MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231085

Dementia Prediction in Older Adults Using Sex-Specific Health Trajectory Clustering

Omar A. IBRAHIM [a], Muskan GARG [a], Sunyang FU [a], Maria VASSILAKI [a], Ronald C. PETERSEN [a], Jennifer ST SAUVER [a] and Sunghwan SOHN [a,1]
[a] *Mayo Clinic, Rochester, Minnesota, USA*

Abstract. With increasing number of people living with dementia, the problem of late diagnosis significantly impacts a person's quality of life while early signs of dementia may provide useful insights to facilitate better treatment plans. With time, this progressive neurodegenerative syndrome could progress from mild cognitive impairment to dementia. A pattern of health conditions can be characterized in unsupervised manner to help predict this progress. As a significant extension to our previous work with streaming clustering model, we consider additional information for predicting dementia onset. With empirical observations, we discover the importance of examining sex and age to predict dementia onset. To this end, we propose a sex-specific model with age-constraint for predicting dementia onset and validate the effectiveness of our models using data from Mayo Clinic Study of Aging (MCSA). The proposed sex-specific models for older adult populations (>=65 years of age) outperformed the previous models with F-score of 77% and 78% for male-specific and female-specific models, respectively. Our experiments of sex-specific temporal clustering of features in older adults demonstrate the potential of more personalized models for early alerts of dementia.

Keywords. Clustering, dementia, mild cognitive impairment, prediction

1. Introduction

Dementia is a major health concern for older adults affecting 50 million people world-wide [1]. As evident from past studies, more than 40% of people with Alzheimer Disease (AD) and other types of dementia received a delayed clinical diagnosis of their condition [2]. Early diagnosis benefits the patient, allowing timely interventions and opportunities to ensure safety and future planning, including caregiver support [3]. Early signs of dementia may be present in patients' evaluations, several years before actual diagnosis [4].

In this study, we further examine data of the participants serial evaluations in the Mayo Clinic Study of Aging (MCSA) dataset to determine whether early changes in evaluations accurately predicted later diagnosis of dementia. To this end, we explored the contribution of routinely collected neuropsychological test score in combination with other clinical data to predict dementia onset.

We first adopted a streaming clustering algorithm to identify distinct groups of dementia patients and to predict early signs of dementia through outliers of clusters. A

[1] Corresponding Author: Sunghwan Sohn, Ph.D., Mayo Clinic, Rochester, Minnesota, USA; email: Sohn.Sunghwan@mayo.edu.

heuristic approach of monitoring cluster change gives useful insights for prediction mechanism. Next, we added the scores of nine neuropsychological tests to enhance the performance of prediction algorithms through clusters. We further investigate model performance in different age groups and propose sex-specific streaming clustering models with age constraint as a new prediction model for dementia patients.

2. Methods

This study was approved by the Mayo Clinic and the Olmsted Medical Center Institutional Review Boards. The data is acquired from the MCSA [5]. The MCSA is a population-based cohort study initiated in 2004 in Olmsted County (MN, USA) to study cognitive aging, mild cognitive impairment (MCI) and dementia and its biomarkers. It comprises 6,185 participants with serial cognitive evaluations at baseline and at follow-up visits approximately every 15 months for a total of 26,807 visits (4.3 average visits per patient). Among these, 3,070 patients were female (49.6%), and 729 patients (11.6%) are diagnosed with dementia. The median age of the cohort at baseline was 73.

We adopted MU streaming clustering (MUSC), which is a Gaussian Mixture Model (GMM) based on streaming clustering algorithm coupled with possibilistic clustering with the goal of identifying signs of early diseases [6]. Since the number of clusters in streaming data points is usually unknown, we must automatically determine the number of initial clusters in the MUSC model. We use probabilistic C-means algorithm [7] and Automatic Merging Possibilistic Clustering Method (AMPCM) [8] to cluster the initial data points, detect anomalies and initialize the GMM. After initializing the streaming clustering model using a small percent of the data, we feed the remaining data samples to the model in a streaming fashion (one sample at a time). When a new sample arrives, we compute the Mahalanobis distance to the mean of each cluster using given features [9].

If the distance is less than the pre-specified threshold and new data is closest to the cluster, this data point will be added to that cluster. The threshold is set as the average distance between data points in existing cluster. When a new data sample is appended in existing cluster, the threshold is updated incrementally. If the distance of a new data sample is greater than the threshold, it is identified as an outlier and moved to the anomaly list. We predict dementia patients by monitoring the visits which are flagged as outliers.

Our previous study with MUSC on the whole MCSA cohort (n = 6,185, 729 dementia) produced reasonable performance as an unsupervised approach (F-score of 0.64) [9]. In this study, we further explored the effect of sex and age (i.e., a sex-specific model with an age constraint (>=65) with additional features (neuropsychological test score, described in the next paragraph) to improve the prediction performance.

For each evaluation visit a psychometrist administers nine neuropsychological tests scores to assess four cognitive domains: (i) memory (Auditory Verbal Learning Test (AVLT) delayed response [10], Wechsler Memory Scale-Revised (WMS-R) [5], Logical Memory II and Visual Reproduction II); (ii) language (Boston Naming Test [11], Category Fluency) [12]; (iii) attention/executive (Trail-Making Test B [6,12], Wechsler Adult Intelligence Scale-Revised (WAIS-R), Digit Symbol); and (iv) visuospatial (WAIS-R Picture Completion and Block Design [7]). These

neuropsychological test score features have been demonstrated to show more effect on machine learning models to predict dementia in a supervised setting [13].

3. Results

The original MUSC model was based on 49 features including demographic, physical, neuropsychiatric, and social characteristics [9]. We added nine neuropsychological test score features to examine their effect on the clustering model. With these added features, the model performance was enhanced in terms of precision while maintaining a similar F1-score (Table 1).

Table 1. Dementia prediction for the original MUSC model and MUSC model with added neuropsychological test score features (MUSC-NS).

Model	Precision	Recall	F1-Score
MUSC	0.56	0.75	0.64
MUSC-NS	0.62	0.68	0.65

We further investigated the effect of age in the performance of MUSC-NS. Although dementia is a condition mainly for older adults, the MCSA participants are 30 years of age and older. We hypothesized that model performance would improve in older adults who are mostly at risk of developing dementia. Therefore, we studied model performance in persons >= 65 years, and results are shown in Table 2. As we remove younger persons of age of 30 to 64, there was a noticeable improvement in the performance against the original MUSC and MUSC-NS (F1-score of 0.64 and 0.65, respectively in Table 1), both of which used the cohort of age >= 30. Age played a pivotal role in dementia prediction in our clustering model.

Table 2. Dementia prediction with an age constraint.

Model	Precision	Recall	F1-Score
MUSC (age>=65)	0.60	0.78	0.68
MUSC-NS (age>=65)	0.65	0.77	0.71

We next developed the MUSC model independently on the separate data for each sex (M: male, F: female) to account for the difference of dementia risk among males and females (Table 3). The experimental results clearly show that the sex-specific models significantly outperformed the original MUSC and MUSC-NS model. We observed improved performance with slightly higher F1 score for the female-based MUSC model as compared to the male-based MUSC model. The sex-specific model with an age constraint produced the highest performance (F1-score of 0.77 and 0.78 for male and female respectively) against all other models. As dementia is a condition prevalent in older adults, a prediction pattern might change due to age differences.

Table 3. Sex-specific dementia prediction.

Model	Precision	Recall	F1-Score
MUSC-NS (M)	0.70	0.73	0.71
MUSC-NS (F)	0.73	0.72	0.72
MUSC-NS (M) (age>=65)	0.78	0.77	0.77
MUSC-NS (F) (age>=65)	0.76	0.80	0.78

4. Discussion

In extension to our previous study for predicting cognitively impaired individuals with temporal stream clustering, we investigated the effect of a sex-specific model with an age constraint adding nine new neuropsychological test scores. Neuropsychological test scores improved model precision showing its discriminatory characteristic but with similar F1-score. In this study, we focused on older individuals as the number of older people is rising and dementia prevalence is also increasing worldwide. As dementia is a prevalent condition in older adults, the cohort with a small portion of younger individuals might cause some degree of confusion to the model predictive capability. With an age constraint (>= 65), both models with and without the neuropsychological test scores showed improved prediction performance (higher for the model including neuropsychological test scores). We further explored the role of sex in our clustering model and demonstrated significant improvement in prediction when it is coupled with an age constraint and neuropsychological test scores.

Our model is based on unsupervised learning, not requiring labeled dementia cases but determining outliers as a dementia case. With limited resources for comprehensive assessment and test for accurate dementia ascertainment, an unsupervised approach would provide additional benefit for early detection of dementia. Considering the current lack of strategies using temporal trends of patients' health conditions in an unsupervised manner for dementia prediction, our models show promising potential as an early warning for older adults at higher risk of dementia.

A limitation of this study includes the use of assessment and test features in the MCSA cohort which may not be available in other studies for exact reproducibility. In the future, we will extend our studies extracting relevant data elements from routine clinical documents for secondary use of EHRs. This study was performed in a single research cohort, and result must be replicated in other settings to assess generalizability.

5. Conclusions

A sex-specific steaming clustering model with an age constraint demonstrated improved capability in dementia prediction when using longitudinal evaluation data. Notable incremental improvement was observed when adding age (>= 65) and sex (male vs. female model) constraint with added neuropsychological test scores, demonstrating a significant role of age and sex in our clustering-based prediction model. Our model could facilitate early warning for dementia as clinical decision support, addressing delayed diagnosis and improving care plans and health outcomes in older adults.

Acknowledgements

This study was supported by NIA R01 AG068007. The Mayo Clinic Study of Aging was supported by National Institutes of Health (NIH) Grants (U01 AG006786, P30 AG062677, R37 AG011378, R01 AG041851), the Alexander Family Alzheimer's Disease Research Professorship of the Mayo Clinic, the GHR Foundation, the Mayo Foundation for Medical Education and Research, the Liston Award, the Schuler

Foundation and was made possible by the Rochester Epidemiology Project (R01 AG034676).

References

[1] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, Brayne C, Burns A, Cohen-Mansfield J, Cooper C, Costafreda SG, Dias A, Fox N, Gitlin LN, Howard R, Kales HC, Kivimäki M, Larson EB, Ogunniyi A, Orgeta V, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020 Aug;396(10248):413-46, doi: 10.1016/S0140-6736(20)30367-6.

[2] Chodosh J, Petitti DB, Elliott M, Hays RD, Crooks VC, Reuben DB, Galen Buckwalter J, Wenger N. Physician recognition of cognitive impairment: evaluating the need for improvement. J Am Geriatr Soc. 2004 Jul;52(7):1051-9, doi: 10.1111/j.1532-5415.2004.52301.x.

[3] Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of Dementia: an Observational Study of Patterns in Diagnosis and Awareness in US Older Adults. J Gen Intern Med. 2018 Jul;33(7):1131-8, doi: 10.1007/s11606-018-4377-y.

[4] Goudarzvand S, Sauver JS, Mielke MM, Takahashi PY, Sohn S. Analyzing early signals of older adult cognitive impairment in electronic health records. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2018 Dec 3; Madrid, Spain: IEEE. p. 1636-40. doi: 10.1109/BIBM.2018.8621278.

[5] Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, Ivnik RJ, Tangalos EG, Petersen RC, Rocca WA. The Mayo clinic study of aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology. 2008 Feb;30(1):58-69, doi: 10.1159/000115751.

[6] Ibrahim OA, Du Y, Keller J. Robust on-line streaming clustering. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations. IPMU 2018; 2018 Jun 11-15; Cham: Springer. p. 467-78. doi: 10.1007/978-3-319-91473-2_40.

[7] Krishnapuram R, Keller JM. A possibilistic approach to clustering. IEEE Trans Fuzzy Syst. 1993 May;1(2):98-110, doi: 10.1109/91.227387.

[8] Yang MS, Lai CY. A robust automatic merging possibilistic clustering method. IEEE Trans Fuzzy Syst. 2010 Sep;19(1):26-41, doi: 10.1109/TFUZZ.2010.2077640.

[9] Ibrahim OA, Fu S, Vassilaki M, Petersen RC, Mielke MM, St Sauver J, Sohn S. Early alert of elderly cognitive impairment using temporal streaming clustering. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2021 Dec;2021:905-12, doi: 10.1109/bibm52615.2021.

[10] Ibrahim OA, Keller J, Popescu M. An unsupervised framework for detecting early signs of illness in eldercare. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2019 Nov:1043-50, doi: 10.1109/BIBM47256.2019.8983010.

[11] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med. 2004 Sep;256(3):183-94, doi: 10.1111/j.1365-2796.2004.01388.x.

[12] Edition F. Diagnostic and statistical manual of mental disorders. Am Psychiatric Assoc. 2013;21(21):591-643.

[13] Ibrahim OA, Fu S, Vassilaki M, Mielke MM, St Sauver J, Petersen RC, Sohn S. Detection of dementia signals from longitudinal clinical visits using one-class classification. 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). IEEE. 2022 Jun;211-6, doi: 10.1109/ICHI54592.2022.00040.