A Thesis for the Degree of Ph.D. in Engineering

Non-Invasive Detection of Dementia Through Language and Diagnosis Video Data: Machine Learning Methods and Applications

September 2025

Graduate School of Science and Technology
Keio University

Chuheng Zheng

別表６
（３）

Keio University

Thesis Abstract

No.

Registration Number	□ "KOU" No. *Office use only	Name	ZHENG, Chuheng

Thesis Title

Non-Invasive Detection of Dementia Through Language and Diagnosis Video Data: Machine Learning Methods and Applications

Thesis Summary

Alzheimer's disease, as the most prevalent form of dementia, accounts for approximately 60–70% of all dementia cases. It is currently incurable with existing medical treatments. Timely diagnosis and intervention are proved to be the effective method to slow down the progression of dementia and improve the quality of life for patients and their families. Therefore, non-invasive and easily accessible methods for early detection are essential.

This dissertation explores the application of machine learning techniques to enhance early dementia detection, focusing on the analysis of language and diagnosis video data. Specifically, we investigated the use of various machine learning models, including long short-term memory networks, transformers, convolutional neural networks, and other models to identify dementia-related patterns in speech data from the DementiaBank dataset and video data from the PROMPT dataset. The goal is to develop non-invasive, easy-to-use and affordable methods that can assist in the diagnosis of dementia, thereby improving patient outcomes and facilitating timely interventions.

Chapter 1 introduces the background of dementia and its impact on individuals and society, emphasizing the need for early detection and the potential of machine learning to address this challenge. It outlines the main objectives of the dissertation, concentrating on developing non-invasive and easy-to-use methods for early dementia detection methods. It also provides an overview of the challenges we encountered and the contributions we have made in overcoming these challenges, which mainly includes data scarcity, lack of explanation, low data quality, and institutional bias.

Chapter 2 focuses on language-based dementia detection using language models adapted from natural language processing. Despite several studies demonstrating the potential of language models in dementia detection, it remains a challenge to provide the patients and clinicians with justifiable and validatable explanations. In this chapter, we propose a novel approach that separates the semantic and syntactic components of language, allowing for a more understandable analysis of dementia-related patterns. Distinctively, we proposed leveraging stop words and part-of-speech sequences, to capture the syntactic structure of language, while context words are used to capture the semantic content. By separately analyzing these two components, we aim to provide clearer post hoc explanations for the model's predictions, providing validatable explanations for the patients and clinicians as to which part of the language is impaired. We evaluated several models, including traditional n-gram models, long short-term memory networks, and attention-based models. Specifically, three models were developed: one using only context words (semantic features), one using stop words and part-of-speech sequences

Thesis Abstract

No.

(syntactic features), and one combining both types of features. Experimental results showed that both semantic and syntactic features contribute to dementia detection, while syntactic features contribute more significantly. The three models achieved accuracies of 70.00%, 76.16%, and 81.54%, respectively.

Chapter 3 leverages the capabilities of recent large language models to generate text embeddings and utilizes autoencoders for anomaly detection in dementia-related data. In this chapter, we proposed a zero-shot learning approach to detect dementia using healthy data only and a binary classification method using both healthy and dementia data. We explored the effectiveness of different text embeddings, including those from RoBERTa and LLaMA models, in capturing the nuances of dementia-related data. We experimented with various autoencoder architectures, including deep autoencoders, denoising autoencoders, and variational autoencoders, to assess their performance in detecting dementia. The results demonstrated that the choice of architecture and latent dimension significantly impacts the performance of the autoencoder in detecting dementia. Denoising autoencoders, particularly with a noise level of 0.3, achieved the highest average accuracy, indicating that adding noise during training can help the model generalize better to unseen data.

Chapter 4 explores the potential of face-related features for dementia detection using the PROMPT dataset. In this chapter, we first identified and proposed methods to reduce the institutional bias in the PROMPT dataset, caused by limited data collection from several limited institutions. After identifying this bias, we extracted the three types of features from the diagnosis video recordings, including face mesh, histogram of oriented gradients features, and action units, and evaluated their effectiveness in distinguishing between dementia patients and healthy controls. Traditional machine learning and deep learning models were trained on these features to investigate their effectiveness. Among them, histogram of oriented gradients features achieved the highest accuracy of 79%, followed by action unit features with 71%, and face mesh features with 76%. These results suggest that facial information serves as a valuable indicator for automated dementia detection and highlight the importance of exploring non-verbal behavioral cues.

Chapter 5 summarizes the dissertation, highlighting its key points and contributions, and suggesting possible future work.