Mild Cognitive Impairment Prediction and Diagnostic Procedure Recommendation using Machine Learning

Sajjad Fouladvand, PhD1, Morteza Noshad, PhD1, Jonathan H Chen, MD, PhD1,2

1Stanford Center for Biomedical Informatics, Stanford University, Stanford, CA, USA;

2Division of Hospital Medicine, Stanford University, Stanford, CA, USA;

Abstract

Without proper prevention and treatment more than 78 million people will be suffering from a type of dementia by 2030. One possible solution is to pivot public health toward identifying patients at the risk of dementia in early stages. In this study, we leverage real-world electronic health records and propose an automated machine learning based framework to predict mild cognitive impairment (MCI) as an important risk factor for dementia. Further, our framework includes a recommender system suggesting necessary diagnostic procedures for patients at the risk of MCI. Our experimental results with logistic regression, random forest and XGBoost models trained and tested on more than 4.2K MCI patients and more than 82K cognitively unimpaired patients show that XGBoost model can predict MCI one year before onset of the disease with ROCAUC of 0.683 and recommend necessary procedures for MCI patients effectively with averagely ROCAUC of 0.784.

# **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 dealing with a type of dementia by the end of this decade (1), placing a tremendous burden on patients 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 a crucial task to help the aging population with their health needs. Early prediction for mild cognitive impairment could afford the option for early intervention. Even if there are limited clinical interventions known to effectively alter the course of MCI and dementia, identifying patients at risk would allow for targeted recruitment of such populations into clinical trials to study developing interventions. Learning and disseminating personalized diagnostic evaluation steps is a further essential process to optimize timely diagnosis of MCI cases, including evaluation and exclusion vs. diagnosis and treatment of potentially reversible causes (e.g., endocrine, nutritional, infectious, and other etiologies).

MCI is mainly characterized by having minor memory impairment (2) and is formally diagnosed by evaluating individual’s cognitive as well as performing clinical examination by a healthcare professional (3). However, in many cases patients don’t routinely get screened for possible MCI risks and as a result MCI diseases are often either under-diagnosed or the diagnosis occurs 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 growing source of information that can be harnessed to identify patients at risk of MCI. 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 mostly using 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 information 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). More sophisticated deep learning models such as graph convolutional neural networks and recurrent neural networks have also been used to predict MCI onset from patients HER data as well as imaging and clinical notes data (3,14,15).

However, using machine learning models in MCI prediction has remained largely unexplored mostly due to the lack of real-world data on MCI diagnosis. Further, previous works have been mainly centered on MCI prediction only. In this study, we use Stanford HER data and seek to implement machine learning models for improving dementia diagnosis through the following objectives:

* To determine if machine learning models trained on patients’ electronic health records can effectively predict mild cognitive impairment.
* To determine if machine learning models can recommend necessary diagnostic procedures for patients at the risk of developing mild cognitive impairment.

# **Materials and Methods**

The proposed framework in this study includes two main components: 1) MCI onset prediction using machine learning, 2) necessary diagnostic procedure recommendation for patients at the risk of MCI. Figure 1 shows the general schema of the cohort extraction and model trainings. Components of this framework are described in detail in the following sub-sections.

Diagram

Description automatically generated

**Figure 1.** General architecture of our proposed framework for MCI prediction and MCI necessary procedures recommendation. (a) MCI patients were identified, and case and control cohorts were created. (b) Diagnosis, procedure, medication and demographic features were used to create training and testing data for both MCI prediction and procedure recommender system. (c) Prediction window for MCI prediction is 1 year and for procedure recommendation is 2 months. Index date for MCI patients is first MCI diagnosis date and for controls is 1 year prior to their last record in the data.

## Data and Cohorts

Our data consist of deidentified EHR records for patients in Stanford Healthcare from 2000 to 2020. Cases include 50 years or older patients with at least one ICD diagnosis of MCI (ICD10s=G31.84, F09 and ICD9s=331.83, 294.9) and two years of data availability. Controls are 50 years or older patients with no ICD diagnosis in their records and two years of data availability. Data availability is the duration time between patients first record date to the index date where the index date for cases is the first MCI diagnosis date and for the controls is one year before their last record in the data. Controls are matched with cases based on age and gender to create a balanced train set including 5,694 patients (2,840 cases and 2,853 controls) and a test set including 81,078 patients (1,227 cases and 79,851 controls). Note, our models are trained using balanced training dataset to predict MCI one year prior to the index date for cases and controls; however, our testing experiments include imbalanced scenarios, and the MCI prediction models are tested on the unseen (held out) imbalanced test set.

## Data Pre-processing

Dataset includes n patients . For a patient their diagnosis, medication, procedure and demographic records are in a format, where is the timestamp for and indicate medications, diagnosis or procedures codes for this patient at . We converted this longitudinal format to a stationary format for each patient by computing the frequencies of the features across all patient’s timestamps . In fact, we created a stationary dataset , where is a 1D vector indicating frequencies of all medication, diagnosis and procedure features concatenated with patient’s demographic features, age, gender and race, and is the target variable. For the MCI prediction task is a binary variable indicating if will develop MCI in a year or not, and for the MCI diagnostic procedure recommendation system is a multi-hot vector indicating the recommended procedure for . Note, the prediction window for MCI prediction task is one year and for the procedure recommendation task is two months. Table 1 and Table 2 respectively describe statistics of the demographic features and prevalence of the top-5 diagnoses, procedures and medications selected by random forest model. Demographic features in terms of age and gender were similar among cases and controls as we matched based on these variables. The average age of cases and controls were 74.36 years (25th and 75th percentiles = 68, 83) and 76.33 years (25th and 75th percentiles = 70, 84), respectively. The majority were female (54.41% among both cases and controls), cases had a higher number of patients with races as black as well as white in cases than in controls, and lower number of Asian patients in cases than controls.

**Table 1**. Patient demographics among Cases and Controls.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Case** | **Control** |
| Age | a74.36 (68, 83) | 76.33 (70, 84) |
| Female | 2,213 (54.41%) | 2,213 (54.41%) |
| Race | | |
| Asian | b458 (11.26%) | 595 (14.63%) |
| Black | 225 (5.53%) | 150 (3.69%) |
| Native American | 14 (0.34%) | 13 (0.32%) |
| Pacific Islander | 32 (0.79%) | 36 (0.78%) |
| White | 2,785 (68.48%) | 2,446 (60.14%) |
| Unknown | 120 (2.95%) | 337 (8.29%) |
| Other | 433 (10.64%) | 494 (12.15%) |

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 in this study include medications, diagnosis, procedure and demographic features. Note, we grouped medications and diagnoses using pharmaceutical class and clinical classification software (CCS) codes, respectively. We included top-100 most frequent medications, diagnoses and procedures codes as well as the demographic features, age, sex and race. The feature set for patient , includes 300 medication, diagnosis and procedure codes plus 3 demographic features. This feature space size further reduced to 30 features using random forest algorithm and used to train machine learning algorithms.

**Targets**

Target variable for the first task, MCI prediction, is a binary variable indicating if a patient will develop MCI one year from the time of prediction or not. Our second task includes predicting necessary procedures for patients at the risk of MCI. Top-100 most frequent procedures among MCI patients were selected reviewed by healthcare professionals in our team. The final target variable set include 46 procedures presented in Table 3. Note, the models were trained using patients’ diagnosis, procedure, medication and demographic records up to two months (prediction window in our first task is 12 and in our second task is 2 months) prior to the MCI onset time for each patient . The trained models then predict necessary diagnostic procedures in a form of a multi-hot vector with 46 elements each representing one of the procedures in Table 3.

**Table 2**. Patient demographics among Cases and Controls.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Case** | **Control** |
| **Diagnoses** | | |
| Nervous system signs and symptoms | 1,459 (31.46%) | 390 (8.41%) |
| Exposure, encounters, screening or contact with infectious disease | 1,655 (35.69%) | 625 (13.43%) |
| Medical examination/evaluation | 2,233 (48.16%) | 993 (21.41%) |
| Musculoskeletal pain, not low back pain | 1,944 (41.92%) | 833 (17.96%) |
| Sleep wake disorders | 1,346 (29.02%) | 492 (10.61%) |
| **Procedure** | | |
| Metabolic panel, comprehensive | 2,844 (61.33%) | 1,219 (26.29%) |
| TSH | 2,196 (47.36%) | 756 (16.30%) |
| CBC with differential | 2,860 (61.68%) | 1351 (29.13%)) |
| External lab results | 42 (0.90%) | 572 (12.33%) |
| Other order scanned report | 2,123 (45.78%) | 925 (19.95%) |
| **Medications** | | |
| Anticonvulsants | 1,204 (25.96%) | 408 (8.80%) |
| Antihyperlipidemic-hmgcoa reductase inhib(statins) | 1,935 (41.73%) | 1,063 (22.92%) |
| Selective serotonin reuptake inhibitor (ssris) | 915 (19.73%) | 263 (5.67%) |
| Opioid analgesic and non-salicylate analgesics | 1,821 (39.27%) | 839 (18.09%) |
| Antiemetic/antivertigo agents | 1,884 (40.63%) | 861 (18.57%) |

Numbers are in N (p%) format, where N is the number of patient and p% shows the percentage in the cohort.

**Models**

The predictors were used to train logistic regression, random forest and xgboost models. Logistic regression uses a logistic function to model the outcome probabilities of a single trial experiment(16). Random forest (17) 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 given the training data set of independent random variables distributed as the independent prototype pair . For each tree in a forest including trees, the predicted value for the input sample is denoted by , where are independent random variables, distributed the same as a generic random variable . Similar to random forest, XGBoost (18) is an ensample model based on decision trees. XGBoost trains tree ensemble models in an additive manner to greedily and efficiently regularize the ensemble tree objective function.

These models were trained using the train set including 5,694 (2,840 cases and 2,853 controls) patients’ data and a randomized parameter search with a 5-fold cross validation. The trained models were tested using a randomly selected held out test set including 81,078 (1,227 cases and 79,851 controls) patients’ data. The optimum logistic regression has a sag as it’s solver with . The optimum random forest model has 1600 estimators with maximum depth of 110. The optimum values for the number of estimators, maximum depth, learning rate and gamma for the xgboost model used are 1600, 16, 0.1, and 10. The optimized models were assessed using accuracy, precision, recall, F1-score and ROCAUC. ﻿Accuracy is the ratio of correct MCI and CU predictions by the models to the total number of samples in the test set. Precision shows how precise a model is when predicting a sample as MCI, and recall measures the performance of the model in retrieving all MCI samples in the test data. F1 is the harmonic mean of precision and recall and ROCAUC shows the model's performance across different decision thresholds.

**Table 3.** List of the 90 diagnosis procedures predicted by the recommender system.

|  |  |
| --- | --- |
| Aemoglobin ALC | Urinalysis, screen for culture |
| Hepatic function panel a | Vitamin b12 |
| STAT creatinine | CT head |
| Lipid panel, non-fasting patient | XR chest 1v |
| Lipid panel, fasting patient | XR chest 2v |
| Magnesium, serum/plasma | MRI brain wo contrast |
| Metabolic panel, basic | Blood culture (aerobic & anaerobic bottle) |
| Metabolic panel, comprehensive | C- reactive protein |
| OT evaluate and treat | Referral to neurology |
| Phosphorus, serum/plasma | Eval/mgmt of new patient level 5 |
| Prothrombin time | Eval/mgmt of est patient level 4 |
| PT evaluate and treat | Eval/mgmt of est patient level 5 |
| PTT partial thromboplastin time | Autonomic testing- cardiovascular |
| OT ongoing treatment | PT ongoing treatment |
| Foley retention catheter | Vitamin d, 25-hydroxyvitamin |
| Sedimentation rate (esr) | TSH w/ reflex ft4 |
| Referral to physical therapy | Miscellaneous processing |
| Interagency referral to home health/addendum to certification | CBC w/o diff |
| T4, free | XR chest 2 views |
| Referral to neuropsychology | Creatinine, serum/plasma |
| Troponin i | ECG 12-lead |
| TSH | Echo - transthoracic echo |
| Urinalysis, complete | Ferritin |

# **Experimental Results**

Table 4 shows the MCI prediction results at one year before the disease onset using logistic regression, random forest and xgboost models. The xgboost model has a slightly higher AUC (=0.683) than random forest and logistic regression on the test set. The recall for this model is 0.733, showing that xgboost could correctly detect and predict majority of MCI patients in the test set at one year before the disease onset. However, precision scores for the xgboost and the other two models are significantly low. This is expected and because of the high control/case ratio in our test set (control/case ratio in the test set is over 65). Precision on a balanced unseen test set including 1,227 cases and 1214 controls for logistic regression, random forest and xgboost were 0.722, 0.762 and 0.780 respectively. All in all, both xgboost and random forest models have reasonably high ROCAUC, which means these models are flexible and adjustable based on clinic care settings. They also perform well in detecting majority of patients at the risk of MCI one year before the disease onset.

**Table 4.** Performance of MCI prediction using machine learning on unseen test sets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-score** | **AUC** |
| Logistic Regression | 0.023 | 0.636 | 0.045 | 0.676 |
| Random Forest | 0.020 | 0.743 | 0.040 | 0.680 |
| XGBoost | 0.022 | 0.733 | 0.043 | 0.683 |

Our second objective in this study is to develop a machine learning model to recommend necessary procedures for patients at the risk of MCI. To this end, we used the MCI cohort (cases) and trained logistic regression, random forest and xgboost models to predict necessary procedures among a list of 46 common diagnostic procedures presented in Table 3. Precision, recall, f1-score and AUC for these experiments are presented in Table 5. Note, we only used random forest and xgboost as they both performed better on our first task, MCI prediction, than logistic regression. Further, we used one-versus-rest micro averaged precision, recall, F1-score and ROCAUC as the target is a multi-hot vector predicting 46 different diagnostic procedures. Micro averaged assessments were computed by counting the total true positives, false negatives and false positives. xgboost model had slightly higher ROCAUC compared to random forest (0.784 vs 0.772). XGBoost also performed better in terms of recall than random forest (0.097 vs 0.062). However, random forest is more precise when flagging a patient as high MCI risk (0.598 vs 0.568). Note, Table 5 also shows the results for random recommendation in which necessary procedures are recommended randomly. The ROCAUC for such system is significantly lower than both random forest and xgboost models.

Table 5. Machine learning prediction performance in diagnostic procedure recommendation for MCI patients.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Micro averaged precision | Micro averaged recall | Micro averaged f1-score | Micro averaged AUC |
| Random recommendation | 0.118 | 0.500 | 0.191 | 0.499 |
| Random forest | 0.598 | 0.062 | 0.113 | 0.772 |
| XGBoost | 0.568 | 0.097 | 0.165 | 0.784 |

**Conclusions**

In this work we analyzed Stanford healthcare EHR data from more than 4.2K MCI and 82K cognitively unimpaired patients over 20 years and created machine learning based models to predict MCI onset as well as creating diagnostic procedure recommendation systems. XGBoost model could predict MCI reasonably effectively (ROCAUC=0.683) and produce a rather effective diagnostic procedure recommendation system for patients at the risk of MCI (ROCAUC=0.784). These machine learning models trained using thousands of patients’ records can automatically screen patients EHR data and detect those at the risk. While there are already a few clinical tools developed to detect MCI such as the Montreal cognitive assessment (MoCA) tool (19), these tools 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. Our automated machine learning based tools may be more effective and feasible to use in predicting MCI and recommending diagnostic procedures in clinical care than traditional tools.

There are a few limitations that need to be considered before using or deploying our proposed framework in this study. Although we thoroughly tested the models using randomly selected held-out test sets, more prospective study is needed to test our models’ performances in a clinical care environment. Testing both the MCI prediction and diagnostic procedure recommender system in real-time and in a real-world environment can re-assure the generalizability of our models. Further, in this study patients EHR data including their past diagnosis, medication and procedure records were used in a structured format. However, other resources such as patients brain MRI images as well as patients’ clinical notes may provide more insights and provide increased power in predicting MCI.

# **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 J-P. 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 Y-Y, 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 P-H, Lien C-W, Wu W-C, Lee L-S, Shaw J-S. 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 Y-C, 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. 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.

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

18. 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

19. 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.