the prognosis of the studied dementia.
Dataset Categories	Classes in which the data units were divided into.
Handling of Censored Data	Description of the way in which censored data was handled.
Follow-up period	Period of time, which the data units were followed.
Data Analysis Techniques	ML or MS techniques that were used to build the prognostic models.
Model Variables	The variables used in building the prognostic models.
Aim of the Study	The goal of the built prognostic models.
Focus of the Study	If the built prognostic models aim its predictions on an individual or population level.

Figure 1: PRISMA flow chart.

In addition to these variables other basic data about the studies was collected, these were: title, authors, journal/source, year and type of publication. No summary measures were used. Summary tables were used for the synthesis of results and no additional analyses were carried out

3 Results

In this section we describe the results.

3.1 Language Analysis of Speakers with Dementia of the Alzheimer's Type - Guinn and Habash (2012)

3.2 Features and Machine Learning Classification of Connected Speech Samples from patients with Autopsy Proven Alzheimer's Disease with and without additional Vascular Pathology - Rentoumi, Raoufian, Ahmed, de Jager, Garrard (2014)

3.3 Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances - Orimaye, Wong and Golden (2014)

3.4 Learning Lingustic Biomarkers for Predicting Mild Cognitive Impairment using Compound Skip-grams - Orimaye, Wong, Wong and Tai (2015)

3.5 Diagnosing People with Dementia using Automatic Conversation Analysis - Mirheidari, Reuber, Blackburn, Walker (2016)

3.6 Formulaic Language in People with Probably Alzheimer's Disease: A Frequency-Based Approach - Zimmerer, Wibrow and Varley (2016)

3.7 Linguistic Features Identify Alzheimer's Disease in Narrative Speech - Fraser, Melzer and Rudzicz (2016

3.8 Declines in Connected Language Are Associated with Very Early Mild Cognitive Impairment: Results from the Wisconsin Registry for Alzheimer's Prevention - Mueller, Koscik, Hermann, Johnson and Turkstra (2018)

3.9 Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia - Orimaye, Wong and Wong (2018)

In this paper, Orimaye et al use deep-deep neural networks language models (D2NNLM) to learn linguistic changes that distinguish the language of

patients with MCI and AD-type dementia from the healthy controls using higher order n-grams. An ordinary DNNLM uses lower order n-gram N-dimensional sparse vectors as descrete feature representations to train the neural network with multiple hidden layers.

3.9.1 Deep Neural Network Language Models

4 Discussion

5 Conclusions

We worked hard, and achieved very little.