Mild Cognitive Impairment: Data-Driven Prediction, Risk Factors, and Workup

Sajjad Fouladvand, PhD[1], Morteza Noshad, PhD[1], Mary Kane Goldstein, MD[3], VJ Periyakoil, MD[4], Jonathan H Chen, MD, PhD[1,2,5]
[1]Stanford Center for Biomedical Informatics Research; [2]Division of Hospital Medicine; [3]Health Policy/Center for Primary Care and Outcomes Research; [4]Department of Medicine; [5]Clinical Excellence Research Center; Stanford University, Stanford, CA, USA

Abstract

Over 78 million people will suffer from dementia by 2030, emphasizing the need for early identification of patients with mild cognitive impairment (MCI) at risk, and personalized clinical evaluation steps to diagnose potentially reversible causes. Here, we leverage real-world electronic health records in the observational medical outcomes partnership (OMOP) data model to develop machine learning models to predict MCI up to a year in advance of recorded diagnosis. Our experimental results with logistic regression, random forest, and xgboost models trained and evaluated on more than 531K patient visits show random forest model can predict MCI onset with ROC-AUC of 68.2±0.7. We identify the clinical factors mentioned in clinician notes that are most predictive of MCI. Using similar association mining techniques, we develop a data-driven list of clinical procedures commonly ordered in the workup of MCI cases, that could be used as a basis for guidelines and clinical order set templates.

Introduction

Dementia is one of the major causes of mortality and morbidity in older people worldwide and it is estimated that 78 million people will be suffering from some form of dementia by the end of this decade (1), placing a tremendous burden on patients, their families, and health care systems. An important risk factor for dementia is mild cognitive impairment (MCI). Identifying early symptoms of MCI and recommending appropriate diagnostic procedures for patients at the risk of developing MCI is thus crucial. Even if there are currently limited clinical interventions known to effectively alter the course of MCI and dementia, identifying patients at risk would allow for targeted recruitment into clinical trials to study developing interventions that may only be effective in early phases of disease. Learning and disseminating personalized diagnostic evaluation steps offers an immediately practical step optimize timely diagnostic workup of MCI cases, including prompt identification of potentially reversible causes (e.g., endocrine, nutritional, and infectious).

MCI is mainly characterized by minor memory impairment (2) and is formally diagnosed by evaluating individual's cognitive capabilities and clinical examination by a healthcare professional (3). Patients do not routinely screen for possible MCI and as a result are often either under-diagnosed or diagnosis is delayed until late in the illness trajectory. One solution to the lack of formal screening for MCI disease is to identify patients otherwise engaged in the health care system by creating automated tools to analyze patients' medical history and detect those at the MCI risk.

Electronic health records are a growing source of information to identify patients at risk. Early and accurate diagnosis of such diseases can be addressed using machine learning based tools and analyzing patients electronic health records (EHR) (4,5). To this end, there have been multiple attempts to predict patients with cognitive impairment using standard machine learning models such as support vector machines (SVMs), logistic regression and random forest (6–9) and public databases such as North American Alzheimer's Disease Neuroimaging Initiative (ADNI) (10) as well as the European's AddNeuroMed Study (11). SVM models have been effectively used to predict MCI using gait analysis of patients (12). Other types of healthcare data such as image-based memory test results along with patients' demographics and medical records have been used to produce MCI prediction tools using naïve bayes models (13). Deep learning models such as graph convolutional neural networks, recurrent neural networks and attention-based model for longitudinal EHR data analysis have also been used to predict MCI onset from patients EHR data as well as imaging and clinical notes data (3,14–16).

Here, we present a multi-component framework to support anticipatory MCI care including three major components: (1) MCI onset prediction, (2) MCI risk factor identification, and (3) care plan recommendation. This design is based

on the idea of an automated MCI prediction system that could screen elderly patients at their primary care visits. We train our machine learning models to predict onset of MCI within one year of patients' visits with their primary care provider for patients 65 years or older. Further, we identify MCI risk factors extracted from patients' diagnoses, medications, procedures, and demographics structured data as well as features extracted from patients unstructured clinical notes. Finally, this study provides a data-driven clinical order set draft in the form of a list of clinical procedures commonly ordered for MCI patients around the time of initial diagnosis.

Materials and Methods

Data and Cohorts

Our data consist of deidentified EHR records for patients in Stanford Healthcare from 1999 to 2022 in the OMOP data model. Our cohort include 531,387 primary care visits where the patient is 65 years or older. We denote these primary care visits as $V = \{v_1, \dots, v_i, \dots, v_n\}$, where n is the number of primary care visits (i.e., n = 531,387). Each primary care visit v_i is assigned a time t_i. For each primary care visit v_i for patient p and at visit time t_i we extracted diagnosis, medication, and procedure concepts as well as key concepts extracted from patient's clinical notes standardized through an existing natural language processing pipeline using ConTex and NegEx algorithms (17–19). These features form a tuple (t_j^i, X^i), where $X^i \in \mathbb{R}^{480}$ is a multi-hot vector representing medication, diagnosis, and procedure concepts as well as clinical notes' key concepts for patient p at or prior to t_i. Demographic information including age, sex, and race were concatenated to X^i. Sex and race were encoded using one-hot vectors. Our final dataset includes 531,387 primary care visits (i.e., sample observations) and 480 variables in total. Each primary care visit v_i was assigned with a binary label indicating if the patient has at least one MCI diagnosis code within one year follow up from the time of their primary care visit, t_i.

Table 1 describes statistics of the demographic features of the patients in our cohort. The average age of patients in MCI encounters and Non-MCI encounters were 74 years (25th and 75th percentiles = 68, 79) and 72 years (25th and 75th percentiles = 66, 77), respectively. The majority were female (59.1% among patients in MCI encounters and 56.3% in patients in Non-MCI encounters).

Table 1. Patient demographics among subsequent MCI and Non-MCI encounters.

Variable	MCI	Non-MCI
n	7,895	146,991
Age	[a]74 (68, 79)	72 (66, 77)
Female	[b]4,669 (59.1%)	82,836 (56.3%)
Race		
Asian	1,196 (15.1%)	24,494 (16.7%)
Black/African American	504 (6.4%)	7417 (5.0%)
American Indian or Alaska Native	19 (0.2%)	431 (0.3%)
Native Hawaiian or Other Pacific Islander	68 (0.9%)	920 (0.6%)
White	5,076 (64.3%)	89,883 (61.1%)
Unknown	100 (1.3%)	6085 (4.1%)
Declines to state	134 (1.7%)	3891 (2.6%)
Others	798 (10.1%)	13870 (9.4%)

[a] Numbers are in V (x, y) format, where V is the average and x and y are 25th and 75th percentile, respectively.
[b] Numbers are in N (p%) format, where N is the number of patient and p% shows the percentage in the cohort.

Predictors

Predictors include diagnosis, medication, and procedure concepts as well as concepts extracted from patient's clinical notes plus demographic features. Concepts extracted from clinical notes have already been pre-processed and provided in OMOP data model using ConTex and NegEx algorithms (17,18). We used data with primary care visit date within 2018 in our train set (our test set data includes visits in 2020 and after) and selected concepts with

highest ratio of occurrence frequency in MCI patients to occurrence frequency in Non-MCI patients. This resulted in 95 diagnosis, 87 medication, 34 procedure, and 5,698 clinical note concepts. We further filtered clinical note concepts and only included concepts with frequencies in top 5 percentile which resulted in 249 clinical note concepts. Thus, our predictor set include 465 diagnosis, medication, procedure, and clinical note concepts. Note, we concatenated 15 demographic features to this predictor set and our final predictor set includes 480 concepts.

Figure 1 describes how historical variables were extracted from structured and unstructured EHR data for one example patient. Their data were de-identified, timepoints were jittered, and only a subset of features is presented in the figure due to high volume of the data and data confidentiality. For example, memory impairment related predictors are among important predictors of MCI and have been inferred from sentences such as "X states that X occasionally forgets things" or "because X has bad memory" in patients clinical notes. Further, conditions and procedures such as hypertension, depressive disorder, and 12 lead ECG were extracted from patients historical structured data. All concepts extracted from structured and unstructured data were concertante to form $X^i \in \mathbb{R}^{465}$. Note, data after the primary care visit time were masked and not used to perform the prediction.

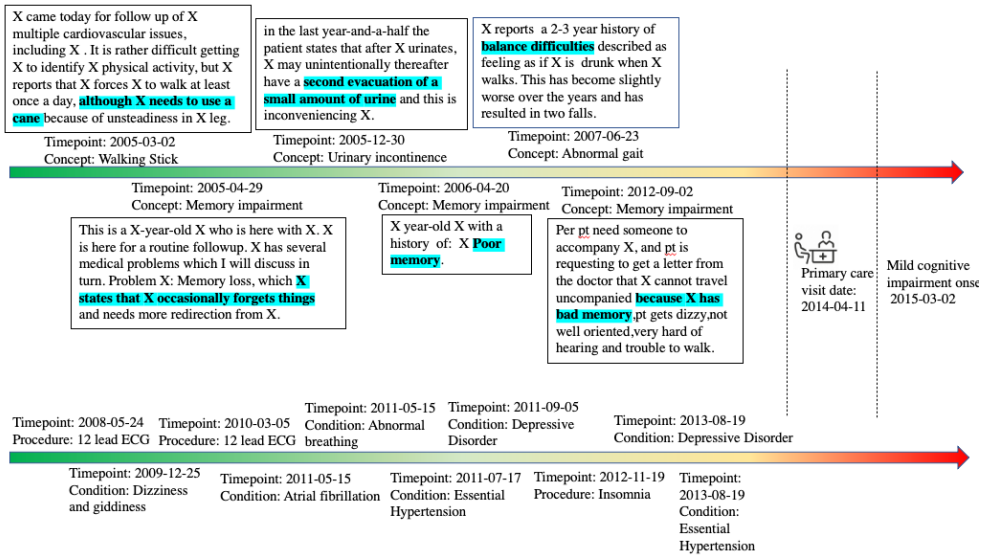

Figure 1. Predictors extracted from structured and unstructured data of an example patient. The models use these clinical factors in patient's historical data on or prior their primary care visit to predict MCI diagnosis within one year of the primary care visit.

Models

Our models include logistic regression, random forest, and xgboost. Logistic regression uses a logistic function to model the outcome probabilities of a single trial experiment (20). Random forest (21) is an ensemble model that operates by constructing a multitude of decision trees at training time and has been used extensively to solve prediction tasks in healthcare data analysis. The goal is to create a predictive model to predict Y^i given the training data set $S = \{(X^1, Y^2), \dots, (X^i, Y^i), \dots, (X^n, Y^n)\}$ of independent random variables distributed as the independent prototype pair (X^i, Y^i) (15). For each tree T_j in a forest including M trees, the predicted value for the input sample X^i is denoted by $m_n(x; \theta_j, D_n)$, where $\theta_1, \dots, \theta_M$ are independent random variables, distributed the same as a generic random variable θ. Similar to random forest, xgboost (22) is an ensample model based on decision trees. Xgboost trains tree ensemble models in an additive manner to regularize the ensemble tree objective function.

Experimental Results

Primary care visits with visit dates in or before 2019 were used as training and primary care visits in 2020 and after were used as testing set. Logistic regression, random forest and xgboost models were trained and optimized using our trainset and through cross-validation method and the models were tested using the unseen test set.

MCI Prediction

Table 2 shows the MCI prediction results on the test set using logistic regression, random forest and xgboost models. The random forest model has a higher ROC-AUC (68.2%±0.7%) than xgboost and logistic regression models on the test set.

Table 2. Performance of MCI prediction using machine learning on unseen test sets.

Model	**AUC**
Logistic Regression	57.0±1.1
Random Forest	**68.2±0.7**
XGBoost	66.8±0.8

MCI Risk Factor Identification

We utilized the trained random forest model to identify important clinical factors in discriminating MCI vs Non-MCI samples based on features' information gain (21). Figure 2 shows top-30 features detected by the random forest model as features that provide the most increased power in MCI prediction. Each bar represents one features and the lengths of the bar represents the importance score for the relevant feature. These features have been selected among 480 demographics, diagnoses, medications, procedures as well as key concepts extracted from patients' clinical notes. The type of feature (demographic, condition, procedure, medication, and clinical note) is noted within parenthesis for each feature in the figure. For example, Memory impairment (clinical notes) indicates that a concept or term related to memory impairment (see Figure 1 for some example terms) has been mentioned in patients' clinical notes prior to their current visit with their primary care provider.
To summarize Figure 2, the clinical factors include patients age at the time of their primary care visit, occurrences of keywords related to memory impairment, senility, and abnormal gait in patients' clinical notes, as well as conditions in patients structured EHR data such as essential hypertension, depressive disorder, hyperlipidemia, malaise, and procedures such as mammography and electrocardiogram.

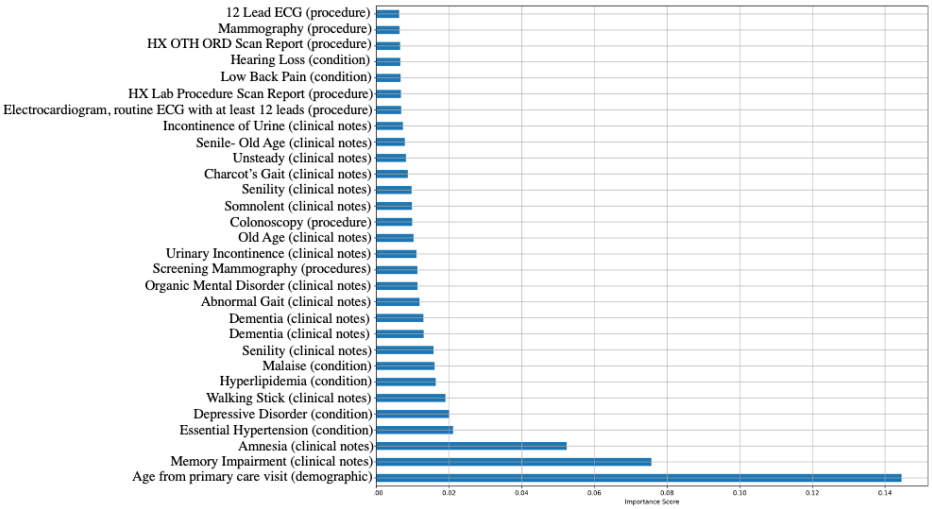

Figure 1. Top 30 Clinical factors identified by random forest model that provide increased power in discriminating MCI and Non-MCI samples.

We further investigated the associations between clinical factors in EHR data and MCI diagnosis. Table 3 described clinical factors that were presented in MCI patients EHR records more compared to Non-MCI patients. All primary care visits within 2018 where the patients were 65 years or older were used to produce Table 3. Patients were followed up for one year from their primary care visit and labeled with MCI if at least one MCI diagnosis was found. Each row in Table 3 represents one concept extracted from patients structured EHR or clinical notes data. Num MCI shows the number and percentage of MCI patients with at least one occurrence of the relevant concept in their historical data prior to their primary care visit. Num Non-MCI shows the same information for Non-MCI patients' cohort. Frequency Ratio shows the ratio of prevalence of each concept in MCI cohort to Non-MCI cohort. For example, Amnesia, which is a concept extracted from patients' clinical notes, was found in 0.22% of MCI patients clinical notes while only in 0.06% of Non-MCI patients notes.
In summary, occurrence of keywords and terms related to amnesia and memory impairment related, injury to head, gait and activity daily living are more prevalent in MCI patients' clinical notes compared to Non-MCI patients. These results are consistent with the clinical factors identified by random forest in Figure 2. Risk factors identified in this study can inform screening criteria and areas of investigation for identifying patients at risk for MCI.

Table 3. Prevalence of clinical factors in MCI patients versus Non-MCI patients. This table is sorted based on the ratio of prevalence of each concept in MCI patients to the Non-MCI patients.

Concept	**Num. MCI[a]**	**Num. non MCI[a]**	**Frequency Ratio[b]**
Amnesia	210 (0.22%)	2917 (0.06%)	3.84
Memory impairment	261 (0.27%)	4689 (0.09%)	2.97
Charcot's gait	169 (0.17%)	3718 (0.07%)	2.42
Historian	128 (0.13%)	2823 (0.05%)	2.42
Head injury	157 (0.16%)	3489 (0.07%)	2.40
MRI of brain and brain stem	177 (0.18%)	3936 (0.08%)	2.40
Intracranial	145 (0.15%)	3232 (0.06%)	2.39
Level of consciousness	151 (0.16%)	3450 (0.07%)	2.33
Injury of head	193 (0.2%)	4424 (0.09%)	2.33
Unsteady	131 (0.14%)	3004 (0.06%)	2.33
Traumatic AND/OR non-traumatic brain injury	189 (0.2%)	4389 (0.09%)	2.30
Restless	119 (0.12%)	2791 (0.05%)	2.27
polyethylene glycols	123 (0.13%)	2892 (0.06%)	2.27
Home health care	169 (0.17%)	4080 (0.08%)	2.21
Mastoid	209 (0.22%)	5092 (0.1%)	2.19
Agitation	117 (0.12%)	2853 (0.06%)	2.19
Toileting	160 (0.17%)	3907 (0.08%)	2.18

[a] The numbers are number of patients (percentage). [b] Numbers show the ratio between percentage of patients in MCI cohort to that percentage in Non-MCI cohort.

Procedure Recommendation

Identifying patients at the risk of MCI is an important step in initiating necessary care plans. Here, we analyzed the MCI patient data around the procedures commonly ordered with within one year from their visit with their primary care provider leading up to the eventual MCI diagnosis. Table 4 shows a list of procedures that have been ordered more for MCI patients compared to Non-MCI patients. Num MCI for shows the number and percentage of patients

who had that relevant procedure within one year of their visit with their primary care provider and Num Non-MCI shows the same information for Non-MCI cohort. Frequency Ratio shows the ratio of prevalence of the procedures in MCI cohort to Non-MCI cohort.

In summary, MRI and imaging of head/brain are procedures that were ordered for patients who were subsequently diagnosed with MCI more frequently compared to patients who did not have an MCI diagnosis within a year. Other key procedures associated with the workup towards an eventual MCI diagnosis include physical therapy, home care plans as well as assessment of cardiac health procedures. The information presented in Table 4 offers a data-driven draft for a clinical order set guide of suggested procedures and reports when considering MCI risk and diagnostic evaluation.

Table 4. Frequency of procedures ordered for patients with subsequent MCI diagnosis compared to Non-MCI patients.

Concept	**Num MCI**	**Num Non-MCI**	**Freq Ratio**
Magnetic resonance (e.g., proton) imaging, brain (including brain stem); without contrast material	1073 (13.5%)	1525 (1%)	13.4
Computed tomography, head or brain; without contrast material	953 (12%)	2050 (1.3%)	8.9
Physical therapy evaluation: moderate complexity, requiring these components: A history of present problem with 1-2 personal factors and/or comorbidities that impact the plan of care; An examination of body systems using standardized tests and measures in	565 (7.1%)	2128 (1.4%)	5.0
Therapeutic procedure, 1 or more areas, each 15 minutes; gait training (includes stair climbing)	612 (7.7%)	2566 (1.7%)	4.6
Self-care/home management training (eg, activities of daily living (ADL) and compensatory training, meal preparation, safety procedures, and instructions in use of assistive technology devices/adaptive equipment) direct one-on-one contact, each 15 minutes	740 (9.3%)	3136 (2.1%)	4.5
Radiologic examination, chest; single view	688 (8.6%)	3024 (2%)	4.3
Therapeutic activities, direct (one-on-one) patient contact (use of dynamic activities to improve functional performance), each 15 minutes	1028 (12.9%)	4602 (3%)	4.3
Physical therapy evaluation	416 (5.2%)	1874 (1.2%)	4.2
Assessment of cardiac status using monitoring device	598 (7.5%)	2744 (1.8%)	4.2
Determination of refractive state	324 (4.1%)	1526 (1%)	4.1
Ophthalmological services: medical examination and evaluation, with initiation or continuation of diagnostic and treatment program; comprehensive, established patient, 1 or more visits	463 (5.8%)	2200 (1.4%)	4.0
Intravenous infusion, for therapy, prophylaxis, or diagnosis (specify substance or drug); initial, up to 1 hour	624 (7.8%)	3161 (2.1%)	3.8
Radiologic examination, chest; 2 views	594 (7.5%)	3370 (2.2%)	3.4
Electrocardiogram, routine ECG with at least 12 leads; tracing only, without interpretation and report	2092 (26.3%)	12021 (7.9%)	3.3

Discussion

In this work we analyzed Stanford healthcare EHR data of over 20 years in OMOP common data model from more than 531K primary care visits including 8,990 primary care visits leading to MCI diagnosis within one year, and 522,478 primary care visits with no incoming MCI diagnosis within one year of primary care visit time. We extracted 480 concepts from patients' structured diagnosis, medication, procedure, and demographic records as well as from patients' clinical notes. These data were used to train MCI prediction tools for patients 65 years or older using machine learning models. Random forest model could predict MCI effectively (ROC-AUC=68.2±0.7).

Clinical factors associated with higher risk of MCI were analyzed using random forest and statistical analysis. Overall, keywords in clinical notes related to memory impairment, abnormal gait, head injury, and activity daily living were found more frequently in MCI patients compared to Non-MCI patients. This shows that patients historical EHR data contain important clinical factors that can be used to identify patients at MCI risk and therefore the potentials to create automated machine learning systems for MCI screening among elderly patients during their primary care visits.

Even though there is only limited standard treatment for MCI and Alzheimer's diseases, identifying patients at the risk of developing MCI is an essential step. This can be used to inform and educate patients about their incoming conditions and can also facilitate decision making for primary care providers when requesting consultation or referring patients to specialty care providers. In addition to identifying MCI risk factors and producing MCI predictive models, in this study we further followed up patients for one year after their primary care visit and identified a list of procedures that have been ordered for patients at higher risk of MCI frequently. These procedures for patients at the risk of MCI include brain and head MR imaging procedures as well as physical therapy and setting up home care plans. This could provide a draft framework to aid primary care providers to include necessary information when requesting consultation with specialists and can help specialists with decision making ordering appropriate procedures for patients at the risk of MCI.

Limitations in the study include that, although we thoroughly tested the models using randomly selected held-out test sets, more study is needed to test our models' performances in a clinical care environment. Here, the models are implemented using data in the OMOP common data model and therefore the models can be feasibly used in any other healthcare system where the data is in OMOP format. Further, in this study patients historical structured and unstructured EHR were used to implement our framework. However, other resources such as patients brain MRI images may provide more insights and provide increased power in predicting MCI.

In conclusion, we presented a multi-component framework including (1) MCI prediction, (2) MCI risk factor identification, and (3) recommendation for diagnostic workup steps. While there are already clinical tools developed to diagnose MCI such as the Montreal cognitive assessment (MoCA) tool (23), these tools are not intended to be *predictive* of future diagnosis, and have typically been produced using small and underrepresented sample size (e.g. only 94 MCI patients data have been used to create MoCA). Further, these tools need to be administered by healthcare professionals with the patient in front of them (or online), which limits their applicability and feasibility. Automated machine learning based tools may be more feasible to use in large population screening for MCI risk.

Acknowledgement

This research was supported in part by Stanford Aging and Ethnogeriatrics (SAGE) Research Center under NIH/NIA grant P30AG059307 and NIH/National Library of Medicine via Award R56LM013365. This research used data or services provided by STARR, Sanford medicine Research data Repository.

References

1. Gauthier S, Rosa-Neto P, Morais JA, Webster C. World Alzheimer Report 2021: Journey through the diagnosis of dementia. London: Alzheimer's Disease International; 2021.

2. Robinson L, Tang E, Taylor JP. Dementia: timely diagnosis and early intervention. BMJ [Internet]. 2015 Jun 16 [cited 2022 Mar 2];350. Available from: https://www.bmj.com/content/350/bmj.h3029

3. Fouladvand S, Mielke MM, Vassilaki M, Sauver JSt, Petersen RC, Sohn S. Deep Learning Prediction of Mild Cognitive Impairment using Electronic Health Records. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2019 Nov;2019:799–806.

4. Goudarzvand S, Sauver JSt, Mielke MM, Takahashi PY, Sohn S. Analyzing Early Signals of Older Adult Cognitive Impairment in Electronic Health Records. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2018. p. 1636–40.

5. Goudarzvand S, St Sauver J, Mielke MM, Takahashi PY, Lee Y, Sohn S. Early temporal characteristics of elderly patient cognitive impairment in electronic health records. BMC Med Inform Decis Mak. 2019 Aug 8;19(Suppl 4):149.

6. Zhang R, Simon G, Yu F. Advancing Alzheimer's research: A review of big data promises. Int J Med Inform. 2017 Oct;106:48–56.

7. Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW, et al. Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging. 2010 Aug;31(8):1429–42.

8. Li M, Oishi K, He X, Qin Y, Gao F, Mori S, et al. An Efficient Approach for Differentiating Alzheimer's Disease from Normal Elderly Based on Multicenter MRI Using Gray-Level Invariant Features. PLOS ONE. 2014 Aug 20;9(8):e105563.

9. van Gils M, Koikkalainen J, Mattila J, Herukka S, Lotjonen J, Soininen H. Discovery and use of efficient biomarkers for objective disease state assessment in Alzheimer's disease. Annu Int Conf IEEE Eng Med Biol Soc. 2010;2010:2886–9.

10. Alzheimer's Disease Neuroimaging Initiative [Internet]. [cited 2022 Mar 3]. Available from: http://adni.loni.usc.edu/

11. Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, et al. AddNeuroMed--the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. Ann N Y Acad Sci. 2009 Oct;1180:36–46.

12. Chen PH, Lien CW, Wu WC, Lee LS, Shaw JS. Gait-Based Machine Learning for Classifying Patients with Different Types of Mild Cognitive Impairment. J Med Syst. 2020 Apr 23;44(6):107.

13. Bergeron MF, Landset S, Zhou X, Ding T, Khoshgoftaar TM, Zhao F, et al. Utility of MemTrax and Machine Learning Modeling in Classification of Mild Cognitive Impairment. Journal of Alzheimer's Disease. 2020 Jan 1;77(4):1545–58.

14. Zhao X, Zhou F, Ou-Yang L, Wang T, Lei B. Graph Convolutional Network Analysis for Mild Cognitive Impairment Prediction. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019. p. 1598–601.

15. Wang L, Laurentiev J, Yang J, Lo YC, Amariglio RE, Blacker D, et al. Development and Validation of a Deep Learning Model for Earlier Detection of Cognitive Decline From Clinical Notes in Electronic Health Records. JAMA Network Open. 2021 Nov 18;4(11):e2135174.

16. Fu S, Ibrahim OA, Wang Y, Vassilaki M, Petersen RC, Mielke MM, et al. Prediction of Incident Dementia Using Patient Temporal Health Status. Stud Health Technol Inform. 2022 Jun 6;290:757–61.

17. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. Journal of Biomedical Informatics. 2001 Oct 1;34(5):301–10.

18. Chapman W, Dowling J, Chu D. ConText: An Algorithm for Identifying Contextual Features from Clinical Text. In: Biological, translational, and clinical language processing [Internet]. Prague, Czech Republic: Association for Computational Linguistics; 2007 [cited 2022 Sep 14]. p. 81–8. Available from: https://aclanthology.org/W07-1011

19. OMOP NOTE_NLP Table [Internet]. [cited 2022 Sep 14]. Available from: https://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:note_nlp

20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12(85):2825–30.

21. Breiman L. Random Forests. Machine Learning. 2001 Oct 1;45(1):5–32.

22. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2022 Mar 3]. p. 785–94. (KDD '16). Available from: https://doi.org/10.1145/2939672.2939785

23. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. Journal of the American Geriatrics Society. 2005;53(4):695–9.