PND98

Patterns and Prediction for Cognitive Decline in Alzheimer's Patients as Assessed by the Mini-Mental Status Exam in an Ambulatory Electronic Medical Record

Buderi R[1], Ransom J[1], Shilnikova A[1], Rusli E[1], Ahmed R[1], Galaznik A[1], Lempernesse B[1], Berger M[1]

[1]Medidata Solutions, Boston, MA, USA

Background

- Disease progression rates, including cognitive decline among patients with Alzheimer's Disease (AD), are variable. The ability to predict progression rates would aid clinicians, patients, and patients' families in decision-making.
- Patient-reported outcomes (PRO) and observer-reported outcomes are important determinants of therapeutic efficacy, safety, and quality of life in AD.[1] Implementation of these measures into clinical workflows, however, can be challenging.
- The Mini-Mental State Exam (MMSE) is a widely used, observer-administered screening tool for cognitive impairment in dementia; it includes tests of orientation, attention, memory, language, and visual-spatial skills.[2]
- Previous studies have used MMSE scores as a measure of AD progression rate.[3]
- Machine-learning approaches are being developed for predicting cognitive decline. Feature engineering, the process of deriving informative features from data, is an important part of machine learning.[4]
- MMSE combined with automated approaches to identify and incrementally add factors to predictive models of AD progression has not been reported previously.

Objectives

- To identify predictors of cognitive decline, as estimated by change in MMSE
- To quantitatively assess their impact on predictive capability of automated feature selection and model performance

Methods

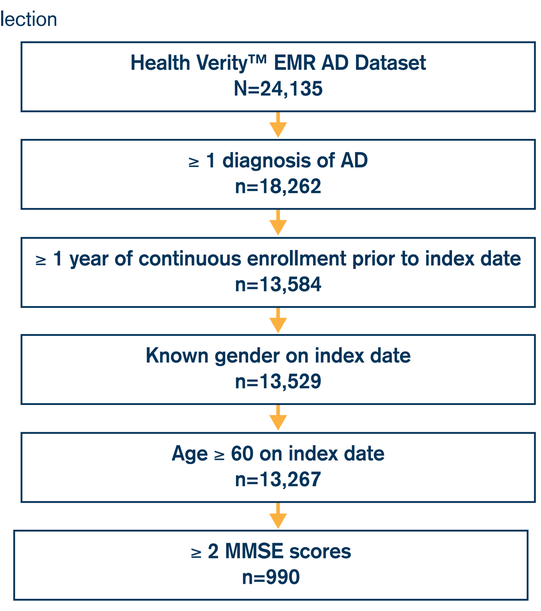
Data Source

- Patient data were extracted from HealthVerity™ Marketplace longitudinal ambulatory electronic medical record (EMR) dataset between Jan 1, 2014 and Dec 31, 2018.
- HealthVerity™ has the most complete coverage of United States healthcare, consumer, and purchase data, with access to over 330 million patients and 30 billion transactions.[5]
  - HealthVerity™ Marketplace is a self-service cloud solution allowing users to build patient and provider cohorts from more than 60 unique data sources.[6]
- Dataset was converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model, version 5.
- Analyses were conducted in SHYFT Quantum v7.1.1.

Study Design

- Inclusion/exclusion criteria (Figure 1):
  - Patients with ≥ 1 diagnoses of AD
  - ≥ 2 MMSE tests with known PRO results ≥ 1 day apart from each other, with the earlier of the 2 test records used as the index date
  - ≥ 1 year continuous enrollment prior to the index date
  - Known gender on index date
  - Age ≥ 60 years on index date

Figure 1: Study Cohort Selection



Study Measures

- Baseline patient characteristics: age, gender, comorbidities, frequency of visits to clinician's office, and average MMSE score
- MMSE:
  - A 30-point questionnaire used to measure cognitive impairment[7]
  - Consists of 7 categories representing different cognitive domains or functions: orientation to time, orientation to place, registration of 3 words, attention and calculation, recall of 3 words, language, and visual construction[7]
  - The maximum score is 30 points, 1 point each for a correct answer. Score of 24-30 = no cognitive impairment, score of 18-23 = mild cognitive impairment, score of 0-17 = severe cognitive impairment[7]
- MMSE change from baseline
  - MMSE change was measured within patients who had ≥ 2 observations.
  - Patients were designated as Maintained (increased or unchanged) or Declined (decreased).

Outcomes & Analyses

- Kaplan-Meier analyses were performed for MMSE change from baseline for age and visit frequency.
- Initial feature selection
  - Common OMOP vocabularies were used to derive clinical (concomitant medications, comorbidities), demographic (age, gender), and outcomes variables.
  - Key clinically relevant variables were identified based on literature and included: initial MMSE score, time between MMSE scores, age at index, dementia (F03), gender, depression (F33), Alzheimer's (G30), Anxiety (F41), bipolar disorder (F31), Parkinson's Disease (G20), and hospitalization pre-index.
- Predictive modeling approach
  - Prediction of MMSE decline was compared using multiple models, including: Linear Regression, Gaussian Naïve Bayes, K-Nearest Neighbor, Linear Discriminant Analysis, Support Vector Machine, Support Vector Machine with principal component analysis (PCA), Decision Tree Regression, Random Forest Regression, Random Forest Regression with PCA, Gradient Boosted Regression, XGBoost, Multilayer Perceptron, and Logistic Regression with PCA.
  - All modeling used open source Sci-kit Learn[8] framework in Python (v3.7) with the exception of XGBoost models, which used the open source XGBoost Python library[9]
  - Model performance was assessed by standard metrics such as Area Under the Curve (AUC), mean squared error, or mean absolute error.
    - A 3:1 training/testing split was employed using cross-validation scoring with a k-fold=6 to train the models. Predictions were made on probabilities from the classifiers without cross-validation.
    - Multi-collinearity was assessed using Pearson's Correlation Matrices.
    - For the sake of interpretability, only classification model results are shown.
    - Feature Importance was assessed for XGBoost models only.
    - Covariates were scaled using standard score scaling.[10]
- Feature selection for predictive modeling
  - Baseline prediction scores were generated using only hand-selected covariates (Table 1).
  - For each modeling exercise performed, RFE (Recursive Feature Elimination),[11] XGBoost was used to add the top 20 covariates from patients' medical histories as Boolean values (had condition, procedure, visit, or drug exposure before index date), based on 3-digit International Classification of Diseases (ICD)–10 code.
  - Predictors were added to the original variables, and models were re-run; incremental changes in Receiver Operating Characteristic (ROC) AUC were assessed.
  - Top 20 predictors were refined to include a frequency count component (number of diagnosis, procedures, exposures a patient had before index date). RFE was used again to select the top 20 frequency covariates.
  - Refined predictors were added again to the original variables, and models were re-run; incremental changes in ROC AUC were assessed.
- Multi-collinearity assessment
  - Correlation matrices were plotted to assess multi-collinearity with addition of RFE-selected incremental medical comorbidity features, followed by addition of frequency-based features (Figure 3a, 3b).
    - Features with correlation ≥ 0.7 were combined, and multi-collinearity was re-assessed (Figure 3c).

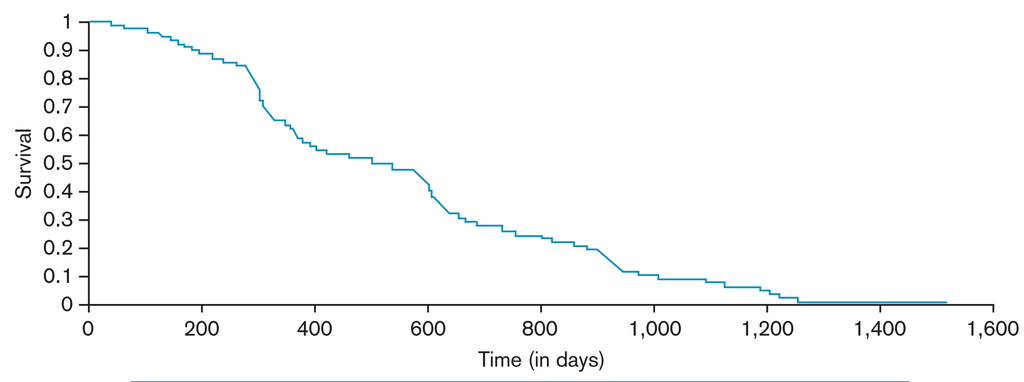
Results

Table 1: Baseline Feature Importance

Characteristic	N=990
Age	
Mean (SD)	82.8 (7.3)
Median	82
Gender	
Male, n (%)	296 (29.4)
Female, n (%)	693 (68.8)
Not reported, n (%)	1 (0.1)
Dementia pre-index, n (%)	787 (78.1)
AD pre-index, n (%)	495 (49.1)
Earliest MMSE test score	
Mean (SD)	20.1 (7.1)
Median	21
Last MMSE test score	
Mean (SD)	17.8 (7.9)
Median	19
MMSE test score change	
Mean (SD)	-2.3 (5.5)
Median	-2
Time between MMSE test scores	
Mean (SD)	509.6 (318.6)
Median	384

SD, standard deviation

Figure 2: Kaplan-Meier Curve of Time to MMSE Score Decline



Number of patients	573
Median time to MMSE decline (95% confidence interval)	580 days (499, 626)

Figure 3: Correlation Matrix

a) Correlation Matrix of Baseline and Incremental Covariates

b) Correlation Matrix of Baseline, Incremental, and Frequency-based Covariates

c) Correlation Matrix of Figure 3b Covariates with Grouping of Multi-collinear Features

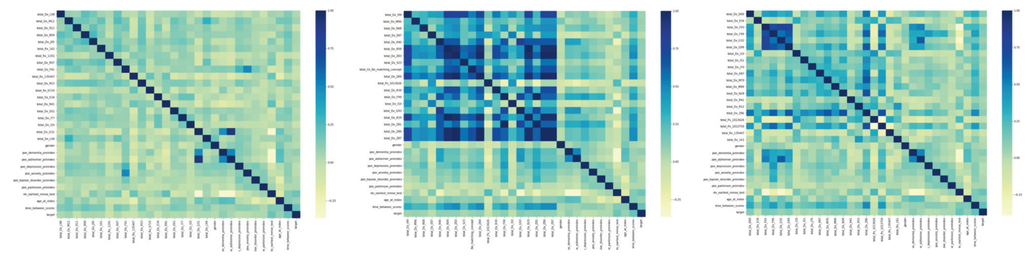


Table 2: RFE Feature Selection and Importance

Feature	Importance
Time between MMSE scores	0.24
Arthropathy (ICD-10 M12)	0.18
Dementia (ICD-10 F03)	0.15
Initial MMSE score	0.14
Donepezil	0.11
Gender	0.11
Age	0.08
Other neurodegenerative disorder (ICD-10 G32)	0
Parkinson's Disease (ICD-10 G20)	0
Bipolar disorder (ICD-10 F31)	0
Anxiety (ICD-10 F41)	0
Depression (ICD-10 F33)	0

Figure 4: ROC Curve – XGBoost with Incremental Covariate

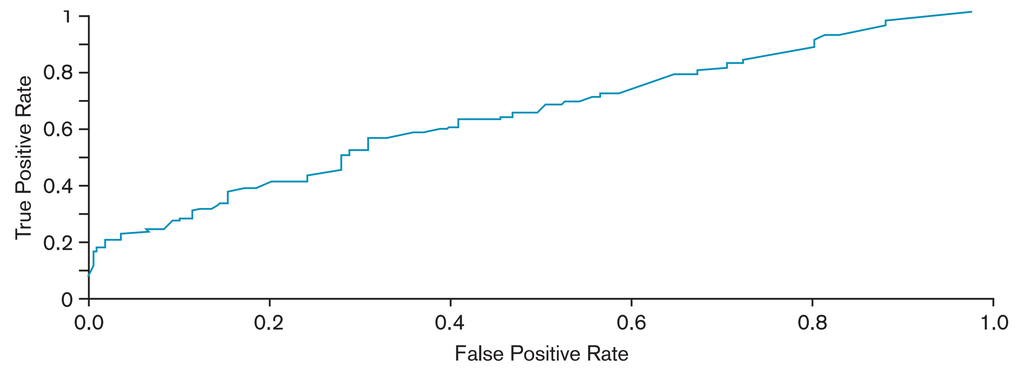


Table : Model Performance with Feature Engineering

Model	Training ROC AUC				
	Baseline - Test	With incremental features	% change from baseline	With incremental features/frequency	% change from baseline
Random Forest	0.69	0.62	-9.63%	0.548	-19.98%
Logistic Regression	0.66	0.69	3.67%	0.684	3.35%
XGBoost	0.66	0.646	-2.36%	0.601	-9.14%
Linear Discriminant Analysis	0.66	0.67	1.43%	0.674	2.07%
Support Vector Machine	0.65	0.67	3.51%	0.622	-4.02%
Gaussian Naïve Bayes	0.643	0.62	-3.62%	0.565	-12.12%
Multilayer Perceptron	0.596	0.55	-7.91%	0.500	-16.13%
Decision Tree Regression	0.575	0.59	2.20%	0.587	2.20%
Logistic Regression w/PCA	0.561	0.52	-7.00%	0.548	-2.44%
Support Vector Machine w/PCA	0.544	0.53	-1.91%	0.514	-5.56%
Random Forest w/PCA	0.539	0.50	-7.17%	0.510	-5.37%
K-Nearest Neighbors	0.533	0.58	8.18%	0.563	5.67%

Summary

- Overall, age and comorbidities of the population were consistent with prior literature describing patients with AD (Table 1).
  - MMSE score change was captured for < 10% of available patients.
  - Patients with worsening MMSE scores had more MMSE measurements (mean 5.6 records per patient) vs. patients with maintained MMSE measurements (mean 5.1 records per patient).
  - Patients with maintained MMSE score had a shorter median time between first and last test though (525 days vs. 345 days).
- The median time to MMSE decline was 580 days (Figure 2).
  - Few patients showed continued assessment after demonstrated decline, hampering the ability to assess decline after diagnosis or treatment initiation.
- Importance of 12 baseline predictor variables are presented in Table 2.
  - Among patients with ≥ 2 MMSE scores, patient demographics, pre-existing dementia, and the earliest MMSE score were important contributors to prediction of MMSE decline.
  - Common AD comorbidities[12], such as major depressive disorder and bipolar disorder, did not contribute meaningfully to prediction of MMSE score change.
- Initial modeling showed best performance with Random Forest (Table 3), with a ROC AUC of 0.69.
  - Incremental addition of XGBoost-identified features resulted in model performance increase, based on ROC AUC, for Decision Tree Regression, Linear Discriminant Analysis, Logistic Regression, and K-Nearest Neighbor (Table 3, Figure 4).
    - Model gains were on the order of 2-6%. Incremental benefit of feature engineering was not impactful in ensemble approaches.
- Multi-collinearity Assessment
  - Correlation matrices for baseline and incremental features showed little evidence of multi-collinearity on visual inspection (Figure 3a).
  - Incorporation of frequency-based features, however, markedly increased the number of correlated features (Figure 3b).
  - Grouping of highly correlated features (≥ 0.7), however, markedly reduced apparent multi-collinearity (Figure 3c).

Limitations

Study Limitations

- Data on MMSE score change was available for a limited patient sample, and may not be representative of general population of patients with AD.
- Ability to assess MMSE decline was limited by sparseness of assessments and low re-assessment rates after AD diagnosis. The resultant predictive model is only applicable to the population of patients represented by the data used to train the model; therefore, generalization may be limited.
- While top predictors were intuitive (dementia diagnosis, treatment initiation, etc.), they could also be seen to imply easier prediction in patients with closer management of their AD. Incorporation of time- and interval- based attributes may help to surface this association more explicitly, by allowing the modeling of features such as time-to-diagnosis, time-to-treatment, and the timing and frequency of assessments.

Conclusions

- Despite overall sparseness with respect to frequency and timing of MMSE scores, predictive modeling approaches were able to yield ROC AUC as high as 0.69.
- Systematic automated feature selection also showed potential to improve predictive ability in a scalable fashion. If larger amounts of data can be added, we expect these models to become even more accurate, rendering explanations of feature importance more valuable.
- Next steps include further refinement of feature engineering to include time- and interval-based attributes, and grouping of similar attributes to reduce multi-collinearity from frequency-based attributes.

References

1. Basch E. NEJM 2017; 376(2):105-108.
2. Folstein MF, et al. J Psychiatr Res. 1975; 12(3): 189-198.
3. Doody RS, et al. Arch Neurol. 2001;58(3):449-454.
4. Ozery-Flato M, et al. Stud Health Technol Inform. 2017;235:181-185.
5. HealthVerity-Overview. https://healthverity.com/wp-content/uploads/HealthVerity-Overview.pdf Accessed October 18, 2019.
6. HealthVerity-Marketplace. https://healthverity.com/wp-content/uploads/HealthVerity-Marketplace.pdf Accessed October 18, 2019.
7. Tombaugh TN, et al. J Am Geriatr Soc. 1992;40(9):922-935.
8. Machine Learning in Python. https://scikit-learn.org/stable/index.html.
9. XGBoost Documentation. https://xgboost.readthedocs.io/en/latest/index.html.
10. StandardScaler. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html.
11. Feature ranking with recursive feature elimination. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html.
12. Garcez ML, et al. An Acad Bras Cienc. 2015;87(2 Suppl):1461-1473.

Disclosures

RB, JR, AS, ER, RA, AG, and BL are employees of Medidata Solutions. MB provided consultancy to Medidata Solutions.

Presented at ISPOR Europe 2019, 2-6 November 2019

Copenhagen, Denmark

QR CODE