AN EHR-BASED COHORT DISCOVERY TOOL FOR IDENTIFYING PROBABLE AD

Donna Tjandra¹, Raymond Migrino², Bruno Giordani³, and Jenna Wiens¹

¹Division of Computer Science and Engineering, University of Michigan, Ann Arbor, MI
²Phoenix Veterans Affairs Health Care System, Phoenix, AZ
³Neuropsychology Section, Department of Psychiatry, University of Michigan, Ann Arbor, MI

EHR-BASED CRITERIA CAN BE USED TO EFFICIENTLY IDENTIFY PATIENTS WITH PROBABLE AD; THESE LABELS CAN BE USED FOR DOWNSTREAM ANALYSES, e.g., MODELING PATIENT TRAJECTORIES OF DISEASE PROGRESSION

Background

Electronic Health Records (EHRs) Contain longitudinal clinical data pertaining to individuals perhaps decades before AD onset

Our goal → Develop an EHR-based cohort discovery tool to identify patients with probable AD

Why?
- Facilitate downstream analysis involving EHR data
- Explore new sources of data in AD research
- ICD billing codes alone may not be reliable

Study Cohorts
- Michigan Medicine’s Research Data Warehouse (RDW): unlabeled EHR data pertaining to 4,253,944 patients from 2000-2019
- Michigan Alzheimer’s Disease Research Center (MADRC) data: contains consensus-based (ground truth) diagnoses for 789 volunteers
- Evaluation cohort: Individuals in RDW and MADRC, 65 and older (n = 624)

Results

Agreement between EHR-based definitions and consensus-based diagnosis

<table>
<thead>
<tr>
<th>Value</th>
<th>Adjusted PPV</th>
<th>PPV</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- A diagnostic code for AD in combination with a visit of moderate to high complexity medical decision making was the best criteria
- This criteria identified patients within RDW on average 3 years before their MADRC diagnosis

Identification Rules
- different EHR-based definitions of probable AD were compared
- definitions consisted of one or multiple components

<table>
<thead>
<tr>
<th>Component</th>
<th>Component Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD Diagnoses</td>
<td>331.0 (ICD9), G30 (ICD10)</td>
</tr>
<tr>
<td>Medications</td>
<td>donepezil, memantine, galantamine, rivastigmine</td>
</tr>
<tr>
<td>Psychological Testing</td>
<td>CPT 96101-96127</td>
</tr>
<tr>
<td>Complex Med. Decisions</td>
<td>CPT 99214, 99215, 99354, 99355</td>
</tr>
</tbody>
</table>

Rule Evaluation
- EHR-based rules were used to label all encounters in RDW
- EHR-based labels were compared to consensus based diagnosis from MADRC (blue circles indicate encounter-level true positives)
- One encounter-level true positive was required for a patient-level true positive

Evaluation Metrics
- population adjusted positive predictive value (PPV)
- PPV: fraction correctly identified, among those identified as AD
- specificity: fraction of non-probable AD patients not identified
- sensitivity: fraction of probable AD patients identified

Approach

Identification Rules
- different EHR-based definitions of probable AD were compared
- definitions consisted of one or multiple components
An EHR-Based Patient Risk Stratification Tool for Probable AD

Donna Tjandra¹, Raymond Migrino², Bruno Giordani³, and Jenna Wiens¹

¹Division of Computer Science and Engineering, University of Michigan, Ann Arbor, MI
²Phoenix Veterans Affairs Health Care System, Phoenix, AZ
³Neuropsychology Section, Department of Psychiatry, University of Michigan, Ann Arbor, MI

Background

Electronic Health Records (EHRs) Contain longitudinal clinical data pertaining to individuals perhaps decades before AD onset

Our goal → Develop a model to predict the onset of AD 10 years in advance using EHR data

Why?
• Observable symptoms of AD occur up to 20 years after physiological changes in the brain begin
• Mining EHR data may aid in early detection

Study Cohorts
• Michigan Medicine’s Research Data Warehouse (RDW): unlabeled EHR data pertaining to 4,263,944 patients from 2000-2019
• Patients aligned between 68-72 years to control for age
• Probable AD labeled using a previously validated cohort discovery tool

Results

Ability to Predict on Large Horizons

• Predictions on large horizons were possible, though more difficult

Predictive Power of EHR Components

• Including longitudinal data from a variety of fields works best
• Important features agree with the literature

Approach

Inclusion/Exclusion Criteria

Model Training

• Linear model with cross entropy loss
• 80%-20% train-test split, feature breakdown below

Evaluation Metrics

EHR DATA CAN BE USED TO PREDICT THE ONSET OF PROBABLE AD 10 YEARS BEFORE CLINICAL ONSET; MINING ROUTINELY COLLECTED DATA COULD SHED LIGHT ON AD PROGRESSION, ESPECIALLY IN THE DECADES PRECEDING CLINICAL ONSET